# CSH Cold Spring Harbor Laboratory
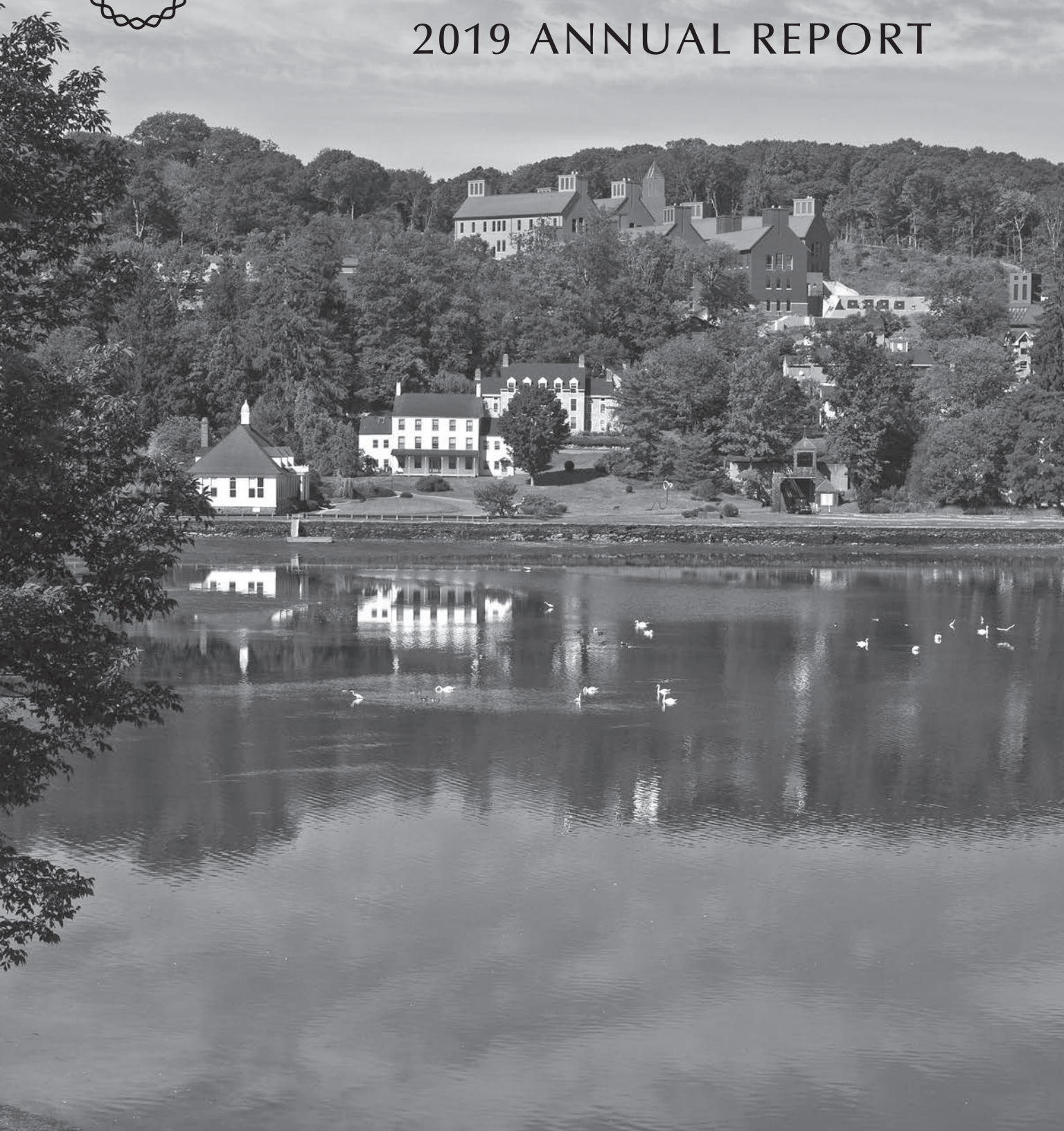
## 2019 ANNUAL REPORT

Cold Spring Harbor Laboratory

# 2019 ANNUAL REPORT

**ANNUAL REPORT 2019**
*© 2020 by Cold Spring Harbor Laboratory*

# Contents

# Officers of the Corporation

# Board of Trustees

## Individual Trustees

# CSHL Scientific Advisory Council

The Scientific Advisory Council (SAC) is an external advisory group that advises the senior management of Cold Spring Harbor Laboratory (CSHL) on matters pertaining to science (both current and future), including the development of a research strategy to maintain CSHL as a world leader. The SAC is a nine-member Council including a Chair of Council who is an individual known for scientific breadth and a detailed understanding of research management at the senior management level. The other eight members are world leaders in their respective fields and as such are able to provide advice on the different research areas of the Laboratory.

**Frederick W. Alt, Ph.D. (Chair)**
*Harvard University*

**Angelika Amon, Ph.D.**
*Massachusetts Institute of Technology*

**Joanne Chory, Ph.D.**
*The Salk Institute for Biological Studies*

**Carol W. Greider, Ph.D.**
*The Johns Hopkins University School of Medicine*

**Eve Marder, Ph.D.**
*Brandeis University*

**Markus Meister, Ph.D.**
*California Institute of Technology*

**Steven Salzberg, Ph.D.**
*The Johns Hopkins University School of Medicine*

**Kevan M. Shokat, Ph.D.**
*University of California, San Francisco; Berkeley*

**Max S. Wicha, M.D.**
*University of Michigan*

# Governance

The Laboratory is governed by a Board of Trustees of up to 35 members that meets three or four times a year. Authority to act for the Board of Trustees between meetings is vested in the Executive Committee of the Board of Trustees. The Executive Committee is composed of the Officers of the Board and any other members who may be elected to the Executive Committee by the Board of Trustees. Additional standing and ad hoc committees are appointed by the Board of Trustees to provide guidance and advice in specific areas of the Laboratory's operations.

Representation on the Board of Trustees itself is divided between business and community leaders and scientists from major educational and research institutions.

The Laboratory is chartered as an educational and research institution by the Board of Regents of the Education Department of the State of New York. It is authorized to operate a graduate program under the name "Cold Spring Harbor Laboratory, Watson School of Biological Sciences" and thereat to confer the degrees of Doctor of Philosophy (Ph.D.), Master of Science (M.S.), and Doctor of Science (Sc.D.), Honorary.

It is designated as a "public charity" under Section 501(c)(3) of the Internal Revenue Code.

# Committees of the Board 2019

# David Hamilton Koch (1940–2019)

David Hamilton Koch, who died on August 23, 2019, was a benefactor and Trustee of Cold Spring Harbor Laboratory from 1991 to 1998, and then he became an Honorary Trustee until his death. He received the Laboratory's prestigious Double Helix Medal in 2007.

David was born on August 3, 1940, in Wichita, Kansas, one of the four sons of Fred and Mary Koch. Fred was an engineer who made a fortune in the 1920s and 1930s developing equipment for refining crude oil and building refineries. By all accounts, Fred was a disciplinarian who fostered competition between his sons, which may account for the conflicts that later roiled the family. After schooling at Deerfield Academy, David followed in his father's footsteps, earning an engineering degree at the Massachusetts Institute of Technology in 1962 and a master's degree in 1963. Among other accomplishments at MIT, David played basketball and set the points record for a single game with a score of 41—he was 6 feet 5 inches and an imposing figure.

Fred died in 1967, and his sons inherited what was already a substantial enterprise. David's brother Charles took on their father's mantle, becoming chairman and chief executive officer, and it was his acumen and enterprise that led what became Koch Industries to expand far beyond the oil business. David joined his brother, and by 1981, he was Executive Vice President, second-in-command to Charles. The company grew to an annual revenue of $100 billion and the wealth of each brother was estimated to be more than $50 billion, placing them among the wealthiest individuals in the world.

David and Charles used their wealth to promote Libertarian polices and support conservative organizations and think tanks such as the Cato Institute and the American Legislative Exchange Council. David ran as the Vice Presidential candidate for the Libertarian party in 1980.

In 1991, David survived a disastrous airplane crash on a runway at Los Angeles International Airport that killed 35 people. This changed his life and resulted in David becoming one of the greatest philanthropists in New York. In 1994, David married Julia Flesher, and they had three children, David, Jr., Mary Julia, and John Mark. Unlike his brother Charles, who shunned publicity, David was well known in society circles in New York City, Aspen, the Hamptons, and Palm Beach, where he had houses. He was a great patron of the arts, including the Metropolitan Museum of Art, and he was passionate about ballet and supported the New York City Ballet. New York State Theater at Lincoln Center, home of the New York City Ballet, was named the David H. Koch Theater in his honor and in recognition of his substantial support. David also supported other academic institutes, including the American Museum of Natural History, Rockefeller University, the Hospital for Special Surgery and the Smithsonian, as well as his alma mater, Deerfield Academy and MIT, providing funds for the Koch Biology Building that housed MIT's expanded biology department.

In 1992, David was diagnosed with prostate cancer, given only a short time to live—but he was treated successfully at Memorial Sloan Kettering Cancer Center (MSKCC). This led him to make contributions to institutes concerned with cancer treatments and research, including MSKCC, the M.D. Anderson Cancer Center in Houston, Johns Hopkins School of Medicine, and New York-Presbyterian Hospital Weill Cornell Medical School, in addition to Cold Spring Harbor Laboratory. He created the David H. Koch Center for Cancer Care at MSKCC, providing outpatient care for cancer patients, and the David H. Koch Institute for Integrative Cancer Research at MIT. There, the MIT Cancer Center implemented a combined approach to cancer research by converging life sciences with engineering, and David continued to support cancer research there for many years.

When meeting with David at his substantial office on East 61st St. in New York or at his wonderful beachfront home in Southampton, one always had to be prepared to inform him of the latest advances in cancer research without sugarcoating the progress. He always asked pointed questions and related answers to what he had learned from others who advised and interacted with him. Thus, David was a very knowledgeable philanthropist when it came to medicine and science, and this was almost certainly also true for the arts.

Here at Cold Spring Harbor, his significant gifts were for a graduate student fellowship in the Cold Spring Harbor School of Biological Sciences, for the renovation of the Demerec Laboratory that now houses a new program in cancer research focused on whole-body physiology and cancer, and for the Preclinical Therapeutics Shared Research Resource at our Woodbury Research Campus. He also provided major support for a new research building on the main campus that bears his name and houses both cancer research and the CSHL program in quantitative biology that merges computer science with research in the life sciences. David was a long-term supporter of many other initiatives at CSHL, including our Child Care Center, our prostate cancer research, and the CSHL Annual Fund that provides much needed support for our operations.

David was introduced to CSHL through his mentor, Ralph Landau, who was also a CSHL Trustee, and his connection was solidified by David's long-term close friend and CSHL supporter Roger Samet. Discussions with David and Roger were always most enjoyable because of their very diverse and broad interests in music, the arts, and science. We very much valued David's contributions to Cold Spring Harbor Laboratory and his friendship that spanned many years. To Julia, her children, and the extended Koch family, and to his friend Roger, we extend our condolences and thanks for sharing David with us.

**Bruce Stillman**
*Cold Spring Harbor Laboratory*

# Joseph F. Sambrook (1939–2019)

It is impossible to measure just how big an impact Joseph Sambrook had on the field of molecular biology. For almost 40 years, well-loved, dog-eared copies of his influential book, *Molecular Cloning: A Laboratory Manual*, have been permanent fixtures in nearly every biology laboratory around the world. Colloquially known as the "cloning bible," its study has been a rite of passage for many researchers.

Sambrook was a true pioneer in a golden era for molecular biology. A proud Scouser, he completed his Bachelors of Science with first-class Honours in 1962 at the University of Liverpool in the United Kingdom. As circumstance would have it, distinguished Australian virologist Frank Fenner had at the same time embarked on a U.K. lecture tour where he would encounter Sambrook on a visit to Liverpool. Sambrook, who even at a young age was unintimidated by status, asked a series of challenging questions following Fenner's lecture. So impressed was Fenner that he offered Sambrook then and there a graduate Ph.D. scholarship to join his laboratory at the Australian National University (ANU) in Canberra. This would prove to be not only the beginning of a stellar career that would span more than 50 years, but also of a lifelong affair with Australia, where he would come to live out his days. In his Ph.D. thesis titled "The genetics of animal viruses," Sambrook isolated and characterized conditional lethal mutants, including temperature-sensitive and host-range mutants, in mammalian poxviruses.

Upon completion of his Ph.D., in 1966 Sambrook returned to England, where he trained as a postdoctoral researcher with future Nobel Prize laureate Sydney Brenner at the Medical Research Council Laboratory of Molecular Biology in Cambridge. When he arrived, the genetic code had been completely solved, except for one codon—UGA. Within just a couple of years, Sambrook had determined that the function of UGA in *Escherichia coli* was in fact not to encode an amino acid; rather, that it served as a chain-terminating codon. With his findings published in *Nature* in 1967, the genetic code was finally complete, helping to put Sambrook well and truly on the map as one of the most promising young molecular biologists of the time.

Sambrook started developing a keen interest in the role that viral genes can play in human disease, including cancer. In the late 1960s, the key to understanding the molecular basis of human cancer was, in the minds of many, by studying animal viruses proven to oncogenically transform cultured animal cells. Renato Dulbecco had recently shown that simian virus 40 (SV40) malignantly transforms cultured mammalian cells, work for which he would later be awarded the Nobel Prize. In 1967 Sambrook uprooted once again, this time to sunny California, to join Dulbecco at the Salk Institute. As a Junior Fellow working in Dulbecco's laboratory, Sambrook

demonstrated that transformed cells carried several copies of SV40 DNA and, importantly, that the virus DNA was integrated into the genome of the host cell.

At around this time James Watson had taken over as Director of CSHL and was anxious to expand its work to include mammalian viruses. Following a chance meeting at Salk, Watson recruited Sambrook to CSHL to establish a Tumor Virus Group. It was here that Sambrook really started to make his mark, and he would later quip that he was recruited for three years and stayed for 16! Between 1969 and 1985 Sambrook's group grew steadily in size, eventually consisting of 60 scientists making up a major force in the study of tumor-associated viral genetics. A primary goal of the Tumor Virus Group was the identification, mapping, and analysis of the major genes of SV40 and adenoviruses. This was no small undertaking, especially considering that the National Institutes of Health (NIH) had at the time imposed a moratorium on cloning genes suspected of having oncogenic potential! CSHL lacked the required containment facility to perform such studies, so a determined Sambrook would

1973

make frequent expeditions to London's Imperial Cancer Research Fund (now part of Cancer Research UK), which had agreed to make its containment facility available to him.

Sambrook's other achievements during his time at CSHL include the first use of restriction enzyme polymorphisms as genetic markers, elucidation of the mechanism of integration of SV40 and adenoviruses, and deciphering the patterns of transcription and posttranscriptional processing of mRNA. He was also the first to exploit ethidium bromide on agarose gels to stain DNA fragments generated by cleavage with newly discovered restriction enzymes.

It was also while at CSHL that the seed was sown for the legendary *Molecular Cloning* manual. Sambrook, with colleagues Edward Fritsch and Tom Maniatis, had started a "how to clone" course to teach the basics of molecular biology to their visiting scientists. Joe was a gifted and enthusiastic writer. He commandeered in turn Maniatis, Fritsch, David Russell, and Michael Green to produce, through four iterations, the nucleic acid–centric instruction manual *Molecular Cloning*. These volumes served as a foundation stone for the growth of CSHL Press, and remain to this day the most popular and influential laboratory manuals ever produced.

In 1977, Sambrook was appointed Assistant Director of CSHL. Sambrook's personal style of leadership was renowned, as was his increasingly formidable scientific reputation. He was fiercely combative with anyone prepared to engage in intellectual jousting. His nose-to-nose style could be quite confrontational and intimidating, particularly to newcomers, although at heart everyone suspected it was good-natured and just Joe being Joe! Indeed, behind his sometimes harsh bark was an unquestionable intent to help others. Always generous with his time, Sambrook would give intense thought to a researcher's work and provide valuable updates and encouragement to keep going. He did this with good humor and a twinkle in his eye, as he loved to surprise. Sambrook taught his postdocs to immerse themselves in their projects, to avoid the trap of following the pack, and to believe in themselves. Such was his impact at CSHL that when he left in 1985, a newly constructed research building was named the Joseph F. Sambrook Laboratory in appreciation of his contributions to the institution.

In 1985, with his wife-to-be Mary-Jane Gething—also a renowned senior scientist—Sambrook was recruited to the University of Texas Southwestern Medical School, he as Chairman of Biochemistry and Gething as its first Howard Hughes Medical Institute Investigator. Building on work started while Gething was still at the Imperial Cancer Research Fund in London, and continued at CSHL, Sambrook and Gething used viral vectors to express cloned influenza virus hemagglutinin and other membrane glycoproteins that would lead to important contributions in the understanding of intracellular traffic and protein folding. Led by Gething, they also devised

and patented a range of mutants of the thrombolytic agent, tissue plasminogen activator (tPA), which they had cloned while at CSHL. These mutants were better tolerated in patients and offered enhanced clinical benefit compared to wild-type tPA, leading to the design and synthesis of a new generation of tissue-type plasminogen activators that are still in clinical use today.

In 1991 Sambrook became Director of the Eugene McDermott Center for Human Growth and Development at UT Southwestern, before he was recruited back to Australia in late 1994 to join the Peter MacCallum Cancer Centre in Melbourne as the Director of Research. In just five short years Sambrook completely revolutionized the research enterprise at Peter Mac, instituting an unflinching requirement for research excellence and a culture of collegiality that permeated the organisation. Through astute recruitment he almost tripled Peter Mac's laboratory research workforce from 75 to more than 200, and set in place a research direction and strategy that grew in size, scope, and impact under his dynamic direction. Sambrook founded and directed the Kathleen Cunningham Foundation Consortium for research into familial breast cancer (KConFab), the first ever clinical cohort of its kind. It brought together experts from across clinical streams to solve pressing clinical, genetic and epidemiological problems of familial breast/ovarian cancer—a project that he was deeply passionate about. He was also one of two founding directors of Cancer 2015, a multiyear, large-scale program funded by the Victorian Government on personalized treatment of cancer.

Among his many honors, Sambrook was an Elected Fellow of the Royal Society (1985), an Elected Fellow of the Australian Academy of Science (1999), a Peter Mac Distinguished Fellow, and the recipient of Honorary Doctorates from the University of Liverpool and the Watson School of Biological Sciences at CSHL.

But Sambrook's legacy extends far beyond his personal achievements. Not only was he a giant in the field of molecular biology, with his wife and long-time collaborator Gething, he and his team were true trailblazers in their characterization of protein trafficking and the protein folding response, studies that formed the foundations for a whole new field of biology. Sambrook took great pride in seeing those he mentored achieve international recognition, and those who worked with him feel very privileged to have been part of his exceptional life.

After a long illness, Joe died peacefully at home on June 14th with Mary-Jane and their daughter Honor at his side.

He also leaves three children by his first marriage to Thelma, eight grandchildren, and countless scientists who were blessed by falling in the shadow of this giant of molecular biology.

Consistent with his lifelong love of music, Sambrook with Gething established the ongoing Joseph Sambrook Opera Scholarships using some of their royalties from the sales of tPA. On July 14th two of the recipients of the scholarships sang at a "Concert for Joe," a fitting send-off in lieu of the traditional funeral that he expressly did not want.

**Ashley Dunn**
*University of Melbourne, Parkville, Victoria, Australia*

**Ricky Johnstone**
*Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia*

**Bruce Stillman**
*Cold Spring Harbor Laboratory*

# PRESIDENT'S REPORT

Last December, I attended a joint meeting of the American Society of Cell Biology (ASCB) and the European Molecular Biology Organization (EMBO) in Washington, D.C., to deliver the keynote address. One of the best things about scientific meetings is the opportunity to encounter ideas to which you might not otherwise be exposed, and I always make it a priority to attend talks beyond my own fields of expertise. In this case, a highlight of the meeting for me was a talk about the importance of active learning in STEM education given by Mary Pat Wenderoth, a biology lecturer at the University of Washington and founder of the Society for the Advancement of Biology Education Research.

Wenderoth, who is known not only for her dynamic teaching in the classroom but also for her rigorous research on biology education itself, was being honored with ASCB's Bruce Alberts Award for Excellence in Science Education. Bruce Alberts—a biochemist at the University of California, San Francisco, who was president of the National Academy of Sciences from 1993 to 2005 and editor-in-chief of *Science* from 2008 to 2013—has long been a tireless advocate for science education reforms, arguing for changes that will better inspire students and prepare them to think critically about the world. In a 2008 *Science* editorial, he wrote, "Rather than learning how to think scientifically, students are generally being told about science and asked to remember facts. This disturbing situation must be corrected if science education is to have any hope of taking its proper place as an essential part of the education of students everywhere" (Alberts 2009).

## What Is Active Learning?

Active learning, Wenderoth told her audience, achieves much more than the simple transmission of facts that occurs in a traditional, lecture-style class. The approach focuses on engaging students directly in the learning process, encouraging them to analyze, synthesize, and evaluate information, rather than merely committing it to memory long enough to pass an exam. And the techniques she talked about can be employed even in large lecture halls: Rather than simply telling students how something works, teachers can use brainstorming sessions, worksheets, group problem-solving, or personal response systems to guide students through the material.

For those of us who know how vital and satisfying real science can be, it is fairly intuitive that throwing lots of complicated information at kids with little context about how it fits into their lives is not the way to share that excitement or for them to learn. That was the type of teaching I encountered as a high school student. It is why in some early years I received C's in science while gaining A's in classes like history and mathematics, where I was challenged to gain, synthesize, and interpret information and think for myself. Wenderoth and her colleagues have data to support this latter type of learning. They have analyzed hundreds of studies on STEM education and found that incorporating active learning strategies into teaching measurably boosts test scores and reduces failure rates in undergraduate science, engineering, and mathematics courses. They found that nearly 34% of students failed STEM courses that relied solely on traditional lectures, whereas the failure rate for STEM classes that incorporated active learning was ~22% (Freeman et al. 2014).

## The DNA Learning Center Approach

If we take an evidence-based approach to science education, it is clear that classrooms have to move beyond the traditional lecture, and CSHL has a long-standing commitment to bringing active learning and hands-on inquiry to science students, both here on Long Island and around the

world. Although I was drawn to Wenderoth's lecture as an opportunity to learn about something beyond my field, many people at CSHL do have deep expertise in science education, particularly those who run our DNA Learning Center (DNALC).

Executive Director David Micklos has steered the DNALC's development since it was established in 1988, and the educational program he conceived has now impacted millions of students worldwide. Our lab field trip program, summer camps, and online resources engage middle school, high school, and college students in active learning experiences that deepen their understanding of biology and the process of scientific inquiry.

Here on Long Island, the DNALC engages about 30,000 students in hands-on genetics experiments every year. Middle school and high school classes visit our eight teaching laboratories for experiences that supplement what the students are learning in class, and DNALC staff visit schools to extend our reach. During the summer, more than 1300 middle and high school students attend week-long camps, where they can explore topics ranging from the fundamentals of cell biology to the power of genomics. These programs use many of the approaches and techniques CSHL faculty and research staff employ in their own research.

At the time of writing this essay, we are in the midst of the COVID-19 pandemic, and most of us are sequestered while working and/or learning from home, because most business, school, and university facilities are closed. True to their entrepreneurial bent, DNALC educators rapidly established online labs for students to access daily and on-demand so that they could continue to learn and perform experiments on DNA using materials found in their homes. This type of in-home active learning has been accessed by thousands already and may well survive the current pandemic era.

## Expanding Access to DNALC Programs

Our newest learning center facility, DNALC NYC at City Tech, currently operates out of a temporary space on the campus of the New York City College of Technology in downtown Brooklyn. In 2021, we will open a new permanent 17,500 square-foot facility, a completely renovated space with six fully equipped teaching labs, two bioinformatics labs, and an interactive exhibit for visitors. This facility is twice as large as the one we have here in Cold Spring Harbor. It will serve a diverse population of New York City middle school and high school students, including many from groups that are underrepresented in the sciences, as well as college students from the City University of New York.

The DNALC's programs are meant for everybody, not just those students who have already excelled in or expressed deep interest in science. We live in an era where personal genomics and precision medicine are becoming more prevalent. Genetics helps doctors identify the best treatment options for patients, and companies like Ancestry and 23andMe share DNA sequencing results directly with consumers—so an understanding of genetics is increasingly relevant to decisions we must make about our health. It is crucial that kids who have never really thought about biology, who might not even know what DNA is, can visit the DNALC (in person and/or online) and suddenly find themselves amplifying and analyzing their own DNA. Seeing their own DNA glow on an electrophoresis gel is the kind of thing that gets most kids really excited. Before they know it, they're actually learning what genetics really is.

In fact, there are a variety of opportunities for students to visualize and analyze their own DNA through the DNALC. One program that is being taught on Long Island and in classrooms worldwide focuses on the genetics that underlies our sense of taste—specifically, variations among individuals' ability to taste a bitter compound called phenylthiocarbamide (PTC).

Variations within the *TAS2R38* gene, which encodes a taste cell receptor that recognizes PTC, influence humans' ability to taste this compound. Depending primarily on three nucleotide

positions within the gene, a person can perceive PTC as mildly or intensely bitter or be unable to taste it at all. These genetics were worked out in the 1930s largely at CSHL, by Albert Blakeslee at the Carnegie Department of Genetics (the forerunner of the contemporary CSHL). Blakeslee showed that the inability to taste PTC is a recessive genetic trait—that is, people who cannot taste PTC inherit mutations from both parents and are thus homozygous. After the gene for the taste receptor was identified in 2003 by Dennis Drayna and colleagues at the National Institutes of Health, the variations, or single-nucleotide polymorphisms (SNPs), in the gene correlated with PTC-tasting ability.

PTC tasting strips are commonly distributed in science classes to help students learn a bit about dominant and recessive traits as they compare their reaction to the compound to those of their classmates. Students who participate in the program developed at the DNALC, however, get a much richer experience. DNALC students are presented with an open-ended experiment in which they make and test their own hypotheses. Prior to encountering the compound itself, DNALC students isolate and analyze DNA from their own cheek cells, using PCR to amplify a short region of the *TAS2R38* gene and finding out whether it can be cut by a restriction enzyme whose recognition sequence includes one of the SNPs. Once they have obtained this genetic information, the students use it to predict their tasting ability. Then the students taste the PTC paper.

This is one of the DNALC's most popular experiments. We have made it available to teachers as a kit sold through Carolina Biological Supply. Like the DNALC's other program modules, it was developed to closely align with the New York State curriculum. It gives high school students real, personal insight into Mendelian genetics and—because through genotyping they accurately predict only about 90% of PTC tasting ability—some of the complexities of genetic testing. It is also a gateway to learning about concepts like precision medicine, which uses genotypes to predict how patients will respond to drugs or other therapies, and even precision agriculture, in which genetics can help farmers identify crops that are most likely to thrive under particular conditions.

## When Student Questions Drive the Program

The DNALC has also pioneered another program that goes still further, fully immersing entire classrooms of students in authentic, open-ended research. Analyses by the University of Texas at Austin and other universities have indicated that authentic research experiences increase both on-time graduation rates and retention in the sciences by 20%. It is unfortunate that traditionally opportunities to do research have been limited to only a fraction of undergraduate biology students, who work one-on-one with faculty members or other mentors. The DNALC's DNA Barcoding Projects broaden access to research experiences by enabling an entire class of students to employ a common set of techniques to address a wide range of questions.

The Barcoding Projects are built around a method for identifying species of plants, animals, bacteria, and fungi (including their viruses) that uses a short sequence of DNA, much like a UPC code uniquely identifies a product. DNA barcoding can be used to survey the wildlife in a local park, test food products for authenticity, or detect the misuse of an endangered species. It is up to the students to decide what they want to investigate with the method.

This is real research. Students come up with their own questions, use sophisticated tools to answer them, including online computational resources and DNA sequencing provided by the DNALC, and contribute new knowledge to the scientific community. Past teams have verified the identity of a fish species from a neighborhood market that shoppers had suspected was fraudulently replaced by a cheaper fish; discovered that supplements marketed in New York herbal medicine stores as *Ginkgo biloba* contained no trace of *Ginkgo* DNA; and identified ant and mosquito species well north of their expected range, suggesting climate change–driven shifts in distribution. Some students are even using a related, data-intensive technique called metabarcoding to identify

not just a single species, but to catalog and compare full sets of microbes or other organisms within environmental samples.

DNA barcoding projects are now offered as a semester-long course on college campuses worldwide. The DNALC has developed a workflow that makes the sophisticated analysis affordable and efficient for classroom use. Once students have collected and isolated DNA from their chosen environment, teachers can ship it off for sequencing on a Monday, and students can begin analyzing the results by Wednesday using an online bioinformatics tool called DNA Subway that was built by the DNALC team. Last year, about 200,000 people used DNA Subway to compare sample sequences to those in a DNA database. Most were students, but this tool is so sophisticated and user-friendly that many instructors have begun using it for their own research, too.

The DNA Barcoding Projects, pioneered at the DNALC in 2011, have become one of the largest systems for course-based undergraduate research in STEM education. Our partners use this program mostly at the college level. It is reaching many students—particularly in their first year, when they have a higher chance of being influenced to remain in science. At James Madison University (JMU) in Virginia, every student who takes freshman biology does the DNA Barcoding Project, and they have already documented hundreds of species of plants, fungi, and invertebrates in their local arboretum. Faculty at JMU have found that the project creates opportunities to cover most of the key topics that would be covered in a traditional introductory biology course, and students not only learn fundamental concepts about genetics, evolution, ecology, and the molecules of life, but they have to integrate and apply those concepts to their own work.

At the DNALC, Urban Barcoding programs are drawing students into research even earlier. These programs target high school students throughout Long Island and New York City, most of whom have never done research before. Students work with their DNALC-trained teachers or mentors at collaborating institutions to document and explore local biodiversity, designing and carrying out the same open-ended type of investigations. Participants in these programs are true citizen scientists, contributing data that will help researchers and policymakers monitor the impacts of climate change and assess local ecosystems.

This form of active learning goes far beyond what most science students will experience in middle school, high school, or even college classrooms. The DNALC is really pushing the envelope of hands-on inquiry and research experiences for kids, especially early on. We want everyone to experience the excitement of discovery, and to know how to use scientific methods and think critically, so they will grow up with an understanding of how science is done and how it fits into their lives.

**Bruce Stillman, Ph.D., F.R.S.**
*President and Chief Executive Officer*

**REFERENCES**

Alberts B. 2009. Redefining science education. *Science* **323:** 437.
Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP. 2014. Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci* **111:** 8410–8415.

# Highlights of the Year

## Research

Hundreds of scientists working in Cold Spring Harbor Laboratory's 50-plus laboratory groups contributed to research that in 2019 was published in the world's major research journals. Their efforts reflect the full spectrum of this institution's scientific activity in cancer, neuroscience, plant biology, quantitative biology, and genomics. The following is a sampling of this year's important findings.

### Machine Learning Better Interprets Gene Regulation

Machine learning algorithms are helping biologists make sense of molecular signals that control how genes function, but algorithms developed to analyze larger and larger sets of data become increasingly complex and difficult to interpret. Associate Professor Justin B. Kinney has an approach to design advanced machine learning algorithms that are easier for biologists to understand.



J. Kinney

The algorithms are a type of artificial neural network (ANN). Inspired by the way neurons connect and branch in the brain, ANNs are the computational foundations for advanced machine learning. Biologists use ANNs to analyze data from an experimental method called a "massively parallel reporter assay" (MPRA), which investigates DNA. Using this data, quantitative biologists can make ANNs that predict which molecules control specific genes in a process called gene regulation.

Unfortunately, the way standard ANNs are shaped from MPRA data is very different from how scientists ask questions in the life sciences. Kinney's approach bridges this gap between computational tools and how biologists think. These custom ANNs mathematically reflect common concepts in biology concerning genes and the molecules that control them, forcing data to be processed in a way that a biologist can understand. Kinney's laboratory is now investigating a wide variety of biological systems, including key gene circuits involved in human disease.

### One Gene Balances Plant Growth versus Immunity

With an unpredictable array of bacteria, fungi, and viruses in the soil and air, a plant needs a robust immune system. But energy spent on pathogen defense cannot be used to grow taller or produce seeds. In the wild, the trade-off is crucial. In a tended field, however, crop plants face fewer threats. One way to boost the productivity of a plant is to redirect some of its resources away from maintaining an overprepared immune system and into enhanced seed production.

As reported in the *Proceedings of the National Academy of Sciences*, Professor David Jackson and his team identified a gene in corn that contributes to both the plant's development and to the control of its immune system. Manipulating this gene could be a way to increase crop yields by reprogramming how a plant balances its investments in growth and defense. Most corn plants cannot survive without the gene, which is called *Gβ* (pronounced GEE beta) and encodes part of an essential signaling complex. Observing that seedlings lacking *Gβ* quickly turn brown and die, the researchers determined the gene helped to keep the plant's immune system in check.



D. Jackson

By experimenting with dozens of genetically diverse lines of corn, the team found that some plants could grow without *Gβ*. Studying these plants, they discovered that *Gβ* impacts the size of a plant's meristems—reservoirs of stem cells from which new growth originates. The team also linked naturally occurring variations in the *Gβ* gene to the production of corn ears with unusually abundant kernels. *Gβ* is involved in both growth and immunity, and it likely mediates cross-talk between the cellular pathways that control these competing functions.

### *Mutant Cells Team to Make a Cancer Deadlier*

A. Krainer

Professor Adrian Krainer and colleagues discovered that two cell mutations, already harmful alone, enhance one another's effects, contributing to the development of the deadly blood cancer acute myeloid leukemia (AML). In *Nature*, they detailed how mutations of the genes *IDH2* and *SRSF2* partner to cause AML.

The presence of the *IDH2* mutation enhances the errors caused by the *SRSF2* mutation, preventing cells in the bone marrow from maturing into the red and white blood cells an AML patient needs to overcome the disease. Prior to this research, the only known similarity between the two mutations was that they are involved in precancerous symptoms. But in many cases, the cause of the symptoms is not the cause of cancer itself. "Just because you see a mutation [in a sick patient's cells] doesn't really show that it's directly contributing to the disease," Krainer said.

The researchers knew that one of the two mutations in question, in the *SRSF2* gene, causes errors in a crucial process called RNA splicing. Splicing converts messages from DNA, in the form of RNA, into readable instructions for a cell. Errors in this process can result in cell malfunction.

Previous research showed that these mutations were present in only 1% of AML patients. But the Krainer laboratory found that this problem is much more common, appearing ~11% of the time in AML patients.

The researchers discovered this by dusting for *SRSF2*'s fingerprints within the gene's actual workplace (RNA), instead of simply searching an entire office block (DNA). Further experiments revealed that the severity of the identified *SRSF2* splicing errors can be enhanced by the presence of a second mutation, in *IDH2*, resulting in even more defective blood cells. This interdependence suggests points of therapeutic intervention.

### *BARseq Builds a Better Brain Map*

A. Zador

Professor Anthony Zador set out a decade ago to map three pillars of brain function: connectivity, gene expression, and physiological activity. His team developed MAPseq, a technique to map the connections of different brain cells and gain a better understanding of how they interact with each other. Improving the technique, Zador's team published BARseq, the next generation of MAPseq, in *Cell*.

The new technology can accurately pinpoint the location of a neuron. BARseq determines not only a neuron's connections, but also its pattern of gene expression and its physiological activity. "We wanted to understand how neurons are connected to one another and relate them to other aspects of neural function, like gene expression and neuronal activity," explained Zador.

The researchers used MAPseq to tag each neuron with a unique barcode composed of genetic sequences. By following those tags across the brain, they can see where the neurons send messages and then map the pathways those signals form between different areas of the brain. But the MAPseq tagging process made it difficult to see how a neuron mediated a particular function such as gene expression. BARseq tags and sequences the neurons in situ, or in their original form and location in the brain, so you can see exactly where the neurons are when the barcodes are sequenced.

The team used BARseq to map the connections of 3,579 neurons in the auditory cortex of the mouse brain. Matching connectivity patterns to gene expression allows scientists to characterize different cell types and define their specific functions in the brain. It could be a valuable tool for studying how neural circuits are formed, providing a foundation for understanding thought, consciousness, decision-making, and how those go awry in neuropsychiatric disorders like autism, schizophrenia, depression, and OCD.

## Detailed New Primate Brain Atlas

Research conducted in Japan with contributing neuroscientists from CSHL resulted in a 3D reconstruction of a marmoset brain, as well as an unprecedented level of detail about neuronal connectivity across the entire brain. The study introduces a new methodology, combining experimental and computational approaches, that helps account for significant variation between individual brains. It allows for synthesizing unique brain connectivity maps into a single reference brain. The resulting data set for the marmoset brain may offer insights into human neural connectivity.

Professor Partha Mitra collaboratively led the study reported in *eLife* as part of Brain/MINDS research conducted at the RIKEN Center for Brain Science in Japan. The brain architecture of marmosets more closely resembles that of humans than does the mouse brain. Although mice are currently the mainstay for modeling human disease, the emergence of marmoset models of human neurological disorders has made marmosets a target of new research.

"Brain connectivity studies have been carried out in the marmoset before," Mitra explains. "But we did not have complete three-dimensional digital data sets, showing connectivity patterns across several entire brains at the light-microscope resolution." With this new data and approach as a basis, neuroscientists are closer to making sense of the complex neural connections in the primate—and human—brain.

P. Mitra

## Novel Approach to MDS Cancer Treatment

CSHL Fellow Lingbo Zhang and colleagues discovered a new drug target for myelodysplastic syndrome (MDS), a common blood cancer with very few treatment options. Published in *Science Translational Medicine*, Zhang's research restores blood cell production in mice genetically engineered to mirror the pathological features of human MDS patients who are resistant to existing treatments.

MDS is sometimes referred to as "bone marrow failure disorder." Bone marrow is designed to produce enough blood for everyday survival. When blood cells are lost via bleeding or when they grow too old to do their job, replacement cells are made and begin to mature. MDS results from those replacements being too few, defective, or both.

L. Zhang

Traditional treatments for MDS symptoms rely on the body's natural ability to make more mature red blood cells, which is driven by a hormone called erythropoietin (EPO). Immature red blood cells developing in the bone marrow must be exposed to EPO to fully mature into red blood cells that can aid the body. Delivering lots of EPO to the bone marrow does not help MDS patients because many do not have enough immature blood cells to begin with.

Zhang and his colleagues looked to an even younger stage of the developing blood cells as a point of intervention. They discovered that activating a specific protein receptor called CHRM4 significantly hampers the maturation of cells responsive to EPO. By blocking this receptor, Zhang and his colleagues were able to restore healthy blood cell production. In mice, this strategy significantly improved survival rates.

## Mice Fidget When Deep in Thought

Almost everyone fidgets, even mice. In *Nature Neuroscience*, Associate Professor Anne Churchland and colleagues observed that the neural activity of mice performing trained tasks indicates that they fidget while making decisions.

Churchland's laboratory investigates the neural circuits that are connected to decision-making. They studied the neural activity all across one part of the brain in a mouse while it was engaged in decision-making tasks. They measured the activity with wide-field

A. Churchland

imaging, like an fMRI (used for mapping brain activity) for a mouse, allowing them to see the activity of neurons across a large portion of the brain all at the same time.

The mice were trained to grab little handles to initiate a trial and lick one way or the other to report their decisions. The scientists expected to see neural activity related to the handle grabbing or with licking. What they saw was that one simple task set off electrical activity across the mouse brain and found that most of that activity was driven by uninstructed movements that the animal was making such as hind-limb movements, pupil dilations, facial movements, nose movements, and whisker movements. "We originally thought the animals were 100% focused on our task—the licking, the grabbing, and the deciding—but it turned out that they had their own set of priorities that involved a lot of movements of all different kinds."

For scientists researching cognition, the results suggest greater attention to discerning signals uniquely related to cognitive processes and signals related to background movements. Researchers will have to work to disentangle the two signals, and this study provides guidelines and computational codes on how to correct for it.

Churchland speculates that movements are more tightly connected to cognition than previously recognized. "Maybe the movements are part of the process of thinking and deciding," she said.

### Quantifying How the Brain Smells



F. Albeanu

To understand how the brain processes and interprets smells, Associate Professor Florin Albeanu and Professor Alexei Koulakov put past odor-classifying models to the test and discovered discrepancies.

Their results, in *Nature Neuroscience*, differ from other published studies that found predictable relations between molecular properties of odors and activity in the early stages of the olfactory system. The new research found that although there were some correlations between some molecular properties of odors and corresponding neuron activity response, they "held little predictive power when new odor pairs or shuffled properties were tested."

Generally, scientists know that odor particles first enter through the nasal cavity, where odorant receptors expressed by olfactory receptor neurons in the sensory tissue bind to them. The olfactory bulb, a structure located in the forebrain of mammals, then processes information from the receptors. Afterwards, the bulb sends out this information to several higher processing brain areas, including the cerebral cortex. There, the olfactory output messages are further analyzed and broadcast across the brain before they are conveyed back to the bulb in a feedback loop.



A. Koulakov

"Rich feedback makes the olfactory system somewhat different from the visual system," Koulakov said. "Olfactory experience is very subjective—perception of smells actually depends on the context, and on an individual's prior experience."

Albeanu and Koulakov suggest that the entry level of olfactory inputs and the further processed bulb outputs reflect different aspects of smell. The unexpected results of their research are an exciting opportunity to build a more comprehensive and testable computational model for the odor space that captures the differences in informational relevance for scent features across the various levels of olfactory processing.

### Architecture of Norovirus Informs Vaccine Development

Noroviruses are a leading cause of food-borne illness outbreaks, accounting for 58% of all outbreaks and causing 685 million cases worldwide each year. There is no effective therapeutic against them. Knowledge of the intricate structure of the outer layer of noroviruses, the capsid, which allows the virus to attach to its human host, is key to vaccine development. In vaccines, specific antibodies recognize the capsids and bind to them so they can no longer interact with human cells.

Professor Leemor Joshua-Tor led a team to solve the high-resolution structures of four different strains of noroviruses using a cryo-electron microscope. This allowed them to see the intricate architecture of virus shells in high definition. Their findings, published in *PNAS*, could help in guiding the development of therapeutics to fight norovirus infection.

L. Joshua-Tor

The team found an unexpected mixture of different shell sizes and shapes. A smaller form consists of just 60 building blocks with 30 surface spikes placed farther apart. Larger shells were made out of 240 building blocks with 120 surface spikes that are lifted significantly above the base of the shell and form a two-layered architecture that could interact differently with human cells. It is the spikes on the shell that interact with the host, and the distance and orientation of the spikes varied across the different strains of noroviruses. This means each strain will interact differently with human cells and the way the antibodies bind is also going to be different. These variations are key to vaccine development.

### Hidden Molecular Pocket Key to Treating Brain Injury

The ideal drug is one that only affects the exact cells and neurons it is designed to treat, without unwanted side effects. Professor Hiro Furukawa revealed a mechanism that could lead to this kind of long-sought specificity for treatments of strokes and seizures.

When the human brain is injured during a stroke, parts of the brain begin to acidify, and this leads to the rampant release of glutamate. The glutamate hits the NMDA receptor, causing it to fire—a lot. In a healthy brain, the NMDA (*N*-methyl-D-aspartate) receptor is responsible for controlling the flow of electrically charged atoms in and out of a neuron. The "firing" of these signals is crucial for learning and memory formation. However, overactive neurons or abnormal NMDA receptor activities have been observed in various neurological diseases and disorders, such as stroke, seizure, depression, and Alzheimer's disease.

H. Furukawa

Furukawa's team looked for a way to prevent overfiring NMDA receptors without affecting normal regions of the brain. Previous work had identified compounds, called the 93-series, suited to this purpose. Eager to join with the NMDA receptor in an acidic environment, these compounds down-regulate the receptor activity, even in the presence of glutamate, thereby preventing excessive neuronal firing.

However, the 93-series compounds sometimes cause the unwanted consequence of inhibiting the NMDA receptors in healthy parts of the brain. The findings of Furukawa and his colleagues that improve on the unique features of the 93-series were detailed in *Nature Communications.*

### Research Faculty

### *Awards*

Described as a "hub of breakthroughs," the 2019 *Nature Index* ranked CSHL the top institution for research output worldwide. Ambition and interdisciplinarity were notable characteristics of smaller institutions like CSHL, which proportionately outstripped larger institutions in the ranking.

The Laboratory's scientists were recognized individually by numerous honors throughout the year.

The Institute for Scientific Information at the Web of Science Group named seven researchers affiliated with CSHL among the scientists producing the top 1% of the most highly cited research in the world. Joining Professor Michael Wigler, one of the most extensively cited researchers of all time, were Tom Gingeras, Josh Huang, Dick McCombie, Michael Schatz, David Tuveson, and Doreen Ware, as well as affiliate Greg Hannon.

A. Krainer



N. Tonks

Professor Adrian Krainer continued to receive honors for his work in RNA splicing and nusinersen (Spinraza®), a treatment for the neurodegenerative disease spinal muscular atrophy (SMA). He won the 2019 Breakthrough Prize in Life Sciences, the 2019 Peter Speiser Award, conferred by the Institute of Pharmaceutical Sciences of ETH Zürich, and the K-J. Zülch Prize from the Gertrud Reemtsma Foundation of the Max-Planck-Gesellschaft Society. Adrian was elected to the National Academy of Medicine.

The RNA Society named Adrian the recipient of the 2019 Lifetime Achievement Award. The RNA Society emerged from a group of scientists who, starting in 1982, met at CSHL for regular "RNA Processing" meetings. In 1993 the society was established formally, and although the annual members' meeting takes place elsewhere, an RNA Processing meeting is still held regularly at the Lab.

Professor Nicholas Tonks was named a 2019 Fellow of the American Association for the Advancement of Science (AAAS). Nick was honored in the field of Pharmaceutical Sciences for his contributions to our understanding of signal transduction, through the discovery of protein tyrosine phosphatases and the characterization of their structure, regulation, and function.

Professor and Howard Hughes Medical Institute (HHMI) Investigator Zachary Lippman was a 2019 MacArthur Fellow. Often referred to as "genius grants," the fellowship provides exceptional individuals in a variety of fields with backing for their intellectual and professional pursuits. Zach's research focuses on the genes that determine when, where, and how many flowers are produced on a plant.

Professor and HHMI Investigator Rob Martienssen was awarded the 2019 Martin Gibbs Medal for his innovative work in the field of plant biology. The award is presented by the American Society of Plant Biology (ASPB) to "an individual who has pioneered advances that have served to establish new directions of investigation in the plant sciences."

Rob's current work focuses on investigating the epigenetic mechanisms of plants and understanding their role in gene regulation and inheritance. He is also an expert on transposable elements, or "jumping genes," studying how they regulate other genes and are in turn regulated during plant development. The Laboratory has a long history with the American Society of Plant Biology, with Charles Shull (who did plant biology work here in the early 1900s) having been part of the ASPB's founding body.

Professor Chris Vakoc was awarded the Paul Marks Prize for Cancer Research in recognition of his significant and ongoing contributions to the understanding of cancer. Presented through Memorial Sloan Kettering Cancer Center (MSKCC), the prize is awarded to up to three young scientists every two years. An expert in how genes are controlled and regulated, Chris studies how dysfunctional gene control can aid and even lead to cancer.



Z. Lippman



R. Martienssen



C. Vakoc



D. McCandlish

Assistant Professor David McCandlish was named a 2019 Sloan Research Fellow. David is a quantitative biologist who develops computational and mathematical tools to analyze genetic data. His lab focuses specifically on analyzing data from so-called "deep mutational scanning" experiments, which determine, for a single protein, the functional effects of thousands of mutations.

Assistant Professor Je H. Lee received the Chan Zuckerberg Initiative (CZI) Seed Networks for the Human Cell Atlas grant, to map gene expression and RNA–protein interactions throughout the formation of breast tissue. The Human Cell Atlas is a global, scientist-led effort to create a reference map of all cell types in the human body as a fundamental reference for biomedical research.

J. Lee

I was humbled by the 2019 Canada Gairdner International Award for "pioneering research on the eukaryotic DNA replication cycles including initiation, regulation, and responses to DNA damage." It was rewarding to share this award with my collaborator Dr. John Diffley, Associate Director of the Francis Crick Institute, who was a former postdoctoral fellow in my lab.

The Gairdner Foundation's announcement states that "by describing the exact sequence of events involved in DNA replication, Stillman and Diffley have provided key insights into how our genome is duplicated and how this process is coordinated with many other essential cellular events, which have implications for understanding genome instability and tumor heterogeneity in cancer."

B. Stillman

I was also elected as a Fellow of the American Association for Cancer Research (AACR) Academy.
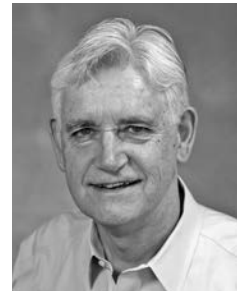
### *New Hires/Promotions*

We recruited three quantitative biologists to CSHL: Associate Professor Saket Navlakha; Assistant Professor Peter Koo; and CSHL Fellow Hannah Meyer.

CSHL Cancer Center Genetics and Genomics Program Leader, Chris Vakoc, was promoted to Professor.

Stephen Monez was recruited as Vice President, Chief Facilities Officer.

## Education Highlights

### *Meetings & Courses Program*

More than 7,200 participants from more than 50 countries attended meetings at CSHL this year. New meetings on Systems Immunology, Microbiome, and Zebrafish Neural Circuits and Behavior were added to the two-year cycle over which CSHL hosts 60 scientific meetings that span between three and five days.

S. Navlakha        P. Koo        H. Meyer        C. Vakoc

The Cold Spring Harbor Asia Program celebrated its 10th anniversary with a new partnership agreement signed by the Suzhou Industrial Park Administrative Committee that extends this program through 2028. In 2019, more than 3,300 participated in meetings at the Suzhou facilities. The next decade will see increased conference and course activities, as well as summer school programs, workshops, and smaller, invitation-only Banbury-style conferences.

At the Long Island campus, advanced scientific courses covered an array of topics in molecular biology, neurobiology, structural studies, and bioinformatics. Nearly 650 trainees, who included advanced graduate students, postdocs, and faculty, benefited from the contributions of the 700+ faculty.

The 84th Cold Spring Harbor Laboratory Symposium focused on RNA Control and Regulation, following previous symposia that have addressed different aspects of RNA biology, including the Symposia on Mechanisms in Transcription (1998), The Ribosome (2001), Epigenetics (2004), and Regulatory RNAs (2006).

Two of the three 2019 Nobel Prize winners in physiology or medicine, Bill Kaelin and Gregg Semenza, have both played active roles in a number of CSHL meetings, whereas the 2017 Nobel Prize winners Michael Rosbash and Michael Young jointly delivered the Seymour Benzer lecture at the 2019 Neurobiology of *Drosophila* meeting.

### Banbury Center

Eighteen Banbury meetings spanned six thematic areas: cancer, neuroscience, technology, public health, plant biology, and science policy. The 475 global experts who participated in these meetings worked to develop strategies for emerging fields or innovate in existing fields, bridge divides across sectors, disciplines, and communities, and address challenging policy issues.

The Center was productive, with 11 articles based on Banbury meetings published in peer-reviewed journals, as well as a journal supplement. Among the publications, a Science Policy Forum article outlined recommendations to increase gender diversity in STEM research, based on 2018's Increasing Gender Diversity in the Biosciences meeting. Banbury continues to impact the Lyme disease field in 2019, with new FDA and CDC policy announcements on diagnostics in line with published recommendations from a 2016 meeting.

Deep conversation at a meeting about HIV at the Banbury Center

A formal initiative to collect information on longer-term outcomes and impact of the Center's program was initiated, with a survey of 2017 meeting attendees. Ninety-nine percent of respondents indicated that new knowledge they gained at Banbury informed their work and 63% developed new collaborations as a result. For 25% of respondents, the meeting led to a new grant proposal, and 53% indicated that the meeting informed a new grant proposal. Although policy issues were part of the objectives for only a handful of the polled meetings, 30% of respondents reported that a meeting informed their policy views, with 13% reporting that the meeting contributed to a change in policy.

### Watson School of Biological Sciences

The 21st incoming class comprised four U.S. and four international students from Armenia, Mexico, China, and the United Kingdom. True to its mission to graduate students faster than students in comparable Ph.D.-granting institutions and position them to secure excellent jobs early in their careers, the program through 2019 counted 114 Ph.D. graduates. Thirty graduates had secured tenure track faculty positions. Twelve have been promoted to associate professor, and two are full professors. Our graduates have also moved into influential positions in administration, publishing, consulting, and industry.

Alumni Zach Lippman was awarded a MacArthur Foundation "Genius" award. Kristen Delevich was awarded a NARSAD Young Investigator Grant from the Brain and Behavior Foundation, and Nilgun Tasdemir was awarded a Pathway to Independence (K990/R00) Award from the National Institutes of Health (NIH).

Current students won prestigious fellowships, awards, and prizes, including the National Science Foundation Graduate Research Fellowship to Lyndsey Aguirre; the Boehringer Ingelheim Fonds Fellowship to Diogo Maia e Silva; and the Gilliam Fellowship for Advanced Study from the Howard Hughes Medical Institute to David Johnson. During the year, scientific papers published by current students of the School appeared in major journals, bringing the cumulative total of papers published by our students on their thesis research to more than 450.

A supplement to the program's NIH T32 training grant was funded in June, allowing for development of new course work and training modules related to career development. The new

WSBS entering class of 2019

curriculum is aimed at increasing student awareness and readiness for careers available to Ph.D.s; communication and negotiation skills and mentor/mentee relationships; and experiential learning in select career paths.

The 60th Undergraduate Research Program welcomed 18 undergraduates from the United States, China, and Ireland. The innovative Partners for the Future Program brought gifted local high school students to CSHL laboratories for hands-on research during their senior year.

### DNA Learning Center

In October, the DNA Learning Center (DNALC) became a lead institution in the InnovATE*BIO* National Biotechnology Education Center funded by the National Science Foundation's Advanced Technological Education (ATE) program. With a goal of workforce competitiveness, the program is focused on two-year colleges. The project is administered from Austin Community College, and the leadership team includes Madison College, Forsyth Technical Community College, Finger Lakes Community College, and the Bay Area Biotechnology Education Community. The DNALC's role is to develop a New York City Genomics hub to foster innovative labs and to support course-based student research at two-year colleges—including DNA barcoding and meta-barcoding.

Furthering this initiative is the announcement that CSHL signed a lease agreement to open the DNA Learning Center NYC at CityTech in Brooklyn, New York. Hosted by the New York City College of Technology (CityTech), programming at this facility builds on CSHL's initial foray into NYC through the Harlem DNA Learning Center established in 2008. The Brooklyn location is an 18,000-square-foot space that will be easily accessible to students and teachers in Brooklyn and the entire city of New York. In addition to programs for middle and high school students and teachers, the partnership with CityTech will develop research experiences and new curriculum for CityTech's two- and four-year degree programs.

In collaboration with Regeneron Pharmaceuticals, Inc., CSHL launched the Regeneron DNALC to serve the Hudson Valley. The new 4,700-square-foot center is located on Regeneron's Sleepy Hollow campus in Westchester County, New York. The interactive

CityTech interior

educational center is equipped with two state-of-the-art teaching labs to host local middle and high school field trips during the academic year, summer camps, and semester or year-long research projects.

During the academic year 20,358 students conducted labs at the Dolan DNA Learning Center, DNALC West, Harlem DNA Lab, Regeneron DNALC, and a temporary lab at CityTech; whereas 1,157 students attended week-long summer camps and 315 participated in research using DNA barcodes. An additional 7,628 students conducted in-school labs led by DNALC staff, and 1,758 used footlocker kits. Also this year, 5,224,126 visitors accessed DNALC's suite of multimedia resources online, including 3,749,711 visits to DNALC websites, 883,944 views of YouTube videos, and 590,471 downloads of smartphone/tablet apps, the *3D Brain, Weed to Wonder,* and *Gene Screen.*

### Cold Spring Harbor Laboratory Press



The open-access journal *Life Science Alliance*

The six research and three review journals published by the Press are the largest component of its publishing program. All of the review journals expanded their audience in 2019 and *Genes & Development, Genome Research*, and *RNA* remained prominent in their fields. The transition of the Laboratory's journal publishing program to an open-access model was signaled with the 2016 launch of the precision medicine journal *Cold Spring Harbor Molecular Case Studies*, which continued to gain submissions and readership in its fourth year. The open access journal *Life Science Alliance* was launched in 2018, published jointly by CSHL, the European Molecular Biology Organization, and Rockefeller University. This journal published 179 new articles in 2019, and monthly usage of its content grew >300% during the year.
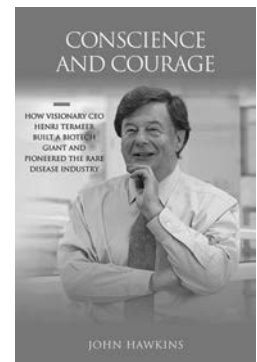


The cover of *Conscience and Courage*

Fewer print books are being published as the content typical of monographs and laboratory manuals is now delivered primarily online. *Conscience and Courage: How Visionary CEO Henri Termeer Built a Biotech Giant and Pioneered the Rare Disease Industry*, by John Hawkins, was a bestseller among the 12 new books published in 2019.

Most Press backlist books are now available in both print and e-book formats and direct sales from the Press website are close to those of the major online retailers. The mission of the Press remains helping scientists succeed, while enhancing the Laboratory's reputation, reach, and financial condition.

### Preprints in Biology and Medicine

A preprint is a research manuscript its authors first distribute via a dedicated server platform. Launched in 2013, the Laboratory's preprint server for life sciences, bioRxiv, made 68,800 papers freely available by the end of 2019 and attracted more than 4.7 million page views per month.

Inspired by bioRxiv's momentum, a preprint server for health sciences, medRxiv, was launched in June 2019 by the Laboratory in a management partnership with Yale University and BMJ, the global health information provider. By December medRxiv had posted more than 900 manuscripts and was receiving close to 1 million page views and downloads each month.

A preprint permits open community assessment of new research, but the formal peer review, endorsement, and publication of that work by research journals remains important. As part of a growing wave of experiments in how peer review is done, bioRxiv launched Transparent Review in Preprints (TRiP) in September 2019, bringing to readers of a preprint the reviews of the research commissioned by journals and other organizations.

By the end of 2019, the Laboratory's preprint platforms had distributed new work from investigators in more than 140 countries and it was evident that this new form of communication is transforming how biomedical research is conducted worldwide.

## Board of Trustees

Under the leadership of Chairman Marilyn Simons, Ph.D., the CSHL Board of Trustees elected four new trustees: Christine Anderson, Lyon Polk, Laura Slatkin, and Diana Taylor.

The Double Helix Medals Dinner (DHMD) honoring Boomer Esiason and Nancy Wexler, Ph.D., raised more than $4 million. Columbia University professor and president of the Hereditary Disease Foundation, Wexler is known for her scientific contributions on Huntington's disease and involvement in public policy, individual counseling, genetic research, and federal health administration. Former NFL quarterback Esiason advocates for cystic fibrosis research through the Boomer Esiason Foundation, a dynamic partnership of leaders in the medical and business communities, to heighten awareness, education, and quality of life for those affected by cystic fibrosis.

The dinner was chaired by Jamie Nicholls and Fran Biondi, Marilyn and Jim Simons, Teresa and Bob Lindsay, Janet and Frank DellaFera, Jenny and Jeff Kelter, and Danielle and Paul Taubman.

C. Anderson          L. Polk          L. Slatkin          D. Taylor

Since the first DHMD honored Muhammad Ali in 2006, the event has raised more than $40 million for the Laboratory's biological research and education programs.

Marilyn and Jim Simons made a significant pledge to support the future expansion of campus facilities, including housing for visiting scientists and additional lab space for neuroscience and quantitative biology research. Schmidt Futures pledged support to seed an Artificial Intelligence Fellows Program. Jenny and Jeff Kelter named a fellow in the graduate school.

## Library and Archives

With a pledge from BGI Group, the global genomics leader headquartered in Shenzhen, China, the BGI Nobel Laureates Archives was established. It comprises all of the current and future personal collections of Nobel laureates held by the CSHL Archives and includes Sydney Brenner, Francis Crick, Walter Gilbert, Carol Greider, Alfred Hershey, Barbara McClintock, Hermann Muller, Richard Roberts, and James Watson.

BGI shares the roots of human genetics history with CSHL, starting from the Human Genome Project (HGP). BGI, originally called the Beijing Genomics Institute, grew out of the vision of its founders to participate in the Human Genome Project, and they led China's contribution to that international effort. The new archive is an asset for the global community of genetics. BGI and its employees who contributed to this pledge strive to join the effort of CSHL to protect the heritage and to digitize the archives so that this resource is more accessible to mankind.

The Archives' History of Science annual meeting was titled Yeast Research: Origins, Insights, and Breakthroughs. Other events included a Special Lecture by Professor Rob Martienssen called Barbara McClintock's Controlling Elements Then and Now, and two Meet the Author events featuring 2009 Nobel laureate Venki Ramakrishnan and Anna Marie Skalka, Professor Emerita at Fox Chase Cancer Center.

## Business Development & Technology Transfer

2019 followed the banner year of 2018 that was dominated by the monetizing of the 2016 FDA-approved spinal muscular atrophy drug Spinraza®. The Business Development & Technology



H. Yang, B. Stillman, L. Pollock

Transfer team focused on building multiple industry relationships that represent significant and long-term value to CSHL.

The team increased licensing and equity revenue to $4 million, leveraging Krainer lab initiatives to expand patent reimbursement costs and liquidate stock from the CSHL spin-out company, Stoke Therapeutics, following a successful IPO. $400,000 in sponsored research funding was received under agreements negotiated and managed by this team.

Positioning around the business of "Innovation," the department worked to partner scientists with companies and investors to bring CSHL discoveries to the public domain through intellectual property and know-how licensing, industry (as well as academic) collaborative research, and new ventures.

The team is developing high-quality relationships between faculty and the best partner companies for their field. This high-level engagement supports the establishment of sponsored research agreements with multiple start-up companies and a large pipeline of opportunity with increasing numbers of faculty involved in translational work.

## Infrastructure

New York State Governor Andrew Cuomo helped open the newly renovated Demerec Laboratory in the fall. "It's good for the economy, but also [this is research] that I believe will improve the quality of life for thousands and thousands of people. I believe this work will save lives," Cuomo said during his visit. Home to four Nobel laureates, this building has been historically central to genetics research in New York and the world. New research in the building will focus on a holistic approach to treating cancer and the disease's impact on the entire body.

"This renovation allows us to really think about where the Lab will take things next," said Stillman. "It will have, I hope, a global impact on the research community, especially in the biomedical sciences."

Other significant projects this year included:

- *Marks Annex construction.* This included an additional postdoc office space, two conference rooms, and collaboration space with expected completion by mid-2020.



Governor A. Cuomo and B. Stillman walking in Demerec Laboratory, looking at images of famous scientists who had worked in that building.

An Open House event attracts participants to experiment for themselves.

- *Woodbury Genome Center Greenhouse construction*. This included an additional 2,500 square feet of greenhouse space with expected completion mid-2020.
- *DNA Learning Center at CUNY CityTech in Brooklyn, New York*. Planned construction includes 17,500 square feet of teaching facilities in the CityTech Pearl Building with construction activities to begin early 2020 and expected completion by spring 2021.
- *Airslie House renovation*. Construction activities continued with expected completion summer 2020.

The Laboratory continued its program of modernizing and improving the heating, ventilation, air conditioning, electrical, and plumbing systems throughout the campus.

## Community Outreach

In June, CSHL welcomed more than 650 neighbors and friends to an Open House event that showcased the many different facets of the Lab and what happens on our campuses every day. Nearly 80 CSHL staff, graduate students, postdocs, and faculty helped visitors explore the campus and learn about our innovative research and education programs.

Scientists shared the latest from their laboratories about DNA, plant biology, cancer research, neuroscience, and quantitative biology. At numerous experimentation stations, scientists showed off corn samples, allowed peeks into microscopes, and gave lessons on how to extract DNA. There were also short talks on topics that ranged from understanding the immune system, tackling pancreatic cancer, and developing biofuels, to how the Banbury Center "think tank" has helped advance science policy.

A team of 18 CSHL graduate students guided 78 public tours throughout the year. More than 1,500 visitors participated in these events. Graduate students and instructors from the DNA Learning Center participated in events to engage with local elementary school children and support local school science fairs and science demonstrations.

CSHL's Public Affairs Department managed these outreach events and continued the annual science lecture series at Grace Auditorium. This team also organized more informal events that bring our science to the general public in other locations throughout the community.

A. Solomon and M. Wigler discuss *Far from the Tree*.

### CSHL Public Presentations

**January 19:** Screening and discussion at Cinema Arts Centre in Huntington Co.; *Far from the Tree*; Panelists: **Michael Wigler, Ph.D.,** professor, CSHL; **Andrew Solomon,** writer; co-presented by Cold Spring Harbor Laboratory and Cinema Arts Centre as a Science on Screen event.

**April 11:** Lecture at Port Washington Public Library; Pancreatic Cancer: Advances in Research; **Lindsey Baker, Ph.D.,** CSHL research investigator; co-presented by Cold Spring Harbor Laboratory and the Lustgarten Foundation.

**April 17:** Public lecture, Seeing with Sequencing; **Molly Gale Hammell, Ph.D.,** associate professor, CSHL; **Justin Kinney, Ph.D.,** assistant professor, CSHL; **David McCandlish, Ph.D.,** assistant professor, CSHL.

**April 24:** *Cocktails & Chromosomes* at Six Harbors Brewing Co.; **Ullas Pedmale, Ph.D.,** assistant professor, CSHL.

**May 29:** Screening and discussion at Cinema Arts Centre in Huntington; *Iceman*; Panelists: **Lindsay Barone, Ph.D.** and **Elna Carrasco-Gottlieb,** DNA Learning Center; co-presented by Cold Spring Harbor Laboratory and Cinema Arts Centre as a Science on Screen event.

**June 24:** Public lecture, Diet and Disease: Exploring the relationship between nutrition and cancer; **Semir Beyaz, Ph.D.,** CSHL fellow; **Jamie Kane, M.D.,** director, Center for Weight Management, Northwell Health, assistant professor, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell; co-sponsored by CSHL, US Trust, Northwell Health, and St. Johnland Nursing Center.

**July 17:** *Cocktails & Chromosomes* at Six Harbors Brewing Co.; **Camila dos Santos, Ph.D.,** assistant professor, CSHL.

**October 6:** Public lecture, Diversity, Ethnicity and Cancer; 2019 Lorraine Grace lectureship on societal issues of biomedical research, presented as part of the ongoing Roy J. Zuckerberg community engagement series from CSHL's NCI-Designated Cancer Center; **Olufunmilayo I. Olopade, M.D., F.A.C.P.,** Walter L. Palmer Distinguished Service Professor of Medicine and Human Genetics, director, Center for Clinical Cancer Genetics & Global Health, University of Chicago Medicine.

**October 10:** *Cocktails & Chromosomes* at Six Harbors Brewing Co.; **Tobias Janowitz, M.D., Ph.D.,** assistant professor, CSHL.

**October 28:** Public lecture, Food and Climate—the Way Forward; Introduction: **Bruce Stillman, Ph.D.,** CSHL president and chief executive officer; Speaker: **Katy Kinsolving**, president of the C-Change Conversations; Panelists: **Rebecca Benner**, director of Conservation and Science for the Nature Conservancy; **Peter Lehner**, director of Earthjustice's Sustainable Food & Farming Program; **Doreen Ware, Ph.D.,** molecular biologist with the USDA Agricultural Research Service (ARS) and CSHL adjunct professor; presented by Cold Spring Harbor Laboratory and North

Shore Land Alliance together with North Country Garden Club, St. John's Church (Cold Spring Harbor), the Nature Conservancy (Long Island chapter), and Three Harbors Garden Club.

**November 13:** *Cocktails & Chromosomes* at Six Harbors Brewing Co.; **Lloyd Trotman, Ph.D.**, professor, CSHL.

*CSHL Public Concerts*

**April 5:** Nathan Lee, piano
**April 26:** Matthew Graybil, piano
**May 3:** Tanya Bannister, piano
**August 23:** Jiayin Shen and Igor Lovchinsky, piano duo
**September 13:** Horszowski Trio
**September 27:** Hanzhi Wang, accordion
**November 8:** Zlatomir Fung, cello



Zlatomir Fung (credit, Matt Dine)

## Looking Forward



CSHL named a Top Workplace

CSHL was proud to be named again by *Newsday* as one of Long Island's Top Workplaces. The anonymous employee surveys of employers across Long Island were analyzed by a third party. As part of the large company category, those with more than 500 employees, the Lab shared the honor with only 12 other institutions. This is CSHL's second year winning this prestigious award. Thank you to our faculty, students and employees for making CSHL a great place to work and to all of you who contribute to the success of this institution. With your support, we will continue to advance biology and genetics to benefit mankind.

**Bruce Stillman, Ph.D., F.R.S.**
*President and Chief Executive Officer*

# CHIEF OPERATING OFFICER'S REPORT

Completing my first full year as COO, I am encouraged by the passion of CSHL's faculty, students, and employees. Thanks to their efforts, 2019 was a successful year for the Laboratory scientifically, operationally, and financially.

## Financial Matters

We budget conservatively and maintain fiscal discipline throughout the year, which allows some flexibility to pursue research and hiring opportunities in the very dynamic environment in which we operate. In 2019 the Laboratory's operating results included higher than budgeted third-party revenues and lower than budgeted expenses, resulting in a surplus of $1.6 million. Operating revenues totaled $185.9 million, which included $117.8 million of external grant funding, $39.8 million of non–research division support and $28.4 million of contributions from the endowment, annual fund, interest, and royalties.

J. Tuke

Operating expenses are dominated by our investment in people, with compensation and benefits totaling $87.4 million, and infrastructure improvements and maintenance of our facilities at $11.2 million. These are relatively fixed costs—required so as to not disrupt our research activities—and we strive to recognize this reality by securing and growing stable funding sources.

Growing the endowment, especially unrestricted funds, remains a strategic priority of the Board as federal and private grant funding becomes increasingly competitive. Endowment balances grew from $595 million to $665 million during 2019, net of the operating draw, gifts, and investment returns. Endowments are meant to provide intergenerational equity and the trailing 10-year return through 2019 of 7.1% exceeded CPI inflation by 5.5%. The Board-approved spending policy is to appropriate 4.5% of the trailing 12-quarter average endowment value for operations.

## Operating Matters

Organizations as complex as CSHL benefit from appropriate longer-range planning and an ongoing review of systems, practices, and policies that provide the framework for efficient, effective, and compliant operations. Last year management commenced several projects with this philosophy in mind.

Enterprise Risk Management (ERM) is now a "best practice" for not-for-profits and an important fiduciary responsibility of the Board, with oversight at CSHL provided by its Audit & Risk Committee. Management established a Risk Committee and, with some guidance from our auditors and insurance broker, engaged with every division and department of CSHL. The purpose is to identify all significant financial, operational, legal, regulatory, or reputational exposures, have the "risk owners" articulate the existing mitigations, and research what other remediations should be considered.

One major development in 2019 was success achieving a long-sought expansion of the DNA Learning Center (DNALC) into New York City, with the signing of a 30-year lease (at $1/yr) in Brooklyn. The expansion project will make it far easier for middle and high school students, especially those from underrepresented communities, to participate in our world-leading genetics education programs. With renovations under way just off the Brooklyn Bridge, in a building owned by our partner, CUNY, we anticipate opening the facility in early 2021. This $20 million project represents the largest investment ever made by the Laboratory in programming outside of our main campus on Long Island.

Over the last year, I have come to appreciate that CSHL is fortunate to have seasoned, dedicated, and stable senior administrators. The newest member of this team is Steve Monez, our VP, Chief Facilities Officer, who joined us from Mt. Sinai Hospital in New York City last fall— bringing with him years of experience working in a scientific and academic environment. The average tenure here of our senior administrators is 20 years, ranging from 1 to 41. Maintaining a culture of collaboration, significant autonomy, and efficient decision making is a very desirable feature in attracting and retaining leadership, along with the chance to further our research and educational missions in a world-class environment.

## Challenges

As I write this report in the spring of 2020, the financial markets are in turmoil, and the longer-term impact on the endowment, our third largest source of funding behind third-party grants and our divisions' revenues, is uncertain. Fortunately, the "smoothing" aspect of CSHL's spending policy will moderate any short-term erosion.

The regulatory and compliance framework of the Laboratory's activities continue to grow, and in 2019 much of it was related to concerns by federal agencies over the integrity and protection of research and intellectual property. CSHL has added personnel to meet these requirements, while honoring the spirit of the sharing of scientific ideas and advances to benefit all.

The Laboratory seems to endure and even flourish in times of rapid change, demonstrated by the resiliency it has shown over more than 130 years because of the dedication of its employees, the importance of its mission, and support of the broader CSHL family.

**John P. Tuke**
*Chief Operating Officer*

# Long-Term Service



Attendees at the 2019 long-term employee dinner. (*Left to right*) Charlene De Poto, Sharon Bense, Alison McDermott, Bruce Stillman, Mila Pollock, Dessie Carter, Jeannette Amato, Jim Hope, Mary Ellen Goldstein, David Spector, Susan De Angelo, Karine Boyce, Rob Gensel, Danny Miller, Chun-hua Yang, Alex Gann, Kevin Donohue, Jan Argentine, Rob Martienssen, Diane Esposito, Carmelita Bautista, Jonathan Parsons, Drew Mendelsohn, Idee Mallardi, John Tuke.

The following employees celebrated milestone anniversaries in 2019:

| | |
|---|---|
| 40 years | Maureen Berejka, Jim Hope, Bruce Stillman |
| 35 years | Carmelita Bautista, Dessie Carter, Rob Gensel, Mary Ellen Goldstein, Danny Miller, Steven Tang |
| 30 years | Sharon Bense, Charlene De Poto, Rob Martienssen, Alison McDermott |
| 25 years | Leslie Allen, Jan Argentine, Susan De Angelo, Carol DuPree, Diane Esposito, Idee Mallardi, Drew Mendelsohn |
| 20 years | Fred Munter, Inessa Hakker, Mila Pollock, Alex Gann, Jeannette Amato, Kevin Donohue, Tony Zador, Chun-hua Yang, Karine Boyce, Jonathan Parsons |
| 15 years | Peter Andrews, Christy Bedell, Sabrina Boettcher, Daniel Chapman, Lina Crawford, Jared Downing, Andres Gonzalez, Bo Li, Gustavo Munoz, Kenneth Orff, Tricia Penner, Maria Smit, Hillary Sussman, Jason Williams |

RESEARCH

# CANCER: GENE REGULATION AND CELL PROLIFERATION

**Camila dos Santos**' laboratory studies the epigenetic regulation of normal and malignant mammary gland development with an emphasis on the alterations brought by pregnancy. Significant changes mark the pre- and postpubescence mammary developmental stages, but those associated with pregnancy have the greatest effect on cellular function, tissue reorganization, and breast cancer susceptibility. dos Santos' group has recently found that mammary glands react differently to a second pregnancy than they do to the first pregnancy, with associated changes in DNA methylation. These findings suggested that pregnancy changes the state of mammary cells and these may permanently alter how they react to the next pregnancy. In addition, the dos Santos laboratory is exploring how the pregnancy-induced epigenetic changes might influence cell transformation and the risk of breast cancer. This research uses genomic and computational approaches to define the pre- and postpregnancy mammary epigenome. An additional objective of the dos Santos laboratory is to use functional genomics to discover novel transcriptional regulators that modulate mammary stem-cell self-renewal, lineage specification, and cell transformation. The long-term objective of dos Santos' group is to improve the notion of the mammary epigenome during normal development and use this information to gain insights into new preventive and curative strategies to target breast cancer.

In **Leemor Joshua-Tor**'s laboratory, researchers study the molecular basis of nucleic acid regulatory processes using the tools of structural biology and biochemistry. One such regulatory process is RNA interference (RNAi), in which a small double-stranded RNA triggers gene silencing. Joshua-Tor and her team offered critical insights when they solved the crystal structure of the Argonaute protein and identified it as the long-sought Slicer. They then went on to explore the mechanism of the slicing event. The structure of human Argonaute 2 (hAgo2) bound to a microRNA (miRNA) guide allowed Joshua-Tor and her colleagues to understand how mRNA is cleaved during RNAi. This past year, members of the Joshua-Tor laboratory explored the function of a very similar protein, called Argonaute 1, which has no slicing ability, even though it is almost identical in structure to the slicing hAgo2. Using biochemical methods and mutational analysis, they were able to identify key parts of the protein that are required for slicing activity. The laboratory also studies the generation of PIWI-interacting RNAs (piRNAs), which serve to protect the genome of germ cells. With colleagues in the Hannon laboratory, Joshua-Tor's team also determined the structure and function of Zucchini, a key nuclease in the initial generation of piRNAs in fruit flies. In other work, the laboratory is exploring the mechanisms of heterochromatin formation and gene silencing through the study of a protein complex called RNA-induced initiation of transcriptional gene silencing (RITS). Joshua-Tor is also well known for her work on the E1 helicase enzyme, which acts to unwind DNA strands during the DNA replication process.

**Adrian Krainer**'s laboratory studies the mechanisms of RNA splicing, ways in which they go awry in disease, and the means by which faulty splicing can be corrected. In particular, Krainer's team studies splicing in spinal muscular atrophy (SMA), a neuromuscular disease that is the leading genetic cause of death in infants. In SMA, a gene called *SMN2* is spliced incorrectly, making it only partially functional. The Krainer laboratory found a way to correct this gene defect using a powerful therapeutic approach. They found it possible to stimulate SMN protein production by altering mRNA splicing through the introduction into cells of chemically modified pieces of RNA called antisense oligonucleotides (ASOs). Following extensive work with ASOs, in mouse models

of SMA, one such molecule, known as nusinersen or Spinraza®, was taken to the clinic, and at the end of 2016 it became the first FDA-approved drug to treat SMA, by injection into the fluid surrounding the spinal cord. The Krainer laboratory is currently using this approach for the study of other diseases caused by splicing defects, including familial dysautonomia. In addition, they are applying antisense technology to stabilize mRNAs that are destroyed by a process called nonsense-mediated mRNA decay (NMD), to learn about the underlying mechanisms and to develop new therapies. The Krainer laboratory has also worked to shed light on the role of splicing proteins in cancer. They found that the splicing factor SRSF1 functions as an oncogene and characterized the splicing changes it elicits when overexpressed in the context of breast cancer; several of these changes contribute to various aspects of cancer progression. Finally, the laboratory continues to study fundamental mechanisms of splicing and its regulation, and they identified novel ways in which the U1 snRNA can recognize natural 5′ splice sites that deviate from the consensus.

**David L. Spector**'s laboratory is focused on characterizing long noncoding RNAs (lncRNAs) that exhibit altered levels of expression in breast cancer progression and during embryonic stem cell differentiation. A major focus of their efforts has been on Malat1 lncRNA, which is one of the most abundant lncRNAs. The Spector laboratory previously identified a novel mechanism of 3′-end processing of this RNA. More recent studies have revealed that increased levels of Malat1 lncRNA impact breast cancer progression and metastasis. Knockout or antisense oligonucleotide knockdown of Malat1 results in the differentiation of mammary tumors and a significant reduction in metastasis. Studies are currently under way to elucidate the mechanism of action of this abundant nuclear retained lncRNA and to implement innovative therapeutic approaches that can impact its function in vivo. In addition, they have identified additional lncRNAs, termed mammary tumor–associated RNAs, that are up-regulated in breast tumors; the team is currently assessing the function of these lncRNAs using 3D tumor organoids as well as mouse models.

A second area of study in the Spector laboratory is based on their earlier discovery of an increase in random autosomal monoallelic gene expression on the differentiation of mouse embryonic stem cells to neural progenitor cells. These data support a model in which stochastic gene regulation during differentiation results in monoallelic gene expression, and, for some genes, the cell is able to compensate transcriptionally to maintain the required transcriptional output of these genes. Therefore, random monoallelic gene expression exemplifies the stochastic and plastic nature of gene expression in single cells. Ongoing studies are examining the relationship of monoallelic gene expression to lineage commitment.

**Bruce Stillman**'s laboratory studies the process by which DNA is copied within cells before they divide in two. Working with yeast and human cells, Stillman and colleagues have identified many of the cellular proteins that function at the DNA replication fork during the S phase, the portion of the cell-division cycle in which DNA synthesis occurs. Among these proteins are those that facilitate the assembly of chromatin, the protein–DNA complexes that form the chromosomes. Current research focuses on the mechanism that initiates the entire process of DNA replication in eukaryotic cells. At the heart of this mechanism is a protein that binds to "start" sites on the chromosomes, called the origin recognition complex (ORC). The Stillman laboratory is part of an ongoing collaboration that determined the cryo-EM structure of ORC proteins in complex with a group of proteins, called helicases, that unwind DNA during replication. These images offer molecular insights into how the helicase is loaded onto DNA. Stillman's research also focuses on the process by which duplicated chromosomes are segregated during mitosis. They found ORC at centrosomes and centromeres, structures that orchestrate chromosome separation during mitosis. At centromeres, ORC subunits monitor the attachment of duplicated chromosomes to the mitotic spindle that pulls the chromosomes apart when they are correctly aligned. Stillman's team has discovered that mutations in the Orc1 protein alter the ability of ORC to regulate both DNA

replication and centrosome duplication. These mutations have been linked to Meier–Gorlin syndrome, a condition that results in people with extreme dwarfism and small brain size, but normal intelligence.

Cancer can be understood as a disease of dysfunctional gene expression control. Research in **Chris Vakoc**'s laboratory investigates how transcription factors and chromatin regulators cooperate to control gene expression and maintain the cancer cell state. This work makes extensive use of genetic screens to reveal cancer-specific functions for transcriptional regulators, as well as genomic and biochemical approaches to identify the molecular mechanisms. One theme that has emerged from their efforts is that blood cancers are often vulnerable to targeting transcriptional coactivators, such as BRD4 and the SWI/SNF chromatin remodeling complex. Vakoc's team demonstrated that chemical inhibition of BRD4 exhibits therapeutic effects in mouse models of leukemia, a finding that has motivated ongoing clinical trials in human leukemia patients. The Vakoc laboratory has also developed a CRISPR-Cas9 screening approach that can reveal individual protein domains that sustain cancer cells. Their laboratory is now deploying this technology in a diverse array of human cancers to reveal therapeutic opportunities and basic mechanisms of cancer gene control.

Stem and progenitor cells of many adult lineages undergo self-renewal and differentiation, and growth signal and nutrient are two major regulators of this process. Exploiting mechanisms, especially the coordination of growth signal and nutrient, holds great promise for the development of regenerative medical strategies and the prevention of abnormal cell proliferation in cancer. The research of **Lingbo Zhang**'s laboratory centers on normal and malignant stem and progenitor cells in the hematopoietic system. They use functional genomics, including chemical genomic and CRISPR-Cas9 functional genomic approaches, to uncover novel regulators of growth signal and nutrient pathways. Through collaborations with medicinal chemists, Zhang's team builds pharmacological approaches to target these novel self-renewal regulators and metabolic vulnerabilities, and the team's findings will help treat devastating hematological diseases and malignancies, including myelodysplastic syndrome and leukemia.

# UNDERSTANDING THE EPIGENETIC REGULATION OF NORMAL AND MALIGNANT MAMMARY GLAND DEVELOPMENT

Camila dos Santos    C. Chen        M. Feigman    M. Trousdell
                     M. Ciccone     M. Moss       A. Varshin Hanasoge Somasundara
                     S. Cyrill      I. Munhoz     S.T. Yang

## Defining a Role for BPTF Inhibition in Blocking Breast Cancer Progression

C. Chen, S.T. Yang, M. Ciccone

During this past year, we focused on defining the effects of BPTF inhibition on controlling the development and progression of mouse mammary tumors. To date, we have established distinct roles for BPTF in a series of murine models of mammary tumorigenesis.

We have monitored mammary tumor development in a cohort of Brca1-deficient female mice, and observed mammary tumors developed within two to nine months after tamoxifen (TAM) administration in all wild-type (wt) mice ($n = 12$). In marked contrast, BPTF knockout (KO) mice did not develop mammary tumors during the one-year observation period, suggesting that *Bptf* depletion significantly blocked mammary tumorigenesis in Brca1-deficient female mice. We are currently setting up a series of cellular and molecular analyses to determine the specific effect of BPTF loss on Brca1-deficient mammary tumorigenesis. In addition, we are developing a CRISPR-Cas9 system to active BPTF expression in normal mouse/human mammary cell lines/organoids, with the goal to define a role for BPTF gain-of-function on cancer progression, especially in cases accompanied by Brca1 deficiency.

Alternatively, tumor-monitoring experiments suggest that *Bptf* deletion delayed the onset and progression of MMTV-PyMT mammary tumors. Histological and immunofluorescence analysis indicated smaller and rare lung metastasis in BPTF KO mice. We are currently characterizing the alterations in metastatic programs and tumor microenvironments that were conferred by *Bptf* deletion and that impact the development of metastatic lesions in MMTV-PyMT mice.

More recently, we developed a $Krt5^{CRE-ERT2}Bptf^{fl/fl}$ CAGMYC transgenic model. In this model, overexpression of human c-MYC, a known inducer of mammary tumor development and a major oncogene associated with up to 60% of all breast cancer subtypes, is driven in a doxycycline-dependent manner. Preliminary data demonstrated that treatment of CAGMYC mammary organoid cultures with small-molecule BPTF inhibitor results in abnormal organoid proliferation after doxycycline-induced c-MYC overexpression. This result is consistent with our previous data showing that loss of BPTF function blocks cell cycle progression and induces apoptosis in human breast cancer cell lines. We are currently monitoring the effects of *Bptf* deletion on malignant mammary lesion progression in this mouse model.

## Characterization of Pharmacological BPTF Inhibition in Human Cancer Cell Lines

C. Chen, M. Ciccone, C. dos Santos [in collaboration with W. Pomerantz, University of Minnesota]

We have co-characterized the first-in-class small-molecule BPTF inhibitor, S-AU-1, in collaboration with the Pomerantz group at the University of Minnesota. Our previous work showed that S-AU-1 effectively treated the growth of several human breast cancer cell lines, such as MCF7 cells, Hs578t cells, MDA-MB-453 cells, and HCC-1937 cells. We also established that S-AU-1 treatment had no effect on the proliferation of HEPG2 cells, a human liver cancer cell line, thus establishing an important control to characterize the immediate and prolonged effects of BPTF inhibition.

Computational and chemical analysis suggested that S-AU-1 may also associate with the kinases CDKL2 and NTKR3, suggesting that inhibition of

these kinases alongside BPTF could influence growth and survival of breast cancer cell lines. To rule this hypothesis out, we engineered CDKL2 and NTKR3 KO (CRISPR-Cas9) cancer cell lines, and initial analysis suggests that breast cancer cell lines do not require CDKL2 and NTKR3 for growth and survival. These results also support the view that S-AU-1 treatment may be exclusively inhibiting BPTF function. We are currently testing newly developed BPTF inhibitors to assess their ability to affect cell growth and the chromatin state of breast cancer cell lines.

## Investigation of the Epigenetic Modifications Brought about by Pregnancy

C. Chen, M. Ciccone, C. dos Santos

Our previous study demonstrated that epigenetic changes conferred on mammary epithelial cells during pregnancy regulate gene expression dynamics, including a more robust up-regulation of genes associated with milk production in a second pregnancy cycle.

We have recently optimized human- and murine-derived organoid culture systems to recapitulate the pregnancy-induced development in vitro, a strategy that has shown post-pregnancy mammary organoids engaging in differentiation and gene expression changes more rapidly than organoid culture derived from pre-pregnancy mammary glands.

Using this system, we are now investigating the role that specific chromatin modifiers play in controlling changes in gene expression and chromatin accessibility in response to pregnancy hormones. Understanding the mechanisms that underlie pregnancy-induced changes would allow us to better determine epigenomic modifications and their contribution to gene regulation across pregnancy cycles and evolutionary scales.

## Defining the Molecular Basis of Pregnancy-Induced Breast Cancer Protection

M. Feigman, C. Chen, M. Moss, M. Trousdell, C. dos Santos

For nearly 100 years, population studies have consistently and definitively found that a full-term pregnancy early in life dramatically reduces the lifetime incidence of breast cancer. Whereas research has shown increased breast cancer risk for roughly five to ten years after, there is a long-term reduction of risk of breast cancer for women completing a full-term pregnancy before the age of 30. A similar risk decrease following pregnancy has been observed in mice, where completion of a pregnancy cycle dampens the frequency of mammary tumor development. To understand the molecular basis for this protective effect, we compared the enhancer landscape of murine pre- and post-pregnancy MECs, and related this analysis to gene regulation, tissue development, and oncogenesis. We found a substantial expansion of the active *cis*-regulatory landscape of post-pregnancy MECs, which associated with a robust activation of pregnancy-related programs and influenced normal MEC development during consecutive exposure to pregnancy hormones.

To characterize the effects of a pregnancy-induced epigenome in response to oncogenic stress, we established a transgenic mouse strain (CAGMYC) in which overexpression of the oncogene *cMYC*, a known inducer of mammary tumor development, is driven in a doxycycline-dependent manner. Using this transgenic mouse strain, we demonstrated that post-pregnancy MECs are less efficient at activating molecular programs driven by cMYC overexpression, a response that interferes with the development of premalignant lesions, but did not perturb the pregnancy-induced epigenomic landscape. Further epigenomic analysis demonstrated that cMYC overexpression activates senescence programs in post-pregnancy CAGMYC MECs, and interruption of such programs increased malignant-like phenotypic changes of organoid cultures in response to *cMYC* overexpression.

## Characterization of Factors That Block Breast Cancer Development and Progression

S. Cyrill, S.T. Yang, C. dos Santos

Our previous definition of the pregnancy-induced enhancer landscape led us to investigate whether such an enhancer landscape is regulated by MLL3 (Kmt2c), an enzyme that deposits a single methyl group on the fourth lysine of histone 3—a mark for enhancer regions. To test the role of MLL3 in enhancer dynamics of MECs, we generated multiple lines of murine MECs (Eph4) with MLL3 loss of function either by employing CRISPR-Cas9 gene editing to induce mutations in the catalytic SET domain of the MLL3

gene or by targeted RNA interference of MLL3 expression. With these lines, we used a combination of in vitro and in vivo assays to understand the role of MLL3 in a pregnancy-dependent manner.

To test the responses of MLL3-deficient cells to pregnancy hormones in a 3D environment, we generated spheroid cultures from the MLL3 mutant cells by embedding them in a Matrigel® matrix. The spheroids were then treated with pregnancy hormones—17β-estradiol, progesterone, and prolactin—to drive proliferation and differentiation. We found that the MLL3-mutated lines had differential spheroid morphologies after treatment compared to MECs with wt, functional MLL3, and are working on quantifying these differences and identifying key targets of the MLL3-regulated parous epigenome.

In parallel, we transplanted the mutant lines into the cleared fat pad of prepubescent female mice that were then put through natural or pseudo-pregnancy to assess the impact of MLL3 dysfunction in vivo at different stages of pregnancy. Using histological analysis and immunofluorescence staining, we analyzed ductal branching during and after pregnancy. We are now exploring the use of CUT&RUN to understand the impact of MLL3 deficiency on epigenetic and gene regulation in MECs experiencing pregnancy-associated changes.

## The Effects of Pregnancy on Reprogramming the Immune Environment of Mammary Glands

M. Feigman, M. Moss, M. Trousdell, C. dos Santos

Using bulk RNA sequencing, we found that post-pregnancy mammary epithelial cells up-regulate genes associated with immune cell communication, suggesting that pregnancy may override the homeostatic control of immune communication in the mammary gland. We identified a unique population of immune cells characterized by expression of cytotoxic genes, *ctsw*, *nkg7*, and *gzma*, which is specifically found in post-pregnancy mammary gland. As pregnancy can alter the predisposition of rodent and human MECs to oncogenesis, we show that loss of post-pregnancy mammary immune resident cells resulted in the development of mammary malignant lesions.

We also discovered that mammary epithelium cells overexpress the antigen presentation CD1d, and we believe that alteration of CD1d levels after pregnancy and during malignant transformation may govern cancer susceptibility in mammary epithelial cells. To address this hypothesis, we have crossed transgenic mouse lines lacking CD1d expression (CD1d KO mice) with our mouse model of pregnancy-induced breast cancer protection (CAGMYC) to probe for a role of CD1d KO MECs in decreasing mammary malignant development. Supporting our hypothesis, we found that CD1d KO cells develop into malignant lesions in response to *cMYC* overexpression regardless of parity.

Moreover, and in partnership with plastic surgeons at Northwell Health, we were able to acquire breast tissue specimens donated by healthy women spanning several parity statuses for molecular analysis. Single-cell RNA sequencing analysis on these samples will provide the basis to investigate patterns of immune-related gene expression profiles in the context of their microenvironment in human tissue. This cell-to-cell variability resolution will also provide insights into how to interlace pregnancy, immune responses, and cancer development in mouse and human tissue.

## Understanding the Effect of an Immune Response to Infections on Pregnancy-Induced Mammary Gland Development and Breast Cancer

S. Cyrill, C. dos Santos

In this project, we set out to understand how the systemic effects of immune responses elicited by urinary tract infections (UTIs) influence a mammary gland involution, tissue homeostasis, and oncogenesis. In a model of pseudo-pregnancy commonly used to study the parity-induced signals, we found that urethral abrasion resulted in the development of UTIs and aberrant mammary histology. This condition is seemingly reversed in mice treated with broad-spectrum antibiotics during the course of the pseudo-pregnancy. Histological analysis of the glands at a post-involution time point revealed the retention of milk proteins inside ductal structures, suggesting that mechanisms involved in the clearance and involution of the mammary gland were delayed in UTI-bearing mice.

To further understand the impact of systemic signals of a UTI on mammary gland biology and

development, we established an infection model using transurethral administration of the uropathogenic *Escherichia* coli strain UTI89. Flow cytometric analysis of the epithelial and immune compartments of mammary glands from nulliparous mice with UTI89 infections revealed alterations in the immune milieu in the mammary glands of infected mice, supporting that UTI-bearing mice have an altered mammary immune microenvironment. We are currently extending this analysis to a model of Brca1 deficiency, with the goal to study the impact of UTI-driven systemic inflammation on mammary oncogenesis.

## The Influence of L-Asparaginase Treatment and Lysosomal Protease Cathepsin B on the Transcriptome of Acute Lymphoblastic Leukemia REH Cell Line

I. Munhoz, C. dos Santos

Acute lymphoblastic leukemia (ALL) is the most common childhood neoplasia among leukemias. Using the enzyme L-asparaginase (ASNase) in treatment, the survival of patients with ALL, within five years, ranges from 80% to 90%. Despite its effectiveness, ASNase presents a large number of adverse effects, which have been associated with the reduction of the half-life of enzymes, altering their activity and causing drug resistance, and may lead to discontinuation of treatment. In solid tumors, lysosomal protease cathepsin B (CTSB) is involved in the progression of the disease, acting on the proliferation, invasion, and metastasis. CTSB has also been associated with ASNase degradation during treatment, but its involvement in disease progression and response to treatment remain largely unknown. To elucidate the influence of CTSB during the treatment of ALL with ASNase, we generated CRISPR-Cas9 *CTSB* gene KO lines using the ALL human cell line REH, which is considered resistant to ASNase treatment. Such cell lines will be used in cellular and molecular assays to elucidate the basis for ASNase resistance and whether CTSB activity interferes with resistance of REH cells to ASNase.

## Investigating the Role of Changes to the Mammary Gland Immune Microenvironment in Pregnancy-Induced Breast Cancer Protection

A. Varshin Hanasoge Somasundara

Our laboratory has previously identified a population of immune cells in the mammary glands of parous female mice and demonstrated that depletion of such cells fosters the development of cMYC-induced malignant transformation of post-pregnancy mammary epithelial cells. We now aim to elaborate on these preliminary results, and to identify the influence of pregnancy signals on clonal expansion and specialization of mammary resident immune cells. In an orthogonal approach, we will test whether pregnancy induces epigenetic and tissue alterations that block the development of *BRCA1*-deficient mammary oncogenesis similar to what we have observed in our model of cMYC-driven oncogenesis. Given that *BRCA1* mutations in humans increase the risk of development of breast cancer by >50%, understanding the biology and function of pregnancy-induced natural killer T cells and *BRCA1*-deficient pregnancy-associated oncoprotection may elucidate strategies to prevent tumorigenesis in high-risk populations.

### PUBLICATIONS

Feigman M, Moss M, Chen C, Cyrill SL, Ciccone FM, Frey W, Yang ST, Wilkinson JE, dos Santos CO. 2019. Pregnancy reprograms the enhancer landscape of mammary epithelial cells and alters the response to cMYC driven oncogenesis. BioRxiv doi:10.1101/642330v1

Kinberger SE, Ycas PD, Johnson JA, Chen C, Ciccone MF, Woo RWL, Urick AK, Zahid H, Shi K, Aihara H, et al. 2019. Selectivity, ligand deconstruction, and cellular activity analysis of a BPTF bromodomain inhibitor. *Org Biomol Chem* **17:** 2020–2027.

Slepicka PF, Cyrill SL, dos Santos CO. 2019. Pregnancy and breast cancer: pathways to understand risk and prevention. *Trends Mol Med* **25:** 866–881.

*In Press*

Fein MR, He XY, Almeida AS, Bruzas E, Pommier A, Eberhardt A, Fearon DT, Wilkinson JE, dos Santos CO, Egeblad M. 2020. Cancer cell CCR2 orchestrates suppression of the adaptive immune response. *J Exp Med* **217:** e20181551.

# NUCLEIC ACID REGULATORY PROCESSES

**L. Joshua-Tor**  D. Adams      J. Ipsaro      K. On
                   B. Bibel      M. Jaremko     V. Pillai
                   E. Elkayam    S. Jones       J. Scheuring
                   A. Garg       J. Jung        R. Tseng
                   S. Goldsmith  K. Meze

We study the molecular basis of nucleic acid regulatory processes—RNA interference (RNAi) and DNA replication in particular. We use the tools of structural biology, biochemistry, and biophysics to study proteins and protein complexes associated with these processes to elucidate how they work. X-ray crystallography, cryo-electron microscopy, and other structural techniques enable us to obtain the three-dimensional structures of these molecular machines. Biochemistry, biophysics, and molecular biology allow us to study properties that can be correlated to their function and biology.

## High-Resolution Cryo-EM Structures of Outbreak Strain Human Norovirus

J. Jung [in collaboration with D. Thomas, CSHL; T. Grant and N. Grigorieff, Janelia Research Campus, HHMI; C. Diehnelt, Arizona State University]

Noroviruses are a leading cause of foodborne illnesses worldwide. Until this work the assembled virus shell structures have been available in detail for only a single strain (GI.1). Assembled shells of viruses without genetic materials enclosed are being used as candidates for vaccine trials. With our new capabilities of determining structures with a state-of-the-art cryo-EM microscope, we determined very-high-resolution (2.6–4.1 Å) structures of virus-like particles (VLPs) of four different strains of noroviruses that are responsible for the vast majority of outbreaks (Fig 1). Although norovirus VLPs have been thought to exist in a single-sized assembly, our structures reveal in near-atomic detail clear polymorphism between and within genogroups. We observed small, medium, and large particle sizes with different spatial and angular arrangements of the antigenic surface spikes. Using asymmetric reconstruction techniques, we were able to resolve a $Zn^{2+}$ metal ion adjacent to the coreceptor binding site, which affected the structural stability of the shell. These noroviruses cause severe gastroenteritis. They exist in three different assemblies, rather than a single one as was proposed in the past, and we have been able to resolve a previously uncharacterized metal binding site on the tip of the norovirus spike that is involved in host interactions.

## Mechanisms of RNAi and Noncoding RNAs

RNAi has made an enormous impact on biology in a very short period of time. Not only are we still discovering new cellular pathways for the regulation of gene expression that use these pathways, but RNAi has become an extraordinarily useful and simple tool for gene silencing. Almost from the beginning, people have used genetics, biochemistry, molecular biology, and bioinformatics to study the mechanism of RNAi and related pathways. To get a true mechanistic understanding of these pathways, however, we must understand how the components of the RNAi machinery work at a molecular level.

## Molecular Basis of piRNA Silencing

J. Ipsaro [in collaboration with A. Haase, NIH; M. Gale Hammell, CSHL]

The stability of germline genomes is critical to the survival of a species. If unchecked, mobile genetic elements (transposons) are able integrate into distant genomic sites and disrupt genome integrity. To protect against this, a germline-specific RNA interference pathway has evolved to specifically and robustly repress transposon expression.

At the center of this critical defense pathway are Piwi proteins and their associated small RNAs (piRNAs). piRNAs are highly variable in sequence, which affords them the ability to silence many of the molecular targets' transposon transcripts. Nonetheless, the diversity
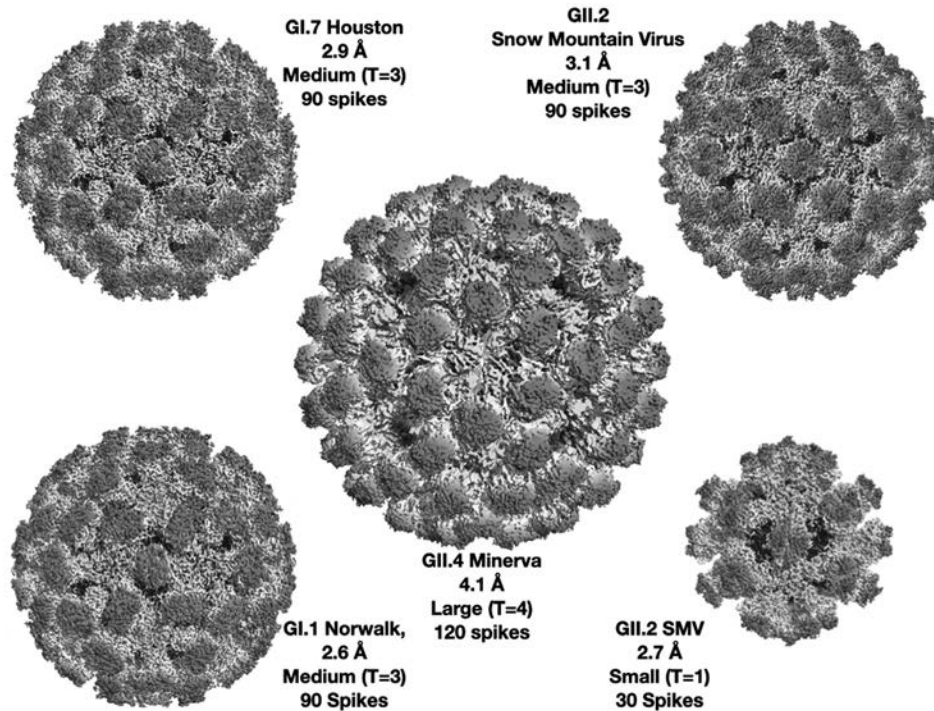
**Figure 1.** Cryo-EM structures of human norovirus outbreak strain capsids.

of Piwi-piRNAs is restricted by their preference of a uridine in the 5′-most position. The basis of this "1U" bias, particularly whether it is established by a Piwi protein or earlier in piRNA biogenesis, was unknown.

We found that the 1U bias of Piwi-piRNAs is established by sequential discrimination against all nucleotides other than U: first during piRNA biogenesis and then through interaction with Piwi's specificity loop. Additionally, sequence preferences during piRNA processing also limit uridine occurrences across the piRNA body. Together, these processes and the resulting 1U bias may modulate the efficacy of transposon silencing by piRNAs and provide a means for purifying selection in the ongoing arms race between germline genomes and their mobile genetic parasites.

## ELTA: An Enzymatic Method to Label Free or Protein-Conjugated ADP-Ribose Monomer and Polymer

E. Elkayam [in collaboration with A. Leung, Johns Hopkins]

ADP-ribosylation, the attachment of one or more ADP-ribose groups onto proteins, is a therapeutically important protein modification. The attached ADP-ribose monomer or polymer, known as PAR, modulates the activities of the modified substrates and/or their binding affinity to other proteins. However, there has been a woeful lack of tools to properly investigate these important modifications and their potential functions. In collaboration with Anthony Leung from Johns Hopkins University we developed a simple, efficient, and versatile enzymatic labeling method, named enzymatic labeling of terminal ADP-ribose (ELTA), to label free or protein-conjugated ADP-ribose monomer and polymers using the enzyme OAS1 and dATP. ELTA can be coupled with a diverse range of chemical analogs of dATP (radioactive, fluorescent, biotin-tag, clickable functional groups, and more) for various applications such as fluorescence-based biophysical measurement and biotin/click chemistry–based enrichment. Of particular importance, ELTA provides a timely tool to directly assess PAR length and distribution following treatment with FDA-approved PARP inhibitors that are now being administered for several cancer indications and others that are being tested. In addition, ELTA could be extended in the future to label other

ADP-ribose derivatives with a free 2′-OH group such as the ADP-ribosylation of the antibiotic rifamycin and the recently discovered modification of DNA ADP-ribosylation.

## PUBLICATIONS

Ando Y, Elkayam E, McPherson RL, Dasovich M, Cheng S-J, Voorneveld J, Filippov DV, Ong S-E, Joshua-Tor L, Leung AKL. 2019. ELTA: enzymatic labeling of terminal ADP-ribose. *Mol Cell* **74:** 845–856. doi:10.1016/j.molec.2018.12.022

Greider CW, Sheltzer JM, NC, Copeland WB, Dasgupta N, Hopkins N, Jansen JM, Joshua-Tor L, McDowell GS, McDowell JL, et al. 2019. Increasing gender diversity in the STEM research workforce. Policies must address harassment and bias. *Science* **366:** 692–695. doi:10.1126/science.aaz0649J

Jung J, Grant T, Thomas DR, Diehnelt CW, Grigorieff N, Joshua-Tor L. 2019. High-resolution cryo-EM structures of outbreak strain human norovirus shells reveal size variations. *Proc Am Natl Acad Sci* **116:** 12828–12832. doi/10.1073/pnas.1903562116

Stein C, Genzor P, Mitra S, Elchert AR, Benner L, Sobti S, Ipsaro JJ, Hammell M, Joshua-Tor L, Haase AD. 2019. Decoding the 5′ nucleotide bias of PIWI-interacting RNAs (piRNAs). *Nature Comm* **10:** 828. doi:10.1038/s41467-019-08803-z

*In Press*

Garg A, Liu L, Ho DD, Joshua-Tor L. 2020. Heterologous expression and purification of SARS-CoV2 nucleocapsid protein. *Bio-101*: e5005. doi:10.21769/BioProtoc.5005

# RNA SPLICING

**A.R. Krainer**
M. Daley   W.K. Ma   L. Wan
N. Garry   L. Manche   M.S. Wong
Y.J. Kim   L. Marasco   Q. Zhang
K-T. Lin   M. Rahman
Y.H. Liu   D. Segovia

## Mechanisms of Constitutive and Alternative Pre-mRNA Splicing

RNA splicing is required for expression of most eukaryotic protein-coding genes. The spliceosome selects authentic splice sites with very high fidelity, relying on limited sequence information present throughout introns and exons. In humans, >90% of genes are expressed via alternative splicing, giving rise to multiple protein isoforms. The choice of alternative splice sites is commonly regulated to alter gene expression, either tissue-specifically or in response to a developmental program or to signaling pathways. The fact that multiple protein isoforms can be expressed from individual genes demonstrates that the classical "one gene–one enzyme" paradigm is no longer valid, and provides an explanation for the unexpectedly small number of protein-coding genes uncovered by genome-sequencing projects.

Both constitutive and alternative splicing mechanisms involve numerous protein components, as well as five noncoding RNA components that are part of small nuclear ribonucleoprotein (snRNP) particles. These components are sequentially assembled with a pre-mRNA substrate into a spliceosome, which catalyzes the two transesterification steps of splicing. There are two kinds of spliceosomes, major and minor, with both distinct and shared snRNA and protein components, which process two different classes of introns. The work in our laboratory focuses on the identification and molecular characterization of protein factors and sequence elements that are necessary for the catalysis and fidelity of splicing and/or for the regulation of alternative splice-site selection. We are interested in how the spliceosome correctly identifies the exons on pre-mRNA, and how certain point mutations in either exon or intron sequences cause aberrant splicing, leading to various human genetic diseases. Related areas of interest include the remodeling of messenger RNP (mRNP) architecture as a consequence of splicing, which influences downstream events such as nonsense-mediated mRNA decay (NMD), the various roles of alternative splicing misregulation in cancer, and the development of effective methods—particularly antisense technology—to correct defective splicing or modulate alternative splicing or for gene/allele-specific inhibition of NMD in a disease context. A summary of our recently published studies is provided below.

## Experimental and Computational Analyses of Splicing

Percent spliced-in (PSI) values are commonly used to report alternative pre-mRNA splicing changes. Previous PSI-detection tools were limited to specific alternative splicing events and were evaluated by in silico RNA-seq data. We developed PSI-Sigma, a computational tool that uses a new PSI index, and we employed actual (nonsimulated) RNA-seq data from spliced synthetic gene transcripts (RNA sequins) to benchmark its performance (i.e., precision, recall, false-positive rate, and correlation) in comparison with leading tools. PSI-Sigma outperformed these tools, especially in the case of alternative splicing events with multiple alternative exons and intron-retention events. We also evaluated its performance in long-read RNA-seq analysis by sequencing a mixture of human RNAs and RNA sequins with nanopore long-read sequencers. PSI-Sigma is available at https://github.com/wososa/PSI-Sigma.

RNA-binding proteins (RBPs) regulate post-transcriptional gene expression by recognizing short and degenerate sequence motifs in their target transcripts, but precisely defining their binding specificity

remains challenging. Cross-linking and immunoprecipitation (CLIP) allows for mapping of the exact protein–RNA cross-link sites, which frequently reside at specific positions in RBP motifs. Chaolin Zhang's laboratory (Columbia University) developed mCross, a computational method to model RBP binding specificity while precisely registering the cross-linking position in motif sites. In collaboration with our laboratory, mCross was applied to 112 RBPs using ENCODE eCLIP data, and the reliability of the discovered motifs was validated by genome-wide analysis of allelic binding sites. This analysis revealed that the prototypical SR protein SRSF1 recognizes clusters of GGA half-sites, in addition to its canonical GGAGGA motif. Therefore, SRSF1 regulates splicing of a much larger repertoire of transcripts than previously appreciated, including HNRNPD and HNRNPDL, which are involved in multivalent protein assemblies and phase separation.

## Splicing and NMD Contributions to Hematological Malignancies

Oncogenic mutations in the RNA splicing factors SRSF2, SF3B1, and U2AF1 are the most frequent class of mutations in myelodysplastic syndromes (MDS) and are also common in clonal hematopoiesis, acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), and a variety of solid tumors. They cause genome-wide splicing alterations that affect important regulators of hematopoiesis. Several mRNA isoforms promoted by the various splicing factor mutants include a premature termination codon (PTC) and are therefore potential targets of NMD. We previously reported that several SR proteins stimulate NMD. In light of the mechanistic relationship between splicing and NMD, we sought evidence for a specific role of mutant SRSF2 in NMD. We found that SRSF2 Pro95 hotspot mutations elicit enhanced NMD more strongly than wild-type SRSF2, and this effect is dependent on sequence-specific RNA binding and splicing. SRSF2 mutants enhanced the deposition of exon junction complexes (EJCs) downstream from the PTC through RNA-mediated molecular interactions. The resulting mRNP architecture favors the recruitment of key NMD factors to elicit mRNA decay. This mechanism of NMD enhancement differs from the one we previously established for SRSF1,

which involves direct recruitment of the NMD RNA helicase UPF1. Gene-specific blocking of EJC deposition by antisense oligonucleotides (ASOs) circumvented aberrant NMD promoted by mutant SRSF2, restoring the expression of PTC-containing transcript. This study uncovered critical effects of *SRSF2* mutants in hematologic malignancies, reflecting the regulation at multiple levels of RNA metabolism, from splicing to decay.

Transcription and pre-mRNA splicing are key steps in the control of gene expression, and mutations in genes regulating each of these processes are common in leukemias. Despite the frequent overlap of mutations affecting epigenetic regulation and splicing in leukemias, how these processes influence one another to promote leukemogenesis is not understood, and functional evidence that mutations in RNA splicing factors initiate leukemia was lacking. By analyzing the transcriptomes from 982 patients with AML in a collaborative study with Omar Abdel-Wahab (MSKCC), we identified a frequent overlap of mutations in *IDH2* and *SRSF2* that together promote leukemogenesis through coordinated effects on the epigenome and RNA splicing. Whereas mutations in either *IDH2* or *SRSF2* imparted distinct splicing changes, coexpression of mutant *IDH2* altered the splicing effects of mutant *SRSF2* and resulted in more profound splicing changes than either mutation alone. Consistent with this finding, coexpression of mutant *IDH2* and *SRSF2* resulted in lethal myelodysplasia with proliferative features in vivo and enhanced self-renewal, which were not observed with either mutation alone. *IDH2* and *SRSF2* double-mutant cells exhibited aberrant splicing and reduced expression of *INTS3*, a member of the integrator complex, consistent with increased stalling of RNA polymerase II. Aberrant *INTS3* splicing contributed to leukemogenesis in concert with mutant *IDH2* and was dependent on mutant SRSF2 binding to *cis*-acting elements in *INTS3* mRNA, as well as increased DNA methylation of *INTS3*. These data identified a pathogenic cross talk between altered epigenetic state and splicing in a subset of leukemias, provided functional evidence that mutations in splicing factors drive myeloid malignancy development, and identified spliceosomal changes as mediators of *IDH2*-mutant leukemogenesis.

RBPs are essential modulators of transcription and translation that are frequently dysregulated in cancer. In a collaborative study with Iannis Aifantis

(NYU) and Omar Abdel-Wahab (MSKCC), RBP dependencies in human cancers were systematically interrogated using a comprehensive CRISPR-Cas9 domain-focused screen targeting RNA-binding domains of 490 classical RBPs. This analysis uncovered a network of physically interacting RBPs upregulated in AML and crucial for maintaining RNA splicing and AML survival. Genetic or pharmacologic targeting of one key member of this network, RBM39, repressed cassette exon inclusion and promoted intron retention within mRNAs encoding HOXA9 targets, as well as in other RBPs preferentially required in AML. The effects of RBM39 loss on splicing further resulted in preferential lethality of spliceosomal-mutant AML, providing a potential strategy for treatment of AML bearing splicing-factor mutations.

Patients with chronic myeloid leukemia (CML) who are treated with tyrosine kinase inhibitors (TKIs) experience significant heterogeneity in the extent and speed of responses. Factors intrinsic and extrinsic to CML cells contribute to the response heterogeneity and TKI resistance. Among extrinsic factors, cytokine-mediated TKI resistance has been demonstrated in CML progenitors, but the underlying mechanisms remain obscure. In a collaborative study with Sin Tiong Ong (Duke-NUS, Singapore), RNA sequencing was used to identify differentially expressed splicing factors in primary CD34+ chronic phase (CP) CML progenitors and controls. SRSF1 expression was elevated as a result of both BCR-ABL1- and cytokine-mediated signaling. SRSF1 overexpression conferred cytokine independence to untransformed hematopoietic cells and impaired imatinib sensitivity in CML cells, whereas SRSF1 depletion in CD34+ CP CML cells prevented the ability of extrinsic cytokines to decrease imatinib sensitivity. Mechanistically, PRKCH and PLCH1 were up-regulated by elevated SRSF1 levels and contributed to impaired imatinib sensitivity. Importantly, very high SRSF1 levels in the bone marrow of CML patients at presentation correlated with poorer clinical TKI responses. In summary, SRSF1 levels are maintained in CD34+ CP CML progenitors by cytokines, despite effective BCR-ABL1 inhibition, and elevated levels promote impaired imatinib responses. Together, these data support the contribution of an SRSF1/PRKCH/PLCH1 axis to cytokine-induced impairment of imatinib sensitivity in CML.

## Alternative Splicing as Driver and Therapeutic Target in Solid Tumors

The M2 pyruvate kinase (PKM2) isoform is up-regulated in most cancers and plays a crucial role in the Warburg effect, which is characterized by the preference for aerobic glycolysis for energy metabolism. PKM2 is an alternative-splice isoform of the *PKM* gene and is a potential therapeutic target. Previously, in collaboration with Ionis Pharmaceuticals, we developed ASOs that switch *PKM* splicing from the cancer-associated PKM2 to the PKM1 isoform and induce apoptosis in cultured glioblastoma cells. We have since explored the potential of ASO-based *PKM* splice switching as a targeted therapy for liver cancer. Normal hepatocytes express a different pyruvate kinase gene, *PKLR*, but this gene is down-regulated and *PKM* turned on in hepatocellular carcinoma (HCC) cells, resulting in PKM2 isoform expression as in other cancers. We demonstrated that our lead ASO induces *PKM* splice switching and inhibits the growth of cultured HCC cells. This *PKM* isoform switch increased pyruvate kinase activity and altered glucose metabolism in HCC cells, promoting the Krebs cycle and decreasing upstream glycolytic intermediates. The lead ASO inhibited tumorigenesis in an orthotopic xenograft HCC mouse model. Finally, a surrogate mouse-specific ASO induced *Pkm* splice switching and inhibited HCC growth, without observable toxicity, in a genetic HCC mouse model. This study demonstrated the relevance of *PKM* and the Warburg effect in HCC and the potential of antisense targeting of this metabolic switch as a therapy for liver cancer.

Misregulation of alternative splicing is a hallmark of human tumors, yet to what extent and how it contributes to malignancy are questions that are only beginning to be solved. In collaboration with former laboratory member Olga Anczuków (JAX, Farmington), we determined which members of the SR protein and SR-like splicing-factor families contribute to breast cancer, and uncovered differences and redundancies in their targets and biological functions. We identified splicing factors that are frequently altered in human breast tumors and assayed their oncogenic functions using breast organoid models. We found that not all SR and SR-like splicing factors affect mammary tumorigenesis in MCF-10A cells. Specifically, up-regulation of SRSF4, SRSF6, or TRA2β disrupted acinar morphogenesis and promoted cell

proliferation and invasion in MCF-10A cells. By characterizing the targets of these oncogenic splicing factors, we identified shared spliced isoforms associated with well-established cancer hallmarks. Finally, we demonstrated that TRA2β is transcriptionally regulated by the MYC oncoprotein, plays a role in metastasis maintenance in vivo, and its levels correlate with breast-cancer patient survival.

## Antisense Technology Development

Spinal muscular atrophy (SMA) is a motor neuron disease caused by loss-of-function mutations of the *SMN1* gene. Humans have a paralog, *SMN2*, whose exon 7 is predominantly skipped, and so *SMN2* cannot fully compensate for the lack of *SMN1*. SMA was the leading genetic cause of infant mortality until a splicing-correcting ASO drug we developed, known as nusinersen or Spinraza®, was approved for clinical use in 2016. Spinraza® targets a splicing silencer located in intron 7 of the *SMN2* pre-mRNA and, by blocking the binding of splicing repressors, it causes higher exon 7 inclusion, resulting in higher SMN protein levels. In collaboration with Alberto Kornblihtt (Universidad de Buenos Aires, Argentina), we found that fast transcriptional elongation elicited by histone deacetylase inhibitors promotes *SMN2* exon 7 inclusion. These drugs acted synergistically with a Spinraza®-like ASO to promote further exon 7 inclusion. Surprisingly, the ASO also elicited the deployment of the silencing histone mark H3K9me2 around its target site in the *SMN2* gene, creating a roadblock for transcriptional elongation that acts negatively on exon 7 inclusion. By removing the roadblock, HDAC inhibition counteracted the untoward chromatin effects of the ASO, resulting in significantly higher exon 7 inclusion. Combined systemic administration of the Spinraza®-like ASO and HDAC inhibitors in neonate SMA mice had strong synergistic effects on SMN expression, growth, survival, and neuromuscular function. Thus, HDAC inhibitors have the potential to increase the clinical efficacy of Spinraza®, and perhaps other splicing-modulatory ASO drugs.

Splice-switching ASOs, which bind specific RNA-target sequences and modulate pre-mRNA splicing by sterically blocking the binding of splicing factors to the pre-mRNA, are promising therapeutic tools to target various genetic diseases, including cancer. However, in vivo delivery of ASOs to orthotopic tumors in cancer mouse models, or to certain target tissues, remains challenging. A viable solution already in use is receptor-mediated uptake of ASOs via tissue-specific receptors. For example, the asialoglycoprotein receptor (ASGP-R) is exclusively expressed in hepatocytes. Triantennary *N*-acetylgalactosamine (GalNAc) (GN3)-conjugated ASOs bind to the receptor and are efficiently internalized by endocytosis, enhancing ASO potency in the liver. In collaboration with Ionis Pharmaceuticals, we explored the use of GalNAc-mediated targeting to deliver therapeutic splice-switching ASOs to cancer cells that ectopically express ASGP-R, both in vitro and in tumor mouse models. We found that ectopic expression of the major isoform ASGP-R1 H1a is sufficient to promote uptake and increase GN3-ASO potency to various degrees in four of five tested cancer cell lines. We also showed that cell type–specific glycosylation of the receptor does not affect its activity. Finally, we showed that, in vivo, GN3-conjugated ASOs specifically target subcutaneous xenograft tumors that ectopically express ASGP-R1 and modulate splicing significantly more strongly than unconjugated ASOs. These results demonstrate that GN3 targeting is a useful tool for proof-of-principle studies in orthotopic cancer models until endogenous receptors are identified and exploited for efficiently targeting cancer cells.

Splice-switching ASOs are typically 15–25-nt-long and considered to be highly specific toward their intended target sequence, typically elements that control exon definition and/or splice-site recognition. However, whether or not splice-modulating ASOs also induce hybridization-dependent mis-splicing of unintended targets had not been systematically studied. We tested the in vitro effects of splice-modulating ASOs on 108 potential off-targets predicted on the basis of sequence complementarity and identified 17 mis-splicing events for one of the ASOs tested. Based on analysis of data from two overlapping ASO sequences, we determined that off-target effects are difficult to predict, and the choice of ASO chemistry influences the extent of off-target activity. The off-target events caused by the uniformly modified ASOs we tested were significantly reduced with mixed-chemistry ASOs of the same sequence. Furthermore, using shorter ASOs, combining two ASOs, and delivering ASOs by free uptake also reduced off-target activity. Finally, we showed that ASOs with strategically placed mismatches can be used to reduce unwanted off-target splicing events.

Based on prior preclinical studies with SMA mouse models, both 2′-*O*-methoxyethyl (MOE) with a phosphorothioate backbone and morpholino with a phosphorodiamidate backbone—with the same or extended target sequence as nusinersen—displayed efficient rescue, but they had not been rigorously compared. In collaboration with former laboratory member Yimin Hua (Nanjing Normal University, China), we compared the therapeutic efficacy of these two modification chemistries in rescue of a severe SMA mouse model, using ASO10-29—a 2-nt-longer version of nusinersen—via subcutaneous injection in newborn mice. Although both chemistries efficiently corrected *SMN2* splicing in various tissues, restored motor function, and improved the integrity of neuromuscular junctions, MOE-modified ASO10-29 (MOE10-29) was more efficacious than morpholino-modified ASO10-29 (PMO10-29) at the same molar dose, as seen by longer survival, greater body-weight gain, and better preservation of motor neurons. Time-course analysis revealed that MOE10-29 had more persistent effects than PMO10-29. On the other hand, PMO10-29 appeared to more readily cross an immature blood–brain barrier following systemic administration, showing more robust initial effects on *SMN2* exon 7 inclusion but less persistence in the central nervous system. We conclude that both modifications can be effective as splice-switching ASOs in the context of SMA and potentially other diseases.

## PUBLICATIONS

Bennett CF, Krainer AR, Cleveland DW. 2019. Antisense oligonucleotide therapies for neurodegenerative diseases. *Annu Rev Neurosci* **42:** 385–406.

Feng H, Bao S, Rahman MA, Weyn-Vanhentenryck SM, Khan A, Wong J, Shah A, Flynn ED, Krainer AR, Zhang C. 2019. Modeling RNA-binding protein specificity in vivo by precisely registering protein-RNA crosslink sites. *Mol Cell* **74:** 1189–1204.

Lin K-T, Krainer AR. 2019. PSI-Sigma: a comprehensive splicing-detection method for short-read and long-read RNA-seq analysis. *Bioinformatics* **35:** 5048–5054.

Park SH, Brugiolo M, Akerman M, Das S, Urbanski L, Geier A, Kesarwani AK, Fan M, Leclair N, Lin K-T, et al. 2019. Differential functions of splicing factors in mammary transformation and breast cancer metastasis. *Cell Rep* **29:** 2672–2688.e7.

Scharner J, Ma WK, Zhang Q, Lin K-T, Rigo F, Bennett CF, Krainer AR. 2019. Hybridization-mediated off-target effects of splice-switching antisense oligonucleotides. *Nucleic Acids Res* **48:** 802–816.

Wang E, Pastore A, Lu SX, Chen X, Imig J, Ghebrechristos YE, Yoshimi A, Bitner L, Kloetgen A, Lin K-T, et al. 2019. Targeting and RNA-binding protein splicing network in acute myeloid leukemia. *Cancer Cell* **35:** 369–384.

Yoshimi A, Lin KT, Wiseman DH, Rahman MA, Pastore A, Wang B, Lee SC, Micol JB, Zhang XJ, de Botton S, et al. 2019. Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis. *Nature* **574:** 273–277.

*In Press*

Rahman MA, Krainer AR, Abdel-Wahab O. 2020. SnapShot: splicing alterations in cancer. *Cell* **180:** 208.e1.

Rahman MA, Lin K-T, Bradley RK, Abdel-Wahab O, Krainer AR. 2020. Recurrent SRSF2 mutations in MDS affect both splicing and NMD. *Genes Dev* **34:** 413–427.

Scharner J, Qi S, Rigo F, Bennett CF, Krainer AR. 2020. Delivery of GalNac-conjugated splice-switching ASOs to non-hepatic cells through ectopic expression of asialoglycoprotein receptor. *Mol Ther Nucleic Acids* **16:** 313–325.

Sheng L, Rigo F, Bennett CF, Krainer AR, Hua Y. 2020. Comparison of the efficacy of MOE and PMO modifications of systemic antisense oligonucleotides in a severe SMA mouse model. *Nucleic Acids Res* **48:** 2853–2865.

Sinnakannu JR, Lee KL, Cheng S, Li J, Yu M, Tan SP, Ong CCH, Li H, Than H, Anczuków-Camarda O, et al. 2020. SRSF1 mediates cytokine-induced impaired imatinib sensitivity in chronic myeloid leukemia. *Leukemia* **34:** 1787–1798.

# REGULATION OF GENE EXPRESSION

**D.L. Spector**  D. Aggarwal   S. Bhatia   K-C. Chang   B. Liu   S. Russo
                  B. Balasooriya  L. Brine   R. Hazra     P. Naik  W. Xu

Most cellular processes can trace their beginnings to the nucleus, where a gene is activated resulting in the production of an RNA molecule, some of which encode proteins, whereas others function as non-protein-coding RNAs. Although much biochemical information is available regarding many of the factors involved in gene expression, the spatial and temporal parameters that influence gene expression, and the role of noncoding RNAs in regulating this multifaceted process, are just beginning to be elucidated. Over the past year, our research has continued to focus on identifying and characterizing the role of long noncoding RNAs (lncRNAs) in breast cancer progression and/or differentiation. In addition, we continue to examine the role of lineage commitment in establishing random autosomal monoallelic (RAM) gene expression, and the role of lncRNAs in lineage commitment. Following is an overview of some of our accomplishments over the past year.

## Identification of lncRNAs Involved in Breast Cancer Progression

S. Bhatia, D. Aggarwal, K-C. Chang, B. Liu, W. Xu, S. Russo, L. Brine [in collaboration with F. Rigo, R. MacLeod, and C. Frank Bennett, Ionis Pharmaceuticals]

Large-scale genome-wide studies have revealed that thousands of RNAs that lack protein-coding capacity are transcribed from mammalian genomes. A subset of these noncoding RNAs are greater than 200 nucleotides in length and are referred to as lncRNAs. With breast cancer being the most frequent malignancy in women worldwide, we aim to identify lncRNAs that play roles in breast cancer progression and evaluate their mechanism of action and potential as therapeutic targets.

*MALAT1* is a nuclear-localized, highly abundant lncRNA that is ubiquitously expressed in mammalian cells and tissues. Overexpression of *MALAT1* has been demonstrated in more than 20 different cancer types, including breast cancer, and as such has been proposed to be a strong prognostic marker and potential therapeutic target. We have previously shown that *Malat1* knockout (KO) in the MMTV-PyMT mammary tumor model resulted in highly differentiated primary tumors and ~70% reduction in metastasis to the lungs. Similar findings were observed upon antisense oligonucleotide (ASO)-mediated knockdown (KD) of *Malat1* in MMTV-PyMT mice.

Over the past year, we have utilized patient-derived breast tumor organoid models to study the effect of *MALAT1* perturbation on breast cancer and to further study its mechanism of action. These organoid models are more representative of the patient diversity and subtype complexity characteristic of breast cancer when compared to established cell lines or genetically engineered mouse models. Additionally, as they are cultured three-dimensionally, they are better models for a disease like cancer where cell–cell and cell–extracellular matrix (ECM) interactions play an important role in progression and molecular pathogenesis. In collaboration with Ionis Pharmaceuticals, we have identified two ASO candidates that are the most potent at knocking down *MALAT1* in the patient-derived breast tumor organoid model system. Thus far, we completed an analysis of *MALAT1* KD efficiency in 10 organoid lines, and we aim to assess more than 20 lines. We have observed differences in KD efficiency among the different patient lines and are investigating multiple plausible causes for the observed differences. We are examining whether there is a correlation between (a) tumor subtype, (b) grade, (c) *MALAT1* expression level, or (d) RNase H1 expression level to the observed variance in *MALAT1* ASO KD efficiency.

In addition, we are performing RNA-seq to identify differentially expressed genes upon *MALAT1* KD versus untreated organoids at an early time point of KD to identify the immediate downstream targets of *MALAT1*. Proliferation and migration assays are being carried out to study the impact of *MALAT1* KD on these processes. We are also assessing *MALAT1*–chromatin interactions and the composition of the *MALAT1* ribonucleoprotein (RNP).

In addition to our studies on *MALAT1*, we previously identified 30 potentially oncogenic lncRNAs, termed mammary tumor–associated RNAs (*MaTARs*), in mouse models of human breast cancer. Our results suggest that *MaTARs* are likely important drivers of mammary tumor progression and represent promising new therapeutic targets. Over the past year, we have continued to focus on *MaTAR25*, which is an ~2,000 nucleotide nuclear-enriched and chromatin-associated lncRNA. *MaTAR25* is overexpressed in mammary tumors in the MMTV-PyMT (luminal B subtype), MMTV-NEU-NDL (HER2 subtype), as well as in triple-negative mammary cancer (TNMC) cells compared to normal mammary (NM) epithelial cells. Using CRISPR-Cas9 genome editing, we previously generated *MaTAR25* KO clones in TNMC 4T1 cells. When comparing *MaTAR25* KO cells to control 4T1 cells, we found a significant 50% decrease in cell proliferation, a 40% decrease in cell motility, and a 45% decrease in invasion ability. Ectopic expression of *MaTAR25* in 4T1 *MaTAR25* KO cells rescued the KO phenotypes, indicating that *MaTAR25* lncRNA plays an important role in these processes. Injection of *MaTAR25* 4T1 KO cells into the mammary fat pad or tail vein of BALB/c mice resulted in a 56% decrease in tumor growth and a 62% decrease in the number of lung metastatic nodules compared to the 4T1 control group. Furthermore, ASO-mediated KD of *MaTAR25* in the MMTV-Neu-NDL mouse model resulted in a 59% decrease in tumor growth compared to the scrambled ASO control group.

Over the past year, we used chromatin isolation by RNA purification and DNA sequencing (ChIRP-seq) to identify the gene interactors of *MaTAR25*, and *Tensin1* (*Tns1*) was identified as the top gene candidate. Importantly, *MaTAR25* KO results in a parallel reduction of *Tns1* mRNA and protein levels, which can be rescued by transient expression of *MaTAR25*. The *Tns1* gene encodes a protein that localizes to focal adhesions and positively regulates cell migration and invasion. Interestingly, high expression of *TNS1* correlates with poor survival of grade 3 human breast cancer patients based on Kaplan–Meier analysis. Next, we performed CRISPR-Cas9 KO of the *Tns1* gene in mouse 4T1 TNMC cells and selected *Tns1* KO clones for in vitro functional assays. We found that the *Tns1* KO cells phenocopied the *MaTAR25* KO cells, exhibiting a significant decrease in cell viability (40% less than control cells) and a decrease in cell migration (30% less than control cells). In addition, ectopic expression of *Tns1* in 4T1 *MaTAR25* KO cells rescued the cell viability phenotype. Together, these data indicated that *Tns1* is a major downstream target of *MaTAR25*.

Because *Tns1* is a key component of focal adhesion complexes and is responsible for cell–cell and cell–matrix interactions as well as cell migration by interacting with actin filaments, we examined the organization of actin filaments as well as the patterns of other focal adhesion complex components paxillin and vinculin in 4T1 control and *MaTAR25* KO cells by immunofluorescence (IF) confocal microscopy. Indeed, F-actin microfilament bundles were disrupted and the distribution of paxillin and vinculin proteins was altered dramatically in *MaTAR25* KO cells as compared to 4T1 control cells. In addition, using transmission electron microscopy, in collaboration with the Microscopy Shared Resource, we identified a significant 81% decrease in microvilli over the cell surface of *MaTAR25* KO cells compared to 4T1 control cells. Interestingly, ectopic expression of either *MaTAR25* or *Tns1* in 4T1 *MaTAR25* KO cells rescued the actin filament phenotype, further supporting our finding that *Tns1* is a critical downstream target of *MaTAR25* regulating mammary tumor progression.

*MaTAR42* was found to be up-regulated in MMTV-PyMT and MMTV-Neu-NDL mammary tumors compared to NM epithelial cells. The *MaTAR42* gene is located on mouse chromosome 4, with a conserved human ortholog identified on human chromosome 9. RACE, qRT-PCR, and northern blot analyses verified that the major transcript expressed in MMTV-PyMT mouse mammary tumor cells is a 2,821-nucleotide, single-exon transcript with low protein coding potential as indicated by computational prediction, and no peptide was generated in an in vitro transcription/translational assay. ASO-mediated KD of *MaTAR42* in MMTV-PyMT-derived mammary tumor organoids reduces cell proliferation and the branching phenotype. Functional studies in MMTV-PyMT mammary tumor-derived cell lines and 4T1 TNMC cells revealed that *MaTAR42* promotes cellular adhesion, migration, and invasion. Mechanistically, *MaTAR42* promotes epithelial-to-mesenchymal transition (EMT) of tumor cells by sensitizing both TGF-β-induced EMT and spontaneous EMT in a 3D culture

system. Combined fluorescence in situ hybridization and IF microscopy revealed that *MaTAR42* expression negatively correlates with expression of the epithelial marker E-cadherin in MMTV-PyMT tumors and mammospheres grown in 3D culture. Interestingly, *MaTAR42* expression is absent in cells expressing the mesenchymal marker vimentin, indicating a potential bipolar regulatory role for *MaTAR42*. Importantly, the human *MaTAR42* ortholog is found to be up-regulated in several TNBC organoid lines. These results suggest that *MaTAR42* is a lncRNA conserved between mouse and human, which may play an important role in EMT and metastasis and thus may serve as an interesting therapeutic target in breast cancer.

TNBC represents 20% of all breast cancer cases and is one of the more aggressive subtypes. TNBCs are characterized by the lack of hormone receptors and the Her2 receptor. As a result, the patients undergo generic chemotherapy treatment, which is harmful, unselective, and results in a poor overall survival and greater chance of relapse. The TNBC subtype is very heterogeneous depending on the underlying molecular signatures of various patients. Our goal is to use patient-derived TNBC organoids to identify specific cancer-associated lncRNA targets, develop new therapeutics, and assess the usage of current therapeutics.

Thus far, we have derived ex vivo organoids from 28 TNBC patient tumors. Most of these samples are from aggressive high-grade tumors (22/28 grade 3, 2/28 grade 2, 1/28 DCIS, 2/28 NA). In addition to the tumor organoids, we have derived organoids from the adjacent normal tissue of these tumors and are using them to do analysis in a patient-specific manner. We also have evaluated nine tissue scrapings from TNBC samples to test whether low amounts of material recovered from tumor scrapings is sufficient to generate organoids. We have successfully generated organoids from 5/9 scrapings, while 3/9 are still in culture. In addition to the TNBC tumor tissue and matched normals, we also derived 10 NM organoids from tissue collected from patients with no history of breast cancer who are undergoing reductive mammoplasty.

We have performed copy number analysis on 6/28 TNBC-derived organoids and targeted gene panel sequencing on 16/28 TNBC-derived organoids to identify the cancer driver mutations in these lines. We have also performed RNA sequencing on 12 of the TNBC organoids in addition to sequencing five NM organoids. We have used the mutation data in concert with the RNA-seq data to identify genes that are significantly and selectively enriched in TNBC organoids compared to NM organoids with a given genomic background. We are currently using this data to validate potential lncRNA targets, which upon KD might result in a reduced tumor organoid growth. We have also performed single-cell RNA-seq on four TNBC and three NM organoids to identify specific cell types in the two organoid types and identify pathways that are misregulated in tumor versus normal cells. Furthermore, we are using the organoid system to perform drug screens with current FDA-approved and experimental drugs to correlate the genomic and transcriptomic profiles of these organoids with their drug response. We have currently performed drug screens on 11 TNBC organoid lines with a panel of 16 drugs.

## Human Breast Tumor Organoid Project

S. Bhatia, S. Russo, P. Naik [in collaboration with K. Kostroff and T. Bhuiya, Northwell Health]

Tumor organoids provide a very innovative and unique platform to study cancer, as they can recapitulate many aspects of the disease with high fidelity. As such, they represent an excellent system for identifying new therapeutic targets and for drug development and screening in a patient-specific manner. Our goal has been to develop a human breast tumor organoid biobank that can be used to study the role of lncRNAs in breast cancer progression and establish a therapeutic platform for rapid screening of cancer-relevant lncRNAs.

We have established 50 breast tumor organoid models along with adjacent or distal normal organoids from a subset of patients. Additionally, we have established NM organoids from individuals who have undergone reductive mammoplasty. We have performed DNA sequencing to identify potential mutations from a panel of 143 pan-cancer driver genes. We currently have mutation data on 23 of the patient-derived tumor organoids. Additionally, to test whether the ex vivo–derived tumor organoids retain the mutation profile of the original tumor, we have sequenced seven tumor tissue and tumor organoid pairs. This data shows that the majority of the mutations of the

original tumor were retained in the cultured organoid system. Furthermore, we find that about 36% of the sequenced tumor organoids have oncogenic mutations in the gene *PIK3CA*, while 14% have mutations in cadherin genes (*CDH1*). Interestingly, all the samples with a *CDH1* mutation belong to the organoids derived from lobular carcinomas and have a loosely organized morphology in culture. In addition to mutation profiling, thus far we have performed copy number analysis on 10 samples and have identified genomic regions of hotspot amplifications and deletions such as in Chr1q and Chr16q. These data confirm that the breast tumor organoid system faithfully recapitulates the tumor properties on a genomic level and thus can be used to study various aspects of tumor growth and progression.

RNA sequencing has been performed on 35 breast tumor organoids and 21 normal organoids. The data shows very high patient specificity in the organoid profiles, where the normal and tumor organoids from the same patient cluster closely together, providing strong evidence for studying breast tumors in a patient-specific manner. We currently have ongoing projects that are using this patient-specific approach to identify and study lncRNAs up-regulated in the luminal A, lobular, and TNBC subtypes.

In addition, we have characterized various organoids based on their morphology and find that breast tumor organoids are extremely heterogeneous and can exhibit a dense cluster, a loosely held cluster, or a more cystic morphology. Additionally, we have performed high-resolution electron microscopy (EM) on tumor and normal organoids, and find that the normal organoids are very organized and exhibit microvilli-expressing luminal cells and flat basal myoepithelial-like cells; however, the tumor organoids do not exhibit such features. We also have performed some pilot single-cell RNA-seq analysis on tumor and normal organoids to identify the cell types present in these cultures and identify potentially targetable cell type(s) that are tumor specific. We are currently expanding this analysis to multiple tumor and normal organoid lines. Using this analysis, we find that normal organoids have a plethora of cells typically present in the mammary gland including keratin 5– and 14–expressing myoepithelial cells and keratin 8– and 18–expressing luminal cells. The tumor organoids are less heterogeneous and lack these well differentiated cell types present within the normal organoids supporting

the EM data. Additionally, using a panel of stem cell markers, CD24, CD44, and ALDH1A1, we have identified organoid lines that express these markers, and we are currently identifying lncRNAs specifically expressed in these potential stem cell types and evaluating whether they are targetable.

## The Role of the Long Noncoding RNA *Platr4* in Lineage Commitment
R. Hazra, L. Brine

lncRNAs are differentially expressed in a development-specific manner across tissues, suggesting regulatory roles in cell fate decision and differentiation. We have identified the functional role of a novel embryonic stem cell (ESC)-specific lncRNA, *Platr4* (pluripotency-associated transcript 4), in cell fate determination. *Platr4* is a 1,035-nucleotide poly(A)+ transcript comprising two exons. Cellular fractionation of mouse ESCs indicates that *Platr4* is mainly present in the nuclear fraction and is associated with chromatin. Using CRISPR-Cas9 genome editing, we have generated mouse ESC (mESC) lines (V6.5 and AB2.2) with deletion of the *Platr4* promoter, resulting in a significant depletion (homozygous deletion, *Platr4*-KO) of the *Platr4* transcript (up to 99%) as measured by qRT-PCR and single-molecule RNA-FISH analysis. Deletion of *Platr4* in mESCs did not affect cell-cycle kinetics or pluripotency, whereas we identified abnormalities in the spontaneous contraction of embryoid bodies (EBs) compared to control cells, which was further confirmed by a decreased level of the major contractile protein transcripts myosin (*myh7b*) and troponin (*Tnnt3*) during ESC differentiation. Further, gene set enrichment analysis has predicted that *Platr4* depletion in ESCs affects mesendoderm (ME) lineage commitment, which was further validated by measuring ME transcripts Sox17, Foxa2, Bra(T), and Eomes.

To determine the expression of *Platr4* in vivo, we have performed single-molecule RNA-FISH in developing mouse embryos (from embryonic day E3.5 to E12.5) and revealed that it is highly expressed at earlier stages, E3.5 to E10.5, and then significantly decreased at E12.5. In addition, we have generated a *Platr4*-KO mouse model using CRISPR-Cas9 technology and the initial phenotypic characterization of the KO mouse is in progress.

To understand the molecular mechanism of *Platr4* function, iRegulon analysis (using differentially expressed genes from RNA-seq data) was performed and predicted the Tead transcription factor, which is one of the known coactivators of ME specification, as a potential regulator of *Platr4*. Further, gene ontology analysis using differentially expressed genes from RNA-seq data predicted a significant enrichment of the Hippo signaling pathway and, interestingly, Tead is an important component of this pathway. Therefore, we have used Tead4 siRNA KD to evaluate its potential role in the regulation of *Platr4* in ESCs. We have found that *Platr4* KO or Tead4 KD did not alter the Tead4 or *Platr4* expression level, respectively. However, Ctgf (connective tissue growth factor), a direct target of Tead4, was down-regulated in both Tead4 KD and *Platr4* KO ESCs. Ongoing experiments are focused on verifying this result. Together, these findings indicate that *Platr4* may be involved in the Hippo pathway by targeting the Ctgf–Tead4 axis, an important regulator of ME specification during mammalian development.

## Random Autosomal Monoallelic Gene Expression and Differentiation

B. Balasooriya

RAM gene expression may partly explain the variable penetrance of disease-associated mutations, and hence this outcome of gene expression warrants extensive investigation in terms of how it is initiated, maintained, and regulated among a clonal population of cells. We previously developed allele-specific live cell imaging systems for two RAM genes (*Cth* and *Ttc4*). Using these systems, we observed allelic pairing and localization at the nuclear periphery when the genes are biallelically expressed in mESCs. However, upon differentiation of the mESCs to neural progenitor cells (NPCs), they are monoallelically expressed and only the active allele localizes to the nuclear periphery. Over the past year, we examined the localization of a cohort of known RAM genes (13), non-RAM NPC-specific genes (6), and monoallelic (imprinted) genes (12) in mESC-derived NPCs using nascent RNA-FISH. Surprisingly, we observed that active alleles of all the tested RAM and monoallelic genes were localized at the nuclear periphery in more than 85% of the cells counted (100

in total) for each gene. We then performed IF labeling for lamin B1 and pSer2-poll2, the transcriptionally active form of poll2, to investigate whether the alleles are actively transcribed at the nuclear periphery. We observed the co-localized labeling of pSer2-poll2 with lamin B1 along the nuclear periphery, suggesting the possibility of active transcription at the nuclear periphery. Following our RNA-FISH and IF observations, we set out to identify the molecular "hook" that anchors the active RAM and monoallelic alleles to the nuclear periphery. First, we performed sequential ChIP-Seq with lamin B1 antibody followed by pSer2-poll2 antibody in C57BL6J and CASTEiJ hybrid (F1) mESC and in vitro–derived F1 NPCs and observed the cohort of expressed genes in the lamin B1-bound fraction. We are in the process of accessing the monoallelically expressed genes in this ChIP-seq data set using an allele-specific ChIP-seq analysis pipeline. The expressed genes (from ChIP analysis) in the lamin B1 compartment will be confirmed by RNA-seq. However, the overlap of RAM and monoallelic genes and the genes enriched by ChIP-seq is low (less than 30%), indicating that lamin B1 may not be the molecular hook that holds RAM and monoallelic genes at the nuclear periphery. Based on the DNA sequences of the RAM genes in NPCs and through literature mining, we identified Sun1 as a potential candidate as the molecular hook for the RAM and monoallelic genes. To test Sun1 as a potential candidate, we performed ChIP-seq experiments for Sun1 in F1 mESCs and F1 mESC-derived NPCs and then performed mass spectrometry (MS) analysis. In MS analysis data, we found 713 highly enriched Sun1 coimmunoprecipitated protein candidates and 78 highly enriched proteins in nucleic acid–depleted ChIP-seq samples in F1 mESCs and F1 NPCs. Within the highly enriched protein set, we identified proteins that are known to be involved in phase separation, DNA binding, and nucleic acid modification (enzymes). From the ChIP-seq analysis, we found a highly enriched DNA sequence motif in the Sun1 bond chromatin fractions. We are currently investigating whether Sun1 binds to DNA directly or via another candidate in the Sun1-bound protein complex using biotin-labeled motif DNA pull-down experiments. Further, we are also evaluating the ChIP-seq data for allele-specific gene enrichment in Sun1-bound chromatin using an allele-specific ChIP-seq analysis pipeline.

## PUBLICATIONS

Abulwerdi FA, Xu W, Ageeli A, Yonkunas MJ, Arun G, Nam H, Schneekloth JS Jr, Dayie TK, Spector D, Baird N, et al. 2019. Selective small-molecule targeting of a triple helix encoded by the long non-coding RNA, *MALAT1. ACS Chemical Biol* **14:** 223–235.

Arun G, Spector DL. 2019. *MALAT1* long non-coding RNA and breast cancer. *RNA Biol* **16:** 860–863.

*In Press*

Chang K-C, Diermeier SD, Yu AT, Brine LD, Russo S, Bhatia S, Alsudani H, Kostroff K, Bhuiya T, Brogi E, et al. 2020. *MaTAR25* lncRNA regulates the *Tensin1* gene to impact breast cancer progression. bioRxiv doi:10.1101/2020.02.03.931881

# DNA REPLICATION AND CHROMATIN INHERITANCE

**B. Stillman**   K. Bhalla      M. Hossain   F. Nasrin      J. Shapp    M. Tramantano
                  H-C. Chou     Y. Hu        M. Quyang    Y-J. Sheu   N. Zali

Since 2004, we have used cryo-electron microscopy to understand the mechanism of the initiation of DNA replication in eukaryotes, using the yeast *Saccharomyces cerevisiae* as the core model, but also comparing the biochemistry and structural biology of DNA replication in human cells. These studies were initiated by Christian Speck (Imperial College London) when he was a postdoctoral fellow at Cold Spring Harbor Laboratory in collaboration with Huilin Li (Van Andel Institute), a structural biologist—this collaboration has been ongoing for 15 years. In addition, to study the structure and function of the origin recognition complex (ORC) in human cells, we have collaborated with Leemor Joshua-Tor's laboratory. Recently, in the last three to four years, improvements in both computational analysis and detection of particles in the electron microscope greatly improved resolution of the structures. The structures of a number of protein complexes have been determined, particularly intermediates in the assembly of pre-replicative complexes (pre-RCs) that license origins of DNA replication prior to the start of S phase. Pre-RC assembly requires the ORC, Cdc6, Cdt1, and the Mcm-7 hexamer that is loaded onto origin DNA as a double hexamer (DH) of Mcm2-7, with each Mcm2-7 hexamer destined to become a DNA helicase at the divergent replication forks following engagement with other proteins (GINS and Cdc45). Low-resolution structures of yeast proteins, including ORC on DNA, ORC-Cdc6 on DNA, the ORC-Cdc6-Cdt1-Mcm2-7 (OCCM), and the ORC-Mcm2-7-Mcm2-7 complex containing ORC and the Mcm2-7 DH were determined. More recently, high-resolution structures of the OCCM on DNA and the Mcm2-7 DH bound onto origin DNA were determined and these higher-resolution structures enabled us to predict how origins of DNA replication are located in the yeast genome and test the predictions with experimental analyses. We have also collaborated with Leemor Joshua-Tor and colleagues to determine the structure of ORC from human cells, including different dynamic forms of ORC.

The yeast OCCM and the Mcm2-7 DH both bound to origin DNA have been determined at high (~3.4 Å–3.9 Å) resolution. The OCCM structure suggested that ORC and Cdc6 likely determine DNA replication origin sequence specificity in *S. cerevisiae* via an 11-amino-acid α-helix in the AAA+ domain of the Orc4 subunit of ORC and a loop in the Orc2 protein, both of which are absent in the human and *Drosophila* ORC structures that we and others have determined. More recent HsORC structures determined in Leemor-Joshua-Tor's laboratory have confirmed that the ORC4 α-helix is absent, forming only a loop that could not insert into the major groove of DNA like the yeast Orc4 α-helix does. Structure-guided amino acid mutations in the Orc4 α-helix and the Orc2 loop were constructed and a series of yeast strains harboring these mutant ORC subunits were created. Some of the resulting mutations were lethal, but others caused a dramatic change in the distribution of origins of DNA replication throughout the genome. The origins that are utilized in early S phase in these mutants were compared to the origins used early in S phase in wild-type cells. The results showed that the Orc4 α-helix is a major determinant of why *S. cerevisiae* has DNA sequence–specific origins of DNA replication, whereas species that lack the Orc4 α-helix lack origin sequence specificity (such as human ORC).

To test this idea rigorously, in collaboration with Justin Kinney, massive parallel section assays were performed with origins of DNA replication that were heavily mutated over a 150-base-pair region (15% base substitution at each nucleotide position). This created a library of many billions of potential origins and they were subjected to selection for functional origins in each of the mutant Orc4 α-helix strains (a total of nine strains were analyzed). The resulting functional origins that were selected in each strain were sequenced and aligned using a custom-designed mutual information maximization software that was based on a new statistical algorithm developed by our quantitative biology colleagues Justin Kinney and Amaar Tareen. This software enabled the statistically robust analysis

of the mutant libraries and the recognition that some mutations caused specific base changes in the selected origins of DNA replication. Three mutants analyzed by this method changed specific base pairs in a region of the origin DNA that was predicted in the structure to have base-specific interactions with Orc4 α-helix amino acids. This was of interest because the structures of ORC on DNA determined by us and by Bik Tye (Hong Kong) and Ning Goa's (Beijing) groups show that a tyrosine at amino acid 486 interacts precisely with the AG dinucleotide in the wild-type origins, and this is the dinucleotide that is changed in the three mutants studied. Interestingly, these amino acid changes convert amino acids in *S. cerevisiae* Orc4 to the amino acids present in *Kluyveromyces lactis* Orc4, and *K. lactis* has an origin sequence that differs from the sequence found in *S. cerevisiae*. These studies support the conclusion from whole genome analysis of the distribution of origins in the genome that Orc4 and Orc2 subunits of ORC are major determinants of DNA replication origin specificity.

The observation that the Orc2 and Orc4 subunits have evolved to acquire the α-helix and loop that interact with origin DNA and that this correlates with DNA sequence-specific origins of DNA replication led to another remarkable observation. These determinants of origin specificity coevolved with the acquisition of ORC-Sir4-mediated transcriptional gene silencing and the loss of RNA interference-mediated transcriptional gene silencing (Fig. 1). Most eukaryotes have an RNAi-mediated mechanism to keep the majority of the genome transcriptionally silent and to maintain the integrity of regions of the genome that have a high density of repeated DNA sequences (representative species are shown in the box outlined in the black rectangle in Fig. 1). We found soon after ORC was discovered that, in addition to binding to origins of DNA replication, it binds to genetic DNA elements that are adjacent to the silent mating type genes in yeast that are normally kept transcriptionally silent via an epigenetically inherited mechanism. ORC binds to silent information regulator (SIR) proteins and silences gene transcription at the silent mating type loci in *S. cerevisiae* and also at the rDNA loci and telomeres that harbor repeated DNA sequences. This mechanism of transcriptional gene silencing also occurs in species that are highlighted in the light gray box in

| | Orc4 α-helix | Orc2 loop | Defined Origin | Orc1 | Sir3 | Sir2 | Hst1 | Sir1 | Sir4 | Esc1 | Dcr | Ago | RdRP | HP1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Saccharomyces cerevisiae* | + | + | + | + | + | + | + | + | + | + | - | - | - | - |
| *Saccharomyces bayanus* | + | + | + | + | + | + | + | + | + | + | * | - | - | - |
| *Candida glabrata* | + | + | + | + | + | + | + | - | + | + | - | - | - | - |
| *Naumovozyma castellii* | + | + | + | + | + | + | + | + | + | + | * | + | - | - |
| *Torulaspora delbrueckii* | + | + | + | + | - | + | - | - | + | + | * | + | + | - |
| *Zygosaccharomyces rouxii* | + | + | + | + | - | + | - | + | + | + | - | - | - | - |
| *Lachancea kluyveri* | + | + | + | + | - | + | - | - | + | + | - | - | - | - |
| *Lachancea waltii* | + | + | + | + | - | + | - | - | + | + | - | - | - | - |
| *Ashbya gossypii* | + | + | + | + | - | + | - | - | + | + | - | - | - | - |
| *Kluyveromyces lactis* | + | + | + | + | - | + | - | - | + | + | - | - | - | - |
| *Candida albicans* | - | - | - | + | - | + | - | - | - | - | * | + | - | - |
| *Debaryomyces hansenii* | - | - | - | + | - | + | - | - | - | - | * | - | - | - |
| *Pichia pastoris* | - | - | - | + | - | + | - | - | - | - | * | - | - | - |
| *Yarrowia lipolytica* | - | - | - | + | - | + | - | - | - | - | - | - | - | - |
| *Neurospora crassa* | - | - | - | + | - | + | - | - | - | - | + | + | + | + |
| *Schizosaccharomyces pombe* | - | - | - | + | - | + | - | - | - | - | + | + | + | + |
| *Drosophila melanogaster* | - | - | - | + | - | + | - | - | - | - | + | + | + | + |
| *Homo sapiens* | - | - | - | + | - | + | - | - | - | - | + | + | + | + |
| *Arabidopsis thaliana* | - | - | - | + | - | + | - | - | - | - | + | + | + | + |

WGD

Tree in unscaled branches

\* Non-canonical Dicer that contains an RNase III domain (like RNT1)

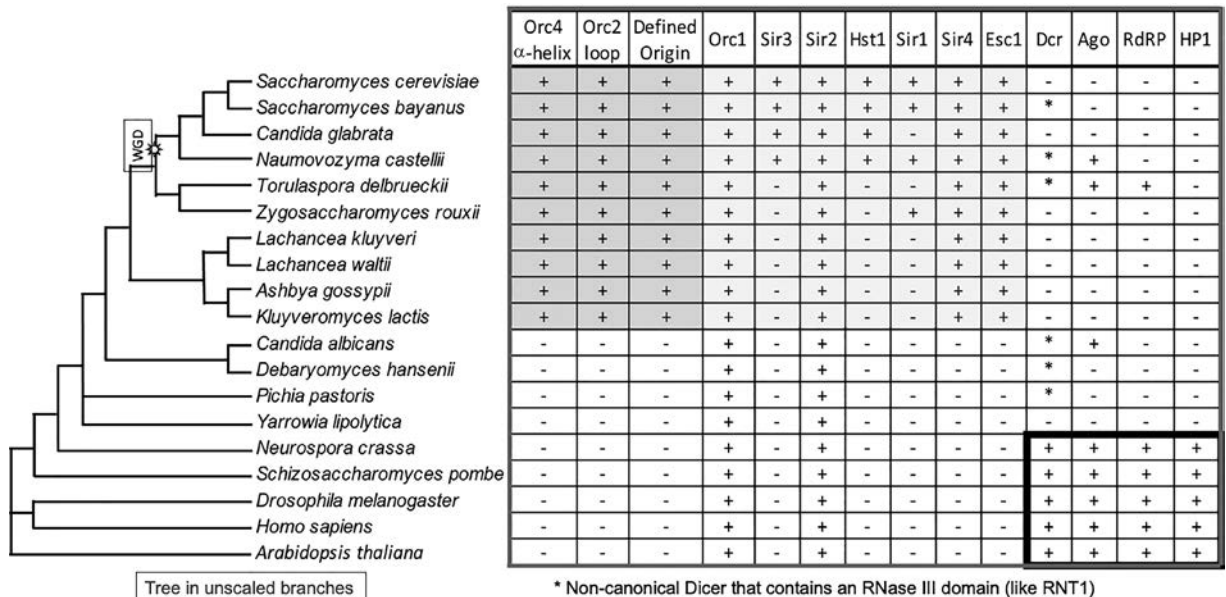**Figure 1.** Coevolution of DNA sequence–specific origins of DNA replication and ORC-Sir4-mediated gene silencing. Sir4 is acquired at the same time during evolution as the Orc2 loop and Orc4 α-helix. Species that do not have DNA sequence–specific origins tend to have RNAi-mediated transcriptional gene silencing. Some transition species harbor a form of Dicer (\*) that has the potential for RNAi.

the figure, and the presence of Sir4, one of the core silencing proteins, correlates with DNA sequence-specific origins that are highlighted in the dark gray box in Figure 1. The species *Torulaspora delbrueckii* retains Ago and Dcr, components of the RNAi pathway, but it has been shown by Jasper Rine's laboratory (UC Berkeley) that mutating genes encoding Ago and Dcr in this species had no effect on transcriptional gene silencing. Thus, *T. delbrueckii* appears to be an interesting transition species in the evolution from RNAi-mediated to ORC-SIR-mediated gene silencing. The species *Candida albicans, Debaryomyces hansenii*, and *Pichia pastoris* have retained a noncanonical Dcr (shown as Dcr* in Fig. 1), which has been shown by the laboratories of Gerry Fink and David Bartel (MIT) to have the potential for carrying out RNAi-mediated gene silencing. Thus, it appears that the transition from RNAi-mediated gene silencing to ORC-Sir4-mediated gene silencing has been accompanied by an acquisition of sequence-specific origins of DNA replication. Because RNAi and Dicer have been shown by Rob Martienssen's laboratory to play a role in termination of gene transcription under circumstances of DNA replication stress, we propose that this helps avoid the conflicts between DNA replication and gene transcription, which cause genome instability. Placing origins of DNA replication in intergenic regions, which is known in *S. cerevisiae*, would aid in avoiding conflicts between RNA polymerase and DNA polymerase.

From this study, one of the most interesting species is *Yarrowia lipolytica*, as analysis to date by others suggests it lacks sequence-specific origins of DNA replication, lacks the Orc2 loop and the Orc4 α-helix, and does not have Sir4; hence, it does not have ORC-Sir4-mediated transcriptional gene silencing. Also, *Y. lipolytica* does not have RNAi, so it is unclear how rDNA- and telomere-repeated sequences are maintained in this species. The lack of the Orc4 α-helix and the Orc2 loop suggests that, like human cells, *Y. lipolytica*

has a mechanism for determining the location along the genome of origins of DNA replication that is more like human cells than *S. cerevisiae*. Therefore, studies have been initiated on the mechanism of DNA replication of origin specification in *Y. lipolytica,* including the genome-wide distribution of origins in this organism and how they are localized and controlled during the cell-division cycle. This yeast is a major species for bioengineering and production of enzymes related to lipid production (oil) and for the production of industrial-scale enzymes. The yeast derives from many environments, including creamy cheeses in which the yeast converts protein into oil.

We continue to study how ORC and CDC6 in human cells combine to form pre-RCs across the genome and how they are involved in transcriptional gene silencing. In previous years, we have shown that HsORC binds to many proteins such as the protein kinases cyclin E-CDK2 and cyclin A-CDK2, the protein phosphatase PP1, the histone methyltransferase SUV39H1, the heterochromatin protein HP1, and the retinoblastoma tumor suppressor protein. We have uncovered many other proteins that are involved in chromatin modifications and organization of DNA in the nucleus and are currently determining how these proteins contribute to DNA replication and gene silencing in human cells. Many, but not all, of these proteins are present in *Y. lipolytica* yeast, but not in the budding yeast *S. cerevisiae*.

## PUBLICATIONS

Dunn A, Johnstone RW, Stillman B. 2019. Joseph Sambrook (1939–2019). *Nat Struct Molec Biol* **26:** 846–847.

Greider CW, Sheltzer JM, Cantalupo NC, Copeland WB, Dasgupta N, Hopkins N, Jansen JJ, Joshua-Tor L, McDowell GS, Metcalf JL, et al. 2019. Increasing gender diversity in the STEM research workforce. *Science* **366:** 692–695.

Hossain M, Bhalla K, Stillman B. 2019. Cyclin binding Cy motifs have multiple activities in the initiation of DNA replication. bioRxiv doi:10.1101/681668

# TRANSCRIPTIONAL DEPENDENCIES IN HUMAN CANCER

**C.R. Vakoc**    L. Benjamin      O. Klingbeil        S. Polyanskaya    Y. Wei
                  C. Fitzpatrick   C. Lopez-Cleary     D. Skopelitis     X. Wu
                  Y. Gao           B. Lu               T. Somerville     Z. Yang
                  Y. Huang         D. Maia-Silva       M. Sroka          T. Yoshimoto
                  S. Hur           J. Milazzo          Y. Tarumoto

Massive genome-wide reprogramming of transcription is critical for malignant transformation. As a consequence, cancer cells are vulnerable to perturbations of the transcriptional apparatus, which includes targeting of DNA-binding transcription factors (TFs)/cofactors and chromatin regulatory machineries. Over the past decade, our laboratory has taken a genetic screening approach to identify transcriptional dependencies in cancer cell lines. Upon identifying cancer-specific patterns of essentiality, we have pursued detailed molecular mechanisms that underpin these cellular phenotypes. By understanding transcriptional dependencies in cancer, we have revealed fundamental mechanisms of gene control, novel processes that drive cancer formation, and new therapeutics that reprogram transcription to eliminate cancer cells. The broad goals of our current research are (1) to identify novel cancer-specific dependencies and evaluate underlying mechanisms, (2) to reveal detailed molecular mechanisms of lineage master regulator TFs that drive cancer cell growth, (3) to develop chemical probes that modulate the function of lineage master regulators, and (4) to explore how lineage cell-of-origin and *trans*-differentiation processes contribute to the pathogenesis and therapy of human tumors.

## Causes and Consequences of Squamous *Trans*-Differentiation in Pancreatic Ductal Adenocarcinoma (PDA)

S. Hur, D. Maia-Silva, T. Somerville [in collaboration with the Tuveson laboratory]

Lineage plasticity is a prominent feature of pancreatic ductal adenocarcinoma (PDA) cells, which can occur via deregulation of lineage-specifying TFs. A major effort in the laboratory is directed toward understanding the causes and consequences of lineage plasticity in PDA. We recently discovered that the zinc finger protein ZBED2 is aberrantly expressed in PDA and alters tumor cell identity in this disease. Unexpectedly, our epigenomic experiments reveal that ZBED2 is a sequence-specific transcriptional repressor of interferon-stimulated genes, which occurs through antagonism of interferon regulatory factor 1 (IRF1)-mediated transcriptional activation at co-occupied promoter elements. Consequently, ZBED2 attenuates the transcriptional output and growth arrest phenotypes downstream of interferon signaling in multiple PDA cell line models. We also found that ZBED2 is preferentially expressed in the squamous molecular subtype of human PDA in association with inferior patient survival outcomes. Consistent with this observation, we found that ZBED2 can repress the pancreatic progenitor transcriptional program, enhance motility, and promote invasion in PDA cells. Collectively, our findings suggest that high ZBED2 expression is acquired during PDA progression to suppress the interferon response pathway and to promote lineage plasticity in this disease.

A highly aggressive subset of PDAs undergo *trans*-differentiation into the squamous lineage during disease progression. In a new study, we investigated whether squamous *trans*-differentiation of pancreatic cancer cells can influence the phenotype of non-neoplastic cells in the tumor microenvironment. Conditioned media experiments revealed that squamous pancreatic cancer cells secrete factors that convert quiescent pancreatic stellate cells into a specialized subtype of cancer-associated fibroblasts (CAFs) that express inflammatory genes at high levels. We used gain- and loss-of-function approaches in vivo to show that squamous-subtype pancreatic tumor models become enriched with inflammatory CAFs and neutrophils in a TP63-dependent manner. These effects occur, at least in part, through TP63-mediated activation of enhancers at proinflammatory cytokine loci, which includes *IL1A* as a key target. Taken together,

our findings reveal enhanced tissue inflammation as a consequence of squamous *trans*-differentiation in pancreatic cancer, thus highlighting an instructive role of tumor cell lineage in reprogramming the stromal microenvironment.

A key question that we have yet to resolve is what triggers TP63 expression to initiate squamous *trans*-differentiation in PDA, which we are now pursuing via TP63 reporters coupled with genome-wide CRISPR screens. In addition, we seek to identify additional TP63-independent mechanisms by which PDA cells achieve a squamous-like cell fate. Our recent studies highlight a role for the TP63 homolog TP73 (TA isoform) in driving squamous cell identity in PDA. The unique attributes of TP63 versus TP73 in this context are being evaluated.

## Lineage Dynamics and Dependencies in Small-Cell Lung Cancer

C. Fitzpatrick, Y. Huang, X. Wu

Small-cell lung cancer (SCLC) is widely considered to be a tumor of pulmonary neuroendocrine cells; however, a variant form of this disease has been described that lacks neuroendocrine features. We applied domain-focused CRISPR screening to human cancer cell lines to identify the TF POU2F3 as a powerful dependency in a subset of SCLC lines. An analysis of human SCLC specimens revealed that POU2F3 is expressed exclusively in variant SCLC tumors that lack expression of neuroendocrine markers and instead express markers of a chemosensory lineage known as tuft cells. Using chromatin- and RNA-profiling experiments, we obtained evidence that POU2F3 is a master regulator of tuft cell identity in a variant form of SCLC. These findings revealed POU2F3 as a cell identity determinant and a dependency in a tuft cell variant of SCLC, which may reflect a previously unrecognized cell of origin or a *trans*-differentiation event in this disease (Huang et al., *Genes Dev* **32:** 915 [2018]).

Because POU2F3-expressing tuft cells exist in the normal mouse lung, we hypothesize that this cell type serves as a novel cell-of-origin in this malignancy. To evaluate this, we are inactivating the tumor suppressor genes *Rb1*, *Trp53*, and *Rb2* in the normal tuft cell lineage of the mouse lung using a Cre knockin at the *Pou2f3* locus, and animals are being characterized to determine whether tumors emerge from this lineage that resemble the human disease we previously identified.

POU2F3 is a powerful dependency in tuft cell lineage lung cancer, yet a Pou2f3$^{-/-}$ mouse is known to exhibit a discrete defect in tuft cell development that results in defective helminth immunity. Thus, we speculate that a target of POU2F3 would have a wide therapeutic index in human patients with tuft cell lung cancer. To advance this idea, we are pursuing the identification of POU2F3 cofactors whose function is critical for the maintenance of tuft cell lung cancer. Using a combination of genetic screening and biochemical evidence, we have recently identified a novel human gene, which we have renamed OCA-T, that functions as a POU2F3 cofactor in tuft cell lung cancer. Ongoing mechanistic studies aim to define the detailed mechanism that connects the OCA-T and POU2F3 function.

In the classical neuroendocrine form of SCLC, tumors express and depend on ASCL1 for their viability. Thus, we hypothesize that molecular mechanisms of ASCL1 might inform novel targeted therapy approaches for the neuroendocrine subtype of SCLC. In a new project, we have devised a genetic screening strategy that seeks to expose ASCL1 cofactors in an unbiased manner.

## Acquired Dependencies in Leukemia Caused by Epigenetic Silencing of Metabolic Genes

Y. Wei, Z. Yang

It has long been recognized that cancer cells acquire aberrant DNA methylation patterns, which often lead to gene silencing when methylation is acquired at CpG islands. Whereas epigenetic silencing can lead to loss of tumor suppressor gene expression, the full range of consequences of aberrant DNA methylation in cancer are unknown. Through two convergent studies, we have discovered an unexpected consequence of aberrant DNA methylation in acute myeloid leukemia (AML). In each of these two studies, we have found that a gene encoding a nonessential metabolic enzyme becomes aberrantly silenced, which leads to an acquired dependence on a compensatory pathway. In one study, the silenced gene is *ISYNA1*, which encodes a rate-limiting enzyme for myo-inositol biosynthesis. Because myo-inositol is an essential metabolite

in cells, the leukemia cells compensate for the lack of *ISYNA1* expression by becoming dependent on the myo-inositol transporter SLC5A3, which we discovered by way of a CRISPR screen. Notably, restoring *ISYNA1* expression rescues SLC5A3 dependence. In a separate study, we discovered via CRISPR screening that a subset of leukemia cell lines is dependent on the FANCONI anemia DNA repair pathway for their viability. Underlying this dependence is the epigenetic silencing of ALDH2, which is a key enzyme responsible for detoxifying endogenous aldehydes. We find that silencing of ALDH2 is compensated for by the FANCONI anemia complex, which becomes essential to repair DNA damage caused by endogenous aldehydes. Notably, in both these studies, the epigenetic silencing appears to be a passenger event in that it does not appear to drive leukemogenesis. Nevertheless, this epigenetic silencing event leads to an acquired vulnerability in this disease.

## Salt-Inducible Kinase Inhibition Suppresses Acute Myeloid Leukemia Progression In Vivo

Y. Tarumoto, S. Polyanskaya, Y. Wei, Z. Yang, B. Lu

Lineage-defining TFs are compelling targets for leukemia therapy, yet they are among the most challenging proteins to modulate directly with small molecules. We previously used CRISPR screening to identify a salt-inducible kinase 3 (SIK3) requirement for the growth of AML cell lines that overexpress the lineage TF myocyte enhancer factor (MEF2C). In this context, SIK3 maintains MEF2C function by directly phosphorylating histone deacetylase 4 (HDAC4), a repressive cofactor of MEF2C. In more recent studies, we evaluated whether inhibition of SIK3 with the tool compound YKL-05-099 can suppress MEF2C function and attenuate disease progression in animal models of AML. Genetic targeting of SIK3 or MEF2C selectively suppressed the growth of transformed hematopoietic cells under in vitro and in vivo conditions. Similar phenotypes were obtained when cells were exposed to YKL-05-099, which caused cell-cycle arrest and apoptosis in MEF2C-expressing AML cell lines. An epigenomic analysis revealed that YKL-05-099 rapidly suppressed MEF2C function by altering the phosphorylation state and nuclear localization of HDAC4. Using a gatekeeper allele of SIK3,

we found that the antiproliferative effects of YKL-05-099 occurred through on-target inhibition of SIK3 kinase activity. Based on these findings, we treated two different mouse models of MLL-AF9 AML with YKL-05-099, which attenuated disease progression in vivo and extended animal survival at well-tolerated doses. These findings validate SIK3 as a therapeutic target in MEF2C-addicted AML and provide a rationale for developing drug-like inhibitors of SIK3 for definitive preclinical investigation and for studies in human patients.

## A Nuclear Phosphatase–Kinase Signaling Pathway that Drives Leukemia Growth

S. Polyanskaya

To identify leukemia-specific vulnerabilities within the human phosphatases, we performed a CRISPR/Cas9-based negative-selection, domain-focused screen in a panel of eight leukemia and four nonleukemia cell lines. Of 211 phosphatases, CTDSPL2 targeting conferred competitive disadvantage to the highest number of leukemia cell lines with minimal impact on the growth of nonleukemia cell lines tested. We confirmed the AML-biased pattern of CTDSPL2 dependence with individual gRNAs in competition assays in ten leukemia and eight nonleukemia cell lines. We also confirmed that CTDSPL2 is required for the proliferation of human AML cell line MOLM-13 in vivo upon tail vein injection of pre-engrafted Cas9+ cells infected with sgCTDSPL2 with ~100% multiplicity of infection (MOI). We showed growth arrest upon targeting Ctdspl2 in murine MLL-AF9/NrasG12D AML cells that could be rescued by the expression of human SCP4 (the protein product of *CTDSPL2*). NIH3T3-immortalized murine fibroblasts were unaffected by Ctdspl2 targeting. In collaboration with the Hematology Department of St. Jude Children's Research Hospital, we assayed whether *CTDSPL2* is required for normal hematopoiesis. We did not observe any effect of the presence of *CTDSPL2* knockout cells in the population on the normal course of in vitro differentiation of either of the hematopoietic lineages. Using the cDNA rescue assay, we found that catalytic mutants of SCP4 were unable to support the proliferation of MOLM-13 cells. In the same manner, we confirmed that SCP4 with amino-terminal 236-aa deletion was sufficient

to rescue the phenotype of endogenous *CTDSPL2* depletion, further underlining the importance of the phosphatase function of SCP4. It has been proven both in our experiments and in the published literature that SCP4 preferentially localizes to chromatin3. However, we did not succeed in mapping it onto the genome via chromatin immunoprecipitation followed by sequencing (ChIP-seq). We profiled gene expression in *CTDSPL2* knockout-sensitive (MOLM-13 and NOMO-1) and -resistant (K-562 and U-937) cell lines using RNA sequencing (RNA-seq). The unsupervised hierarchical clustering of the RNA-seq data demonstrated that the cell lines that are sensitive or resistant to *CTDSPL2* knockout consistently cluster separately from each other, and thus *CTDSPL2* has a differential effect on gene expression in sensitive versus resistant cell lines. This body of evidence makes *CTDSPL2* an attractive target for possible therapeutic intervention studies. However, we did not succeed so far in identifying diagnostic markers for *CTDSPL2* dependency that would make it difficult for patient stratification. The need for understanding the molecular mechanism of the SCP4 function is apparent. We were very intrigued by our finding that catalytically active SCP4 interacts with both homologs PDIK1L and STK35L. These are almost unstudied nuclear kinases, and their functions in cells are largely unknown. Our advantage compared to the other groups is that we have data indicating that their genetic redundancy might conceal the requirement for them in some cancer cells. We are planning to explore the relationship between SCP4, PDIK1L, and STK35, further elucidating the molecular mechanism they participate in to ensure the proliferation of AML cells.

## Targeting Master Regulator Transcription Factors in Sarcoma

Y. Gao, C. Lopez-Cleary, M. Sroka, T. Yoshimoto

Alveolar rhabdomyosarcoma (aRMS) is a rare muscle cancer that affects primarily children and adolescents. Whereas the disease bears a low overall mutational burden, over 60% of aRMS patients harbor the t(2;13)(q35;q14) translocation that leads to expression of the PAX3-FOXO1 fusion oncoprotein, in which the DNA-binding domain of PAX3 is linked with the transactivation domain of FOXO1. Fusion-positive

tumors are the most aggressive, with a four-year overall survival rate of 8% for metastatic tumors. Numerous studies showed that aRMS cancers are dependent on the presence of the PAX3-FOXO1 chimera and that withdrawal causes cell death or differentiation. However, no therapies targeting the fusion protein exist to date, partly because of the challenges associated with designing drugs that target TFs. The goal of our ongoing efforts is to elucidate the entire PAX3-FOXO1 fusion oncoprotein pathway in aRMS. To this end, we are characterizing how aRMS cells respond to fusion inactivation as well as the molecular signatures and dynamics of each response. We have developed an unbiased, reporter-based, FACS-assisted CRISPR screening strategy to identify genes that phenocopy fusion inactivation upon knockout. These efforts will impact our understanding of the basic biology of fusion-positive aRMS. Further, the findings might allow the development of therapeutics that indirectly silence the pathway by interfering with factors that cooperate with the fusion oncoprotein in the sustenance of the transformed state.

Two other fusion proteins are also of interest to our laboratory, including EWS-FLI1 in Ewing's sarcoma and CIC-DUX4 in Ewing's-like sarcoma. We are attempting to establish a molecular reporter of these fusion proteins to be assessed by CRISPR screening. In addition, we are pursuing mechanistic studies of MyoD, which is a powerful lineage dependency in rhabdomyosarcoma cells.

## Paralog Screening Identifies MARK2/MARK3 Kinases as Carcinoma Dependencies

O. Klingbeil, D. Skopelitis [in collaboration with O. El Demerdash, Functional Genomics Core Facility]

One limitation of our previous CRISPR screening strategy is that only single genes are inactivated in our pooled genetic screens. Importantly, evolution often produces novel genes via duplication events, which can produce gene pairs that function redundantly to support cellular functions. We have been concerned about whether redundancy conceals essential gene functions in our essentiality screens. To address this issue, we developed a CRISPR screening strategy in which two sgRNAs are expressed from a single lentiviral vector backbone. This allows us to produce single and double knockouts within a single genetic

screen. We generated sgRNA libraries that cotarget homologous kinase-, phosphatase-, and chromatin-modifying enzymes. These studies led us to make the discovery that several carcinoma cell lines are dependent on MARK2/MARK3, which function in a redundant manner to support cancer cell line growth. Notably, several hematopoietic and neuroendocrine lineage tumor lines do not require MARK2/MARK3 for survival. We are currently pursuing the mechanisms underlying this pattern of kinase essentiality. In addition, we are developing peptide-based inhibitors that target MARK kinase activity for use as probes of biological function in vivo.

## PUBLICATIONS

Aho ER, Wang J, Gogliotti RD, Howard GC, Phan J, Acharya P, Macdonald JD, Cheng K, Lorey SL, Lu B, et al. 2019. Displacement of WDR5 from chromatin by a WIN site inhibitor with picomolar affinity. *Cell Rep* **26:** 2916–2928.e13.

Bell CC, Fennell KA, Chan YC, Rambow F, Yeung MM, Vassiliadis D, Lara L, Yeh P, Martelotto LG, Rogiers A, et al. 2019. Targeting enhancer switching overcomes non-genetic drug resistance in acute myeloid leukaemia. *Nat Commun* **10:** 2723.

Bharathy N, Berlow NE, Wang E, Abraham J, Settelmeyer TP, Hooper JE, Svalina MN, Bajwa Z, Goros MW, Hernandez BS, et al. 2019. Preclinical rationale for entinostat in embryonal rhabdomyosarcoma. *Skelet Muscle* **9:** 12.

Böttcher J, Dilworth D, Reiser U, Neumüller RA, Schleicher M, Petronczki M, Zeeb M, Mischerikow N, Allali-Hassani A, Szewczyk MM, et al. 2019. Fragment-based discovery of a chemical probe for the PWWP1 domain of NSD3. *Nat Chem Biol* **15:** 822–829.

Caiado F, Maia-Silva D, Jardim C, Schmolka N, Carvalho T, Reforço C, Faria R, Kolundzija B, Simões AE, Baubec T, et al. 2019. Lineage tracing of acute myeloid leukemia reveals the impact of hypomethylating agents on chemoresistance selection. *Nat Commun* **10:** 4986.

Ehrenhöfer-Wölfer K, Puchner T, Schwarz C, Rippka J, Blaha-Ostermann S, Strobl U, Hörmann A, Bader G, Kornigg S, Zahn S, et al. 2019. SMARCA2-deficiency confers sensitivity to targeted inhibition of SMARCA4 in esophageal squamous cell carcinoma cell lines. *Sci Rep* **9:** 11661.

Foronda M, Tarumoto Y, Schatoff EM, Leach BI, Diaz BJ, Zimmerman J, Goswami S, Shusterman M, Vakoc CR, Dow LE. 2019. Tankyrase inhibition sensitizes cells to CDK4 blockade. *PLoS One* **14:** e0226645.

Gryder BE, Pomella S, Sayers C, Wu XS, Song Y, Chiarella AM, Bagchi S, Chou HC, Sinniah RS, Walton A, et al. 2019. Histone hyperacetylation disrupts core gene regulatory architecture in rhabdomyosarcoma. *Nat Genet* **51:** 1714–1722.

Lan X, Khandros E, Huang P, Peslak SA, Bhardwaj SK, Grevet JD, Abdulmalik O, Wang H, Keller CA, Giardine B, et al. 2019. The E3 ligase adaptor molecule SPOP regulates fetal hemoglobin levels in adult erythroid cells. *Blood Adv* **3:** 1586–1597.

Li F, Deng Z, Zhang L, Wu C, Jin Y, Hwang I, Vladimirova O, Xu L, Yang L, Lu B, et al. 2019. ATRX loss induces telomere dysfunction and necessitates induction of alternative lengthening of telomeres during human cell immortalization. *EMBO J* **38:** e96659.

Rudin CM, Poirier JT, Byers LA, Dive C, Dowlati A, George J, Heymach JV, Johnson JE, Lehman JM, MacPherson D, et al. 2019. Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. *Nat Rev Cancer* **19:** 289–297.

Yohe ME, Heske CM, Stewart E, Adamson PC, Ahmed N, Antonescu CR, Chen E, Collins N, Ehrlich A, Galindo RL, et al. 2019. Insights into pediatric rhabdomyosarcoma research: challenges and goals. *Pediatr Blood Cancer* **66:** e27869.

# CANCER: GENETICS

**Christopher Hammell**'s laboratory is interested in understanding gene regulatory processes that give rise to robust phenotypes associated with normal development in animals (specifically, how the timing of developmental processes is controlled), as well as the alterations in these pathways that give rise to diseases such as cancer (as in the alterations in mitogenic pathways in melanoma). Hammell and colleagues approach this elemental problem by using a variety of model organisms and patient-derived cancer cell lines. To directly identify the components that function in controlling normal developmental timing, they use the small nematode *Caenorhabditis elegans*, applying forward and reverse genetic approaches. In contrast to the extreme robustness of cell-fate lineage in *C. elegans*, in which specification of developmental programs is hard-wired, mutations that alter conserved signaling pathways in melanoma create relatively plastic developmental landscapes that allow these lesions to become aggressive tumors. Notably, the gene regulatory architecture of melanoma cells allows them to acquire resistance to therapeutic agents. Hammell's team is interested in epigenetic mechanisms that contribute to resistance, specifically dramatic changes in gene expression patterns and intracellular signaling pathways. They are performing high-throughput screens to identify cellular factors that allow these rewiring events to occur, with the idea that these components would make ideal therapeutic targets to complement existing clinical strategies.

The **Je H. Lee** laboratory studies how cells interact with their microenvironment to regulate gene expression during development. Single-cell heterogeneity in gene expression can result from spatial differences in cell–cell and cell–extracellular matrix interactions. Such differences contribute to stochastic evolution of tumor cells as well as morphogenesis during normal development. However, the spatial control of gene expression in complex tissues, embryos, or tumors remains poorly understood, because most genome-wide studies sample bulk tissues or dissociated single cells.

The Lee laboratory has recently developed a method to sequence RNA molecules directly within single cells and tissues using subcellular resolution imaging, and they showed subtle differences in cell–cell/ECM signaling and gene expression genome-wide in situ. By clustering transcripts into functionally or morphologically discrete regions, they find many unique spatial markers and signaling pathways. The Lee laboratory focuses on the role of noncoding RNA in chromatin remodeling and tumor progression using mouse and organoid models of human cancer. They use in situ sequencing, cell lineage tracing, and single-cell profiling to understand how noncoding RNA affects tumor cell evolution in their native context. The laboratory's long-term goal is to develop better tumor classification tools and anticancer therapeutics using our understanding of the tumor microenvironment.

**Alea Mills** is studying genetic pathways important in cancer, aging, and autism, identifying the genetic players and determining how aberrations in their functions culminate in human disease. Through innovative use of a technique called "chromosome engineering," the Mills group discovered that one of the most common genetic alterations in autism—deletion of a 27-gene cluster on chromosome 16—causes autism-like features in mice. These autism-like movement impairments can be identified just days after birth, suggesting that these features could be used to diagnose autism. The Mills group has also used chromosome engineering to identify a tumor-suppressor gene that had eluded investigators for three decades. The gene, called *Chd5*, was shown by the Mills group to regulate an extensive cancer-preventing network. This year, the Mills laboratory uncovered how *Chd5* acts as a tumor suppressor: It binds to a protein found within chromatin to turn specific genes on or off, halting cancer progression. The epigenetic role of *Chd5* in development, cancer, and stem-cell maintenance is currently being investigated. The Mills laboratory is also studying p63 proteins,

which regulate development, tumorigenesis, cellular senescence, and aging in vivo. They succeeded in halting the growth of malignant tumors by turning on production of one of the proteins encoded by the *p63* gene, called TAp63. TAp63 also exerts other protective effects. This year, the Mills laboratory generated a mouse model that allowed them to find that TAp63 is required to prevent a genetic disorder, known as EEC (ectrodactyly-ectodermal dysplasia cleft lip/palate syndrome), which is characterized by a cleft palate and major deformities of the skin and limbs in infants. In addition, they recently discovered that a different version of *p63*, called ΔNp63, reprograms stem cells of the skin to cause carcinoma development—the most prevalent form of human cancer. Modulation of these proteins may offer new ways to treat human malignancies in the future.

Human cancers show a diverse array of genomic gains and losses that alter the dosage of hundreds of genes at once. About 90% of solid tumors display whole-chromosome aneuploidy, whereas many tumors with diploid karyotypes nonetheless harbor segmental or arm-length aneuploidies that also result in significant gene copy number alterations. Despite the prevalence of aneuploidy in cancer, its functional consequences for cell physiology remain poorly understood. **Jason Sheltzer** and colleagues have shown the existence of several surprising phenotypes that are shared among cells with different chromosomal imbalances. They showed that aneuploidy can function as a novel source of genomic instability, as aneuploid cells tend to display elevated levels of mutation, mitotic recombination, and chromosome loss. Sheltzer and colleagues have also identified a transcriptional signature of aneuploidy that is associated with cellular stress and slow proliferation and is found in aneuploid primary and cancer cells across a host of organisms. More recently, they have investigated the link between aneuploidy and cellular transformation. Using a series of genetically matched euploid and aneuploid cell lines, they have shown that aneuploidy can paradoxically function as a barrier to tumor growth. They are currently continuing their investigation of the role in aneuploidy in cancer. They are also applying CRISPR-Cas9-mediated genome engineering to develop novel mouse models for exploring the impact of gene dosage alterations on tumor development in vivo. Although aneuploidy is a ubiquitous feature of human tumors, it occurs rarely in somatic cells. Thus, differences between aneuploid and euploid cells may represent crucial therapeutic vulnerabilities in cancer. By identifying phenotypes that are shared among tumors with different aneuploidies, Sheltzer and colleagues hope to discover pathways that can be manipulated to selectively eliminate aneuploid cells or to block aneuploidy's non–cell autonomous effects. Drugs that target these pathways may have broad utility against a wide range of aneuploid cancers while showing minimal toxicity in euploid tissue.

**Michael Wigler**'s work provides a new paradigm for understanding and exploring human disease. The Wigler laboratory studies human cancer and the contribution of new mutation to genetic disorders. The cancer effort (with James Hicks, Alex Krasnitz, and Lloyd Trotman) focuses on breast and prostate cancers. It involves collaborative clinical studies to discover mutational patterns predicting treatment response and outcome and the development of diagnostics to detect cancer cells in bodily fluids such as blood and urine. The major tools are single-cell DNA and RNA analysis. The single-cell methods, which are in development, are also being applied to problems in neurobiology (with Josh Huang and Pavel Osten) to characterize neuronal subtypes, somatic mutation, and monoallelic expression. The Wigler laboratory's genetic efforts are in collaboration with Ivan Iossifov and Dan Levy, and this team focuses on determining the role of new mutations in pediatric disorders. In a large-scale population sequencing project with W. Richard McCombie and the Genome Sequencing Center at Washington University in St. Louis, and supported by the Simons Foundation, the team has proven the contribution of this mechanism to autism. The work further suggests a relationship between the mutational targets in autism and the process of neuroplasticity that lies at the heart of learning. Smaller-scale population studies of congenital heart disease and pediatric cancer (collaborations with scientists at Columbia University and Memorial Sloan Kettering Cancer Center, respectively) also point to new mutation as a causal factor in these disorders.

# CONTROL OF GENE DOSAGE DURING ANIMAL DEVELOPMENT

**C.M. Hammell**   K. Hills-Muckey   C. Scher
                   O. Huiwu          N. Stec
                   B. Kinney         J. Wang

Dynamic changes in gene expression are a hallmark of developmental biology in which cell fates are determined in specific orders to produce functional tissues and organs. Our laboratory studies how genes are turned on and off during development. We are specifically focused on understanding how the precise timing and sequence of developmental events are organized and how the correct expression levels of these key regulatory proteins are established.

## Gene Dosage of Cyclically Expressed Genes Is Controlled by Blmp-1 and Elt-3 That Function as Transcriptional Amplifiers

N. Stec, C.M. Hammell

Using the model organism *Caenorhabditis elegans*, we have identified a developmental clock that organizes the sequence of developmental events by driving repetitive patterns of gene expression. Many of the genes that are targets of this clock (including conserved microRNAs and signaling molecules) directly control aspects of cell fate specification. At the center of this molecular clock is *lin-42*, encoded by the *C. elegans* ortholog of the *Period* gene in humans that is implicated in circadian gene expression. Like Period, LIN-42 functions to dampen the expression of cyclically expressed genes. Through genetic approaches, we have determined that *lin-42* controls the transcription of key microRNAs that are required to specify distinct cell fate transitions during development. Many of these microRNAs are conserved in humans where they also limit cell proliferation and are often times misregulated in cancer.

One of the important functions of the developmental clock is to control the precise transcriptional output of its target genes. This feature is essential because many of the clock's transcriptional targets operate in dosage-dependent manners to control cell fate specification. Precocious or overabundant expression

of clock-controlled genes leads to temporal cell fate transformations in which intermediate steps in development are skipped. In previous years, we have leveraged the genetic tractability of the *C. elegans* organism and the quantifiable precocious developmental phenotypes associated with *lin-42* mutations to identify additional clock components. These efforts identified two conserved transcription factors, BLMP-1 and ELT-3, that function as transcriptional amplifiers of clock-controlled genes. Using a custom-designed microfluidics apparatus built in 2019, Hammell laboratory members established an experimental platform to directly monitor gene expression (at cellular resolution) in living animals throughout development. This allowed the characterization of continuous, dynamic gene expression patterns in *C. elegans* for the first time. Analysis of a variety of clock-controlled reporter transgenes in *blmp-1* and *elt-3* mutants demonstrated that these mutants retain periodic expression of target genes but display altered transcriptional kinetics. Specifically, pulse duration (how long the gene is turned on) and amplitude of transcription are reduced in these mutants. These changes in gene expression directly result in temporal patterning defects.

Using these phenotypes as a foundation, we focused on defining the molecular mechanisms by which BLMP-1 and ELT-3 contribute to transcriptional regulation. This was accomplished using two orthogonal approaches. BLMP-1 and ELT-3 ChIP-seq data, combined with an analysis of chromatin (ATAC-seq) in developing larva, suggested that the binding sites for these transcription factors are associated with an open chromatin conformation (i.e., the genomic location of BLMP-1 and ELT-3 targets lack histone occupancy). This led to the hypothesis that BLMP-1 and ELT-3 may function as pioneer transcription factors that "prime" gene expression of target genes. By interacting directly with nucleosomal bound DNA, BLMP-1 (and ELT-3) binding would lead to the opening of local chromatin by dissociating histone octamers. This "priming" feature would allow the access of

additional transcription factors that mediate transcriptional activation. Using recombinant proteins, we (in collaboration with the Luk laboratory at Stony Brook University) demonstrated that BLMP-1 can bind its cognate recognition sequences in the context of the nucleosome (a defining feature of pioneer factors) (Fig. 1A). We then developed a novel in vivo approach to monitor chromatin dynamics in living animals (Fig. 1B,C). This strategy employed fluorescent proteins that specifically associate with individual clock target gene loci within the nucleus of all somatic cells of the developing animal. By monitoring the internuclear localization and dynamics of these loci during development, the Hammell team determined that the genomic regions of target clock genes are normally tightly compacted in most somatic cell types. Importantly, these loci are specifically decompacted in cells that express BLMP-1. These mechanistic interpretations were validated by demonstrating that this decompaction is lost in animals that lack BLMP-1 expression (Fig. 1D,E).

These findings are important for several reasons. First, we have demonstrated that transcriptional output is modulated by pioneer transcription factors, and this feature directly controls the duration of transcriptional pulses. Second, Blimp1, the human ortholog of *C.e.*BLMP-1, hBlimp1, plays an essential role in human development and transcriptional regulation. Mutations or misexpression of hBlimp1 have also implicated in a number of different cancer types and is essential for the maintenance of blood-related cancers (multiple myeloma) and whose overexpression in pancreatic ductal adenocarcinoma (PDAC) is essential for metastasis. Whether the molecular features we have described for the *C. elegans* version of BLMP-1 (decompaction of chromatin and nucleosomal binding) play a role in cancer pathology is currently being tested in human cancer models and in vitro through collaborations with other CSHL Cancer Center members.

## The Phased Expression of Key Developmental Genes Is Controlled in a Modular Fashion by Multiple Nuclear Hormone Receptors

K. Hills-Muckey, B. Kinney

Through analysis of RNA-sequence data, we have previously established that thousands of *C. elegans*

mRNAs and microRNAs (~18% to the transcriptome) exhibit highly periodic patterns of expression and that these cyclical patterns are tied to the larval molting cycle. Many of the transcripts in this class are encoded by genes that control specific cell fate specifications and whose expression is essential for normal development. A significant amount of effort in our laboratory has been focused on identifying the positive arm of the developmental clock that initiates transcription of cyclically expressed genes. In the past few years, we have been trying to identify these factors using our and other publicly available ChIP-seq data sets to try to identify putative transcription factors that may fit this role. This global approach has been limited by the fact that there are more than 900 putative transcription factors encoded in the *C. elegans* genome and currently ChIP-seq data for only ~10% of these factors exist, limiting the ability to determine and/or correlate putative roles for most transcription factors in our GRN.

To solve this problem using an orthogonal and unbiased approach, we began focusing our efforts on understanding how key targets of this GRN are regulated. This strategy employed two convergent approaches. First, we constructed a series of transcriptional reporters that drive the expression of a GFP-pest reporter using the upstream regulatory regions of genes that exhibit oscillatory expression patterns and control temporal cell fate of specific cell lineages during larval development. These reporters were designed to incorporate highly conserved upstream sequences (identified by comparing synthetic regions found in other evolutionarily related nematode species) that likely contribute to their spatiotemporal expression. Using our novel microfluidics imaging platform that enables us to monitor gene expression in real time over the entire development of the animal, we delineated the expression features for a number of these reporters. We then focused on identifying which conserved regions of these promoters are essential for these expression patterns. These efforts identified regulatory elements that are essential for the normal cyclical expression of these genes. A more detailed dissection of these elements demonstrated that the spatial and temporal information in these sequences is likely encoded by the similar small and/or highly overlapping nucleotides.

The second approach we employed took advantage of the above, detailed characterization of the individual regulatory elements of the *lin-4* gene. We

**A**



**B**



**C**

*cshIs119 (lin-4pro::mCherry-pest; LacO)*

*let-858pro::GFP::LacI*
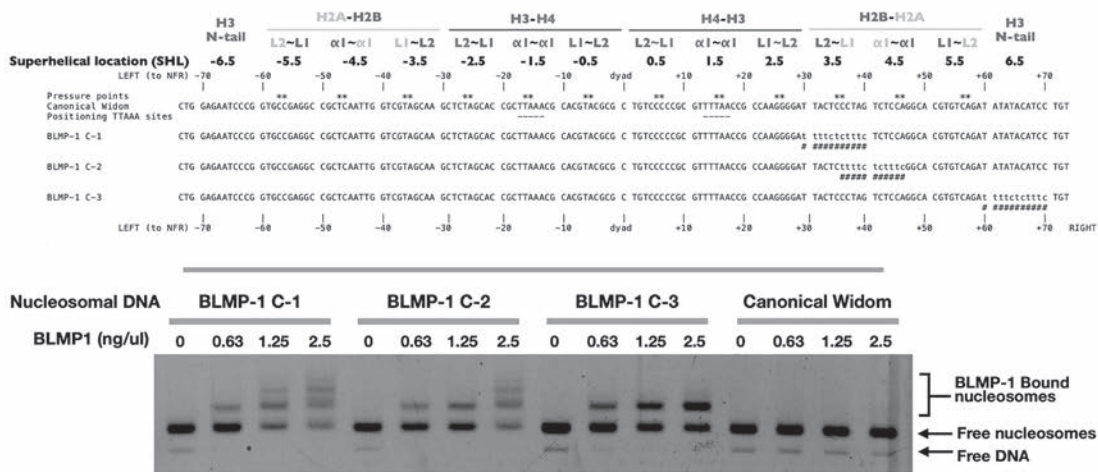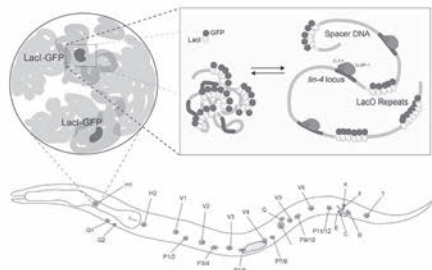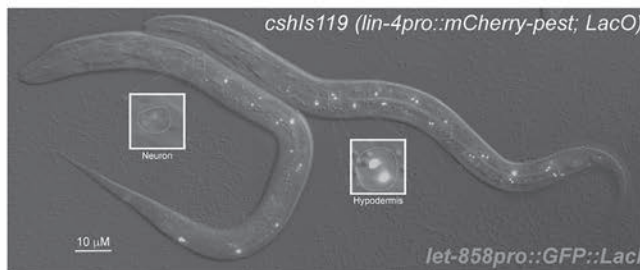
**D**
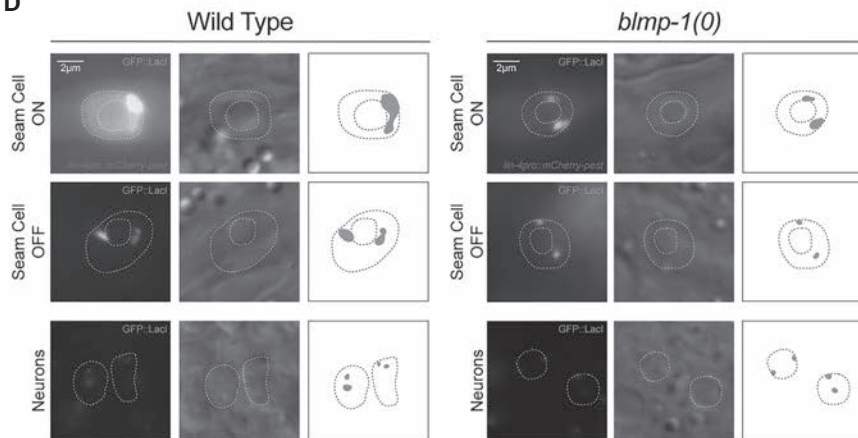


**E**
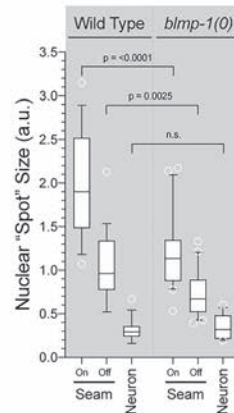


**Figure 1.** BLMP-1 functions as a pioneer factor in vitro and in vivo. (*A*) Recombinant BLMP-1 zinc-finger domains can specifically interact with their cognate binding sites while wrapped in nucleosomal DNA. (*Top*) The various DNA substrates harboring BLMP-1 consensus binding sites (demarcated with ### symbols) at different locations relative to the nucleosome dyad. (*Bottom*) Representative gel shifts of nucleosomal DNA complexes and recombinant BLMP-1 (ZnF domains). BLMP-1 only interacts and alters the mobility of nucleosomes that harbor BLMP-1 binding sites (C-1, C-2, and C-3) and not nucleosomes that lack them (Windom sequence). (*B*) Diagram of the in vivo assay to localize specific loci in living animals. (*C,D*) Representative images of whole animals and individual nuclei of transgenic animals indicate that the lin-4 loci is "decompacted" in cells that express BLMP-1. (*D,E*) Images and quantification of lin-4 loci "puffs" sizes in various tissues of wild-type and BLMP-1 animals.

obtained a complete recombinant, arrayed cDNA library of all (948) *C. elegans* transcription factors, and these coding sequences were cloned into yeast expression vectors as fusion with a general transcriptional activation domain. This library enabled us to probe interactions between each *C. elegans* transcription factor and the various individual regulatory elements identified above. These experiments identified two highly conserved nuclear hormone receptors (NHR-23 and NHR-85) and three additional transcription factors (see below). nhr-23 and nhr-85, encode the *C. elegans* orthologs of RORγ and Rev-ERBβ, respectively—two genes implicated in the human circadian gene regulation. Although these two human nuclear hormone receptors are known to play a role in controlling circadian gene expression, the regulation of, identity, and function of their putative hormone ligands are unknown.

To characterize the function of these nuclear hormone receptors for roles in development and transcription, we needed to create alleles of these genes to assay loss-of-function and gain-of-function phenotypes in vivo. Using conventional, reverse genetic approaches we determined that many of these nuclear hormones are essential for development. Animals harboring null

alleles of *nhr-23* die during embryogenesis, thus preventing us from studying their function in postembryonic development. To circumvent this problem, we adapted the auxin-induced degradation system to control their expression during development. The system takes advantage of previous observations in yeast and human cells in which the heterologous expression of a plant E3 ligase can illicit the degradation of almost any fusion protein that harbors the small plant epitope recognized by this ligase. Importantly, the heterologously expressed E3 ligase, TIR1, can only recognize the epitope in the context of the plant hormone auxin, making this system a highly tractable system to control protein expression. One issue limiting this system's utility in our model is the fact that most natural sources of this plant hormone are not very soluble in aqueous solution. In collaboration with the Matus laboratory at Stony Brook University, we explored the activity of a variety of auxin analog compounds by generating a wide array of transgenic *C. elegans* strains that express. We demonstrated that derivatives of these analogs, specifically 1-naphthaleneacetic acid (NAA), are very effective in our system and can even be adapted to our microfluidics device (Fig. 2A–C). We have also constructed a number of transgenic *C.*
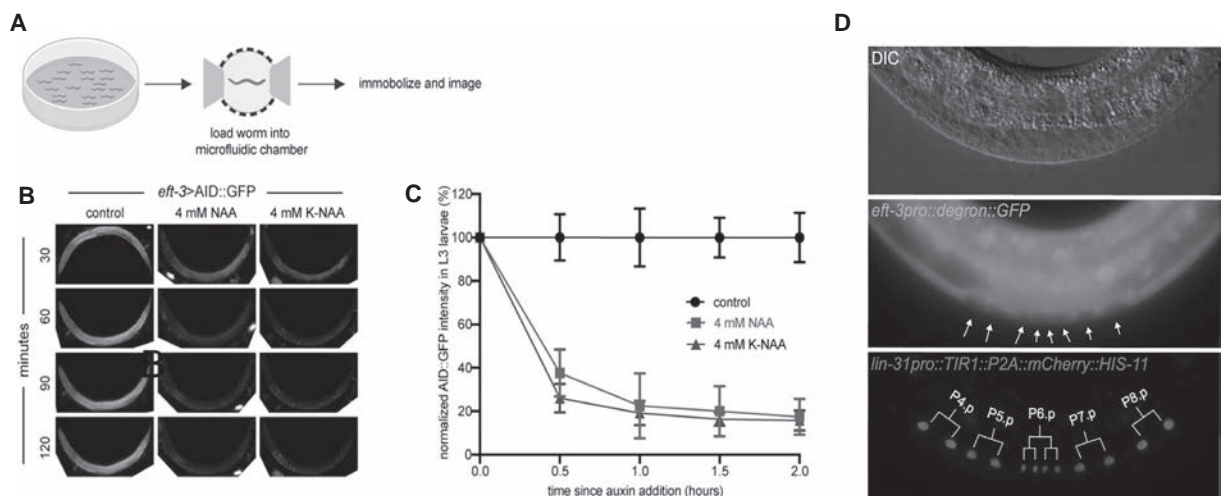


**Figure 2.** K-NAA degradation kinetics in a *Caenorhabditis elegans*–based microfluidic device and traditional solid growth media. (*A*) Schematic representation of the microfluidics-based approach (Keil et al. 2017). (*B*) Images of *eft-3*>AID::GFP expression from mid-L3 stage animals in control conditions (M9 buffer containing NA22 only) or conditions where a 4-mM NAA or K-NAA solution in M9 buffer containing NA22 was perfused through the microfluidic chamber for the time indicated. Anterior is *left* and ventral is *down*. (*C*) Rates of degradation were determined by quantifying AID::GFP in whole animals following auxin treatment. Data presented as the mean +/- SD ($n > 4$ animals were examined for each time point). (*D*) Tissue-specific expression of the TIR1 adaptor from *Arabidopsis thaliana* can deplete the expression of a ubiquitously expressed *eft-3*>::AID:GFP reporter specifically in vulval precursor cells enabling protein to be directly regulated by auxin or NAA in a spatially and temporally defined manner.
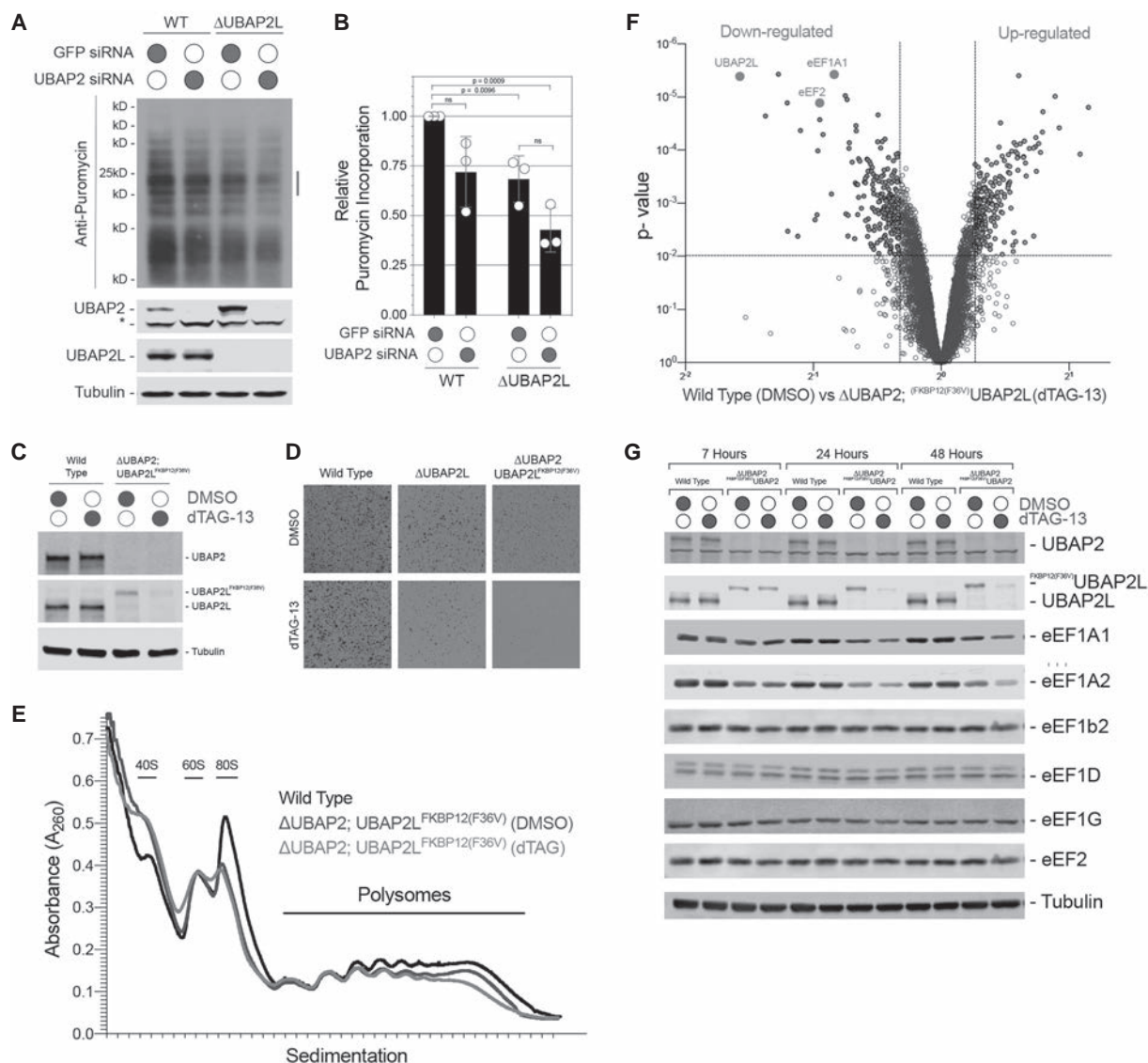
**Figure 3.** UBAP2 and UBAP2L function to promote translation and ensure the translation of TOP mRNAs regulated by mTORC1 signaling. (*A*) RNAi depletion of UBAP2 and UBAP2L result in the reduction of global translation. Puromycin incorporation into nascent polypeptides was used to determine the relative rates of protein synthesis in the indicated cell lines and conditions. (*B*) Quantification of the puromycin-labeling experiments indicate that doubly depleted cells have a significant reduction in translation. (*C*) The d-TAG degradation system efficiently depletes UBAP2 and UBAP2L. (*D*) dTAG-depleted cells exhibit a reduced ability to form colonies in soft agar assays. (*E*) dTAG-depleted cells also show a strong reduction in polysome loading in sucrose gradients consistent with a strong reduction in protein translation. (*F*) The indication by quantitative proteomics (isobaric tags for relative and absolute quantitation [iTRAQ]) is that approximately 120 proteins are extremely sensitive to UBAP2 and UBAP2L depletion. Many of these proteins are encoded by TOP mRNAs that are sensitive to mTORC1 inhibition. (*G*) Western blot analyses of dTAG-depleted cells validate the iTRAQ findings.

*elegans* strains that express the TIR1 E3 ligase in specific cell types. These reagents will enable us to study the developmental clock in individual cells without altering other aspects of development in essential tissues. These experiments will test the requirement of these nuclear hormone receptors for the normal cyclical transcription of clock target genes and proper temporal cell fate specification.

## PQN-59/UBAP2/2L Modulate Transcriptional Output by Maintaining the Levels and Activity of Translational Elongation Factors

O. Huiwu, J. Wang

One of the major developments in the past year has been our characterization of UBAP2 and UBAP2L activity in gene regulation. This project initially began through our identification of PQN-59, the *C. elegans* ortholog of UBAP2 and UBAP2L, as a component that modulates the expression of several key heterochronic genes. Although significant progress has been made in the past toward integrating the activity of PQN-59 in various genetic circuits, we, in general, were unable to determine its molecular function. As a major transition, we decided to study the functions of these genes in human cell culture, in which a variety of more sophisticated molecular and proteomic approaches could be applied to the problem. To this end, we generated conditional expression alleles of UBAP2 and UBAP2L in a variety of human cell lines. First we demonstrated that depletion of UBAP2 and UBAP2L leads to the overall reduction in protein translation as measured by metabolic labeling of nascent translational products in vivo (Fig. 3A,B). We then developed the dTAG system to experimentally deplete UBAP2 and UBAP2L expression through the addition of a compound that specifically and rapidly targets these proteins for degradation (Fig. 3C). Depletion of these proteins leads to a rapid reduction in cell proliferation and failure to form colonies in soft agar (Fig. 3D). Consistent with a reduction of protein translation when UBAP2 and UBAP2L are depleted via RNAi, dTAG-mediated depletion of UBAP2 and UBAP2L reduces overall polysomal loading (Fig. 3E). We next employed the dTAG system to deplete the expression of UBAP2/2L in order to identify gene products that are most dependent on these proteins for normal expression. These experiments employed quantitative proteomics (isobaric tags for relative and absolute quantitation [iTRAQ]) and revealed a highly reproducible set of proteins (approximately 120) that are acutely reduced upon UBAP2/2L depletion (Fig. 3F,G). Analysis of these UBAP2/2L-dependent proteins revealed a remarkable relationship between UBAP2/2L-dependent protein targets: A significant number of these proteins are encoded by the TOP mRNA class, a class of transcripts whose translation is hypersensitive to a variety of growth signals, including those transmitted by the mTORC1 pathway. Future experiments will probe the relationship between UBAP2/2L activities and the regulation of these mRNAs.

*In Press*

Aguirre-Chen C, Stec N, Ramos OM, Kim N, Kramer M, McCarthy S, Gillis J, McCombie WR, Hammell CM. A *Caenorhabditis elegans* model for integrating the functions of neuropsychiatric risk genes identifies components required for normal dendritic morphology. *G3 (Bethesda)* **10:** 1617–1628.

Martinez MAQ, Kinney BA, Medwig-Kinney TN, Ashley G, Ragle JM, Johnson L, Aguilera J, Hammell CM, Ward JD, Matus DQ. 2020. Rapid degradation of *Caenorhabditis elegans* proteins at single-cell resolution with a synthetic auxin. *G3 (Bethesda)* **10:** 267–280.

# SPATIAL FUNCTIONAL GENOMICS IN ORGANOGENESIS AND CANCER

**J. Lee**  D. Chitty    R. Peters
         D. Fürth     X. Wang
         D. Ghosh     X. Yuan
         C. McLaughlin

Our laboratory is developing tools to systematically identify the mechanisms that regulate gene expression during the formation of complex tissue patterns. We recently developed a method to detect and colocalize functional *cis*-regulatory elements and the molecular mechanisms within individual cells across the whole tissue in situ. Our goal is to understand the basic principle of how cells integrate cell–cell interaction to make gene expression decisions in organogenesis. We are now developing a related method to colocalize tumor-specific genetic features in single cells to label and study disseminated cancer cells.

## Identifying *cis*-Regulatory Mechanisms that Interpret Cell–Cell Interactions In Situ

The regulation of spatially coordinated gene expression is a fundamental feature of pattern formation in higher animals. Yet, little is known about the molecular mechanisms that orchestrate the process. In the early 1900s, D'Arcy Thompson argued that cells and their surface geometry could be thought of as bubble-like aggregates in his influential book *On Growth and Form*. Nearly a century later, scientists found that cells in the *Drosophila* compound eye could be modeled after soap bubbles involving homotypic interactions between cadherin proteins; however, this is the exception rather than the rule. In most cases, a network of signaling proteins, mechanical cues, and gene expression programs is thought to determine the organ size or pattern; however, our understanding of how these pathways converge at the genetic level remains limited.

The regulation of gene expression depends on *cis*-acting elements, *trans*-acting factors, and upstream signaling cascades. Modern tools can now map thousands of *cis*-regulatory elements (CREs) across various molecular steps genome-wide. For example,

expression quantitative trait loci (eQTL) analyses have mapped more than 1,000 *cis*-regulatory loci throughout *Drosophila* development in promoter or enhancer regions; however, an additional approximately 7,000 loci are also associated with posttranscriptional RNA processing (Cannavò et al., *Nature* **541:** 402 [2017]). We reasoned that the colocalization of functionally active CREs genome-wide across multiple steps in gene regulation in all cells across the entire tissue is needed to reveal the population-level cell decision-making process in vivo. Hence, we developed spatial single-cell functional genomics, in which the functional loci are directly identified in situ. When fluorescent in situ sequencing (FISSEQ) was first developed (Lee et al., *Science* **343:** 1360 [2014]), it was hampered by short read length, slow turnaround time, and low sensitivity. We have now eliminated these barriers, enabling the identification of functional CREs and the molecular mechanisms genome-wide inside the *Drosophila* retina with subcellular resolution in 3D (Fürth et al. 2019).

In our method, the combination of cDNA start-stop sites serves to index in situ sequencing reads, enabling subsequent tissue lysis and Illumina sequencing linked to a specific subcellular location. To do so, we developed an off-the-shelf in situ sequencing reagent to read the indexing barcode in only six imaging cycles. As a proof of concept, we used the *Drosophila* retina from third-instar larvae. In our experiment, 7,000 cells were imaged, after which the in situ cDNA reads were mapped onto the reference *Drosophila* retina. We detected 340,000 in situ sequencing reads per retina, out of which 33,000 were unambiguously mapped to the paired-end 250-base MiSeq reads. Most poly(dT)-anchored cDNA reads were 90 bases long and aligned mostly to the 3′ UTR. When we compared our result to single-cell RNA-Seq data from the eye imaginal disc, our method was enriched in genes that regulate cytoskeletal

proteins, membrane signaling, and RNA-binding in eye development (Ariss et al., *Nat Commun* **9**: 5024 [2018]). We did not observe housekeeping mRNAs that dominate bulk RNA-Seq or scRNA-Seq. In fact, our result correlated highly with that of ribosome profiling (RIBO-seq) from the imaginal disc. When our method was applied to the purified total RNA, the data structure no longer resembled that of RIBO-seq (Chen and Dickman, *PLoS Genet* **13**: e1007117 [2017]), indicating that the protein or polysome-bound mRNAs were preferentially detected. For these reasons, we have named our method "in situ transcriptome accessibility" sequencing, or INSTA-seq.

When we applied INSTA-seq to larger mammalian cell types, we detected up to 10,000 unique cDNAs (mean = 4,000) per cell. Because of the high density and spatial resolution, multiple subnuclear, ER-associated, or cytoplasmic granules were revealed. This result opens up the possibility of profiling phase-separated granules in developing tissues in situ in an unbiased manner. In addition, our preliminary results suggest that mRNAs bound to the polysomes are better immobilized in situ after cross-linking, providing a plausible explanation of why INSTA-seq reads resemble those of RIBO-seq. We are now attempting to confirm these findings using orthogonal assays so that our method could be used to detect the nascent proteome in situ.

We noted that single-base errors occurred when multiple reads from the same gene terminate at the same site, as in CLIP-seq. For 15% of the reads, the termination site occurred ~10 bases upstream of RNA-binding protein (RBP) motifs. Out of the 15 genes discovered to have functional CREs, eight were known to have developmental phenotypes, including the *Drosophila* homolog of human β-actin, *Act5C*. By examining the nucleotide composition from the last aligned position, we determined probable alternative polyadenylation (APA) sites and searched for RBP motifs that were differentially enriched in APA isoforms. This resulted in the identification of a new motif similar to the zipcode motif (ZIP1/IMP1). The motif was present exclusively in the long isoform, implying that the APA-mediated 3′-UTR cleavage regulates subcellular localization of *Act5C* mRNA. Subsequently, we found that the long isoform was enriched in the optical stalk, whereas the short isoform was enriched in the epithelium, consistent with the role of dIMP in neural migration. We have since validated our

findings using isoform-specific smFISH. These results show that INSTA-seq can delineate multiple layers of regulatory mechanisms across different tissue regions in an unbiased manner. Over the next five years, our goal for INSTA-seq is as follows.

1. Defining *cis*-acting gene regulatory mechanisms with spatial resolution genome-wide: Our goal is to discover the *cis*-regulatory molecular mechanisms that interpret morphogen gradients in the *Drosophila* imaginal disc and generate spatially defined gene expression patterns. The wing imaginal disc has been studied extensively with respect to morphogen signaling and tissue organogenesis, and we will use the conditional alleles of *Decapentaplegic* (*Dpp*) to control tissue region– or developmental stage–specific Dpp expression (Bosch et al., *eLife* **6**: 22546 [2017]) to identify novel pre- or post-transcriptional *cis*-regulatory mechanisms. We are currently automating our sequencing, data analysis, and user-interaction workflows. This will provide a foundation for building a comprehensive in situ database of gene expression, translation, and molecular mechanisms, including 3′-UTR processing, splicing, functional RBPs, and RNA–RNA interactions, enabling one to start unraveling the molecular basis of how cells interpret positional information during animal development.

2. Uncovering new regulatory principles in RNA organization and translation: We find a striking correlation between INSTA-seq and RIBO-seq at the global and molecular levels. We speculate that it might have to do with the enhanced anchoring of mRNAs trapped in the polysomes, protecting specific RNA regions from downstream nucleases. In cultured cells, INSTA-seq reads are also enriched around the rough endoplasmic reticulum (ER), further strengthening the association with the polysomes. To investigate the link between RNA organization and translation, we will compare INSTA-seq to various forms of nascent proteomics measurements, including RIBO-seq and protein mass spectrometry measurements. If these observations could be validated, INSTA-seq could emerge as a potent tool in the field of spatial proteomics for studying the regulation of localized translation in vivo. In the future, it might even be possible to classify clinical specimens based on the nascent proteome signature in an unbiased manner.

## Identifying Genetic Abnormalities That Distinguish Early or Residual Cancer Cells In Situ

Disseminated residual tumor cells can lay hidden in the normal tissue for months to years until they trigger cancer relapse and death. These cells can have altered gene or protein expression patterns from the primary tumor, preventing their detection based on immunophenotype. For instance, multipanel flow cytometry (MFC) is one of the main ways of assessing "minimal residual disease" (MRD) in AML, and it can technically detect one cell out of a million using leukemia-associated immunophenotype (LAIP) antigens. Nevertheless, 50% of MRD-negative patients relapse within a couple of years, suggesting that many residual cells go undetected. Similarly, immunological criteria can also be used to capture circulating tumor cells from solid cancers. However, the molecular phenotype of disseminated single cells can differ substantially from that of primary tumors, making it difficult to quantify MRD based on immunological profiles alone.

Given these challenges, liquid biopsy and NGS-based methods have garnered much attention in recent years, as their sensitivity is limited only by the sequencing depth, at least in theory. What is less discussed is the fact that cell-free methods cannot distinguish oncogenic mutations from a small number of cancer cells versus similar mutations from far more abundant nonmalignant cells. As a result, their clinical sensitivity is capped by the level of preexisting mutations in the body and requires a predisease baseline to interpret the clinical relevance of low-level mutations in each patient. In sum, flow cytometry–based methods are straightforward and compatible with frequent follow-up testing, but they rely on surrogate biomarkers that may not be present on a clinically important subset of cancer cells. On the other hand, sequencing-based molecular tests can detect definitive mutations across a large number of loci, but they are unable to pinpoint the tissue source for low-level mutations, limiting their clinical utility in residual disease detection. What is needed is a sequencing method that can colocalize multiple cancer mutations in single cells for flow cytometry–like quantification.

To do so, we combined key features from antibody-based flow cytometry and in situ sequencing to fundamentally transform the way cancer cells are labeled, quantified, and isolated. Briefly, fixed single cells undergo targeted reverse transcription polymerase chain reaction (RT-PCR), molecular cluster generation, and in situ sequencing. Our workflow differs from INSTA-seq in that it interrogates three-base triplets in one step rather than multiple sequencing cycles one base at a time. When the interrogated codon triplet is wild-type or benign, no label is incorporated, whereas pathogenic codons are fluorescently tagged to label the whole cell. Multiple codon variants can be colocalized using different colors, and the labeled cells are analyzed using flow cytometry. In essence, we have converted sequencing by ligation (SBL) into synthetic anticodon tRNAs capable of recognizing cancer neoantigens, bypassing the need for surrogate biomarkers or antibodies. As a result, any combination of neoantigens can be interrogated in different types of cancers, and their colocalization in specific clones permits one to assess the clinical relevance. Thus, we have named the method CloneSeek. Over the next two years, we will examine disease progression in AML patients whose bone marrow and peripheral blood samples have been assessed by the CloneSeek MRD test. We are also developing ways to colocalize mutations on a glass slide, and we hope to test CloneSeek for tumor cytology in blood, urine, or sputum samples in the future.

### PUBLICATION

Fürth D, Hatini V, Lee JH. 2019. In situ transcriptome accessibility sequencing (INSTA-seq). bioRxiv doi:10.1101/722819

# GENETIC AND EPIGENETIC BASIS OF CANCER AND NEURODEVELOPMENTAL SYNDROMES

**A. Mills**    S. Balinth    M. Fisher    P. Shrestha
              C. Ballon    D. Johnson    S. Sun
              Y. Chang    H. Liang    C. Wu

The goal of our work is to determine the genetic/epigenetic basis of cancer and neurodevelopmental syndromes. We have discovered genes impacting these conditions, have determined how the encoded proteins work in normal cells, and have determined how their deregulation contributes to disease. Our findings have had a major impact and have affected how clinicians analyze and treat patients with these syndromes.

Major discoveries:

- Identifying *p63* as a gene affecting development, aging, and cancer
- Defining the genetic basis of autism
- Discovering *CHD5* as a gene that prevents cancer
- Determining that *CHD5* impacts male infertility

## *p63* in Development, Aging, and Cancer

Having discovered *p63*, a gene related to *p53*—which encodes a tumor suppressor and is defective in most human cancers—my laboratory has been focused on how the p63 protein normally works. Although p63 looked very similar to p53, it was so new that its function was completely unknown. We discovered that loss of p63 (the mouse version of human p63) leads to premature aging, and that it is required for replenishing normal stem cells. Indeed, the lack of p63 causes features of aging such as curvature of the spine, hair loss, and severe skin lesions. Yet, the levels are crucial, as we discovered that the excess of one version of p63 (ΔNp63α) causes carcinoma—the most prevalent type of human cancer. In contrast, we found that a different version of p63 (TAp63) prevents cancer. Our work showing that TAp63 inhibits tumor growth, even when p53 is absent, was surprising: It had previously been assumed that p53 was needed to prevent cancer. We showed that TAp63 can do the job alone.

We had initially discovered that p63 was needed for development, as its loss causes loss of stem cells leading to birth defects of the limbs, skin, and palate. Our findings led others to interrogate p63, and to reveal that its mutation causes seven different human syndromes characterized by birth defects of the limbs, skin, and palate. By generating mouse models for one of these syndromes—ectrodactyly, ectodermal dysplasia, clefting (EEC) syndrome—we found the reason why some children with EEC have symptoms that are very severe and even life-threatening, whereas other children with EEC—even those in the same family with the same p63 mutation—have symptoms that are barely noticeable. Within the past year, we have used gene-editing screens to identify new P63-modulated pathways that drive the cancer stem cell population. Importantly, we found that genetic and/or pharmacological inhibition of these pathways shuts down tumor phenotypes. To extend these findings, we are collaborating with clinical oncologists at Northwell Health (NWH), enabling us to study how p63 impacts carcinomas of the head, neck, cervix, and salivary gland—tissues in which we found p63 to be essential.

## Revealing the Genetic Basis of Autism

We found that one of the most common genetic alterations in autism—deletion of a 27-gene cluster on human chromosome 16—causes autism-like features. By generating mouse models with chromosome deletions corresponding to those at human chromosome 16 using chromosome engineering—a technology that allows us to generate precise chromosome rearrangements in the mouse—we provided the first functional evidence that inheriting fewer copies of genes in this region leads to features resembling those used to diagnose children with autism. Our mouse models had autism-like behaviors such as higher activity, difficulty adapting to change, sleeping problems, and repetitive/

restrictive behavior—each of these are clinical criteria used to diagnose autism in humans. These mice also had changes in brain architecture that were detectable by magnetic resonance imaging (MRI). Our work provides functional evidence for the genetic basis of autism that was not previously appreciated. Within the past year, our collaborative work revealed that genes deleted in autism affect sleeping patterns by impacting rapid eye movement and neural oscillation (Lu et al. 2019). We have also been investing significant effort into generating novel mouse models with subdeletions of regions analogous to the human 1p36 region. We are generating these models for the scientific community, as we believe these models will be invaluable for pinpointing the genes responsible for autism and for identifying more effective treatment regimens.

## CHD5, a New Tumor Suppressor

We discovered *CHD5* as a tumor suppressor mapping to human 1p36—a region of our genomes frequently deleted in a variety of cancers. Although the frequency of cancer-associated 1p36 deletions suggested a tumor suppressor was located in this region, the gene responsible was unknown. By generating mice with deletions and duplications of the genomic region corresponding to 1p36, using chromosome engineering, we discovered *CHD5* as the tumor-suppressor gene in the region and found that its product turns on a network of tumor suppressors. In addition, we found that *CHD5* is frequently deleted in human glioma.

We have been focusing on defining the role of CHD5 in chromatin dynamics and deciphering how dysregulation of CHD5 and the pathways it regulates leads to disease. We found that mouse Chd5 uses its plant homeodomains to bind to histone 3, and that this is essential for tumor suppression. Our work paved the way for further discoveries, and *CHD5* is now known to be mutated in human cancers of the breast, ovary, and prostate, as well as in melanoma,

glioma, and neuroblastoma. Importantly, recent reports indicate that patients with high levels of *CHD5* have much better overall survival rates than those with low levels. We found that Chd5 is essential for packaging DNA, and that loss of Chd5 causes improperly packaged DNA that is prone to DNA damage. Intriguingly, Chd5's absence is particularly important during the process of sperm maturation—an event in which the DNA is first unpackaged and then repackaged using an elaborate series of steps—and that deficiency of Chd5 causes male infertility. We discovered that Chd5 is highly expressed in neurons, and that Chd5 plays a pivotal role in the brain, suggesting that inappropriate DNA packaging contributes to neurodevelopmental syndromes such as autism. We discovered that Chd5 regulates a ribosome biogenesis switch that dictates neuronal cell fate and that Chd5 deficiency leads to an excessive number of astrocytes at the expense of neurons. Within the past year, we collaborated with NWH clinicians to reveal that *CHD5* mutations occur in infertile men, providing an intriguing link between cancer and infertility and highlighting the role of epigenetic processes in these syndromes (Hershlag et al. 2020). We are currently delving deeper into the mechanisms whereby Chd5-mediated regulation of chromatin affects gene expression cascades regulating neuronal stem cells and how deregulation of these processes sets the stage for neurodevelopmental syndromes and cancer.

## PUBLICATIONS

Lu HC, Pollack H, Lefante JJ, Mills AA, Tian D. 2019. Altered sleep architecture, rapid eye movement sleep, and neural oscillation in a mouse model of human chromosome 16p11.2 microdeletion. *Sleep* **42:** zsy253. doi:10.1093/sleep/zsy253

*In Press*

Hershlag A, Peyser A, Bristow SL, Puig O, Pollock A, Niknazar M, Mills AA. 2020. The potential impact of tumor suppressor genes on human gametogenesis: a case-control study. *J Assist Reprod Genet* **37:** 341–346. doi:10.1007/s10815-019-01634-3

# CANCER AND HUMAN GENETICS

| M. Wigler | J. Alexander | J. Kendall | C. Podszus | D. Trimboli |
| | P. Andrews | S. Li | M. Riggs | Z. Wang |
| | C. Danyko | B. Ma | M. Ronemus | M. Wroten |
| | A. Gruet | A. Moffitt | J. Rosenbaum | Z. Yu |
| | I. Hakker | S. Park | A. Stepansky | |

Our research efforts combine methodology development, computational analysis, and clinical and population discovery. We work in three major areas: autism, cancer, and genomics. We collaborate with the groups of Ivan Iossifov (in autism), Alexander Krasnitz (in cancer), and Dan Levy (in all three) at CSHL and with Kenny Ye (in autism) at the Albert Einstein College of Medicine.

## Cancer Genetics

J. Alexander, P. Andrews, C. Danyko, A. Gruet, I. Hakker, J. Kendall, S. Li, B. Ma, A. Moffitt, S. Park, C. Podszus, M. Riggs, J. Rosenbaum, A. Stepansky, D. Trimboli, Z. Wang, Z. Yu [in collaboration with S. Allen, J. Kolitz, G. Goldberg, N. Vincoff, G. Ho, and D. Budman, Northwell Health; L. Muthuswamy, Beth Israel Deaconess Medical Center; T. Baslan, Memorial Sloan Kettering Cancer Center; A. Zetterberg, Karolinska Institutet; D. Levy, A. Krasnitz, J. Sheltzer, and J. Tollkuhn, CSHL]

We have developed and published MASQ (multiplex accurate sensitive detection) to follow tumor load in blood from patients, both in cellular[1] and cell-free DNA. MASQ employs varietal tags to error-correct DNA sequence, thus achieving the high degree of sensitivity and specificity needed. We use the primary tumor to identify tumor-specific variants that we can measure quantitatively and with great sensitivity, even in the presence of a vast excess of normal genomic DNA from patients. The application of MASQ to acute myeloid leukemia (in collaboration with Steve Allen and Jonathan Kolitz at Northwell Health) suggests it will be useful in predicting relapse.[1] We also find signal in blood from ovarian and endometrial cancer (in collaboration with Gary Goldberg at Northwell Health), breast cancer (in collaboration with Nina Vincoff, Gloria Ho, and Daniel Budman at Northwell Health), and pancreatic cancer (in collaboration with Lakshmi Muthuswamy at Boston Deaconess Hospital). Cell-free DNA yields a stronger signal than circulating tumor cells, and so we are in the process of optimizing the method for cell-free signal detection. Collaborations with additional medical centers are being established to further these efforts.

We have developed and published a new method for single-cell nuclear DNA and RNA analysis that allows greater depth of coverage than previous methods, with greater flexibility in its applications and at lower cost.[2] We term this method "BAG" because we embed cells into a ball of acrylamide gel, and then cross-link the cellular nucleic acid to the polyacrylamide matrix. We are using BAG to define the host cells within tumors, to characterize tumor heterogeneity, and to classify the expression patterns of neurons (the latter in collaboration with Jessica Tollkuhn and others at CSHL).

Our method for visualizing the results of single-cell cancer analysis—integrating the profiles with phylogenetic trees, anatomical site, and histopathology—has also been recently published.[3] A study led by Timour Baslan, while he was at CSHL and that uses our methods for examining clonal heterogeneity and phylogeny in breast cancer, has recently been published.[4] The study shows that tumor heterogeneity can be readily monitored by these techniques, and the degree of copy number heterogeneity may be useful in predicting disease outcome. A study led by Jason Sheltzer of CSHL that is based on past years' efforts on copy number measurement—namely, the SMASH technique—has recently been published.[5] It shows that although chromosome imbalance is common in cancer, merely having extra chromosomes actually suppresses tumorigenesis.

Finally, we have continued our collaboration with Anders Zetterberg of the Karolinska Institutet on the role of copy number alterations in the outcome of breast cancer. This work confirms our earlier work on the predictive power of copy number assessment, and

it now points to even greater power when the cancers are classified by their hormone status (in progress).

## Autism Genetics

J. Kendall, M. Ronemus, M. Wroten [in collaboration with I. Iossifov, CSHL/New York Genome Center; D. Levy and T. Janowitz, CSHL; K. Baldwin, Scripps Institute; K. Ye, Albert Einstein College of Medicine]

Continuing our analysis of the contribution of de novo mutation in autism incidence, we have finally determined successfully its contribution in high-risk autism families, the risk class that by definition comprises the entirety of multiplex collections. The contribution of de novo mutation in such families is much smaller than its contribution in simplex families. Because simplex collections are a known proportion of high- and low-risk families, we now estimate the contribution of de novo mutation in low-risk families to be 60%–80%.

We achieved this new analysis using high-coverage whole-genome sequencing (WGS) of multiplex and simplex families. From the high quality of such data, we have also found strong evidence ($p$ values < 0.001) for contribution from de novo mutations striking entirely within the introns between coding exons, as both copy number alterations and small indels. From our work and the work of others, we now estimate the contribution from noncoding mutations to autism to be as much as half the contribution from coding mutations—a significant and largely underappreciated source of disease causality.

To better understand these findings in terms of mechanism, we have initiated and conducted pilot studies on the effects of de novo mutation upon gene expression (in collaboration with the New York Genome Center). We are examining gene expression in cultured blood cell lines from autistic individuals, as well as in the neuronal cell cultures into which they can be induced to differentiate (in collaboration with Kristin Baldwin of the Scripps Institute). These studies are beginning to confirm our previous findings on the impact of intron mutations, but are also pointing to a new and possibly interesting finding: rare sporadic unexplained monoallelic expression. This represents a plausible explanation for some of the unresolved questions of variable penetrance observed with autism genetics, such as discordance between identical twins.

Previous work has shown that the genetic contribution to autism comes from transmitted as well as de novo variation. We have measured this source in the past, as have others, but in recent years we have focused on direct measurements of parental sharing by concordant and discordant siblings. Our results clearly show a role for transmission, surprisingly, most strongly from the father. The paternal source is in direct conflict with predictions we previously made based on the much lower rate of autism in females. We expected that mothers would be the primary carriers of autism risk because of the presumed resilience of females to the cognitive defects caused by disruptive variants. However, it is fathers who appear to be the major carriers of transmissible autism risk. In collaboration with Tobias Janowitz of CSHL, we are now testing the hypothesis that this paternal effect arises from maternal–fetal antigenic incompatibility.

## Genomics

P. Andrews, S. Li, A. Moffitt, Z. Wang [in collaboration with D. Levy and I. Iossifov, CSHL]

Almost all of what we do can be viewed as genomics: single-cell sequencing, measuring and displaying copy number variation, discovery of de novo mutation, detecting tumor-specific variation, phylogenetic analysis of cancers, and even characterizing gene expression. There are, however, two projects that are in a sense purely genomic. These have been described in previous years, but are still ongoing.

MUMDEX is a method using exact sequence matching for characterizing the discontinuity in one genome compared to another. We use it to discover de novo mutation and find tumor variants. We have also used it to discover recent pseudogene formation in humans. It has not yet been fully written up, but it is our intention to return to this task, as we still rely on the method for many applications.

MUSEQ is a method based on template mutagenesis to perform long-range assembly of transcripts and genome segments from short-read sequence platforms. We have published papers on its theory and implementation in past years, but still continue to work through its bench execution and informatics pipeline. It has application to the analysis of complex RNA splicing, to the assembly of regions of the genome for which short-read sequencing fails (e.g., because of excessive repetitive structure), and

to the determination of haplotype phasing over regions where sequence variants are too sparse.

## REFERENCES

1. Moffitt AB, Spector MS, Andrews P, Kendall J, Alexander J, Stepansky A, Ma BC, Kolitz J, Chiorazzi N, Allen SL, et al. 2020. Multiplex accurate sensitive quantitation (MASQ) with application to minimal residual disease in acute myeloid leukemia. *Nucleic Acids Research* **48:** e40.

2. Li S, Kendall J, Park S, Wang Z, Alexander J, Moffitt A, Ranade N, Danyko C, Gegenhuber B, Fischer S, et al. 2019. Copolymerization of single-cell nucleic acids into balls of acrylamide gel. *Genome Res* **30:** 49–61.

3. Chorbadjiev L, Kendall J, Alexander J, Zhygulin V, Song J, Wigler M, Krasnitz A. 2020. An integrated computational pipeline for single-cell genomic profiling. *JCO Clinical Cancer Informatics* **4:** 464–471.

4. Baslan T, Kendall J, Volyanskyy K, McNamara K, Cox H, D'Italia S, Ambrosio F, Riggs M, Rodgers L, Leotta A, et al. 2020. Novel insights into breast cancer copy number genetic heterogeneity revealed by single-cell genome sequencing. *eLife* **9:** e51480.

5. Vasudevan A, Baruah PS, Smith JC, Wang Z, Sayles NM, Andrews P, Kendall J, Leu J, Chunduri NK, Levy D, et al. 2020. Single-chromosomal gains can function as metastasis suppressors and promoters in colon cancer. *Dev Cell* **4:** 413–428.

# CANCER: SIGNAL TRANSDUCTION

Cells respond and adapt to the signals that they receive from their environment. Environmental factors such as nutrients affect cellular states by altering cell state–specific gene expression or metabolic programs. The **Semir Beyaz** research group investigates the causal cellular and molecular mechanisms that link nutrition to organismal health and disease. For example, diets that lead to obesity, such as high-fat diets, are significant environmental risk factors that influence cancer incidence and progression in several tissues. Although the interactions between tumor cells and the immune system play a significant role in tumorigenesis, little is known about how dietary perturbations impact immunity against cancer. The studies of the Beyaz laboratory interrogate the functional consequences of diets on immune recognition and response pathways that play critical roles in cancer immunity. By identifying the altered gene expression and metabolic programs in the immune system in response to dietary perturbations, their goal is to uncover mechanistic links that can be therapeutically exploited for the treatment of diseases associated with immune dysfunction, such as cancer.

**Mikala Egeblad** and colleagues study cancer and, in particular, the microenvironment in which the cancer cells arise and live. Solid tumors are abnormally organized tissues that contain not only cancer cells, but also various other stromal cell types and an extracellular matrix, and these latter components constitute the microenvironment. Communications between the different components of the tumor influence its growth, its response to therapy, and its ability to metastasize. Among the tumor-associated stromal cells, the laboratory's main focus is on myeloid-derived immune cells, a diverse group of cells that can enhance angiogenesis and metastasis and suppress the cytotoxic immune response against tumors. The Egeblad group is interested in how different types of myeloid cells are recruited to tumors and how their behaviors—for example, their physical interactions with cancer cells and other immune cells—influence cancer progression, including metastasis. The Egeblad laboratory studies the importance of the myeloid cells using mouse models of breast and pancreatic cancer and real-time imaging of cells in tumors in live mice. This enables them to follow the behaviors of and the interactions between cancer and myeloid cells in tumors during progression or treatment. This technique was instrumental when the laboratory showed that cancer drug therapy can be boosted by altering components of the tumor microenvironments, specifically reducing either matrix metalloproteinases (enzymes secreted by myeloid cells) or chemokine receptors (signal receptors on myeloid cells). Most recently, the Egeblad laboratory has shown that when a specific type of myeloid cell, called neutrophil, is activated during inflammation, it can awaken sleeping cancer to cause cancer recurrence. The neutrophils do so by forming so-called neutrophil extracellular traps—structures of extracellular DNA—and these alter the extracellular matrix surrounding the sleeping cancer cells to provide a wake-up signal.

**Douglas Fearon**'s laboratory studies the interaction between cancer and the immune system. The underlying premise is that the tumor microenvironment is immune-suppressive because cancer cells elicit responses characteristic of wound healing and tissue regeneration. This approach has led to the finding that activated fibroblasts in the tumor stroma mediate immune suppression in several mouse models of cancer, including the autochthonous model of pancreatic ductal adenocarcinoma of the Tuveson laboratory. Their understanding of the basis of immune suppression is evolving, but they know that it involves the production of the chemokine CXCL12 by the fibroblastic stromal cells, binding of this CXCL12 by pancreatic cancer cells, and exclusion of T cells from the vicinity of the cancer cells. T-cell exclusion, which also occurs in several types of human adenocarcinomas, causes antagonists of T-cell checkpoints to be ineffective, despite the presence of cancer-specific CD8[+] T cells. This immune suppression is interrupted by

administering AMD3100, an inhibitor of CXCR4, the receptor for CXCL12, which leads to the rapid accumulation of T cells among cancer cells, thereby restoring the efficacy of anti-PD-L1 and eliminating cancer cells. Because human pancreatic cancer has certain immunological characteristics of the mouse model, a phase 1 clinical trial of AMD3100 in patients with pancreatic cancer was initiated in 2015; the results were due to be reported in 2020. A phase 2 trial of this immunotherapy for pancreatic cancer is ongoing at Johns Hopkins University School of Medicine in Baltimore, Maryland. Some of the next steps are to determine the biological process that causes cancer cells to express nonmutated, shared antigens, and the means by which dormant metastases escape immune elimination.

The **Tobias Janowitz** laboratory asks the questions, "How do tumors interact with the biology of the host system?" and "What can we learn from studying the physiology and biochemistry of the host system in the context of cancer?" These are principal questions that drive the research in the laboratory. For example, they investigate the convergence of systemic metabolic stress, endocrinology, and suppressed anticancer immunity to discover mechanism-based strategies for combination therapy for patients with cancer. They have shown that interleukin-6-induced metabolic stress is sufficient to down-regulate hepatic ketogenesis. This causes significant systemic stress during periods of caloric deficiency that are often part of the cancer care pathway. The resulting elevation of glucocorticoids suppresses antitumor immunity in model systems of pancreatic cancer. Using clinical samples and data, they have shown correlative findings of weight loss, reduced ketogenesis, and elevated glucocorticoids in patients with pancreatic cancer. Their work, therefore, confirms that cancer cannot be understood and probably cannot be treated by investigating tumors in isolation. They use findings like these to develop strategies for interventional studies with the aim of improving outcome for patients with cancer.

Our genome can encode hundreds of thousands of different proteins, the molecular machines that do the work that is the basis of life. **Darryl Pappin** and colleagues use proteomics, a combination of protein chemistry, mass spectrometry, and informatics, to identify precisely which proteins are present in cells—cells from different tissues, developmental stages, and disease states.

Despite their large variety of genetic abnormalities, cancer cells have been found to be extremely sensitive to the reversal of certain mutations. **Raffaella Sordella** and colleagues study why cells in certain cancers are responsive to the inhibition of one particular gene or gene product. Why, for instance, do non-small-cell lung cancer (NSCLC) cells that have a particular mutation in the epidermal growth factor receptor (EGFR) respond dramatically to its inhibition by the drug Tarceva? This occurs in 15%–20% of patients, the great majority of whom, within 1–3 years, develop resistance. Various mutations have been implicated in about half of resistant patients. Sordella and colleagues have discovered a new resistance mechanism in a subpopulation of NSCLC cells that are intrinsically resistant to Tarceva. These tumor cells were observed to secrete elevated amounts of a growth factor called transforming growth factor-β (TGF-β), which in turn increases secretion of interleukin-6 (IL-6), an immune-signaling molecule. Significantly, these effects were independent of the EGFR pathway. The team therefore hypothesizes that inflammation is one of the factors that can render a tumor cell resistant to treatment with Tarceva. In other work, Sordella collaborates with the Krainer laboratory to study whether alternative splicing has a role in the failure of p53-mediated senescence to halt oncogenesis in certain lung cancers.

**Nicholas Tonks** and colleagues study a family of enzymes called protein tyrosine phosphatases, or PTPs, which remove phosphate groups from proteins and other signaling molecules, such as lipids, in cells. Disruption of PTP function is a cause of major human diseases, and several of the PTPs are potential therapeutic targets for such diseases. Tonks' group seeks to fully characterize the PTP family, understanding how PTP activity is controlled and how PTPs modify signaling pathways. In addition, they are working to determine how those pathways are abrogated in serious

illnesses, including cancer, diabetes, and Parkinson's disease. The overall goal is to identify new targets and strategies for therapeutic intervention in human disease. Tonks and colleagues have defined new roles for PTPs in regulating signaling events in breast cancer, identifying three PTPs as novel potential tumor suppressors. They have characterized the regulation of PTP1B by reversible oxidation, demonstrating that it is regulated by covalent modification of the active site by hydrogen sulfide ($H_2S$) under conditions of endoplasmic reticulum (ER) stress that are linked to protein folding–related pathologies, such as Parkinson's and Alzheimer's. In addition, they have generated recombinant antibodies that selectively recognize the oxidized conformation of PTP1B; these antibodies display the ability to promote insulin signaling in cells and suggest novel approaches to therapy for diabetes. Finally, they have also discovered a novel mechanism for allosteric regulation of PTP1B activity, offering the possibility of developing small-molecule drugs that could inhibit the phosphatase and thereby modulate signaling by insulin and the oncoprotein tyrosine kinase HER2, potentially offering new ways to treat insulin resistance in type-2 diabetes and breast cancer.

**Lloyd Trotman**'s recent research path begins at his discovery some years ago that the loss of a single copy of a master tumor-suppressing gene called *PTEN* is sufficient to permit tumors to develop in animal models of prostate cancer. His team later found that complete loss of *PTEN* paradoxically triggers senescence, an arrested state that delays or blocks cancer development in affected cells. These findings explained why many patients only display partial loss of this tumor suppressor when diagnosed with prostate cancer. Now the team is researching ways to restore the PTEN protein levels in these patients. This therapeutic approach could slow disease progression and thus greatly reduce the need for surgical removal of the prostate or similar drastic interventions that carry the risks of incontinence and impotence. Their second approach to combat prostate cancer is to model the lethal metastatic disease in genetically engineered mice. They are developing a novel approach that allows for quick generation and visualization of metastatic disease. The efficacy of existing and novel late-stage therapies, such as antihormonal therapy, can then be tested and optimized in these animals. At the same time, the Trotman laboratory is exploring the genome alterations associated with metastatic disease and with resistance to therapy. To this end, they use single- and multicell genome sequencing techniques developed at CSHL by Drs. Wigler and Hicks.

**David Tuveson**'s laboratory uses murine and human models of pancreatic cancer to explore the fundamental biology of malignancy and thereby identify new diagnostic and treatment strategies. The laboratory's approaches run the gamut from designing new model systems of disease to developing new therapeutic and diagnostic approaches for rapid evaluation in preclinical and clinical settings. The laboratory's studies make use of organoid cultures—three-dimensional cultures of normal or cancerous epithelia—as ex vivo models to probe cancer biology. Current projects in the laboratory explore changes in redox metabolism associated with pancreatic cancer tumorigenesis, dissect signaling by the Ras oncogene, discover new biomarkers of early pancreas cancer, and identify mechanisms of cross-talk between pancreatic cancer cells and the tumor stroma. Novel treatment approaches suggested by these studies are then tested by performing therapeutic experiments in mouse models. To dissect molecular changes associated with pancreatic tumorigenesis, the Tuveson laboratory has generated a large collection of human patient–derived organoid models. By measuring the therapeutic sensitivities of patient-derived organoids, the laboratory is working to identify novel strategies to treat patients as well as markers of therapeutic response. The Tuveson laboratory maintains strong links to clinical research, and the ultimate goal is confirmation of preclinical findings in early phase trials. Collectively, the laboratory's bench-to-bedside approach is codified as the "Cancer Therapeutics Initiative," and this initiative will provide these same approaches to the entire CSHL cancer community.

Dr. Tuveson serves as Director of the Cold Spring Harbor Laboratory Cancer Center and is the Chief Scientist for the Lustgarten Foundation.

**Linda Van Aelst**'s laboratory studies how aberrations in intracellular signaling involving enzymes called small GTPases can result in disease. They are particularly interested in Ras and Rho GTPases, which help control cellular growth, differentiation, and morphogenesis. Alterations affecting Ras and Rho functions are involved in cancer and various neurodevelopmental disorders. Van Aelst's team has extended its prior study of mutations in a Rho-linked gene called *oligophrenin-1* (*OPHN1*), part of an effort to connect the genetic abnormalities associated with mental retardation to biological processes that establish and modify the function of neuronal circuits. In addition to a role for *OPHN1* in activity-driven glutamatergic synapse development, laboratory members have obtained evidence that *OPHN1* has a critical role in mediating mGluR-LTD (long-term depression), a form of long-term synaptic plasticity, in CA1 hippocampal neurons. Their findings provide novel insight not only into the mechanism and function of mGluR-dependent LTD, but also into the cellular basis by which mutations in *OPHN1* could contribute to the cognitive deficits observed in patients. Defects in cortical neurogenesis have been associated with cerebral malformations and disorders of cortical organization. The Van Aelst team discovered that interfering with the function of the Rho activator DOCK7 in neuronal progenitors in embryonic cerebral cortices results in an increase in the number of proliferating neuronal progenitors and defects in the genesis of neurons. In an extension of these studies, the Van Aelst team this year showed that DOCK7 has a central regulatory role in the process that determines how and when a radial glial cell progenitor "decides" to either proliferate (i.e., make more progenitor cells like itself) or give rise to cells that will mature, or "differentiate," into pyramidal neurons. These lines of research provide novel insight into mechanisms that coordinate the maintenance of the neural progenitor pool and neurogenesis.

# THE TUMOR MICROENVIRONMENT AND METASTASIS: TUMOR–HOST INTERACTIONS THAT CONTROL OR PROMOTE METASTASIS

M. Egeblad   J. Albrengues   C. Evans   D. Ng   L. Sun
             E. Bružas      X. He      J.D. Plenker   P. Vempati
             J. Curtis      L. Maiorino   M. Shevik

Solid tumors are aberrant tissues. Like organs, solid tumors are composed of cancer cells and stroma. The stroma is the supportive framework of the organs and includes the extracellular matrix (ECM) as well as immune cells, fibroblasts, adipocytes, and cells of the vascular system. Interactions between epithelium and stroma are essential for normal organ development as well as for tumor progression. In solid tumors, the stroma is also known as the tumor microenvironment.

We study how the tumor microenvironment influences cancer cells in the context of tumor initiation, growth, drug resistance, and metastasis. We use mouse models of breast, ovarian, and pancreatic cancer together with real-time spinning disk confocal and multiphoton microscopy in living mice, known as intravital imaging. This allows us to study how cancer cell proliferation, migration, and survival are influenced by stromal components in real time.

## Cancer Cell Chemokine Receptor CCR2 Orchestrates Suppression from the Adaptive Immune Response

X. He, P. Vempati, C. Evans, E. Bružas, J. Curtis
[in collaboration with C. dos Santos, D. Fearon, and L. Van Aelst, CSHL]

The immune system is very efficient at eliminating pathogens that can cause harm to the organism. The immune system also has the potential to eliminate neoplastic cells. The concept of "tumor immune surveillance" was first described more than 50 years ago and refers to the ability of immune cells to detect tumor cells and destroy them. T cells, part of the adaptive immune system, are critical for tumor immune surveillance. Immune surveillance may lead to a period in which cancer cells are kept in check by the immune system, and the tumor neither expands nor regresses. Eventually, tumors develop means to escape

immune control. Tumors have multiple mechanisms of escaping immune control, including cancer cell-intrinsic changes that alter how the cancer cell is recognized by the immune system and extrinsic changes that suppress immune cell activities.

Chemokines, or chemotactic cytokines, have critical roles in mediating recruitment of immune cells to sites of inflammation and to tumors. The C-C chemokine ligand 2 (CCL2) recruits CC chemokine receptor (CCR2)-expressing immune cells to tumors. The primary role of CCR2 in cancer has therefore been considered to be the regulation of immune cell infiltration, and we previously showed that CCL2 causes recruitment of CCR2-expressing monocytes to tumors after treatment with chemotherapy and furthermore, that these newly recruited monocytes inhibited the chemotherapy response. CCL2-mediated recruitment of CCR2[+] inflammatory monocytes to the lung has also been shown to promote breast cancer extravasation and metastasis in mice. Furthermore, elevated levels of CCL2 in tumors and in serum are associated with advanced disease and poor prognosis in breast carcinoma patients. These findings have sparked interest in targeting the CCR2 pathway for therapeutic benefit in breast cancer.

It is not just immune cells that express CCR2, breast cancer cells also express CCR2. The potential role(s) of CCR2 signaling in cancer cells have, however, not been well studied, largely because they were thought to be minor compared with the roles of CCR2 in myeloid cells. To test the function of CCR2 in breast cancer cells, we used orthotopic transplantation of MMTV-PyMT breast cancer cells as our breast cancer mouse model. We found that *Ccr2* deletion in cancer cells led to reduced tumor growth and twofold longer survival. The longer survival was accompanied by multiple alterations associated with better immune control: increased infiltration and activation of cytotoxic T lymphocytes

(CTLs) and CD103[+] cross-presenting dendritic cells (DCs), as well as up-regulation of MHC class I and down-regulation of checkpoint regulator PD-L1 on the cancer cells. Pharmacological inhibition of CCR2 increased cancer cell sensitivity to CTLs and enabled the cancer cells to induce DC maturation toward the CD103[+] subtype. The combination of all these changes likely results in more effective immune surveillance and reduced growth of tumors derived from the *Ccr2*[−/−] cancer cells. Indeed, tumors from *Ccr2*[−/−] cancer cells were not growth-restricted in *Batf3*[−/−] mice lacking the CD103[+] DC subtype or in mice lacking CTLs.

Our results establish a novel role for CCR2 signaling in cancer cells in orchestration of the suppression of the immune response. These new data, together with our previous findings regarding the role of CCR2 in recruitment of monocytes that reduce chemotherapy response, make CCR2 a potential therapeutic target in combination with both chemotherapy and immunotherapy.

## Activating a Collaborative Innate-Adaptive Immune Response to Control Metastasis

L. Sun, X. He, D. Ng [in collaboration with S. Adams, NYU]

It was discovered more than a century ago that intratumoral injection of dead bacteria, "Coley's toxin," named after the physician who devised the treatment, led to durable antitumor responses in some patients. It is thought that the responses to the injected bacteria were caused by the activation of Toll-like receptors (TLRs) on immune cells, including macrophages. Many cancers recruit monocytes/macrophages and polarize them into tumor-associated macrophages (TAMs). TAMs promote tumor growth and metastasis and inhibit cytotoxic T cells. Yet, macrophages can also kill cancer cells after polarization by, for example, lipopolysaccharide (LPS, a bacteria-derived TLR4 agonist) and interferon gamma (IFNγ). They do so via nitric oxide (NO), generated by inducible NO synthase (iNOS). Altering the polarization of macrophages could therefore be a strategy for controlling cancer.

We recently revisited Dr. Coley's immunotherapeutic approach: in proof-of-concept studies, we used the TLR4 agonist monophosphoryl lipid A (MPLA) and IFNγ. First, we determined that MPLA with IFNγ activated macrophages isolated from metastatic pleural effusions of breast cancer patients to kill the corresponding patients' cancer cells in vitro. We found that TAMs activated with the combined MPLA + IFNγ treatment (but not with either MPLA or IFNγ given alone) killed ~70%–90% of cancer cells in 48 hours. Next, we determined that intratumoral injection of MPLA with IFNγ not only controlled local tumor growth, but also reduced metastasis in mouse models of luminal and triple negative breast cancers. Furthermore, in an ovarian carcinoma mouse model, intraperitoneal administration of MPLA with IFNγ reprogrammed peritoneal macrophages, suppressed metastasis, and enhanced the response to chemotherapy. Specifically, median survival time of the mice after tumor cell transplantation was increased from 63 days with control treatment, to 106 days when mice were treated with MPLA and IFNγ, and extended beyond our study end point of five months when MPLA and IFNγ treatment was combined with chemotherapy.

In vitro and in vivo analysis revealed the cellular mechanisms responsible for the effects of the combined MPLA + IFNγ treatment: the treatment reprogrammed the immunosuppressive tumor microenvironment to become immunostimulatory by recruiting leukocytes, stimulating type I interferon signaling, decreasing tumor-associated (CD206[+]) macrophages, and increasing tumoricidal (iNOS[+]) macrophages. In vitro, CD8[+] T cells were stimulated through macrophage-secreted interleukin 12 (IL-12) and tumor necrosis factor α (TNFα), suggesting that MPLA + IFNγ may also activate CD8[+] T cells in vivo. Indeed, we found that both macrophages (of the innate immune system) and T cells (of the adaptive immune system) were critical for the antimetastatic effects of MPLA + IFNγ in vivo. MPLA and IFNγ are already used individually in clinical practice, so our strategy to reprogram the tumor immune microenvironment to engage the antitumor immune response, which requires no knowledge of unique tumor antigens, could be tested in clinical trials in the near future.

## Stress-Induced Metastasis

X. He, D. Ng [in collaboration with L. Van Aelst, CSHL]

Increasing evidence suggests that both intrinsic (genetic and epigenetic) changes in the cancer cells and extrinsic changes occurring to the organism or the

microenvironment can drive disseminated, dormant cancer cells to form proliferative metastases. We recently showed that lung inflammation is one extrinsic factor that can cause metastasis, but now we have discovered another: stress. Stress is a complex physiological process initiated by environmental or psychosocial factors leading to a cascade of systemic effects, starting with the processing of information in the central nervous system. The stress response involves release of corticotropin-releasing hormone from the hypothalamus, leading to secretion of the adrenocorticotropic hormone from the anterior pituitary, in turn resulting in release of glucocorticoids from the adrenal gland. Although epidemiological and clinical studies have provided strong evidence for links between chronic stress, depression, social isolation, and cancer development and recurrence, very little is known about the mechanisms by which stress promotes metastasis.

To investigate the impact of stress on metastasis, we have generated a new model of disseminated breast cancer. In brief, breast cancer cells from MMTV-PyMT mice were transplanted into the mammary fat pad of mice. Once primary tumors had grown to a small size, allowing time for dissemination but not for development of macrometastases, we surgically removed the primary breast tumors and started subjecting mice to chronic stress (the restraint stress model). Strikingly, in this model, chronic stress caused mice to develop a fourfold higher metastatic burden compared with unstressed control mice.

An elevated ratio of neutrophils-to-lymphocytes in the blood is associated with poor prognosis in breast and many other cancers, and intriguingly, is also observed in animals and humans subjected to stress. Because we previously identified major pro-metastatic roles for neutrophils, we investigated possible links between stress, neutrophils, and metastasis in our model. We found a fourfold increase in neutrophil recruitment to lungs in stressed mice, and, importantly, depletion of neutrophils reduced stress-induced metastasis. Neutrophils' major normal function is to kill harmful microorganisms. The most peculiar mechanism by which they do so is through formation of neutrophil extracellular traps (NETs)—scaffolds of DNA with associated enzymes that are released into the extracellular space to trap and kill microorganisms. However, we and others have shown that NETs activated by infections,

tobacco smoke, or even cancer cells themselves, also promote metastasis of breast, ovarian, and lung cancer in mice. Strikingly, we now have found that chronically stressed mice have elevated NET levels in their blood and that daily injections with DNase I, which degrades NETs, significantly reduce stress-induced metastasis.

We next asked how stress caused NET formation. Two types of stress hormones, glucocorticoids and adrenaline, are produced by the adrenal gland. We found that NETs were no longer induced after chronic stress exposure in mice that had their adrenal glands removed. We therefore tested whether the stress hormones could directly induce neutrophils to form NETs. We found that glucocorticoids (including dexamethasone used in cancer treatment), but not adrenaline, induced neutrophils to form NETs in vitro.

The signaling mechanisms leading to NET formation are still poorly understood, but often involve activation of the protein arginine deiminase 4 (PAD4) enzyme, which citrullinates histones to initiate chromatin decondensation and nuclear membrane disintegration. However, we found that glucocorticoid-induced NET formation was PAD4-independent, suggesting the involvement of other mechanisms. It was recently shown that neutrophils can repurpose molecules involved in proliferation—cyclin-dependent kinase (CDK) 4/6—to instead form NETs. We found that two CDK4/6 inhibitors used to treat metastatic breast cancer, palbociclib and abemaciclib, inhibited glucocorticoid-induced NET formation in vitro.

Although epidemiological and clinical studies have provided strong evidence for links between chronic stress, depression, social isolation, and cancer development and recurrence, very little has been known about the mechanisms by which stress promotes metastasis. Our study is starting to provide insights into the cellular and molecular mechanisms that mediate stress-induced metastasis. Ongoing work aims to further delineate how glucocorticoids induce NETs, as well as how exactly stress-induced NETs promote metastasis of disseminated cancer cells.

## PUBLICATIONS

Egeblad M, de Visser KE. 2019. Sticking together helps cancer to spread. *Nature* **566:** 459–460. doi:10.1038/d41586-019-00341-4

He XY, Ng D, Van Aelst L, Egeblad M. 2019. Stressing out about cancer immunotherapy. *Cancer Cell* **36:** 468–470. doi:10.1016/j.ccell.2019.10.013

Maiorino L, Egeblad M. 2019. Tumours pick the path to cancer inflammation. *Nat Cell Biol* **21:** 1055–1057. doi:10.1038/s41556-019-0386-2

Rozenblit M, Hendrickx W, Heguy A, Chiriboga L, Loomis C, Ray K, Darvishian F, Egeblad M, Demaria S, Marincola FM, et al. 2019. Transcriptomic profiles conducive to immune-mediated tumor rejection in human breast cancer skin metastases treated with Imiquimod. *Sci Rep* **9:** 8572. doi:10.1038/s41598-019-42784-9

# FIBROBLASTS WITH IMMUNE FUNCTIONS

**D.T. Fearon**    J. Li          R. Yan
                   P. Moresco    M. Yao
                   K. Wang

The Fearon laboratory is expanding our understanding of how the fibroblast regulates the immune microenvironment of diverse reactions, including secondary lymphoid tissues, autoimmune tissues, and tumors.

This past year we have defined the developmental lineage of the two fibroblasts that organize the accumulation and positioning of the two types of lymphocytes that reside in the lymph node, the fibroblastic reticular cell (FRC) that supports T cells, and the follicular dendritic cell (FDC) that supports B cells (Denton et al. 2019). Using a new fate-mapping mouse model, we traced the developmental origin of mesenchymal lymph node stromal cells to a previously undescribed embryonic cell that can be identified by its expression of fibroblast activation protein-$\alpha$ (FAP). FAP+ cells of the lymph node anlagen express lymphotoxin $\beta$ receptor, suggesting they are early mesenchymal lymphoid tissue organizer cells. Clonal labeling shows that FAP+ progenitors locally differentiate into FRCs and FDCs. This process is also coopted in nonlymphoid tissues in response to infection to facilitate the development of tertiary lymphoid structures, thereby mimicking the process of lymph node ontogeny in response to infection. These findings are relevant to our research on the immune microenvironment of cancers because the same fibroblast lineage is found in all solid tumors. Paradoxically, we have shown that the FAP-expressing fibroblasts in a mouse model of pancreatic ductal adenocarcinoma suppress immune reactions in the tumor microenvironment.

We also collaborated with investigators in the United Kingdom to demonstrate that the FAP+ lineage of fibroblasts that supports the development of lymph nodes also participates in the immune reactions in a model of Sjögren's syndrome, an autoimmune response in the salivary gland (Nayar et al. 2019), and in human rheumatoid arthritis (Croft et al. 2019).

In summary, the Fearon laboratory continues to define the fibroblast lineage that supports or suppresses immunity. We aim to discover the signals that cause this cell type to *trans*-differentiate between these diametrically opposed functional states.

## PUBLICATIONS

Croft AP, Campos J, Jansen K, Turner JD, Marshall J, Attar M, Savary L, Wehmeyer C, Naylor AJ, Kemble S, et al. 2019. Distinct fibroblast subsets drive inflammation and damage in arthritis. *Nature* **570:** 246–251. doi:10.1038/s41586-019-1263-7

Denton AE, Carr EJ, Magiera LP, Watts AJB, Fearon DT. 2019. Embryonic FAP+ lymphoid tissue organizer cells generate the reticular network of adult lymph nodes. *J Exp Med* **216:** 2242–2252. doi:10.1084/jem.20181705

Nayar S, Campos J, Smith CG, Iannizzotto V, Gardner DH, Mourcin F, Roulois D, Turner J, Sylvestre M, Asam S, et al. 2019. Immunofibroblasts are pivotal drivers of tertiary lymphoid structure formation and local pathology. *Proc Natl Acad Sci* **116:** 13490–13497. doi:10.1073/pnas.1905301116

# HOST RESPONSE TO CANCER

**T. Janowitz**    B. Demestichas        C. Koza
                   M. Ferrer Gonzalez    E. von Gablenz
                   Y. Gao                E. Williams
                   A. Inguscio           X. Zhao

Cancer is a systemic disease. Using laboratory and clinical research, we investigate the connections between metabolism, endocrinology, and immunology to discover how the body's response to a tumor can be used to improve treatment for patients with cancer.

We have shown that tumors can reprogram the host metabolism, induce a systemic endocrine stress response, and thereby suppress immunity systemically to a degree that is sufficient to cause failure of cancer immunotherapy. Our work, therefore, confirms that cancer cannot be understood and probably cannot be treated by investigating tumors in isolation. We use findings like these to develop strategies for interventional studies with the aim of improving outcomes for patients with cancer.

Based on large-scale clinical data sets and statistical modeling, we also develop models for patient physiology and organ function with the aim of stratifying patients better and to understand when they would benefit from host-targeted treatment. To translate our work, we perform clinical studies and trials that combine host reprogramming with cancer immunotherapy.

## Validation of a New Model to Predict Glomerular Filtration Rate in Patients with Cancer

E. Williams [in collaboration with an international team of oncologists from the United Kingdom, Sweden, Switzerland, and the United States]

Important oncological management decisions rely on kidney filtration function assessed by serum creatinine–based estimated glomerular filtration rate (eGFR). However, no large-scale multicenter comparisons of methods to determine eGFR in patients with cancer are available. To compare the performance of models and equations for eGFR based on routine clinical parameters and serum creatinine not calibrated with isotope dilution mass spectrometry, we studied 3,620 patients with cancer and 166 without cancer who had their glomerular filtration rate (GFR) measured with an exogenous nuclear tracer at one of seven clinical centers. The mean measured GFR was 86 mL/min. Accuracy of all models was center dependent, reflecting intercenter variability. A model developed by our group, CamGFR (Fig. 1), was the most accurate model for eGFR (root-mean-square

$$GFR = \Big(1.81395 + 0.01914{\times}Age + 4.73278{\times}BSA - 0.02970{\times}Age{\times}BSA - 3.71619{\times}\log(Scr)$$
$$+ 1.06284{\times}\log(Scr)^2 - 0.91420{\times}\log(Scr)^3 + (0.02020 + 0.01247{\times}Age)[\text{if male}]\Big)^2$$

where

- GFR [mL/min] - Glomerular Filtration Rate
- Age [years]
- BSA [m$^2$] - Body Surface Area
- Scr [mg/dL] - Serum Creatinine Concentration

**Figure 1.** The equation of the CamGFR model for estimation of glomerular filtration rate in patients with cancer.
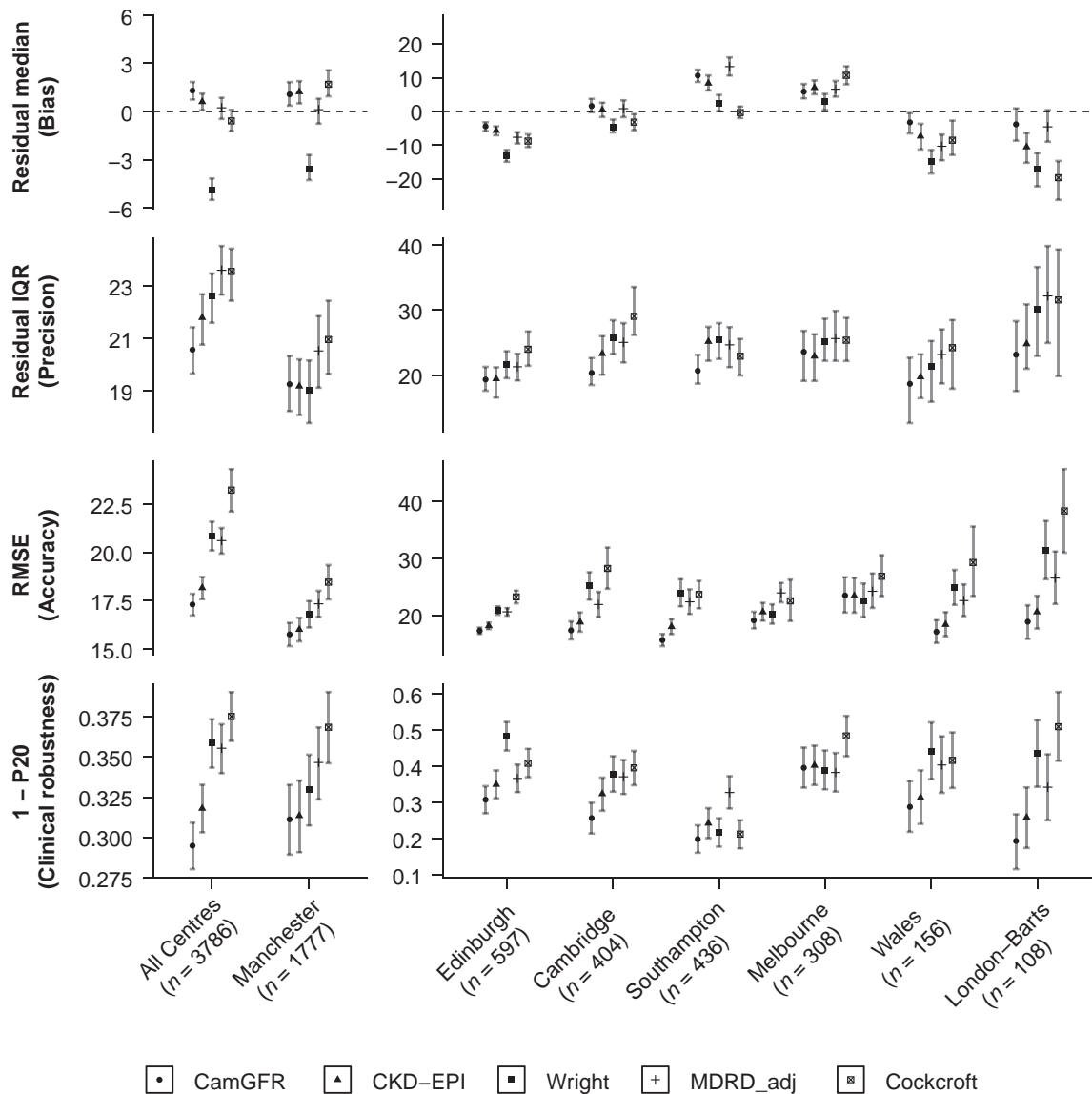
**Figure 2.** Performance analysis of commonly used estimated glomerular filtration rate (eGFR) models. Results for the five best-performing models (CamGFR, CKD-EPI, Wright, MDRD-186, and Cockcroft-Gault) for the 3,776 patients from the non–isotope dilution mass spectrometry (IDMS)–creatinine validation data set are displayed. A pooled analysis of data from all centers and the individual center analyses are shown along the *x*-axis. (*First row*) The residual (measured GFR–eGFR) median, which is a measure of a model's bias, is displayed. (*Second row*) The residual interquartile range (IQR), which is a measure of a model's precision, is displayed. (*Third row*) The root-mean-square error (RMSE), which is a measure of a model's accuracy, is displayed. Accuracy is a combination metric of bias and precision. (*Fourth row*) The proportion of patients who have an absolute percentage error of >20% (1-P20), which reflects clinical robustness by illustrating the proportion of patients with a clinically relevant error, is displayed. The best results are closest to zero for the residual median and the smallest value for IQR, RMSE, and 1-P20. All error bars are 95% confidence intervals calculated using bootstrap resampling with 2,000 repetitions and a normal distribution approximation. (CKD-EPI) Chronic Kidney Disease Epidemiology Collaboration, (MDRD-186) Modification of Diet in Renal Disease version 186.

error 17.3 mL/min), followed by the Chronic Kidney Disease Epidemiology Collaboration model (root-mean-square error 18.2 mL/min) (Fig. 2). We propose that our model should be used as a new international standard for eGFR using serum creatinine not calibrated with isotope dilution mass spectrometry and are planning to expand this work to develop a universal model that can be used for all laboratory-based creatine measurements. We have made our model freely available in the form of an online tool (Williams et al. 2019).

## The Connected Metabolic, Endocrine, and Immunological Response to Pancreatic Cancer

M. Ferrer Gonzalez, E. von Gablenz, Y. Gao, A. Inguscio

Using murine model systems of primary and metastatic cancer, we determined the effect of cancer progression on the processing of nutrients in different organ systems. Ph.D. student Miriam Ferrer explores how cancer progression alters lipid and carbohydrate metabolism in the liver and tumor and how these alterations affect the metabolism after lipid-rich diet intake. M.D. student Eva von Gablenz and postdoctoral immunologist Alessandra Inguscio study whether metabolic stress is sufficiently immunosuppressive to engender cancer initiation and progression of metastatic cancer.

## Clinical Effects of CXCR4 Inhibition

This work was done in collaboration with the Fearon Laboratory (CSHL); and Jodrell Laboratory (Cambridge, UK)

Research technician Ya Gao and Dr. Janowitz have worked with collaborating laboratories and clinicians to complete the laboratory and clinical analyses for a clinical study that targets the CXCR4 chemokine receptor. Patients with pancreatic cancer and colorectal cancer were treated for seven consecutive days with AMD3100, a small-molecule inhibitor of CXCR4. The study has revealed that CXCR4 inhibition induces an integrated immune response in pancreatic and colorectal cancer metastases that is predictive of response to immune checkpoint inhibition. A phase 2 clinical trial building on these results is currently being set up.

## Collaborations at CSHL

Research technician Breanna Demestichas and postdoctoral biochemist Xiang Zhao are setting up a research collaboration with the Wigler Laboratory to study the similarities and differences of the immunological responses to cancer and to normal tissue.

## Collaborations with Northwell Health

To further strengthen the strategic partnership between CSHL and Northwell Health, Dr. Janowitz has joined the newly formed Cancer Institute at Northwell Health as an academic medical oncologist.

PUBLICATIONS

Luo Y, Yang Y, Liu M, Wang D, Wang F, Bi Y, Ji J, Li S, Liu Y, Chen R, et al. 2019. Oncogenic KRAS reduces expression of FGF21 in acinar cells to promote pancreatic tumorigenesis in mice on a high-fat diet. *Gastroenterology* **157:** 1413–1428.

Williams EH, Connell CM, Weaver JMJ, Beh I, Potts H, Whitley CT, Bird N, Al-Sayed T, Monaghan PJ, Fehr M, et al. 2019. Multicenter validation of the CamGFR model for estimated glomerular filtration rate. *JNCI Cancer Spectr* **3:** pkz068. doi:10.1093/jncics/pkz068

# MASS SPECTROMETRY

**D. Pappin**   S. Costa   A. Makarenko   K. Marquart   K. Rivera

## Regulation of PTP1B Activation through Disruption of Redox-Complex Formation

D. Pappin, K. Rivera [in collaboration with N. Tonks, CSHL; B. Boivin, A. Londe, S. Curley, A. Kannan, G. Coulis, and S. Rizvi, SUNY Polytechnic Institute, Albany; R. Linhardt and F. Zhang, Rensselaer Polytechnic Institute, Troy; A. Bergeron, Montreal Heart Institute, Montreal, Canada; S.J. Kim, Korea Research Institute, Daejeon, South Korea]

In a collaboration with the Boivin and Tonks laboratories, the Mass Spectrometry (MS) laboratory identified a molecular interaction between the reversibly oxidized form of protein tyrosine phosphatase 1B (PTP1B) and 14-3-3ζ that regulates PTP1B activity. Destabilizing the transient interaction between 14-3-3ζ and PTP1B prevented PTP1B inactivation by reactive oxygen species and decreased epidermal growth factor receptor (EGFR) phosphorylation. The data suggest that destabilizing the interaction between 14-3-3ζ and the reversibly oxidized and inactive form of PTP1B may establish a path to PTP1B activation in cells.

## The Glycan CA19-9 Promotes Pancreatitis and Pancreatic Cancer

D. Pappin, K. Rivera [in collaboration with D. Engle, H. Tiriac, A. Pommier, T. Oni, B. Alagesan, E.J. Lee, M. Yao, M. Lucito, B. Spielman, B. Da Silva, C. Schoepfer, K. Wright, B. Creighton, L. Alfinowicz, Y. Park, D. Tuveson, CSHL; S. Whalen and K. Pollard, Gladstone Institute, San Francisco; K. Yu, Memorial Sloan Kettering Cancer Center; New York; R. Gruetzmann and C. Pilarsky, University Klinikum Erlangen, Erlangen, Germany; D. Aust, University Klinikum Dresden, Dresden, Germany; P. Gimotty, University of Pennsylvania, Philadelphia; R. Hruban and M. Goggins, Johns Hopkins University, Baltimore; M. Hollingsworth, University of Nebraska, Omaha]

Glycosylation alterations are thought to be indicative of tissue inflammation and neoplasia, but whether these alterations contribute to disease pathogenesis is largely unknown. To study the role of glycan changes in pancreatic disease, the Tuveson laboratory had previously inducibly expressed human fucosyltransferase 3 and β1,3-galactosyltransferase 5 in mice, reconstituting the glycan sialyl-Lewis[a] (CA19-9). Remarkably, CA19-9 expression in mice resulted in rapid and severe pancreatitis with hyperactivation of EGFR signaling. The Pappin laboratory was able to identify the mechanism, in which CA19-9 modification of a matricellular protein, fibulin-3, increased its interaction with EGFR. Blockade of fibulin-3, EGFR ligands, or CA19-9 prevented EGFR hyperactivation in organoids. CA19-9-mediated pancreatitis was shown to be reversible and could be suppressed with CA19-9 antibodies. CA19-9 also cooperated with the $Kras^{G12D}$ oncogene to produce aggressive pancreatic cancer. These findings were the first to implicate CA19-9 in the etiology of pancreatitis and pancreatic cancer, and nominate CA19-9 as a therapeutic target.

## Muller's Ratchet and Ribosome Degeneration in Microsporidia

K. Rivera, A. Makarenko, D. Pappin [in collaboration with S. Melnikov, K. Manakongtreecheep, and D. Soll, Yale University, New Haven]

Microsporidia are fungal-like parasites that have the smallest known eukaryotic genome and are often used as a model to study the phenomenon of genome decay. Similar to other intracellular parasites that reproduce asexually in an environment with alleviated natural selection, microsporidia experience continuous genome decay that is driven by Muller's ratchet—an evolutionary process of irreversible accumulation of deleterious mutations that lead to gene loss and the miniaturization of cellular components. Microsporidia have remarkably small ribosomes in which the rRNA is reduced to the minimal enzymatic core. Through MS of the microsporidian proteome and parallel analysis of microsporidian genomes, it was found that massive rRNA reduction in microsporidian ribosomes appears to annihilate the binding sites for ribosomal proteins eL8, eL27, and eS31, suggesting that these proteins were no longer bound to the ribosome

in microsporidian species. The study illustrated that, although microsporidia carry the same set of ribosomal proteins as nonparasitic eukaryotes, some ribosomal proteins that are no longer participating in protein synthesis are preserved from genome decay by having extraribosomal functions.

**PUBLICATION**

Engle DD, Tiriac H, Rivera KD, Pommier A, Whalen S, Oni TE, Alagesan B, Lee EJ, Yao MA, Lucito MS, et al. 2019. The glycan CA19-9 promotes pancreatitis and pancreatic cancer in mice. *Science* **364:** 1156–1162. doi:10.1126/science.aaw3145

# DRUGGING THE "UNDRUGGABLE": AN ALTERNATIVE APPROACH TO CANCER TARGET THERAPIES

**R. Sordella**    P. Sefaric

Since the advent of molecular biology and more recently of next-generation genome sequencing technology, considerable progress has been made in the quest to understand the mechanisms underlying tumorigenesis. An astonishing number of cancer mutations and genome-phenotypic associations have been described. It is now possible to imagine that an almost complete catalog describing the functional relevance of all human cancer genetic variations—deleterious, advantageous, or neutral—will be completed in the near future.

Some of these findings have been exploited for the design of new cancer therapies. More than 10 years ago now, we and others showed that certain mutations driving a subset of lung cancers could be targeted by small molecules. This finding revolutionized the treatment of cancer and gave rise to the field of "cancer-targeted therapy."

Despite the natural excitement emerging from such a seminal discovery, daunting challenges remain about the possibility of extending the "genomic" "targeted" therapy revolutions into the development of new therapeutic strategies beyond a little more than a handful of cancer treatments. In fact, it turned out that druggable driving mutations are rare. So far, only ~10% of identified cancer mutations are actionable by either small molecules or antibodies.

In the past year, to address this hurdle, we created toolkits and a defined pipeline that enabled us to leverage the principles, rules, and functional modules of cellular networks to predict the response of the network to a particular perturbation and ultimately to identify novel druggable targets in cancer driven by "unactionable" mutations.

One interesting aspect of cellular networks is that the activity of a given genetic mutation or perturbation is not restricted only to that specific gene product, but it can diffuse, migrate throughout the intricate network of connections that define cellular functions, modify the activity of other network elements, and impact the network's overall topology. We argue that this could provide new opportunities for targeting "undruggable" cancer mutations. In other words, having a map of the intricate wiring diagram of cancer cellular components, similar to a mechanic trying to fix a car, could enable us to think "globally" and to act in points of the network that are unique and essential. Although with these approaches part of the richness and complexity of gene product interactions such as posttranslational modifications, allosteric changes, and activities are often lost, this simplification enables us to model gene interaction at scale and to more easily visualize and identify critical nodes in the network.

As a paradigm we used a particular subset of bone cancer called mesenchymal chondrosarcoma (MCS). Bone tumors are highly heterogeneous and are driven by different mutations. Mesenchymal chondrosarcoma in particular is a rare high-grade variant of chondrosarcoma that was first described in 1959. It only accounts for 1%–10% of all chondrosarcomas, with less than 800 cases described so far. Histologically, MCS presents as small cells with mesenchymal-like features disseminated in islands of atypical cartilage. Differently from typical chondrosarcomas, MCS is more frequent in young adults (age 10–30) and in females. MCS tends also to be more aggressive, with 5- and 10-year survival rates of 54.6% and 27.3%, respectively.

Because of its infrequency, MCS remains poorly understood and often misdiagnosed. Recently it has been shown that the large majority of MCSs are characterized by a deletion of chromosome 8q that brings together part of the transcriptional repressor HEY1 and the epigenetic modifier NCOA2. Although the precise molecular function of the HEY1-NCOA2 fusion is still unknown, this specific gene fusion has been used as a genetic marker to diagnose MCS.

Currently the primary form of treatment for a mesenchymal chondrosarcoma is surgery and adjuvant chemotherapy. The latter is based on standard chemotherapy regimens that are known to be active against most sarcomas such as alternating cycles of etoposide

with ifosfamide and adriamycin with vincristine plus cyclophosphamide. Yet the high relapse rate suggests that MCSs are not very sensitive to these treatments and emphasizes the need for new therapies.

Genotype–phenotype relationships are highly contextual and rarely depend on the abnormality of a single effector gene product—they require specific network components and arrangements.

Hence, to study MCS driven by the HEY1-NCOA2 fusion, our laboratory created a unique cellular system in which we expressed this mutation within human primary chondrocytes from different donors via lentiviral infection.

We observed that only the fusion, but none of our controls, was able to modify the morphology of cells and increased their growth and motility. When injected into immunocompromised mice, the cells expressing this mutation formed tumors that histologically resembled MCS.

Gene expression profile analysis of these cells revealed that, differently from control cells, the primary human chondrocytes expressing the fusion expressed genes that typify cells in a mesenchymal and proliferative state. Of particular interest, the HEY1-NCOA2 cells also express genes that participate in wound-healing responses. These observations are notable because they seem to recapitulate at the molecular level features observed in primary tumors.

Next, to identify possible druggable targets, our laboratory "reverse-engineered" gene expression data sets to generate functional cellular networks and infer regulatory interactions among genes using ad hoc computation algorithms.

Thinking of the complexity of cellular networks is intimidating. The expected number of nodes and of functionally relevant interactions between the network components is in fact on the order of hundreds of thousands. In the past, various interactome and functional relationship networks comprising all genes expressing in cells in different conditions have been compiled and integrated to generate functional networks. These studies revealed that cellular networks are not random but follow a series of basic organizing principles in their structure and evolution that distinguish them from randomly connected networks. In fact, it has been observed that most nodes have a different number of links, and although some are highly connected (e.g., hubs), others are characterized by singular connections (e.g., edges). Furthermore, not all

hubs are identical. Their role in coordinating specific processes within or between modules of a functional cellular network allowed their classification into "party" and "date" hubs. Cellular networks also display areas with "small world" property, which means that most proteins or network elements in these areas are only a few interactions from any other network components. This implies that hubs with a high centrality are ideal actionable nodes as they often hold the whole network together.

When we generated a functional gene-cellular network based on HEY1-NCOA2 fusion or control data sets, we observed several network articulation points (e.g., nodes that are critical for communication within the network and that when removed create breaks and isolated modules) that were unique to the cells expressing the fusion.

When we mapped all FDA-approved drug targets on these functional networks, we observed that although at first glance the targets of FDA-approved drugs appeared widely distributed, detailed inspection showed that they were concentrated in certain areas, often clustered together. This was the case with Neuropillin1, Abl 1/2, PDGFR, and VEGFR.

This was particularly interesting to us because increasing attention has recently been paid to therapies involving multiple targets that may be potentially more effective in reversing the disease phenotype than single drugs. Most disease phenotypes are in fact difficult to reverse through the use of a single "magic bullet" targeting a single node in the network. The possibility of targeting simultaneously multiple hubs of the cellular network could provide a more robust and efficacious drug treatment by preventing the rewiring of the network, decreasing the adaptability of the cells to the drug perturbation, and increasing the efficacy by composite target inhibition.

This prediction paralleled our experimental findings. Although treatment of cells expressing the HEY1-NCOA2 fusion with standard chemotherapy or inhibitors targeting only PDGFR or ABL showed very limited efficacy, treatment with PDGFR and ABL inhibitors together or with drugs such as imatinib or rerogafenib targeting simultaneously PDGFR, Abl1/2, and VEGFR dramatically reduced the number of viable cells.

To increase the relevance of our observations to the human disease, we extended these studies to a unique

MCS patient–derived xenograft (i.e., PDX) model we created.

Also, in this case, we were able to show that 15 days of treatment with imatinib resulted in a decrease in the tumor mass and density compared to control. Histological analysis confirmed this observation and in addition showed a dramatic reduction of cancer cells in the treated cohort.

In summary, the studies we conducted uncovered and validated a new possible treatment for MCS—an incurable tumor—and opened the possibility of using the tools we have developed and our conceptual framework to identify new treatments for incurable cancers, especially for the ones that are driven almost exclusively by a single mutation.

# PROTEIN TYROSINE PHOSPHATASES AND THE CONTROL OF SIGNAL TRANSDUCTION

**N.K. Tonks**   C. Bonham      C. Felice    V. Mandati   P. Venkataramani
                 Y. Cen         C. Gauss     Z. Qian      S. Zhang
                 L. Christensen Q. Hu        D. Song      I. Zubiete Franco

As cells encounter stimuli, such as growth factors, cytokines, and hormones, receptors on the cell surface modulate the activities of protein kinases and phosphatases. The functions of these enzymes, which promote the addition and removal of phosphate groups, are coordinated in signal transduction pathways to mediate the cellular response to the environmental stimuli. These pathways are of fundamental importance to control of cell function and their disruption frequently underlies major human diseases. Consequently, the ability to modulate such signal transduction pathways selectively with drugs holds enormous therapeutic potential. In the area of tyrosine phosphorylation-dependent signal transduction, drug discovery efforts to date have emphasized the protein tyrosine kinases (PTKs). Although there have been spectacular successes, challenges remain, including the acquisition of drug resistance. Considering the reversibility of protein tyrosine phosphorylation, there is the potential to manipulate signal transduction pathways at the level of both PTKs and protein tyrosine phosphatases (PTPs). Although the PTPs have been garnering attention as potential therapeutic targets, they remain largely an untapped resource for drug development. The long-term objectives of the Tonks laboratory are to characterize the structure, modes of regulation, and physiological function of members of the PTP family of enzymes. Through basic research to understand the mechanism of action and function of PTPs, the Tonks laboratory is trying to devise creative new approaches to exploiting these enzymes as targets for therapeutic intervention in major human diseases, including diabetes, obesity, and cancer.

During the last year, we were joined by Vinay Mandati, a new postdoctoral fellow, and Lisa Christensen, who will be helping with various animal studies.

## PTPN23

Although most studies of the PTP family have focused on their enzymatic function as phosphatases, there are several examples in which nonenzymatic functions have been recognized as important. To survey systematically the genetic dependency of a cancer model on PTP function, we collaborated with the Vakoc laboratory to use the CRISPR-Cas9 approach to ablate the PTPs in a murine acute myeloid leukemia (AML) cell line. The screen identified *Ptpn23* as one of the top hits, which was specifically required for mouse and human AML cells but dispensable for mouse bone marrow cells. Interestingly, this reflects a positive role for PTPN23 in AML cell proliferation and survival, which contrasts with the tumor suppressor function we showed for this enzyme in breast cancer. PTPN23 is an ESCRT-associated protein (endosomal sorting complexes required for transport). ESCRTs are multimeric protein complexes mediating a number of important physiological processes, including multivesicular body (MVB) formation, cytokinetic abscission, autophagy, and membrane repair. In addition to the carboxy-terminal phosphatase domain, PTPN23 has multiple functional domains, including amino-terminal Bro1, V, and His domains that associate with ESCRT components. In a rescue experiment, we showed that a truncated mutant form of PTPN23, which comprised residues 1–872 and lacked the phosphatase domain, was sufficient to sustain AML cell survival, thus illustrating a function of PTPN23 that was independent of catalytic activity. Upon ablation of *PTPN23*, we observed increased NF-κB signaling in AML cells. In collaborative studies with Darryl Pappin (Mass Spectrometry Shared Resource), we conducted structure–function analyses to identify the segments of PTPN23 that were critical for these effects and used BioID to identify important binding

proteins. Going forward, we plan to define the role of NAP1 as an adaptor protein to link PTPN23 to trafficking of TNFR1, and to define the importance of TNFRs in the mechanism of AML cell death caused by depletion of PTPN23.

## Targeting PTP1B Therapeutically

A current focus of the laboratory remains an examination of the effect of our allosteric inhibitors of PTP1B, such as MSI-1436, on tyrosine phosphorylation in HER2-positive breast cancer cells, to define which signaling pathways downstream from HER2 are regulated by PTP1B. Furthermore, a major problem exists with current therapeutic strategies in that 70% of HER2-positive breast cancer patients display de novo resistance to the frontline targeted therapy Herceptin (trastuzumab), and the remaining 30% that do respond initially acquire resistance in ~2 years. Consequently, the identification of alternative or combinatorial targets for therapeutic intervention is desperately needed. Preliminary data suggest that our allosteric PTP1B inhibitors may overcome both de novo and acquired resistance to Herceptin in cell models. We continue to investigate the mechanisms underlying these effects.

## Novel Copper Chelators

Although our original allosteric inhibitor, MSI-1436, shows efficacy in an injectable format, like many PTP1B inhibitors identified to date it also shows limited oral bioavailability. In 2018, we published two papers describing the identification and characterization of an orally bioavailable analog of MSI-1436 that is a potent and specific inhibitor of PTP1B and crosses the blood–brain barrier. This compound, termed DPM-1001, shows a unique specificity and high affinity for copper that enhances its potency as an inhibitor of PTP1B. This has focused our attention on the ability of DPM-1001 to function as a copper chelator, particularly in a cancer context.

The phrase "oncogene addiction" is used to describe a situation in which some cancers become dependent on (addicted to) a small number genes for both maintenance of the malignant phenotype and cell survival. Copper represents a newly discovered form of such addiction. The activities of critical enzymes in the control of cell growth and survival, as well as in metastasis, are enhanced by copper. This offers a unique and novel point of therapeutic intervention in cancer. Interestingly, Linda Vahdat (Memorial Sloan Kettering Cancer Center) has shown proof of concept for chelation of copper as an effective therapeutic strategy in a phase 2 study of patients with breast cancer. Spectacular increases in disease-free survival were reported in this trial. What is needed are new, potent, and specific chelators of copper to exploit this opportunity fully.

In collaboration with a small company, DepYmed, we have shown that elevated copper is a defining feature of a wide array of different cancers and is associated with increased levels of a copper transporter, CTR1. In fact, we anticipate that measurements of tumor copper levels will represent a powerful tool for prioritizing patients for future clinical trials. We have shown that DPM-1001 induced cytotoxicity in those cancer cells that feature, and are addicted to, elevated copper. Efficacy has been shown in triple negative breast cancer and melanoma models, including cancer cell lines in culture, in tumor xenografts, and now in organoid models derived from human tumor material. Our data indicate that copper chelation may offer a new approach to treatment of a wide variety of different cancers and we are working with Dr. Vahdat to take one of these copper chelator/PTP1B inhibitors into clinical trials.

SAR syntheses around the structure of DPM-1001 (methyl 4-[7-hydroxy-10,13-dimethyl-3-({4-[(pyridin-2-ylmethyl)amino]butyl}amino)hexadeca hydro-1H-cyclopenta[a]phenanthren-17-yl] pentanoate) have allowed us to define which parts of the molecule are important for copper binding, which parts confer specificity for copper over other metals, and which parts are important for inhibition of PTP1B. We have now succeeded in resolving copper chelation from inhibition of PTP1B and have generated small molecules that chelate copper with high affinity (10–20 nM) and specificity, but do not inhibit PTP1B in vitro. This opens up a clean mechanism of action focused on chelation of copper, without potentially confounding effects of inhibition of PTP1B. Future studies will focus on further optimizing and characterizing such molecules in various cancer models.

In addition, we have validated a further analog of DPM-1001, called DPM-1013, and showed that it blocks tumor growth in xenografts of A2058 melanoma cells, exerting comparable effects to DPM-1001. In

immunohistochemistry analyses, we have shown that DPM-1001 and DPM-1013 arrest proliferation (Ki67 staining) and trigger apoptosis (cleaved PARP staining) in A2058 melanoma cell xenografts. A collaboration has been initiated with Scott Lyons (Animal Imaging Shared Resource) to visualize these changes in copper in tumor models.

We have shown that the activity of enzymes such as pyruvate kinase and cytochrome c oxidase, which bind copper tightly, were not altered by DPM-1001. In contrast, the activity of kinases such as MEK and PAK was enhanced in cells in which copper had accumulated to aberrantly high levels—this effect was reversed by treatment with DPM-1001. Future studies, in collaboration with Darryl Pappin (Mass Spectrometry Shared Resource) will focus on defining more precisely the effects of elevated copper on signaling in cancer cells.

## Developing Inhibitors of the Protein Kinases PIM and DYRK

In a long-standing collaboration with the laboratories of Darryl Pappin and Leemor Joshua-Tor, we have purified and characterized CSH-4044, a small-molecule natural product that we isolated from fermented wheat germ extract. We showed that CSH-4044 has a unique structure and a unique specificity for PIM and DYRK family kinases. PIMs and DYRKs have been implicated in a wide variety of hematological and epithelial tumors, with the expectation that inhibitors of these kinases may have broad therapeutic utility. In collaboration with Elad Elkayam and Leemor Joshua-Tor, the crystal structure of the kinase PIM-1 was determined in a complex with CSH-4044 at 1.95 Å resolution. CSH-4044 binds to the ATP binding site of PIM-1 in two alternate conformations due the symmetrical nature of the inhibitor. We identified several specific interactions between the inhibitor and the protein that would help to guide the synthesis of analogs of the compound. A formal collaboration was established with Vichem Chemie to support a medicinal chemistry program required for optimizing CSH-4044. A synthetic route was established and ~180 analogs were produced. We generated inhibitors that were ATP-competitive and displayed both improved potency and selectivity relative to CSH-4044. We identified compounds that inhibited

DYRK preferentially, inhibited PIM preferentially, or inhibited both. Specificity was confirmed by profiling against a panel of 140 kinases and predictive ADME analysis confirmed drug-like properties.

Now, we are focusing on the DYRKs as our targets because we believe they offer the best opportunities for therapeutic development. The DYRKs belong to the CMGC family of protein kinases, which includes cyclin-dependent kinases, MAP kinases, and glycogen synthase kinases. Newly translated DYRKs undergo intramolecular autophosphorylation on a single Tyr residue in the activation loop, but the mature proteins recognize Ser and Thr residues in their target substrates. There are five DYRKs (1A, 1B, 2, 3, and 4), which have been implicated in the etiology of several major diseases. The *DYRK1A* gene is probably the best characterized because it is located at the Down syndrome critical locus, is overexpressed in trisomy 21, and has been implicated in the neurodegeneration and cancer susceptibility of Down syndrome patients. It has been implicated as a therapeutic target in many cancers, including EGF-dependent glioblastoma. DYRK1B is also overexpressed in several cancers. Importantly, there are currently no DYRK-directed drugs; although inhibitory compounds, such as INDY and the natural product harmine, have received attention, they display off-target effects that negate any therapeutic potential. Consequently, this provides an opportunity for exploiting our DYRK-directed inhibitors as cancer therapeutics.

Going forward, we are developing further two of the inhibitors identified in our SAR program (IC50 20–50 nM). Although it was a challenging target, Elad Elkayam determined the crystal structure of DYRK1A in a complex with one of these at 2.4 Å resolution. By combining sequence analysis, structural insights and biochemical assays, we have generated point mutants of DYRK1A in which catalytic function was maintained, but in which affinity for the inhibitors was markedly attenuated—from nanomolar to micromolar. We have used these mutants to establish that the effects of our small-molecule drug candidates are caused by "on-target" inhibition of DYRK, rather than "off-target" effects, in cancer cells. We have shown that our compounds were inhibitory to growth of glioblastoma cell lines in culture. Furthermore, they inhibited neurosphere formation by U87MG glioblastoma cells, showing improved effects compared with INDY, an established inhibitor of DYRK. The plan is to test

the existing inhibitors in animal models of glioblastoma, both to investigate further their mechanism of action as proof of concept and to test their potential as therapeutics. Preliminary data indicate synergistic effects of these DYRK inhibitors with inhibitors of the EGF receptor. This will be pursued as the basis for a potential combinatorial approach to EGFR-dependent glioblastoma. In addition, the group has established a collaboration with Dr. Yousef Al-Abed and his team at the Feinstein Institute to optimize the compounds further—both to improve bioavailability and to improve potency and selectivity. In addition, we will examine the possibility of exploiting unique structural features of the DYRKs to generate new inhibitor classes.

## PTP1B and Rett Syndrome

Rett syndrome (RTT) is an X-linked neurological disorder presenting with autistic features that is caused primarily by mutations in a transcriptional regulator, methyl CpG binding protein 2 (MECP2). It has been reported that some RTT mouse models display obesity and leptin resistance, with insulin resistance also noted in some RTT patients. We showed that glucose metabolism and insulin signaling in the brain were attenuated in *Mecp2*-mutant mice, which suggested to us that PTP1B function might be altered in RTT. We showed that the *PTPN1* gene, which encodes

PTP1B, is a target of MECP2 and that disruption of MECP2 function was associated with increased levels of PTP1B in RTT models. Pharmacological inhibition of PTP1B, with multiple structurally and mechanistically distinct small-molecule inhibitors, ameliorated the effects of MECP2 disruption in mouse models of RTT, including improved survival in young male (*Mecp2*$^{-/y}$) mice and improved behavior in female heterozygous (*Mecp2*$^{-/+}$) mice. Furthermore, we showed that the elevated levels of PTP1B in RTT represent a barrier to BDNF signaling. Inhibition of PTP1B led to increased tyrosine phosphorylation of TRKB in the brain, which augmented BDNF signaling. Taken together this work presents PTP1B as a mechanism-based therapeutic target for the treatment of Rett syndrome, validating a novel strategy for treating the disease by modifying signal transduction pathways with small-molecule drugs. Currently, we are conducting detailed mechanistic analyses of the effects of PTP1B inhibitors in animal models, with a view to taking such inhibitors into clinical trials in Rett syndrome patients.

*In Press*

Londhe AD, Bergeron A, Curley SM, Zhang F, Rivera KD, Kannan A, Coulis G, Rizvi SHM, Kim SJ, Pappin DJ, et al. 2020. Regulation of PTP1B activation through disruption of redox-complex formation. *Nat Chem Biol* **16:** 122–125. doi:10.1038/s41589-019-0433-0

# MECHANISMS AND MODELING OF METASTASIS

**L. Trotman**    A. Aziz          S. Escobar-Avelar   M. Swamynathan
                 I. Casanova-Salas   M. Lee            M. Walsh
                 H. Cox             G. Mathew          K. Watrud

Resistance to therapy of metastatic prostate cancer (CRPC) is responsible for the deaths of some 30,000 U.S. men each year. Although there is considerable progress in the development of improved antihormone therapy for treatment of metastatic disease, this standard-of-care approach will invariably fail at some point.

Our focus is to understand the mechanisms driving human prostate cancer in its most lethal form: metastatic disease. We have studied the human genetics behind the transition from indolent to lethal metastatic prostate cancer and combined it with viral transgene delivery into prostate. With this approach, we have succeeded in generating a unique, fast and faithful mouse model for metastatic prostate cancer. We have termed this system RapidCaP as it allows us to generate any genetically mutant mouse prostate cancer with a much-accelerated time frame compared to breeding-based approaches. Now, we use RapidCaP for analysis and therapy of metastatic disease, and we use human genomics analysis for discovery of new candidate drivers of metastasis.

At the same time, we aim to better understand how the PTEN tumor suppressor works. This has given us unique insights into how the process of endocytosis is intimately associated with tumor suppressor function of PTEN, allowing us to redefine this pathway.

## Endocytosis and Cancer

M. Lee, G. Mathew, H. Cox, M. Swamynathan

The transduction of signals in the PTEN/PI3-kinase (PI3K) pathway is built around a phosphoinositide (PIP) lipid messenger, phosphatidylinositol triphosphate, PI(3,4,5)P3 or $PIP_3$. Another, more ancient role of this family of messengers is the control of endocytosis, in which a handful of separate PIPs act like postal codes. Prominent among them is PI(3)P, which helps to ensure that endocytic vesicles, their cargo, and membranes themselves reach their correct destinations. Traditionally, the cancer and the endocytic

functions of the PI3K signaling pathway have been studied by cancer and membrane biologists, respectively, with some notable but overall minimal overlap. This is because cancer rarely mutates the endocytic pathway as the process is essential.

The discovery that PTEN contains an autonomous PI(3)P reader domain, fused to the catalytic $PIP_3$ eraser domain, has prompted us to explore the relationship between PI3K signaling and endocytosis. We have preliminarily shown that PTEN function can be enhanced by a compound that inhibits clathrin-mediated endocytosis, revealing for the first time that PTEN activity can be enhanced by small molecule-based approaches. As tumors frequently present with haploinsufficiency for PTEN and therefore reduced PTEN activity, this compound represents the starting point for further research into the therapeutic potential of targeting endocytosis in cancer. Our work now aims to elucidate how PTEN is recruited during endocytosis and what protein players are involved in this process.

## Genomics of Lethal Human Prostate Cancer

G. Mathew, H. Cox, I. Casanova-Salas, A. Jhaveri, K. Watrud, A. Ambrico [in collaboration with S. Hall, M. Vira, C. Metz, O. Yaskiv, O. Rodriguez, and M. Ryan, Northwell Health; L. Kollath and C. Morrissey, University of Washington, Seattle]

Next-generation sequencing techniques have provided the ability to incorporate cutting-edge genomic profiling in understanding, prognosis and treatment of various tumor types. However, heterogeneity of the prostate tissue during tumorigenesis makes it difficult to conduct exhaustive transcriptome analyses. Spontaneous genetic changes arising in PC are a crucial imprint of this variability. Therefore one of our aims in the lab is to shed some light on key genomic drivers of metastatic progression utilizing genomic information obtained from nuclear DNA (nDNA) of prostate cancer cells. The project goal was to obtain

an extensive copy number landscape of visceral and bone metastases from 10 patients who had consented to subject their bodies to rapid autopsy (RAP) after death from prostate cancer. Samples were obtained from University of Washington, Seattle and Northwell Health, NY. We have successfully established a pipeline for processing of frozen tissue samples for single-nucleus sequencing to determine copy number alterations (CNAs) at the single-cell level. Recurrent CNAs involving cancer genes have emerged as the primary driver of lethal metastatic PC, whereas recurrent missense mutations are infrequent. After processing all bone metastases, we analyzed matched visceral metastases (liver, lung, lymph node, etc.) from each of the 10 patients. Normal muscle sample from each patient was obtained and processed as a baseline control for CNAs.

Based on the data collected on 2914 cells from 31 metastatic sites of 10 patients, we first answer these general questions on the metastatic landscape seen at single-cell resolution:

1. What is the CNA-based clonality of metastasis (within the tumor site and between sites)?
2. Can we infer fitness of clones based on representation (within a site and/or between sites of a patient)?
3. Are there recurrent CNAs that have been missed by bulk sequencing of metastatic PC?

Given our expertise and the emergence of *PTEN* deletion as the most prominent feature of lethal metastatic PC, we place special emphasis on the below questions.

1. What genes are most significantly co-deleted with *PTEN* at the single-cell level?
2. Does loss of *PTEN* dominate clonality as expected from a strong driver event?

These data are complemented by our analysis of tumors from primary PC patients using the same approach. Samples from these early patients are collected through our collaboration with clinician scientists at Northwell Health. Analysis of genome-wide DNA and RNA alteration in primary and metastatic Rapid-CaP samples is used for cross-species prioritization of results. Based on our preliminary results this project allows us to discover novel markers of metastasis and new drivers of the lethal disease, which have escaped our notice based on bulk sequencing analysis.

The precise experimental modeling of gene dosage is a critical step in understanding how genes collaborate during cancer progression. This principle is an important yardstick in understanding the genetics of the disease and thereby administering drugs that are effective. Our review of the manuscript by Hermanova et al. discusses the genetic handle on how dosage of *LKB1* gene in the context of Pten loss is critical in prostate cancer metastasis.

## Blood-Based Monitoring of Prostate Cancer

I. Casanova-Salas, G. Mathew, H. Cox, A. Jhaveri, M. Swamynathan [in collaboration with S. Hall, M. Vira, C. Metz, O. Yaskiv, O. Rodriguez, and M. Ryan, Northwell Health; L. Kollath and C. Morrissey, University of Washington, Seattle]

We have developed a method to infer CNAs from exosomal DNA (exoDNA) in patient blood. Because CNAs are the major driver behind metastatic PC, we use this low-cost approach to discover new biology that can be translated into tests that will help patients. First, we find that the whole genome is represented in exoDNA, whereas overt CNA events are mostly detectable only in metastatic patients. These results are consistent with the notion that exoDNA from normal cells far outnumbers that from a tumor site until there is overt metastasis. Therefore, we explore other parameters in these samples to define biomarkers and new biology that allows us to get insights on the nature of the tumor in a patient. Based on our preliminary results, our approach can lead to tests that are directly applicable for validation studies in patients that are under active surveillance for signs of prostate cancer progression.

## The PHLPP2 Phosphatase Protects MYC and Is a Target for Prevention of Prostate Cancer Progression

D.G. Nowak (current address Weill Cornell Medicine, NY), K. Watrud, I. Casanova-Salas, C-H. Pan, A. Ambrico, M. Swamynathan [in collaboration with K.C. Katsenelson and A.C. Newton, University of California, San Diego; J.E. Wilkinson, University of Michigan, Ann Arbor]

Metastatic PC commonly presents with targeted, biallelic mutation of *PTEN* and *TP53* tumor-suppressor genes. In contrast however, most candidate tumor suppressors are part of large recurrent hemizygous deletions,

such as the common chromosome 16q deletion, which involves the AKT-suppressing phosphatase, PHLPP2.

Using RapidCaP, a genetically engineered mouse model of Pten-Trp53-mutant metastatic PC, we found that complete loss of Phlpp2 effectively blocks prostate tumor growth and progression to otherwise lethal metastasis. We show that Phlpp2 activates Myc, a key driver of prostate cancer and metastasis. Mechanistically, Phlpp2 dephosphorylates the Thr-58 site of Myc, thus directly increasing MYC stability. Finally, we show that small-molecule inhibitors of PHLPP2 can suppress MYC and cause cell death. Our findings reveal how PTEN-deficient tumors can thrive in the absence of AKT activation, driven by PHLPP2 stabilization of MYC. They also suggest that the frequent hemizygous deletions on chromosome 16q present a druggable vulnerability for targeting the MYC protein through PHLPP2 phosphatase inhibitors.

## PUBLICATIONS

Carbonetti G, Converso C, Clement T, Wang C, Trotman LC, Ojima I, Kaczocha M. 2020. Docetaxel/cabazitaxel and fatty acid binding protein 5 inhibitors produce synergistic inhibition of prostate cancer growth. *Prostate* **80:** 88–98. doi:10.1002/pros.23921

Froeling FEM, Swamynathan MM, Deschênes A, Chio IIC, Brosnan E, Yao MA, Alagesan P, Lucito M, Li J, Chang AY, et al. 2019. Bioactivation of napabucasin triggers reactive oxygen species-mediated cancer cell death. *Clin Cancer Res* **25:** 7162–7174. doi:10.1158/1078-0432.CCR-19-0302

Nowak DG, Katsenelson KC, Watrud KE, Chen M, Mathew G, D'Andrea VD, Lee MF, Swamynathan MM, Casanova-Salas I, Jibilian MC, et al. 2019. The PHLPP2 phosphatase is a druggable driver of prostate cancer progression. *J Cell Biol* **218:** 1943–1957. doi:10.1083/jcb.201902048

*In Press*

Lee MF, Trotman LC. 2019. PTEN: bridging endocytosis and signaling. *Cold Spring Harb Perspect Med* **10:** a036103. doi:10.1101/cshperspect.a036103

# CANCER MEDICINE

| D. Tuveson | A. Adler | A. Deschênes | R. Koike | H. Patel | H. Tiriac |
| --- | --- | --- | --- | --- | --- |
| | P. Alagesan | A. Doshi | A. Leonhart | D. Plenker | M. Tokov |
| | L. Baker | E. Elyada | M. Lucito | M. Rubin | C. Tonelli |
| | G. Biffi | D. Engle | K. Miyabayashi | F. Scheufele | B. Traub |
| | E. Brosnan | F. Froeling | S. Moran | B. Spielman | M. Yao |
| | G. Caligiura | Z. Galluzzo | T. Oni | J. Thalappillil | G. Yordanov |
| | D. Cheng | M. Kheir Gouda | Y. Park | H-C. Ting | K. Yu |

Our laboratory investigates pancreatic ductal adenocarcinoma (PDAC), the primary form of pancreatic cancer and the third leading cause of cancer-related deaths in the United States. More specifically, we aim to generate insights into the molecular underpinnings of PDAC that could inform novel strategies to detect and treat this currently incurable cancer. Recently, we developed three-dimensional organoid cultures as ex vivo models of PDAC biology. Organoids have enabled new insights into the factors driving PDAC development that have the potential to improve patient care. Through analysis of patient-derived organoid cultures, we have identified signatures of genes whose expression predicts patient response to chemotherapy agents, and we are now developing clinical trials to refine and evaluate those signatures. By studying organoid and mouse models of PDAC, we have also begun to disentangle the cell types and signals responsible for patterning the PDAC microenvironment. Finally, we are using organoids and mouse models to identify better biomarkers to aid in the earlier diagnosis of pancreatic cancer. Through these studies, we have identified a surprising role for the biomarker CA19-9 in promoting inflammatory changes that hasten pancreatic cancer development.

## Characterizing a Novel Fibroblast Present in the Pancreatic Cancer Microenvironment

This work was done in collaboration with P. Robson (Jackson Laboratory) and A. Califano (Columbia Medical School).

PDAC is distinct from other cancers because of its high content of nonneoplastic tissue. This tissue is composed of a large number of noncancerous cell types recruited by the cancer cells. Among these cell types, cancer-associated fibroblasts (CAFs) were long believed to be recruited to support the cancer cells. However, recent studies have suggested conflicting roles for this population, hinting at the existence of more than one kind of CAF. To better characterize the types of CAFs present in the PDAC microenvironment, we initiated a collaboration with Dr. Paul Robson at the Jackson Laboratory Cancer Center and Dr. Jonathan Preall at CSHL to perform single-cell analysis of mouse and human PDAC tumors. Analysis of these data has revealed a novel CAF population that expresses major histocompatibility complex (MHC) class II molecules, proteins usually only present on professional antigen-presenting cells (APCs) of the immune system. Professional APCs use these molecules to present antigens to other cells, including T cells, so we initially thought the antigen-presenting CAFs (apCAFs) might serve a similar function. Consistent with that idea, using an in vitro T-cell activation assay, the team demonstrated that apCAFs have the capacity to present antigen to CD4+ T cells. Professional APCs express costimulatory molecules on their cell surface, which provide the second signal necessary to induce clonal proliferation and further activation of CD4+ T cells. However, apCAFs expressed low levels of the costimulatory genes compared to professional APCs. Together, these results, which were published in *Cancer Discovery* in 2019 (Elyada et al. 2019), demonstrate the presence of at least three CAF subtypes in PDAC and suggest that unlike professional APCs, apCAFs are only capable of partial T-cell activation. The team is continuing to study the role of apCAFs in PDAC tumors and has identified another difference between apCAFs and professional APCs: Unlike professional APCs, apCAFs are not able to induce IL2 secretion from T cells, again confirming that apCAFs are only capable

of partial T-cell activation. We hypothesize that partial activation of T cells mediated by apCAFs could alter the T-cell response and thereby attenuate the function of CD8[+] T cells that are capable of recognizing peptides bound to MHC class I on PDAC cells, resulting in immune suppression. If this hypothesis holds true, targeting apCAFs might help make pancreatic tumors more sensitive to immunotherapy. Future studies will be needed to test this hypothesis and to better understand the role of apCAFs in PDAC development and progression.

## Characterizing a Novel Role for the Glycan CA19-9 in Promoting the Development of Pancreatitis and Pancreatic Cancer

This work was done in collaboration with T. Hollingsworth (Nebraska Epley Cancer Center), D. Pappin (CSHL), and R. Hruban and M. Goggins (Johns Hopkins Cancer Center).

Our laboratory has also been studying the functions of the PDAC biomarker CA19-9. For many patients with pancreatic cancer, increases in CA19-9 levels are observed following tumor development. Mice are normally incapable of producing this glycan, making its role difficult to study in mouse models of PDAC. To address this, the team engineered a mouse model in which CA19-9 production can be selectively induced. Expression of CA19-9 in mice resulted in pancreatitis with accompanying increases in serum levels of the pancreatitis markers amylase and lipase. The team confirmed that CA19-9 elevation is also commonly present in the pancreatic tissues obtained postsurgically from patients suffering from chronic pancreatitis. These results directly implicate CA19-9 as a causative factor in the development of this inflammatory condition. Additionally, in a preventive setting, two antibodies directed against CA19-9 both reduced immune infiltration, ductal metaplasia, and fibrosis following induction of CA19-9 in the mouse model. Moreover, in an intervention setting of existing acute pancreatitis, treatment of these mice with a CA19-9 antibody reduced pancreatitis histology, amylase production, levels of activated EGFR in the pancreas, and immune infiltration, suggesting that CA19-9-targeted therapy may benefit patients with pancreatitis. Finally, when inducible CA19-9 mice were crossed with the oncogenic *Kras*[LSL-G12D] allele,

CA19-9 induction was found to accelerate pancreatic cancer development, resulting in decreased survival. Together, these results suggest that rather than being simply a biomarker of pancreatic cancer, CA19-9 has a direct role in promoting this aggressive disease. Prophylactic intervention to block CA19-9 in the setting of recurrent pancreatitis in patients might help to reduce the severity of pancreatitis and to prevent the development of PDA. This study was published in *Science* in 2019 (Engle et al. 2019).

## PUBLICATIONS

Baker LA, Tuveson DA. 2019. Generation and culture of tumor and metastatic organoids from murine models of pancreatic ductal adenocarcinoma. *Methods Mol Biol* **1882:** 117–133.

Baker LA, Tiriac H, Tuveson DA. 2019. Generation and culture of human pancreatic ductal adenocarcinoma organoids from resected tumor specimens. *Methods Mol Biol* **1882:** 97–115.

Biffi G, Tuveson DA. 2019. A FATal combination: fibroblast-derived lipids and cancer-derived autotaxin promote pancreatic cancer growth. *Cancer Discov* **9:** 578–580.

Biffi G, Oni TE, Spielman B, Hao Y, Elyada E, Park Y, Preall J, Tuveson DA. 2019. IL1-induced JAK/STAT signaling is antagonized by TGFβ to shape CAF heterogeneity in pancreatic ductal adenocarcinoma. *Cancer Discov* **9:** 282–301.

Bott AJ, Shen J, Tonelli C, Zhan L, Sivaram N, Jiang Y-P, Yu X, Bhatt V, Chiles E, Zhong H, et al. 2019. Glutamine anabolism plays a critical role in pancreatic cancer by coupling carbon and nitrogen metabolism. *Cell Rep* **29:** 1287–1298.e6.

Dent P, Booth L, Roberts JL, Liu J, Poklepovic A, Lalani AS, Tuveson D, Martinez J, Hancock JF. 2019. Neratinib inhibits Hippo/YAP signaling, reduces mutant K-RAS expression, and kills pancreatic and blood cancer cells. *Oncogene* **38:** 5890–5904.

Driehuis E, Kolders SD, Spelier S, Löhmussaar K, Willems SM, Devriese LA, de Bree R, de Ruiter EJ, Korving J, Begthel H, et al. 2019. Oral mucosal organoids as a potential platform for personalized cancer therapy. *Cancer Discov* **9:** 852–871.

Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhart RA, Teinor JA, Belleau P, Biffi G, Lucito MS, et al. 2019. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. *Cancer Discov* **9:** 1102–1123.

Engle DD, Tiriac H, Rivera KD, Pommier A, Whalen S, Oni TE, Alagesan B, Lee EJ, Yao MA, Lucito MS, et al. 2019. The glycan CA19-9 promotes pancreatitis and pancreatic cancer in mice. *Science* **364:** 1156–1162.

Froeling FEM, Swamynathan MM, Deschênes A, Chio IIC, Brosnan E, Yao MA, Alagesan P, Lucito M, Li J, Chang A-Y, et al. 2019. Bioactivation of napabucasin triggers reactive oxygen species–mediated cancer cell death. *Clin Cancer Res* **25:** 7162–7174.

Humpton TJ, Alagesan B, DeNicola GM, Lu D, Yordanov GN, Leonhardt CS, Yao MA, Alagesan P, Zaatari MN, Park Y, et al. 2019. Oncogenic KRAS induces NIX-mediated mitophagy to promote pancreatic cancer. *Cancer Discov* **9:** 1268–1287.

Mirabelli CK, Nusse R, Tuveson DA, Williams BO. 2019. Perspectives on the role of Wnt biology in cancer. *Sci Signal* **9:** 12.

Neesse A, Bauer CA, Öhlund D, Lauth M, Buchholz M, Michl P, Tuveson DA, Gress TM. 2019. Stromal biology and therapy in pancreatic cancer: ready for clinical translation? *Gut* **68:** 159–171.

Ponz-Sarvise M, Corbo V, Tiriac H, Engle DD, Frese KK, Oni TE, Hwang C-I, Öhlund D, Chio IIC, Baker LA, et al. 2019. Identification of resistance pathways specific to malignancy using organoid models of pancreatic cancer. *Clin Cancer Res* **25:** 6742–6755.

Sugarman R, Patel R, Sharma S, Plenker D, Tuveson D, Wasif Saif M. 2019. Pharmacokinetics and pharmacodynamics of new drugs for pancreatic cancer. *Expert Opin Drug Metab Toxicol* **15:** 541–552.

Tian C, Clauser KR, Öhlund D, Rickelt S, Huan Y, Gupta M, Mani DR, Carr SA, Tuveson DA, Hynes RO. 2019. Proteomic analyses of ECM during pancreatic ductal adenocarcinoma progression reveal different contributions by tumor and stromal cells. *Proc Natl Acad Sci* **116:** 19609–19618.

Tiriac H, Plenker D, Baker LA, Tuveson DA. 2019. Organoid models for translational pancreatic cancer research. *Curr Opin Genet Dev* **54:** 7–11.

Tuveson D, Clevers H. 2019. Cancer modeling meets human organoid technology. *Science* **364:** 952–955.

# Ras AND Rho GTPase AND THE CONTROL OF SIGNAL TRANSDUCTION

**L. Van Aelst**    F. Boato        N.B. Gallo        S.C. Tsaniras
                    J. Freedman     E. Michalopoulou  M. Wang

Research in my laboratory is focused on the role of Ras and Rho GTPase family members in signal transduction. Ras and Rho family members play key roles in cellular activities controlling cell growth, differentiation, and morphogenesis. Alterations in Ras and Rho functions have been implicated in cancer as well as brain/mental disorders; the latter include intellectual disability, autism, schizophrenia, epilepsy, and mood disorders. Our interests lie in understanding how defects in Ras- and Rho-linked proteins contribute to the development of these diseases/disorders. Toward this end, my laboratory has continued to define the functions of selected GTPases, their regulators, and effectors, using animal models of cancer and neurodevelopmental/neurological disorders. Below we highlight our key projects.

## *Oligophrenin-1* Moderates Behavioral Responses to Stress via Regulation of Parvalbumin Interneuron Activity in the Medial Prefrontal Cortex

*Oligophrenin-1* (*OPHN1*), which encodes a Rho-GTPase activating protein, was the first identified Rho-linked intellectual disability (ID) gene. It was initially identified by the analysis of a balanced translocation t(X;12) observed in a female patient with mild ID. Subsequent studies revealed the presence of *OPHN1* mutations in families with a syndromic form of ID, with affected individuals exhibiting mild to moderate/severe ID, vermis and/or hemispheric cerebellar hypoplasia, as well as behavioral problems. The latter include hyperactivity, emotional imbalance, and intolerance to frustration, which can trigger helpless/depressive reactions and are often precipitated or exacerbated by stressful events. To date, the function of OPHN1 has been mostly studied in the hippocampus, with multiple studies unveiling key roles for OPHN1 in the regulation of hippocampal synaptic structure/function and plasticity as well as learning and memory. Despite the fact that stress-related behavioral problems/symptoms are observed in *OPHN1* patients, its role in the modulation of maladaptive behavioral responses and resilience to stress remains unexplored.

To address this, we genetically ablated *Ophn1* either globally or locally in distinct brain regions/neuronal cell types in mice and assessed the effects on performance in the learned helplessness procedure to determine the involvement of OPHN1 in the establishment of adaptive versus maladaptive behavioral responses to inescapable/uncontrollable stress. We found that mice lacking *Ophn1* globally or selectively in the prelimbic (PL) region of the medial prefrontal cortex (mPFC) display a marked increase in learned helplessness, a "depression-like" phenotype whereby animals show reduced escape from escapable foot shocks. Strikingly, deletion of *Ophn1* exclusively in PL-mPFC parvalbumin (PV) interneurons, but not somatostatin (SOM) interneurons or Emx1-expressing pyramidal neurons (PyNs), was sufficient to induce helplessness, underscoring the importance of intact PL-mPFC PV interneuron function in mediating adaptive behavioral responses (i.e., resilience) to stress. At a cellular level, we found that excitatory synaptic transmission onto PL-mPFC PV interneurons lacking OPHN1 is considerably weakened, leading to decreased spike output of inhibitory PV interneurons and consequently increased activity of neighboring mPFC PyNs. Importantly, suppression of mPFC neuronal activity using inhibitory DREADDs (designer receptors exclusively activated by designer drugs) reversed the stress-induced helpless behavior phenotype of *Ophn1*-deficient mice. Finally, we uncovered that OPHN1's effect on neuronal activity/stress-related behavior is critically dependent on its inhibition of the RhoA/Rho-kinase signaling pathway. In particular, we found that suppression of this pathway using the Rho-kinase inhibitor fasudil

normalizes the elevated neuronal activity and alleviates the helpless behavior in *Ophn1*-deficient mice. It is noteworthy that fasudil is currently in clinical trials in patients with primary pulmonary hypertension. Taken together, our results uncover a previously unknown role of OPHN1 in the regulation of mPFC PV interneuron activity required for shaping adaptive behavioral responses in the face of stress and suggest a cellular mechanism by which mutations in *OPHN1* may contribute to the behavioral deficits/problems in *OPHN1* patients.

## Axo-Axonic Innervation of Neocortical Pyramidal Neurons by GABAergic Chandelier Cells Requires AnkyrinG-Associated L1CAM

Proper assembly and functioning of cortical circuits rely on the formation of specific synaptic connections between excitatory pyramidal neurons (PyNs) and different types of GABAergic interneurons. Among the various cortical interneuron subtypes, chandelier cells (ChCs), in particular, have a powerful influence over the output of excitatory PyNs because of their unique morphology and type of connections they make. Specifically, ChCs possess a very distinctive axonal arbor with multiple arrays of short vertical sets of cartridges, each harboring a string of synaptic boutons. This unique architecture enables a single ChC to couple to a large population of PyNs. Furthermore, ChC cartridges make exclusive contact with the axon initial segment (AIS) of PyNs, which is the most excitable part of a neuron where action potentials are initiated. Importantly, aberrant ChC/PyN AIS innervation has been reported in several disease states associated with altered cortical excitability, including schizophrenia, epilepsy, and autism spectrum disorder. Despite the importance of ChCs, very little is known about the mechanisms governing ChC structure and connectivity. To date, the only molecules implicated in neocortical ChC morphogenesis are the atypical Rac activator DOCK7 and the receptor tyrosine kinase ErbB4. In particular, we uncovered that silencing of DOCK7 in ChCs via a novel vMGE-directed in utero electroporation (IUE) approach markedly decreases ChC cartridge bouton size and density—and intriguingly does so by modulating the activity of ErbB4. Although noteworthy, we found that DOCK7- and ErbB4-depleted ChCs still make contact with PyN

AISs, indicating that other molecules must regulate ChC/PyN axo-axonic innervation.

To identify the molecular factors required for neocortical ChC/PyN AIS innervation, we initiated an in vivo RNA interference (RNAi) screen of PyN-expressed axonal cell adhesion molecules (CAMs) and select Ephs/ephrins. More specifically, we devised a strategy taking advantage of IUE and RNAi to individually knock down these molecules in neocortical PyNs while concurrently labeling ChCs using the recently generated *Nkx2.1-CreER* mouse line, which enables tamoxifen-dependent ChC red fluorescent protein labeling via Nkx2.1-driven CreER expression. Strikingly, of all the candidates tested, we found the panaxonally expressed CAM L1CAM to be the only molecule required for neocortical ChC/PyN AIS innervation, as knockdown of PyN L1CAM, but none of the other screened candidates, significantly reduced PyN AIS innervation by ChCs. In line with this, we observed the number of VGAT and gephyrin puncta at the AIS, but not along the somatodendritic compartment, to be concomitantly reduced in L1CAM-depleted PyNs, indicating that PyN L1CAM selectively regulates ChC/PyN AIS synaptic innervation and not the subcellular targeting of other PyN subcellular domains by other interneuron subtypes. Furthermore, we showed that L1CAM is required during both the establishment and maintenance of neocortical ChC/PyN AIS innervation. Finally, we provided evidence that anchoring of L1CAM at the AIS by the ankyrin-G/βIV-spectrin AIS cytoskeletal complex is essential for ChC subcellular innervation of PyN AISs. Taken together, our findings identify L1CAM as the only PyN-expressed CAM known to date to regulate axo-axonic innervation of PyNs by ChCs in the neocortex.

## Target Discovery for Lung Cancer Multiple-Organ Metastasis

To gain insight into the mechanisms that mediate multiple-organ metastases for lung adenocarcinoma (ADC), we implemented orthotopic xenograft transplantation techniques to model lung cancer multiple-organ metastasis in mice. Using this model system, we uncovered a critical role for the atypical Rho activator DOCK4 in mediating TGF-β–driven lung ADC metastasis. Of note, DOCK4 is rapidly and robustly

induced by TGF-β in a Smad-dependent manner and high DOCK4 expression correlates with activated TGF-β signaling and poor prognosis in human lung ADC. Specifically, we found that blockade of TGF-β–mediated DOCK4 induction attenuates the ability of lung ADC cells to extravasate into distant organ sites, resulting in a marked reduction in metastatic burden. At a cellular level, our evidence supports a model in which TGF-β–induced DOCK4 facilitates extravasation by stimulating lung ADC cell protrusive activity, motility, and invasion, without promoting epithelial-to-mesenchymal transition (EMT), and, intriguingly, that it does so by driving Rac1 activation. So far, Rac1 has only been linked to TGF-β via a noncanonical pathway. Thus, our findings identified the atypical Rac1 activator DOCK4 as a novel key component of the TGF-β/Smad pathway that promotes lung ADC cell extravasation and metastasis.

Recently, we have expanded this line of research toward identifying genes that mediate lung ADC organ-specific metastases, with a particular focus on genes that mediate colonization of specific organs. In collaboration with Kenneth Chang, we initiated an in vivo RNAi screen to determine how 15 high-priority genes affect organ-specific metastasis. Interestingly, the in vivo RNAi screen identified genes that modulate cell homing to and/or colonization of lung ADC metastases in the brain and bone. We then examined specific roles for these genes and found that knockdown of *EphA4* enhanced the formation of brain metastases, whereas *Arhgap26* knockdown promoted metastasis formation in multiple organs, including bone, adrenal gland, and liver, with the highest penetrance

occurring in bone. Thus, our data identify *EphA4* and *Arhgap26* as potential candidate suppressor genes of brain and multiple-organ metastases from lung ADC. More recently, we also expanded our studies of the mechanisms that drive metastasis to include breast cancer. In collaboration with the Egeblad laboratory, we developed an experimental immune-competent mouse model of metastatic breast cancer. We engineered the 4T1 breast cancer cells to express Akaluc (a modified version of luciferase that produces much brighter emissions in vivo than conventional luciferase) and the TRAP transgene (EGFP-L10a). The latter permits isolation of transcripts from rare cell populations embedded in intact tissue. This animal model enables us to not only monitor multiple-organ metastases, but to also identify candidate genes that are involved in organ-specific colonization and drug resistance of metastatic breast cancer in the context of an intact immune system.

## PUBLICATIONS

He XY, Ng D, Van Aelst L, Egeblad M. 2019. Stressing out about cancer immunotherapy. *Cancer Cell* **36:** 468–470. doi:10.1016/j.ccell.2019.10.013

Tai Y, Gallo NB, Yu JR, Van Aelst L. 2019. Axo-axonic innervation of neocortical pyramidal neurons by GABAergic chandelier cells requires AnkyrinG-associated L1CAM. *Neuron* **102:** 358–372. doi:10.1016/j.neuron.2019.02.009

*In Press*

Gallo NB, Paul A, Van Aelst L. 2020. Shedding light on chandelier cell development, connectivity, and contribution to neural disorders. *Trends Neurosci* **43:** 565–580. doi:10.1016/j.tins.2020.05.003

# NEUROSCIENCE

How does the brain encode stimuli from the outside world, within and across sensory modalities, to generate specific perceptions that trigger complex behaviors? How is the brain shaped by sensory experience, and what modifications occur in neuronal circuits that allow us to learn and remember? These are the questions guiding the work of **Florin Albeanu**, who is using the olfactory bulb and olfactory cortex of mice as the subject of his current studies. Airborne chemicals are translated into neuronal signals by specific receptors in the nose, and the signals are sent directly to the olfactory bulb. Advances in optical imaging and optogenetics, combined with electrophysiological recordings, enable Albeanu and colleagues to monitor and/or alter patterns of activity at unprecedented synaptic and millisecond resolution, in real time, as animals are engaged in various behaviors. For survival, rodents need to identify the smells of objects of interest such as food, mates, and predators, across their recurring appearances in the surroundings, despite apparent variations in their features. Furthermore, animals aptly extract relevant information about their environment across different sensory modalities, combining olfactory, visual, or auditory cues. By recording neuronal activity in the input and output layers of the olfactory bulb, as well as feedback from olfactory cortical areas and neuromodulatory signals, Albeanu and his team aim to understand computations the bulb performs and how this information is decoded deeper in the brain. They have recently published evidence suggesting that the mouse olfactory bulb is not merely a relay station between the nose and cortex, as many have supposed. Using optogenetic tools and a novel patterned illumination technique, they discovered that there are many more information output channels leaving the olfactory bulb for the cortex than there are inputs received from the nose. They are currently investigating how this diversity of bulb outputs is generated, as well as how downstream areas, such as the piriform and parietal cortexes, make use of such information during behaviors.

The study of decision-making provides a window into the family of brain functions that constitute cognition. It intervenes between perception and action and can link one to the other. Although much is known about sensory processing and motor control, much less is known about the circuitry connecting them. Some of the most interesting circuits are those that make it possible to deliberate among different interpretations of sensory information before making a choice about what to do. **Anne Churchland**'s laboratory investigates the neural machinery underlying decision-making. Laboratory members use carefully designed paradigms that encourage experimental subjects to deliberate over incoming sensory evidence before making a decision. Recent results show that rats and humans have a statistically similar decision-making ability. To connect this behavior to its underlying neural circuitry, the researchers measure electrophysiological responses of cortical neurons in rodents as they perform designated tasks. The laboratory's current focus is on the parietal cortex, which appears to be at the midpoint between sensory processing and motor planning. Churchland and colleagues also use theoretical models of varying complexity to further constrain how observed neural responses might drive behavior. This approach generates insights into sensory processing, motor planning, and complex cognitive function.

The brain's activity is in constant motion. It ebbs and flows in big waves when we are in a deep slumber, turns into small ripples when we reawaken, and flows in orchestrated streams when we perceive, decide, and remember. These complex dynamics are driven by intricate networks of microscopic interactions between millions of neurons and thus are only vaguely observed in spike trains of single neurons. Fortunately, recent advances in recording techniques enable us to monitor the activity of large neural populations in behaving animals, offering the opportunity to investigate how dynamic variations of collective neural activity states translate into behavior. To gain

insights from these large-scale recordings, **Tatiana Engel** and colleagues develop and apply computational methods for discovering collective neural dynamics from sparse, high-dimensional spike-train data. They also develop models and theory to explain how collective neural dynamics support specific network computations and how these dynamics are constrained by biophysical properties of neural circuits. In these endeavors, they use and extend tools and ideas from diverse fields, such as statistical mechanics, machine learning, dynamical systems theory, and information theory. Their work benefits from close collaborations with experimental neuroscience laboratories that are collecting neurophysiological data in animals engaged in sophisticated tasks, such as attention, decision-making and learning.

**Hiro Furukawa**'s laboratory studies receptor molecules involved in neurotransmission. Its members mainly focus on the structure and function of NMDA (*N*-methyl-D-aspartate) receptors—ion channels that mediate excitatory transmission. Dysfunctional NMDA receptors cause neurological disorders and diseases, including Alzheimer's disease, Parkinson's disease, schizophrenia, depression, and stroke-related ischemic injuries. The Furukawa laboratory is working to solve the three-dimensional structure of the very large NMDA receptor by dividing it into several domains. They seek to understand the pharmacological specificity of neurotransmitter ligands and allosteric modulators in different subtypes of NMDA receptors at the molecular level. Toward this end, they use cutting-edge techniques in X-ray crystallography to obtain crystal structures of the NMDA receptor domains and validate structure-based functional hypotheses by a combination of biophysical techniques, including electrophysiology, fluorescence analysis, isothermal titration calorimetry, and analytical centrifugation. Crystal structures of NMDA receptors serve as a blueprint for creating and improving the design of therapeutic compounds with minimal side effects for treating neurological disorders and diseases. During the last several years, the team discovered and mapped several regulatory sites in specific classes of NMDA receptors—progress that now opens the way to the development of a new potential class of drugs to modulate receptor activity.

**Josh Huang** and colleagues study the assembly and function of neural circuits in the neocortex of the mouse. The neocortex consists of a constellation of functional areas that form a representational map of the external (sensory, social) and internal (visceral, emotional) world. These areas are strategically interconnected into elaborate information processing networks that guide behavior. The group's overarching hypothesis is that, at the cellular level, cortical processing streams and output channels are mediated by a large set of distinct glutamatergic pyramidal neuron types, and functional neural ensembles are regulated by a diverse set of GABAergic interneuron types. Understanding cortical circuit organization requires comprehensive knowledge of these basic cellular components. The Huang laboratory uses state-of-the-art genetic approaches to systematically target cell types and facilitate the application of a full set of modern techniques for exploring cortical circuits. Among GABAergic interneurons, the chandelier cell is one of the most distinctive cell types that controls pyramidal neuron firing at the axon initial segment. Huang and colleagues are studying the developmental specification, activity-dependent circuit integration, and functional connectivity of chandelier cells—an entry point toward understanding a local circuit module. Regarding pyramidal neurons, they are systematically characterizing the developmental origin, axon projection pattern, and input connectivity of multiple classes of pyramidal neuron types, focusing on the forelimb motor cortex. They combine a range of approaches that include genetic and viral engineering, genetic fate mapping, gene expression profiling, cellular imaging, electrophysiology, and behavior analysis. Recently, they began to integrate their studies in the context of the motor cortex control of forelimb movements.

**Adam Kepecs** and colleagues are interested in identifying the neurobiological principles underlying cognition and decision-making. They use a reductionist approach, distilling behavioral questions to quantitative behavioral tasks for rats and mice that enable the monitoring and manipulation of

neural circuits supporting behavior. Using state-of-the-art electrophysiological techniques, they first seek to establish the neural correlates of behavior and then use molecular and optogenetic manipulations to systematically dissect the underlying neural circuits. Given the complexity of animal behavior and the dynamics of neural networks that produce it, their studies require quantitative analysis and make regular use of computational models. The team also has begun to incorporate human psychophysics to validate its behavioral observations in rodents by linking them with analogous behaviors in human subjects. Currently, the team's research encompasses study of (1) the neural basis of decision confidence, (2) the division of labor among cell types in the prefrontal cortex, (3) how the cholinergic system supports learning and attention, and (4) social decisions that rely on stereotyped circuits. A unifying theme is the use of precisely timed cell-type and pathway-specific perturbations to effect gain and loss of function for specific behavioral abilities. This year, the Kepecs laboratory was able to link foraging decisions—the choice between staying or going—to a neural circuit and specific cell types in the prefrontal cortex. In other work, they identified a class of inhibitory neurons that specializes in inhibiting other inhibitory neurons in the cerebral cortex and conveys information about rewards and punishment. Through manipulations of genetically and anatomically defined neuronal elements, the team hopes to identify fundamental principles of neural circuit function that will be useful for developing therapies for diseases such as schizophrenia, Alzheimer's disease, and autism spectrum disorder.

**Alexei Koulakov** and colleagues are trying to determine the mathematical rules by which the brain assembles itself, with particular focus on the formation of sensory circuits such as those involved in visual perception and olfaction. The visual system of the mouse was chosen for study in part because its components, in neuroanatomical terms, are well understood. What is not known is how projections are generated that lead from the eye through the thalamus and into the visual cortex, how an individual's experience influences the configuration of the network, and what parameters for the process are set by genetic factors. Even less is known about the assembly of the neural net within the mouse olfactory system, which, in the end, enables the individual to distinguish one smell from another with astonishing specificity and to remember such distinctions over time. These are among the challenges that engage Koulakov and his team.

Understanding the link between neural circuits and behavior has been the focus of research in **Bo Li**'s laboratory. Li and colleagues are particularly interested in studying the synaptic and circuit mechanisms underlying reward processing, attention, and learning and memory, as well as synaptic and circuit dysfunctions responsible for maladaptive behaviors that are related to major mental disorders. They integrate in vitro and in vivo electrophysiology, imaging, molecular, genetic, optogenetic, and chemogenetic techniques to probe and manipulate the function of specific neural circuits—with a focus on the fear and reward circuits—in the rodent brain and determine how these circuits participate in adaptive or maladaptive behavioral responses in various tasks.

**Partha Mitra** is interested in understanding intelligent machines, which are products of biological evolution (particularly animal brains), with the basic hypothesis that common underlying principles may govern these "wet" intelligent machines and the "dry" intelligent machines that are transforming the present economy. Dr. Mitra initiated the idea of brain-wide mesoscale circuit mapping, and his laboratory is involved in performing such mapping in the mouse (http://mouse .brainarchitecture.org) and the marmoset (in collaboration with Japanese and Australian scientists at the RIKEN Brain Science Institute and Monash University).

Dr. Mitra spent 10 years as a member of the theory department at Bell Laboratories and holds a visiting professorship at IIT Madras, where he is helping establish the Center for Computational Brain Research. He has an active theoretical research program in machine learning and control theory, wherein he is using tools from statistical physics to analyze the performance of distributed/ networked algorithms in the "thermodynamic" limit of many variables.

**Pavel Osten**'s laboratory works on identification and analysis of brain regions, neural circuits, and connectivity pathways that are disrupted in genetic mouse models of autism and schizophrenia. Osten hypothesizes that (1) systematic comparison of multiple genetic mouse models will allow determination of overlaps in pathology—neural circuit endophenotypes—responsible for the manifestation of neuropsychiatric disorders, and (2) neural circuit–based classification of autism and schizophrenia will provide key circuit targets for detailed mechanistic studies and therapeutic development. Osten and colleagues have developed the first systematic approach to the study of neural circuits in mouse models of psychiatric diseases, based on a pipeline of anatomical and functional methods for analysis of mouse brain circuits. An important part of this pipeline is high-throughput microscopy for whole-mouse brain imaging, called serial two-photon (STP) tomography. This year, they used this method to describe the first whole-brain activation map representing social behavior in normal mice. They are currently focusing on using this approach to study brain activation changes in two mouse models of autism: the 16p11.2 df/+ mouse model, which shows an increased propensity to seizures and hyperactivity, and the CNTNAP2 knockout mouse model, which shows abnormal social behavior.

**Stephen Shea**'s laboratory studies the neural circuitry underlying social communication and decisions. They use natural social communication behavior in mice as a model to understand circuits and processes that are evolutionarily conserved and therefore shared broadly across species, likely contributing to disorders such as autism. Shea and colleagues have examined how emotion and arousal enable mice, via their olfactory systems, to store memories of other individuals and related social signals. The team has exploited the intimate relationship between memory and emotion to effectively create memories in anesthetized mice, allowing them unprecedented access to neurobiological processes that typically only occur during behavior. The laboratory has been making a detailed analysis of the changes in neural connections that underlie odor memory. The team is particularly focused on an enigmatic cell type (granule cells or GCs) that has long been hypothesized to be crucial for memories, but has resisted direct study. They have developed methods for recording, giving them the first glimpse of the dynamics of these cells while the animal is learning an odor. The results show unexpectedly complex population dynamics among the GCs, which were independently predicted by a model of odor learning developed in Alexei Koulakov's laboratory. The two laboratories are collaborating to discern how GC population activity gets integrated by olfactory bulb output neurons and to pinpoint the synaptic circuit that underlies this form of learning. In parallel, another member of the laboratory is using imaging techniques to determine how memories are stored among broad neuronal ensembles, at a different level of the system. Recently, the laboratory made a key breakthrough, developing the ability to record from GCs in awake animals and discovering that their activity is dramatically modulated by state of consciousness. Finally, the Shea laboratory completed a series of studies of a different form of social recognition: auditory recognition of pup vocalizations by their mothers. Through this research, they have shown that a mouse model of Rett syndrome shows deficits in communication and learning not unlike those in human patients. Grants from the Simons and Whitehall foundations are allowing the laboratory to extend this work by directly linking these deficits to the action of the gene *MeCP2* in the auditory cortex.

**Jessica Tollkuhn**'s laboratory seeks to understand how transient events during brain development exert lasting effects on gene expression, circuit function, and, ultimately, behavior. They study how sex-specific neural circuits in rodents are established and modulated by the gonadal hormones estrogen and testosterone. The cognate receptors for these hormones are nuclear receptor transcription factors, which orchestrate modification of local chromatin environment and thus exert long-term effects on gene expression. However, the genes regulated by these receptors, as well as the specific mechanisms they use, remain poorly understood in the brain. This is in part because the extraordinary cellular heterogeneity of the brain complicates analysis of the small subpopulations of neurons that mediate sex-specific behaviors.

Having recently identified sex differences in both gene expression and chromatin in brain regions known to regulate sex-specific behaviors, the Tollkuhn laboratory is now working to understand how hormones generate these molecular sex differences during development, through the use of biochemical, genomic, and behavioral analyses. They have developed a method that permits genome-wide analysis of histone modifications or DNA methylation in genetically defined populations of neurons. They hypothesize that these epigenetic data, combined with gene expression profiling, define the molecular signature of the critical period for sexual differentiation of the brain. Their goal is to provide a mechanistic link between the transcriptional effects of hormone signaling during development and the subsequent social behaviors displayed in adulthood.

**Anthony Zador** and colleagues study how brain circuitry gives rise to complex behavior. Work in the laboratory is focused on two main areas. First, they ask how the cortex processes sound, how that processing is modulated by attention, and how it is disrupted in neuropsychiatric disorders such as autism. Recently, the laboratory found that when a rat makes a decision about a sound, the information needed to make the decision is passed to a particular subset of neurons in the auditory cortex, wherein axons project to a structure called the striatum. In the second major line of work in the Zador laboratory, they are developing new methods for determining the complete wiring instructions of the mouse brain at single-neuron resolution, which they term the "connectome." In contrast to previous methods, which make use of microscopy, these methods exploit high-throughput DNA sequencing. Because the costs of DNA sequencing are plummeting so rapidly, these methods have the potential to yield the complete wiring diagram of an entire brain for just thousands of dollars.

# UNDERSTANDING THE NEURONAL BASIS OF INTERNAL MODELS OF THE WORLD AND MAPPING ODOR SPACE

**F. Albeanu**   W. Bast   D. Cowan   D. Hernandez Trejo
H. Chae   M. Davis   S.S.K. Teja
Y. Chen   M. Dussauze   P. Villar
S. Choi   P. Gupta

The focus of our research group is twofold: (1) understanding the logic of odor space and olfactory neuronal representations underlying olfactory perception, and (2) understanding how brain-wide neuronal circuits learn the statistics of the world (internal models) and solve fundamental sensorimotor challenges in closed-loop behaviors. We investigate how the brain encodes and interprets inputs from environment as a function of context and prior expectations to implement meaningful behaviors. We focus on neuronal circuits in the olfactory stream (olfactory bulb, olfactory cortex, and striatum), as well as on inputs to sensory circuits from the motor cortex and association areas, in wild-type mice and models of psychiatric disorders. The broad scope of our research is understanding how actions relate to perception.

## The Logic of Olfactory Bulb and Piriform Cortex Outputs Revealed by High-Throughput Single-Neuron Projection Mapping Using DNA Barcode Sequencing

The nature of odor object representations in olfactory networks remains elusive. This stems from the apparent complexity of the olfactory system that relies on hundreds of odorant receptors (ORs) to recognize numerous volatile compounds and from our lack of insight into how the brain extracts and sorts these representations from the sensory periphery. In contrast to the organized long-range connectivity and spatial representations of other sensory modalities (vision, audition, and somatosensation), studies based on sparse labeling and axonal tracing found that the olfactory bulb (OB) outputs, the mitral and tufted cells (MTCs), project in a highly distributed and seemingly random fashion to their largest cortical target, the piriform cortex (PC). This has inspired several influential computational models over the past five decades that have proposed random connectivity schemes between the olfactory bulb and rest of the brain as a necessary basis for learning arbitrary combinations of odorants as needed during the lifetime of animals.

In collaboration with the Koulakov and Zador laboratories, we explored the brain-wide projections of 4,500 MTCs and 22,000 PC output neurons using single-cell resolution, sequencing-based mapping techniques (BARseq and MAPseq, developed in the next-door Zador laboratory). We identified several distributed OB projection modules, reproducible across individuals, as well as systematic and differential logarithmic tiling of cortical areas such as PC and the anterior olfactory nucleus (AON). Each module targets a distinct territory along the anterior-posterior (A-P) axis of PC and a select set of neighboring non-piriform target regions specifically co-innervated by the same cells (Fig. 1). Thus, the location of an OB cell projection along the A-P axis of the piriform is highly predictive of its projections to other target areas.

Furthermore, we described several input–output piriform cortex circuit motifs that span the anterior–posterior axis of PC, and each connects with downstream brain regions within the same, but not across, OB projection modules (Fig. 1). Specifically, each cortical output motif is densely connecting to downstream target brain regions within the same OB projection module, but only sparsely projecting to a different OB projection module. Therefore, MTCs differentially innervate in a precisely organized (*nonrandom*) manner distinct downstream circuits, which may extract specific features of odorants.

### Impact and Significance

These results challenge the popular random connectivity model of olfactory processing and encourage formulation of novel computational models to account for the logic of information flow in the
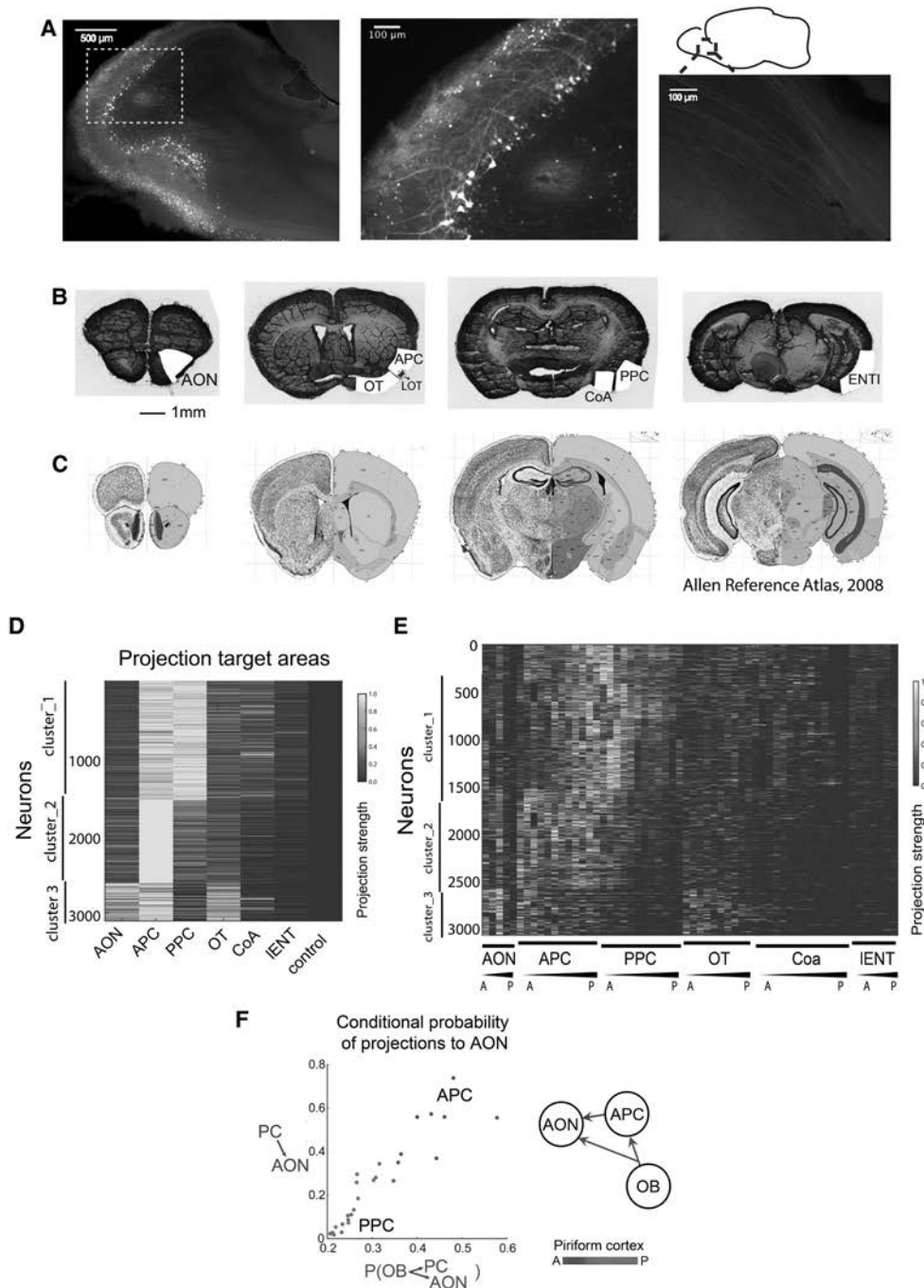
**Figure 1.** MAPseq experiments reveal the organization of the olfactory bulb and olfactory cortex projection patterns. (*A*) Fluorescence image of barcoded virus injection site labeling mitral cell somata and axons in the olfactory bulb (OB). (*B,C*) Slicing and laser microdissection of tissue from the six major OB target areas (*B*) and the corresponding Allen mouse brain reference atlas locations (*C*). (*D,E*) Heatmaps of projection patterns of 3,000 OB output barcodes (BCs) from six mice across the major target areas of the bulb shown with single-area resolution (*D*) or 200-μm slice resolution along the A-P axis of the brain (*E*). Projection density is color-coded. BCs are clustered according to similarity. Clusters are sorted according to the specificity of their projections. (*F*) Conditional probability of projections from piriform cortex to AON versus the distribution of projections from the olfactory bulb along the A-P axis of the piriform cortex given that they also project to the AON. OB neurons that project to AON target the anterior part of the piriform cortex and less so the posterior piriform cortex. Similarly, anterior piriform cortex output neurons are more likely to target the AON than those originating in the posterior piriform cortex.

olfactory system. The large amount of single-cell resolution projection data that MAPseq promises is revolutionary. Within a few experiments, this approach enabled gathering more data toward this end than the whole field over several decades and obtained novel fundamental insight. By further combining MAPseq with multiphoton imaging of functional responses to large odor sets from the barcoded mitral and tufted cells in awake mice, we aim to connect the tuning of the OB outputs to their projection patterns, a long-sought-after goal for any neuronal circuit in the brain. In addition, by analyzing the odor tuning of MTC cells in awake mice naive, or engaged in odor discrimination and navigation tasks, we would like to determine how any arising projection and functional biases are shaped by brain state and behaviorally relevant variables. These projects will open exciting venues for understanding how odor information is demultiplexed from the sensory periphery to the rest of the brain and for probing the specificity of interplay between feedforward and feedback processing.

## Neuronal Substrates of Olfactory Perception

Olfactory perception is critically dependent on responses of a large cohort of olfactory receptors (ORs) to an even larger set of odorants. Progress in olfaction has been substantially impeded by failure in the field to accumulate information about the OR/odorant binding affinity matrices in vivo with sufficiently high throughput, and to relate them across different conditions in a predictive manner. Our approach relates the molecular identity of ORs to their in vivo responses and to downstream activity and connectivity patterns for hundreds of odorants and thousands of individual neurons per brain (Fig. 1). To this end, we use in vivo functional imaging of large odor panels in conjunction with fluorescent in situ RNA sequencing via BARseq, MAPseq developed at CSHL, and modern machine-learning techniques. We aim to predict the perceptual quality of an odorant based solely on its molecular structure, a long-sought-after dream in olfaction. In vision and other senses, substantial progress was achieved by understanding the features of the stimulus space that are represented by the brain. The realization that color perception is based on three types of cone photoreceptors enabled

the invention of cameras and displays that faithfully reproduce any natural stimulus by mixing a basis set of just three lights. In the case of smell, we lack any comparable conceptual understanding. We do not understand what properties of odorants lead to particular percepts and how these properties are represented in the neuronal activity. Our approach is ideally suited to provide the answer, by first building a framework for relating the physical and neuronal spaces and further connecting them to odor perception using modern machine learning techniques (deep neural networks or DNNs). Altogether, we aim to identify the analogs of the red–green–blue basis set of odor perception. (This work is performed in collaboration with the Zador and Koulakov laboratories.)

## Long-Range Functional Specificity of Parallel Processing Loops in Mammalian Olfaction

Long-range connectivity across different areas is a highly recurrent feature of brain architecture. For example, in sensory processing streams, canonical feedforward thalamocortical connections are complemented by massive cortical feedback projections to the corresponding thalamic areas. The computational roles of such feedforward–feedback loops remain largely unknown, although cortical feedback has been implicated in a variety of functions ranging from gain control to predictive coding. To address this issue, we have recently begun to decipher the logic of information flow within the early mammalian olfactory system. Specifically, we investigated whether different types of olfactory bulb (OB) projection neurons carry nonredundant signals to particular cortical areas, and to what degree feedback from those target areas to the OB is specific to the input they receive.

We find that the two classes of OB outputs, the mitral and tufted cells (MC, TC) which innervate distinct sets of higher brain areas (including piriform cortex [PC] vs. anterior olfactory nucleus [AON]) are in turn specifically regulated by differential negative feedback from these areas. Cortical feedback from AON preferentially controls the gain of TC odor representations, whereas PC feedback specifically shapes mitral cell odor responses. These feedback signals are highly precise, acting in an odor-cell pair-specific manner, proportionally to the strength of feedforward
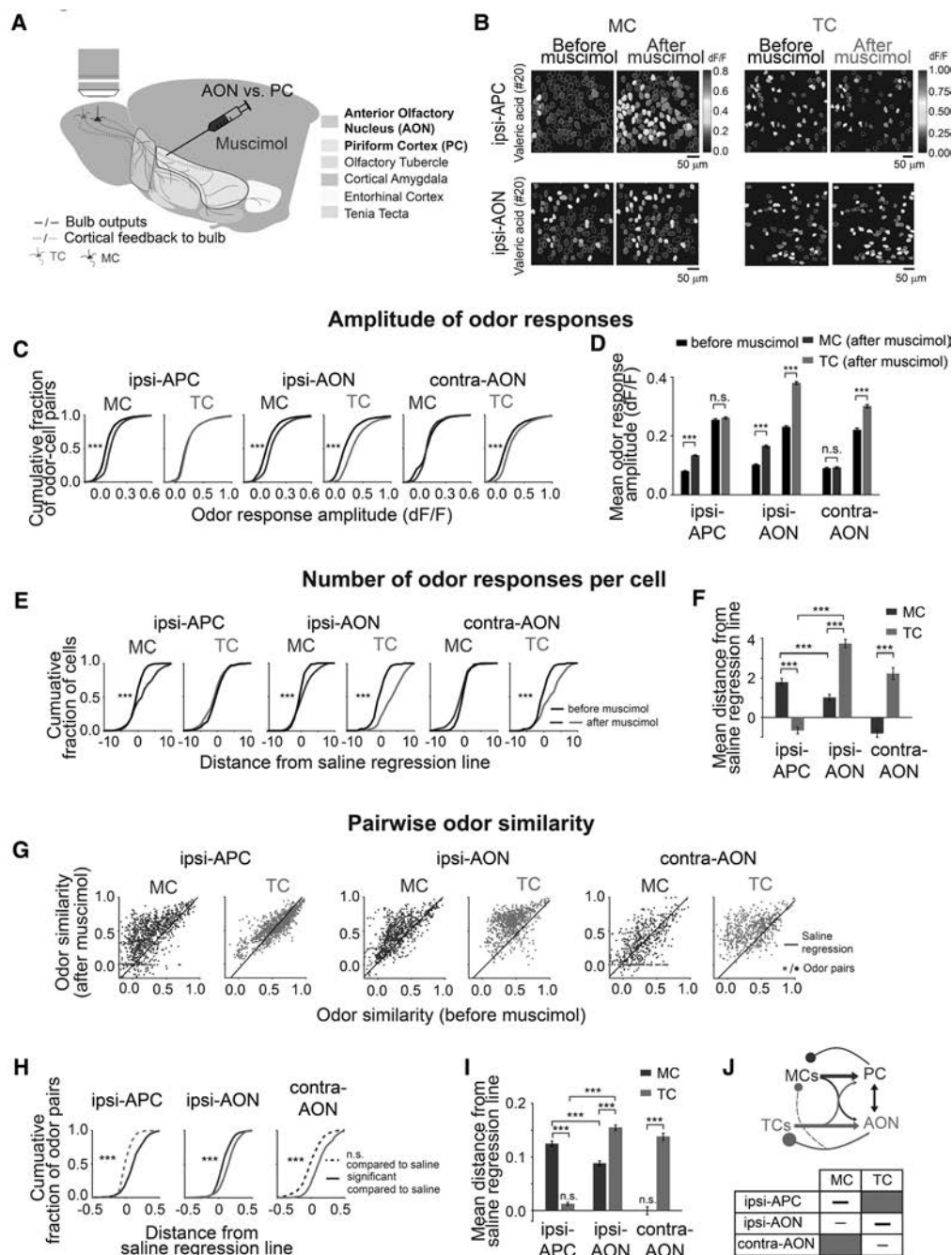
**Figure 2.** Two long-range feedforward–feedback functional loops in the olfactory system. (*A*) Schematic of the experimental framework: monitoring the activity of the olfactory bulb (OB) outputs, the mitral and tufted cells (MC and TC), in awake head-fixed mice via multiphoton imaging, while suppressing the feedback from the anterior olfactory nucleus (AON) and anterior piriform cortex (APC). (*B*) Color maps showing average fluorescence change in response of MCs and TCs to valeric acid (1:100 dilution) in example fields of view before (*left*) and after (*right*) muscimol injection in the APC (*top*) and AON (*bottom*). (*C–F*) (*Left*) Response amplitude (*C,D*) and number of odor responses (*E,F*) per cell before and after muscimol injection in the ipsi-APC, ipsi-AON and contra-AON; (*D,F*) quantification of response amplitude and average number of responses across conditions (before vs. after muscimol injection). (*G–I*) Pairwise odor similarity across cells before and after suppression of cortical feedback; (*G*) each dot corresponds to one odor–odor pair. (*H,I*) Cumulative distributions and quantification of distances from saline regression line of pairwise odor similarity across conditions (before versus after muscimol injection). (*J*) Cartoon representations of specific feedforward–feedback loops engaging the mitral cells and APC and the tufted cells and the AON.

drive (Fig. 2). Furthermore, we find that robust and odor-specific sensory representations emerge already in the bulb outputs, and are distinct across the mitral and tufted populations. Surprisingly, tufted cell ensembles substantially outperform mitral cells in several tasks, including decoding of concentration-invariant stimulus identity as well as concentration calling (Fig. 3). Moreover, suppression of PC feedback substantially impairs odor decoding by mitral cell ensembles. In contrast, performance of decoders based on tufted cell responses was only mildly affected by AON silencing, consistent with a gain control action of the AON-to-TC feedback.

### Impact and Significance

These results identify two interconnected feedforward–feedback loops in the early olfactory system and suggest they have specialized roles in odor processing—one related to sensory processing and the other well positioned for flexible learning of contextual associations. In addition, taken together, the findings show that invariant odor identity representations emerge already in the OB outputs, preferentially in tufted cell ensembles, which project largely to AON and less to the piriform cortex. This challenges a popular model that posits that recurrent circuitry intrinsic to the piriform cortex computes intensity-invariant odor identity drawing from the mitral cells' inputs. Instead, the results indicate that the piriform cortex receives odor identity information indirectly from the bulb's tufted cells via the AON. To test this model, we are currently assessing the effect of AON suppression on PC neural activity and odor decoding capacity. Finally, to probe the roles of cortical feedback loops during behavior, we are manipulating feedback input to the OB from the APC and AON respectively in mice engaged in concentration-invariant odor discrimination and contextual learning tasks.

## Understanding the Relationship between Olfactory Perceptual Discriminability and Glomerular Response Features

For rodents, the ability to recognize and discriminate particular combinations of volatile compounds is essential for their survival. Mice can easily report the difference between weak, similar odors in rich sensory scenes, even when stronger odorants fluctuate in the background. To date, the neural mechanisms underlying such behavior remain unknown.

To understand the neural basis of odor discrimination, we measured and manipulated the activity of the input nodes of the olfactory system, the glomeruli. By using wide-field optical imaging in conjunction with odor stimulation, we tracked the position of glomeruli and quantified their odor-response properties; this allowed us to define different sets of affine and nonaffine glomeruli with variable number of components. We aim to determine the relationship between the discriminability of olfactory stimuli and the similarity of glomerular odor response profiles. We additionally quantify the discriminability of the stimuli with the degree of overlap between different sets of glomeruli, as well as the physical separation of glomeruli on the bulb surface.

Toward this end, and to assess the specificity of photostimulation, we express red-activatable channelrhodopsin1 (ReaChR) in all mature olfactory sensory neurons and GCaMP6f in the OB output neurons. We use digital micromirror device (DMD)-based patterned illumination to selectively stimulate combinations of glomeruli on the dorsal surface of the bulb with subglomerular resolution (~10 μm) and high temporal precision (3 msec) in awake, head-fixed mice. Before optogenetic stimulation, using a large odor panel (up to approximately 100 stimuli), we identified the exact locations of glomeruli, revealing their shapes and response tuning to the odors sampled. We further create glomerular light patterns of known odor response similarity (within the range of our panel) and project specific glomerular inputs. In a two-alternative forced-choice discrimination task, we systematically relate the similarity of these light patterns to the perceived difference between them. Further, using a novel strategy to decouple patterned photostimulation and two-photon imaging across different axial planes, we are monitoring the responses of mitral and tufted (M/T) cells in the deeper layers of the bulb.

We are further implementing strategies that will enable noninvasive, functional dissection of neuronal networks with cellular resolution in behaving animals. This will be brought about via a closed-loop strategy involving real-time control of activity of select neurons with simultaneous monitoring of the concomitant effects of these manipulations on neuronal outputs within the circuit and elsewhere in the brain. Briefly, we are employing digital holography
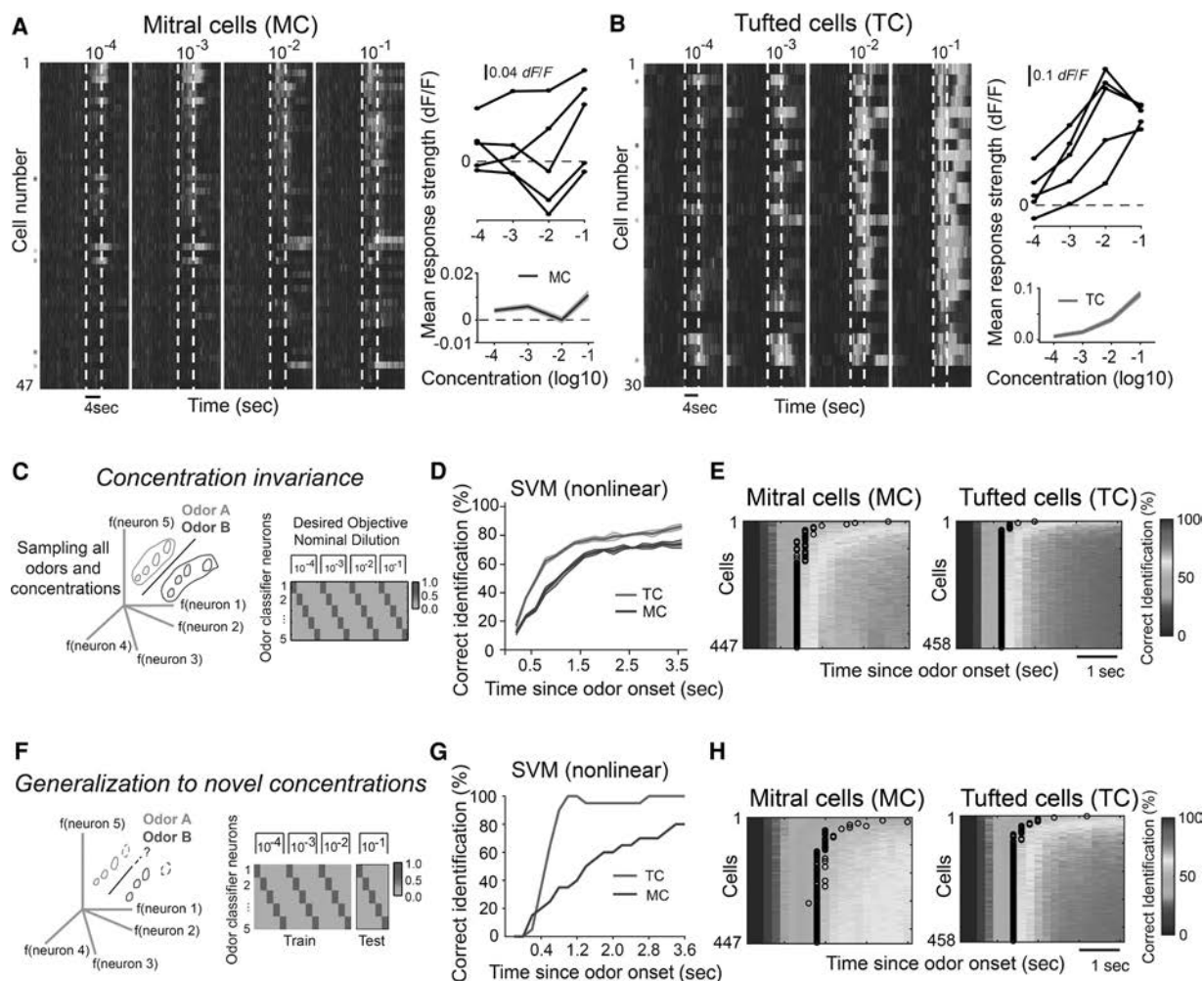
**Figure 3.** Tufted cells are superior to mitral cell ensembles in decoding concentration-invariant odor identity. (*A*) Mean peristimulus time histogram of 47 exemplar simultaneously recorded mitral cells to increasing concentrations of valeraldehyde within the same FOV (*left*). Color indicates normalized change in fluorescence with respect to pre-odor baseline (*dF/F*$_0$). Dotted lines mark odor presentation. Mean odor responses of four example cells in *A* indicated by the colored fiduciary marks (*right*, *top*). Mean population concentration response across five odors and all MC (*right*, *bottom*). (*B*) Same as in *A*, except for 30 example tufted cells from a different FOV. (*C*) Cartoon of a first decoding task—concentration-invariant odor identification. Each stimulus occupies a distinct portion of the neural state space and all concentrations of a given odorant need to be grouped together by the classifier. (*Right*) The decoding objective function in which one hypothetical classifier neuron signals the presence (value = 1) of its corresponding odorant for each of four concentrations sampled and absence (value = 0) for all other stimuli in the panel. (*D*) Cross-validated classification performance of a nonlinear support vector machine (SVM) as a function of time for mitral cells (blue) and tufted cells (red). (*E*) 2D classification performance color map for all four experimental conditions as a function of time (abscissa; bin size, 200 msec), while varying the number of neurons included in the analysis using bootstrap resampling (ordinate, bin size, five neurons). Black dots indicate the first occurrence of 50% performance in each row. Chance performance is 20% (five odors). (*F*) (*Left*) Cartoon of a second decoding task—generalization to a novel concentration. The neural network learns to group any three of four concentrations sampled for a given odorant together. The cross-validated performance is tested on the ability to classify the fourth concentration previously not used to train. (*Right*) Decoding objective function where training is done using three concentrations (10-4, 10-3, 10-2 nominal oil dilutions) and performance is evaluated on the fourth concentration (10-1 nominal odor dilution). (*G*) Cross-validated classification performance of a nonlinear support vector machine (SVM) as a function of time for mitral cells (blue) and tufted cells (red). (*H*) Same as *E*. Each plot reflects the classifier performance averaged across the four test concentrations. The classifier was trained independently using any three concentrations and tested on the fourth one (not included in the training set).

methods via spatial light modulators (SLMs) to optogenetically control neurons of interest at single-cell level and DMD-based methods to control cell type–specific populations across large brain regions (Fig. 4). This allows us to both replicate and systematically manipulate stimulus-evoked activity patterns in a circuit. We are simultaneously employing two-photon calcium imaging and electrophysiology within the same and different brain regions (olfactory bulb vs. olfactory cortex) to dissect how the alteration of select circuit elements, or their specific properties, affects the output of the network. This closed-loop approach will make it possible to determine the spatiotemporal integration rules within the bulb and the olfactory cortex, investigate the relevance of spike-time codes, and reveal underlying decoding schemes. (This work is performed in collaboration with the Engel laboratory.)

## Mosaic Representations of Odors in the Input and Output Layers of the Olfactory Bulb

The elementary stimulus features encoded by the olfactory system remain poorly understood. We examined the relationship between 1,666 physical–chemical descriptors of odors and the activity of olfactory bulb inputs and outputs in awake mice. Glomerular and mitral/tufted cell (MTC) responses were sparse and locally heterogeneous, with only weak dependence of their positions on physical–chemical properties. Odor features represented by ensembles of MTCs were overlapping but distinct from those represented in glomeruli, consistent with extensive interplay between feedforward and feedback inputs to the bulb. This reformatting was well described as a rotation in odor space. The physical-chemical descriptors accounted for a small fraction in response variance, and the similarity of odors in physical–chemical space was a poor predictor of similarity in neuronal representations (Fig. 5).

### Impact and Significance

These results suggest that commonly used physical–chemical properties are not systematically represented in bulbar activity and encourage further studies for better descriptors of odor space. (This work is performed in collaboration with the Koulakov laboratory.)

## Understanding the Neuronal Substrates of Internal Models of the World

During behavior, sensation and action operate in a closed loop. Movements shape sensory input, and sensory inputs guide motor commands: Where one looks determines what one sees. Through experience, the brain learns the reciprocal relationship between sensory inputs and movements to build internal models that predict the sensory consequences of upcoming actions (sensorimotor predictions). Comparing internal sensory predictions to actual sensory observations generates prediction errors that can be minimized by learning increasingly accurate models of the world. This exchange of sensory inputs and egocentric expectations is at the core of active perception. Experimental investigation of this idea has been sparse and split between behavioral interrogation of sensory-guided, precise motor control in primates (visuomotor adaptation tasks) and the search for neuronal substrates of sensory predictions in rodents via simpler running-based closed-loop behaviors.

To study internal models, both at behavioral and circuit level, we developed a novel behavioral task in which head-fixed mice are trained to steer the left–right location of an odor source by controlling a lightweight lever with their forepaws. In this manner, we (1) link a precise motor action to well-defined sensory expectations (odor location) and (2) subsequently violate the learnt expectations via online feedback perturbations in trained animals. Expert mice showed precise movements that were locked to the instantaneous odor feedback during normal closed-loop coupling. However, when sensory feedback was transiently interrupted (halting of odor source) or distorted (displacement of odor source or change in movement gain), movements were initially guided by each animal's learnt internal model, in the absence of sensory feedback, and, further, quickly adapted (within a few sniffs in single trials) in accordance with the instantaneous sensory error (Figs. 6 and 7).

In an open-loop instantiation of the task, the odor feedback cannot be controlled by the mouse and is rather generated via a replay of the odor movements generated by the mouse during a closed-loop block of the same session. Mice tended to follow similar paw/lever movement patterns to those observed in the closed-loop trials, matching their sensory experience
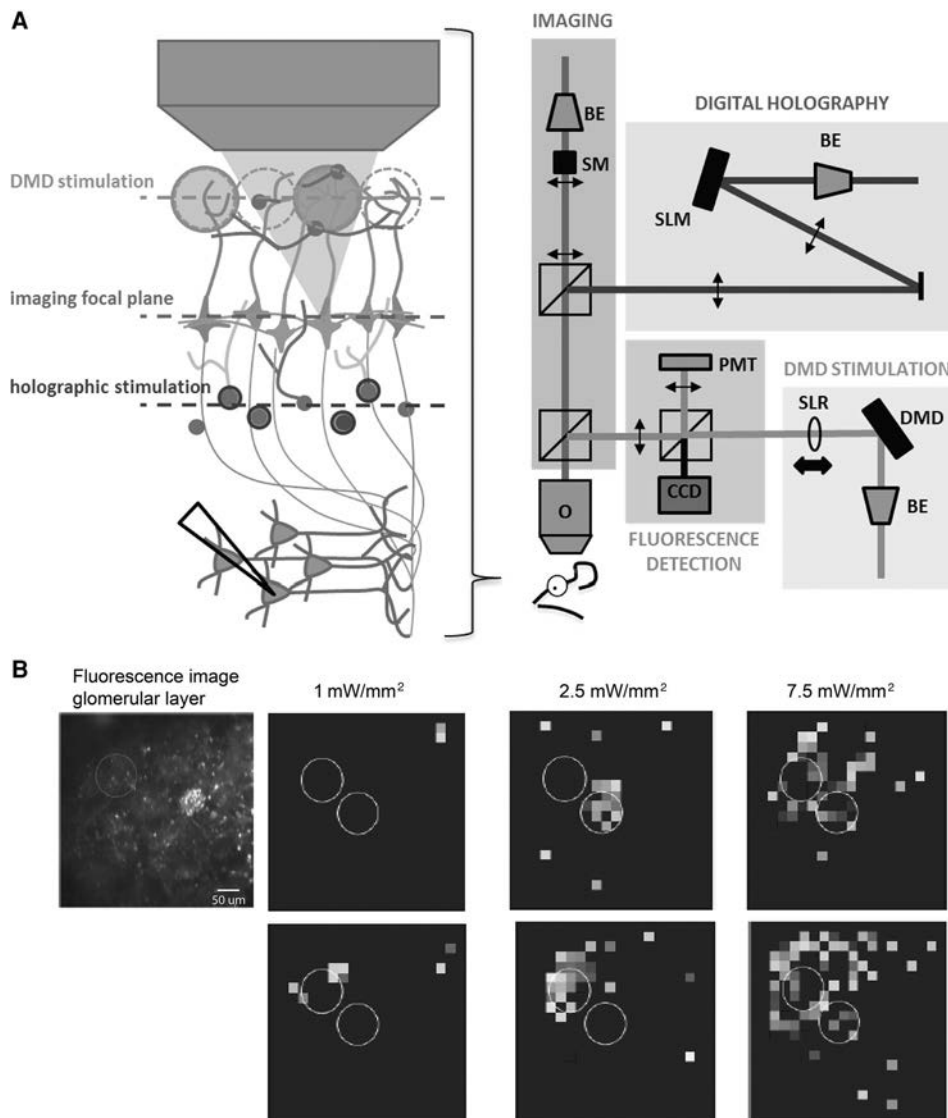
**Figure 4.** Combined imaging and photostimulation. (*A*) (*Left*) Experimental configuration, combining scanning two-photon imaging (*center*), DMD photostimulation (*top*), and holographic photostimulation (*bottom*). (*Right*) Schematic of the experimental microscope. DMD stimulation is used to create spatiotemporal light patterns on the surface (<100 μm). Digital holography is used to photostimulate deeper (<500 μm) in the brain with cellular resolution. Calcium activity is monitored in an independent optical plane by two-photon imaging and by electrodes in downstream brain regions. (BE) Beam expander, (SM) scan mirrors, (O) objective, (PMT) photomultiplier, (SLR) camera lens, (DMD) digital micromirror device, (SLM) spatial light modulator, (CCD) charge-coupled device. (*B*) Specificity of optogenetic patterned stimulation of neighboring glomeruli in TBET-Cre mice expressing Chrimson-tdtomato in the mitral and tufted cells versus increasing light intensities.

with their previous sensorimotor model. Nevertheless, frequent replay of open-loop trials further induced sensorimotor decoupling.

Specific manipulations of olfactomotor feedback in a series of online perturbations, as well as in the open- versus closed-loop versions of the task, identified neurons that responded selectively or modulate substantially their response in the face of mismatch between the expected and the actual olfactory consequences of motor output. Such error signals were prominent in several parts of the olfactory cortex and striatum, consistent with models of predictive processing.
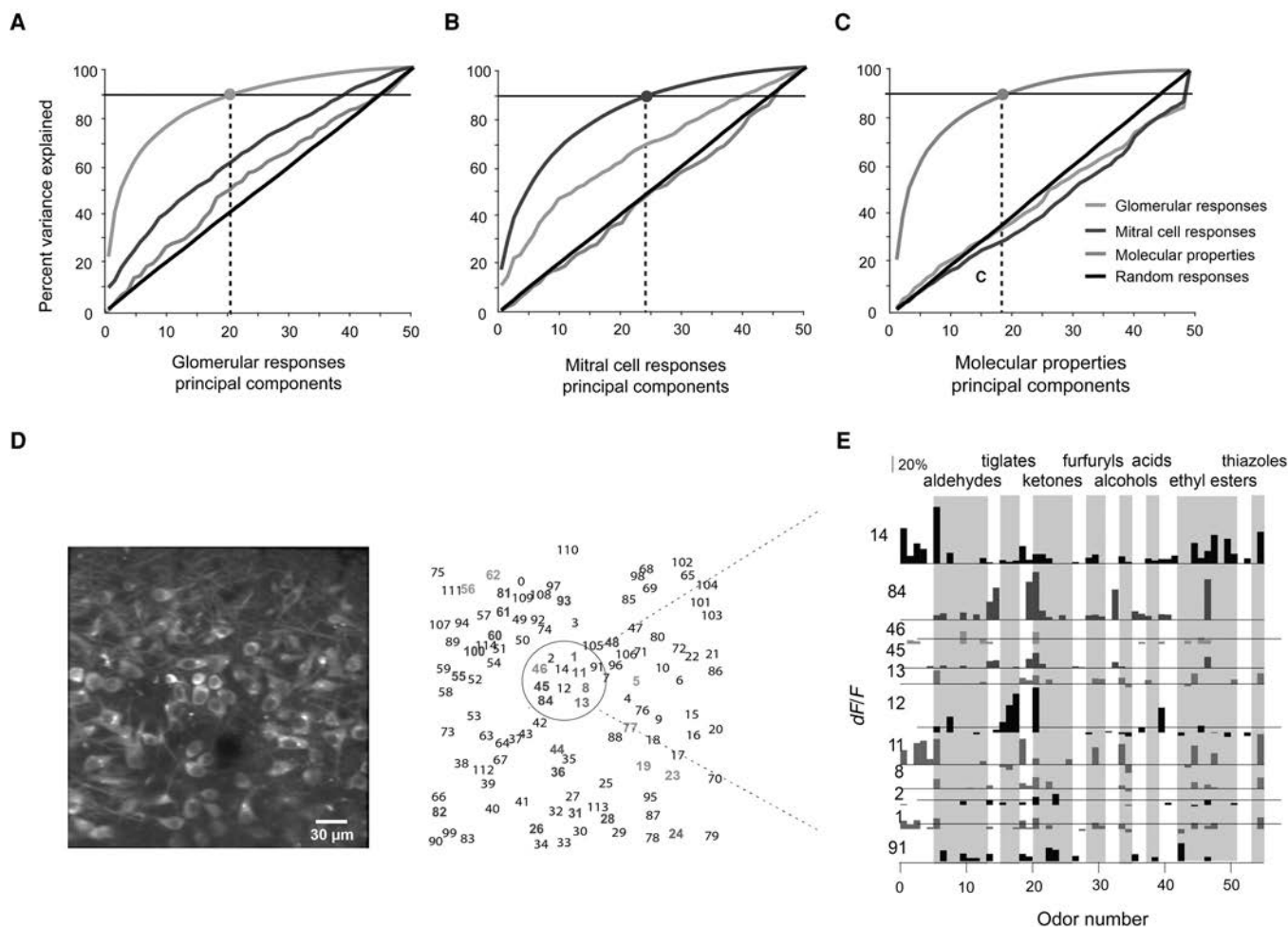
**Figure 5.** Diversity of odorant representations in the glomeruli and mitral cells, independent of molecular properties. (*A–C*) Results of principal component analysis (PCA) for glomerular (*A*), mitral cell responses (*B*), and molecular properties (*C*) (*n* = 871 glomeruli, *n* = 639 mitral cells, *n* = 1,320 properties). Percent of variance explained is shown as a function of the number of included principal components (PCs). (*A*) Percent variance explained of glomerular (green), mitral cell (blue) odor responses, molecular property strength vectors (gray), and random data controls (black) shown as a function of the number of included glomerular responses principal components. (*B,C*) Percent variance explained of glomerular (green), mitral cell (blue) odor responses, molecular property strength vectors (gray), and random data controls (black) shown as a function of the number of included mitral cell responses principal components (*A*) and of molecular properties principal components (*B*). Note that neural responses of both glomeruli and olfactory bulb outputs (mitral cells) are poorly tuned to the physical–chemical properties analyzed and instead reflect odor features that are not well captured by these molecular properties commonly used in computational chemistry, and by previous studies in olfaction. (*D*) (*Left*) Average resting fluorescence multiphoton image in the mitral cell layer. (*Right*) Numbers indicate the relative positions of 112 mitral cell bodies in the imaged field of view. (*E*) Example odor responses of mitral cell bodies within an arbitrarily picked 75-µm diameter region circled in *D* sorted according to functional chemical groups (aldehydes, tiglates, ketones, furfuryls, alcohols, acids, ethyl esters, thiazoles). Note that neighboring neurons (enclosed in the circle in *D*) have diverse as well as similar odor tuning (e.g., cells 45 and 84 vs. cell 12). We probed the local diversity further, by rearranging the odor response spectra (ORS) so as to group odors in terms of chemical classes (ketones, acids, etc.). Mitral cells responding to many different classes of chemicals can be found even within a small region of glomerular size.
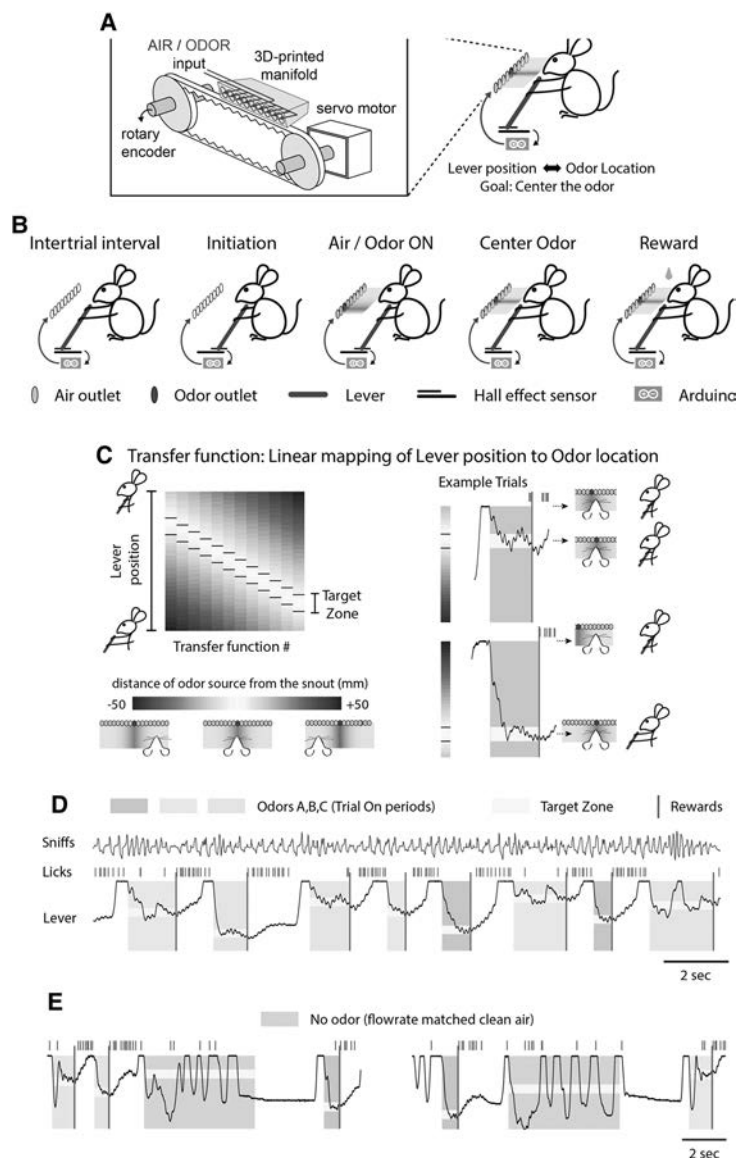
**Figure 6.** A novel method for fast, closed-loop control of odor source location. (*A*) Schematic of the system: 1D movement of the lever (read by a Hall effect sensor) is transformed to left–right displacement of an odor source. The odor location is manipulated by displacing a 3D-printed manifold, affixed to a timing belt servo system. The manifold comprises a central odor outlet and 16 air outlets on either side. Flowrate (0.2 L/min) is matched across all outlets. To obtain rewards, animals are required to align the odor outlet to their snout. (*B*) Trial structure: Mice initiate trials by retracting and holding the lever to activate air/odor flow and place the odor source at a fixed starting location. Mice can then steer and hold the odor, centered on their snout, by bringing the lever within a narrow target zone. Maintaining target zone hold for ~300 msec triggers reward and terminates the trial. (*C*) (*Left*) Different transfer functions map lever position to odor source displacement. Colors indicate relative lateral distance from the snout. Black lines demarcate the target zone, a set of lever positions that place the odor <3 mm from the snout. All transfer functions have the same gain; unit displacement of the lever results in unit displacement of odor, but the odor source location at trial start differs across trials. Using different transfer functions across trials ensures that reward availability is not associated with any specific lever position. (*Right*) Two example trials with different transfer functions and corresponding lever trajectories in time (black). Gray bar indicates trial on period. Yellow demarcates the target zone. Schematics show odor locations at trial start and in the target zone. (Green ticks) Water reward, (red ticks) licks. (*D*) Example behavioral trace from an expert mouse showing that the animal reliably centers the odor in each trial despite varying target zone locations (yellow bands). (*E*) Lever movements are unstructured in catch trials in which olfactory feedback is not provided, confirming that animals rely on closed-loop odor feedback to guide their movements during this task.
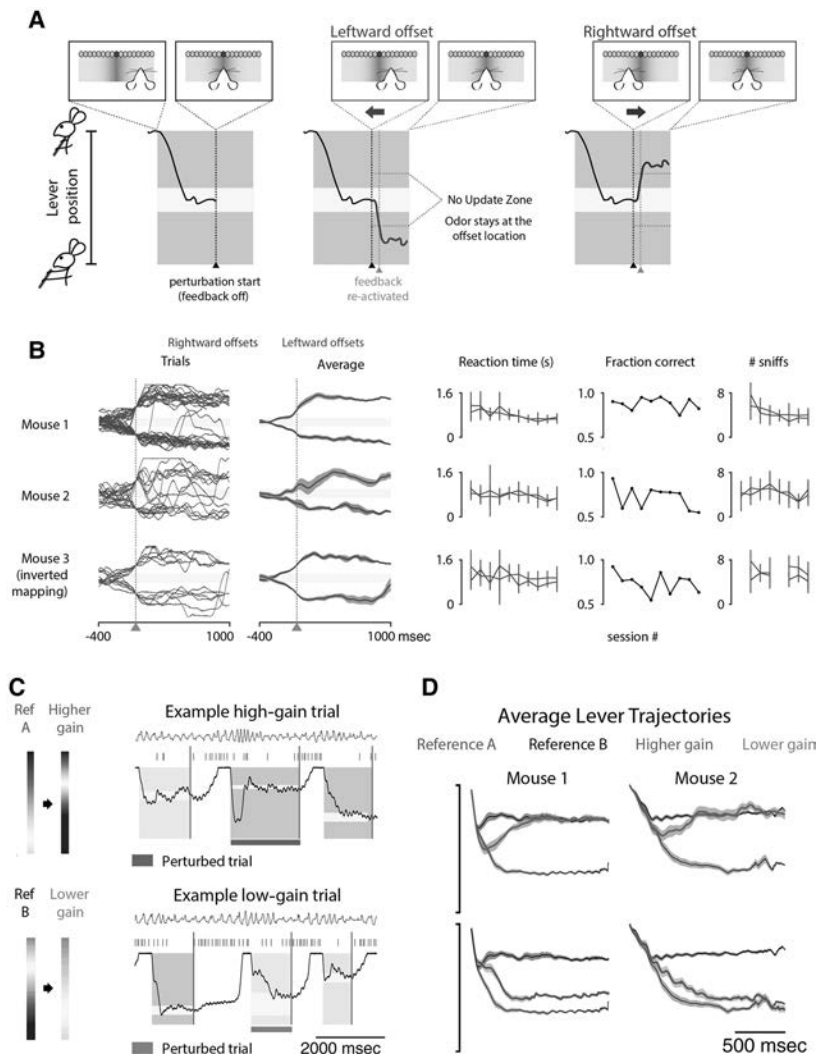
**Figure 7.** Feedback perturbations trigger corrective movements that reveal the animal's internal model. (*A*) Schematic of within-trial perturbations of odor location and expected corrective movements (red, leftward; blue, rightward odor offset). This example illustrates the scenario for a mouse that was trained to move the odor from left → right via forward motion of the lever (away from the body). Vertical black dotted line indicates time of odor displacement (perturbation start). No update in odor location is provided until the corrective movement crosses threshold (2x width of the target zone). This ensures that observed corrective movements purely reflect the animal's internal model and are not refined via instantaneous odor feedback. Vertical orange dotted line marks time of reactivation of closed-loop feedback. (*B*) (*Left*) Distinct trial-averaged and single-trial corrective movements in three mice, sorted by leftward (red) and rightward (blue) odor offsets. Mouse# 1-2 were trained to move the odor left → right by forward motion of the lever, whereas Mouse 3 was trained on the opposite (forward lever movement = right → left odor displacement). Note that corrective movements for left and right odor offsets diverge well before closed-loop feedback is reactivated. (*Right*) Summary plots showing session-averaged reaction times (time from offset start to threshold crossing), success rates (fraction of trials in which the corrective response was in the correct direction), and number of sniffs during perturbation period (offset start to feedback reactivation). (*C*) Schematic of gain-change perturbations. (*Top*) Example trial with higher (3x) lever gain compared to reference. As a result, centering the odor requires smaller lever displacement (target zone closer to the body) despite the same odor start location in perturbed and reference trials. Note how the animal initially overshoots and steers the lever to where the target zone would be given normal gain (1x) and then quickly corrects to successfully center the odor, despite the higher gain. (*Bottom*) Example trial with lower gain (0.4x) compared to reference. Here, the target zone is lower than expected. Thus, the animal initially stalls the lever (undershoots) but then quickly corrects. (*D*) Average lever trajectories during gain perturbation trials in comparison to reference control trials showing consistent overshoots upon higher-gain and undershoots upon lower-gain perturbations.

## Impact and Significance

A fundamental question in neuroscience is how the brain actively predicts its inputs and builds increasingly accurate models of the world to best match its predictions to reality. Active perception can be viewed as the continuous comparison of expected (sensorimotor predictions) and actual sensory inputs. Although the idea of mental simulation dates back as early as Plato, and is echoed by modern theories of cortical function, little is known about how the brain builds flexible internal models of the world and predicts upcoming inputs. This work provides an experimentally robust, high-throughput, and flexible platform to investigate the neural circuit–level substrates of internal representations.

## Other Collaborative Projects with CSHL Groups

*Huang*: Understanding the neuronal substrates of internal models across sensory modalities.

*Li*: Relating goal-directed behaviors to motivation and neuronal representation of internal models.

*Osten*: Optimizing light sheet–based approaches for fast optical reconstruction of neuronal circuits.

*Kepecs*: Modulation of VIP neurons by reward and punishment assessed using fiber photometry.

*Zador*: Optical monitoring and manipulation of neuronal activity in genetically and anatomically defined cortical circuits in animal models of cognition.

### PUBLICATION

Chae HG, Kepple DR, Bast WG, Murthy VN, Koulakov A, Albeanu DF. 2019. Mosaic representations of odors in input and output layers of the mouse olfactory bulb. *Nat Neurosci* **22**: 1306–1317. doi:10.1038/s41593-019-0442-z

# INVESTIGATING NEURAL CIRCUITS FOR MULTISENSORY DECISION-MAKING

**A.K. Churchland**    G. Bekheet    S. Gluf    S. Pisupati    X.R. Sun    C. Yin
                       L. Chartarifsky    A. Khanal    J. Roach    A. Urai

Understanding how brains make decisions remains a major challenge because decision-making involves multiple, diverse computations. These computations include assembling samples of sensory information, interpreting signals in light of the current internal state and ultimately selecting the best action to carry out the choice. This multistep process likely recruits numerous neural structures and activates specific cell types with precision within those structures. As a result, decision-making is well-suited to benefit from new experimental methods for targeting, measuring, and manipulating neurons that have proven critical for understanding parenting, aggression, fear conditioning, reward seeking, and navigation.

Our laboratory has played a critical role in bringing new experimental methods to decision-making. Importing and modifying new experimental methods allowed us to tackle fundamental biological problems that were infeasible or impossible to address using the traditional animal models for decision-making—human and nonhuman primates. We enhanced the breadth and precision of the experimental methods for studying decision-making by deploying them in conjunction with powerful mathematical tools.

Although previous experiments have begun to reveal how neural systems combine evidence to make decisions, major gaps remain. Specifically, very little is currently known about the neural mechanisms that make it possible to combine information from multiple sensory modalities for decisions. The gap is apparent even though it is clear from behavioral observations that neural systems can combine multisensory information. When parsing speech in a crowded room, for example, the listener makes use of both auditory information (the speaker's vocal sounds) and visual information (the speaker's lip movements). Understanding the neural mechanisms of multisensory decisions is critical for two reasons. First, it is essential for a complete understanding of sensory perception because real-world stimuli rarely activate a single sense in isolation. Therefore, understanding how the brain interprets incoming information requires understanding how the brain merges information from different senses. Second, it is likely of clinical importance as several developmental abnormalities appear to be related to difficulties in integrating sensory information. For example, abnormalities in multisensory processing are a hallmark of subjects with autism spectrum disorder. Impairments in multisensory processing are also observed in subjects with a collection of sensory abnormalities known together as sensory processing disorder and may also be evident in patients with Rett syndrome and dyslexia. Understanding the neural mechanisms of multisensory integration could inform treatment of those conditions. Our long-term goal is to understand how the brain can make decisions that integrate inputs from our multiple senses, stored memories, and innate impulses.

## Understanding Lapses in Perceptual Decision-Making

L. Chartarifsky, S. Pisupati (current address: Princeton University), A. Khanal, C. Yin

During perceptual decisions, subjects often display a constant rate of errors independent of evidence strength, referred to as "lapses." Their proper treatment is crucial for fitting models to behavior; however, they are often treated as a nuisance arising from motor errors or inattention. We have proposed a new explanation—that lapses can reflect a dynamic form of exploration. We showed that uncertainty about sensory stimuli modulates the probability of lapses in rats. These effects cannot be accounted for by traditional models in the field and are instead concisely explained by a normative model of uncertainty-guided exploration. Further, we showed that

changing the reward associated with one of the decisions selectively affects the lapses associated with that decision in uncertain conditions, while leaving unchanged very easy "sure-bet" decisions, as predicted by the model. Finally, we showed that pharmacological inactivation of secondary motor cortex and posterior striatum affect lapses across modalities. Together, our results suggest a novel value-based account for lapses, and that far from being a nuisance, lapses are informative about an individual animal's exploration–exploitation trade-off. We have placed a preprint of this work on bioRxiv (Pisupati et al. 2019). A new laboratory member, Chaoqun Yin, is building on the success of the original project. He will measure neural activity in frontal and striatal regions that we previously identified as being relevant to action value computations.

## A Novel Device to Support Electrophysiological Recordings during Ethological Behavior

G. Bekheet [in collaboration with A. Juavinett, University of California, San Diego]

The starting point for this project was a major problem we saw in the field: The field has recently begun to benefit from Neuropixels probes, a unique biosensor with the ability to measure neural activity from far more neurons than was previously possible. These probes are being released to the community and afford the opportunity to make high-quality recordings of hundreds of neurons simultaneously spanning multiple brain areas. However, these probes are essentially unusable for scientists studying freely moving or ethological behaviors, which together constitute a critical component of neuroscience research. Further, Neuropixels were not designed to allow reuse after chronic implantation. This is a devastating problem for any researcher who needs chronically implanted probes in any experimental setup or animal.

We devised a solution to move the field forward: We designed a new device, the apparatus to mount individual electrodes (AMIE). The AMIE can be used in conjunction with Neuropixels probes. The AMIE holds the Neuropixel probe safely and securely in place for weeks until the experiment is finished, at which point it can be explanted and reused

in another animal. This transformative technology will bring Neuropixels probes to an entirely new community who have, until now, been excluded from the opportunity to benefit from Neuropixels probes.

The paper was published this past year (Juavinett et al. 2019) and additional information was published in protocol format (Juavinett et al. 2020). Critically, we have made all the materials related to this device available to the community: The technical drawings, the methodological instructions, the photographs, and supporting code will, together, allow any researcher to rapidly adopt this new technology and begin to benefit from Neuropixels probes.

## Population Dynamics of Neurons during Decision-Making

This work was done in collaboration with F. Najafi (Allen Institute for Brain Sciences, Seattle).

The starting point of this project is that mathematical models of decision-making have long relied on inhibitory neurons, but the role of such neurons has never been tested in vivo. This stands in stark contrast to the neural circuits field in which the responses of inhibitory neurons have been carefully studied, but mainly during passive conditions. Our approach joins the study of decision-making with the field of neural circuits: During two-choice perceptual decisions, we used two-photon imaging to measure neural activity in large neural populations. We ran these experiments in transgenic mice in which we could distinguish excitatory from inhibitory neurons. Surprisingly, inhibitory population activity predicted the animal's trial-by-trial choice as accurately as excitatory population activity. The ability to predict choice emerged jointly in the two populations during learning.

Our experimental observations, combined with our powerful network simulations, argue that inhibitory neurons form subnetworks with excitatory neurons according to their functional properties, conferring stability and robustness. This conclusion is in stark contrast with leading decision-making models. Our discovery of this architecture in mouse decision-making structures suggests that it may reflect a canonical computation that is relevant to many behaviors. These results were published this year (Najafi et al. 2019).

## Brain-Wide Macrocircuits That Support Decision-Making

S. Musall, S. Gluf, X.R. Sun [in collaboration with M. Kaufman, University of Chicago]

Cognition and action are typically studied separately and are assumed to recruit largely nonoverlapping neural structures. Little is known about how cognitive and action processes interact, especially for actions beyond simple instructed movements used in laboratory behaviors. Here, we tracked a vast array of movements in mice making decisions about auditory/visual stimuli, measured neural activity across the entire dorsal cortex, and developed a model to connect single-trial neural activity to movement and cognitive variables. Neural activity reflected both kinds of variables, but was dominated by movements. Using new methods, we then partitioned average neural activity into cognitive and movement components. This revealed that neurons with similar average responses could reflect utterly different combinations of cognitive and movement variables. Taken together, our observations argue that cognitive functions and movements can be tightly intertwined, and that during cognition, movements are a much higher priority than previously believed. This work was published this year (Musall et al. 2019a). We also highlighted our work alongside others' in an influential review about behavioral diversity and neural computations (Musall et al. 2019b).

## International Brain Laboratory

A. Urai

I cofounded this organization and secured funding in 2017 from the Wellcome Trust and the Simons Collaboration on the Global Brain. The International Brain Laboratory is a virtual laboratory, unifying a group of 21 experimental and theoretical neuroscience groups distributed across the world to understand the neural computations supporting decision-making. During the past year, we have made tremendous progress. We presented 10 posters at the Society for Neuroscience Annual meeting and placed a preprint on bioRxiv about the data architecture system we have developed to support this collaboration (the International Brain Laboratory et al. 2019).

## Neural Dynamics

J. Roach [in collaboration with T. Engel, CSHL]

Classical circuit models of decision-making focus solely on the effects of recurrent excitation, treating inhibitory neurons as agnostic facilitators of competition between excitatory subpopulations. Inspired by recent experimental results reporting highly selective inputs to and firing by inhibitory neurons within cortical circuits, we have developed a mean-field firing rate model of a cortical decision-making circuit that parameterizes selectivity in connection strengths between subgroups within the excitatory and inhibitory populations. We found that to produce network selectivity for a single choice outcome excitatory–excitatory selectivity (recurrent excitation) must always be high, whereas excitatory–inhibitory selectivity can be varied over a wider range, as long as inhibitory selectivity is changed as well. Specifically, for every decrease in excitatory–inhibitory selectivity, there must be a corresponding increase in inhibitory selectivity. From this work, we can make an inference as to the circuit structures that support reliable perceptual decision-making with equal excitatory and inhibitory selectivity, providing greater insight into the role of inhibition in cortical computation. We will now test these experimentally. We presented a poster on this project at the Society for Neuroscience Annual Meeting (Roach et al. 2019).

### PUBLICATIONS

Juavinett AL, Bekheet G, Churchland AK. 2019. Chronically-implanted Neuropixels probes enable high yield recordings in freely moving mice. *eLife* **8:** 47188. doi:10.7554/eLife.47188

Musall S, Kaufman MT, Juavinett AL, Gluf S, Churchland AK. 2019a. Single-trial neural dynamics are dominated by richly varied movements. *Nature Neurosci* **22:** 1677–1686. doi:10.1038/s41593-019-0502-4

Musall S, Urai AE, Sussillo D, Churchland AK. 2019b. Harnessing behavioral diversity to understand neural computations for cognition. *Curr Opin Neurobiol* **58:** 229–238. doi:10.1016/j.conb.2019.09.011

Najafi F, Elsayed GF, Cao R, Pnevmatikakis E, Latham PE, Cunningham JP, Churchland, AK. 2019. Excitatory and inhibitory subnetworks are equally selective during decision-making and emerge simultaneously during learning. *Neuron* **105:** 165–179.e8. doi:10.1016/j.neuron.2019.09.045

Pisupati S, Chartarifsky-Lynn L, Khanal A, Churchland AK. 2019. Lapses in perceptual judgments reflect exploration. bioRxiv doi:10.1101/613828

Roach JP, Churchland AK, Engel TA. 2019. Inhibitory selectivity and choice performance in a circuit model of perceptual decision making. *Society for Neuroscience Abstracts* (2019).

The International Brain Laboratory, Bonacchi N, Chapuis G, Churchland AK, Harris KD, Rossant C, Sasaki M, Shen S, Steinmetz NA, Walker EY, et al. 2019. Data architecture and visualization for a large-scale neuroscience collaboration. bioRxiv doi:10.1101/827873

*In Press*

Juavinett AL, Bekheet G, Churchland AK. 2020. Implanting and recycling neuropixels probes for recordings in freely moving mice. *Bio-protocol* **10:** e3503. doi:10.21769/BioProtoc.3503

# DYNAMICS OF BRAIN NETWORKS FOR PERCEPTION AND COGNITION

**T.A. Engel**   C. Aghamohammadi   C. Langdon
               M. Genkin          J. Roach
               R. Kwapich         Y. Shi

Core brain functions—perception, attention, and decision-making—emerge from complex patterns of neural activity coordinated within local micro-circuits and across brain regions. Recent advances in massively parallel neural recording technologies enable monitoring activity from hundreds of neurons simultaneously. These data present the opportunity to investigate how activity is orchestrated across large neural populations to drive behavior. Now, progress is mainly limited by the availability of computational methods to interpret high-dimensional neural activity data and by the lack of theories linking dynamic features of neural activity to circuit computations and behavior. The goal of our research is to understand how coordinated activity in distributed neural circuitry gives rise to behavioral and cognitive functions. To achieve this goal, we develop theory and computational methods and apply them to analyze large-scale neural activity recordings from our experimental collaborators.

## Interpretable Machine Learning for Large-Scale Neural Population Dynamics
M. Genkin, C. Langdon

Significant advances have been made recently to develop powerful machine-learning methods for finding predictive structure in large-scale neural activity data. However, most of these techniques compromise between flexibility and interpretability. On one hand, simple ad hoc models are interpretable by design but likely to distort defining features in the data. On the other hand, flexible models, such as artificial neural networks, can capture various data features but are difficult to interpret. We develop flexible and interpretable machine-learning methods to enable biological insights into behaviorally relevant neural dynamics based on large-scale activity recordings.

In one project led by M. Genkin, we developed a flexible, yet intrinsically interpretable, framework for inference of neural population dynamics from spike data. In our framework, population dynamics are governed by a nonlinear dynamical system defined by a potential function. Spiking activity of each neuron is related to the population dynamics through a unique firing-rate function. The shapes of potential and firing-rate functions are inferred from spike data. Unlike simple ad hoc models, our framework is flexible as it covers many different dynamics within a single model architecture. At the same time, the potential shape is intrinsically interpretable—for example, the potential minima reveal attractors. Leveraging our framework, we showed that optimizing flexible models for data prediction—the standard practice in machine learning—often produces overfitted models with spurious features. This behavior arises because standard techniques assume that overfitted models predict data poorly, which, however, does not hold for flexible models with high capacity. As a result, flexible models optimized for data prediction cannot be reliably interpreted. We developed an alternative strategy for identifying models with correct interpretation by comparing model features discovered from different data samples to separate true features from noise. As flexible models are gaining popularity for interpreting biological data, our results urge developing new model selection principles that prioritize accurate interpretation. We demonstrated one such principle within our framework.

In another project led by C. Langdon, we develop mathematical frameworks for interpreting dynamics of recurrent neural networks that perform behavioral tasks. Training recurrent neural networks on a behavioral task is a popular machine-learning method for constructing biologically realistic models of brain functions that can be compared with neural activity data. Although recurrent networks can be trained on

various tasks, it is not easy to understand how the connectivity and dynamics of trained networks produce the behavioral output. Recurrent networks are typically high-dimensional, as they comprise many units that all interact with each other—giving rise to mixed representations of task variables that are difficult to interpret. We developed an approach to project dynamics of a high-dimensional network onto a low-dimensional network trained to perform the same task. The low-dimensional network is designed to be interpretable, with each unit representing a single task variable. This dimensionality reduction takes advantage of the interpretability of the low-dimensional network to uncover the circuit mechanisms at play in the high-dimensional network. This dimensionality reduction technique can also be applied to interpret large-scale neural activity data, in which representations of task variables are mixed and distributed across many cortical areas.

## Specificity of Inhibition in Cortical Decision-Making Circuits

J. Roach [in collaboration with A. Churchland, CSHL]

During decision-making, excitatory and inhibitory cortical neurons equally modulate their firing rates to reflect the animal's choices. However, existing circuit models of decision-making treat inhibitory neurons as an untuned, nonspecific pool that facilitates competition between excitatory neurons. To test how choice tuning of inhibitory neurons affects decision-making, we extended circuit models to account for the specificity of inputs to and outputs from inhibitory neurons. Inhibitory output specificity falls into two general classes: ipsiselective, in which inhibitory neurons preferentially feed back to the excitatory neurons of the same tuning, and contraselective, in which inhibitory neurons preferentially output to oppositely tuned excitatory neurons (Fig. 1). We found that changing from an ipsiselective to a contraselective inhibitory class leads to faster and less accurate decisions by altering the attractor dynamics that underlie decision-making in the circuit. Our model predicts that when cotuned excitatory neurons are weakly coupled, contraselective inhibition facilitates decision-making. When cotuned excitatory neurons are strongly coupled, ipsiselective inhibition is required to stabilize circuit activity. Further, the model predicts divergent patterns of firing-rate correlations between choice-selective populations in contraselective versus ipsiselective circuits. The model reveals that ipsiselective circuits are weakly interacting pairs of cotuned excitatory and inhibitory pools, whereas contraselective circuits are well integrated and primed for competition between choice-selective populations. These model predictions
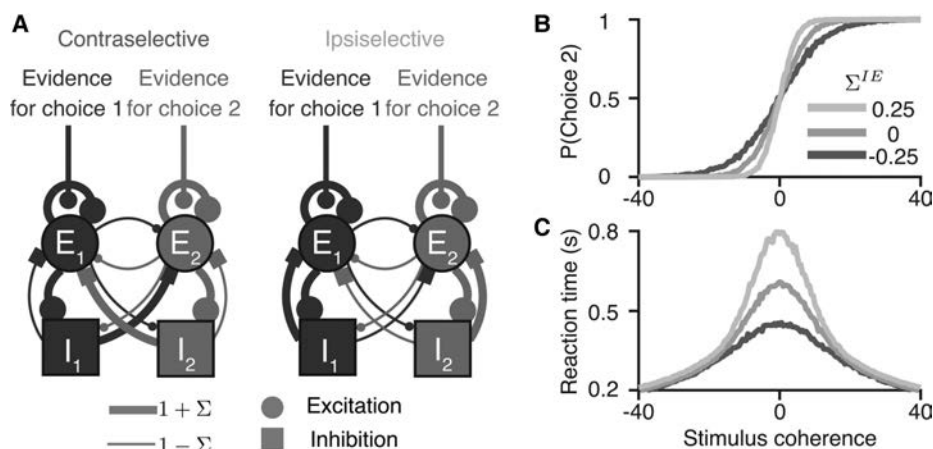


**Figure 1.** Selectivity of inhibitory outputs defines the speed versus accuracy tradeoff in decision-making circuits. (*A*) Circuit models for decision-making with contraselective and ipsiselective inhibition. The two circuits have the same inhibitory choice tuning but differ in the specificity of inhibitory outputs. (*B,C*) Speed versus accuracy tradeoff. Psychometric functions in *B* show differences in response accuracy between contraselective (dark gray), ipsiselective (light gray), and nonselective (gray) circuits. Choice 2 is the correct choice for positive stimulus coherence. Chronometric functions in *C* show that reaction times are slower in ipsiselective circuits that exhibit higher response accuracy (same colors as in *B*).

provide a measure to identify which inhibitory motifs are present in cortical decision-making circuits.

## Unsupervised Identification of Brain States from Multichannel LFP Recordings

R. Kwapich [in collaboration with S. Chauvette and I. Timofeev, Laval University]

The alternation between sleep and wakefulness is observed in all animals. The main neocortex electrophysiological hallmarks of mammalian sleep are slow-wave sleep (SWS) and rapid eye movement (REM) sleep. These brain states are traditionally defined by applying handcrafted criteria to neurophysiological recordings based on differences in spectral power of local field potentials (LFPs). However, accumulating evidence suggests that the conventional division of global brain states into the wake, SWS, and REM is insufficient. During sleep, neural activity in some brain regions can exhibit awake signatures and vice versa. Multilevel characterization of brain states is needed, which can leverage newly available large-scale longitudinal recordings to capture the spatiotemporal complexity of neural dynamics on a global scale. We developed a computational framework to characterize multidimensional brain-state dynamics from multisite

LFP recordings in mice (Fig. 2). Our framework combines wavelet transform of LFPs with a convolutional auto-encoder (CAE). A CAE performs a nonlinear dimensionality reduction of the data to reveal the low-dimensional features characteristic of mouse brain states. We analyzed continual 14-channel LFP recordings across an entire hemisphere of mouse neocortex and found rich sleep and wake dynamics. First, the latent representation from a single LFP channel revealed three major clusters, which largely agreed with human-expert heuristic labeling of the classical wake, SWS, and REM states. Next, we applied the model to the multichannel LFP data and found a richer structure of brain-state dynamics and relationships among brain regions. Our approach can serve as a unifying framework for identifying the robust components of sleep and wake cycle and their reflection in heterogeneous spatiotemporal activity across the brain.

## Variability and Correlations in Cortical Networks

Y. Shi, C. Aghamohammadi

The neocortex is the most evolved and specialized of all brain structures. Neocortical activity fluctuates endogenously, with much variability shared among
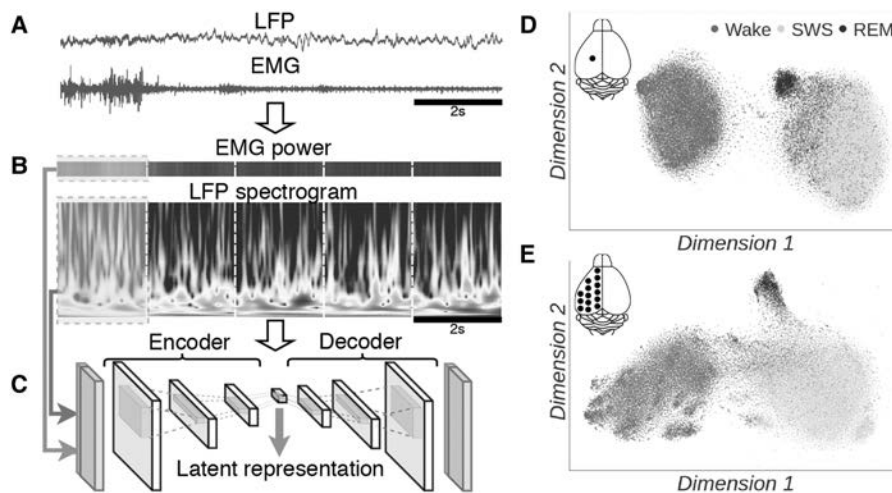


**Figure 2.** Unsupervised identification of brain states from multichannel recordings. (*A*) Raw local field potential (LFP) and electromyography (EMG) signals. (*B*) EMG power and wavelet spectrogram of LFPs provide inputs to the convolutional auto-encoder (CAE). (*C*) CAE architecture. (*D*) Low-dimensional representation of the EMG and LFP activity from a single channel largely agrees with human-expert labeling (color code). Each dot represents a 2-sec segment of the longitudinal data. (*E*) Low-dimensional representation of the EMG and LFP activity from 14 simultaneously recorded channels agrees with human-expert labeling; color code indicates the proportion of channels in wake, slow-wave sleep (SWS), and rapid eye movement (REM) sleep and also exhibits finer subclusters.

neurons. The spatial and temporal structure of these fluctuations impacts sensory processing and behavior. This spatiotemporal structure is defined by the anatomical organization of cortical circuits and external inputs. We analyze spatiotemporal structure of fluctuations in neocortical activity and construct network models to explain how this structure emerges from connectivity and inputs.

In one project led by Y. Shi, we developed a theoretical model to investigate how the spatial and temporal modes of correlations between neurons relate to the network connectivity and to the operating regime of network dynamics that is controlled by external inputs. The theory reveals that because of spatial dependence of interactions in the network, each activity timescale is associated with fluctuations of a particular spatial frequency and makes hierarchical contributions to the correlations. We showed how local versus distributed spatial connectivity shapes the timescales and spatial patterns of neural correlations and how external inputs affect the timescales by changing the network's operating regime. We confirmed model predictions by analyzing how timescales of endogenous activity fluctuations change during spatial attention in the primate visual cortex (data sets from T. Moore, Stanford University, and A. Thiele, Newcastle University).

In another project led by C. Aghamohammadi, we investigated variability of spike generation in individual neurons. Variability of spike generation can be quantified using CV2, the variance to squared mean ratio of interspike intervals. However, estimating CV2 from experimental data is challenging because of nonstationary variations in the underlying firing rates. Through theoretical analysis and simulations, we found that previous methods of estimating CV2 from data suffer from statistical biases that result in misleading interpretation of inferred firing-rate dynamics. We developed an unbiased method for estimating CV2, which explicitly compares spike-count variance across time bins of different sizes while taking the renewal nature of spiking into account. Using this method, we analyzed variability of spike generation in experimental data recorded from the parietal (data set from A. Churchland, CSHL) and premotor cortex of behaving monkeys (data set from K. Shenoy, Stanford University). We discovered that CV2 is a neuron-specific constant independent of the firing rate. Moreover, we found that during decision-making, CV2 is constant across different phases and conditions of the behavioral task. These observations provide strong constraints for microscopic neural network models.

## PUBLICATIONS

Engel TA, Steinmetz NA. 2019. New perspectives on dimensionality and variability from large-scale cortical dynamics. *Curr Opin Neurobiol* **58:** 181–190. doi:10.1016/j.conb.2019.09.003

Genkin M, Engel TA. 2019. Beyond generalization: enhancing accurate interpretation of flexible models. bioRxiv doi:10.1101/808261v1

*In Press*

van Kempen J, Gieselmann MA, Boyd M, Steinmetz NA, Moore T, Engel TA, Thiele A. 2020. Top-down coordination of local cortical state during selective attention. bioRxiv doi:10.1101/808261v

# STRUCTURES AND REAGENTS OF NEURONAL RECEPTORS AND CHANNELS

**H. Furukawa**   E. Chou      N. Simorowski
                 Z. Dawood    J. Syrjanen
                 K. Michalski J.X. Wang

Neurotransmission and neuromodulation are the fundamental currencies for brain development and functions including learning and memory. These processes are largely driven by permeation of ions and substrates across the biological membranes via a number of molecular machines, including neurotransmitter-gated ion channels and large-pore channels. The research in the Furukawa laboratory attempts to establish a molecular basis for cellular signal transductions involved in neurotransmission and neuroplasticity in the mammalian brain, with a scope to develop therapeutic compounds for treatment of neurological diseases and disorders including schizophrenia, depression, stroke, and Alzheimer's disease. To achieve our goals, we conduct structural and functional studies on cell surface receptors and ion channels that regulate intracellular calcium signaling on stimulation by voltage and/or neurotransmitters. These ion channels regulate strength of neurotransmission—the fundamental process for neuronal communication. Dysfunction of the ion channels studied in our group is highly implicated in neurological disorders and diseases noted above. The Furukawa laboratory has been working on a class of ligand-gated ion channel called *N*-methyl-D-aspartate (NMDA) receptors and two families of large-pore channels, calcium homeostasis modulators (CALHMs) and pannexin. The abnormal activation of the above channels is caused by a number of factors, including excessive neurotransmitters and cellular signaling. To unravel the molecular basis for normal and abnormal NMDA receptor activities, we employ structural biology techniques including X-ray crystallography and single-particle cryo-electron microscopy (cryo-EM) and biochemical and biophysical techniques such as electrophysiology. In 2019, we have advanced our understanding and development of reagents that target NMDA receptors in a subtype-specific manner.

## Structural Basis for the Binding of pH-Sensitive Compound 93-31 onto GluN1-GluN2B NMDA Receptors

We obtained crystal structures of the amino-terminal domain (ATD) of the GluN1b-GluN2B NMDA receptors in complex with a series of pH-sensitive compounds, the 93-series, at high resolution ranging from 2.1 Å to 2.5 Å. These compounds were developed by Dr. Liotta and Dr. Traynelis at Emory University. The high-resolution crystal structures allowed us to unambiguously map the compound binding site and identify a new binding pocket that we called a "hydrophobic cage" (Fig. 1), in which the *N*-alkyl motif unique to the 93-series compounds is accommodated. The extent of the van der Waals contacts in this site depends on the orientation and the size of the *N*-alkyl group of the 93-series compounds. Among all the 93-series compounds tested, the *N*-butyl group of 93-31 most closely matches the shape of the hydrophobic cage by aligning in such a way as to form a hydrophobic contact with the side chain of GluN1b Ile133.

To validate the unique binding mode of the 93-series compounds observed in our crystallographic studies and to gain mechanistic insights into the pH sensitivity of 93-31, we conducted site-directed mutagenesis of residues around the binding pocket and measured the ion channel activity by two-electrode voltage-clamp (TEVC) recordings in *Xenopus* oocytes. Toward this end, we obtained concentration–response curves of 93-31 at pH 6.9 and pH 7.6 on the mutant GluN1-4a-GluN2B NMDA receptors. As pH sensitivity appears to be a function of 93-series N-substituent volume, we tested residues in the immediate vicinity of the *N*-alkyl group within the protein hydrophobic cage. As described above, our crystal structures clearly identified GluN1b Ile133 as the major interacting residue with the *N*-alkyl group of the 93-series compounds, and the strength of van
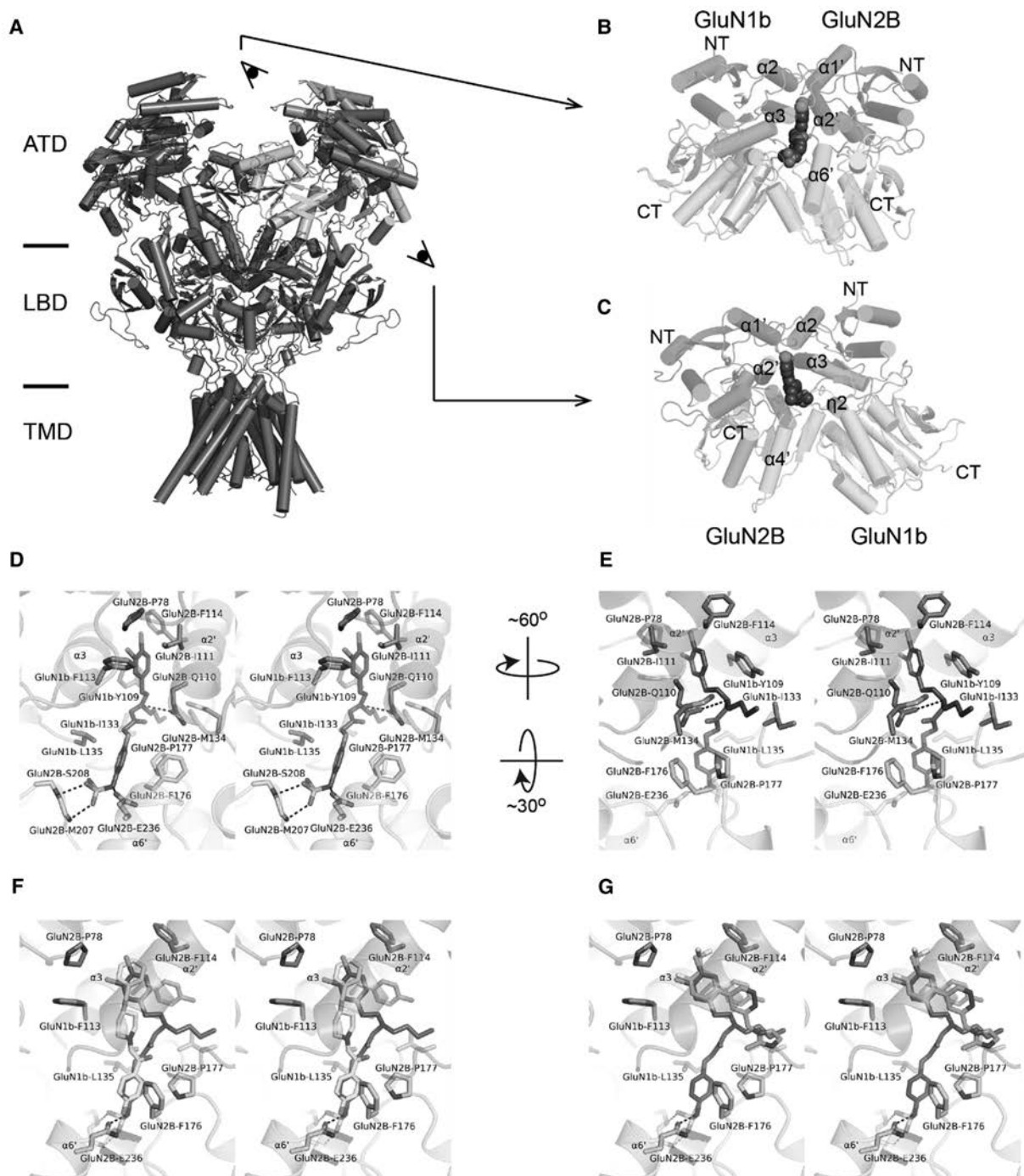
**Figure 1.** Structure of the 93-series binding site. (*A*) The intact heterotetrameric GluN1b/GluN2B NMDA receptor is composed of three structured domains, with the ATD farthest from the cell membrane (PDB code: 6CNA). (*B,C*) The crystal structure of the isolated GluN1b/GluN2B ATD heterodimer bound to the pH-sensitive 93-31 reveals the ligand binding site at the heterodimer interface, as viewed from two angles. (*D,E*) The key residues surrounding the *N*-alkyl chain of 93-31 are primarily hydrophobic, shown here in stereo view. (*F*) Overlay of 93-31 and ifenprodil (yellow, PDB code: 3QEL). (*G*) Overlay of 93-31 and EVT-101 (green, PDB code: 5EWM).

der Waals interactions with GluN1b Ile133 will vary depending on the nature of the *N*-alkyl substituent. The GluN1b Ile133Ala mutation, which decreases the volume of the hydrophobic side chain, consistently reduces both potency and efficacy of inhibition by 93-31 as well as the pH boost (Fig. 2). The placement of the *N*-alkyl substituent is also likely to be stabilized by the side chain of GluN2B(Met134) and GluN2B(Pro177), which are located within 5 Å of the *N*-alkyl group. The GluN2B(Met134Ala) mutation does not affect potency, but increases the current at pH 6.9 (36%) and reduces pH boost (3.2-fold). The GluN2B(Pro177Gly) mutation had a robust effect on potency, efficacy, and pH boost (Fig. 2). Because the
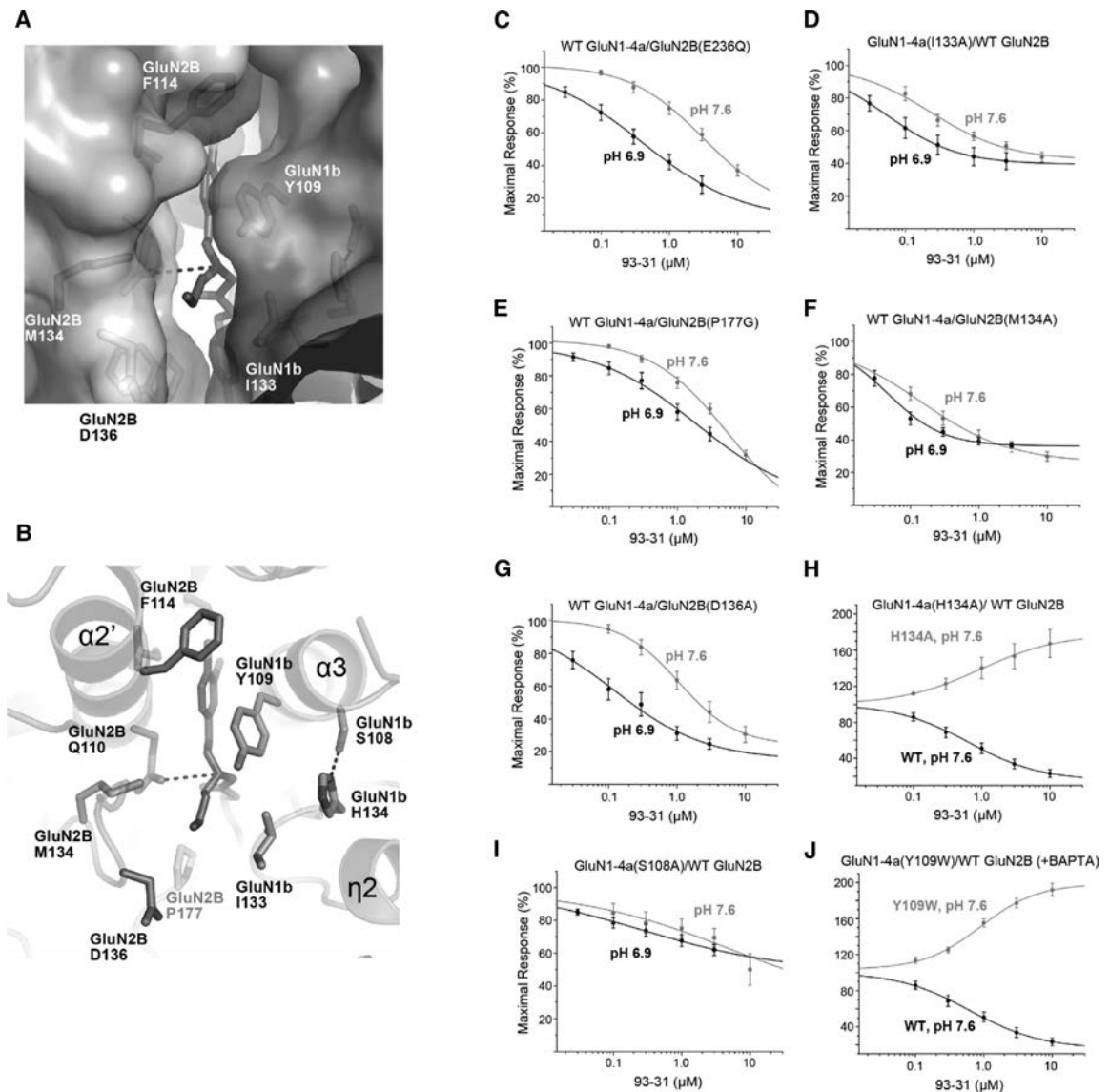


**Figure 2.** TEVC concentration–response curves for GluN1-4a-GluN2B mutants. (*A*,*B*) Key residues at the GluN1/GluN2B ATD dimer interface that interact with the *N*-alkyl chain of 93-31 were mutated and evaluated for effects on pH sensitivity of 93-31. (*C*–*H*) Current responses to maximally effective concentration of glutamate and glycine (100 μM glutamate, 30 μM glycine) are shown in the presence of varying concentrations of 93-31 as a proportion of maximal response. (*I*–*J*) Two mutations, GluN1-4a(His134Ala) and GluN1-4a(Tyr109Trp), convert 93-31 into a potentiator. Error bars represent mean ± S.E.M.

chemistry of pH sensitivity typically involves changes in protonation states associated with altered solution pH, we speculated that the hydrophobic interaction involving the *N*-alkyl group of the 93-series compounds is controlled by the protonation of residues near the hydrophobic pocket. One such residue that caught our attention was GluN1b(His134), which is positioned to form a hydrophobic interaction with GluN1b(Ile133) when it is not protonated. When protonated at low pH, GluN1b(His134) has a considerably weaker hydrophobic interaction with GluN1b(Ile133), which in turn allows GluN1b(Ile133) to form stronger interactions with the *N*-alkyl group of the 93-series compounds and serve as a local pH sensor. GluN1b(His134) is located at the exit of the GluN1b-GluN2B subunit interface and is solvent accessible, allowing it to sense solution pH. The pKa of His is 6 in free solution and likely different in the context of adjacent residues. Thus, the His134 pKa could be tuned to show a maximal change in the ionization state with changes in extracellular pH. For example, a pKa of 6.4 would yield a fourfold change and a pKa of 6.8 would yield a threefold change in the ionization state of His134 with a change in extracellular pH from 7.6 to 6.9. Further supporting the important role of the protonation state of GluN1(His134) is the observation that breakage of the GluN1(His134)-GluN1(Ser108) interaction by mutating GluN1(Ser108Ala) robustly reduces 93-31 potency, efficacy, and pH boost. Situated in the middle of these pH-sensitive elements is GluN1(Tyr109), proximal to the backbone and the *N*-alkyl group of the 93-series compounds. The GluN1(Tyr109Ala) mutation to remove the bulky side chain mildly lowers potency and efficacy of inhibition, but not pH boost.

## Targeting GluN1-GluN2C and GluN1-GluN2D NMDA Receptors

Subunit diversity is a hallmark of NMDA receptors (NMDARs) Undefined and can be potentially exploited to target specific diseases. Different combinations of GluN1, GluN2, and GluN3 subunits give rise to specific di- and triheteromeric NMDAR subtypes with a wide spectrum of electrophysiological and pharmacological properties. Extensive research has shown distinct spatiotemporal distribution of NMDAR subtypes in the brain, implying unique roles of different NMDAR subtypes in specific aspects of brain development and functions, and suggesting therapeutic potential for subtype-specific targeting of NMDARs. Thus, development of highly subtype-specific reagents will advance our understanding of the biological roles of NMDAR subtypes in brain function and development and may provide possible treatments for the aforementioned diseases and disorders. Although GluN2A/2B-containing NMDARs are dominant subtypes that are expressed in the adult brain, the expression of GluN2C/2D-containing NMDARs is restricted to discrete regions critical for diseases. For example, in schizophrenia, recent evidence has pointed to critical involvement of NMDAR hypofunction in cortical GABAergic neurons where GluN2D subunits are highly expressed. In Parkinson's disease, in which overfiring of subthalamic nucleus (STN) neurons occurs as a result of loss of dopaminergic neurons in the substantia nigra pars compacta (SNc), GluN2D-containing NMDARs may be the relevant target because they are present and mediate synaptic neurotransmission in the STN.

One of the critical limitations in studying GluN2C- and GluN2D-containing NMDARs in both pre- and postsynaptic processes has been the lack of highly potent and subtype-specific agonists and antagonists. This is in stark contrast to the GluN2B- and GluN2A-containing NMDARs, where highly subtype-specific compounds, ifenprodil and TCN-201, respectively, are available. Still, compounds with twofold to 10-fold GluN2C and GluN2D selectivity over GluN2A/2B, such as PPDA (($2S^*,3R^*$)-1-(phenanthrene-2-carbonyl) piperazine-2,3-dicarboxylic acid) and its derivatives UBP141–($2R^*,3S^*$)-(1-(phenanthrene-3-carbonyl) piperazine-2,3-dicarboxylic acid)—and UBP145–($2R^*,3S^*$)-1-(9-bromophenanthrene-3-carbonyl)piperazine-2,3-dicarboxylic acid–were frequently used to show that presynaptic GluN2D-containing receptors contribute to the major component of short-term potentiation and spike timing–dependent long-term depression, and mediate synaptic currents in the juvenile hippocampus. Although there has been much improvement in more GluN2C- and/or GluN2D-specific compounds in recent years, all of these allosteric compounds have shown IC$_{50}$ values in the high nanomolar to micromolar range and still lack well-defined binding sites on the NMDAR, limiting the prospects for rational compound optimization. In this study, we present the new PPDA derivative UBP791

(2*S*\*,3*R*\*)-(1-(7-(2-carboxyethyl)phenanthrene-2-carbonyl)piperazine-2,3-dicarboxylic acid), which showed 47-fold and 16-fold preference of GluN2C/2D over GluN2A- and GluN2B-containing NMDARs, respectively, and we use UBP791 to study the key molecular determinants within GluN2D that confer GluN2C/GluN2D-selective compound binding. Through X-ray crystallography and electrophysiology, we determined that a combination of a methionine and a lysine unique to GluN2C/2D (rat GluN2D-Met763/Lys766, GluN2C-Met736/Lys739) confers subtype-selective binding of UBP791. Rationally modifying UBP791 then led to a greatly improved compound, UBP1700 ((2*S*\*,3*R*\*)-1-(7-(2-carboxyvinyl)phenanthrene-2-carbonyl)piperazine-2,3-dicarboxylic acid), which had 63-fold and 52-fold selectivity for GluN2D (50- and 40-fold for GluN2C) over GluN2A and GluN2B, and still showed high potency with $K_i$ values in the low nanomolar range. Hence, our study demonstrates that despite high conservation of the GluN2 ligand binding domains (LBDs) (with sequence identities of 69%-82% in rat LBDs), particularly at the L-glutamate binding pocket, understanding the exact binding mode of a compound like UBP791 allows exploration of the LBD as a potential subtype-specific target.

To understand the exact binding mode of UBP791 and the molecular basis for the GluN2C/2D subtype selectivity and to identify unexplored possibilities for compound development, we next sought to obtain structures of the GluN2A LBD and GluN2D LBD complexed to UBP791. Although crystal structures of GluN2D LBD in complex with agonists and partial agonists have been successfully obtained, an antagonist-bound GluN2D LBD structure has been technically difficult to capture. The only competitive antagonist-bound crystal structures of NMDAR GluN2 LBDs to date are the ones complexed to GluN1-GluN2A LBDs, which have been obtained by soaking glycine- and L-glutamate-bound GluN1-GluN2A LBD crystals against the crystallization buffer containing glycine and GluN2-antagonists to substitute L-glutamate with the antagonists within the GluN2A LBD (see Methods). Similarly soaking the agonist-bound GluN2D LBD crystals or cocrystallization with antagonists did not result in antagonist-bound GluN2D LBD structures. We therefore attempted to use GluN2A-4m LBD as a structural mimic of GluN2D LBD and a tool to capture antagonist binding by GluN2D.

Here, we found that GluN2A-4m LBD protein can be recombinantly expressed, purified, and cocrystallized with GluN1 LBD protein as in the case of GluN1-GluN2A LBD heterodimers (Fig. 3).

Structural comparison between the GluN1-GluN2A LBD and the GluN1-GluN2A-4m LBD complexed to UBP791 provided important insights into preferential binding of UBP791 to GluN2D over GluN2A. In both crystal structures, only the (2*S*,3*R*)-enantiomer of UBP791 was observed in the inter-D1-D2 domain cleft, consistent with the previous finding that (2*S*,3*R*)-PPDA has approximately 10-fold higher potency compared to (2*R*,3*S*)-PPDA. The binding modes of the piperazine and phenanthrene moieties are identical to those observed in the structure of the GluN1-GluN2A LBD complexed to PPDA. That is, the piperazine ring interacts via polar interactions involving residues GluN2A/GluN2A-4m-Thr513, Arg518, Ser511, and Ser689, whereas the phenanthrene ring participates in a van der Waals interaction with GluN2A/GluN2A-4m-Tyr730, Val734, Tyr761, Val713, Phe416, Tyr737, and Ala414 in GluN2A (Arg414 in GluN2A-4m) (Fig. 3B,E). In contrast, we observed the major differences in the protein-ligand binding mode between the carboxyethyl group of UBP791 and the GluN2A LBD or the GluN2A-4m LBD (Fig. 3C,F).

In the GluN1-GluN2A-4m LBD, several elements favor accommodation of the carboxyethyl group. The most notable is the specific polar interaction with the amino group of GluN2A-4m-Lys741 (Lys766 in GluN2D) (Fig. 3E,F). This interaction is made possible by the hydrophobic interaction of GluN2A-4m-Lys741 with GluN2A-4m-Met738 (Met763 in 2D), which orients the GluN2A-4m-Lys741 side chain towards UBP791. GluN2A-4m-Met738 is positioned to form sulfur-aromatic interactions with the phenanthrene moiety of UBP791 as well as GluN2A-4m-Tyr737 to stabilize the binding pocket. Furthermore, GluN2A-4m-Met738 and the methylene group closest to the phenanthrene ring of UBP791 may form hydrophobic interactions. The other mutated residues GluN2A-4m-Arg414 and –Arg740 are not further involved in binding of UBP791.

The equivalent residues to GluN2A-4m-Met738 and GluN2A-4m-Lys741 in the GluN1-GluN2A LBD are GluN2A-Lys738 and GluN2A-Arg741, which are not involved in binding of UBP791. The largest difference here is that the GluN2A-Arg741 side chain faces
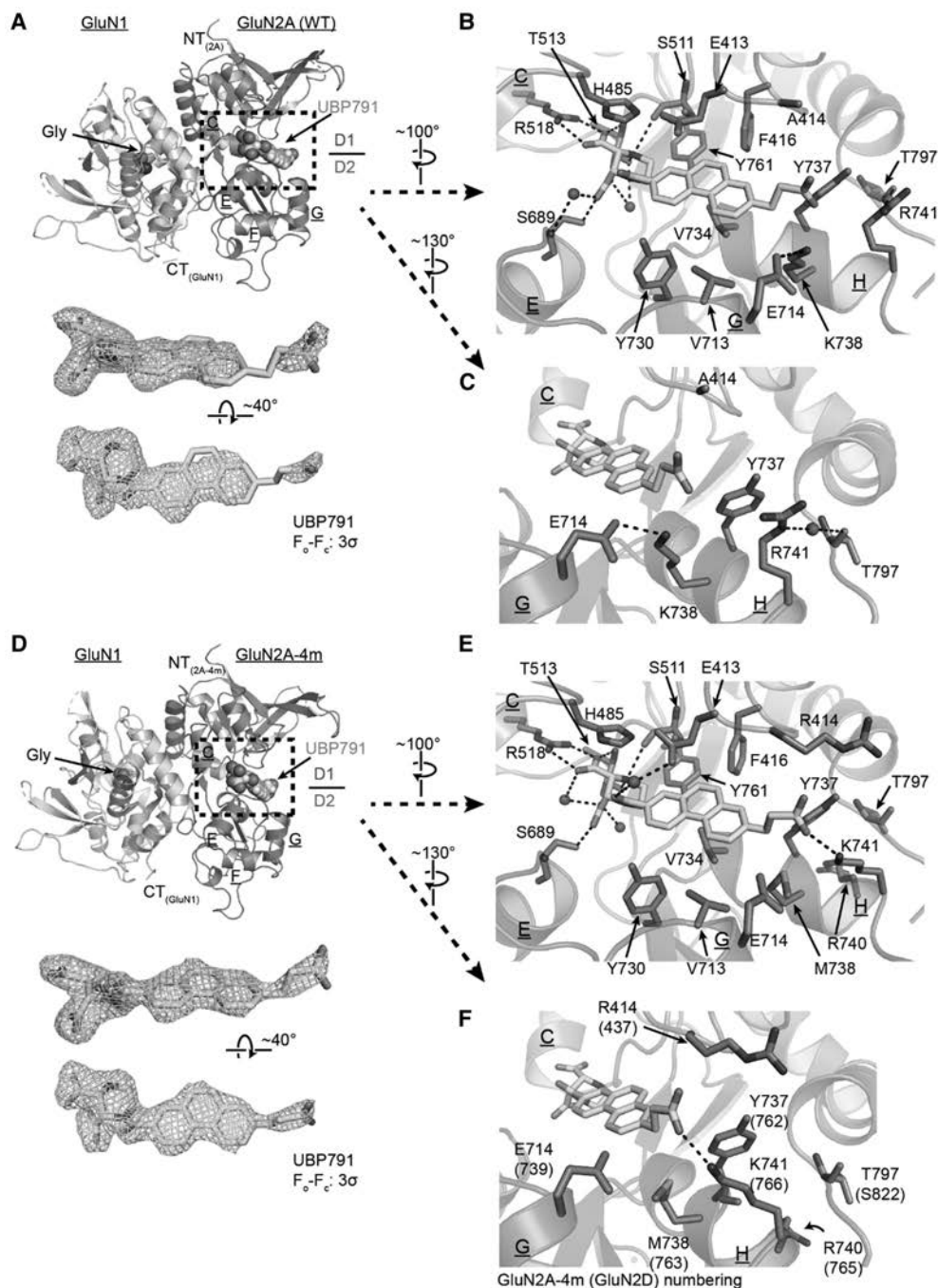
**Figure 3.** Structures of glycine/UBP791 complexed to GluN1-GluN2A (WT) and GluN1-GluN2A-4m LBD. (*A*) Overall structure of the GluN1-GluN2A LBD complexed to glycine and UBP791 (spheres). Shown in mesh below is the $F_o$-$F_c$ omit map of UBP791 contoured at 3σ. (*B,C*) The binding site of UBP791 showing polar (dashed lines) and hydrophobic interactions. GluN2A-Lys738 and -Glu714 form a hydrogen bond, whereas GluN2A-Arg741 and –Thr797 form a water-mediated hydrogen bond. (*D*) Overall structure of the GluN1-GluN2A-4m LBD complexed to glycine and UBP791 (spheres). Note that the $F_o$-$F_c$ omit map of UBP791 contoured at 3σ (mesh) here is more ordered and continuous compared to that in the GluN1-GluN2A LBD in *A*. (*E*) The binding site of UBP791 (sticks) showing similar polar (dashed lines) and hydrophobic interactions with the piperazine and phenanthrene moieties to those in the GluN1-GluN2A LBD in *A*. (*F*) In contrast to the GluN1-GluN2A LBD, GluN2A-4m-Met738 forms sulfur-aromatic interactions with the ligand and Tyr737, while GluN2A-4m-Lys741 forms a hydrogen bond with the carboxyethyl group of UBP791.
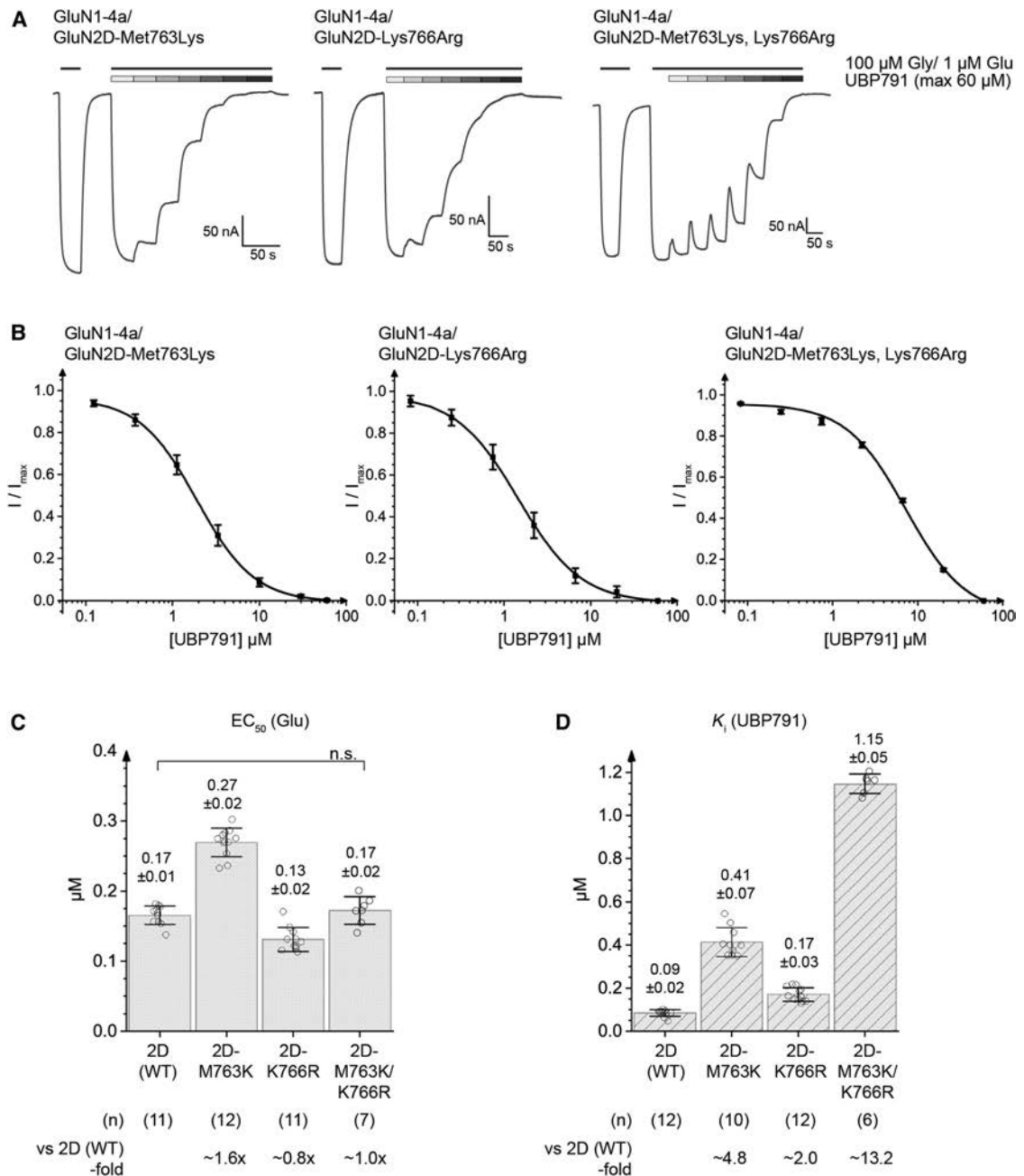
**Figure 4.** Effect of mutations in GluN2D on L-glutamate and UBP791 sensitivity. (*A*) Representative TEVC dose–response traces of single mutant GluN1-4a/GluN2D Met763Lys (*left* panel), Lys766Arg (*middle* panel), or double mutant Met763Lys/Lys766Arg (*right* panel) NMDARs held at –60 mV. Currents were evoked by application of 100 µM glycine and 1 µM L-glutamate and inhibited by varying concentrations of UBP791 (Met763Lys: concentration increments: 0.12/0.37/1.1/3.3/10/30/60 µM; for Lys766Arg and double mutant: threefold increments from 0.08–60 µM). (*B*) Averaged dose–response curves (mean ± s.d.) for inhibition with UBP791 from eight, twelve, and six recordings for GluN1-4a/GluN2D Met763Lys, GluN1-4a/GluN2D Lys766Arg, and GluN1-4a/GluN2D Met763Lys/Lys766Arg, respectively, fit with the Hill equation to calculate $IC_{50}$ and Hill coefficient ($n_H$). (*C*) $EC_{50}$ for L-glutamate and (*D*) $K_i$ for UBP791 for the mutants were obtained by TEVC recordings as in Figure 2. Single data points are shown as open circles, the bar graph represents the mean with error bars for s.d., the number of recordings (*n*) and the fold-difference to $EC_{50}$ and $K_i$ of GluN2D (WT) are as shown. Pairwise comparison shows WT and mutants have different potencies ($p < 0.05$ with two-tail *t*-test and Bonferroni correction) except where stated (n.s.).

away from the binding pocket, which is likely facilitated by charge repulsion between the amino group of GluN2A-Lys738 and the guanidinium group of GluN2A-Arg741. GluN2A-Arg741 instead forms water-mediated hydrogen bonds with GluN2A-Thr797 (Fig. 3C). GluN2A-Lys738 is also not ideally positioned to interact with the carboxyethyl group and instead forms a polar interaction with GluN2A-Glu714 from Helix G (Fig. 3C). The unfavorable binding of UBP791 in the GluN2A binding cleft is reflected by the discontinuous and disordered electron density of the carboxyethyl group (Fig. 3A), which is in stark contrast to the density observed in GluN2A-4m (Fig. 3D). In summary, structural comparison between the GluN2A LBD and GluN2A-4m LBD (our GluN2D mimic for this study) implied that the key molecular determinants for preferential binding of UBP791 to GluN2D over GluN2A lie in the 738 and 741 positions (numbering in GluN2A), where they are lysine and arginine in GluN2A and methionine and lysine in GluN2D. Together, the methionine and lysine residues in GluN2D favorably accommodate the carboxyethyl group of UBP791 by forming both polar and hydrophobic interactions. The methionine and lysine residues are also conserved at the equivalent positions of GluN2C; thus, GluN2C specificity is also mediated via a similar mechanism.

To test the validity of the structural observation for the critical involvement of the methionine/lysine residue combination (GluN2D-Met763/Lys766, and GluN2A-4m-Met738/Lys741) in preferential binding of UBP791 to GluN2D over GluN2A, we conducted site-directed mutagenesis and assessed inhibition potencies of the mutant channels by TEVC. Specifically, we converted the GluN2D residues to the equivalent ones in the GluN2A subunit and measured macroscopic currents of the GluN1-4a/GluN2D mutant NMDARs. We first tested the single point mutants GluN2D-Met763Lys and GluN2D-Lys766Arg, which showed approximately fivefold and twofold increases in $K_i$ compared to the wild-type (WT) GluN2D,

respectively (Fig. 4). The mutant GluN2A-Lys738Met (the reverse mutant of GluN2D-Met763Lys) was previously shown to increase PPDA potency by 5-fold compared to the WT GluN2A. Thus, our present result on GluN2D-Met763Lys strongly supported the interaction between GluN2D-Met763 and the phenanthrene backbone contained in both PPDA and UBP791. The modest change in the $K_i$ value of GluN2D-Lys766Arg may be attributed to the possibility that, in the absence of potential charge repulsion as seen in the UBP791-bound GluN2A WT LBD structure, the arginine side chain could still orient itself to form some interaction with the carboxyethyl group; hence, we next tested the double mutant GluN2D-Met763Lys/Lys766Arg. In line with our structural observations, this double mutant lowered UBP791 potency by 13-fold compared to the WT GluN2D, demonstrating that GluN2D-Met763 and -Lys766 synergistically contribute to subtype-selective UBP791 binding (Fig. 4D).

## PUBLICATIONS

Regan MC, Zhu Z, Yuan H, Myers SJ, Menaldino DS, Tahirovic YA, Liotta DC, Traynelis SF, Furukawa H. 2019. Structural elements of a pH-sensitive inhibitor binding site in NMDA receptors. *Nat Commun* **10:** 321. doi:10.1038/s41467-019-08291-1

Wang JX, Furukawa H. 2019. Dissecting diverse functions of NMDA receptors by structural biology. *Curr Opin Struct Biol* **54:** 34–42. doi:10.1016/j.sbi.2018.12.009

*In Press*

Michalski K, Syrjanen JL, Henze E, Kumpf J, Furukawa H, Kawate T. 2020. The Cryo-EM structure of pannexin 1 reveals unique motifs for ion selection and inhibition. *eLife* **9:** e54670. doi:10.7554/eLife.54670

Syrjanen JL, Michalski K, Chou TH, Grant T, Rao S, Simorowski N, Tucker SJ, Grigorieff N, Furukawa H. 2020. Structure and assembly of calcium homeostasis modulator proteins. *Nat Struct Mol Biol* **27:** 150–159. doi:10.1038/s41594-019-0369-9

Wang JX, Irvine MW, Burnell ES, Sapkota K, Thatcher RJ, Li M, Simorowski N, Volianskis A, Collingridge GL, Monaghan DT, et al. 2020. Structural basis of subtype-selective competitive antagonism for GluN2C/2D-containing NMDA receptors. *Nat Commun* **11:** 423. doi:10.1038/s41467-020-14321-0

# ASSEMBLY AND ORGANIZATION OF CORTICAL CIRCUITS UNDERLYING MOTOR CONTROL

**J. Huang**  X. An  J. Hatfield  H. Kondo  Y. Qian
S. Fredericks  D. Huilgol  Yi. Li  R. Raudales
S.W. Galbavy  K. Kelly  K. Matho  S. Srivas
E. Gamache  G. Kim  H. Mohan  B. Wang

Our research program aims to understand the general principles underlying the neural circuit organization of the cerebral cortex that enables cognitive function. The neocortex is the seat of human mental prowess as well as the origin of devastating neuropsychiatric disorders. An enduring challenge in neuroscience is to decipher how mental activities, from sensorimotor control to cognitive processing, emerge from the cellular constituents of the cortex that assemble progressively higher-level functional architectures. This knowledge is key to a biological explanation of the mind and is necessary to guide the diagnosis and treatment of brain disorders.

Despite its daunting cellular and wiring complexity, the fundamental organization and construction plans of the elementary cortical architecture have been conserved since its humble origin in small mammals. Indeed, the basic scaffold of major output channels, intracortical processing networks, and local circuit templates can be readily recognized across mammalian species and are nearly identical among individuals of the same species. Our overarching hypothesis is that these conserved features of cortical organization are implemented by a large set of "cardinal neuron types," which are reliably generated through developmental programs rooted in the genome. Grounded on these evolutionary and genetic principles, the overarching theme of our research is a genetic dissection of cortical circuits in the mouse through systematic targeting of its diverse basic elements, the neuronal cell types.

Over the past decade, we have pioneered multiple rounds of genetic targeting of GABAergic inhibitory interneurons and glutamatergic pyramidal neurons. These tools establish reliable experimental access to neuronal subpopulations and have transformed the study of cortical circuits. In parallel we have made a fundamental discovery on the seemingly intuitive yet surprisingly elusive notion of "neuron types": We found that neuron type identity can be delineated by their synaptic communication styles that are rooted in key transcriptional signatures, an emerging conceptual framework that will guide cell type discovery and classification. With our recent progress on genetic tools for projection neuron types that mediate cortical processing streams and output channels, we began exploring the cell type basis of cortical circuit function in the context of sensorimotor control. A unique strength of our research program is leveraging experimentally accessible neuron types as a solid middle ground to navigate across levels of organization: from exploring the molecular and developmental genetic basis of cell types on the one hand to their roles in circuit function and behavior on the other.

## Genetic Dissection of Glutamatergic Neuron Subpopulations and Developmental Trajectories in the Cerebral Cortex

Over the past 15 years, we have made sustained progress toward a systematic genetic targeting of cortical neuron types (Fig. 1). We designed multiple combinatorial strategies that engage molecular markers, cell lineage program, and anatomy to target increasingly more specific GABAergic interneurons. These genetic tools (more than four dozen driver lines) have transformed the study of GABAergic circuits. More recently, we have extended this effort to glutamatergic neurons. Diverse types of glutamatergic pyramidal neurons (PyNs) mediate the myriad processing streams and output channels of the cerebral cortex, yet all derive from neural progenitors of the embryonic dorsal telencephalon. Here, we establish genetic strategies and tools for dissecting and fate mapping PyN subpopulations based on their developmental and molecular programs. We leverage key transcription factors and effector genes to systematically target the temporal patterning programs in progenitors
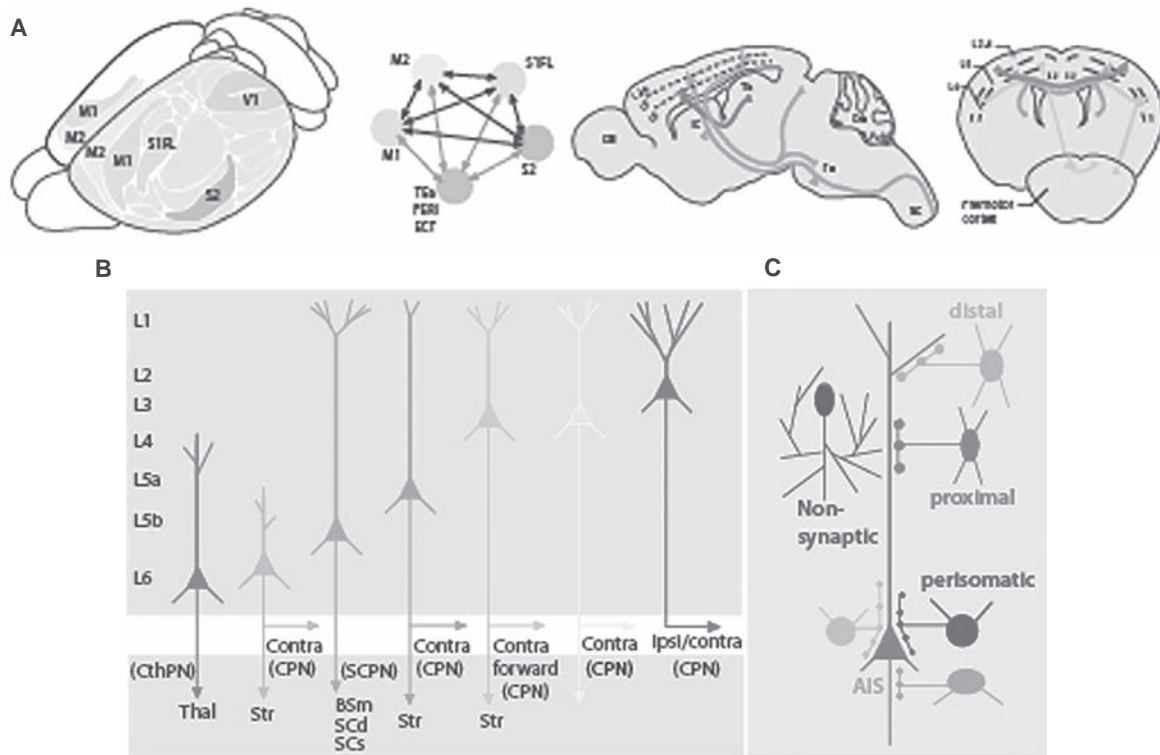
**Figure 1.** (*A*) Cortical processing networks and output channels mediated by PyNs. (*B*) Several major types of PyNs distinguished by laminar position and axon projection. (*C*) Several cardinal types of GABAergic neurons that target different subcellular compartments of PyNs.

and differentiation programs in postmitotic neurons. We generated more than a dozen temporally inducible mouse Cre and Flp knock-in driver lines to enable combinatorial targeting of major progenitor types and projection classes. Intersectional converter lines confer viral access to specific subsets defined by developmental origin, marker expression, anatomical location, and projection targets. These strategies establish an experimental framework for multimodal characterization of PyN subpopulations and tracking their developmental trajectories toward elucidating the organization and assembly of cortical processing networks and output channels. We are leading a Center for Mouse Brain Cell Atlas in the BRAIN Initiative Cell Census Network (BICCN). We continue to systematically generate cell type tools targeting cortical, striatal, and thalamic projection neurons. These cell type tools will greatly accelerate studying the cortical–striatal–thalamic system, the most prominent network in the mammalian brain, which mediates a wide range of sensory, motor, emotional, and cognitive functions.

## Assembly and Function of a Chandelier Cell–Pyramidal Cell Microcircuit Module

Unlike the retinal circuits, which consist of readily identifiable connectivity modules (e.g., retinal mosaics), the complexity of cortical networks often precludes the recognition of simpler modules. The chandelier cell (ChC)–PyN connectivity may represent a rare exception (Fig. 2). ChCs are the most distinctive GABAergic interneurons that specifically innervate PyNs at their axon initial segment (AIS) and likely control spike initiation. Thus, ChC is likely a key entry point to explore the organization of cortical microcircuits as well as global networks through discovering their dynamic control of PyN functional ensembles. Our genetic targeting of ChCs established a powerful experimental

system to integrate development and functional studies of cortical circuits. First, through genetic fate mapping with the Nkx2.1-CreER driver, we discovered the lineage origin of ChCs (Fig. 2A). Specified at cell birth, young ChCs appear endowed with cell-intrinsic programs that guide their long-distance migration and laminar deployment toward innervating PyNs at AIS. Second, we made the surprising discovery that ChC density at the border between primary and secondary visual cortex is regulated by a massive cell death process driven by contralateral callosal inputs and spontaneous retinal activity shortly before eye opening. This activity-dependent ChC elimination process likely reflects activity-regulated wiring of ChCs into cortical circuitry and is necessary for the proper development of binocular vision. Third, combining genetic labeling and high-resolution large-volume imaging, we found that ChCs consist of multiple fine-grained subtypes likely delineated by their input–output connectivity (e.g., to subsets of PyNs defined by projection targets; Fig. 2B). Fourth, we have discovered that a subset of layer 2 ChCs mediates highly specific and directional

connectivity between two PyN ensembles and cortical subnetworks (Fig. 2C).

Based on novel markers revealed by scRNA-seq (Unc5b, Pthlh), we have generated new intersectional driver lines that specifically and robustly target most if not all ChCs (Fig. 2D). These genetic tools enable a multifaceted and integrated study of ChC connectivity and function in the context of behavior.

## Functional Organization of Cortical Circuits

With increasingly precise and comprehensive genetic tools for GABA interneurons and GLU PyN, we began to explore the functional organization of cortical circuitry, initially in the context of motor control. Among brain functions ranging from perception to cognition to action, the deployment of adaptive and complex movement stands out as the only one through which animals impact the world. Although much progress has been made in understanding the spinal sensorimotor circuits controlling simpler movements
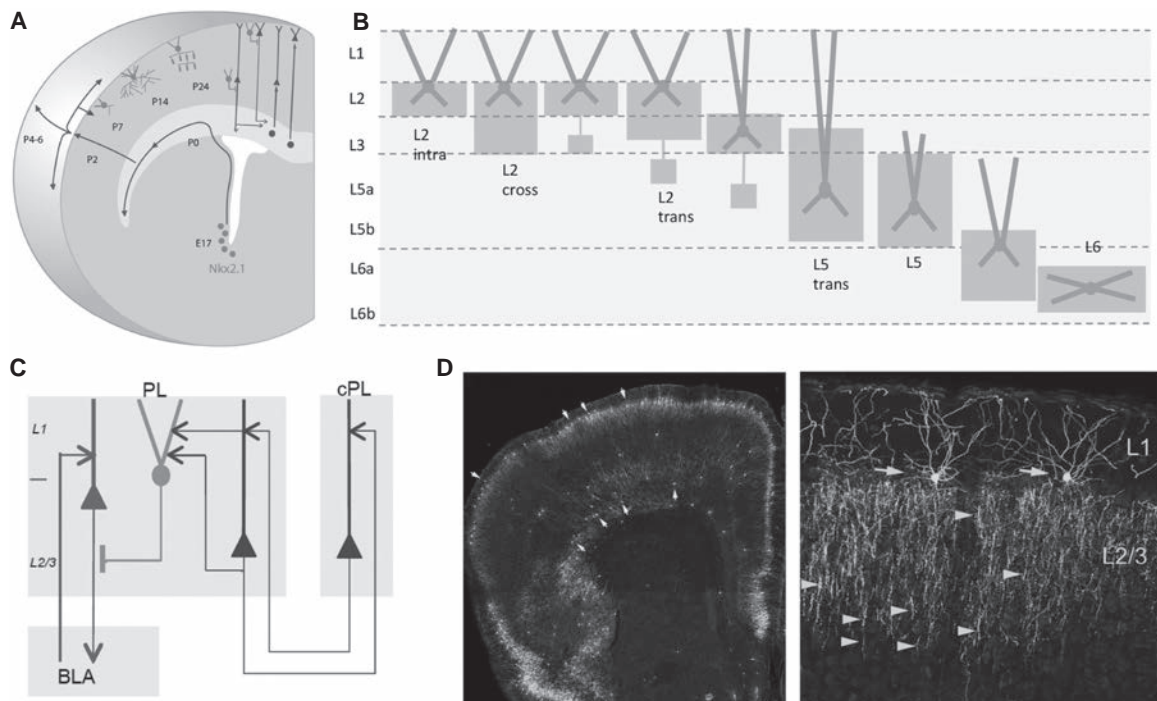


**Figure 2.** The ChC-PyN module. (*A*) Genetic fate mapping reveals the developmental trajectory of ChCs. (*B*) Single-neuron anatomy reveals multiple fine-grained ChC subtypes likely distinguished by input–output connectivity. (*C*) L2 ChCs mediate directional inhibitory control between PyN ensembles and cortical subnetworks. (*D*) The Unc5b-CreER captures highly specific ChC subsets.
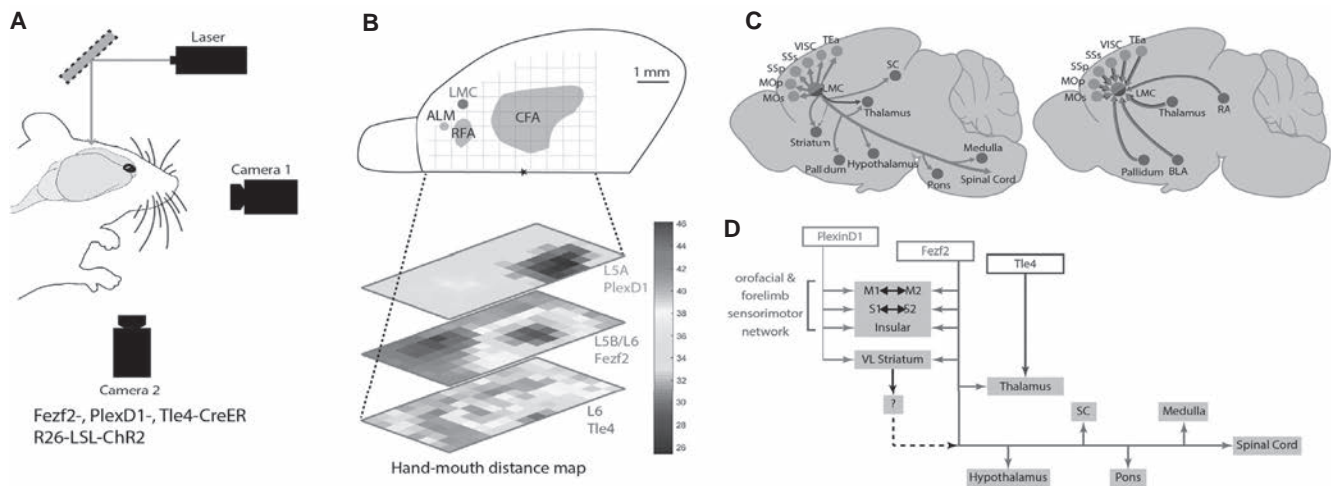
**Figure 3.** (*A*) Optogenetic activation screen of PyN types in forelimb and orofacial movements. (*B*) PlexD1, Fezf2, and Tle4 PyNs construct different set of cortical "motor maps." The rostral forelimb-orofacial area (RFO) is indicated in relation to other known motor areas. (*C*) The output–input connectivity patterns of PlexD1, Fezf2, and Tle4 PyNs in lateral motor cortex (LMC) revealed by anterograde and retrograde viral tracing. (*D*) Summary schematic showing that the three PyN types in RFO form a cortico–striatal–thalamic subnetwork that emphasizes the forelimb-orofacial sensorimotor system, with subcortical output mainly from Fezf2 neurons.

(e.g., locomotion), the supraspinal neural mechanisms that orchestrate volitional and skilled movements are far from clear. Decades of behavioral, anatomical, and physiological studies in primates have implicated the "motor cortex" in various aspects of forelimb motor control, but a framework of motor cortex organization has yet to emerge, and the underlying neural circuits remain largely unexplored. Rodents display a rich repertoire of skilled forelimb movements from reach and grasp to manipulate, which likely have derived from and elaborated through food-handling behaviors adapted to their ethological niche. This capacity for coordinated and dexterous sensorimotor control in rodents provides the opportunity to explore the underlying neural circuit mechanisms, including those in the cerebral cortex.

Using our multiple PyN driver lines, we carried out a systematic optogenetic screen of PyN projection types and cortical areas that induce forelimb and orofacial movements in head-fixed behaving mice (Fig. 3). We discovered a rostral forelimb orofacial area (RFO), where activation of Fezf2 PyNs (PyN$^{Fezf2}$, pyramidal track-PT) and PlexinD1 PyNs (PyN$^{PlexD1}$, cortico-striatal/intratelencephalic-IT) induces highly coordinated forelimb and orofacial movements resembling feeding. Anterograde and retrograde tracing from these PyN types in RFO reveals a highly connected cortical network involving primary and secondary sensory and motor areas of the forelimb and orofacial regions; this cortical network is embedded in the cortico–striatal–thalamic system, with outputs to numerous subcortical targets from midbrain to hypothalamus to pons and spinal cord. Wide-field GCaMP6 imaging during a head-fixed feeding behavior revealed that PyN$^{Fezf2}$ and PyN$^{PlexD1}$ activity patterns in the RFO closely correlate with food handling and manipulation. We have developed a "mouse restaurant" behavior paradigm in which free-moving mice feed on automatically delivered food items while being video-recorded for behavior tracking and analysis. Inactivation of RFO PyNs impairs hand maneuvering and hand–mouth coordination during handling and manipulation without disrupting biting, eating, and grasping per se. These results uncover a specific cortical area and distinct neuron types that orchestrate object handling and manipulation, providing one of the most compelling evidences for distinct roles of molecular and anatomically defined cortical neuron types in dexterous sensorimotor control.

In a more general context of exploring the functional architecture of cortical processing networks, fMRI of metabolic activities in large brains has been the dominant approach but is severely limited in spatiotemporal resolution and in inferring relationship to neuronal activity. Our PyN type-based wide-field Ca²⁺ imaging in mice enables, for the first time, dorsal cortex–wide, real-time (without averaging), and cell

type resolution monitoring of neural activation patterns in behaving animals. We have made the surprising observation that IT-PyNs$^{PlexD1}$ and PT-PyNs$^{Fezf2}$ often display distinct, in addition to overlapping, spatiotemporal activation patterns during a range of sensorimotor behaviors. This finding questions the long-held view of a strictly columnar information flow, which predicts correlated activation of PyNs$^{PlexD1}$ and PyNs$^{Fezf2}$. Instead, our finding suggests that PyN types involved in intracortical processing (PyNs$^{PlexD1}$) and cortical output (PyNs$^{Fezf2}$) may engage in distinct spatiotemporal subnetworks, and information flow from PyNs$^{PlexD1}$ to PyNs$^{Fezf2}$ is likely dynamically regulated. Our approach thus may provide novel insight into the functional architecture of cortical processing networks with cell-type resolution.

## Tracking the Developmental Trajectory of PyNs from Lineage Origin to Circuit Function

A fundamental question in neuroscience is How do diverse functional neuron types (e.g., grid cells, border cells) emerge from developmental genetic programs and activity-regulated processes that all initiate from neural progenitors? Addressing this question requires being able to track the trajectory of functionally defined neuron types from their lineage progression to their circuit operations. We have established a set of cell-type resolution genetic fate mapping tools that enable tracking the developmental trajectories of PyNs. We have made two major discoveries on the developmental specification of PyN types, with relevance to circuit function and implications in cortical evolution.

All PyNs are generated from radial glial progenitors (RGs) either directly or indirectly through intermediate progenitors (IPs) (Fig. 4). Whereas direct neurogenesis from RGs is a universal mechanism throughout the neural tube for the entire nervous system, indirect neurogenesis through IPs is restricted to the telencephalon that gives rise to the forebrain—especially the neocortex. During vertebrate evolution, whereas RG-direct neurogenesis is conserved across species, IP-indirect neurogenesis only becomes prominent in mammals and continues to expand in primates and humans. It is widely assumed that indirect neurogenesis is a mechanism to amplify neuronal production and contribute to the expansion of cortical neuron numbers. However,



| | Cre-Driver | RGC-Direct | IPC-Indirect |
|---|---|---|---|
| All | Emx1 | 21.80% | 78.20% |
| PT | Fezf2 | 50% | 50% |
| IT | PlexinD1 | 0% | 100% |
| | Satb2 | Ongoing | Ongoing |
| CT | Tle4 | 100% | 0% |
| | Tbr1 | Ongoing | Ongoing |
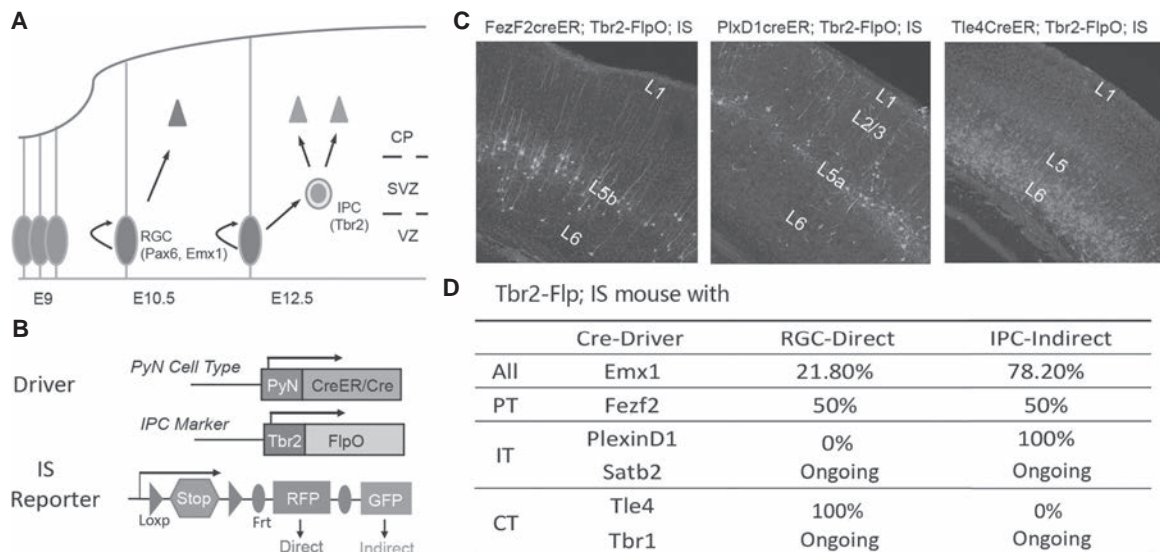
**Figure 4.** (*A*) A schematic of direct neurogenesis from radial glial cells (RGC) and indirect neurogenesis from intermediate progenitors (IPC). Tbr2 is specific to IPC. (VZ) Ventricular zone, (SVZ) subventricular zone, (CP) cortical plate. (*B*) Genetic strategies for differential labeling and manipulation of directly and indirectly generated PyNs. (*C*) Roughly equal ratio of Fezf2 PyNs derive from direct (red) and indirect (green) neurogenesis (*left*); all PlexD1 PyNs derive from the indirect pathway (*middle*); and all Tle4 PyNs derive from the direct pathway (*right*). (IS) Intersection/subtraction reporter. (*D*) Summary of current and ongoing studies on the neurogenic origins of different classes of PyNs.

whether IPs may contribute to the diversification of PyN types is unknown. Using a Tbr2-Flp driver targeting all IPs and intersection-subtraction strategy to distinguish RG- and IP-derived PyNs (Fig. 4), we have made the first discovery that IPs not only amplify but also diversify PyN types. Whereas ~20% of all PyNs derive from RGs and ~80% derive from IPs, this ratio varies widely according to projection types: whereas PyNs$^{Fezf2}$ (PT) derive from RG and IP in equal ratio, PyNs$^{PlexD1}$ (IT) are only generated from indirect neurogenesis, and PyNs$^{Tle4}$ (corticothalamic) are only generated from RGs (Fig. 4C,D). Importantly, direct and indirect PyNs$^{Fezf2}$ differ in their projection targets, implying differential connectivity and function. These results may begin to link developmental, functional, and evolutionary considerations of cortical neuron types.

Classic embryonic cell birth-dating studies revealed an inside-out trend of cortical lamination, but the progenitor mechanisms, and the role of IPs in particular, in the generation-distinct PyN projection types are unclear. Using an inducible Tbr2-CreER driver, we performed comprehensive fate mapping of IPs throughout neurogenesis. In addition to quantifying the laminar location of fate-mapped PyNs, we combine anterograde and retrograde viral labeling to reveal their axon projection patterns. We found that (1) individual IPs are fate-committed to produce "twin PyNs" with near-identical location and morphology, (2) both upper and lower layer PyNs are generated at both early and late embryonic times, (3) nonconsecutive layers are generated at the same time, and (4) the same layer is generated at different times. By analyzing the axon projection patterns of fate-mapped PyNs, our results reveal the orderly production and deployment of PyN projection types, instead of a strict inside-out sequence, as a developmental basis underlying the construction of cortical architecture. These findings will lead to a significant revision of a foundational concept of cortical development.

## PUBLICATIONS

Chen X, Sun YC, Zhan H, Kebschull JM, Fischer S, Math K., Huang ZJ, Gillis J, Zador AM. 2019. High-throughput mapping of long-range neuronal projection using in situ sequencing. *Cell* **179:** 772–786.e19.

Huang ZJ, Paul A. 2019. The diversity of GABAergic neurons and neural communication elements. *Nat Rev Neurosci* **29:** 563–572.

Kelly SM, Raudales R, Moissidis M, Kim G, Huang ZJ. 2019. Multipotent radial glia progenitors and fate-restricted intermediate progenitors sequentially generate diverse cortical interneuron types. bioRxiv doi:10.1101/735019

Wang B-S, Sol Bernardez Sarria M, He M, Crair MC, Huang ZJ. 2019. Retinal and callosal activity-dependent chandelier cell elimination shapes binocularity in primary visual cortex. bioRxiv doi:10.1101/612358

Wang X, Tucciarone J, Jiang S, Yin F, Wang BS, Wang D, Jia Y, Jia X, Li Y, Yang T, et al. 2019. Genetic single neuron anatomy reveals fine granularity of cortical axo-axonic cells. *Cell Rep* **26:** 3145–3159.e5.

*In Press*

Lau BYB, Krishnan K, Huang ZJ, Shea SD. 2020. Maternal experience–dependent cortical plasticity in mice is circuit- and stimulus-specific and requires MECP2. *J Neurosci* **40:** 1514–1526.

Matho K, Huilgol D, Galbavy W, Kim G, He M, An X, Lu J, Wu P, Di Bella DJ, Shetty AS, et al. 2020. Genetic dissection of glutamatergic neuron subpopulations and developmental trajectories in the cerebral cortex. bioRxiv doi:10.1101/2020.04.22.054064v1

# NEUROBIOLOGY OF COGNITION AND DECISION-MAKING

**A. Kepecs**
A. Arora
D. Bonda
M. Bosc
Q. Chevy
E. Demir
E. Gibson
T.S. Gouvea

Q.A. Li
S. Li
P. Masset
T. Ott
T. Pinkhasov
L. Ramirez Sanchez
S. Ren

K. Schmack
A. Siebels
S. Starosta
J.F. Sturgill
M. Wulf
W. Zhong

Our long-term goal is to reverse engineer the computational and neurobiological processes underlying cognition and decision-making and apply the resulting insights to biological psychiatry. We start with quantifiable behavioral tasks, for both humans and rodents, which enable us to isolate and study distinct cognitive processes. We probe the neural basis of these processes in rodents using state-of-the-art electrophysiological, imaging, and optogenetic techniques to establish the underlying neural mechanisms. Given the complexity of behavior and the dynamics of neural networks producing it, we also develop new algorithms for analyzing data and models to help us interpret them. By understanding how brains accomplish cognitive tasks, often beyond the capacity of current machine-learning algorithms, we also expect to uncover new computational principles. Using this integrated approach over the past decade, we have established the neurobiological basis of decision confidence and identified several cell type–specific cortical principles. Going forward, the laboratory will be working to build a bridge from animal studies to psychiatry to understand what goes awry in the brain during mental illness.

## Confidence in Rats, Humans, Brains, and Statistics

Every decision we make is accompanied by a sense of confidence about its likely outcome. This sense informs subsequent behavior, such as investing more—whether time, effort, or money—when the payoff is more certain. Conversely, the pathological misevaluation of confidence contributes to a wide range of neuropsychiatric conditions, including anxiety, obsessive–compulsive disorder, and addiction. Our long-term goal has been to understand how the brain implements confidence judgments and acts on these.

Over the past decade, we have developed a set of behavioral tasks and theoretical frameworks that rigorously translates the psychological concept of confidence into a formally defined decision variable. Using this approach, we have identified orbitofrontal cortex (OFC) neurons that encode the confidence associated with a perceptual decision. Furthermore, we have derived a mathematical framework for decision confidence from first principles of statistics (Hangya et al., *Neural Comput* **28:** 1 [2016]). We showed that key properties of statistical decision confidence match human self-reported confidence (Sanders et al., *Neuron* **90:** 499 [2016]), providing a deep link between objective and subjective notions of confidence.

Previously we showed that the firing of many OFC neurons encodes statistical confidence about olfactory-discrimination decisions and OFC inactivation specifically impairs rats' ability to optimally invest time waiting for reward. However, a neural representation of abstract confidence should not just (1) reflect a confidence computation, as we have shown, but also (2) predict multiple confidence-guided behaviors and (3) be independent of the source of information used to make a choice (i.e., independent of sensory modality). Using a new task design, we have been able to show now that single orbitofrontal cortex neurons encode statistical confidence and predict two confidence-guided behaviors (trial-by-trial time investment and confidence-guided updating) and do so irrespective of whether the sensory discrimination was olfactory or auditory. Therefore OFC appears to contain a modality-general representation of confidence that

could provide an information source–independent probability estimate, useful for confidence-driven adaptive behaviors, such as learning and time investment (Masset et al. 2020).

Confidence or uncertainty has been long suggested to modulate the degree of learning. We identified a novel form of reinforcement learning during perceptual decisions that depends on the confidence of past sensory judgments. We showed that these outcome-dependent biases depend on the strength of past sensory evidence, suggesting that they are consequences of confidence-guided updating of choice strategy. We illustrate that this form of choice updating is a widespread behavioral phenomenon that can be observed across various perceptual decision-making paradigms in mice, rats, and humans. This trial-to-trial choice bias was also present in different sensory modalities and transferred across modalities in an interleaved auditory/olfactory choice task. To explain these observations, we have formulated a class of reinforcement-learning models that compute prediction errors scaled by decision confidence and produce confidence-guided updating of choice bias (Lak et al. 2020a,b).

Neurons in the orbitofrontal cortex, like in other regions of frontal cortex, display baffling complexity, responding to a mixture of sensory, motor, and other variables. We developed a new approach to understand the representational content and architectural logic of higher-order cortical areas. We found that discrete groups of orbitofrontal cortex neurons encode distinct decision variables and these categorical representations map directly onto decision-variables, such as reward size, decision confidence, and integrated value, in a choice model explaining our task. This suggests that, like sensory neurons, frontal neurons form a sparse and overcomplete population representation aligned to the natural statistics of the world—in this case spanning the space of decision variables required for optimal behavior (Hirokawa et al. 2019).

Beyond our rodent work, we are now in a position to use these quantitative measures of decision confidence in humans, through collaborations, with neuroimaging and genetic approaches. Moreover, by showing that a single confidence measure is applicable to humans and rodents, our results strengthen the case for using the rat as a model system for translational studies in cognition and psychiatry.

## Cell Type–Specific Cortical Architecture

This work was done in collaboration with Z.J. Huang and J. Tollkuhn (CSHL), B. Rozsa (Hungarian Academy of Sciences), and A. Nectow (Princeton).

Reverse engineering cortical algorithms requires understanding the contributions of a multitude of different cortical neuron types, each with unique properties and connectivity. Thus, to identify neural circuit implementations of computations among diverse cortical elements, we must couple sophisticated tasks with targeted, high-throughput monitoring and manipulation of neural circuits at cell-type resolution. Because there is no consensus definition of neural types, we use complementary techniques based on genetic and projection targeting with optogenetics-assisted cell-type identification. We continued our efforts to expose the cell type–specific logic of cortex by focusing on a few key inhibitory and projection cell-types (Fishell and Kepecs 2020).

Chandelier cells (ChCs) constitute perhaps the most unique GABAergic interneurons in cortex. They specialize in innervating the axon initial segment of excitatory pyramidal neurons, the site for action potential generation—yet it remains unclear whether they function to inhibit their targets. Taking advantage of a genetic and viral approach developed in the Huang laboratory, we are able to target a subgroup of ChCs in the prefrontal cortex (PFC). By combining electrical stimulation at the basolateral amygdala (BLA) and extracellular recordings in PFC, we are aiming to determine whether ChCs inhibit or excite their BLA-projecting pyramidal neurons. We are also studying the role of these circuits during approach-avoidance conflict to gain insights into the principles of their recruitment during behavior.

We have recently identified a disinhibitory cortical circuit motif that appears to be a conduit for fast neuromodulatory action in the cortex. This circuit is controlled by a class of inhibitory interneurons that express vasoactive intestinal polypeptide (VIP) and inhibit other interneurons, thereby disinhibiting a subpopulation of principal neurons. Functionally, we showed that VIP interneurons in the auditory cortex are recruited in response to specific reinforcement signals such as reward and punishment. To explore the generality of these observations across cortex regions, we are collaborating with B. Rozsa (KOKI, Hungarian Academy of Sciences), using a state-of-the-art 3D random-access inertia-free acousto-optic deflector (AOD) two-photon

imaging system to recording the sparse VIP population across large regions.

Reward and punishment elicited rapid activation of most VIP interneurons cortex-wide, which could not be explained by arousal state modulation alone. Visual cortical VIP interneurons also showed weak sensory tuning that did not predict their reinforcer-triggered activation. These studies revealed that a global response mode of VIP interneurons provides a means for reinforcement signals to influence local circuit computations and their plasticity.

To understand OFC neuron types, we use two complementary approaches. First, we target specific projection neurons using retrograde viruses and use optogenetic stimulation to identify these in electrophysiological recordings (Li et al., *Neuron* **97:** 481 [2018]). Using this technique, we have found that OFC projections to ventral striatum show a characteristic response in which negative value signals are sustained throughout the intertrial interval to the beginning of the subsequent trial (Hirokawa et al. 2019). Second, in collaboration with the Klausberger laboratory, we use juxtacellular labeling to target neurons based on their functional response profiles. Once labeled, neurons are subjected to detailed ex vivo analysis of the axonal projection patterns. Using these techniques, we have begun to record and identify OFC neurons that specifically signal the confidence-dependent time investments. We expect that the combination of juxtacellular and optogenetically identified extracellular recordings will enable us to reverse engineer the cell type–specific circuit logic of orbitofrontal cortex.

## Computational Logic of Neuromodulation

This work was done in collaboration with B. Hangya (Hungarian Academy of Sciences) and Y. Li (Peking University).

Neuromodulators constitute central brain systems with cell bodies located in deep brain areas that project across large areas of the brain, providing broadcast signals to reconfigure circuits through a unique set of neurotransmitters, such as acetylcholine (ACh), dopamine (DA), serotonin (5HT), and norepinephrine (NE). These neuromodulators have been implicated in a broad range of behavioral functions, many overlapping across modulators, such as learning (ACh, DA), arousal (NE, ACh), and impulsivity (5HT, DA). Based on these differences, each

neuromodulator has been suggested to have a distinct computational function. Using a new class of genetically encoded sensors, we are investigating the roles of multiple neuromodulators in simple learning tasks to understand their similarities. We are also probing the roles of dopamine in foraging, impulsivity, and motor behaviors, to understand how its functions generalize across behaviors.

The basal forebrain cholinergic (CBF) systems constitute a major neuromodulator implicated in normal cognition functions, as well as cognitive deficits in Alzheimer's disease, Parkinson's dementia, and age-related dementias. CBF releases ACh across the cortical mantle and mediates seemingly disparate functions, including attention and learning. It remains unclear whether the CBF achieves this versatility through distinct projections conveying target-specific messages, or through a coordinated broadcast that reflects a single underlying computational principle. Here, we show that both the spiking of CBF neurons and ACh release at multiple cortical targets (amygdala, medial prefrontal and auditory cortices) show hallmarks of a prediction error signal. CBF neurons responded to unpredictable primary reinforcers, acquired responses to reinforcement-predictive stimuli, and concomitantly showed diminished responses to predicted reinforcers. Reward and punishment both activated the CBF, revealing the prediction error signal to be free of hedonic valence. During a reversal learning task, ACh tracked conditioned behavior as swiftly as DA, and they also shared fluctuations of activity from trial to trial. We identified shared inputs to forebrain ACh and midbrain DA neurons from regions known to carry outcome-predictive signals, exposing an overlapping circuit for prediction error computations. We propose that these predictive and valence-free characteristics explain how ACh can prospectively (attention) and retrospectively (learning) promote adaptive responses to behaviorally salient events.

## Social Reward and Cognition

This work was done in collaboration with R. Axel (Columbia University).

Social behavior is integral to animals' survival and reproduction—social deficits are at the heart of psychiatric disorders such as autism spectrum disorder that

have proven profoundly difficult to study in model organisms. We would like to understand how social information is represented, computed, and used by mice. In rodents, the primary source of information for social decision-making and reward valuation is the chemosensory system.

One component of socially/behaviorally important scents is darcin, a mouse urinary protein that is sufficient to induce innate attraction in sexually receptive female mice and also acts as an unconditioned stimulus in associative-learning paradigms. We have identified a neural circuit that extends from the vomeronasal organ to the medial amygdala and mediates the innate response to darcin. Silencing of either accessory olfactory bulb or the medial amygdala eliminates the innate attraction to darcin, whereas optogenetic reactivation of darcin-activated medial amygdala neurons elicits attraction behavior. This medial amygdala circuit thus acts as a central hub for integrating pheromonal information and conveying it to other structures to promote mate encounters and produce reinforcements (Demir et al. 2020).

We are also interested in understanding basic rules that mice use to choose partners. For this purpose, we have developed a psychophysical social behavior task, the "social carousel," inspired by perceptual psychophysics and game theoretic traditions that have been instrumental in the study of other facets of cognition. Our task enables reliable, quantitative, and high-throughput analysis of social interactions in mice. Mice can choose to engage in extended social interactions with the caged mice at the expense of delaying the water reward. This task allows us to infer the "social value" of a mouse based on the trade-off between social interactions and appetitive rewards. In addition, this task is compatible with our electrophysiological studies because it is devised for precise stimulus delivery and reproducible behavioral contingencies.

## Computational and Circuit Psychiatry: Hallucinations and Impulsivity

Our behavioral research is aimed at determining the computational principles of cognition. The starting point of our studies is the observation that behavior is often described using folk psychological categories, even though almost everyone agrees that the neural architecture that supports behavior performs computational functions. Hence, there is a large gap between the well-established molecular and neural circuit–level descriptions of the brain and much less developed computational descriptions of behavior produced by the brain. We try to fill this gap using two main sources of insight. First, we use psychophysical and behavioral economic approaches to design tasks and collect large data sets. Second, we use machine-learning and theoretical neuroscience models in an attempt to reverse engineer the algorithms. Our studies have mainly focused on decision confidence in rats, mice, and humans, and are beginning to expand into impulsivity and cognitive control and computational phenotyping of human behavior, including psychiatric populations.

Hallucinations, a core symptom of schizophrenia, can be operationalized as false percepts that are experienced with the same certainty as veridical percepts. Based on this insight we have developed a behavioral task to quantitatively measure hallucination-like percepts in mice. We trained mice to perform a psychometric auditory detection task with a time investment–based confidence report. Using computational modeling, we found that choice behavior closely followed the predictions of an ideal observer model, suggesting that mice indeed reported their perception and their confidence. We are currently testing the translational validity of this behavioral test by reproducing previous results and setting up patient testing. This mouse model will enable a circuit-level description of the causal link between dopamine and schizophrenia symptoms.

Impulsivity is a behavioral trait present in many psychiatric disorders that significantly increases the risk of suicide, violence, and criminal behavior. Whether a particular decision made too early is impulsive is challenging to determine, because misevaluation of expected outcome or misestimation of time could lead to similar consequences. Therefore, as an initial step toward understanding the underlying neural circuits, we devised a behavioral task that isolates the contribution of impulsivity to individual choices and separates it from reward valuation. Our goal is to elucidate the contribution of neural circuits that are involved in impulsivity, particularly the anterior cingulate cortex's control of the serotonergic and dopaminergic systems, using photometry recordings and optogenetic manipulation. Ultimately, we hope our circuit-based understanding of impulsivity will con-

tribute to the design of circuit-specific treatments for impulsivity disorders.

## Brain–Body Interactions: Cancer Neuroscience and the Immune Bridge to the Brain

This work was done in collaboration with T. Janowitz, D. Tuveson, L. Trottman, and P. Osten (CSHL).

The intimate connection between the brain and the body has long been recognized in popular culture as well as by mystics. The brain and all organ systems of the body evolved together to support the survival of an organism. Yet, compared to our extensive scientific knowledge of the many systems of the body and the brain, there remains an enormous gap in our biological understanding of how the brain and body interact.

Understanding the interactions across the multitude of organ systems that support survival is also essential for the progress of medicine. Lacking such systems-level understanding creates challenges for medical treatments for many conditions, as illustrated by the COVID-19 pandemic. Although the initial steps of the infection by the novel coronavirus are understood at a molecular level, how it induces its deadly multisystem pathology from acute respiratory distress syndrome, thromboembolic disorder and extreme fatigue is shrouded in mystery.

We have begun a new line of research to break the disciplinary boundaries, starting with studies on the neural control of cancer and cancer-induced behavioral changes. We are using viral approaches to trace the neural connections from the brain to the pancreas and the prostate in order to understand whether and how the autonomic nervous system can drive tumor growth and proliferation.

Cancer is often accompanied by mood disorders such as depression, which have been assumed to be psychological reactions to the life-threatening disease. Emerging research, however, points to the possibility that tumors themselves might influence the brain and produce apathy. Therefore, we have begun to investigate the behavioral and neural changes that accompany a devastating metabolic condition called cancer cachexia. Cancer cachexia is classically characterized by aberrant eating behavior, weight loss, and muscle wasting, and is often accompanied by changes in mood. Using cachetic mice, we are studying

behavioral correlates of apathy as deficits in effort-related decision-making. To understand the neural basis of cachexia-induced apathy, we are using brain-wide c-fos mapping and dopamine photometry measurements. We aim to identify causal links between cancer-induced neuroinflammation and the observed behavioral deficits. Ultimately, we hope to develop a new etiologically valid mouse model for apathy and gain a new entryway toward a biological understanding of mood disorders.

## Neurotechnologies: Viral Targeting, Behavioral Language, and Nanophotonics

This work was done in collaboration with M. Lipson (Columbia University).

Progress in neuroscience relies on continual technique development and improvement. Part of our efforts is devoted to developing, improving, and adopting instrumentation that enables the study of neural circuits and behavior. For instance, we developed an open-source behavioral control system (Bpod) that is used by more than 80 laboratories around the world. Building on this, we are also developing a formal behavioral description language that is hardware-independent and brings rigor and reproducibility to complex behavioral tasks.

We also continue to develop our viral complementation strategy that enables tropism-free retrograde viral delivery for targeting long-range projections (Li et al., *Neuron* **98:** 905 [2018]). This approach overcomes the major limitation of traditional viral techniques; traditional viral techniques rely on cell type–specific molecules for uptake and transport and, as a result, may fail to infect neurons that do not express the requisite complement of surface receptors (viral tropism). We have developed a receptor-complementation strategy to overcome this problem for the canine adenovirus type 2 (CAV-2). We are currently developing variants that will enable brain-wide expression through systemic delivery of a viral receptor, CAR.

We collaborate to develop and test a new class of nanophotonics silicon probes for high-density optical stimulation. The ability to activate neural populations using optogenetics has revolutionized the study of neural circuits; however, optogenetic stimulation typically relies on a single fiber to flood light into a large volume of the brain. With the Lipson group we designed an

implantable silicon-based probe that can switch and route multiple optical beams to stimulate identified sets of neurons across cortical layers and simultaneously record the produced spike patterns. Using an eight-beam probe, we can independently stimulate small groups of single neurons to produce multineuron spike patterns at submillisecond precision. We have tested an integrated nanophotonic silicon probe with cofabricated electrical recording sites to simultaneously optically stimulate and electrically measure deep-brain neural activity (Mohanty et al., arXiv:1805.11663 [2018]).

## PUBLICATIONS

Browning M, Carter CS, Chatham C, Den Ouden H, Gillan CM, Baker JT, Chekroud AM, Cools R, Dayan P, Gold J, et al. 2020. Realizing the clinical potential of computational psychiatry: report from the Banbury Center Meeting, February 2019. *Biol Psychiatry* **88:** e5–e10. doi:10.1016/j.biopsych.2019.12.026

Hirokawa J, Vaughan A, Masset P, Ott T, Kepecs A. 2019. Frontal cortex neuron types categorically encode single decision variables. *Nature* **576:** 446–451. doi:10.1038/s41586-019-1816-9

Kepecs A. 2019. Summary: order and disorder in brains and behavior. *Cold Spring Harb Symp Quant Biol* **6:** 038935. doi:10.1101/sqb.2018.83.038935

Richards BA, Lillicrap TP, Beaudoin P, Bengio Y, Bogacz R, Christensen A, Clopath C, Costa RP, de Berker A, Ganguli S, et al. 2019. A deep learning framework for neuroscience. *Nat Neurosci* **22:** 1761–1770. doi:10.1038/s41593-019-0520-2

*In Press*

Demir E, Li K, Bobrowski-Khoury N, Sanders JI, Beynon RJ, Hurst JL, Kepecs A, Axel R. 2020. The pheromone darcin drives a circuit for innate and reinforced behaviours. *Nature* **578:** 137–141. doi:10.1038/s41586-020-1967-8

Fishell G, Kepecs A. 2020. Interneuron types as attractors and controllers. *Annu Rev Neurosci* **43:** 1–3. doi:10.1146/annurev-neuro-070918-050421

Kepecs A, Lipson M. 2020. Reconfigurable nanophotonic silicon probes for sub-millisecond deep-brain optical stimulation. *Nat Biomed Eng* **4:** 223–231. doi:10.1038/s41551-020-0516-y

Lak A, Hueske E, Hirokawa J, Masset P, Ott T, Urai AE, Donner TH, Carandini M, Tonegawa S, Uchida N, Kepecs A. 2020a. Reinforcement biases subsequent perceptual decisions when confidence is low, a widespread behavioral phenomenon. *eLife* **9:** e49834. doi:10.7554/eLife.49834

Lak A, Okun M, Moss MM, Gurnani H, Farrell K, Wells MJ, Reddy CB, Kepecs A, Harris KD, Carandini M. 2020b. Dopaminergic and prefrontal basis of learning from sensory confidence and reward value. *Neuron* **105:** 700–711.e6. doi:10.1016/j.neuron.2019.11.01

Masset P, Ott T, Lak A, Hirokawa J, Kepecs A. 2020. Behavior- and modality-general representation of confidence in orbitofrontal cortex. *Cell* **182:** 112–126.e18. doi:10.1016/j.cell.2020.05.022

Mohanty A, Li Q, Tadayon MA, Roberts SP, Bhatt GR, Shim E, Ji X, Cardenas J, Miller SA, Monje M, et al. 2020. Roadmap for the emerging field of cancer neuroscience. *Cell* **181:** 219–222. doi:10.1016/j.cell.2020.03.034

# MATHEMATICAL PRINCIPLES OF NEURAL COMPUTATION

**A. Koulakov**   B. Başerdem   D. Kepple   N. Tran
H. Giaffar   S. Shuvaev

Our laboratory works on theories of neural computation. Our overall strategy is to use methods developed in mathematics, physics, machine learning, computer science, and statistics to build experimentally testable models of neural networks and their function. In most cases, we base our theories on what is known about particular biological systems; however, given that the principles of brain function remain unclear, in many cases, we resort to building computational theories. This means that we formulate the problems solved by the brain in a mathematically rigorous fashion and hypothesize how an engineer would solve the problem, given the biological and experimental constraints. We then use these solutions to form experimentally testable predictions. Testing these predictions in collaboration with our experimental colleagues helps us refute or refine our theories. For example, we are interested in understanding how connectivity is established in the brain. We have proposed several theories that may determine the rules of making connections between neurons based on a limited set of instructions contained in the genome. These theories address several levels of organization, including computational, biological, engineering, and evolutionary. Our theories may explain the differences between connectivities in normal and abnormal brain circuits. We are also interested in understanding the principles of perceptual invariance—that is, how can sensory systems represent objects in the environment despite substantial variations in intensity and background. Visual percepts, for example, retain basic features, such as perceived shape and color composition, despite variable luminance, spectral composition, scale, and position of the stimuli. Although we study the question of perceptual invariance in application to well-defined problems, we believe that the principles that we will uncover may generalize across sensory modalities. Finally, we are pursuing the question of how modern theories of machine learning and artificial intelligence

can apply to brain function. Although reinforcement learning, deep learning, long short-term memory networks, etc., are successful in solving a variety of artificial intelligence problems, their mapping onto brain circuits remains unclear. We attempt to bring these systems closer to satisfying the constraints imposed by biology. We hope that the convergence of theoretical constructs and their biological underpinning will help us learn more about brain function.

## Mosaic Representations of Odors in the Input and Output Layers of the Mouse Olfactory Bulb

D. Kepple, A. Koulakov [in collaboration with H.G. Chae, W. Bast, and F. Albeanu, CSHL; V. Murthy, Harvard]

The elementary stimulus features encoded by the olfactory system remain poorly understood. We examined the relationship between 1,666 physical–chemical descriptors of odors and the activity of olfactory bulb inputs and outputs in awake mice. Glomerular and mitral and tufted cell responses were sparse and locally heterogeneous, with only a weak dependence of their positions on physical–chemical properties. Odor features represented by ensembles of mitral and tufted cells were overlapping but distinct from those represented in glomeruli, which is consistent with an extensive interplay between feedforward and feedback inputs to the bulb. This reformatting was well described as a rotation in odor space. The physical–chemical descriptors accounted for a small fraction in response variance, and the similarity of odors in the physical–chemical space was a poor predictor of similarity in neuronal representations. Our results suggest that commonly used physical–chemical properties are not systematically represented in bulbar activity and encourage further searches for better descriptors of odor space.

## Network Cloning Using DNA Barcodes

S. Shuvaev, B. Baserdem, A. Koulakov [in collaboration with T. Zador, CSHL]

The connections between neurons determine the computations performed by both artificial and biological neural networks. Recently, we have proposed SynSeq, a method for converting the connectivity of a biological network into a form that can exploit the tremendous efficiencies of high-throughput DNA sequencing. In SynSeq, each neuron is tagged with a random sequence of DNA—a "barcode"—and synapses are represented as barcode pairs. SynSeq addresses the analysis problem, reducing a network into a suspension of barcode pairs. Here, we formulate a complementary synthesis problem: How can the suspension of barcode pairs be used to "clone" or copy the network back into an uninitialized tabula rasa network? Although this synthesis problem might be expected to be computationally intractable, we find that, surprisingly, this problem can be solved efficiently, using only neuron-local information. We present the "one-barcode–one-cell" (OBOC) algorithm, which forces all barcodes of a given sequence to coalesce into the same neuron, and show that it converges in a number of steps that is a power law of the network size. Rapid and reliable network cloning with single-synapse precision is thus theoretically possible.

## Deconstructing Odorant Identity via Primacy in Dual Networks

D. Kepple, H. Giaffar, A. Koulakov [in collaboration with D. Rinberg, NYU]

In the olfactory system, odor percepts retain their identity despite substantial variations in concentration, timing, and background. We study a novel strategy for encoding intensity-invariant stimulus identity that is based on representing relative rather than absolute values of stimulus features. For example, in what is known as the primacy coding model, odorant identities are represented by the conditions that some odorant receptors are activated more strongly than others. Because, in this scheme, odorant identity depends only on the relative amplitudes of olfactory receptor responses, identity is invariant to changes in both intensity and monotonic nonlinear transformations of its neuronal responses. Here, we show that

sparse vectors representing odorant mixtures can be recovered in a compressed sensing framework via elastic net loss minimization. In the primacy model, this minimization is performed under the constraint that some receptors respond to a given odorant more strongly than others. Duality is a concept frequently used in applied mathematics, machine learning, elementary particle physics, field theories, and other fields of theoretical physics. We argue here that neural networks are ideally posed to solve a variety of dual problems. Using duality transformation, we show that the constrained optimization problem of sparse olfactory signal recovery can be solved by a neural network whose Lyapunov function represents the dual Lagrangian and whose neural responses represent the Lagrange coefficients of primacy and other constraints. The connectivity in such a dual network resembles known features of connectivity in olfactory circuits. We thus propose that networks in the piriform cortex implement dual computations to compute odorant identity, with the sparse activities of individual neurons representing Lagrange coefficients. More generally, we propose that sparse neuronal firing rates may represent Lagrange multipliers, which we call the dual brain hypothesis. We show such a formulation is well-suited to solve problems with multiple interacting relative constraints.

## Reinforcement Learning Basis of Social Conflict

S. Shuvaev, A. Koulakov [in collaboration with K.U. Venkataraju and P. Osten, CSHL; E. Amelchenko, D. Smagin, and G. Enikolopov, SUNY Stony Brook; N. Kudryavtseva, Institute of Cytology and Genetics, Novosibirsk, Russian Federation]

Establishment of social hierarchy through intermale aggressive behavior helps to deflect excessive violence and injury, protects a group's valuable resources, and molds the societal structure. Successful acts of aggression may be rewarding, with a series of wins increasing aggressive motivation and propensity to engage in aggressive behavior, and a series of defeats having an opposite, aversive, effect. Social hierarchy is a dynamic system that may be altered after encounters between animals of a comparable rank. Several brain regions and circuits involved in aggressive behavior and social dominance have been identified; however, the quantitative principles describing social conflict are

unknown. Here, we used an animal model of male-on-male social conflict and applied global mapping of c-Fos expression to determine network activity across the brain regions and infer the circuits involved in aggression, defeat, and reversal of social status. Male mice were trained in a model of chronic social conflicts and analyzed in several behavioral paradigms: (1) aggression (3, 10, or 20 days of consecutive agonistic interactions in pairs, producing winner and loser animals); (2) deprivation (20 days of agonistic interactions followed by 14 days of fight deprivation); and (3) inversion (20 days of agonistic interactions followed by placing the animals in new pairs: winner vs. winner and loser vs. loser). Overall, our data set included 116 whole-brain samples for 20 conditions, including four types of controls. Each sample was represented by about 40 million voxels (Fig. 1). We show that this data set can be described by two axes representing social status and experience, respectively. We then train deep neural Q-networks to maximize rewards in the conditions of chronic social conflict by self-play. We specify the rules of social conflict by defining the reward function that yields diverse behaviors, such as aggression or cowardice. We perform training in two steps. First, to model evolution, we place networks in a large number of conflicts, adjust their weights using Q-learning via self-play, and pool their weights to yield similar prototypical structure. Second, we place the evolutionary optimized networks in various antagonistic conditions similar to the ones used in our experimental design—that is, pure aggression, deprivation, and inversion. We observe that the arrangement of samples along the social dominance axis is similar between brain activation patterns observed experimentally and the Q-learning networks trained via self-play. This arrangement of
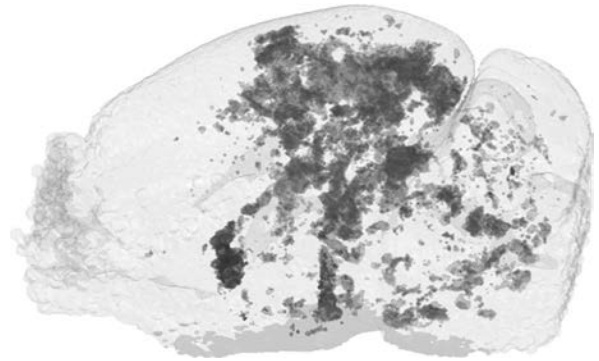


**Figure 1.** Regions in the mouse brain that allow differentiation of losing from winning animals engaged in chronic social conflict.

networks along the social hierarchy axis predicts certain correlations in the temporal sequence of outcomes of a series of social conflicts. We then examined data on the results of 3,569 bouts between 1,561 Ultimate Fighting Championship (UFC) fighters. We find that the correlation following from the mouse data and reinforcement learning networks is indeed observed in the outcomes of these matches. Finally, we identify functional subnetworks of coactivated brain regions that may yield these results.

## PUBLICATIONS

Chae H, Kepple DR, Bast WG, Murthy VN, Koulakov AA, Albeanu DF. 2019. Mosaic representations of odors in the input and output layers of the mouse olfactory bulb. *Nat Neurosci* **22:** 1306–1317. doi:10.1038/s41593-019-0442-z

Kepple DR, Giaffar H, Rinberg D, Koulakov AA. 2019. Deconstructing odorant identity via primacy in dual networks. *Neural Comput* **31:** 710–737. doi:10.1162/neco_a_01175

Shuvaev SA, Başerdem B, Zador AM, Koulakov AA. 2019. Network cloning using DNA barcodes. *Proc Natl Acad Sci* **116:** 9610–9615. doi:10.1073/pnas.1706012116

# THE FUNCTION AND PLASTICITY OF CENTRAL SYNAPSES IN ADAPTIVE AND MALADAPTIVE BEHAVIORS RELATED TO PSYCHIATRIC DISORDERS

**B. Li**   A. Bouhuis       A. Furlan         L. Kimoto        T. Yang
           S. Boyle          J. Giovanniello   R. Sharma        K. Yu
           C. Bravo-Rivera   O. Gschwend       D. van de Lisdonk   X. Zhang
           H. Deng           W. Guan           X. Xiao

The focus of research in our laboratory has been to understand the link between neural circuits and behaviors. We are particularly interested in studying the synaptic and circuit mechanisms underlying aspects of motivated behaviors, such as attention, motivation, and learning and memory, as well as synaptic and circuit dysfunctions that may underlie the pathophysiology of mental disorders, including anxiety disorders, depression, autism, and drug addiction. We integrate in vitro and in vivo electrophysiology, imaging, molecular, genetic, optogenetic, and chemogenetic methodologies to probe and manipulate the function of specific neural circuits in the rodent brain and to determine their roles in adaptive or maladaptive responses in various behavioral paradigms. We are currently undertaking the following major lines of research.

## The Role of Amygdala Circuitry in Motivated Behaviors

Our previous studies demonstrate that the central amygdala (CeA) has a key role in learning and expression of defensive responses to threats. In particular, our studies indicate that somatostatin-expressing (SOM$^+$) neurons in the lateral division of the central amygdala (CeL) are essential for the acquisition and recall of fear memories. Another major class of CeL neurons, the protein kinase C-$\delta$-expressing (PKC-$\delta^+$) neurons, is essential for the synaptic plasticity underlying learning in the lateral amygdala, as it is required for lateral amygdala neurons to respond to unconditioned stimulus, and moreover carries information about the unconditioned stimulus (US) to instruct learning. Furthermore, we demonstrate that enhanced excitatory synaptic inputs onto SOM$^+$ CeL

neurons and the resulting reduction in inhibition onto downstream SOM$^+$ neurons in the bed nucleus of the stria terminalis (BNST) plays an important role in the generation of anxiety-related behaviors. Notably, our results indicate that an increase in dynorphin signaling in SOM$^+$ CeA neurons mediates the paradoxical reduction in inhibition onto SOM$^+$ BNST neurons, and that the consequent enhanced activity of SOM$^+$ BNST neurons is both necessary for and sufficient to drive the elevated anxiety. Our results unravel previously unknown circuit and cellular processes in the central extended amygdala that can cause maladaptive anxiety.

In parallel, we also investigate the role of the basolateral amygdala (BLA) in motivated behaviors. In a recent study (Zhang et al. 2020), we targeted a genetically defined BLA pyramidal neuron population and identified two functionally distinct classes in behaving mice—the negative valence neurons and positive valence neurons. These two populations innately represent the respective valences and through learning acquire responses predicting negative or positive outcomes for promoting punishment avoidance or reward seeking. Notably, these two classes of neurons receive inputs from separate sets of sensory and limbic areas and convey punishment and reward information through projections to the nucleus accumbens and olfactory tubercle, respectively, in the ventral striatum to drive negative and positive reinforcement. Thus, valence-specific BLA neurons are wired with distinctive input/output structures, forming a circuit framework that supports BLA's roles in encoding, learning, and executing motivational behaviors. These new findings provide novel insights into how BLA circuits contribute to behaviors driven by reward and punishment and how dysfunctions in these circuits may cause mental disorders such as depression.

## The Basal Ganglia Circuit in Motivated Behaviors

The basal ganglia, a group of subcortical nuclei, play a crucial role in motivated behaviors. Recently we showed that neurons in the habenula-projecting globus pallidus (GPh), an output of the basal ganglia, are essential for evaluating action outcomes. Our current study addresses the roles of another basal ganglia output, the ventral pallidum (VP), in motivated behaviors.

The VP is critical for invigorating reward seeking and is also involved in punishment avoidance, but how it contributes to such opposing behavioral actions remains unclear. In a recent study (Stephenson-Jones et al. 2020), we show that GABAergic and glutamatergic VP neurons selectively control behavior in opposing motivational contexts. In vivo recording, combined with optogenetics in mice, revealed that these two populations oppositely encode positive and negative motivational value, are differentially modulated by an animal's internal state, and determine the behavioral response during motivational conflict. Furthermore, GABAergic VP neurons are essential for movements toward reward in a positive motivational context, but suppress movements in an aversive context. In contrast, glutamatergic VP neurons are essential for movements to avoid a threat, but suppress movements in an appetitive context. Our results indicate that GABAergic and glutamatergic VP neurons encode the drive for approach and avoidance, respectively, with the balance between their activities determining the type of motivational behavior.

## Circuit Mechanisms of Cortical Dysfunction in a Genetic Model of Schizophrenia

Altered cortical excitation–inhibition (E-I) balance resulting from abnormal parvalbumin interneuron (PV IN) function is a proposed pathophysiological mechanism of schizophrenia and other major psychiatric disorders. Preclinical studies have indicated that *disrupted-in-schizophrenia-1* (*Disc1*) is a useful molecular lead to address the biology of prefrontal cortex (PFC)-dependent cognition and PV IN function. To date, PFC inhibitory circuit function has not been investigated in depth in *Disc1* locus impairment (LI) mouse models. Therefore, in a recent study (Delevich et al. 2020), we used a *Disc1* LI mouse model to investigate E-I balance in medial PFC (mPFC) circuits. We found that inhibition onto layer 2/3 excitatory pyramidal neurons in the mPFC was significantly reduced in *Disc1* LI mice. This reduced inhibition was accompanied by decreased GABA release from local PV, but not somatostatin (SOM) INs, and by impaired feed-forward inhibition (FFI) in the mediodorsal thalamus (MD) to mPFC circuit. Our mechanistic findings of abnormal PV IN function in a *Disc1* LI model provide insight into biology that may be relevant to neuropsychiatric disorders, including schizophrenia.

### PUBLICATIONS

Li B. 2019. Central amygdala cells for learning and expressing aversive emotional memories. *Curr Opin Behav Sci* **26:** 40–45. doi:10.1016/j.cobeha.2018.09.012

Yan H, Li B, Wang J. 2019. Non-equilibrium landscape and flux reveal how the central amygdala circuit gates passive and active defensive responses. *J R Soc Interface* **16:** 20180756. doi:10.1098/rsif.2018.0756

*In Press*

Delevich K, Jaaro-Peled H, Penzo M, Sawa A, Li B. 2020. Parvalbumin interneuron dysfunction in a thalamo-prefrontal cortical circuit in *Disc1* locus impairment mice. *eNeuro* **7:** ENEURO.0496-19.2020. doi:10.1523/ENEURO.0496-19.2020

Stephenson-Jones M, Bravo-Rivera C, Ahrens S, Furlan A, Xiao X, Fernandes-Henriques C, Li B. 2020. Opposing contributions of GABAergic and glutamatergic ventral pallidal neurons to motivational behaviors. *Neuron* **105:** 921–933. doi:10.1016/j.neuron.2019.12.006

Zhang X, Guan W, Yang T, Furlan A, Xiao X, Yu K, Huang ZJ, Li B. 2020. Genetically defined amygdala-striatal circuits for valence-specific behaviors. *Nat Neurosci* (in revision).

# THE STUDY OF INTELLIGENT MACHINES

**P.P. Mitra**    S. Abbu      L. Kimoto    J. O'Rourke
            S. Banerjee   C. Lee       S. Savoia
            B. Harman    X. Li        F. Xu
            C. Hsu       M. Lin
            B. Huo       L. Lodato

## EXPERIMENTAL

### Mesoscale Circuit Mapping in the Mouse and Marmoset Brains

The core research thrust in the laboratory is to uncover the principles of brain architecture by mapping brain circuitry at the mesoscopic scale, using tracer injections, across vertebrate species. This is a research program initiated by Dr. Mitra starting with a Banbury workshop in 2008 and a position paper in 2009 that proposed such mapping and introduced the concept of a "mesoscopic scale" into the neuroscience literature. The mesoscopic scale can be defined as the transitional length scale from a microscopic scale at which individual variation is prominent, to a macroscopic scale at which one can see species-typical patterns. The mesoscale roughly corresponds to the scale of interest in classical neuroanatomical reference atlases. Reference to mesoscale neuroanatomy has now become commonplace, and this proposal inspired projects at the Allen Institute for Brain Sciences as well as in other laboratories, including at CSHL.

The Mitra laboratory has collected data for mesoscale circuit mapping in the mouse (at CSHL) and also in the marmoset (at RIKEN, Japan) using a uniform neurohistological pipeline based on the tape-transfer method to transfer tissue sections cut with a cryomicrotome to glass slides. This method is compatible with automated instrumentation used in clinical anatomical pathology as well as whole-slide image scanners, is scalable to large format human brains, permits conventional histochemical and immunohistochemical process, and preserves tissue geometry—permitting reassembly of the sections into 3D brain volumes.

These two experimental efforts have produced a unique and unprecedented joint data set for comparative analysis of whole-brain mesoscale circuitry across rodents and primates (Fig. 1). The analysis of these data sets is a primary focus of current research in the laboratory, entailing development of new analytical methods drawn from computational geometry and topology as well as machine learning, described further below.

### Postmortem Human Whole-Brain Histology

One important advantage of the tape transfer–based neurohistological pipeline is scalability to larger format brains. The eventual goal of the laboratory is to understand mesoscale circuit architecture in the human brain. In previous work with Dr. L. Latour (NINDS) and colleagues, the tape transfer–based pipeline was applied to human brain samples with traumatic brain injury, in work that was published in 2019. This research direction is being taken forward in two ways: scaling up to whole human brain postmortem histology, in collaboration with colleagues at IIT-Madras, and an effort in the laboratory to map the detailed cytoarchitecture of the human hippocampus. In collaboration with David Nauen at Johns Hopkins Medical Institute, we have produced an unprecedented histological data set of a human hippocampus, with 20-micron alternating serial sections (Nissl and myelin), leading to ~15 TB of compressed image data. Preliminary visual examination of this unprecedented data set shows structure that has not been previously reported in the literature, and the analysis of this data set is currently in progress using the computational techniques developed for mouse and marmoset.

### Atlas-Mapped Single-Nucleus RNA Sequencing

Although there has been rapid progress in single-cell and single-nucleus RNA sequencing (snRNAseq) to obtain a full characterization of cell types in brains and other tissues, these techniques typically start from tissue microdissections that are difficult to map precisely to brain atlas space. In collaboration with Dr. Evan Macosko at the Broad Institute of MIT, we developed a new method for performing single-nucleus RNA sequencing of small tissue punches, with simultaneous Nissl histology on the same sections as well as
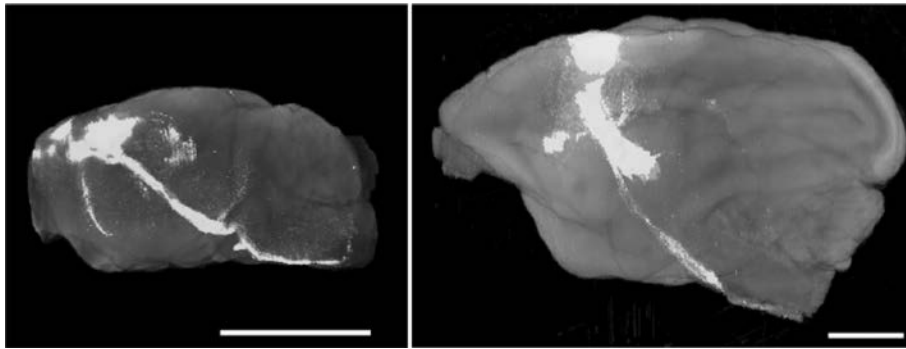
**Figure 1.** Sagittal views of a mouse brain (*left*) and a marmoset brain (*right*), both with tracer injections (AAV-GFP) in the primary motor cortex. The marmoset motor cortex appears to show a comparatively stronger projection to the thalamus. Scale bars, 5 mm.

adjacent sections, that permits precise localization in atlas space. A number of brains processed through this pipeline are now posted on a web resource, combining the Broad Institute's Single-Cell Portal for transcriptome-derived clusters of single nucleus data and high-resolution Nissl histology on the Brain Architecture portal (see www.brainarchitecture.org). This data set is now being analyzed. The atlas-mapped snRNAseq method is promising for comparing cell types across brains from different vertebrate taxa with mesoscale spatial resolution. A pilot project in the laboratory is pursuing this aim in an avian brain (the zebra finch), with a goal of addressing questions of comparative brain evolution.

### Histology Core for UCSD U19 Project

Despite the advances in whole-brain imaging techniques utilizing brain clearing and light sheet–based volumetric microscopy, classical Nissl-stained thin sections remain the most accurate overall method for cytoarchitectural delineation. This led our laboratory to become a histology core for a Brain Initiative U19 project led by David Kleinfeld (UCSD) focused on the mouse brainstem. In a methodological development, in collaboration with Dr. Kleinfeld, we are developing an arm of the pipeline utilizing water-based coverslipping media (as opposed to xylene-based media) for better preservation of fluorescent signal.

## COMPUTATIONAL INFORMATICS

### Virtual Neuroanatomist

The computational/informatics arm of the laboratory performs management, analysis, and web dissem-

ination of the data gathered in the experimental arm of the laboratory. An important part of this work is the development of a new class of data analysis techniques for neuroanatomical and transcriptomic data, combining methods from computational geometry, topology, and machine learning, to automate tasks normally performed by human experts. We refer to this task as the development of a "Virtual Neuroanatomist." The Virtual Neuroanatomist has to perform three tasks: "atlas mapping" (brain to atlas registration), "semantic segmentation" of the images to extract neuronal compartments of interest (soma, dendrites, axons, etc.), and "skeletonization" of single neurons and tracer injections in order to characterize the mesoscale circuit anatomy.

### *Brain To Atlas Registration*

This work was done in collaboration with M. Miller, D. Tward, and B. Lee (Johns Hopkins University).

Mapping information from different brains gathered using different modalities into a common coordinate space corresponding to a reference brain is an aspirational goal in modern neuroscience. In a continuing collaboration with colleagues at Johns Hopkins University, we have solved this problem in its full generality by developing and implementing a rigorous, nonparametric generative framework, which learns unknown mappings between contrast mechanisms from data and infers missing data. A quantitative, scalable, and streamlined workflow was developed for unifying a broad spectrum of multimodal whole-brain light microscopic data volumes into a coordinate-based atlas framework. This registration pipeline was successfully applied to

multimodal data sets across species and proved robust to a variety of artifacts and data distortions, similar to the capabilities of a human neuroanatomist.

## Semantics Segmentation

This work was done in collaboration with J. Huang (CSHL); J. Jayakumar (India Institute of Technology, Madras); and Y. Wang, D. Wang, and L. Magee (Ohio State University).

Understanding the morphology and the connectivity of neurons is an important step in determining the neural circuitry of the brain. A computational framework was developed for a systematic treatment of the semantic segmentation problem for neuroanatomical image data. Within this framework, supervised learning–based segmentation of neurites (axons and dendrites) was combined with topological data analysis using discrete Morse (DM) theory in whole-brain histology sections. The method was shown to outperform state-of-the-art deep nets, with precision/recall generally of >90%.

## Topological Skeletonization

This work was done in collaboration with Y. Wang, D. Wang, and L. Magee (Ohio State University).

To capture the topological features of tree-like shapes of neurons, a novel skeletonization method based on DM theory was developed and applied to extract the tree skeletons of individual neurons from volumetric brain image data and to summarize collections of neurons labeled by tracer injections. The conceptually elegant DM approach lacks hand-tuned parameters and captures global properties of the data—as opposed to previous approaches that are inherently local. For individual skeletonization of sparsely labeled neurons, substantial performance gains were noted over state-of-the-art nontopological methods. The consensus-tree summary of tracer injections incorporates the regional connectivity matrix information, but in addition captures the collective collateral branching patterns of the set of neurons connected to the injection site and provides a bridge between single-neuron morphology and tracer-injection data.

## Informatics/Web Portal

In 2019, the Brain Architecture data portal (see www.brainarchitecture.org) continued to expand its database to incorporate cell type–specific data and

marmoset whole-brain connectivity data, in addition to the previous mouse brain connectivity data. For cell type–specific data, a portal for BICCN projects (see www.brainarchitecture.org/cell-type) has been serving cell projection data and cell distribution data collected using serial two-photon microscopy, as well as snRNAseq data collected in the Mitra laboratory. The marmoset brain architecture portal (see www.marmoset.brainarchitecture.org) serves data collected in the RIKEN collaboration and in Dr. Marcello Rosa's laboratory at Monash University. Continuous streams of data sets have been added to the data portal as they were collected in various projects.

## BICCN-CSHL U19 Data Core

As one of the Data Cores of the NIH Brain Initiative Cell Census Network (BICCN), the Mitra laboratory performed (1) data ownership, stewardship, and metadata management; (2) high-performance data analytics development; and (3) coordination with the BICCN data archives for data sharing and format standardization.

## Virtual Pathologist Project

This work was done in collaboration with A. Kepecs (CSHL); J. Crawford and M. Nassim (Northwell); B. Gallas and Q. Gong (FDA); and C. Abbey (UCSB).

With exponential growth in AI/machine learning, there is significant interest in computer-aided diagnoses (CADs). The Virtual Pathologist project was designed as a system for acquiring extensive behavioral data from a pathologist as they read a physical slide on a microscope, which constitutes their standard working environment. The system for capturing this data is based on the FDA's Evaluation Environment for Digital and Analog Pathology (eeDAP) platform running on a CSHL workstation. The goals of the project are to better understand how pathologists query a vast amount of data available to them on slides containing thin tissue sections stained according to the standards of the indicated medical exam (breast tissue assessment, GI tissue assessment, etc.), and then evaluate machine learning approaches using the same information as image annotations.

In 2019, the experimental protocols were refined and finalized based on trial runs of the study. An interactive web-based tool was developed to facilitate initial review of the multimodal data captured using the eeDAP apparatus from a pathologist carrying out diagnostic tasks. The resulting platform is more

user-friendly than the original eeDAP tool and is suitable for usage by study coordinators and pathologists.

## THEORETICAL

### Understanding Overfitting Peaks in Generalization Error: Analytical Risk Curves for $l_2$ and $l_1$ Penalized Interpolation

Traditionally, in regression one minimizes the number of fitting parameters or uses smoothing/regularization to trade training (TE) and generalization error (GE). Driving TE to zero by increasing fitting degrees of freedom (dof) is expected to increase GE. However, modern big-data approaches, including deep nets, seem to overparameterize and send TE to zero (data interpolation) without impacting GE. In previous theoretical work with M. Belkin and D. Hsu, an explanatory framework was proposed for this phenomenon, by introducing data interpolation algorithms (such as the weighted interpolating nearest neighbor algorithm) that can generalize well even in the presence of noise. In subsequent work, Belkin and collaborators demonstrated a related phenomenon called "double descent," showing that increasing the number of parameters beyond the data interpolation point can lead to a decrease in generalization error after an initial increase. A peak or divergence in generalization error is observed precisely at the data interpolation point. Belkin and colleagues suggested that this overfitting peak separates a classical regime from a modern (interpolating) regime in which overparameterization improves performance.

We challenged this picture by introducing a generative and fitting model pair (misparameterized sparse regression or MiSpaR) that dissociates data interpolation by overparameterization from model overspecification. Data interpolation occurs when the fitting model has enough parameters to fit the observations. However, this by itself does not guarantee that the fitting model has enough flexibility to specify the underlying data generative model. We suggest that "good interpolation" only occurs when the model is well specified or overspecified, something that cannot be guaranteed by data interpolation per se. The phenomena involved were studied within the MiSpaR model using $l_2$ and $l_1$ regularization terms, using random matrix theory and methods from statistical physics to obtain analytical formulas for the generalization error curves in the high-dimensional limit. The resulting GE curves show that the point at which proper model specification occurs (by increasing fitting parameters) does not, in general, coincide with the data interpolation point at which the overfitting peak occurs. Thus, our work further clarifies the phenomena surrounding overparameterized model fitting as present for deep networks.

## PUBLICATIONS

Griffin AD, Turtzo LC, Parikh GY, Tolpygo A, Lodato Z, Moses AD, Nair G, Perl DP, Edwards NA, Dardzinski BJ, et al. 2019. Traumatic microbleeds suggest vascular injury and predict disability in traumatic brain injury. *Brain* **142:** 3550–3564. doi:10.1093/brain/awz290

Huo B-X, Zeater N, Lin MK, Takahashi YS, Hanada M, Nagashima J, Lee BC, Hata J, Zaheer A, Grünert U, et al. 2019. Relation of koniocellular layers of dorsal lateral geniculate to inferior pulvinar nuclei in common marmosets. *Eur J Neurosci* **282:** 20–14. doi:10.1111/ejn.14529

Kleinfeld D, Luan L, Mitra PP, Robinson JT, Sarpeshkar R, Shepard K, Xie C, Harris TD. 2019. Can one concurrently record electrical spikes from every neuron in a mammalian brain? *Neuron* **103:** 1005–1015. doi:10.1016/j.neuron.2019.08.011

Miller CT, Hale ME, Okano H, Okabe S, Mitra P. 2019. Comparative principles for next-generation neuroscience. *Front Behav Neurosci* **13:** 12. doi:10.3389/fnbeh.2019.00012

Mitra PP. 2019. Understanding overfitting peaks in generalization error: analytical risk curves for $l_2$ and $l_1$ penalized interpolation. arXiv 1906.03667.

Ramezanali M, Mitra PP, Sengupta AM. 2019. Critical behavior and universality classes for an algorithmic phase transition in sparse reconstruction. *J Stat Phys* **175:** 764–788. doi:10.1007/s10955-019-02292-6

Tegling E, Mitra P, Sandberg H, Bamieh B. 2019. On fundamental limitations of dynamic feedback control in regular large-scale networks. *IEEE Trans Automat Contr* **64:** 4936–4951. doi:10.1109/TAC.2019.2909811

Tward DJ, Mitra PP, Miller MI. 2019. Estimating diffeomorphic mappings between templates and noisy data: variance bounds on the estimated canonical volume form. *Q Appl Math* **77:** 467–488. doi:10.1090/qam/1527

Tward D, Li X, Huo B-X, Lee B, Mitra P, Miller M. 2019. 3D mapping of serial histology sections with anomalies using a novel robust deformable registration algorithm. In *Multimodal brain image analysis and mathematical foundations of computational anatomy*, pp. 162–173. Springer, Cham.

Wang X, Tucciarone J, Jiang S, Yin F, Wang B-S, Wang D, Jia Y, Jia X, Li Y, Yang T, et al. 2019. Genetic single neuron anatomy reveals fine granularity of cortical axo-axonic cells. *Cell Rep* **26:** 3145–3159.e5. doi:10.1016/j.celrep.2019.02.040

Zaki A, Mitra PP, Rasmussen LK, Chatterjee S. 2019. Estimate exchange over network is good for distributed hard thresholding pursuit. *Signal Processing* **156:** 1–11. doi:10.1016/j.sigpro.2018.10.010

*In Press*

Tward D, Brown T, Kageyama Y, Patel J, Hou Z, Mori S, Albert M, Troncoso J, Miller M. 2020. Diffeomorphic registration with intensity transformation and missing data: application to 3D digital pathology of Alzheimer's disease. *Front Neurosci* **14:** 52. doi:10.3389/fnins.2020.00052

# HOW BRAINS ARE BUILT: LEARNING THE PRINCIPLES OF BRAIN ARCHITECTURE TO UNDERSTAND COMPLEX BEHAVIORS AND NEURODEVELOPMENTAL DISORDERS

**P. Osten**   A. Bludova   C. Elowsky   R. Muñoz-Castañeda   R. Palaniswamy   X. Qi
             R. Drewes    J. Mizrachi  A. Narasimhan                J. Palmer       J. Taranda

Despite more than a century of neuroanatomical efforts, the relationship between brain structure and its function remains more poorly understood than for any other system in the human body. This has been a significant barrier to understanding not only the basic principles of brain functions, but also how to treat or modify neurological diseases. The goal of our laboratory is to overcome this barrier by constructing a comprehensive anatomical model of the mammalian brain, including the classification and quantitative description of all its cell types and all its connections. We accomplish this by pairing our new microscopy and computational methods with genetic and behavioral manipulations in mice and other rodents to create atlases of differing scales that can be used to quantify changes in brain structure arising in evolution and in different disease states. Recent progress in the laboratory spans technological advances, new paradigms for integrating mesoscale and super-resolution imaging, and discoveries of structural underpinnings in neurodevelopmental genetic disease models.

## Mapping the Brain at Cellular and Super-Resolution Scale

Over the last six years our laboratory has been a part of a large National Institutes of Health (NIH)-sponsored effort, named BRAIN Initiative Cell Census Network (BICCN), which brings together more than two dozen research laboratories from across the United States to work collaboratively on creating a detailed cell type–based understanding of the mouse, marmoset, and human brain. Our laboratory has been making major contributions to this effort, both by developing new microscopes and data analysis methods and by applying these new tools to rigorous, large-scale data production.

### Research Program 1. Technologies for Automated Whole-Brain "Mesoscale" Imaging and Embedded Volumetric Super-Resolution Imaging

We have developed two major microscopy methods for imaging brain tissue: serial two-photon tomography (STPT), which we introduced in 2012 (Ragan et al., *Nat Methods* **9:** 255 [2012]; Kim et al., *Cell Rep* **10:** 292 [2015]; Kim et al., *Cell* **171:** 456 [2017]), and oblique light-sheet tomography (OLST), which we first described in 2017 (Narasimhan et al., bioRxiv 132423 [2017]). Although these instruments are used heavily in data production for brain atlasing (described below), we have also continued to innovate on the more recent OLST technology by improving the spatial resolution in a second-generation OLST v2.0 instrument as well as by introducing an option to combine whole-brain imaging by OLST with super-resolution imaging for selected brain areas.

The project focusing on developing a super-resolution capacity for OLST imaging, named OLST[SR], has been led by recent SBU graduate Judith Mizrachi working together with computational science analyst Xiaoli Qi, who built the OLST v2.0 instrument. In this work, Mizrachi used a computational approach, called super-resolution optical fluctuation (SOFI) imaging, for the calculation of super-resolved images from a recorded image time series by OLST (Fig. 1). This method, once fully implemented and integrated in the OLST v2.0 instrument, will enable highly novel studies of the relationships between local brain features visualized at super-resolution, such as the postsynaptic distribution of glutamate receptors representing synaptic strength, and whole-brain tracing of long-range connectivity of the same cells.
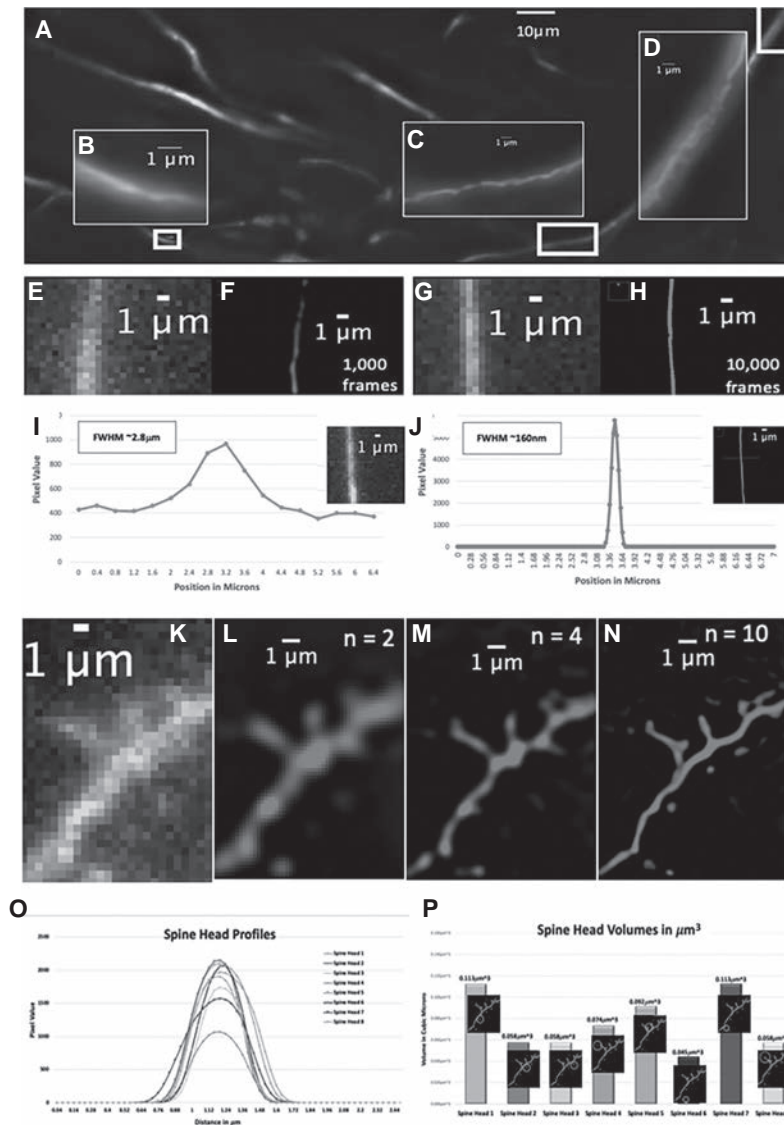
**Figure 1.** SR-OLST examination of cortical axons and dendritic spines. (*A–D*) Visualization of super-resolution (zoom-in) and mesoscale channels. (*C,D*) Zoom-in of boxed regions in *A*. (*F,H*) Resolution improvement with SOFI 10th-order cumulant analysis of the corresponding OLST images in *E* and *G* with 1,000 and 10,000 frame time series stacks. Spectral analysis before (*I*) and after (*J*) super-resolution. The cortical axon's diameter in this region is ~160 nm, as shown in *J*. (*K–N*) Resolution improvement of dendritic spines with SOFI cumulant orders 2–10. Spine head morphologies are analyzed by cross-sectional profiles (*O*) and spine head volumes (*P*) of indicated regions. These spine heads have cross-sectional diameters of ~0.4 µm and volumes of ~0.04–0.12 µm³. The average spine head volume is ~0.076 µm³.

### Research Program 2. Cell Type–Based Atlasing of the Mouse Brain

Our BRAIN Initiative project "Collaboratory for atlasing cell type anatomy in the female and male mouse brain" currently focuses on building a neuronal and glia-based cell type atlas of the mouse brain, including

mapping the distribution and ratios of brain cell types and their wiring into neuronal circuits that underlie the vast diversity of mammalian behaviors. To date we have mapped the brain-wide distribution of more than 30 cell types in the male and female brains by STPT, comprising more than 400 whole-brain data
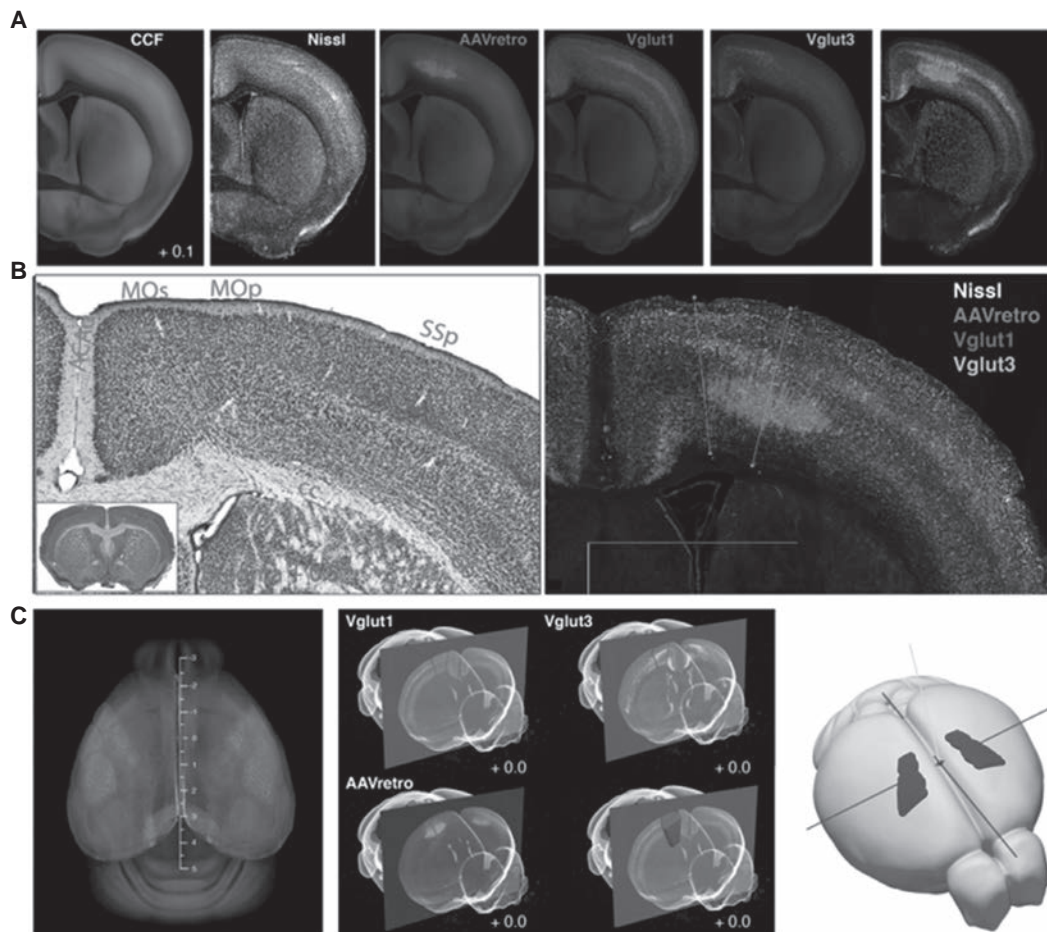
**Figure 2.** Delineation of upper limb area of the primary motor cortex (MOp-ul) using cloud-based Neuroglancer web platform. (*A*) Multiple anatomical data sets are coregistered in Neuroglancer for collaborative data analysis. (*B*) Nissl, cell type, and tracing data sets used for delineation of the MOp-ul area shown in purple lines. (*C*) 3D delineation of the MOp-ul area in the mouse brain.

sets. In addition, we have imaged 15 mouse brains at high spatial resolution by OLST, generating an atlas of more than 20,000 cortical neuron morphologies. Finally, we have built a cloud-based computational pipeline for collaborative data analysis that is being used by the BICCN community for delineating and refining anatomical structures in the mouse brain (Fig. 2). Taken together, these data sets offer an unprecedented wealth of quantitative information about the brain-wide distribution of neuronal and glia cell types, representing a highly unique resource for the neuroscience community that is accessible via our web portal at http://mouse.brainarchitecture.org/ost/—which was built in a collaboration with Partha Mitra's laboratory at CSHL.

## The Structure and Function of Neural Circuits Involved in Social Bond Formation

A second area of research interest in our laboratory focuses on the study of the neurobiology of social attachment—social bond forming—in monogamous prairie voles (this work is done in a collaboration with the laboratory of Steven Phelps at University of Texas, Austin). As this rodent species has not been previously studied with modern tools of systems neuroscience, we first developed a pipeline of imaging and computational methods for analysis of brain anatomy and function, a project led by postdoctoral researcher Rodrigo Muñoz-Castañeda. As a next step, our colleagues in the Phelps laboratory carried out a detailed time course during the behavior
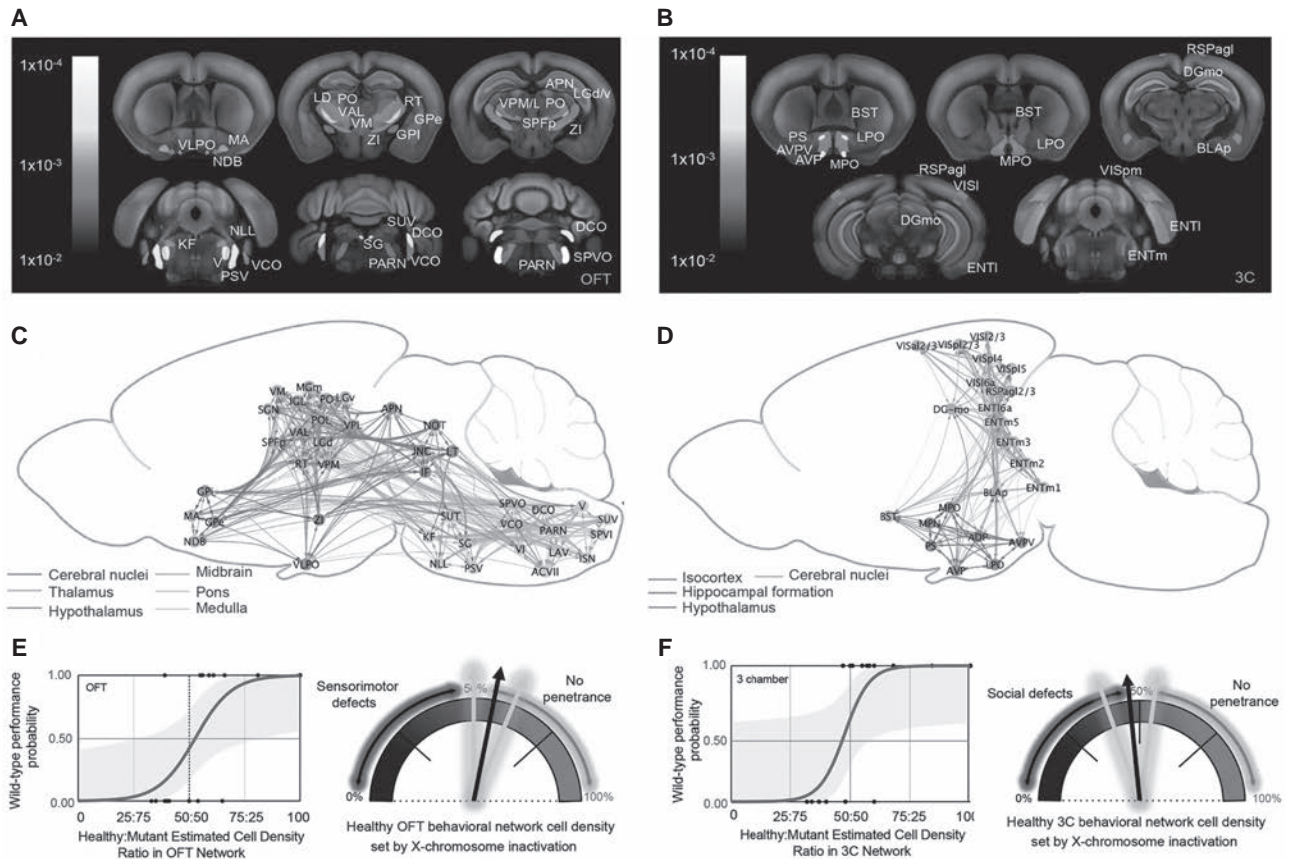
**Figure 3.** Linking brain circuits to disease symptoms in fragile X syndrome. **(***A*,*B***)** Identification of distinct sets of brain areas (highlighted in heatmap by statistical significance) linked to sensorimotor (*A*) and social behavior (*B*) deficits in a mouse model of fragile X syndrome (FXS). (*C*,*D*) Computational analysis of whole-brain connectivity reveals that these two sets of brain represent two distinct brain networks. (*E*,*F*) Computational analysis of the ratios of the mutant versus healthy X chromosome distribution reveals that only ~55% density of the mutant allele is needed for disease penetrance.

and collected the brains of these animals at 1, 3, 6, and 24 hours after initiating the formation of a stable social bond. Muñoz-Castañeda then used our whole-brain mapping pipeline, based on detection of the immediate early gene *c-fos* (Kim et al*., Cell Rep* **10:** 292 [2015]), to identify brain areas activated at the different time points—generating the first comprehensive map of brain structure to function relationships underlying this highly sophisticated and evolutionary significant social behavior.

## Deciphering Mouse Models of Human Neurodevelopmental and Psychiatric Disorders

Finally, a third major area of research interest in our laboratory is the study of vulnerable brain circuits affected in genetic neurodevelopmental disorders. This research area was a major motivation for the development of the whole-brain imaging and analysis methods described above.

In a first application of our methods to a clinically relevant question, we have generated the first whole-brain map of X-chromosome inactivation (XCI) —a process by which one X chromosome is inactivated in the female brain. This work showed that a combination of systematic bias in XCI toward higher silencing of the X chromosome inherited from the father and local stochastic variation of XCI per brain regions significantly influences the penetrance and severity of disease symptoms for the X-linked fragile X syndrome (Fig. 3), providing an important etiological insight into the source of phenotypic variability in human X-linked disorders.

In a second study focused on genetic neurodevelopmental disorders, postdoctoral researcher Julian Taranda identified a unique condition of an incomplete penetrance of structural, epileptiform, and behavioral phenotypes in a mouse model of 16p11.2 deletion, a recurrent copy number variation (CNV) linked to developmental delays, intellectual disability, autism, and childhood seizures. Taranda was able to show that approximately half of the genetically identical 16p11.2 deletion (16p11.2 del/+) mice display a number of prominent phenotypes, including spontaneous epileptiform episodes of cortical activity, increased propensity to convulsant-evoked seizures, pronounced volume reductions in cortical areas correlating with increased convulsant-evoked local cortical activity, disrupted sleep, hyperactivity, and increased repetitive behaviors. In contrast, the remaining 16p11.2 del/+ mice showed only moderate brain volume changes and hyperactivity, but otherwise appeared phenotypically normal. The phenotypic discordance was observed in isogenic 16p11.2 del/+ mice within the same litter and across multiple generations and parents; this strongly implies that the clustered phenotypes share a common mechanism of origin and their divergence reflects a bifurcating developmental choice occurring stochastically in the presence of the CNV, without the requirement for secondary mutations or environmental factors. Our

current efforts focus on understanding these stochastic processes in a mechanistic way by carrying out a detailed developmental analysis of changes in RNA expression in the brain in the 16p11.2 del/+ affected, unaffected, and wild-type littermates.

Motivated by the methodological difficulties in quantifying cellular phenotypes across the entire mammalian brain, our laboratory has pioneered a series of high-throughput and high-resolution methods to enable new discoveries relevant to brain circuit functions, gender dimorphism, social behavior, brain evolution, and abnormal neurodevelopment caused by human genetic risk factors.

## PUBLICATIONS

Mandelbaum G, Taranda J, Haynes TM, Hochbaum DR, Huang KW, Hyun M, Umadevi Venkataraju K, Straub C, Wang W, Robertson K, et al. 2019. Distinct cortical-thalamic-striatal circuits through the parafascicular nucleus. *Neuron* **102:** 636–652 e637. doi:10.1016/j.neuron.2019.02.035

*In Press*

Němec P, Osten P. 2020. The evolution of brain structure captured in stereotyped cell count and cell type distributions. *Curr Opin Neurobiol* **60:** 176–183. doi:10.1016/j.conb.2019.12.005

Ueda HR, Dodt H-U, Osten P, Economo MN, Chandrashekar J, Keller PJ. 2020. Whole-brain profiling of cells and circuits in mammals by tissue clearing and light-sheet microscopy. *Neuron* **106:** 369–387. doi:10.1016/j.neuron.2020.03.004

# NEURAL CIRCUITRY FOR SOCIAL COMMUNICATION

**S. Shea**    A. Corona    C. Kelahan    A. Pagliaro    L. Shen
          R. Dvorkin    A. Nowlan    D. Rupert    Y. Xie

The broad goal of our laboratory's research is to understand how the brain detects and interprets sensory stimuli to guide flexible behavior. We are particularly interested in how neural activity and plasticity in olfactory and auditory brain circuits facilitate communication and social behavior. We are revealing neural mechanisms that allow organisms to detect and recognize familiar individuals, to gather information about their identity and social status, and to select appropriate behaviors. The smells and sounds emitted by a mouse during a social encounter convey a remarkable amount of information to its partners in the encounter. For example, they can signal the mouse's sex, genetic identity, reproductive state, levels of distress, or sexual interest. Their friends can even determine what food they recently ate and whether it was good. As you might imagine, proper interpretation of social signals is indispensable for survival and mating success.

This research program has two broad intellectual goals. First, we want to identify the fundamental principles that govern how the brain adaptively controls complex behavior. At the core of this capacity is an interplay between innate predispositions and experience. Therefore, most organisms are endowed with a menu of species-typical behaviors and also possess flexible control over when to implement these behaviors and the choice of targets. Natural social behaviors are well suited for this first goal because they are fundamental to survival, sculpted by evolution, and malleable to experience. Second, we hope to pinpoint and repair neural circuitry defects that impair appropriate use of social information. Difficulty with social perception and cognition are core features of the autism spectrum disorders (ASDs). For example, patients may have trouble perceiving and interpreting communication gestures such as speech, facial expressions, and "body language." This broad feature is also evident in many mice that carry genetic variants identified in human ASD populations. Therefore, if we can ascertain the neural circuit substrates of social behavior in normal mice, we can make and test predictions for how the

circuitry is affected in the mouse models. The results are likely to tell us more about the synaptic modifications that occur in human autism.

Increasingly, our research is focusing on not just the detection of sensory information related to social behavior, but on how that information is combined and interfaced with behavioral decisions. This is often achieved by fluctuating levels of neuromodulators and hormones that modify large-scale electrical activity patterns in the brain. We are particularly interested in how all stages of neural processing are flexible according to experience and behavioral state. Consequently, experiments in our laboratory frequently involve observing and manipulating brain activity online during ongoing free interactions between mice.

## Vocal Communication between Mothers and Pups Regulates Maternal Behavior

For several decades, we have understood that mice are continuously "speaking" (or vocalizing) to one another in a "language" that we have only just begun to understand. Many distinct vocalization types are produced by males, females, juveniles, and adults in a range of behavioral contexts. We are working to understand the meaning of these calls and how they are used to guide behavioral choices. For example, to properly interpret the message these calls are intended to convey, mice must possess the neural circuitry to distinguish between different messages and link each of them to the appropriate behavioral responses. Identifying and monitoring this circuitry and its plasticity across development and life events is a major goal of our research program.

One form of vocalization for which we have a solid understanding of both the message and the appropriate response is the ultrasonic distress vocalization (USV). Young mice prior to vision and full mobility will occasionally become separated from the nest. This is stressful for them, and they will therefore call out

to their mother with a very high frequency squeak. New mothers develop sensitivity to these cries and respond by moving toward their source to retrieve the pup. Moreover, females who have never given birth do not innately show approach responses to pups or their calls; however, they can learn to perform the behavior with experience. We refer to this group as "surrogates." Accurate maternal learning by mothers and surrogates is suspected to involve rewiring (or "plasticity") in the auditory cortex.

## Auditory Plasticity Sharpens Vocal Perception during Parental Learning

L. Shen

We are working to understand how mice discriminate between different vocalization types and distinguish them from nonvocal sounds. Mice are constantly surrounded by many sounds, some of which may share pitch or timing features with important vocal signals. To which features do they attend? How much tolerance do they have for variability in those features? And do these limits change after maternal experience? Graduate student Luqun Shen has developed an innovative, high-throughput behavioral paradigm that combines our focus on natural communication behavior with techniques developed by some of our CSHL colleagues to systematically probe behavioral responses to many stimuli. In this behavioral task, Luqun teaches female mice that if they hear a pup call, they can lick a spout for a water reward, but if they hear a very different sound, such as a synthetic beep, they should withhold licking or they will be subjected to a "time out." Once the mouse has learned this rule, Luqun will "ask it questions" by periodically sneaking in a novel and ambiguous stimulus. This way he can learn whether the mouse responds to the stimulus as if it were a call or not. He then compares the responses of female mice that have no experience caring for pups ("naïve") with mice that have cohabited with a mother and her pups ("surrogate"). Figure 1 shows data from these experiments comparing responses to some of these ambiguous stimuli between naïve and surrogate females. Note that naïve mice respond more strongly to nonpup call stimuli than surrogate mice, showing that mice become more picky and selective about which sounds they respond to. In parallel, Luqun is recording from individual neurons in a part of the brain that is important for detecting vocalizations (auditory cortex) in naïve and surrogate mice. Early data suggest that neural responses to pup calls also become more distinct after maternal experience.
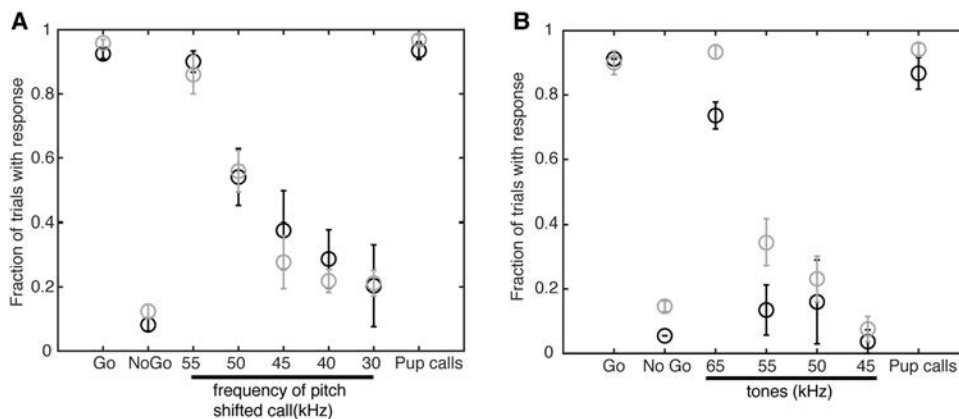


**Figure 1.** Female mice show systematic changes in behavioral responses to parametrically modified pup call and tone stimuli. Naïve and surrogate female mice were trained on a head-fixed Go/No Go behavioral task to respond to pup calls by licking a spout for a water reward, and to withhold licks in response to a low synthetic tone. They were intermittently given unrewarded catch trials which consisted of parametrically modified tones and pup calls to measure their licking behavior to these ambiguous stimuli. The fraction of trials on which they licked to one of these stimuli was taken as an indication of how similar the mouse judged that stimulus to be to a pup call. (*A*) In addition to the trained pup calls (Go) and the No Go stimulus, mice were presented with unfamiliar pup calls (*right*) and pitch-shifted pup calls digitally reduced to the denoted mean frequency. (*B*) This set of experiments was identical, but instead of pitch-shifted calls, the mice were presented with synthetic pure tones of the denoted frequency. Note that surrogate females were less likely to respond to the pure tones than naïve females.

## Father Mice Show Greater Behavioral and Neural Variability Than Mother Mice

A. Corona

The vast majority of studies of parental behavior in mice have focused entirely on females and maternal behavior. Nevertheless, graduate student Alberto Corona has shown that males exhibit paternal behavior. He is trying to identify brain regions in males that overlap with or are distinct from circuits that govern maternal behavior. Therefore, Alberto is performing behavioral experiments and whole-brain imaging (in collaboration with P. Osten, CSHL) from male and female mice to screen for paternal behavior–specific neural activity. These experiments have identified regions such as the anterior cingulate cortex and the bed nucleus of the stria terminalis that exhibit activity specifically during retrieval behavior. However, Alberto has made another interesting discovery. He has found that fathers are much less reliable than mothers in their performance of parental behavior, and this variability is mirrored by variability in neural activation. Fathers performing a retrieval behavior assay are much less reliable than mothers, especially when challenged by being placed into an unfamiliar cage. Our whole-brain imaging experiments show that the number of activated cells across the whole brain of fathers performing parental behavior is highly variable. Interestingly, this neural variability is highly correlated with variability in attention to the pups.

## Auditory Plasticity and Maternal Behavior Are Impaired in a Mouse Model of Rett Syndrome

D. Rupert, A. Pagliaro

Consistent with our objective of identifying impairments in neural circuitry that underlie social communication difficulties in ASD, former postdoctoral fellows Billy Lau and Keerthi Krishnan led a collaboration with CSHL professor Dr. Josh Huang on vocal perception in a mouse model of Rett syndrome. This collaboration resulted in its second publication late last year. Rett syndrome is caused by loss of function of a gene called *Mecp2*. Female mice that possess only a single copy of *Mecp2* are not able to develop proficiency at gathering pups. We previously showed that this likely happens because MeCP2 (the protein product of the gene) plays

a critical role in maintenance and plasticity of the auditory cortex by acting on inhibitory interneurons.

Based on changes in the pattern of the expression of certain molecular markers that we observed after maternal experience, we predicted that a network of inhibitory neurons expressing a protein called parvalbumin (PV) are central to auditory cortical plasticity. Notably, deletion of *MECP2* only in this minority of neurons is sufficient to disrupt pup care. Therefore, Billy made neuronal recordings in awake behaving females of both genotypes that differed in their maternal experience. The data show that when a normal adult female mouse is first exposed to pups, her auditory cortex becomes "disinhibited" (i.e., there is suppression of the inhibitory network). Interestingly, we found that the removal of inhibition was selective for deep layers of the cortex, preferentially acted on late components of inhibition, and only affected responses to behaviorally meaningful calls, as opposed to synthetic tones. In contrast to typical mice, we observed no disinhibition in *Mecp2*-deficient mice. The results of this study are consistent with our model that MeCP2 regulates plasticity in adults and juveniles through its effects on inhibitory interneurons. The article describing these findings was recently published in the *Journal of Neuroscience* (Lau et al. 2020).

Graduate student Deborah Rupert is following up on this work by performing behavior assays, making neural recordings, and examining marker expression in mice that have had *MECP2* deleted only in PV neurons. Her goal is to assess whether these mice exhibit the same deficits that the mice lacking *MECP2* in all cells exhibit. If so, that would strongly suggest that the lack of *MECP2* in PV neurons–specifically is the cause of these deficits. PV-specific knockout mice exhibit the same impairments in behavioral performance, neuronal activity, and marker expression as do the comprehensive mutants. For example, Figure 2 shows a comparison of the disinhibitory changes in the responses of PV neurons to pup calls in nonmutant mice, nonspecific mutants, and PV neuron–specific mutants. Although the nonmutants show the expected reduction in PV neuron activity following maternal experience, neither the nonspecific mutants nor the PV neuron–specific mutants show evidence of disinhibition. Deborah has also found, in preliminary experiments, that restricting deletion of *mecp2* to a different type of interneuron called somatostatin (SST) neurons does not cause behavioral performance deficits,
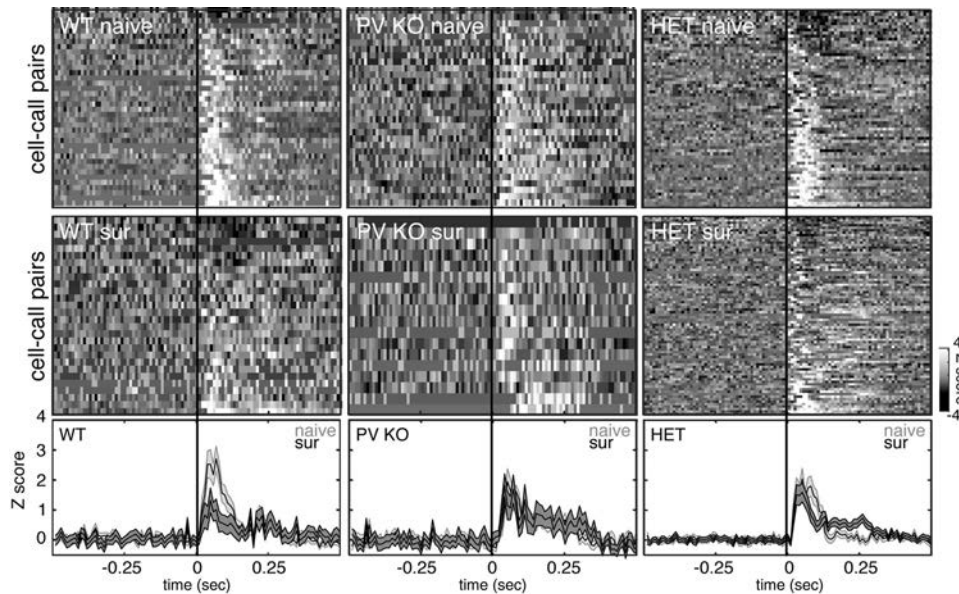
**Figure 2.** Parvalbumin (PV) neuron output is reduced after maternal experience in WT females but not in mice that have had *Mecp2* deleted throughout the brain or selectively in PV neurons. (*Top* panels) Heatmaps depicting *Z*-score-normalized responses for all PV cell-call pairs in naïve and surrogate female WT, PV-specific *Mecp2* knockout, and complete heterozygous *Mecp2* knockout mice. Each row in the heatmap represents a peristimulus time histogram (PSTH) of the mean response for a cell-call pair (bin size, 10 msec). Rows are sorted by response magnitude from least to most excitatory. (*Bottom* panels) Mean traces for all cell-call pairs for each genotype comparing naïve females (gray) and surrogate females (black). Center and outside lines represent mean and SEM, respectively.

further underscoring the central role of PV neurons for maternal behavior impairments in the mouse model of Rett syndrome. Another graduate student, Alexa Pagliaro, is developing approaches using miniaturized head-mounted microscopes to observe the activity of PV neurons during active pup retrieval.

## Multisensory Integration in Maternal Retrieval Behavior

A. Nowlan, C. Kelahan

Social encounters are inherently multisensory experiences. Communication from other mice is carried by auditory vocal signals and social odors, as well as potentially other modalities. Accordingly, pup retrieval behavior depends heavily on both vocalizations and olfactory cues from the pups. Surprisingly, pup odors actually change neural processing of sounds by mothers, an example of what is known as multisensory integration. Nevertheless, neither the functional significance of this integration nor the neural pathways involved are well understood. Pup odors may have short-term as well as long-term effects on how the auditory system

responds to pup calls. For example, pup odors may serve as an immediate contextual cue that elevates the behavioral significance of vocalizations so that mothers are more attentive to USVs. Another nonmutually exclusive possibility is that olfaction and audition work together to trigger long-term synaptic modifications in the auditory cortex that permanently alter neuronal responses to vocalizations. With assistance from technician Clancy Kelahan, graduate student Alexandra Nowlan has discovered a rich projection from the basal amygdala (BA) to the auditory cortex. This discovery allowed her to tag these neurons, despite the fact that they are comingled among other types of neurons, with tools to optically monitor and control their activity. Alexandra found that neurons in the BA that project to the auditory cortex are active during pup retrieval and likely convey information about odors to the auditory system. Cells in this pathway exhibit consistently elevated activity during olfactory investigation and search, and they respond with increased firing to odors, including pup odors (Fig. 3). In separate experiments, Alexandra used light to excite the connections of BA neurons in the auditory cortex and clearly showed that this dramatically affects the responses of auditory
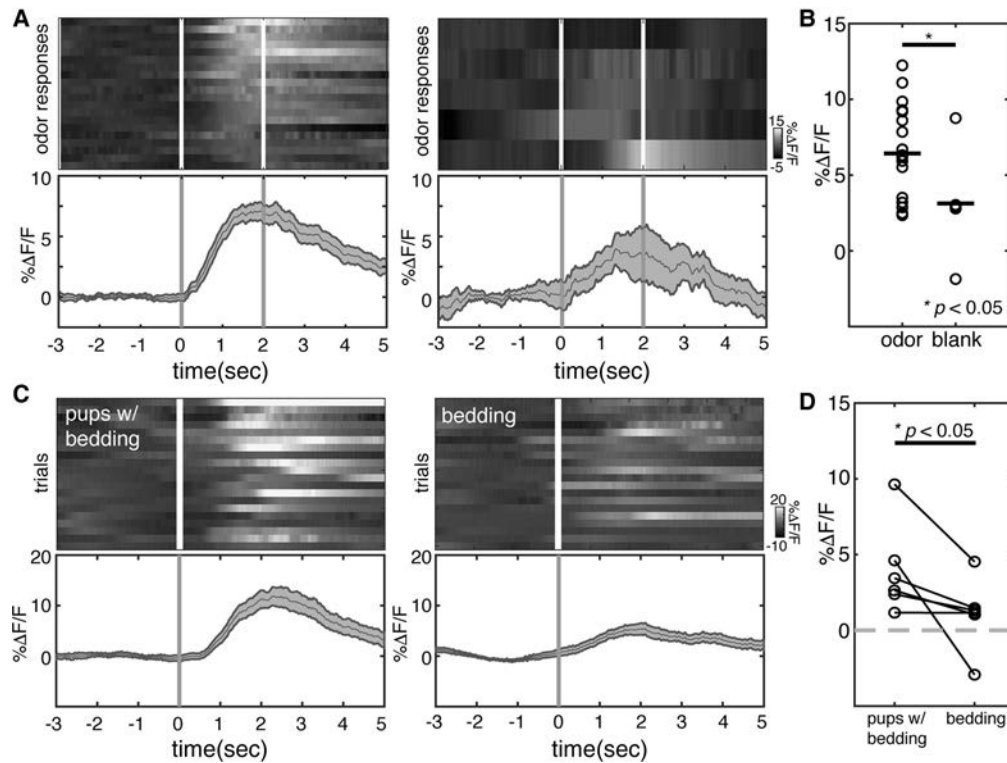
**Figure 3.** Basal amygdala (BA) neurons that project to the auditory cortex respond to pup and other odors. (*A*) Mean responses of AC-projecting BA neurons to multiple odors. The *top* panels are heatmaps in which each row is the mean response of one mouse to one odor. Five different odors from five different mice are depicted in the *top left* panel, sorted from the strongest to the weakest response. The *top right* panel shows the mean response to the blank (mineral oil control) for the same mice. The vertical bars show the time of odor presentation (0–2 sec). The gray traces below depict the mean and SEM of the responses. (*B*) Scatterplot comparing the mean strength of the responses to odors with the mean strength of the blank response. Odor responses were significantly stronger than blank responses ($p < 0.05$, Mann–Whitney *U*-test). (*C*) Responses of AC-projecting BA neurons to 20 trials of pup odor presentation. The *top* panels are heatmaps in which each row is the response on one trial to either pups with clean bedding (*top left*) or clean bedding alone as a control (*top right*). The vertical lines show the onset of odor presentation. The gray traces below depict the mean and SEM of the responses across trials. (*D*) Scatterplot comparing the mean amplitude of the response to pup odor with the mean amplitude of the response to bedding only for six different mice. Pup odor responses were significantly stronger than bedding responses ($p < 0.05$, paired *t*-test).

neurons to pup vocalizations. These data suggest that the BA may provide the auditory cortex with access to odor information during parental behavior. Taken together, all of these findings lead us to conclude that the BA facilitates multisensory integration of odors and sounds during maternal search and retrieval of pups.

## Neural Activity Signaling Emotion, Arousal, and Reward during Social Encounters

R. Dvorkin, Y. Xie

Organisms are constantly being bombarded by an overwhelming number and variety of stimuli from all

of their senses. Therefore, one of the greatest challenges faced by the nervous system is to make sensible and efficient choices about which stimuli to attend and remember. Decades of evidence have established that this calculation is achieved in large part with the contribution of a class of neurochemicals that enable neuronal communication and are collectively referred to as "neuromodulators." Neurotransmitters of this type generally do not participate in fast, moment-to-moment communication between neurons. Instead, their levels fluctuate more slowly and modify the function of larger groups of neurons or circuits. These properties place neuromodulatory systems in a central role with regard to interfacing between social stimuli and

behavioral decisions. Recent data from our laboratory suggest that the neuromodulators noradrenaline (NA) and dopamine (DA) play an important role in motivating and reinforcing pup retrieval behavior.

Postdoctoral fellow Roman Dvorkin is interested in a small nucleus in the brainstem called locus coeruleus (LC) that sends neuromodulatory signals throughout the brain by releasing the chemical NA. Evidence strongly suggests that in artificial, operant learning tasks, the activity of LC neurons both influences behavior and signals the appearance of behaviorally interesting stimuli. In parallel, LC activity is also closely associated with arousal and emotion, which is thought to be related to LC's role in social behavior. Nevertheless, there has been no direct observation of how these neurons participate in unstructured social interaction. Therefore, Roman has made both electrical recordings of individual neurons and optical recordings of neuronal populations in LC during social encounters between mice, including courtship and parental interactions. Roman has succeeded in observing LC neural activity, day after day, as the mouse has repeated social interactions using two independent methods. The first conclusion to emerge from this work is that each time a maternal caregiver retrieves a pup, there is a large, precisely timed burst of firing in LC neurons just as she makes contact with the pup (Fig. 4). High-resolution kinetic analysis of the mouse's subsequent trajectory back to the nest shows that LC population activity precedes and correlates very closely with the mouse's velocity. This suggests that the release of a bolus of the powerful neuromodulator noradrenaline may help motivate robust maternal behavior, and it underscores the profound emotional content of interactions with offspring.

Yunyao Xie, another postdoctoral fellow, is performing similar experiments recording from neurons that release the neurotransmitter dopamine. DA is very important for motivated behavior because it both stimulates movement and helps evaluate rewards. Central to dopamine's role in processing reward is its property of signaling reward prediction error. What this means is that dopamine does not signal reward per se, but rather the difference between expected and encountered
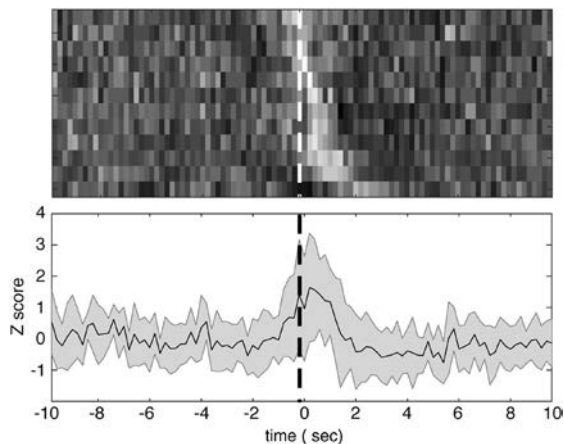


**Figure 4.** Locus coeruleus (LC) neurons consistently show precisely timed and transient firing events around the time of pup retrieval. The *top* panel is a heatmap in which each row is the *Z*-scored mean peristimulus time histogram (PSTH; bin size, 200 msec) for a different well isolated single unit from LC. Each response is sorted by latency to the moment the female lifts the pup. The *lower* panel is a trace of the mean of all 12 LC neurons. Center and outside lines represent mean and SD, respectively.

rewards. Therefore, when an organism receives an unexpected reward, DA neurons will fire briskly, reinforcing the action that led to that reward. As the organism continues to be rewarded for that action, the reward becomes expected, and DA neurons accordingly fire less. Yunyao has collected some very exciting preliminary data that suggest that DA neurons may also exhibit this property with respect to maternal retrieval of pups. Two observations support that hypothesis. First, firing in DA neurons in initial retrieval trials show a "ramp up" to retrieval that peaks just after the female lifts the pup, which may reflect anticipation. Second, as the female becomes proficient at retrieval, DA neurons begin to exhibit smaller responses to retrieval. We are planning a number of experiments in the near term to more directly test this idea.

Lau BY, Krishnan K, Huang ZJ, Shea SD. 2020. Maternal experience-dependent cortical plasticity in mice is circuit- and stimulus-specific and requires MeCP2. *J Neurosci* **40:** 1514–1526. doi:10.1523/JNEUROSCI.1964-19.2019

# HORMONAL REGULATION OF GENE EXPRESSION IN THE BRAIN

**J. Tollkuhn**    K. Broekman    B. Gegenhuber    M. Wu
                   R. Bronstein    J. Ruiz-Ortiz

Our laboratory seeks to understand the mechanisms that shape and regulate sex differences in the brain. Females and males differ in many behaviors and are differentially affected by mental health disorders, but the distinct developmental trajectories that give rise to these sex differences remain poorly understood. Much of our knowledge about the cellular and molecular differences between the sexes in the mammalian brain has been obtained through studies of the hormonal regulation of the differentiation and function of neural circuits underlying innate, sex-typical behaviors and physiology in rodents. Paradoxically, estrogen is required to both feminize and masculinize the brain. Males undergo a transient perinatal testosterone surge, and this circulating testosterone is converted to estradiol (the most abundant endogenous estrogen) locally in the brain. Treating females with estradiol at birth irreversibly masculinizes both adult behaviors and gene expression patterns, suggesting that perinatal estrogen directs gene regulatory events that organize persistent sex differences in the brain. The receptors for estrogen, estrogen receptor alpha (ERα) and estrogen receptor beta (ERβ), are nuclear receptor transcription factors that are recruited to DNA in the presence of hormone. Although the regulatory strategies used by estrogen receptors are well understood in the context of breast cancer, they remain obscure in the brain. Defining the gene programs regulated by estrogen in the developing and adult brain is the focus of our current research program. Our central hypothesis is that hormone-responsive genes underlie sex differences in the incidence and etiology of psychiatric and neurological diseases.

## Identification of Estrogen-Regulated Genes in the Brain

Although the effects of estrogen on behavior, neuroprotection, and mood are well established, the genes that ERα activates to produce these effects have remained unknown. We aim to understand how ERα regulates the expression of genes involved in these behaviors and ultimately whether dysregulation of sex-specific transcriptional programs contributes to the pathology of neurodevelopmental disease. To this end, we have identified the first direct targets for ERα in the brain, through the use of "cleavage under targets and release under nuclease" (CUT&RUN). Comparison to ChIP-seq data sets from peripheral tissue reveal that 60% of our ERα targets are unique to the brain. Brain-specific ERα target genes are associated with synapse organization and axon development; some of these genes, like *Maoa* and *Scn2a*, are implicated in neurodevelopmental and psychiatric disorders. Moreover, we show that ERα is the primary driver of sex differences in gene expression in the brain and cooperates with the NFI factor Nfix to define a male-biased neuronal population. This is the first description of neural gene regulation by any gonadal hormone receptor; our results reveal multiple mechanisms that estrogen activates to modulate brain function. We also anticipate that our approach may be broadly used to determine transcription factor binding sites in defined neuron populations throughout the brain, thereby providing functional links between genes and resulting behaviors.

One prototypical example of such a link is the subject of a recent collaboration with Holly Ingraham's laboratory at UCSF. It has been known for decades that females are more active when estrogen levels are high. The Ingraham laboratory identified melanocortin receptor 4 (Mc4r) as a direct regulator of physical activity. *Mc4r* is expressed in sparse regions of the hypothalamus and increases during proestrus, the highest estrogen phase of the estrus cycle. We found that *Mc4r* is a direct target of ERα. By designing a guide RNA to the estrogen response element in the *Mc4r* promoter, the Ingraham laboratory was able to employ CRISPRa to increase Mc4r function in a specific hypothalamic region, the VMHvl, in adult animals. Overexpression of this single ERα target was sufficient to increase locomotion in females, even females that have undergone ovariectomy and

otherwise have low activity levels. This exciting result is the first example of altered behavior from a targeted CRISPRa experiment in adult animals and highlights the power of identifying direct ERα binding sites.

## Identification of a Male-Specific Neuronal Population Using Single-Nucleus Sequencing

Males have more neurons than females in the bed nucleus of the stria terminalis (BNST), a key node in social behavior circuitry. This sex difference is conserved in humans, and in mice it is specified by neonatal estrogen—in the absence of locally derived estradiol the female BNST undergoes increased cell death. We have collaborated with the laboratory of Michael Wigler (CSHL) to determine the identity of the surviving neurons in males, with the long-term goal of manipulating gene expression and neural activity to alter sex-typical behaviors. Siran Li in the Wigler laboratory has recently developed a new droplet-based single-cell sequencing approach, BAG-seq, which is ideal for profiling of our limited neuronal populations. We optimized BAG-seq for neuronal nuclei and identified a cluster of GABAergic neurons that is overrepresented in males. We are now performing follow-up experiments to determine when in development this cluster arises, and if it is lost in ERα or Nfix mutant animals. In addition, we have found that several autism candidate genes show sex differences in gene expression in the BNST, which may underlie the increased susceptibility to developing this condition in males.

## Sex Differences in Stress Responses

As part of our analysis of sex differences in gene expression, we have found that the gene for glucocorticoid receptor (GR), *Nr3c1*, is expressed at higher levels in the female brain and that expression decreases when estrogen is high. This result may explain why females across species have more reactive behavioral stress responses compared to males, as well as the increased incidence of mood and anxiety disorders in women. To understand sex differences in GR function in the brain, we have initiated CUT&RUN studies in hippocampus and BNST. Our hypothesis is that GR occupies distinct subsets of targets in females during unstressed baseline conditions and activates region-specific genes to coordinate behavioral responses and homeostasis.

We are extending the line of inquiry into another rodent species, the prairie vole *Microtus ochrogaster*. Prairie voles have 10 to 20 times the levels of baseline corticosterone compared to other rodents, and it is thought that their hyperactive HPA access evolved in parallel with their unique prosocial behaviors. In humans, many patients with major depressive disorder (MDD) also have a hyperactive HPA axis, but the actions of GR in this hyperactive state are not known. We will determine whether social stress that results from separation from partners activates distinct gene programs compared to standard experimental stressors such as physical restraint. Collectively, these GR experiments will define the molecular signature of stress response in the brain.

### PUBLICATIONS

Gegenhuber B, Tollkuhn J. 2019. Signatures of sex: sex differences in gene expression in the vertebrate brain. *Wiley Interdiscip Rev Dev Biol* **20:** e348. doi:10.1002/wdev.348

Gegenhuber B, Tollkuhn J. 2019. Sex differences in the epigenome: a cause or consequence of sexual differentiation of the brain? *Genes* **10:** 432. doi:10.3390/genes10060432

Krause WC, Rodriguez R, Gegenhuber B, Matharu N, Rodriguez AN, Padilla AM, Herber CB, Correa SM, Ahituv N, Tollkuhn J, Ingraham HA. 2019. Estrogen drives melanocortin neurons to reduce sedentary behavior. bioRxiv 794792. doi:10.1101/794792

Li S, Kendall J, Park S, Wang Z, Alexander J, Moffitt A, Ranade N, Danyko C, Gegenhuber B, Fischer S, et al. 2019. Copolymerization of single-cell nucleic acids into balls of acrylamide gel. *Genome Res* **30:** 49–61. doi:10.1101/gr.253047.119

# CIRCUITS UNDERLYING DECISIONS IN AUDITORY CORTEX

**A. Zador**  V. Aguillon-Rodriguez   S. Ghosh      Y. Li         A. Vaughan        A. Zhang
          X. Chen            H. Gilbert    S. Lu         W. Wadolowski
          E. Fong            L. Huang      D. Maharjan   L. Yuan
          A. Funamizu        C. Krasniak   Y-C. Sun      H. Zhan

Our laboratory is interested in how neural circuits underlie cognition, with a focus on the cortical circuits underlying auditory processing and decision-making. To address these questions, we use a combination of molecular, electrophysiological, imaging, and behavioral approaches. In addition, we have developed a method that allows us to use high-throughput DNA sequencing to determine the wiring diagram of the brain at single-neuron resolution rapidly and efficiently, at low cost.

## The International Brain Laboratory
V. Aguillon-Rodriguez, C. Krasniak

Progress in neuroscience is hindered by poor reproducibility of mouse behavior. Here, we show that in a visual decision-making task, reproducibility can be achieved by automating the training protocol and by standardizing experimental hardware, software, and procedures. We trained 101 mice in this task across seven laboratories at six different research institutions in three countries and obtained three million mouse choices. In trained mice, variability in behavior between laboratories was indistinguishable from variability within laboratories. Psychometric curves showed no significant differences in visual threshold, bias, or lapse rates across laboratories. Moreover, mice across laboratories adopted similar strategies when stimulus location had asymmetrical probability that changed over time. We provide detailed instructions and open-source tools to set up and implement our method in other laboratories. These results establish a new standard for reproducibility of rodent behavior and provide accessible tools for the study of decision-making in mice.

## Citric Acid Water as an Alternative to Water Scheduling in Behaving Mice
V. Aguillon-Rodriguez [in collaboration with A.E. Urai, CSHL]

Powerful neural measurement and perturbation tools have positioned mice as an ideal model species for probing the neural circuit mechanisms of cognition. Crucial to this success is the ability to motivate animals to perform specific behaviors. One successful strategy is to schedule their water intake, rewarding them with water during a behavioral task. However, water scheduling requires rigorous monitoring of an animal's health and hydration status and can be challenging for some animals.

Here, we present an alternative that allows mice more control over their water intake—free home-cage access to water to which a small amount of citric acid (CA) has been added. CA makes water taste slightly sour but still palatable. Mice reduce their intake of CA water while still maintaining a healthy weight. Home-cage access to CA water has only subtle impacts on their willingness to perform a decision-making task, in which they are rewarded with sweetened water. CA water is thus a promising alternative that allows animals more control over their water intake without impacting behavioral performance.

## Shared Molecular Logic Underlying Long-Range Projections across Cortical Areas Revealed by In Situ Sequencing
X. Chen, Y-C. Sun

In the vertebrate brain, functional circuits are composed of neurons with enormous diversity in various neuronal characteristics, such as long-range projections and gene expression. Revealing the molecular signature of long-range projections in such a diverse population of neurons remains a central goal of neuroscience. Achieving this goal requires interrogating both projections and multiplexed gene expression in single neurons with high throughput. To solve this problem, we combine BARseq, a high-throughput projection mapping technique based on in situ barcode sequencing, and in situ sequencing of endogenous mRNAs. Using BARseq, we determined the

expression of classical cadherins, nonclustered proto-cadherins, and cell type markers in 117,397 cells and the projection patterns of 3,119 projection neurons in both the mouse primary motor cortex and auditory cortex. Further analyses on 1,381 neurons with both gene expression and projection information revealed a shared cadherin signature that correlates with homologous projections across the two cortical areas. By relating multiplexed gene expression to combinatorial projection patterns at cellular resolution with high throughput, BARseq provides a path to reveal the molecular logic underlying neuronal circuits.

## Neural Encoding and Decoding of Auditory Cortex during Perceptual Decision-Making
A. Funamizu

Neurons in auditory cortex encode auditory stimuli, but the precise encoding can depend strongly on task-relevant variables such as stimulus or reward expectation. This raises the question: If the cortical representation of the stimulus varies with task-relevant variables, how can areas downstream from auditory cortex decode these representations? One possibility is that decoding in downstream areas also depends on these task-relevant variables. To address this question, we developed a two-alternative choice auditory task for head-fixed mice in which we varied either reward expectation (by varying the amount of reward, in blocks) or stimulus expectation (by varying the probability of different stimuli). We then used calcium imaging to record populations of neurons in auditory cortex while mice performed the task. We found that varying either reward or stimulus expectation changed neural representations (i.e., stimulus encoding), sometimes dramatically. However, the optimal decoder was remarkably invariant to different encodings induced by different expectations. Our results suggest that stimuli encoded by auditory cortex can be reliably read out by downstream areas, even when the encoding is modulated by task-relevant contingencies.

## Corticostriatal Plasticity Underlying Learning and Reversal of Auditory-Motor Association in Mice
S. Ghosh

Animals use complex sensory cues from their environment to make a variety of decisions in their lives. Such behavior requires integration of sensory discrimination, decision-making, and appropriate motor actions. We used an auditory discrimination task to understand the brain circuits involved in such decision-making and how these circuits evolve during learning of sensorimotor associations. Previous studies from our laboratory have shown that the connections between auditory cortex and the auditory striatum in rats are instrumental for an animal to perform this task (Znamenskiy and Zador, *Nature* **497:** 482 [2013]). Moreover, learning of this discriminatory task results in formation of a memory of the learned association in this circuit (Xiong et al., *Nature* **521:** 348 [2015]). These findings suggest that parts of the striatum receiving predominantly sensory inputs might be involved in promoting contralateral movements. We tested this model in a reversal paradigm using the "tonecloud" task, in which an animal initially trained to form a specific auditory-motor association was then forced to reverse its association to obtain reward successfully. We find that mice can indeed learn to reverse these associations, taking comparable training times and reaching similar performances in both training epochs. We then investigated the pattern of plasticity in this circuit following the reversal and found that reversing the association does not result in a simple reversal of the memory trace. In fact, we observe a strong persistence of the synaptic plasticity pattern that only reflects the initial association. Given this observation, we then asked whether reversal trained animals can effectively revert to their original behavior when challenged with a second reversal. We find that this second reversal is not significantly faster than the first reversal. Our results are consistent with a model in which forming new memories does not erase previous ones. These results suggest that the sensory striatum might not be simply transducing sensory information into a contralateral motor output. Moreover, these findings raise questions regarding which other brain areas are involved in reversal learning of the tonecloud task and how task-related context information is integrated in this circuit.

## Hydrogel Grafting in Brain Slices
H. Gilbert

Embedding biological tissue in 3D polyacrylamide hydrogel networks allows the preservation of structural and molecular detail at nanometer resolution.

Because the hydrogel network is porous, it is possible to selectively remove objects from the system, which is the basis of the CLARITY method in which the removal of lipid and protein from thick samples (e.g., whole brain) generates an optically clear preparation that can be imaged at high resolution. Interestingly, a copolymer of acrylamide and acrylic acid can be isotropically swollen (3–20X) in water, which is the foundation of expansion microscopy (ExM), a form of super-resolution imaging. Both of these methods rely on standard radical polymerization of acrylamide/acrylate monomers. However, because the initiators that trigger polymerization are randomly distributed throughout the sample, it is not possible to drive gelation in a spatially restricted manner. In the last year, we have developed a technology that solves this problem by spatially restricted and controlled radical polymerization in brain slices. We are currently integrating this method with a scheme for spreading barcodes in virus-infected neurons.

## BRICseq Bridges Brain-Wide Interregional Connectivity to Neural Activity and Gene Expression in Single Animals

L. Huang

Comprehensive analysis of neuronal networks requires brain-wide measurement of connectivity, activity, and gene expression. Although high-throughput methods are available for mapping brain-wide activity and transcriptomes, comparable methods for mapping region-to-region connectivity remain slow and expensive because they require averaging across hundreds of brains. We have developed BRICseq, which leverages DNA barcoding and sequencing to map connectivity from single individuals in less than four weeks and at a total cost of less than $10,000. Applying BRICseq to the mouse neocortex, we find that region-to-region connectivity provides a simple bridge relating transcriptome to activity: The spatial expression patterns of a few genes predict region-to-region connectivity, and connectivity predicts activity correlations. We also exploited BRICseq to map the mutant BTBR mouse brain, which lacks a corpus callosum, and recapitulated its known connectopathies. BRICseq allows individual laboratories to compare how age, sex, environment, genetics, and species affect neuronal

wiring and to integrate these with functional activity and gene expression.

## Automated High-Throughput Microscopy Platform for In Situ Neuronal Projection Mapping

Y. Li, X. Chen, A. Vaughan

BARseq, previously described by our laboratory, has employed in situ sequencing to map neuronal projections in a high-throughput manner. The barcode sequences are read out using a four-channel sequencing by synthesis (SBS) technology. Each cycle of the sequencing readout requires at least 1 hr of chemical reactions and 30 min of imaging. Because there is no commercially available platform for in situ sequencing using SBS technology, this sequencing process was previously done manually in our laboratory.

The intensive labor investment during sequencing in generating in situ BARseq data motivated us to transit into the automation of this process. In the past year, we have established an automated in situ sequencing pipeline built around a one-of-a-kind automated microfluidics microscope. Compared to the manual sequencing, the automated sequencing platform has a significantly increased throughput and cost efficiency with minimal human-induced accidental error. Using this instrument, we have implemented new protocols for in situ sequencing and improved its overall reliability and throughput. This system has now become the workhorse of our laboratory. We are currently supporting various large collaborations through this work.

## Pairing Gene Expression and Spatial Location to Explore Whole-Brain Patterns of Transcription

S. Lu [in collaboration with J. Gillis, CSHL]

What is the relationship between gene expression and brain areas defined by conventional neuroanatomy? To study this relationship, we are using data from a recently developed spatial sequencing approach, spatial transcriptomics, in conjunction with the Allen Brain Atlas adult mouse in situ hybridization data. Combining these data will provide a foundation for understanding how gene expression relates to neuronal connectivity.

## Role of Corticostriatal Plasticity in Auditory Decision-Making

D. Maharjan

Animals rely on sensory information to guide decision-making. Over the past few decades, the neural mechanisms responsible for associating stimuli have been studied extensively. However, neuronal mechanisms responsible for associating stimulus representations with behavioral actions have not been investigated in equal depth. Using a novel auditory discrimination task, previous work from our laboratory has revealed that sensory information relayed from auditory cortex to the auditory striatum plays a pivotal role in guiding auditory decision-making. Furthermore, acquisition of this task leads to synaptic changes in the connections between these regions based on the tuning properties of the corticostriatal projections (Znamenskiy and Zador, *Nature* **497:** 482 [2013]; Xiong et al., *Nature* **521:** 348 [2015]). Whether this synaptic strengthening occurs selectively onto a specific cell type in the auditory striatum is still unknown. Using in vitro and in vivo approaches in transgenic mouse lines, we will investigate the role of cell type–specific synaptic strengthening of corticostriatal projections in auditory decision-making.

## Axonal BARseq, a Novel Technique for High-Throughput Mapping Single-Cell Projections In Situ

L. Yuan, X. Chen

The wiring of neural circuits across and within brain regions is crucial for understanding brain function. However, even within a single brain region, neuronal connections are surprisingly heterogeneous, and even adjacent cells often make distinct projections. Therefore, we need a method to map projections with single-cell resolution.

MAPseq (multiplexed analysis of projections by sequencing) is a high-throughput brain mapping method. In MAPseq, individual neurons are labeled with unique RNA barcodes and their projections are resolved with bulk DNA sequencing. Recently, we introduced a novel method, BARseq, in which in situ sequencing is used to read out somatic barcodes.

We are now extending these approaches to axons. Axonal BARseq is able to rapidly map thousands of single-cell projections across different target areas with micron resolution. Currently, we are applying axonal BARseq on mouse auditory cortex to study the effects of cortical layers and topographic organization on neural circuit wiring. Meanwhile, we are developing a data process and analysis pipeline for the significant amount of data generated from the axonal BARseq experiments. We anticipate these tools and results can uncover further spatial details regarding the heterogeneity of neural wiring.

## Developing Synaptic PLA (Proximity Ligation Assay) as a Method to Detect Recently Potentiated Synapses among the Corticostriatal Synapses of Mice

H. Zhan, S. Ghosh

The formation of memory and the process of learning have long been attributed to changes in synaptic strength between specific cell populations in the brain. Yet, it remains highly challenging to identify which specific synapses undergo such changes when an animal acquires a new memory or learns to perform a certain task. Popular techniques such as staining for "immediate early genes" (e.g., c-fos and arc) can only indicate which cells in the brain are activated during a behavior and fail to provide answers at a subcellular resolution (i.e., at the level of a synapse). Two-photon imaging techniques may provide the necessary resolution, but remain extremely low-throughput and technically challenging. Therefore, we developed SYNPLA (synaptic proximity ligation assay), a sensitive, specific, and high-throughput method for detecting the synaptic plasticity between candidate neuronal populations (Dore et al. 2020). This technique leverages the well established finding that during long-term potentiation (LTP) of synapses, GluA1 receptors move into synapses. It results in a greatly amplified, punctate fluorescent signal at recently potentiated synapses, allowing their easy and high-throughput identification using light microscopy. We have demonstrated that PLA-based detection of existing synapses is possible in the mouse corticostriatal circuit. We are now training animals on tonecloud tasks to detect the synapses potentiated by learning on this task.

# Representations in Primary Visual Cortex Underlying Visual Discrimination

A. Zhang

Although the hierarchical pathways underlying visual decision-making in primates have been studied extensively, the neural circuits underlying visual decision-making in rodents are not as well understood. We therefore set out to investigate the representations and contributions of rodent primary visual cortex (V1) in a visual decision-making task. We designed a novel visual discrimination task for freely moving rats that asks animals to judge the dominant spatial location of a visual stimulus by comparing between two regions in visual space. This task is directly analogous to an existing auditory discrimination task used in the laboratory that asks animals to judge the dominant frequency of an auditory stimulus (Znamenskiy and Zador, *Nature* **497:** 482 [2013]; Xiong et al., *Nature* **521:** 348 [2015]). We control viewing angle at the center port by implementing virtual head position control using online video tracking. Subjects readily learn to perform this task at high levels of accuracy. Although subjects are able to find and stably carry out the comparison rule when it is necessitated by the statistics of the stimulus distribution over trials, behavioral experiments revealed that at baseline, subjects reliably converged on an abbreviated strategy such that only half of the full stimulus was necessary and sufficient to drive the behavior. We use tetrode recordings to interrogate neuronal responses in V1 during behavior.

In addition to classically responsive visual neurons, we find overlapping subpopulations of V1 single neurons that are not only responsive but also selective to choice side and outcome. Further, we find that the organization of representations is distributed across the population with little correlation between stimulus tuning and choice or outcome tuning. We continue to investigate how stimulus use or behavioral context shapes stimulus encoding in the early visual system in this task and in a modified task in which the animal's strategy is independent of the sensory stimulus. To understand the causal role of how V1 contributes to this task, we are using optogenetic techniques to inactivate V1 during different task epochs, and testing the effect on behavior.

## PUBLICATIONS

Chen X, Sun YC, Zhan H, Kebschull JM, Fischer S, Matho K, Huang ZJ, Gillis J, Zador AM. 2019. High-throughput mapping of long-range neuronal projection using in situ sequencing. *Cell* **179**: 772–786.e19. doi:10.1016/j.cell.2019.09.023

Shuvaev SA, Başerdem B, Zador AM, Koulakov AA. 2019. Network cloning using DNA barcodes. *Proc Natl Acad Sci* **116**: 9610–9615. doi:10.1073/pnas.1706012116

Zador AM. 2019. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat Commun* **10**: 3770. doi:10.1038/s41467-019-11786-6

*In Press*

Dore K, Pao Y, Soria Lopez J, Aronson S, Zhan H, Ghosh S, Merrill S, Zador AM, Malinow R, Kebschull JM. 2020. SYNPLA, a method to identify synapses displaying plasticity after learning. *Proc Natl Acad Sci* **117**: 3214–3219. doi:10.1073/pnas.1919911117

# PLANT BIOLOGY

**David Jackson** and colleagues study genes and signals that regulate plant growth and architecture. They are investigating a unique way in which plant cells communicate by transporting regulatory proteins via small channels called plasmodesmata. These channels, which direct the flow of nutrients and signals through growing tissues, are regulated during development. The team discovered a gene encoding a chaperonin, *CCT8*, that controls the transport of a transcription factor called SHOOTMERISTEMLESS (STM) between cells in the plant stem-cell niche, or meristem. STM is critical for stem-cell maintenance, and studies of the *CCT8* gene indicate that movement of STM between cells is required for this function. The laboratory also continues to identify other genes that control plant architecture through effects on stem-cell maintenance and identity, and their work has implications for crop yields. Recent examples include discovery of a subunit of a heterotrimeric G protein that is conserved throughout the animal and plant kingdoms, and their studies indicate that this gene controls stem-cell proliferation. They have found that in plants the G protein interacts with a completely different class of receptors than in animals. Their discovery helps to explain how signaling from diverse receptors is achieved in plants. Last year, they also showed that weak mutations in one of the receptor proteins can enhance seed production in maize, which could lead to yield increases. Separately, the laboratory has characterized system-wide networks of gene expression, using next-generation profiling and chromatin immunoprecipitation methods that have led to many new hypotheses regarding developmental networks controlling inflorescence development. They are also developing a collection of maize lines that can drive expression of any reporter or experimental gene in any tissue type—tools of great interest to maize researchers that are being made available to the broader scientific community, enabling experiments never before possible in crop plants.

**Zachary Lippman**'s research focuses on the processes of flowering and flower production in plants, which are major contributors to reproductive success and crop yield. Specifically, Lippman's research program integrates development, genetics, genomics, and gene editing to explore the mechanisms that determine how plant stem cells become shoots and flowers. The laboratory takes advantage of extensive natural and mutant variation in inflorescence production and architecture in tomato and related nightshade species (e.g., potato, pepper, groundcherry) to explore how differences in these processes explain the remarkable diversity in the architectures of flower-bearing shoots (inflorescences) observed in nature and agriculture. Recent discoveries regarding the genes and networks underlying this diversity have led to broader questions about the significance of genomic structural variation, gene redundancy, and epistasis in development, domestication, and breeding. Based on their fundamental discoveries, the group is developing and applying innovative concepts and tools for crop improvement.

Epigenetic mechanisms of gene regulation—chemical and conformational changes to DNA and the chromatin that bundles it—have had an important impact on genome organization and inheritance and cell fate. These mechanisms are conserved in eukaryotes and provide an additional layer of information superimposed on the genetic code. **Robert Martienssen,** a pioneer in the study of epigenetics, investigates mechanisms involved in gene regulation and stem-cell fate in yeast and model plants, including *Arabidopsis* and maize. He and his colleagues have shed light on a phenomenon called position-effect variegation, caused by inactivation of a gene positioned near densely packed chromosomal material called heterochromatin. They have discovered that small RNA molecules arising from repeating genetic sequences program that heterochromatin. Martienssen and colleagues have described a remarkable process by which "companion cells" to sperm

in plant pollen grains provide them with instructions that protect sperm DNA from transposon damage. They found that some of these instructions, or epigenetic marks, could be inherited in the next generation. These marks, and the small RNAs responsible for guiding them, can sense the number of chromosomes inherited from pollen and may allow *Arabidopsis*, a flowering plant, to produce egg cells without meiosis, an important step toward a long-time goal of plant breeding: generating clonal offspring to perpetuate hybrid vigor. The laboratory has also shown that when RNA polymerase II has transcribed a stretch of DNA, the RNA interference mechanism causes the enzyme to release its hold on the DNA and fall away. This allows the replication fork to progress smoothly and the DNA strands to be copied; histone-modifying proteins, which follow right along, establish heterochromatin. Martienssen's group also continues to work on problems related to the creation of plant-based biofuels. As part of a collaborative project to generate a high-quality full genome map of the oil palm plant, Martienssen and his colleagues identified a transposon whose modification controls the yield of oil palm trees. This discovery will increase yields and should lessen the environmental burden of oil palm production, which often threatens already endangered rainforest lands.

Plants and animals interact with their environment. Because plants are incapable of moving around, they are sensitive to their surrounding environment and modify their development according to external signals. Plants face variability in growth conditions—temperature, light quality and quantity, herbivores, pathogens, water availability, etc. Yet, plants respond to these biotic and abiotic factors and survive substantial fluctuations in their environment. Plants also must balance the range of potential threats and benefits confronting them and should make appropriate decisions on resource allocation. Such adaptability is essential given the sessile nature of plants. The mechanisms that underlie this adaptability likely involve complex signaling to generate the appropriate response. In some adaptive responses (e.g., when the plants have to cope with climate change and increased competition for light), there is a decrease in productivity (yield, biomass) as the plant reallocates resources to better adapt.

**Ullas Pedmale**'s laboratory seeks to determine the mechanisms behind how a plant perceives and successfully adapts to its environment. They also aim to understand how a plant must integrate intrinsic and extrinsic cues and "decide" how best to respond to environmental cues. Understanding how plants deal with and respond to a multitude of environmental signals could help to develop crops that cope with unfavorable growth conditions without significant changes in yield.

# DEVELOPMENTAL BIOLOGY—STEM-CELL SIGNALING AND CROP ARCHITECTURE

**D. Jackson**

R. Chen
H. Claeys
E. Demesa-Arevalo
C. Fugina
J. Huang

Y. Jian
M. Kitagawa
J-H. Lee
P. Lindsay
L. Liu

B. Siegel
T. Skopelitis
A. Steinberger
T. Tran
M. Venezia

Z. Williams
Q. Wu
F. Xu
X. Xu

Our research asks how the growth and shape of plants is controlled, with the ultimate goal of improving crop yields. We identify the genes, signals, and pathways that regulate plant architecture and development. Like all organisms, plants grow and develop by manipulating the passage of information between cells. We are interested in discovering the signals that carry this information, how they are transmitted, and how they function. A major focus has been identification of genes that control stem-cell signaling. In the past year, we reported a new way to increase yield in maize, by balancing a tradeoff between growth and defense controlled by G proteins in stem cells. We also identified a new way in which plants control branching through use of enzymes that function in sugar metabolism, but our research suggests these enzymes play a more important role in the cell nucleus. In other studies, we continue to expand our use of CRISPR genome editing to uncover new gene functions as well as to increase allelic diversity to improve maize yield traits.

## The Control of Meristem Size in Maize

Q. Wu, F. Xu, P. Lindsay [in collaboration with B.I. Je, Busan University, Korea]

All plant organs derive from populations of stem cells called meristems. These cells have two purposes: to divide and maintain themselves and to give rise to daughter cells, which will differentiate into plant organs. Consequently, meristems must precisely control the size of the stem-cell niche, via a network of positive- and negative-feedback signals. A loss of function in a negative regulator of stem-cell fate results in an enlarged or fasciated meristem phenotype and a dramatic alteration in the morphology of the maize ear and tassel. Maize is an excellent model system for these studies because of a large collection of developmental mutants and a diverse genome. Our laboratory uses genetics to identify key regulators of stem-cell homeostasis and meristem size. Two previously cloned mutants, *fasciated ear2* and *thick-tassel dwarf1*, encode orthologs of the *Arabidopsis thaliana* genes *CLAVATA1* and *CLAVATA2*, indicating the well-known *CLAVATA-WUSCHEL* regulatory feedback loop is conserved in monocot crops. However, little else is known about the control of this important developmental process in maize. Here, we describe progress in identifying additional genes contributing to stem-cell homeostasis.

A common class of proteins that signal directly downstream from cell surface receptors is the heterotrimeric G proteins, consisting of Gα, Gβ, and Gγ subunits. Our previous work found that the maize *COMPACT PLANT2* (*CT2*) gene, which encodes the α subunit of a heterotrimeric GTPase (Gα), functions in the CLAVATA pathway to control meristem size through its interaction with the FEA2 receptor. To further study the mechanism of G-protein signaling in meristem development, we knocked out the sole *Gβ* gene of maize, *ZmGB1,* using CRISPR-Cas9. Using a functional translational fusion of *ZmGB1* and *YFP* under the control of its native promoter, we detected the protein in both cytosol and plasma membrane throughout the meristem, suggesting that it acts in receptor complexes, consistent with our previous findings for Gα. However, to our surprise, we found that the *Zmgb1ᶜʳⁱ* null mutant plants died at an early stage of seedling development, with overaccumulation of reactive oxygen species and salicylic acid, constitutive activation of MAP-kinases, and upregulation of *PATHOGENESIS-RELATED* immune marker genes. These results suggest that *ZmGB1* mutation causes autoimmune symptoms. We therefore crossed the *Zmgb1ᶜʳⁱ* null alleles into different maize lines to see if the lethal phenotype can be suppressed and indeed found that it was partially suppressed

in a tropical maize background. In this lethality-suppressed background, the *Zmgb1^{cri}* mutants had significantly larger meristems and fasciated ears, suggesting that ZmGB1 plays an important role in development. We mapped the suppressor of *Zmgb1* lethality in CML103 and identified a disease resistance (*R*) gene as a candidate, suggesting that *G*β acts as an immune sensor, or "guardee," in maize, unlike in *Arabidopsis*, in which *G*β mutants are viable. We have also performed a genetic screen to identify lethality suppressors and identified a candidate that we are now characterizing.

By chance, we identified a second allele of *Zmgb1* with a single amino acid change in a conserved residue of one of the WD40 domains. This protein was impaired in forming a complex with the α and γ subunits. The mutants had fasciated ears and thick tassels and were dwarf with enlarged meristems. They also developed necrotic lesions, reminiscent of the *Zmgb1^{cri}* null mutants, and failed to complement *Zmgb1^{cri}* mutants, suggesting that they are allelic. Double-mutant analysis suggested that ZmGB1 functions in the same pathway as CT2, and *fea2* mutants were epistatic to *Zmgb1*, so ZmGB1 and CT2 might function together downstream from FEA2. In contrast, double mutants with another mutant, *fea3*, made massively enlarged meristems (Fig. 1), suggesting that ZmGB1 functions in a parallel pathway to FEA3.

We also continue to study the mechanism of action of *FEA3*, which encodes a predicted leucine-rich repeat receptor-like protein related to *FEA2*. FEA3 is of particular interest because it is expressed in the organizing center of the shoot apical meristem (SAM) and in leaf primordia, and expression of maize WUSCHEL, a marker for the stem-cell niche organizing cells, spreads downward in *fea3* mutants, which is strikingly different from its response in the known *CLAVATA* stem-cell mutants.

To further understand the role of FEA3 in stem-cell signaling, we are using immunoprecipitation-mass spectrometry (IP-MS) with FEA3-tagged plants. Our preliminary analysis found some candidate interactors, which we are now validating. One candidate interactor, protein phosphatase 2A, has been verified by co-immunoprecipitation experiments in tobacco cells. We are now knocking out this gene using CRISPR-Cas9 to see if it has an effect on inflorescence development. As a complementary approach, we are also cloning constructs for a new method, proximity labeling, to assess transient interactions between FEA3 and interacting proteins. This technique may prove particularly useful for identifying receptor interactions.



**Figure 1.** Highly fasciated meristems of *fea3; Gβ* double mutants.

## Genetic Redundancy in Circuits Controlling Meristem Development

L. Liu, E. Demesa-Arevalo, F. Xu, T. Skopelitis, R. Chen [in collaboration with Z. Lippman, CSHL; M. Bartlett, University of Massachusetts, Amherst; Z. Nimchuck, University of North Carolina, Chapel Hill]

The CLE (CLAVATA3/endosperm surrounding region-related) peptides are fundamental players in meristem maintenance in plants, acting as mobile signals that establish feedback signaling to control the balance of stem-cell division and differentiation. Disruption of this pathway causes overproliferation or fasciation in meristems. We have described different fasciated mutants encoding leucine-rich repeat receptor-like kinases or receptor-like proteins (LRR-RLKs or RLPs); however, the signals perceived by many of these receptors remain elusive. Forty-nine CLE peptide genes that are candidate ligands for LRR receptors have been identified in maize, suggesting either specialization or redundancy in these ligands.

To decipher the roles of CLE peptide ligands in maize meristem regulation and their involvement in redundant circuits, we generated CRISPR-Cas9 knockouts. We first analyzed expression patterns from publicly available data sets combined with our transcriptional profiles from meristematic tissues and identified 31 candidates expressed in meristems.

CRISPR frameshift mutations have been induced in 27 of them, and mutants from the same subgroup are being crossed together to evaluate their functional redundancy. To identify *Zm*CLE peptides that may regulate meristem size in a redundant way, we performed RNA-Seq in dissected inflorescence meristems from mutants in the maize CLV3 ortholog, *Zmcle7*. Three CLE genes, including *Zm*FCP1 and *Zm*CLE7, were significantly up-regulated. We found that *Zm*FCP1 partially compensates for loss of *Zmcle7*$^{cr}$. However, the mechanism of compensation was distinct from that observed in *Arabidopsis* and tomato, suggesting that different mechanisms underlie the evolution of genetic buffering in stem-cell pathways. We identified another maize CLE with a similar expression pattern to *ZmCLE7* that was up-regulated in *Zmcle7* mutants (Fig. 2A-B), suggesting a potentially redundant role in meristem homeostasis. CRISPR-Cas9 null alleles of this novel CLE gene had significantly longer ears with more kernel rows (Fig. 2C–E).

Redundancy in stem-cell signaling is also seen on the receptor side, as *Arabidopsis clv1* mutant phenotypes are enhanced by mutations in the related *BARELY ANY MERISTEM 1*, *2*, and *3* (*BAM1*, *BAM2* and *BAM3*) receptors. In the past year, we found that one maize BAM gene, *ZmBAM1d*, controls seed size—an exciting finding that could help us address the tradeoff between seed number and size. Maize has seven *BAM*-like genes and a single *CLAVATA1* gene (*TD1*), and to characterize their function and redundancy, higher-order CRISPR mutants have been made. The genetic interactions between different *BAM* genes and CLEs will allow us to dissect additional signaling pathways for meristem maintenance in maize.

We previously found that weak alleles of fasciated ear mutants can improve maize yield traits, such as kernel row number, by increasing meristem size and number of primordia, while maintaining structural integrity of the meristem. We found that *fea3* weak allele hybrids also enhance overall yield in laboratory strains of maize. These results are particularly exciting because in our previous studies of weak *fea2* alleles, we found an increase in kernel row number but no overall increase in ear weight—because of a compensatory reduction in kernel size. Therefore, the newly identified *FEA3* signaling pathway could be used to develop new alleles for crop improvement. Because *Zmfcp1* and *Zmcle7* single mutants are also fasciated, we mutated their promoters using CRISPR-Cas9 editing to create weak alleles to ask if they could also enhance yield traits. The promoters (~2kb) of *ZmFCP1* and *ZmCLE7* were targeted by multiplex sgRNAs (single guide RNAs), and a variety of edited haplotypes were obtained. Some promoter weak alleles showed a significantly enlarged but nonfasciated ear; for example, the *pcle7-1* allele led to an increase of ~25% in grain yield, indicating potential utility of these favorable alleles in maize breeding (Fig. 3).
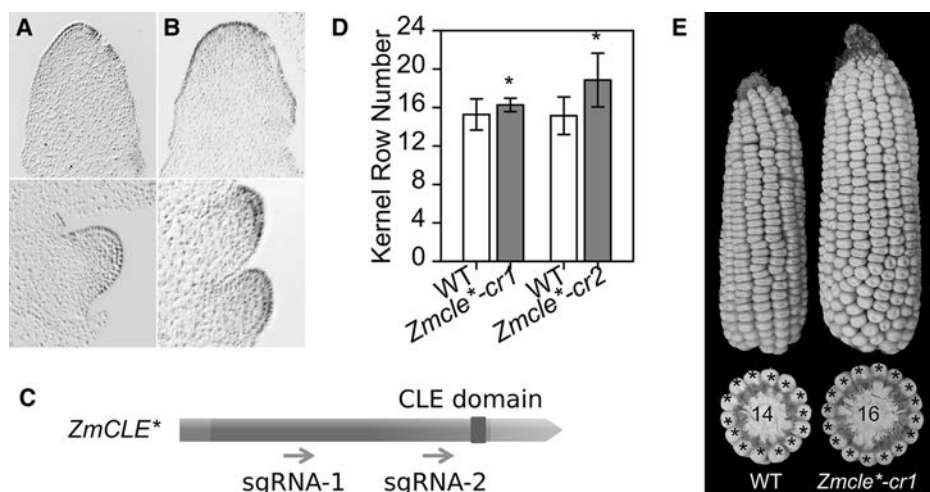


**Figure 2.** Role of a novel maize *CLE* gene in maize yield traits. (*A*) In situ hybridization of the new *CLE* gene reveals expression very similar to *ZmCLE7* (*B*). (*C*) Generation of CRISPR-Cas9 alleles. (*D*) Two CRISPR-Cas9 alleles have greater kernel row number, and the mutants had longer ears (*E*).
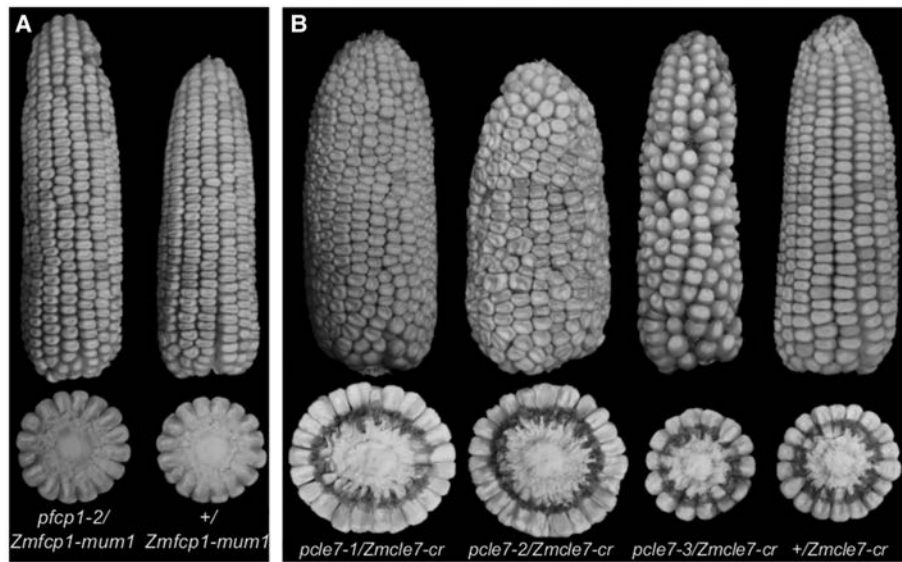
**Figure 3.** Promoter-edited alleles of *ZmFCP1* (*A*) and *ZmCLE7* (*B*) enhance maize ear size and seed number. In each panel, the standard control is on the *right*, and the CRISPR-enhanced ears are on the *left*.

## Control of Shoot Branching and Determinacy

H. Claeys, E. Demesa-Arevalo, X. Xu, T. Skopelitis, T. Tran, J-H. Lee, Z. Williams [in collaboration with A. Steinberger, Williams College; M.J. Abraham-Juárez, Instituto Potosino de Investigación Científica Y Tecnológica (IPICYT), Mexico]

The *RAMOSA* (*RA*) genes function to impose determinacy on axillary meristem growth; consequently, *ra* mutants (*ra1*, *ra2*, and *ra3*) have highly branched inflorescences. *RA3* encodes a trehalose phosphate phosphatase, an enzyme that converts trehalose-6-phosphate (T6P) to trehalose. T6P is an important regulatory metabolite that connects sucrose levels, and thus sugar status, to plant growth and development—but its mode of action is still unclear. *RA3* is expressed at the base of axillary inflorescence meristems and localizes to distinct puncta in both nuclear and cytoplasmic compartments, suggesting that its effect on development may not be simply metabolic. These data support the hypothesis that *RA* genes serve as mediators of signals, maybe a sugar signal, originating at the boundary domain and regulating determinacy.

We aim to identify factors that act in the same pathway by screening for enhancers of the *ra3* phenotype. Typically, *ra3* mutants have three to eight ectopic branches at the base of the ear. We mutagenized *ra3* and looked for plants that have more branches and/or branches in the upper part of the ear. So far, four independent alleles of *TPP4* (*TREHALOSE-6-PHOS-PHATE PHOSPHATASE*), a *RA3* paralog, were identified, and we confirmed that *TPP4* is the causative gene using additional CRISPR-Cas9–generated alleles. *TPP4* was expressed in the same domain as *RA3*, and on mutation of *ra3* its expression was up-regulated, suggesting that it acts as a redundant backup to compensate for loss of *RA3*. All ethyl methanesulfonate (EMS)-induced alleles contained single amino acid substitutions, and some of the resulting mutant proteins still had considerable enzymatic activity, despite all having similar phenotypic strength. Therefore, there is no straightforward relationship between TPP activity and phenotype, and additional regulatory functions of TPP4 may be important. Using CRISPR-Cas9, we also generated knockouts of *TPP12*, a more distant family member that is also expressed in developing inflorescences; but unlike *TPP4*, mutating *TPP12* did not enhance the RA3 phenotype, showing functional divergence within the TPP family.

Another *ra3* enhancer mutant was mapped to a gene encoding an RNA-binding protein that functions in inflorescence development. We have made maize lines carrying functional yellow fluorescent protein (YFP) fusions of this gene, and will use these

to further characterize its role in meristem determinacy, and its relationship with RA3, by looking at protein–protein and protein–RNA interactions. More recently, we mapped a third enhancer mutant to *IDS1 (INDETERMINATE SPIKELET1)*. *ids1* mutants have more floral meristems, and *IDS1* expression was significantly increased in *ra3* mutants. We are testing whether the enhancement of RA3 by IDS1 may be due to their physical interaction. For the other *ra3* enhancer mutants, we are currently at various stages of mapping and confirming the causal genes.

To further probe RA3 functions, we also screened for interactors using IP-MS and probed their subcellular localization. Of particular interest, given the nuclear speckle localization of RA3, we found two predicted RNA binding proteins. Bimolecular florescence complementation assays revealed that their interaction with RA3 occurs in the nucleus, forming speckles—implying a potential regulatory function in meristem determinacy. Our ongoing genetic and biochemical analysis aims to uncover the biological meaning of this association in modulating meristem determinacy. Using different nuclear markers, we are also defining the nature of the RA3 nuclear speckles.

We also seek to understand the contribution of the enzymatic function of RA3 to its biological mechanism. We achieved partial complementation of *ra3* mutants using a catalytically dead mutant version of RA3, supporting a moonlighting hypothesis, and are now refining this experiment by mutating endogenous RA3 and TPP4 loci to catalytically dead versions using CRISPR-Cas9–mediated base editing. So far, we have identified several catalytically dead TPP4 alleles. Our recent study revealed that TPP4 acts as a redundant backup for RA3. Therefore, we will introgress the catalytic dead TPP4 allele into *ra3* mutant backgrounds to test whether it also enhances *ra3* branching. The results will allow for a better understanding of TPP enzyme function and its relevance to developmental mechanisms.

IP-MS experimentation also found that RA3 interacts with TPS1, a noncatalytic trehalose phosphate synthase (TPS). TPSs catalyze the step preceding RA3 in the trehalose metabolic pathway. *tps1* mutants enhanced *ra3* phenotypes, suggesting the interaction between TPS1 and RA3 is biologically significant. Interestingly, a yeast-2-hybrid experiment found that TPS1 also interacts with two catalytic TPSs, TPS14 and TPS11. Our working model is that the noncatalytic TPS1 might bind and affect the enzyme activity of a catalytic TPS complex in maize. To test this model, we made CRISPR-Cas9 mutant alleles of these catalytic TPSs, and double and triple *tps1, tps11,* and *tps14* mutations are being generated. These mutants will be used to characterize the function of noncatalytic and catalytic maize TPS. In addition, other approaches will be applied to test RA3 interactions with catalytic and noncatalytic TPS proteins, to further understand the enigmatic role of trehalose signaling in plants.

Last, we are attempting to knock out the entire TPP gene family in *Arabidopsis* using CRISPR. So far, mutation of two RA3 homologs, *TPPI* and *TPPJ*, did not reveal any obvious phenotype, so we mutated all 10 TPP genes using CRISPR-Cas9 and are combining them to overcome the likely redundancy in this gene family.

## Natural Variation in Inflorescence Architecture

H. Claeys, T. Tran

Maize inflorescence architecture has been a target for extensive selection by breeders since domestication, and the maize genome is highly diverse; hence, different maize inbreds vary greatly in these traits. The genetic basis underlying this diversity is largely unknown, but is of great interest for both fundamental and applied work. To identify natural variation relevant to inflorescence traits, we screened for inbred backgrounds that can enhance or suppress the phenotypes of different mutants. We focused on the 25 nested association mapping (NAM) founder inbreds, because they were selected to capture the diversity of maize germplasm and because of the available genetic tools. We crossed these 25 inbreds to our collection of mutants (in a B73 background) and screened the F2s for mutants with suppressed or enhanced phenotypes.

*fea2-0* was strongly enhanced in the NC350 background, and we mapped this enhancement to a single major-effect locus on chromosome 5 using both bulked segregant analysis and crosses to recombinant inbred lines (RILs). Interestingly, we used B73-NC350 heterogeneous inbred families (HIFs) to show that the NC350 allele of *TD1* positively affects kernel row number (KRN), demonstrating its potential usefulness in breeding. Fine mapping of the NC350 enhancer locus led us to a small region containing only

about five genes, with a prominent candidate *THICK TASSEL DWARF1* (*TD1*), a known meristem regulator. Consequently, a *td1* loss-of-function mutant cannot rescue the enhanced phenotype. There are no obvious causal changes in the coding sequences of the two alleles of TD1, but we found regulatory changes that may explain the enhanced phenotype; the B73 allele of *TD1* is up-regulated in a *fea2* mutant background, whereas the NC350 allele is insensitive to loss of *FEA2*. Our results reveal changes in the wiring of regulatory networks controlling compensation in meristem size between ecotypes. To test this candidate, we have been generating alleles with mutations in the *TD1* promoter through CRISPR-Cas9-mediated promoter editing and are asking if they can recapitulate the behavior of the NC350 allele. We identified one allele with a ~600-bp insertion; however, it did not lead to any phenotype, so we are now identifying additional alleles.

## Mechanism of Active Transport of Transcription Factors through Plasmodesmata

M. Kitagawa, T. Skopelitis

In plants, some transcription factors (TFs) are actively and selectively transported between cells to specify their fates. These TFs are transported through plasmodesmata (PD), membrane-lined channels traversing the cell wall. To this date, however, the mechanism underlying the active and selective transport of TFs through PD has been largely unknown. Previously, we established a system for evaluating the capacity of the active transport of TFs in *Arabidopsis* seedlings using a mobile homeodomain TF called knotted (KN1). Using this system, we isolated two mutants, *rb31-7* and *mk5-140*, that are mutated in the same gene, encoding rRNA processing protein (RRP) 44A. This protein is a subunit of the RNA exosome complex and functions in the processing and degradation of specific RNAs. Both mutants have amino acid substitutions in the conserved catalytic domain. RRP44A fused to the TagRFPT fluorescent protein (RRP44A-TagRFPT) under its native promoter in *Arabidopsis* was able to fully complement the mutant phenotype. RRP44A-TagRFPT expression was observed in meristem tissues, overlapping with the *Arabidopsis* KN1 homolog shoot meristemless (STM). Double *rrp44;stm* mutants had enhanced fusions between leaves or petioles, and RNA-immunoprecipitation RT-PCR assays found that RRP44A binds to *STM* mRNA. Additionally, we found that RRP44A colocalized with PD. Collectively, these data support a hypothesis that RRP44A regulates cell-to-cell trafficking of *KN1/STM* mRNA through PD for proper meristem functions.

To ask whether and how RRP44A regulates mRNA trafficking, we next generated a system to visualize *KN1* mRNA using the MS2 system. In this system, fluorescent protein–tagged MS2 bacteriophage coat protein binds to a recognition RNA motif that we incorporated into the *KN1* mRNA, leading to its visualization. Using this system, we found that *KN1* mRNA can move between cells, forms speckles in the cytoplasm and locates at PD (Fig. 4). We will compare *KN1* mRNA trafficking between wild type and *rb31-7* or *mk5-140* to ask if RRP44A controls *KN1* mRNA trafficking via PD.
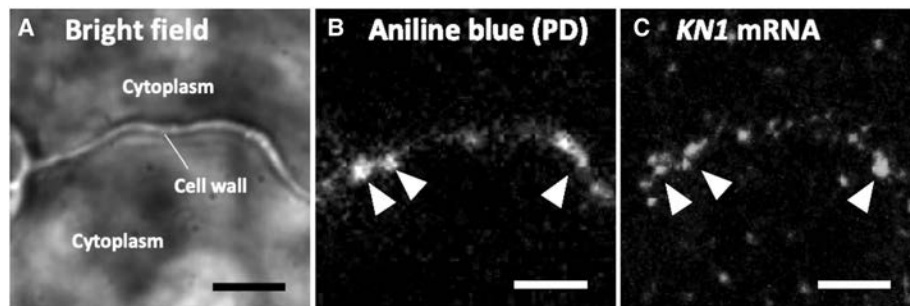


**Figure 4.** *KN1* mRNA colocalizes with PD. (*A*) Bright-field image of cells in an *Arabidopsis* seedling. (*B*) PD were stained by aniline blue that stains a PD component, callose. (*C*) *KN1* mRNA was visualized by the MS2 system. *KN1* mRNA forms speckles and colocalizes with PD (arrowheads). Scale bars, 5 μm.
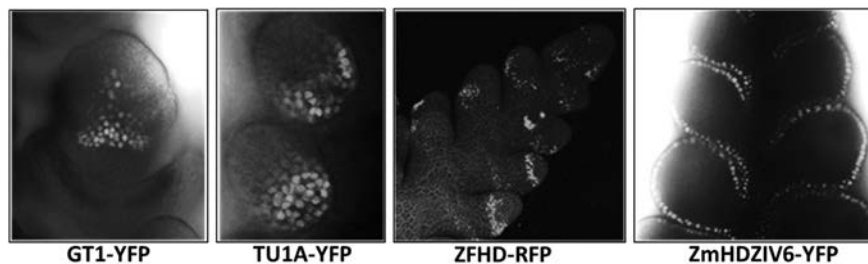
**Figure 5.** Examples of fluorescent protein (FP) tagged transgenic lines for diverse transcription factors. Examples are GRASSY TILLERS1(GT1), TUNICATE1A (TU1A), ZINC FINGER HOMEODOMAIN (ZFHD), and ZmHOMEODO-MAIN-LEUCINE ZIPPER IV6 (ZmHDZIV6).

## Functional Annotation of the Maize Genome by ChIP-Seq and FACS

X. Xu, D. Jackson [in collaboration with F. Li, L. Wang, X. Wang, J. Drenkow, T. Gingeras, and D. Ware, CSHL; Z. Lu and R. Schmitz, University of Georgia, Athens; A. Luo and A. Sylvester, University of Wyoming, Laramie; C.O. Ramirez and K. Birnbaum, New York University, New York]

To develop functional annotation of the maize genome, we are conducting genome-wide transcription factor (TF) binding analysis by chromatin immunoprecipitation-sequencing (ChIP-Seq) and expression profiling of cell types by fluorescence activated cell sorting (FACS). These are important goals of the MaizeCODE project, an initial analysis of functional elements in the maize genome.

To perform genome-wide binding site analysis, we generated fluorescent protein (FP) tagged transgenic lines for TFs that function in different aspects of maize development (Fig. 5). These TFs belong to several different families, such as the MADS box, TUNICATE1A (TU1A), functioning in flower development changes during maize domestication; the homeodomain TF, ZmWUSCHEL1 (ZmWUS1), functioning in meristem maintenance; and the GATA TF, TASSELSHEATH1 (TSH1), functioning in bract suppression. To overcome limited tissue availability for conducting ChIP-Seq, we crossed these lines into a double mutant, *branched silkless*;*Tunicate* (*bd*;*Tu*), that transforms the maize ear into a "cauliflower" with overproliferating meristems. We conducted ChIP-Seq for eight TFs and detected the expected canonical binding motifs for five of them. For example, we detected a MADS box motif for TU1A and ZmMADS16 (ZmM16) and HDZIP motifs for GRASSY TILLERS1(GT1) and ZmHOMEODOMAIN-LEUCINE ZIPPER

IV6 (ZmHDZIV6). We also identified many interesting binding targets—for instance, *ZmWUS1* and *ZmWUS2* were identified as binding targets for ZmM16, and many other *ZmHDZIV* genes were identified as binding targets for ZmHDZIV6.

To profile cell type–specific expression by FACS, we also generated tissue-specific promoter FP lines to isolate cell types or domains for RNA-Seq and ATAC-Seq analysis (Fig. 6). The promoter lines cover several different types of cells or tissues in the meristem, such as the organizing center (p*ZmWUS1*-mRFP1), lateral organ zone (p*ZmYABBY14*-tagRFPt), and epidermal zone (*pZmHDZIV6*-tagRFPt). RNA-Seq of FACS-sorted cells revealed that p*ZmYABBY14*-tagRFPt and *pZmHDZIV6*-tagRFPt cells were enriched for all known *YABBY* and *ZmHDZIV* genes, respectively, indicating that the experiments were successful. The FACS profiling data are now available through the CyVerse MaizeCODE project portal as a public resource.
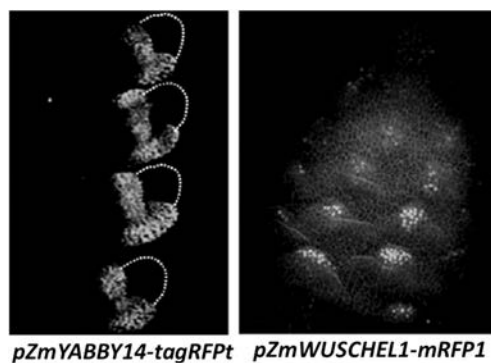


**Figure 6.** Examples of tissue-specific promoter FP lines. *ZmYABBY14* is marking organ primordia; *ZmWUSCHEL1* is marking the organizing center.

# New Insights into Maize Ear Development Using Single-Cell (sc) RNA-Seq

X. Xu, L. Liu, E. Demesa-Arevalo, D. Jackson [in collaboration with M. Crow, N. Fox, J. Preall, and J. Gillis, CSHL]

Productivity of maize depends on development of the ear, requiring a programmed series of meristem fate decisions between different cell populations. A fundamental understanding of development requires insight into the full diversity of cell types and developmental domains and the gene networks required to specify them. However, current studies have classified cell types and domains mainly by morphology and signatures derived from bulk tissue or organ RNA sequencing. Major insights into genes required to specify cell types and developmental domains have come from classical genetics; however, this approach is limited by genetic redundancy and pleiotropy. Single-cell transcriptome profiling of heterogeneous tissues, such as the maize ear inflorescence, can provide high-resolution windows to understand genome-wide transcriptional signatures of specific cell types and identify new developmental domains in a quantitative and comprehensive manner.

We investigated the transcriptional profiles of 15,000 single cells from B73 ear primordia across six replicates and a similar number from roots. In ears, we detected expression from 28,000 genes, with an average of 5,600 transcripts detected per cell, and clustering identified 18 reproducible cell groups via meta-analysis. Many of the groups expressed known marker genes, such as an L1/epidermal layer group, marked by *LIPID TRANSFER PROTEIN* and *OUTER CELL LAYER* genes; an L2 layer meristem group, marked by *KNOTTED1*; a lateral primordium group, marked by *YABBY* genes; and a vasculature group, marked by *RAN BINDING PROTEIN2*. Importantly, we could map 76 of the 77 known maize inflorescence development genes defined by mutant phenotype to specific cell clusters. Each of these groups contains an additional 20–100 new cell type– or domain-enriched genes that are novel candidates for markers or developmental regulators. To validate the accuracy of clustering and determine where each cell population was located, we selected marker genes

to perform mRNA in situ hybridization. Strikingly, we identified novel markers for specific developmental domains, such as meristem branching sites and meristem tips. Our resource can inform genetic analysis by accurately predicting genetic redundancy, such as finding a functional paralog of *RAMOSA3* (*RA3*), and aid in building co-expression networks at a single-cell level (e.g., to identify transcriptional regulators of an ear length quantitative trait loci [QTL]). In summary, we developed a community resource identifying hundreds of novel candidate regulators of cell fate or development, in the maize inflorescence and root, to inform maize genetics at a fundamentally new level.

## PUBLICATIONS

Claeys H, Vi SL, Xu X, Satoh-Nagasawa N, Eveland AL, Goldshmidt A, Feil R, Beggs GA, Sakai H, Brennan RG, et al. 2019. Control of meristem determinacy by trehalose 6-phosphate phosphatases is uncoupled from enzymatic activity. *Nat Plants* **5:** 352–357. doi:10.1038/s41477-019-0394-z

Kitagawa M, Jackson D. 2019. Control of meristem size. *Annu Rev Plant Biol* **70:** 269–291. doi:10.1146/annurev-arplant-042817-040549

Kitagawa M, Balkunde R, Bui H, Jackson D. 2019. An aminoacyl tRNA synthetase, OKI1, is required for proper shoot meristem size in *Arabidopsis*. *Plant Cell Physiol* **60:** 2597–2608. doi:10.1093/pcp/pcz153

Rodriguez-Leal D, Xu C, Kwon CT, Soyars C, Demesa-Arevalo E, Man J, Liu L, Lemmon ZH, Jones DS, Van Eck J, et al. 2019. Evolution of buffering in a genetic circuit controlling plant stem cell proliferation. *Nat Genet* **51:** 786–792. doi:10.1038/s41588-019-0389-

Wu Q, Xu F, Liu L, Char S, Ding Y, Je B, Schmelz E, Yang B, Jackson D. 2019. The maize heterotrimeric G protein β subunit controls shoot meristem development and immune responses. *Proc Natl Acad Sci* **117:** 1799–1805; commentary **117:** 1842–1844. doi:10.1073/pnas.1917 577116

Yang N, Liu J, Gao Q, Gui S, Chen L, Yang L, Huang J, Deng T, Luo J, He, L, et al. 2019. Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat Genet* **51:** 1052–1059. doi:10.1038/s41588-019-0427-6

*In Press*

Jia H, Li M, Li W, Liu L, Jian Y, Yang Z, Shen X, Ning Q, Du Y, Zhao R, et al. 2020. A serine/threonine protein kinase encoding gene *KERNEL NUMBER PER ROW6* regulates maize grain yield. *Nat Commun* **11:** 988. doi:10.1038/s41467-020-14746-7

Liu H, Jian L, Xu J, Zhang Q, Zhang M, Jin M, Peng Y, Yan J, Han B, Liu J, et al. 2020. High throughput CRISPR/Cas9 mutagenesis streamlines trait gene identification in maize. *Plant Cell* **32:** 1397–1413. doi:10.1105/tpc.19.00934

# MECHANISMS OF STEM CELL CONTROL AND QUANTITATIVE TRAIT VARIATION IN PLANTS

**Z.B. Lippman**

| L. Aguirre | J. Kim | S. Soyk |
| M. Benoit | A. Krainer | X. Wang |
| D. Ciren | C-T. Kwon | S. Zebell |
| J. He | S. Qiao | |
| A. Hendelman | G. Robitaille | |

Our research is focused on two main areas of plant biology. First, we are interested in the genes and mechanisms underlying flowering and flower production in nature and agriculture. Second, we are investigating mechanisms of quantitative trait variation, with an emphasis on the role of *cis*-regulatory mutations in evolution, domestication, and crop improvement. We use tomato and multiple related Solanaceae species as our primary model systems, supplemented with the classical plant model system *Arabidopsis thaliana*. Using genetics, genomics, and genome editing, we are creating libraries of mutations in both coding and *cis*-regulatory sequences of multiple developmental genes that control the production and maturation of stem cell populations (meristems). Such genes that we and others have characterized are major regulators of plant architecture—in particular, the branching number and pattern of reproductive shoots known as inflorescences, and also flower and fruit organ number and size. We have found that modifying the function of stem cell regulating genes individually and in combination can provide continuums of quantitative trait variation for these and other traits, mimicking natural alleles that impact gene expression. More recently, we have uncovered the extent of structural variation in a diverse collection of wild and domesticated tomato genotypes, which has revealed major roles for variation in gene and expression dosage on trait variation. In all of our work, genome editing is a key tool enabling a wide exploration and dissection of fundamental principles underlying stem cell control and plant productivity, which has resulted in direct demonstrations of how manipulation of plant developmental programs can be used to improve the productivity of major crops and often overlooked orphan crops.

## The Tomato Pan-SV Genome and Its Impact on Gene Expression and Quantitative Trait Variation

X. Wang, M. Benoit, S. Soyk, J. Kim, D. Ciren, G. Robitaille, A. Krainer

One of the greatest challenges in biological research is decoding the genetic changes that underlie important quantitative traits. The introduction of high-throughput short-read sequencing a decade ago accelerated the discovery of canonical genetic variants (i.e., single-nucleotide polymorphisms [SNPs] and small indels). Although these resources have helped reveal the genetic basis for some phenotypic diversity, the majority of quantitative variation remains unexplained. It is expected that a high-quality pan-genome that incorporates notoriously difficult-to-capture structural changes in DNA (large deletions, insertions, duplications, and chromosomal rearrangements) will expose a vast amount of hidden genetic variation underlying trait diversity.

Previous studies from our laboratory and others have shown examples in which structural variants (SVs) play important roles in shaping important crop domestication and improvement traits. However, characterizing the extent, diversity, and quantitative impact of SVs has been challenging because of limitations in their detection. Taking advantage of the third-generation long-read sequencing and genome-editing technologies, we constructed the most comprehensive pan-SV genome for a major crop and studied its significance in evolution, domestication, quantitative genetics, and breeding. By applying single-molecule Oxford Nanopore long-read sequencing on a panel of 100 domesticated and wild tomato genomes, we resolved more than 200,000 SVs. Genome distribution of SVs revealed extensive admixture, including

cases of near-entire chromosomes from wild species introgressed into modern tomatoes through breeding. RNA sequencing of multiple tissues from more than 20 genotypes revealed that SVs near genes impact expression quantitatively. To meet the need for multiple reference genomes for plant biology and crop improvement, we established 14 new genomes from across the tomato phylogeny. This resource, along with our global SV analyses, resolved genomic complexity that was masking the genes and variants underlying quantitative variation for previously studied, but poorly characterized, domestication and improvement traits. In the first example, we found the missing enzymatic gene and mutations for a genome-wide association study (GWAS) volatile that causes an undesirable fruit aroma—the "smoky" flavor. Next, we discovered that a major domestication fruit weight quantitative trait locus (QTL) is caused by higher expression from a gene within a tandem duplication, rather than a previously proposed promoter SNP. Finally, we dissected a remarkably complex case of epistasis involving four SV mutations, including duplications and transposon insertions, in three MADS-box genes, which was required to prevent a loss of productivity following the introduction of an important harvesting trait in modern tomato production. Notably, in the last two examples, modest changes in expression due to changes in gene copy number were predicted from our de novo assembled genomes, and we used CRISPR-Cas9 genome editing to directly link quantitative relationships between gene dosage and phenotype.

Our findings demonstrate the prevalence and importance of SVs at the population scale and highlight the underexplored roles of SVs in quantitative trait variation. As sequencing technologies continue to advance, SV discovery in other organisms will follow. Our work establishes a foundation for integrating computational, bioinformatic, quantitative genetic, and genome-editing tools to better understand genotype-to-phenotype relationships.

## Redundancy and Compensation in the Control of Plant Stem Cell Proliferation

C-T. Kwon, L. Aguirre

The classical CLAVATA (CLV) signaling pathway controls plant stem cell proliferation and meristem size in multiple species, and mutations in components of this pathway are associated with crop domestication and improvement. The *CLV3* small signaling peptide and the homeobox gene *WUSCHEL* constitute the core of this conserved feedback circuit. Mutations in *CLV3* result in enlarged meristems and additional flower and fruit organs because of the overproliferation of stem cells that causes enlarged shoot meristems. We previously found that the tomato ortholog of *CLV3* (*SlCLV3*) has close paralog *SlCLE9*, which is absent in *Arabidopsis* despite having 32 CLE family members. Interestingly, both *SlCLE9* and *SlCLV3* are highly up-regulated in *slclv3* mutant meristems, indicating that *SlCLE9* may act as a backup signaling peptide in the absence of *SlCLV3* during meristem proliferation. Single mutants of *slcle9* resemble wild-type plants, but *slclv3 slcle9* double mutants show a substantial enhancement of meristem size. Additional genetic and molecular experiments showed that *SlCLE9* is the main player in an "active compensation" mechanism whereby up-regulation of *SlCLE9* buffers stem cell homeostasis.

To understand how the peptide compensation has evolved in tomato and related Solanaceae, we mutated the orthologs of *SlCLV3* and *SlCLE9* in the Solanaceae fruit crop *Physalis grisea* (groundcherry) using CRISPR-Cas9 technology. Surprisingly, groundcherry *pgclv3* mutants were largely unaffected, suggesting *PgCLE9* could be a better compensator than tomato *SlCLE9*. This evolutionary distinction could be due to differences in peptide amino acids or variation in transcriptional compensation, which we are currently testing. Notably, *pgcle9* single mutants were normal, similar to tomato *slcle9* single mutants. We are now generating higher-order mutants and also targeting multiple ligand receptors of the CLV pathway in both species. Finally, to decipher the evolution of both peptides within Solanales, some of which lack the *cle9* paralog, we are also targeting the orthologs of *SlCLV3* in other Solanaceae species. By precise phenotypic comparisons and transcriptome profiling of the CRISPR knockouts from various species, we can resolve the mechanisms underlying how CLE redundancy and compensation evolves in closely related plants.

Taking advantage of a large set of CRISPR *SlCLV3* promoter alleles generated in the laboratory (see below), we are also studying the dosage response of compensation. We have combined a subset of *slclv3* promoter alleles with *slcle9,* and phenotyping the resulting double mutants showed that compensation activates when *slclv3* is moderately compromised.

In contrast, weak alleles of *slclv3* show no activation of compensation. This has led to the hypothesis that compensation might scale linearly with allelic strength, meaning that *slclv3* alleles that are stronger than moderate will exhibit more compensation from *SlCLE9* and thereby a more pronounced enhancement by the presence of *slcle9* null mutations. These findings have implications for compensation in other signaling circuits and perturbations to those circuits that might result from hypomorphic natural mutations.

## Dissecting *cis*-Regulatory Control of Gene Expression and Quantitative Trait Variation

X. Wang, L. Aguirre

Whereas the CLV-WUS circuit is conserved among different angiosperm lineages, little is known about the *cis*-regulatory control of the core components. Previous work in our laboratory has shown that CRISPR multiplex targeting of the upstream promoter region in the tomato homolog of *CLV3*, *SlCLV3*, can generate a range of different lesions that presumably perturb sets of *cis*-regulatory elements (CREs). These alleles provide a quantitative phenotypic readout in the form of increased chambers of seeds in fruits, known as locules. Interestingly, although the majority of hypomorphic promoter alleles generated by this multiplex targeting changed *SlCLV3* expression and increased locule number, the size and type of lesion created did not correlate directly (linearly) with the degree of expression change. This phenomenon indicates a disproportionate input by different CREs to the expression of *SlCLV3*. A detailed examination of the *SlCLV3* promoter revealed multiple conserved noncoding sequences between related members of the Solanaceae family that likely contain CREs that contribute to the regulation and expression of *SlCLV3*. By selectively perturbing individual CREs located within these regions and combining them in the promoter of *SlCLV3*, a better understanding is gained of if and how CREs contribute in a linear or nonlinear manner between gene expression and phenotypic changes.

We have taken further advantage of our CRISPR promoter mutagenesis drive system to generate more than 30 unique *SlCLV3* promoter alleles. A detailed quantitative phenotypic analysis of this allelic series pointed to specific proximal and distal *cis*-regulatory regions that affect *SlCLV3* function and phenotypic

outputs. The most striking effects were observed in lines having distal deletions >3 kb upstream where multiple conserved noncoding sequence (CNS) regions reside. Moreover, taking advantage of *trans*-targeting by CRISPR transgenes, we could combine lesions in *cis*, which showed both additive and synergistic effects from proximal and distal CNS deletions. Our generation of additional lines with combined CNS mutations will help to further elucidate interactions between *cis*-regulatory regions influencing *SlCLV3* expression, whether those interactions are additive, epistatic, or synergistic. This work is providing important insights into how the CLV3-WUS circuit may be rebalancing following *cis*-regulatory mutagenesis.

## Genetic and Molecular Dissection of the Inflated Calyx Syndrome

Jia He

The evolution of morphological novelties has been a long-standing interest for biologists. The diverse floral organ traits among Solanaceae species are ideal for studying this question. One spectacular but understudied trait is the so-called "Chinese lantern" or Inflated Calyx Syndrome (ICS) found in genera like *Physalis*, *Withania*, and *Nicandra*, in which sepals continue growth after anthesis, forming balloon-like structures encapsulating the fruits. Previous studies in *Physalis* suggested the heterotopic expression of a MADS-box transcription factor (MPF2) to be key to the evolution of the ICS. However, a later study showed that MPF2 expression in the calyx was widespread across species with or without ICS, suggesting other factors to be the determinants of this developmental process. A recent phylogenetic analysis of the ICS in the Physalideae tribe revealed that the inflated calyx had evolved many times in a stepwise and irreversible fashion. These results encouraged us to revisit the genetic and developmental mechanisms of the ICS.

Using CRISPR-Cas9 genome editing, we generated MPF2 coding sequence mutations resulting in premature stop codons in *Physalis grisea*. Compared to wild type, these mutants showed mildly altered fruit shape but no difference in calyx development, suggesting that MPF2 is not essential to the ICS. This evidence raised questions to previous conclusions and theories regarding ICS, highlighting the need for a comprehensive study of this phenomenon.

To begin dissecting the molecular and evolutionary mechanisms underlying ICS, we are performing global expression analysis during calyx development. Multiple MADS-box genes in tomato and other species are known to determine floral development, and played critical roles in the evolution of flowering plants. By analyzing the expression of MADS-box genes along the calyx developmental stages, we are narrowing our focus down to a small number of MADS-box genes that showed notable expression patterns. CRISPR-Cas9 editing of these genes is being carried out to provide genetic evidences for their potential roles in ICS.

## Engineering Fruit Crops for Urban Agriculture

C-T. Kwon

A major frontier in modern agriculture is achieving sustainable food production. Expanding farming to urban environments holds great promise to help reach this goal. However, a major limitation of urban agriculture (e.g., rooftop farms, vertical farms in warehouses) is that crops must be extremely compact and

rapid cycling. Thus, leafy green vegetables such as lettuce dominate these highly restrictive production systems. To have a meaningful impact, urban agriculture must be expanded to more crops. Fruit crops are highly desired, but developing new varieties whose architectures and productivities are optimized for these specific growth parameters is challenging. Based on our discovery of key genes that control important productivity traits in tomato, we aimed to develop tomato for urban agriculture production.

Using a forward genetics approach, we discovered a new regulator of tomato stem length (*SlER*) and devised a trait-stacking strategy along with two regulators of flowering (*SP* and *SP5G*) to combine mutations causing precocious growth termination, rapid flowering, and condensed shoots (Fig. 1). Using CRISPR-Cas9 genome editing of these three genes restructured vine-like growth of both large- and small-fruited tomato plants into a compact, early-yielding "triple-determinate" plant. We confirmed yields and fruit quality were maintained in field-based productivity trials, and we further demonstrated cultivation in an indoor vertical farm with LED lights and hydroponic growth systems. Our approach provides a simple genetic solution to transform any tomato genotype into
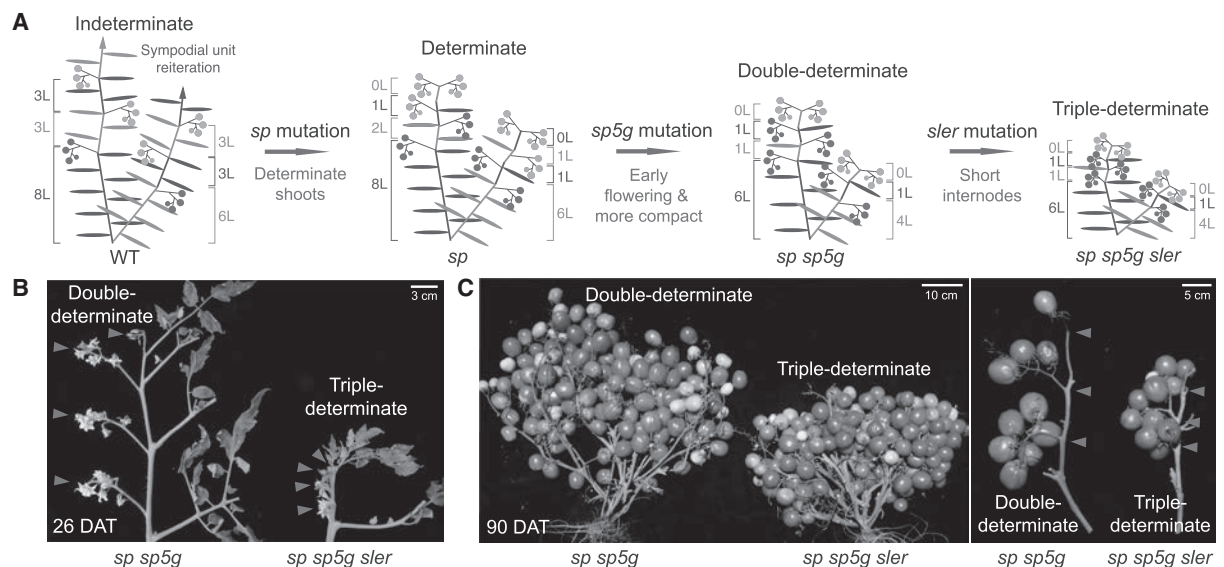


**Figure 1.** Creating highly compact, rapid flowering tomatoes by genome editing. (*A*) A trait-stacking strategy that combines mutations that cause precocious growth termination, rapid flowering, and shorter stems to create "triple-determinate" tomato varieties. (*B*) A comparison of double (*sp sp5g*) and triple (*sp sp5g sler*) determinate tomato genotypes. Basal axillary shoots of *sp sp5g* and *sp sp5g sler*. Arrowheads indicate inflorescences. (*C*) Mature plants and fruits (*left*) and associated shoots and inflorescences (*right*) from field-grown plants of double and triple determinate genotypes. Leaves were removed to expose fruits. Arrowheads indicate inflorescences.

a highly compact, rapid-cycling variety adapted for urban agriculture. Notably, we further showed that because of functional conservation, the same stem length regulator can be used to customize groundcherry (*Physalis grisea*), a berry fruit related to tomato, for urban agriculture. Interestingly, *pger* mutants alone had a dwarf phenotype that was more severe than tomato *sler* mutants and resembled *sp sp5g sler* triple-mutant plants, suggesting even closely related species may require different genetic solutions to create new varieties suitable for urban farms. For both fruit crops, our strategy allowed rapid generation of compact plants that maintain high productivity under high-density conditions, and which we show can easily be implemented into traditional breeding programs.

## *cis*-Regulatory Dissection of Pleiotropy

A. Hendelman, S. Zebell

Our development of a multiplex CRISPR-Cas9 drive system to generate series of promoter alleles has opened the door to address questions not only on *cis*-regulatory control of gene expression and phenotypic variation, but also the control of pleiotropy. In the last year, we generated promoter allelic series for more than 10 developmental genes, and in all cases we resolved multiple transgene-free homozygous alleles with a range of quantitative effects. These results show that many developmental genes are dose-sensitive, suggesting targeting promoters of such genes and also downstream *cis*-regulatory regions is a powerful approach to quantitatively tune traits linked to productivity. Notably, we found one transcription factor gene in which promoter alleles revealed hidden

pleiotropy. Previous characterized natural alleles that were assumed to be null mutations affected inflorescence complexity; however, a deeper phenotypic characterization revealed a low penetrance of vegetative meristem defects. By generating nearly 30 promoter alleles of this gene, we resolved this pleiotropy by showing that proximal mutations affect embryonic meristem function, whereas distal mutations affect reproductive meristem function. Meristem ATAC-seq assays along with CNS analyses described above show that these sequences are important *cis*-regulatory regions underlying the pleiotropy. Further genetic and molecular analyses will help to pinpoint the contributions of specific predicted CREs, and additional CRISPR-Cas9 of this gene and regulatory regions in related Solanaceae and *Arabidopsis* will show the text that this pleiotropy, and the underlying promoter elements, are evolutionarily conserved.

**PUBLICATIONS**

Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol* **20:** 224.

Eshed Y, Lippman ZB. 2019. Revolutions in agriculture chart a course for targeted breeding of old and new crops. *Science* **366:** eaax0025.

Kwon CT, Heo J, Lemmon Z, Capua Y, Hutton S, Van Eck J, Park SJ, Lippman ZB. 2019. Rapid customization of Solanaceae fruit crops for urban agriculture. *Nat Biotechnol* **38:** 182–188.

Rodriguez-Leal D, Xu C, Kwon CT, Soyars C, Demesa-Arevalo E, Man J, Lei L, Lemmon Z, Jones DS, Van Eck J, et al. 2019. Evolution of buffering in a genetic circuit controlling plant stem cell proliferation. *Nat Genet* **51:** 786–792.

Soyk S, Lemmon Z, Sedlazeck FJ, Jiménez-Gómez JM, Alonge M, Hutton S, Van Eck J, Schatz MC, Lippman ZB. 2019. Duplication of a domestication locus neutralized a cryptic variant that caused a breeding barrier in tomato. *Nat Plants* **5:** 471–479.

# EPIGENETIC INHERITANCE IN PLANTS AND FISSION YEAST

**R. Martienssen**    C. Alves         K.H.(T.) Cheng    H.S. Kim       A. Schorn
                      E. Anderson      W. Dahl           S-C. Lee       T. Sharma
                      B. Berube        J. Dolata         U. Ramu        A. Shimada
                      S. Bhattacharjee E. Ernst          M. Regulski    J. Simorowski
                      J. Cahn          M. Gutbrod        B. Roche       J. Steinberg

Plants and fission yeast provide excellent models for epigenetic mechanisms of transposon regulation, heterochromatic silencing, and gene imprinting, important both for plant breeding and for human health. We are investigating the role of RNA interference (RNAi) in heterochromatic silencing in the fission yeast *Schizosaccharomyces pombe* and in the model plant *Arabidopsis thaliana*, as well as continuing our work on *Zea mays*. In fission yeast, we have found that RNAi promotes DNA replication and repair, as well as histone modification required for centromere function. In quiescence, RNAi becomes essential because it is required for release of RNA polymerase I. We have found that long noncoding RNA plays an important role in this process, which may reflect an ancient role for RNAi. In plants, we have shown that histone modifications are reprogrammed in pollen (the male germline) to prevent epigenetic inheritance by incorporating a histone variant resistant to Polycomb. We have also found that epigenetically activated small interfering RNAs (easiRNAs) in pollen depend on RNA polymerase IV and are essential for pollen development in *Capsella*, a relative of *Arabidopsis*. Finally, we have shown that somatic easiRNAs control transposition of long terminal repeat (LTR) retrotransposons, which are related to HIV and other retroviruses. We have developed a method to sequence retrotransposon and retroviral replication intermediates contained in virus-like particles, which can readily detect active retrotransposons in very large genomes. We continue to develop duckweeds for biofuel by sequencing the genomes of several species and by developing an efficient transformation system in the clonally propagated aquatic macrophyte *Lemna minor*. We are beginning to obtain significant yields of oil in transgenic duckweed.

This year we said goodbye to graduate student Michael Gutbrod, who graduated and started his postdoc at MIT, and to postdocs Sonali Bhattacharjee, who joined *Genes and Development*, and Andrea Schorn, who started her own group at CSHL. Will Dahl graduated from Cornell and started his Ph.D. at Brandeis after five years of duckweed research in the laboratory. We welcomed graduate student Teri Cheng.

## Reprogramming of Histone H3K27me3 Resets Epigenetic Memory in Plant Paternal Chromatin

Y. Jacob,* C. LeBlanc,* R. Martienssen [in collaboration with P. Voigt, University of Edinburgh; J. Becker, Gulbenkian Institute, Lisbon; T. Higashiyama, Nagoya University, Japan; F. Berger, Gregor Mendel Institute, Vienna]; *present address: Yale University

Epigenetic marks are reprogrammed in the gametes to reset genomic potential in the next generation. In mammals, paternal chromatin is extensively reprogrammed through the global erasure of DNA methylation and the exchange of histones with protamines. We have previously shown that the paternal epigenome is also reprogrammed in flowering plants, in that DNA loses small RNA–directed DNA methylation, although most DNA methylation and histones are retained in sperm. Along with our collaborators, we have uncovered a multilayered mechanism by which histone H3K27me3 is lost from sperm chromatin in *Arabidopsis*. This mechanism involves the silencing of H3K27me3 writers, activity of H3K27me3 erasers, and deposition of a sperm-specific histone, H3.10 (MGH3), which we have shown is immune to lysine 27 methylation because of an altered amino-terminal amino acid sequence. The loss of H3K27me3 facilitates the transcription of genes essential for spermatogenesis and preconfigures sperm with a chromatin state that forecasts gene expression in the next generation. Thus, plants have evolved a specific mechanism to simultaneously differentiate male gametes and reprogram the paternal epigenome.

## *Arabidopsis* DNA Replication Initiates in Intergenic, AT-Rich Open Chromatin

U. Ramu, C. LeBlanc,* R. Martienssen [in collaboration with M, Vaughn, Texas Advanced Computer Center; H. Bass, Florida State University; L. Hanley-Bowdoin and B. Thompson, North Carolina State University]; *present address: Yale University

The selection and firing of DNA-replication origins play key roles in ensuring that eukaryotes accurately replicate their genomes. This process is not well documented in plants, largely because of a new functional assay to label and map very early replicating loci that must, by definition, include at least a subset of replication origins. *Arabidopsis* (*Arabidopsis thaliana*) cells were pulse labeled with 5-ethynyl-2′-deoxy-uridine (EdU), and nuclei were subjected to two-parameter flow sorting. We identified more than 5,500 loci as initiation regions (IRs), the first regions to replicate in very early S phase. These were classified as strong or weak IRs based on the strength of their replication signals. Strong initiation regions were evenly spaced along chromosomal arms and depleted in centromeres, whereas weak initiation regions were enriched in centromeric regions. IRs are AT-rich sequences flanked by more GC-rich regions and located predominantly in intergenic regions. Nuclease sensitivity assays indicated that IRs are associated with accessible chromatin. Based on these observations, initiation of plant DNA replication shows some similarity to, but is also distinct from, initiation in other well-studied eukaryotic systems.

## RNA Polymerase IV Plays a Crucial Role in Pollen Development in *Capsella*

F. Borges,* R. Martienssen [in collaboration with C. Kohler, Swedish Agricultural University, Uppsala]; *present address: INRA Versailles, Paris

In *Arabidopsis* (*Arabidopsis thaliana*), DNA-dependent RNA polymerase IV (Pol IV) is required for the formation of transposable element (TE)-derived small RNA transcripts. These transcripts are processed by DICER-LIKE3 into 24-nucleotide small interfering RNAs (siRNAs) that guide RNA-directed DNA methylation. In the pollen grain, Pol IV is also required for the accumulation of 21/22-nucleotide epigenetically activated siRNAs, which likely silence TEs via posttranscriptional mechanisms. Despite

this proposed role of Pol IV, its loss of function in *Arabidopsis* does not cause a discernible pollen defect. Here, we show that the knockout of *NRPD1*, encoding the largest subunit of Pol IV, in the Brassicaceae species *Capsella* (*Capsella rubella*), caused postmeiotic arrest of pollen development at the microspore stage. As in *Arabidopsis*, all TE-derived siRNAs were depleted in *Capsella nrpd1* microspores. In the wildtype background, the same TEs produced 21/22-nucleotide and 24-nucleotide siRNAs; these processes required Pol IV activity. Arrest of *Capsella nrpd1* microspores was accompanied by the deregulation of genes targeted by Pol IV-dependent siRNAs. TEs were much closer to genes in *Capsella* compared with *Arabidopsis*, perhaps explaining the essential role of Pol IV in pollen development in *Capsella*. Our discovery that Pol IV is functionally required in *Capsella* microspores emphasizes the relevance of investigating different plant models.

## *Arabidopsis* Retrotransposon Virus-Like Particles and Their Regulation by easiRNA

S-C. Lee, E. Ernst, B. Berube, F. Borges,* J.S. Parent,** P. Ledon, A. Schorn, R. Martienssen; present addresses: *INRA Versailles, Paris; **Agriculture Canada, Ottawa

In *Arabidopsis*, LTR retrotransposons are activated by mutations in the chromatin remodeling gene *DECREASE in DNA METHYLATION 1* (*DDM1*), giving rise to 21- to 22-nt easiRNA that depend on RNA-dependent RNA polymerase 6 (RDR6). We purified virus-like particles (VLPs) from *ddm1* and *ddm1rdr6* mutants in which genomic RNA is reverse transcribed into complementary DNA. High-throughput short-read and long-read sequencing of VLP DNA (VLP DNA-Seq) revealed a comprehensive catalog of active LTR retrotransposons without the need for mapping transposition, as well as independent of genomic copy number (Fig. 1). Linear replication intermediates of the functionally intact *COPIA* element *EVADE* revealed multiple central polypurine tracts (cPPTs), a feature shared with HIV in which cPPTs promote nuclear localization. For one member of the *ATCOPIA52* subfamily (which we named *SISYPHUS*), cPPT intermediates were not observed, but abundant circular DNA indicated transposon "suicide" by futile autointegration within the VLP (Fig. 1). easiRNA targeted *EVADE* genomic RNA, polysome association

A



B
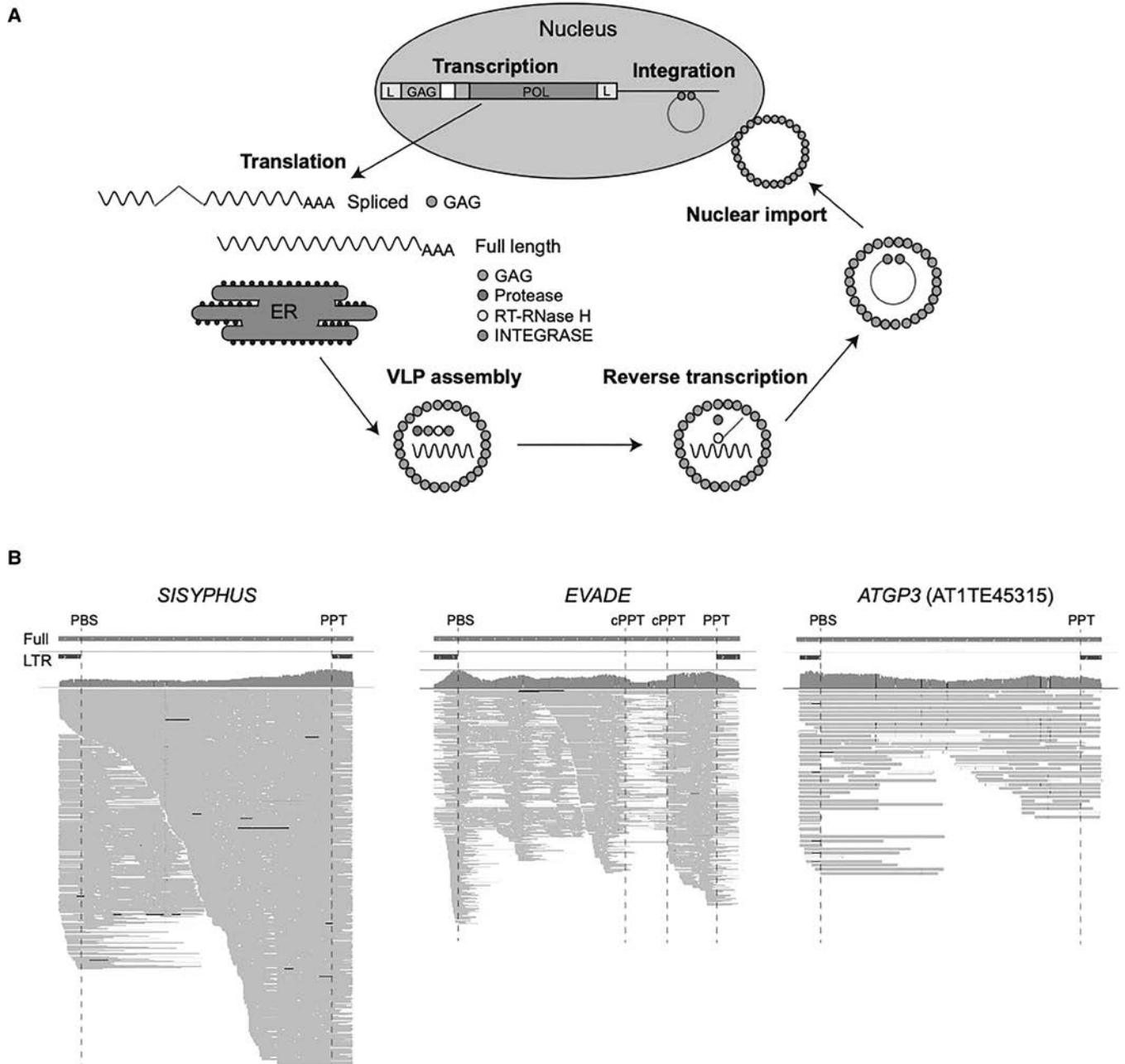
**Figure 1.** Long-read sequencing of virus-like particle (VLP) cDNA. (*A*) A schematic diagram of production of virus-like particles from functional LTR retrotransposons. For *COPIA* elements in *Arabidopsis*, introns are spliced from subgenomic RNA to make only GAG proteins. Polyproteins encoded by full-length RNA comprise enzymes essential for reverse transcription and integration. After polyproteins are cleaved by protease, RT-RNase H processes full-length genomic RNA by reverse transcription. Double-stranded VLP DNA enters the nucleus and inserts into new genomic loci mediated by integrase. (L) LTR, (ER) endoplasmic reticulum. (*B*) Alignments of Oxford Nanopore Technologies long reads from *ddm1* VLP DNA. The central polypurine tract (cPPT), PBS, and PPT positions are indicated as dashed lines relative to full and LTR annotation of *SISYPHUS* (AT3TE76225), *EVADE* (AT5TE20395), and *ATGP3* (AT1TE45315). Gaps in individual reads are indicated with black horizontal lines, and sequence mismatches are shown as dots in the read alignments. Pileups of linear intermediates are observed for *EVADE*, whereas a continuous distribution of fragment lengths is observed for *SISYPHUS*. Many of these reads are circular permutations, indicating futile auto-integration of *SISYPHUS* within the VLP.

of *GYPSY* (*ATHILA*) subgenomic RNA, and transcription via histone H3 lysine-9 dimethylation. Interestingly, there were partial VLP DNA fragments from nonfunctional *ATHILA* elements accumulated only in *ddm1rdr6*, suggesting potential roles of easiRNAs in controlling reverse transcription. *EVADE* is the most active LTR copia retrotransposon in *Arabidopsis* and produces easiRNAs from the *GAG* gene when it is desilenced. Our translatome data suggested easiRNAs do not generally regulate translation. Two micro RNA (miRNA) target sites flank the easiRNA cluster in the *GAG* gene, but no cleavage sites were detected, suggesting RDR6 was recruited without cleavage. *EVADE* easiRNAs were not detected in *ddm1* pollen, consistent with maternal but not paternal silencing previously reported. *EVADE* copy number and RNA levels in *ddm1rdr6* were strikingly higher than *ddm1*, suggesting easiRNAs control retrotransposition by direct RNA interference. VLP DNA-Seq provides a comprehensive landscape of LTR retrotransposons and their control at transcriptional, posttranscriptional, and reverse transcriptional levels and can be used to detect active retrotransposons in very large genomes.

## Conserved Chromosomal Functions of RNA Interference

M. Gutbrod, R. Martienssen

RNA interference (RNAi), a cellular process through which small RNAs target and regulate complementary RNA transcripts, has well-characterized roles in posttranscriptional gene regulation and transposon repression. Recent studies have revealed additional conserved roles for RNAi proteins, such as Argonaute and Dicer, in chromosome function. By guiding chromatin modification, RNAi components promote chromosome segregation during both mitosis and meiosis and regulate chromosomal and genomic dosage response. Small RNAs and the RNAi machinery also participate in the resolution of DNA damage. Interestingly, many of these lesser-studied functions seem to be more strongly conserved across eukaryotes than are well-characterized functions such as the processing of microRNAs. These findings have implications for the evolution of RNAi since the last eukaryotic common ancestor, and they provide a more complete view of the functions of RNAi.

## RNA Polymerase I Regulation by RNA Interference in Cellular Quiescence Relies on a Novel Class of Long Noncoding RNAs

B. Roche, R. Martienssen [in collaboration with B. Arcangioli, Institut Pasteur, Paris]

Most cells in nature, and in the human body, are present in a nondividing state ($G_0$). Cellular quiescence is an important $G_0$ state characterized by its reversibility and metabolic activity, found, for example, in stem cells and memory lymphocytes. To establish and maintain quiescence, cells undergo a major transcriptional reprogramming. However, this phenomenon is still largely unexplored at the molecular level and little is known about the mechanisms involved in maintaining viability in quiescent states. Because cellular quiescence is in many aspects an epigenetic transition, we hypothesized that specific epigenetic pathways would be involved. Indeed, we have found that several of these pathways are rewired in cellular quiescence and become essential specifically in $G_0$. Using fission yeast as a model system for the fundamental biology of quiescence, we have designed a specific strategy to identify $G_0$-specific suppressors of these new functions. We previously reported that RNA interference acquires a novel, and essential, nucleolar function in $G_0$ cells, in which it regulates the epigenetic state of ribosomal DNA (rDNA) repeats by regulating RNA polymerase I transcription (Roche et al., *Science* **354:** aah5651 [2016]). Mutants in key RNAi factors, such as Dicer and Argonaute, display a major loss of viability specifically in quiescent cells because of an overaccumulation of H3K9 methylation on rDNA repeats. Strikingly, this phenomenon can be suppressed by specific mutants in the RNA polymerase I holoenzyme itself. Dicer catalytic activity is required, showing that specific target RNAs are likely to mediate this function. We have discovered a novel group of long noncoding RNAs that are strongly up-regulated in Dicer-deficient $G_0$ cells. In particular, we have determined that at least one of these noncoding RNAs is necessary for

accumulating H3K9 methylation at rDNA, thus appearing to be the first *trans*-acting functional homolog of the promoter-associated RNA (pRNA) in human cells, a classical RNA-mediated rDNA silencing factor.

## The Genomics and Epigenomics of *Lemnaceae* for Biofuel Applications

E. Ernst, J. Simorowski, U. Ramu, R. Martienssen
[in collaboration with T. Michael, Salk Institute; J. Birchler, University of Missouri; E. Lam, Rutgers University; J. Shanklin, Brookhaven National Laboratory]

Genomics has greatly impacted breeding in domesticated crops, through genome- and marker-assisted selection, as well as finding genes underlying key traits. But crops with only minimal domestication, such as oil palm and duckweed, present challenges to this approach—especially when they are propagated asexually as clones. Although clones have the advantage of potentially fixing hybrid vigor, lack of germline passage, where epigenetic reprogramming occurs, can lead to epigenetic variation. Advances in single-molecule genomic sequencing and epigenomic profiling have enabled the rapid generation and assembly of large and highly repetitive plant genomes. We have employed single-molecule sequencing from Oxford Nanopore to complete the first long-read chromosomal assemblies of the *Lemna gibba* 7742 and *Lemna minor* 8627 genomes, as well as *S. polyrhiza* and *W. australiana* in collaboration with our colleagues, with updated gene annotations informed by full-length cDNA sequences from the ONT MinION. We have found that all four genomes lack CHH methylation, reflecting loss of CHROMOMETHYLASE2 (CMT2), and three have reduced 24nt siRNA,

reflecting most likely a novel DICER-LIKE 3 gene. We are now exploring potential implications for natural variation, genetic modification, and next-generation biofuels in *Lemnaceae*, aquatic plants with rapid clonal growth habit. We have overexpressed several genes involved in triacylglycerol production and have achieved substantial increases in TAG. We are testing a number of strategies to overcome subsequent defects in growth and clonal partitioning.

## PUBLICATIONS

Bhattacharjee S, Roche B, Martienssen RA. 2019. RNA-induced initiation of transcriptional silencing (RITS) complex structure and function. *RNA Biol* **169:** 1133–1146. doi:10.1080/15476286.2019.1621624

Schorn A, Martienssen RA. 2019. Getting in LINE with replication. *Mol Cell* **4:** 415–417. doi:10.1016/j.molcel.2019.04.023

*In Press*

Borg M, Jacob Y, Susaki D, LeBlanc C, Buendia D, Axelsson E, Kawashima T, Voight P, Boavida L, Becker J, et al. 2020. Targeted reprogramming of H3K27me3 resets epigenetic memory in plant paternal chromatin. *Nat Cell Biol* **22:** 621–629. doi:10.1038/s41556-020-0515-y

Gutbrod MJ, Martienssen RA. 2020. Conserved chromosomal functions of RNA interference. *Nat Rev Genet* **21:** 311–331. doi:10.1038/s41576-019-0203-6

Lee SC, Ernst E, Berube B, Borges F, Parent JS, Ledon P, Schorn A, Martienssen RA. 2020. *Arabidopsis* retrotransposon virus-like particles and their regulation by epigenetically activated small RNA. *Genome Res* **30:** 576–588. doi:10.1101/gr.259044.119

Wang Z, Butel N, Santos-Gonzalez J, Borges F, Yi J, Martienssen RA, Martinez G, Kohler C. 2020. Polymerase IV plays a crucial role in pollen development in *Capsella*. *Plant Cell* **32:** 950–966. doi:10.1105/tpc.19.00938

Wheeler E, Brooks AM, Concia L, Vera D, Wear EE, LeBlanc C, Ramu U, Vaughn MW, Bass H, Martienssen R, et al. 2020 *Arabidopsis* DNA replication initiates in intergenic, AT-rich open chromatin. *Plant Physiol* **183:** 206–220. doi:10.1104/pp.19.01520

# MOLECULAR SIGNALING EVENTS UNDERLYING ENVIRONMENTAL CONTROL OF PLANT GROWTH

U. Pedmale    M. Biggiani    Y. Hu    A. Matheus
              H. Gumber      S. Jang  S. Sankaranarayanan
              J. Hart        L. Lindbäck  O. Spassibojko

A fundamental question in biology, which remains unanswered, is how the environment of the organism regulates its growth and development. Both plants and animals interact with their environment; however, plants grow postembryonically as they are incapable of moving around. Unlike animals, plants do not have specific organs that see or that hear various stimuli, yet plants are sensitive to their surrounding environment and modify their growth according to various external and internal signals. Plants regularly face variability in growth conditions—temperature, light quality and quantity, herbivores, pathogens, water availability, etc. A plant responds to these biotic and abiotic factors and survives substantial fluctuations in its environment. A plant also must balance the range of potential threats and benefits confronting it and should make appropriate decisions on resource allocation. Remarkably, lacking a brain, plants can successfully integrate various cues and make appropriate decisions about growth. Such adaptability is essential to the sessile nature of plants. In some adaptive responses—for example, when the plants have to cope with climate change and increased competition for light—there is a decrease in productivity (yield, biomass) as the plant relocates resources to adapt better.

The goal of our laboratory is to determine the mechanisms behind how a plant perceives and successfully adapts to its environment. We also aim to understand how a plant must integrate intrinsic and extrinsic cues and "decide" how best to respond to environmental cues. Understanding how plants deal with and respond to a multitude of environmental signals could help to develop crops that cope with unfavorable growth conditions without significant changes in yield. Our laboratory primarily studies the effect of light environment on plant growth and development. Light is among the most relevant environmental signals because it not only drives photosynthesis, but also provides critical information about the local growth environment as well as seasonal time. Light is perceived by a complex array of photoreceptors, which include phytochromes (PHYA-E), cryptochromes (CRY1-2), phototropins (PHOT1-2), zietlupe family (FKF1, LKP2, and ZTL), and UVR8. Plants have developed various adaptive responses to interpret and utilize light directionality, quantity, and quality. In vegetational shading, when plants are under the shade of another plant, they perceive a decrease in the ratio of red to far-red light (R:FR) caused by the absorption of red light by chlorophyll and reflection of far-red light by the neighboring foliage. Simultaneously, there is also decrease in blue light and the available photosynthetically active radiation (PAR).

We focus on blue light–absorbing CRYs; apart from being an excellent genetic and molecular tool to tease out the complexities of growth and adaptation, there are still many open questions about the molecular function of CRYs in plants. Understanding the role of CRYs is not only appealing for agriculture, but it also has an impact on human health, which could make this field appealing to diverse funding agencies. CRYs regulate growth and development and provide circadian entrainment to both plants and animals. In metazoans, disruption of CRY activity is linked to cancer, altered behavior, magnetoreception, and metabolism. Therefore, understanding CRY function in plants is not only important as they are integral to growth of plants, but also can have an impact on human health.

## Decoding the Nature of Long-Distance Communication in Plants—The Case of Shoot Control of Root Growth

During shading, many aerial organs elongate rapidly, whereas the root growth is reduced with the delay in the emergence of the lateral roots. Roots not only serve as a mechanical anchor, but also play a vital role in the well-being of the entire plant. Therefore, a robust and well-developed root system is required for healthy

plant growth. As one can imagine, there is a negative cycle occurring during shading; shoot-perceived shade leads to reduced root growth, which in turn is unable to support the shoot—leading to unproductive plants. However, this phenomenon is an excellent model to understand growth at a systems level because of the different growth phenotypes observed in the various organs of the same plant, as well as enabling exploration of the nature of the interorgan and long-distance communication that is used to signal when a distant organ is exposed to an adverse environment. Unfortunately, and surprisingly, not much is known about the mechanisms that underlie reduced root growth seen during shading.

In this context, we performed a transcriptomic analysis of *Arabidopsis thaliana* seedlings exposed to 30 min to 5 days of shade. We found that stress-induced genes were up-regulated in shaded roots, compared to those grown under nonshading conditions. Furthermore, a group of 1,175 genes specifically induced in shaded roots were found to have W-box motifs in their promoters. Such elements are recognized by WRKY transcription factors, the known modulators of plant stress responses (Fig. 1). Specifically, a group of 12 WRKY genes were found to be consistently up-regulated in shaded roots throughout our time course. Among them, some WRKYs have already been characterized as regulators of root development in response to biotic and abiotic stresses, however, without any known involvement in light responses. Therefore, these 12 WRKYs were selected as likely regulators of shade avoidance responses in roots. To test this hypothesis, WRKY-overexpressing lines and CRISPR-mediated mutants are being generated in *Arabidopsis*. Root development of these lines is being tested under shading and nonshading conditions. Those WRKYs identified as regulators of root development will be further characterized by RNA-Seq and ChIP-seq of the mutant and overexpressing lines, respectively. With this approach, we intend to generate gene-regulatory networks that will help understand how root development is regulated under optimal and suboptimal growth conditions imposed by light. Another ongoing approach to identify shade response regulators in roots is a forward genetic screen of mutants. *CMT6* gene was identified in our transcriptome as a suitable marker for shade responses in roots and not in the shoot. A transgenic line harboring a reporter called GUS (β-glucuronidase) under the control of *CMT6* promoter was generated. These plants, under shade, express GUS in their roots.
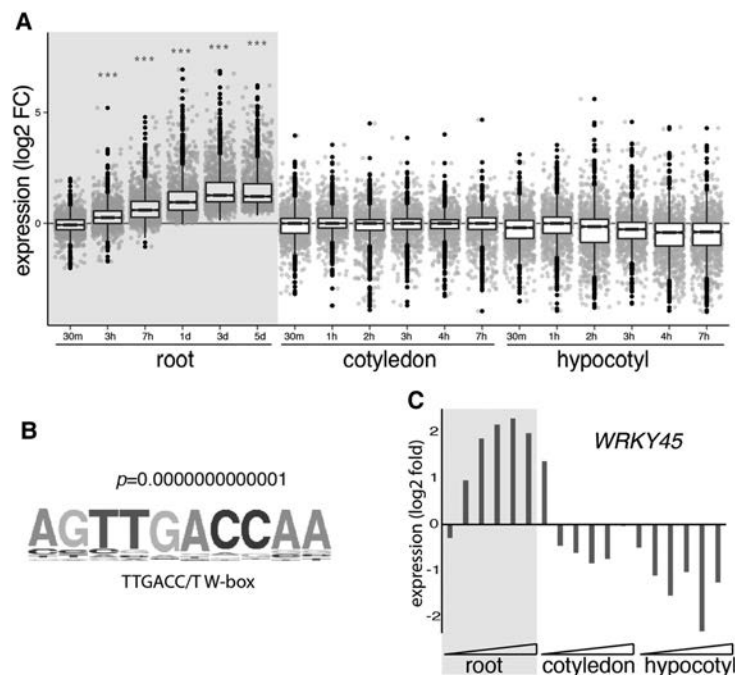


**Figure 1.** (*A*) Our RNA-Seq analysis identified a group of 1,175 genes that were expressed only in the root in shade (shaded) in comparison to the hypocotyl and the cotyledons. (*B*) The promoters of these genes contain a W-box to which WRKYs generally bind. (*C*) *WRKY45* is induced in the roots.

EMS-mutagenized *CMT6::GUS* lines were then generated to identify other factors using a forward genetic screen. We are currently screening the third generation of mutants that do not display GUS expression indicating mutants. These mutants with confirmed loss of GUS expression and increased lateral roots, similar to unshaded plants, will be sequenced to find the causative loci. Next, complementation assays and functional characterization of these genes will be performed. We hope that these combined approaches will help reveal the molecular mechanisms regulating shade responses in roots and provide molecular tools to bypass the negative effect of shade avoidance in root systems.

## The Role of RNA-Binding Proteins in Cryptochrome-Mediated Signaling

We also identified many RNA-binding proteins that copurified with CRY2. Recent studies reported copurification of RNA binding proteins with human CRY1/2. However, the significance of CRYs associating with RNA binding proteins is not known. Interestingly, the CRY2 nuclear speckles resemble those formed by pre-mRNA splicing factors, SR proteins, and other RNA-binding proteins in plants and animals, indicating that CRYs likely have a role in RNA metabolism. This observation may provide mechanistic insights into posttranscriptional control, known to be essential for animal and plant circadian biology, and into control of alternative splicing in plants by light. We are currently focusing on two unknown proteins that are hypothesized to bind to modified RNA; their mutants resemble *cry2* mutant plants, indicating an epistatic relationship between them. Furthermore, similar to our experiments in plants, we are also studying their orthologs in animals to determine whether they can also interact directly with mammalian CRY2. We have started profiling RNA modification in different mutant backgrounds using genome-wide approaches (Fig. 2). RNA modifications are emerging to be important regulators of various cellular processes not limited to protein translation, mRNA degradation, alternative splicing, and nuclear export. We are generating loss-of-function mutants of these RNA-binding proteins to understand their impact on plant growth, circadian rhythms, alternative splicing, and other physiological responses.

We recently discovered that CRY2 interacts with a plethora of RNA-binding proteins and CRY2 forms nuclear speckles that are highly similar to those
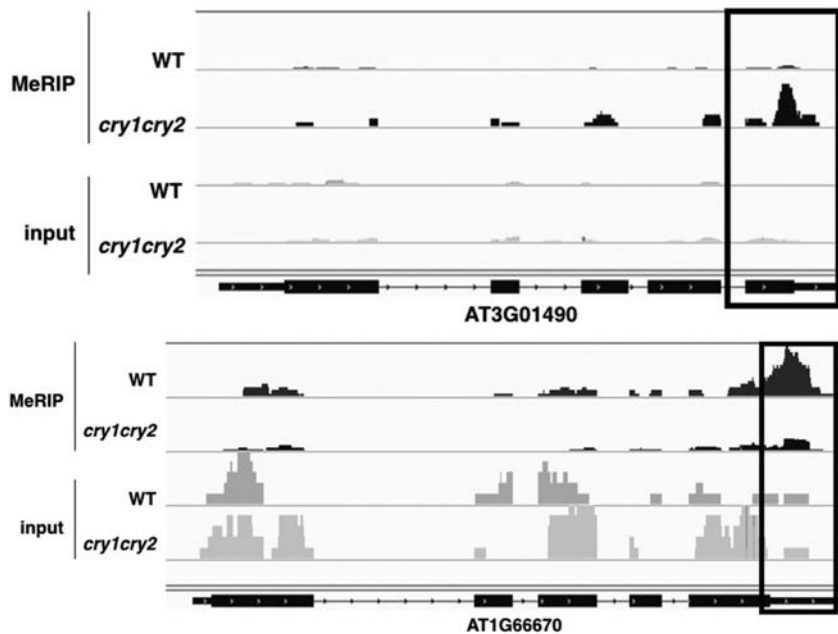


**Figure 2.** Examples of m6A peak identified by transcriptome-wide m6A profiling (MeRIP). Our results indicate that m6A is mainly deposited on 3′ ends of transcripts. We identified transcripts that were specifically methylated in wild-type or *cry1cry2* mutant plants.

observed for proteins involved in RNA processing and metabolism. One of the proteins directly interacting with CRY2 is a protein that binds to m6A-modified RNA, a so-called m6A reader. m6A is the most abundant and a highly dynamic mRNA modification. It has gained a lot of traction in the field of RNA biology lately because it plays a major role in various processes spanning different disciplines, such as cell proliferation in human cancers, entrainment of the circadian clock in mice, neuronal development in flies, or embryonic development in plants. Genetic analyses in *Arabidopsis* mutants provide strong evidence that CRY2 and the m6A reader act in the same pathway. This suggests that CRY2 might put the epitranscriptome of plants under the control of light, a novel and, so far, unexplored mechanism of gene regulation in plants. We are currently establishing methods for transcriptome-wide profiling of m6A depositions to elucidate the underlying molecular mechanisms governing signal transduction. First results indicate that RNA methylation is differentially regulated by the CRYPTOCHROME photoreceptors and m6A is mainly deposited on 3′ ends of transcripts (Fig. 2). Notably, very little is known about m6A regulation in other organisms. Because of the high conservation of the m6A machinery, we will likely make an impact on our understanding of this RNA modification beyond the field of plant biology.

## Molecular Determinants of CRY2 Protein Signaling and Stability

CRYs were first identified in plants and then discovered in animals. CRY2 protein accumulates in the dark and in vegetational shade and is readily degraded by the 26S proteasome under high intensities of blue light. Therefore, it is obvious that the CRY2 protein level and activity are tightly regulated by light to ensure proper signaling and response. However, the signaling events from the photoactivated CRYs to growth and development programs are not known. In animals and plants, CRY protein levels and activity are tightly modulated to influence signaling outcome. Therefore, to elucidate CRY signaling pathway, our laboratory has purified CRY2-containing protein complexes from *Arabidopsis thaliana* seedlings exposed to low-intensity blue light, which is typically encountered under shading. We identified CRY2-associated proteins by tandem affinity purification and mass spectrometry. Interestingly, the orthologs of some of the CRY2-associated proteins were also present in CRY protein complexes purified from human cells. This indicates that there could be a similar signaling mechanism in these two different evolutionary lineages.

Two of the highly enriched proteins in the CRY2-associated protein complex were deubiquitinases called UBP12 and UBP13. Deubiquitinases remove the ubiquitin protein covalently bound to a target protein. We found out that CRY2 and these deubiquitinases can interact directly in the nucleus of the cell. We hypothesize that CRY2 is ubiquitinated continuously, but, in certain situations, it recruits deubiquitinases to protect itself from proteasomal degradation such that downstream signaling can proceed. Next, we tested CRY2 protein levels in the deubiquitinase mutants and in plants in which they are overexpressed. Surprisingly, we found that CRY2 protein levels were very high in the deubiquitinase mutant and lower
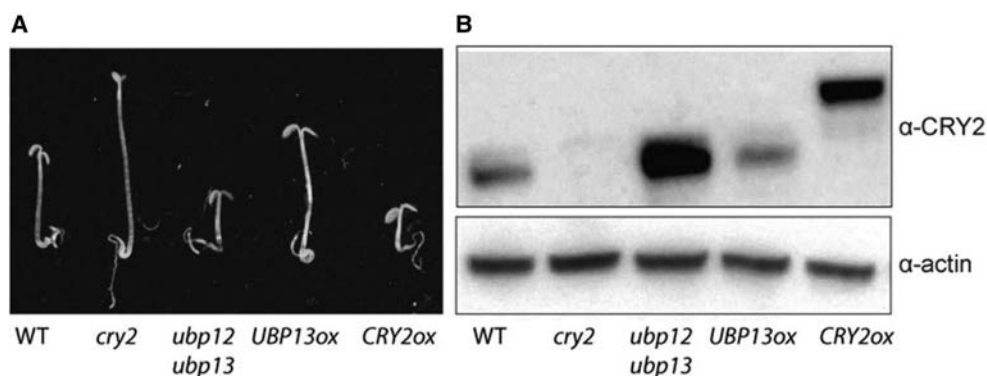


**Figure 3.** (*A*) Hypocotyl phenotype of seedlings of indicated genotypes grown in blue light for four days and (*B*) the CRY2 protein levels determined by immunoblotting using its specific antibody.

when overexpressed (Fig. 3). This matches with the physiological response exhibited by the seedling stem length of these genetic backgrounds. The deubiquitinase mutant seedling had a short hypocotyl when compared with the wild-type and the *cry2* mutant, and the overexpression line had a longer hypocotyl similar to *cry2* (Fig. 3). This observation indicates that the deubiquitinase–CRY2 protein complex likely modifies a protein that affects CRY2 protein levels which is likely to be a E3 ubiquitin ligase. Efforts are under way to identify ubiquitylated residues in the CRY2 protein and to evaluate the effect of catalytically dead deubiquitinases in plants. Using seedlings expressing catalytically dead deubiquitinase, we have identified several candidate E3 ligases that may affect CRY2 protein levels. Unlike animals, substrates for the large number (approximately 50) of plant deubiquitinases remain unidentified, except for histones. In parallel, to identify the E3 ubiquitin ligase responsible for targeting plant CRY2 for degradation, we are undertaking a forward genetic screen. Therefore, we are presented with a unique opportunity to study how deubiquitinases participate in the CRY2 signaling pathway and also their role in plant growth and development. In conclusion, our findings will provide novel insight into the regulation of CRYs by reversible ubiquitination, as well as the role of deubiquitinases in plant growth and development, which is largely unknown so far.

# GENOMICS

Advances in high-throughput sequencing technologies resulted in an explosive growth of multi-omics data. While presenting a tremendous opportunity for quantitative studies of numerous biological processes, crucial for both fundamental research and clinical applications, it also created a set of unique bioinformatics challenges for processing, integrating, and interpreting vast amounts of data.

Dr. **Alexander Dobin** and colleagues are biological data scientists working to resolve these challenges by developing highly efficient and accurate algorithms such as STAR, the popular RNA-Seq analysis software used by thousands of researchers worldwide. We are conceiving novel computational approaches to process data from emerging sequencing technologies, such as single-cell RNA-Seq and long-read nanopore sequencing, with a special emphasis on detecting RNA and DNA aberrations in tumors.

Another exciting research area in our group is functional annotation of the noncoding genome via integration of multi-omics data generated by the ENCODE, Roadmap Epigenomics, and GTeX consortia, essential for the deciphering of gene regulation mechanisms, interpretation of disease-associated variants in genome-wide association studies (GWASs), and understanding epigenetic effects in cancer biology.

There has been a growing appreciation in recent years that gene function is frequently context-dependent, with a large part of that context provided by the activities of other genes. But trying to understand how genes interact to produce function is a hugely complicated problem and one that appears likely to become more so as genomic information becomes more detailed. **Jesse Gillis** and colleagues are computational biologists who are challenging an oft-taken approach to the problem in which the functions of genes are interpreted in the context of networks derived from gene association data. Such networks consist of millions of interactions across thousands of genes, derived from protein-binding assays, RNA coexpression analysis, and other sources. Historically, many attempts to understand gene function through networks have leveraged a biological principle known as "guilt by association." It suggests that genes with related functions tend to share properties (e.g., physical interactions). In the past decade, this approach has been scaled up for application to large gene networks, becoming a favored way to grapple with the complex interdependencies of gene functions in the face of floods of genomics and proteomics data. Gillis' work centers on identifying the limits of the approach and making fundamental improvements to its operation, as well as applying those improvements to neuropsychiatric gene network data.

**Thomas Gingeras** and colleagues study where and how functional information is stored in genomes. These efforts help explain the biological and clinical effects of disease-causing gene mutations in humans and other organisms. Gingeras is a leader of the ENCODE (ENCyclopedia of DNA Elements) and the mouseENCODE and modENCODE (model genome ENCODE) projects of the National Institutes of Health. His research has altered our understanding of the traditional boundaries of genes, revealing that almost the entire lengths of genomes in organisms ranging from bacteria to humans can be transcribed into RNA (pervasive transcription) and that most RNA products made by a cell are not destined to be translated into proteins (noncoding, or ncRNAs). In fact, ncRNAs are proving to be involved in a variety of other important biological functions. Some have been shown to be critical components in the pre- and posttranscriptional and translational processes, as scaffolds upon which large protein complexes are assembled and as

extracellular signals. The initial studies that led to these observations have been extended to cover the entire human genome.

The insights of **W. Richard McCombie** and colleagues have led to the introduction and optimization of novel methods of high-throughput genome sequencing. His team has made it possible to catalog variation among individual organisms in a way that would have been unthinkable 10 years ago. They have brought online a new generation of Illumina sequencers and optimized their function to a level at which eight to 10 trillion DNA bases can be sequenced in a month. McCombie's team has been involved in international efforts culminating in genome sequences for maize, rice, and bread wheat—three of the world's most important food crops. They have also had an important role in projects to sequence the flowering plant *Arabidopsis thaliana* (the first plant genome to be sequenced), the fission yeast *Schizosaccharomyces pombe*, and the human genome and other important genomes. McCombie's group is currently involved in several important projects to resequence genes in patient samples that are of special interest to human health, including *DISC1* (a strong candidate gene for schizophrenia), looking for genetic variants implicated in bipolar illness and major recurrent depression. They are also looking for genes that contribute to cancer progression using whole-genome sequencing or a method called exome sequencing, which they developed with Greg Hannon to look at mutations in the regions of the genome that code for proteins.

Using multidisciplinary approaches that combine computational analysis, modeling, and prediction with experimental verification, **Doreen Ware**'s lab seeks a deeper understanding of the evolution of genome sequences in plants and their implications for agricultural improvement. By looking comparatively across the genomes of plants in the same lineage, they seek answers to the following questions: How are genes conserved and lost over time? What are the fates of duplicated genes? What is the impact of structural variation on phenotypic variation? Ware's team also studies gene regulation in plants, focusing on gene regulatory networks, targeting transcription factors and microRNA genes with the objective of understanding how these parts of the plant genome work together in determining spatial and temporal expression of genes. The lab had an important role in the project to produce a haplotype map reference genome of maize, spearheading the most comprehensive analysis of the crop yet. This has provided important information on the variation of the reference genome, as well as comparative data showing changes in the genome acquired through domestication and breeding. They have devoted special attention to examining diversity within maize, grape, and tomato, aiming to accelerate the development of strategies to introduce new germplasm that is needed to meet demands of increasing population and a changing environment. The lab also has brought fully sequenced genomes into an integrated data framework to enhance the power of their comparative studies. This past year, Ware was named as its principal investigator for the National Science Foundation–funded Gramene project, a comparative genomics resource for agriculturally important crops and models to support sustainable food and fuel production. Ware, as principal investigator for plants, has also helped lead an effort funded by the Department of Energy to create—out of many separate streams of biological information—a single, integrated cyber-"knowledgebase" for plants and microbial life.

# DEVELOPING NOVEL ALGORITHMS FOR COMPUTATIONAL GENOMICS

**A. Dobin**    A. Blibaum        B. Kaminow
                N. Castro-Pacheco    F. Mirhaj

In 2019, our team had continued developing state-of-the-art tools for genomics data analyses. We were mainly focused on the following projects.

## Quantifying Isoform Expression in Single-Cell RNA-Seq Data with STARsolo-Quant

High-throughput single-cell RNA sequencing (scRNA-Seq) is revolutionizing the analysis of complex biological systems, providing unprecedented insight into transcriptomic profiles of individual cells. The most current analyses concentrate on cell-type-dependent gene expression profiles, leaving behind other rich information contained in scRNA-Seq datasets. Popular scRNA-Seq mapping and quantification tools (10× Cell Ranger, Alevin, Kallisto/bustools, STARsolo) quantify gene expression by estimating the number of unique molecular identifiers (UMIs) for each gene. We have developed STARsolo-Quant, an expectation-maximization maximum likelihood (EM-ML) algorithm for the estimation of relative transcript isoform abundances from scRNA-Seq data. Our algorithm takes into account the idiosyncrasies of certain scRNA-Seq protocols; in particular, it models UMIs and 3′ or 5′ end cloning biases.

The main drawbacks of the current scRNA-Seq technologies are low sequencing depth (number of reads) in individual cells and massive variation of sequencing depth between cells. These obstacles prevent accurate quantification of isoforms in each cell, except for the most abundant isoforms inside the most sequenced cells. To mitigate this problem, STARsolo-Quant conducts isoform quantification over cell clusters. First, the standard STARsolo gene quantification is performed, including barcode demultiplexing, UMI collapsing, gene overlap, and counting. Next, the cells are split into clusters via uniform manifold approximation and projection (UMAP). Finally, the EM-ML isoform quantification step is performed for each cluster using reads from all cells in the cluster.

## Mercury: High-Quality Visualization and Reproducible Exploratory Analysis of scRNA-Seq Data

scRNA-Seq has recently become an extremely popular methodology for the analysis of biological systems, providing transcriptome profiles at single-cell resolution. However, scRNA-Seq data is challenging to analyze due to its high dimensionality and sparsity, making visual and supervised analysis of the data crucially important. Although hundreds of tools exist for statistical analysis of scRNA-Seq data, very few interactive processing and visualization solutions are available, and all have limitations in both their functionality and workflow. Here we present Mercury, a tool for high-quality visualization and reproducible exploratory analysis of scRNA-Seq data that addresses the shortcomings of current visualization software, unifies multiple scRNA-Seq analysis pipelines, and integrates a novel automated cell-type prediction method called ProtoCell. Mercury's close integration with analysis tools allows the user to define and run a pre-processing pipeline using Seurat or Scanpy in an intuitive user interface and without leaving the application.

Mercury also enhances reproducibility by tracking user actions, which can allow another researcher, or the users themselves, to replay the steps they took in their analysis. In addition to core workflow improvements, Mercury provides important analysis functions beyond existing tools, including high-quality 3D visualization and interactive integration with a novel automated cell-type identification method called ProtoCell. ProtoCell can make accurate predictions for cell types with few training examples, identify and cluster novel cell types, and make predictions for new cell types at interactive speeds. Mercury closes the gap in the current scRNA-Seq analysis pipeline, providing a sophisticated, usable, and open-source tool for reproducible visual scRNA-Seq analysis.

## Accuracy Assessment of Fusion Transcript Detection via Read-Mapping and De Novo Fusion Transcript Assembly-Based Methods

This work was done in collaboration with B. Haas (Broad Institute).

Accurate fusion transcript detection is essential for comprehensive characterization of cancer transcriptomes. Over the last decade, multiple bioinformatics tools have been developed to predict fusions from RNA-Seq, based on either read mapping or de novo fusion transcript assembly. We benchmark 23 different methods including applications we developed: STAR-Fusion and TrinityFusion, leveraging both simulated and real RNA-Seq. Overall, STAR-Fusion, Arriba, and STAR-SEQR are the most accurate and fastest for fusion detection on cancer transcriptomes. The lower accuracy of de novo assembly–based methods notwithstanding, they are useful for reconstructing fusion isoforms and tumor viruses, both of which are important in cancer research.

## Is It Time to Change the Reference Genome?

This work was done in collaboration with S. Ballouz and J. Gillis (CSHL).

The use of the human reference genome has shaped methods and data across modern genomics. This has offered many benefits while creating a few constraints. In Ballouz et al. (2019), an opinion piece, we outlined the history, properties, and pitfalls of the current human reference genome. In a few illustrative analyses, we focused on its use for variant calling, highlighting its nearness to a "type specimen." We suggested that switching to a consensus reference would offer important advantages over the continued use of the current reference with few disadvantages.

### PUBLICATIONS

Ballouz S, Dobin A, Gillis JA. 2019. Is it time to change the reference genome? *Genome Biol* **20:** 159. doi:10.1186/s13059-019-1774-4

Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. 2019. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol* **20:** 213. doi:10.1186/s13059-019-1842-9

# GENE EXPRESSION, FUNCTION, AND META-ANALYSIS

**J. Gillis**  M. Crow    B. Harris    J. Lee    H. Suresh
         S. Fischer  J. Hover    S. Lu     J. Werner
         N. Fox      R. Kawaguchi  S. Muzumdar

A dominant interest within computational biology is the analysis of gene networks to provide insight into diverse levels of functional activity, typically starting with regulatory interactions and moving up to more diffuse associations important for understanding systemic dynamics. Gene associations (of various sorts) are believed to encode functional interaction, and this interaction is frequently shown to be able to substantially predict gene function. This approach, commonly called "guilt by association," is embedded in everything from prioritization of de novo variants to uncovering novel regulatory interactions or mechanisms of disease. While black box–style network analyses are common, explaining the basis of how and why methods work is more rarely attempted. In the Gillis laboratory, we are developing network-based methods and software that improve both the sophistication and breadth of data available for determining how genes interact to produce function, particularly focusing on how genes interact to cause disease or cell phenotypes. Broadly, our research can be divided into methods development and our own research applications—often carried out in collaboration with other laboratories to test computational predictions experimentally. In addition to Jesse Gillis, the members of the Gillis laboratory are postdoctoral fellows Maggie Crow, Stephan Fischer, Risa Kawaguchi, Hamsini Suresh, and Sukalp Muzumdar; graduate students Shaina Lu, Ben Harris, and Jonathan Werner; and computational science developers Nathan Fox, John Lee, and John Hover.

## Neuronal Cell-Type Replicability and Discovering Interneuron Signatures

The mammalian brain is a complex organ involving millions to billions of neurons. Identifying neuronal cell types and how these cell types interact is an essential step toward understanding the organization of the brain. The Brain Initiative Cell Census Network (BICCN) is an NIH initiative that aims at defining a taxonomy of robust neuronal cell types that can be characterized through multiple modalities (e.g., transcriptional similarity, morphology, and electrophysiology). In the past two years, the BICCN generated several large molecular data sets; preliminary analyses of the motor cortex data yielded more than 100 potential cell types. Most of these cell types are novel and await experimental validation. We refined these results by identifying cell types that replicate across data sets. A replicable biological signal is more likely to remain consistent across modalities, brain regions, and species. To accomplish this, we extended MetaNeighbor, a statistical framework previously developed in the laboratory to quantify network cluster replicability, to scale up to one million cells. We also showed that MetaNeighbor's heuristic approach for defining informative genes is near optimal. We then applied MetaNeighbor across seven data sets comprising more than half a million cells. We first showed that broad classes of excitatory neurons and interneurons are highly replicable across data sets, independently of the clustering strategy or sequencing technology used. At a finer level, out of the more than 100 original cell types, 59 neuronal clusters replicated among two out of seven data sets, with 33 clusters recurring in at least half the data sets. These results depended more strongly on clustering and technology, suggesting uncertainty about the exact contour of cell types and highlighting the importance of manual curation in current computational analyses. Despite this wealth of data, the exact definition of a "cell type" is still fluid. With our framework, we will be able to further characterize BICCN cell types in terms of molecular markers and biological properties of neurons, such as molecules involved in synaptic communication.

We also used the improved MetaNeighbor to assess the robustness and conservation of interneuron signatures. We determined which gene families are distinctively expressed across cell types within each species, and evaluated the degree of conservation in gene family member expression across species.

On average, we found that 11% of gene families contributed to cell identity in all species, 10% contributed only in primates, and 22% in any one species. However, there was substantial variability across cell types, so the majority of gene families (66%) contributed to at least one type in all species and half were primate-specific at least once. This indicates that no single process contributes to differences in cell phenotypes across these species. Furthermore, we also found that glutamatergic cell types were typically more diverged between primates and mouse than GABAergic cell types. These results are in good agreement with previous work to show that GABAergic cell-type profiles are maintained across distant species (reptiles/rodents) and provide a strong foundation for ongoing work to define meta-analytically robust interneuron signatures.

## Using Coexpression as a Measure for Functional Relationships

Coexpression of genes is thought to reflect the shared function of the gene products. We investigate the informative power of coexpression in a number of contexts. First, we aggregated nearly all published bulk RNA-seq coexpression data for 14 species to provide a gold-standard data resource with a high, well-powered signal-to-noise ratio. We have made this aggregated data resource available online on a web server called CoCoCoNet. CoCoCoNet also allows users to identify gene modules conserved across species and to discover coexpression relationships with their gene(s) of interest.

We then developed a novel metric of gene functional conservation based on coexpression. We predicted that genes are more functionally conserved across species if their coexpressed gene relationships inside the species are consistent with the orthologs in another species. This "coexpression neighborhood" or "fingerprint" allows us to augment sequence-based ortholog annotations with functional activity information. We found that with sufficient data, as few as 10 genes can be used to form a stable fingerprint of an individual gene's activity. This allows us to accurately identify 1-1 orthologs and provide a ranked estimate of the likelihood that a gene has maintained its functional activity across species. We showed that we can use this ranking to accurately predict whether or not

a gene from one species will compensate for loss of its ortholog in another species. Additionally, we showed that marker genes that are conserved across species tend to have higher functional conservation scores than markers that are species-specific.

We used these 14 aggregate coexpression networks and this new metric for functional conservation to investigate functional conservation/divergence of paralogous genes resulting from duplication events. Duplication events are believed to contribute to neofunctionalization in plants, as well as being implicated in the origin of metazoan multicellularity. We found that a few highly conserved gene modules can reconstruct known phylogenetic relationships, that there is higher functional conservation among 1-1 orthologs, and that there are bimodal functional conservation patterns among 1-many orthologs, indicating potential sub- or neofunctionalization among these paralogs.

We also evaluated the value of single-cell coexpression networks, as opposed to bulk RNA-seq coexpression networks. Unfortunately, coexpression relationships from bulk samples are influenced by compositional effects created when multiple cell types exist in a single bulk sample. This can create artifactual results. scRNA-seq, on the other hand, allows sorting cells into cell types, thus completely removing any compositional artifactual signal. We compared network topology and coregulatory modules generated from two sources: a bulk RNA-seq coexpression network built from more than 2,000 mouse brain samples from 52 studies, and an aggregate scRNA-seq coexpression network built from the 500,000 cells/nuclei from the seven BICCN data sets. We find that despite compositional effects in the bulk data, the results are consistent between the two. Differential signals between broad cell classes persist in driving variation at finer levels, indicating that convergent regulatory processes affect cell phenotype at multiple scales. These results support future coexpression analysis for studying coregulatory modules, and also discourage unnecessary deconvolution of bulk expression data by cell type.

## Spatial Expression Data in the Brain

A fundamental goal of the NIH BRAIN Initiative is to integrate information from multiple modalities into a cohesive model. Spatial transcriptomics (ST) is an extremely new technology; it is not yet clear how

much agreement exists between it and scRNA-seq data. We addressed this open question by evaluating our ability to predict brain area of origin from a cell's expression profile alone. We obtained two single-cell transcriptomics data sets with spatial labels. The first is a whole-brain, adult mouse spatial gene expression data set generated with ST, an array-based transcriptome-wide mRNA assay that maintains the spatial origin of transcripts. The second is the Allen Institute's transcriptome-wide adult mouse in situ hybridization data (ABA ISH). We quantified how well we can learn canonical Allen Brain Atlas brain area labels using only gene expression for each data set using a Lasso classifier. We found the performance to be higher in the ABA ISH compared to ST. We also tested the models trained in one data set to test in the opposite data set. When trained on ST data, the prediction quality distribution looks similar when tested on either ST or ABA ISH data. However, models trained on ABA ISH data exhibit reduced performance when tested on ST data. We identify potential explanations for these differences in performance across brain areas to suggest that canonical brain area labels are meaningful in expression space and that patterns of spatial expression are not merely capturing physical distance in the brain. Emerging spatial gene expression data sets from the mouse brain will allow further characterization of the potential for cross–data set generalization of integrated scRNA-seq and ST data.

## Interpreting scATAC-seq Data

Single-cell assay for transposase accessible chromatin using sequencing (scATAC-seq) measures genome-wide chromatin accessibility for the discovery of cell-type-specific regulatory networks and new cell subpopulations. Current scATAC-seq data is extremely sparse and exhibits significant technical variation, making it challenging to interpret. Because the sequenced reads are only stochastically observed around cell-type-specific marker genes, cell-type identification by scATAC-seq is difficult, even when cells can be successfully clustered. We developed an approach to overcome the problem of scATAC-seq sparseness by using reference knowledge obtained from other scATAC-seq or scRNA-seq data sets for cell typing. By using marker genes combined with coexpression information, we are able to interpret scATAC-seq

data at a much higher resolution. Furthermore, we developed a novel pipeline for consensus scATAC-seq analysis (Catactor), to optimize the coexpression information for annotated scATAC-seq data using a supervised Lasso alternate algorithm. We expect Catactor to generalize to new unannotated data types, beyond scATAC-seq. To investigate the existence of coaccessible gene signatures in scATAC-seq data, we collected seven scATAC-seq data sets from mouse brain and cortex. In this collection, two data sets are based on joint profiling methods with scRNA-seq data for the exact same cells and one is from the reference atlas BICCN. The detected marker genes for each cluster from six well-annotated data sets consistently overlapped with our gene list, which contains marker genes and their coexpressed genes learned from multiple scRNA-seq data sets. In evaluating the prediction performances for major cell types using chromatin signals, we found the gene marker set showed substantially higher performance at both cluster and single-cell levels than those markers from individual studies. Our results demonstrate the value of coexpression and cross–data set comparison to capture subtle signals from scATAC-seq profiles and to improve alignment to known biology.

## Allele-Specific Expression

In mammalian females, during early development, one of the two X chromosomes is transcriptionally inactivated in each cell. The inactivated chromosome is chosen at random in each cell, resulting in a mosaic of cells expressing either the maternal or paternal X chromosome. The functional consequences of chromosome-level allelic variation within individuals remains poorly understood. Female mammals provide an ideal model to investigate cellular heterogeneity and differential expression because the allelic variation occurs without any confounded genetic background effects. Additionally, we have the opportunity to study escape genes, which are genes that are persistently expressed from the inactivated X chromosome, creating a female-specific source of transcriptional variation. We obtained scRNA-seq data from monocytes obtained from female carriers and male probands of X-linked chronic granulomatous disease (X-CGD), an X-linked immune disease. We used this data to begin designing a phasing pipeline to assign cells to one of

two X chromosome haplotypes, reconstructed based on the allele-specific coexpression.

## PUBLICATIONS

Ballouz S, Dobin A, Gillis JA. 2019. Is it time to change the reference genome? *Genome Biol* **20:** 159.

Chen X, Sun YC, Zhan H, Kebschull JM, Fischer S, Matho K, Huang ZJ, Gillis J, Zador AM. 2019. High-throughput mapping of long-range neuronal projection using in situ sequencing. *Cell* **179:** 772–786.e19.

Crow M, Gillis J. 2019. Single cell RNA-sequencing: replicability of cell types. *Curr Opin Neurobiol* **56:** 69–77.

Crow M, Lim N, Ballouz S, Pavlidis P, Gillis J. 2019. Predictability of human differential gene expression. *Proc Natl Acad Sci* **116:** 6491–6500.

*In Press*

Li S, Kendall J, Park S, Wang Z, Alexander J, Moffitt A, Ranade N, Danyko C, Gegenhuber B, Fischer S, et al. 2020. Copolymerization of single-cell nucleic acids into balls of acrylamide gel. *Genome Res* **30:** 49–61.

# GENOME ORGANIZATION, REGULATION, AND FUNCTIONAL ROLES OF NONCODING RNAs

T.R. Gingeras    C. Danyko    A. Dobin    J. Drenkow    G. Nechooshtan    D. Yunusov

## ENCODE Phase 3

C. Danyko, A. Dobin, J. Drenkow [in collaboration with C. Zaleski, A. Scavelli, former members who worked on Phase 3]

During phase 3 of the Encyclopedia of DNA Elements (ENCODE) project, a total of 4,834 human and 1,158 mouse experiments were performed to define and annotate diverse classes of functional elements in the human and mouse genomes. This scale of experimental data has provided new insights concerning questions of genome organization and inherent function as well as catalyzed new capabilities for deriving biological insights and principles, as detailed below and in recent publications (The ENCODE Project Consortium et al. 2020). In summary, during this phase of the ENCODE project we contributed directly and indirectly to the following set of conclusions:

- We defined core gene sets corresponding to major histopathological features using extensive new maps of RNA transcripts in a broad range of primary cell types (Breschi et al. 2020).
- We described an expansive new genomic compartment of DNA elements that encode recognition sites for RNA binding proteins, providing new insights into posttranscriptional regulation.
- We defined human genome-wide polymerase-III-transcribed Alu elements indicating the location of cell type–specific enhancer functions (Zhang et al. 2019).
- We deeply mapped the co-occupancy patterns of human transcription factors in reference cell types and connected these to key biological features of promoters and distal enhancers.
- We greatly increased the cell and tissue range, genomic resolution, and biological annotation of human DNase I hypersensitive sites and transcription factor footprints.

- We systematically defined cell-selective topological regulatory domains.
- We expanded the annotation of mouse chromatin modification, DNA accessibility, DNA methylation, and RNA transcription landscapes in early developmental stages not readily accessible in human.
- We systematically integrated DNA accessibility and chromatin modification data to create a categorized and expandable registry of candidate *cis*-regulatory elements in the human and mouse genomes.

## ENCODE Tissue Expression (EN-TEx) Project

C. Danyko, A. Dobin, J. Drenkow

The EN-TEx project is a subproject of the ENCODE project and has as its main goal to determine the genome sequences for four individuals (two male, two female) and to augment these DNA sequences (DNA-seq) with information concerning the transcriptional profiles (RNA-seq and RAMPAGE), DNase hypersensitive regions (DNase-seq), chromatin modification profiles (ChIP-seq), and 3D chromosomal interacting maps (HiCseq) of the study subjects. These studies were performed to generate and describe an individualized and precisely annotated human genome for each of the four individuals. In turn, these individualized data would be used to better understand the functional effects of the genetic variations (single nucleotide variation [SNV] and structural variation [SV]) found in each tissue type obtained from each donor.

## Genome Sequences

Multiple sequencing technologies and analysis software were employed to determine the DNA sequences for each genome. Short-read Illumina genome sequence data for each of the four donors were collected at a depth of >100× coverage. This was used to determine at high accuracy the single-nucleotide

sequence variations and short insertions and deletions (indels). Long-read sequencing was carried out for the purpose of detecting larger insertions and to provide biallelic information for each genome using PacBio (PB) and Oxford Nanopore Technologies (ONT) platforms. In addition, full chromosome length phasing of the genomes and the identification of chromosomal interactions were determined using HiC sequence analysis and 10X analyses. Unique and common single-nucleotide polymorphism (SNP) and structural variations for each genome were determined for each allele of each genome. An average of 3.9 million SNVs and 5.5 hundred thousand short SVs (<50 nt) were detected using Illumina sequencing and 23.7 thousand long SVs (10–50 kb) were detected by PB technology. The use of each sequencing technology revealed inherent biases. Only PB technologies could detect tandem expansion and contraction sequences, whereas the error rate of PB in the detection of SNV was 100-fold that of the Illumina technology.

## Reference versus Precision Mapping of the Transcriptional Landscape

Of the current ~19,000 and ~43,600 protein-coding and -noncoding genes, respectively, a total of ~15,000 and ~5,000 protein-coding and -noncoding genes were expressed (TPM 1–10) in at least one of the 80 tissues obtained from the four donors. Only a handful (25–50) of the expressed genes were seen to be expressed in neither the reference (hg38) nor the diploid precision genomes, indicating that there are a few errors in either genome sequence assembly. The remainder of the 15,000–20,000 expressed genes were highly correlated in their expression levels when mapped to either the reference and/or diploid genomes. A total of 10,000–80,000 more reads were mapped to the diploid genome. Thus, as expected, there were more multiple and fewer unique reads mapping to the diploid genome, indicating that the differences in sequence between the versions of the genome were distributed across the genome rather than being collected in large regions. Extreme differences in expression after mapping the RNA-Seq results to the reference versus precision versions of the genomes (e.g., no expression mapped in mapping to the reference genome compared to detected expression in the precision genome and vice versa) for each of 60,000 coding and noncoding genes

was observed for 728 genes. These results are likely to have been caused by errors in the either of the genome assemblies or in the mappings of the RNA-Seq data.

## Effects on Gene Expression When Structural Variations Overlap Functional Elements of the Genome

Considering the cases of complete or partial (at least 1-bp) overlap with any 22 classes of functional elements, negative selection is observed for all classes except for ENSEMBL CTCF binding sites in all four donors, thus indicating that functional elements are not often the sites of SV location. The expression levels of 87 genes and 75 promoters are affected significantly in two of the donors when these genes and promoters are covered completely by an SV. Affected genes of this group with TPM > 5 show no decrease in expression. Similarly, affected promoters with TPM > 5 all appear to not have their expression affected. Analyses of these deleted genes and promoters indicate that these involve heterozygous deletions, leaving one functional allele. Because these are biallelically expressed genes in genomes with no SVs, these results suggest that compensatory mechanisms are contributing to maintaining the levels of expression for the heterozygous genes. The effects of SVs on epigenetic and methylation marks are under analysis.

## MaizeCode: Construction of an Encyclopedia of DNA Elements of the Maize Genome

This work was performed in collaboration with the CSHL laboratories of W.R. McCombie, R. Martienssen, D. Jackson, D. Ware, M. Schatz, D. Micklos, and K. Birnbaum.

The overarching goal of the MaizeCode project is to produce empirical data sets for the identification of biochemically active and biologically functional elements encoded in the genome sequences of four maize lines (B73, NC350, W22, and Til11). A comprehensive catalog of these elements has been to construct a critical link between the genotypes of each of the lines with phenotype in this classical plant system. The project began with the genomic sequencing of the NC350 and Til11 lines. These data have been used to construct a plant/maize genomics database with associated metadata, workflows, and analysis

pipelines for the project and community. These are being disseminated using CyVerse, an NSF-funded platform and therefore accessible to all researchers to use in an unrestricted fashion, to process their data through and/or to reproduce our results and update them as desired.

1.  One hundred percent of RNA-Seq experiments (>200 nt, <200 nt, RAMPAGE) for five tissues present in three of MaizeCode Maize lines and four tissues in one of the lines (Til11) were completed.

2.  All tissues were analyzed by ChIP-seq to study four epigenetic elements present in five tissues obtained from the NC350, B73, and W22 lines and four tissues from the Til11 line.

3.  A first preliminary assembly of Til11 using Oxford Nanopore long-read sequencing has been made and we continue to finalize the NC350 assembly to ensure the highest quality and contiguity.

4.  Using newly developed means of cell wall digestion to generated protoplasts, coupled with efficient cell capture using combinations of the fluorescent dyes, a transcriptome map of the maize seminal root has been generated for six cell types and tissues. Tissue sample "slices" along the maturation gradient of the root have been taken to profile developmental stages. Single-cell profiles of the root were generated and markers form the six-tissue reference and developmental slices were used to reconstruct the maize root meristem in high resolution.

5.  Different tissue types in the targeted maize strains (B73, W22, NC350, and Til11) have been collected for distribution to the project and external collaborators as desired for additional assays.

6.  Using tagged strains provided by NSF-funded Maize Cell Genomics project (http://maize.jcvi.org/cellgenomics/index.php) allowed us to collect nine tagged transcription factor (TF) lines in the B73 background to study the chromatin profiles of each TF. For one line, ChIP-Seq experiments have been completed; for others, the tissues have been collected and ChIP experiments are in progress.

7.  Using RFP-tagged tissue-specific promoter lines from the NSF-funded Maize Cell Genomics project and protoplast generation techniques developed by the Birnbaum Laboratory, FACS-sorted protoplasts for four targeted specific shoot cell types have been analyzed by RNA-Seq to obtain the transcriptional profile of each of these cell types.

8.  RNA-Seq and RAMPAGE workflow/pipelines, processed the MaizeCode data have been built and the computational and experimental metadata have been integrated on SciApps.org with the newly developed SciApps RESTful API. We hosted two Annotation Jamborees on the latest B73 RefGen_ v4 assembly.

9.  We designed the MaizeCode project website (http://www.maizecode.org/), administered a needs assessment survey at Maize Genetics Conference, and have begun redeveloping the DNA Subway Red Line for the community annotation of the four maize genomes.

10. Two Maize Annotation Jamborees were held to involve primarily undergraduate institution (PUI) faculty and plant researchers in community curation of maize gene models on the latest B73 *Zea mays* v4 assembly.

The v1.2 draft NC350 genome is available on the CyVerse and MaizeCode portals: http://www.maizecode.org/data/ and http://datacommons.cyverse.org/browse/iplant/home/shared/maizecode/NC350/assembly/hybrid/Zm-NC350-MAIZECODE-1.2.fasta.

## Sorting and Processing of Various Types of RNAs in the Extracellular Environment

G. Nechooshtan, D. Yunusov, J. Drenkow

RNA is not only found within the boundaries of cells. We had turned our attention to the identification of a collection of RNAs that undergo site-specific processing in the extracellular milieu. Examples of these RNAs include tRNAs and Y RNAs. One of our first goals of this project is to identify and to isolate the factors involved in this processing activity. Following proteomic and biochemical work, we have identified RNase 1 as one agent responsible for this processing (Nechooshtan et al. 2020).

Several studies in recent years showed that tRNA halves and distinct Y RNA fragments are abundant in the extracellular space, including in biofluids.

Although their regulatory and diagnostic potential has gained a substantial amount of attention, the biogenesis of these extracellular RNA fragments remains largely unexplored. Over the past year, we have demonstrated that these fragments are produced by RNase 1, a highly active secreted nuclease. We used RNA sequencing to investigate the effect of a null mutation of RNase 1 on the levels of tRNA halves and Y RNA fragments in the extracellular environment of cultured human cells. We complemented and extended our RNA sequencing results with northern hybridization studies, showing that tRNAs and Y RNAs in the nonvesicular extracellular compartment are released from cells as full-length precursors and subsequently cleaved to distinct fragments. In support of these results, formation of tRNA halves is recapitulated by recombinant human RNase 1 in our in vitro assay. These findings assign a novel function for RNase 1 and position it as a strong candidate for generation of tRNA halves and Y RNA fragments in biofluids.

## PUBLICATIONS

Rahmanian S, Murad R, Breschi A, Zeng W, Mackiewicz M, Williams B, Davis CA, et al. 2019. Dynamics of microRNA expression during mouse prenatal development. *Genome Res* **29:** 1900–1909.

Zhang X-O, Gingeras TR, Weng Z. 2019. Genome-wide analysis of polymerase-III transcribed Alu-elements suggests cell-type specific enhancer function. *Genome Res* **29:** 1402–1414.

*In Press*

Breschi A, Munoz-Aguirre M, Wucher V, Davis CA, Garrido-Martin D, Djebali S, Gillis J, Pervouchine D, Vlasova A, Dobin A, et al. 2020. A limited set of transcriptional programs define major histological types and provide the molecular basis for a cellular taxonomy of the human body. *Genome Res* **30:** 1047–1059.

Nechooshtan G, Yunusov D, Chang K, Gingeras TR. 2020. Processing by RNase 1 forms tRNA halves and distinct Y RNA fragments in the extracellular environment. *Nucleic Acids Res* **48:** 8035–8049.

The ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, et al. 2020. Expanded encyclopedias of DNA elements in the human and mouse genomes. *Nature* **583:** 699–710.

Wang L, Lu Z, delaBastide M, Van Buren P, Wang X, Ghiban C, Regulski M, Drenkow J, Xu X, Ortiz-Ramirez C, et al. 2020. Management, analyses, and distribution of the MaizeCODE data on the cloud. *Front Plant Sci* **11:** 289.

# LEVERAGING LONG-READ SEQUENCING TECHNOLOGIES TO FACILITATE GENOMIC DISCOVERY

**W.R. McCombie**    M. Anzaldi    S. Iyer    S. Mavruk Eskipehlivan    A. Qui
                     E. Ghiban    M. Kramer    O. Mendivil Ramos    R. Wappel
                     S. Goodwin    J. Lihm    S. Muller

We continue to make progress in understanding both plant and animal genomes with long-read sequencing as well as in developing better ways to use this capability in specific instances, such as with cancer. In addition, we continue to provide and enhance our ability to support other research efforts at Cold Spring Harbor with advanced sequencing capabilities that facilitate their completion. For the purpose of this overview we will divide our efforts into animal genomics (mainly cancer genomics), plant genomics, technology development, and genomics support projects.

In animal genomics we have had three main projects in the past year, all cancer-related. These have been continuing our analysis of breast cancer organoids for structural variants, long-read sequencing of patient pedigrees that have a high burden of cancer but no detectable genetic variations likely to cause cancer based on short-read sequencing (in cooperation with Memorial Sloan Kettering), and completing the sequencing of two bat genomes to study from a comparative genomics standpoint. In plant genomics we have carried out long-read sequencing on several very large and unusual plant genomes in collaboration with the New York Plant Genome Consortium, of which we are a member. We have also, in collaboration with the Lippman laboratory at CSHL and others, sequenced a large number of tomato genomes with long-read sequencing. In technology development we have made substantial progress in using CRISPR-based targeted sequencing with very long-read sequences. In doing this we have been able to capture genomic targets from human cell lines and sequence them with individual reads up to 100 kilobases long. In our genomics support capacity, we continue to provide help with validating organoids for the Leidos project and with other projects as well. A number of these projects have manuscripts in preparation or submitted. Details of these projects are described below.

## CANCER GENETICS

### Application of Long-Read Technologies to Probe the Genomic Architecture of Tumor Organoids

S. Goodwin, M. Kramer, R. Wappel, W.R. McCombie
[in collaboration with G. Arun and D. Spector, CSHL;
K. Kostroff, Northwell Health; M. Schatz, R. Sherman,
I. Lee, and W. Timp, Johns Hopkins University;
F. Sedlazeck, Baylor College of Medicine]

Organoids are thought to effectively recapitulate tumor characteristics in a controlled, reproducible manner, making them essential tools for modeling cancer and personalized cancer treatment. Tumor cells are separated from the surrounding normal tissue and grown into homogeneous cell cultures that can be used for an array of assays that would not be possible from primary tumor tissue. In 2018 and into 2019 we completed high-depth sequencing of two breast tumor–derived organoids and the SKBR3 breast cancer model cell line. In 2018 we completed the PacBio sequencing of SKBR3 and published those results in *Genome Research* (Nattestad et al., *Genome Res* **28:** 1126 [2018]). This data will be deposited into dbGap in 2020. Sequencing was carried out via PacBio, Oxford Nanopore, and 10X Genomics methods. Over the course of this study, we achieved more than 50× coverage of the SKBR3 and tumor organoid genomes on both PacBio and Oxford Nanopore Technologies (ONT). The N50 of each flow cell was between 10 and 18 kb. Our collaborators furthered the project by carrying out in-depth analysis of the data. When comparing the number of structural variations (SVs) called via the different platforms and different SV calling tools, we found that there is high concordance between SVs called using ONT and PB technology (13,891 in SK-BR-3) (Fig. 1). There are many more SVs called via long-read technology than short-read technologies, even when employing 10X Genomics technology (Fig. 1).
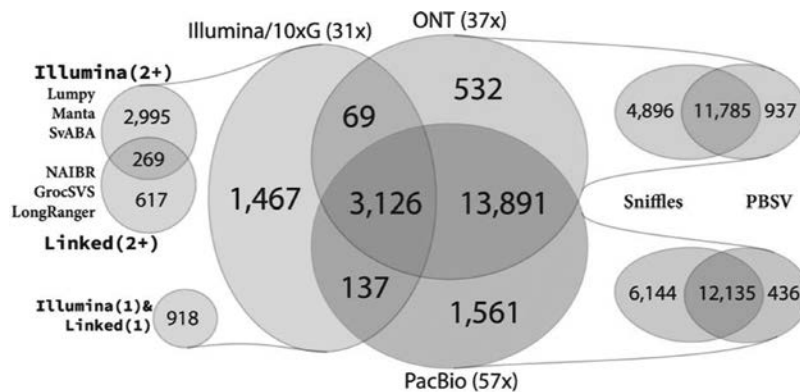
**Figure 1.** Comparison of the number of structural variations (SVs) in SKBR3 identified between Illumina, PacBio, and ONT via various SV calling tools.

In addition to the exploration of SV identification, we explored calling cytosine methylation directly from the ONT data. We found global hypomethylation in all cancer samples (Fig. 2A), as has been observed in previous studies. The notable exception to this trend is found in the promoter region of many genes (Fig. 2C). The hypomethylated state data was compared to the SV data. We found a modest correlation between hypomethylated promoters and SV incidence (Fig. 2E). This work was submitted to bioRxiv in 2019 and will be submitted to *Genome Research* in 2020.

## Quality Assurance of Organoid Cancer Models

W.R. McCombie, S. Goodwin, M. Kramer, S. Muller, E. Ghiban [in collaboration with D. Tuveson, D. Spector, P. Sridevi, D. Plenker, G. Arun, and S. Bhatia, CSHL; V. Corbo, University of Verona]

In 2019 we continued our collaboration to characterize and validate organoid models of several cancers for the Leidos/National Cancer Institute (NCI) initiative to enable individualized drug therapy and improve cancer treatment response. We performed targeted
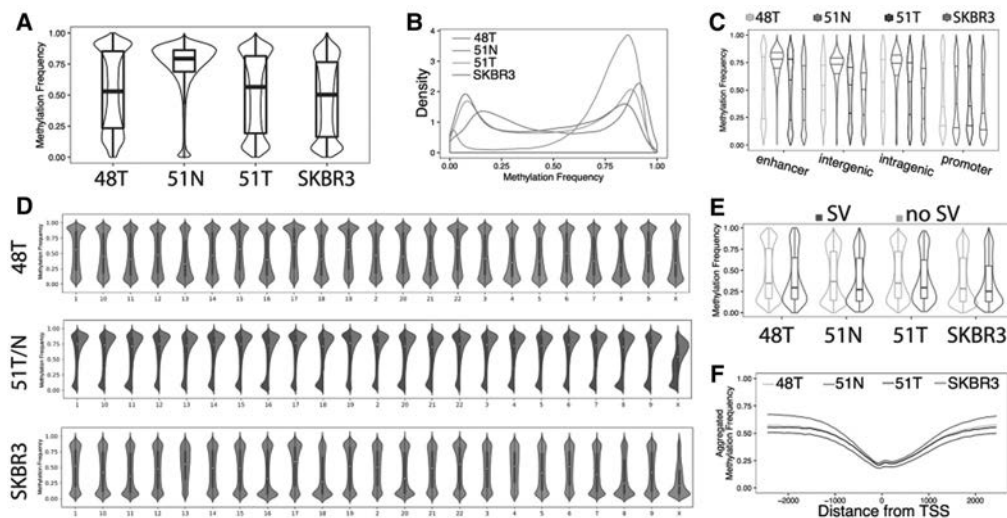


**Figure 2.** Methylation analysis of nanopore data for SKBR3 and organoids. (*A*,*B*) Distribution of methylation frequencies and frequency density functions. (*C*) Methylation frequency distributions over different genomic regions. (*D*) Per-chromosomal methylation frequency distributions. (*E*) Methylation frequency distributions of promoter regions with and without SVs in the gene body. (*F*) Averaged aggregation methylation frequencies around transcription start sites (TSS).

capture and sequencing of 140 known cancer genes in 244 samples (including organoid models and matched normal or tumor tissue) for phase II of this project. Analysis of the models was performed using our previously described variant annotation pipeline to verify driver mutations curated from ClinVar, TCGA, and COSMIC in the models and ensure that they are tumor-derived. This year, 59 positive models were selected for submission to Leidos, and additional models are pending further review. Including our previous models, this brings the total number of confirmed models to 163—a valuable resource for colorectal, breast, pancreatic, head and neck, endometrial, and other cancers. A new gene panel was also designed with NimbleGen, including coverage of six additional driver genes and additional probes to target copy number variation (CNV) regions. Testing of this panel is under way and will be used to enhance detection of drivers in more diverse samples.

## Long-Read Sequencing of Early-Onset Cancer Pedigrees

W.R. McCombie, S. Goodwin, M. Kramer, R. Wappel, S. Muller [in collaboration with Z. Stadler and Z. Patel, MSKCC]

Despite advances in the determination and detection of the genetic factors driving cancer, a subset of cases remains unresolved after screening with standard tests. In collaboration with Zsofia Stadler, we have sequenced the genomes of 10 individuals from three families with early-onset testicular or colorectal cancer whose previous screening did not uncover explicit driver mutations. We employed Oxford Nanopore long-read sequencing of high-molecular-weight DNA derived from blood to uncover structural variants that may contribute to cancer progression. Long-read sequencing has been shown to be superior to typical short-read next-generation sequencing (NGS) applications for discovery of genomic alterations >50 bp. The inclusion of family members enables us to filter the large number of individual variants by family structure and affected status to prioritize potential drivers. We sequenced four family members (a quad) including two unaffected parents and two affected brothers with early-onset testicular cancer, one of whom had bilateral tumors. We also sequenced a second family of three male cousins with early-onset testicular cancer. Finally, we sequenced three family

members (a trio) including two unaffected parents and their son with early-onset colorectal cancer.

Our ONT sequencing produced N50 read lengths of ~18–30 kb, ensuring ample coverage of long reads to span genomic repeats and rearrangements. The long reads were aligned to the human reference (UCSC hg38) with NGMLR (Sedlazeck et al., *Nat Methods* **15:** 461 [2018]), which uses a gap penalty model adapted to allow long reads to span true variant events while tolerating sequencing errors. Sample coverage ranged from 20- to 35-fold aligned depth. Structural variants were called using Sniffles (Sedlazeck et al., *Nat Methods* **15:** 461 [2018]), requiring a minimum of eight reads to support the alternate allele. Approximately 21,000–28,000 total SVs were called per genome, with deletions and insertions being the dominant type of variation. Variants were subjected to group genotyping using SURVIVOR (Sedlazeck et al., *Nat Methods* **15:** 461 [2018]) in order to avoid missing the presence of variants in family members that fell below the imposed coverage threshold. SVs were filtered according to family structure to uncover events that differed between healthy and affected members. Initial analysis revealed several variants of interest including intronic deletions in cancer-related genes such as *PTEN* and *ACVR1*, and an intronic insertion in *ALK*. Continuing work includes stratification of variants by population frequency and overlap with regulatory regions to predict functional impact.

In 2020 we plan to further refine the analysis of the SVs, include analysis of single-nucleotide polymorphisms (SNPs) and small indels, and perform methylation analysis using the raw nanopore signal data. Additional early-onset pedigrees may also be sequenced, or additional samples may be selected to enhance the current pedigrees. Integration of these analyses could provide a comprehensive picture of the disparate genomic alterations that may foster aggressive cancer development.

## Long-Read Sequencing of Bats as a Model for Cancer Resistance Mechanisms

O. Mendivil Ramos, M. Kramer, S. Goodwin, R. Wappel, W.R. McCombie [in collaboration with N. Simmons and S. Oppenheim, AMNH; M. Schatz, Johns Hopkins University]

Long-lived mammals, in particular the bat clade, offer an excellent point of comparison for studies of cancer resistance and genome stability given their smaller size and long life span. In bats, apart from

telomere-dependent tumor suppressor mechanisms, genome maintenance mechanisms, and effective immune response, additional lineage-specific tumor suppressor mechanisms remain unknown.

In collaboration with groups at the AMNH and JHU, we have sequenced two new bat species, *Artibeus jamaicensis* and *Pterotus mesoamericanus*, which are above the median age in years of all species of bats noted so far and have noncarnivorous diets in comparison with other bats. We used ONT long-read sequencing along with short-read sequencing to provide long-read error correction. We have accomplished a total coverage of 131× ONT, with 38× of 30-kb fragments, and ~30× Illumina for *A. jamaicensis* and a total of 156× ONT, with 37× of 30-kb fragments, and ~30× Illumina for *P. mesoamericanus*. We deployed the wtbg2 (Ruan and Li, *Nat Methods* **17**: 155 [2020]) assembly pipeline for the long reads as first-pass assembly by the end of 2019 for each bat, which we will continue to polish and refine in early 2020.

Additionally in 2020, with the Siepel lab at CSHL we will be delving into a computational analysis investigating large-scale structural variation (deletions, duplications, inversions, and translocations over 100 kb) and positive gene selection in a panel of bats and more cancer-prone mammals and mice. We are anticipating the detection of gene families either positively selected or duplicated that are linked to immunity and DNA damage repair mechanisms in cancer resistance. This may explain their exceptional longevity and cancer resistance.

## Targeted Capture for Long-Read Sequencing [ACME]

S. Iyer, M. Kramer, S. Goodwin, W.R. McCombie

Targeted sequencing dramatically improves our ability to study the genome by providing the depth necessary and accuracy necessary to explore specific targets of interest and detect rare alleles in a heterogenous population of cells. Current enrichment strategies often involve fragmentation of genomic DNA prior to amplification, resulting in short (<1,000 bp) templates, and cannot be used to study large, variable regions of the genome. To address this gap, we looked into CRISPR-Cas-based strategies, specifically the ONT CRISPR-Cas9-mediated polymerase chain reaction (PCR)-free enrichment method. Briefly, this approach works by dephosphorylating gDNA fragment ends, followed by using crRNA guides to target regions upstream and downstream of the region of interest (ROI) for Cas9 cleavage, freeing the target DNA and making its ends available for adapter ligation.

Although the ONT Cas9 targeting approach is fast and effective, we identified two major limitations.

1. Relatively shorter target regions (<30 kb) covered by single reads. Single reads covering whole target regions minimize mapping errors due to SVs. So, the shorter the target sizes captured in a single read, the greater the number of crRNA guides required to tile through larger target regions.

2. The lack of a background reduction step, which results in non–target DNA fragments competing with targets for sequencing, effectively reducing target depth. To address these gaps, we designed a size titration panel consisting of 10 prominent cancer genes that belonged to different size ranges—10 kb, 20 kb, 40 kb, 80 kb, and 150 kb—to identify the largest possible target that could be effectively captured and spanned with single reads using the ONT Cas9 approach. We also made modifications to the existing ONT Cas9 protocol to facilitate work with longer targets by reducing background fragments.

Most significantly, we developed ACME, an affinity-based Cas9-mediated enrichment method that uses His-Tag Isolation and Pulldown to remove background DNA. The ONT Cas9 method uses a HiFi Cas9 nuclease that contains a 6-histidine tag at its carboxyl terminal. After the Cas9 cleavage and dA tailing step in protocol, the Cas9 enzyme remains bound to the PAM-distal end (i.e., non–target DNA side), protecting it from subsequent sequencing adapter ligation. At this step, we introduced His-Tag-specific magnetic beads and were able to pull down Cas9-bound non-target fragments from the sample, allowing more target DNA to make it onto the flow cells. On using ACME with our cancer gene panel in two breast cell lines—MCF 10A and SK-BR-3—we were able to capture targets as large as 100 kb in single reads, pushing existing size limits by at least twofold. Looking specifically at the 91-kb target *BRCA2* (Fig. 3), we saw a 70- to 100-fold enrichment across both cell lines, giving us a coverage of 100 × –115× of the gene. The ability to capture this gene in its entirety gives us the opportunity to look for variants in this gene that could have been missed by standard short-read-based gene panels.
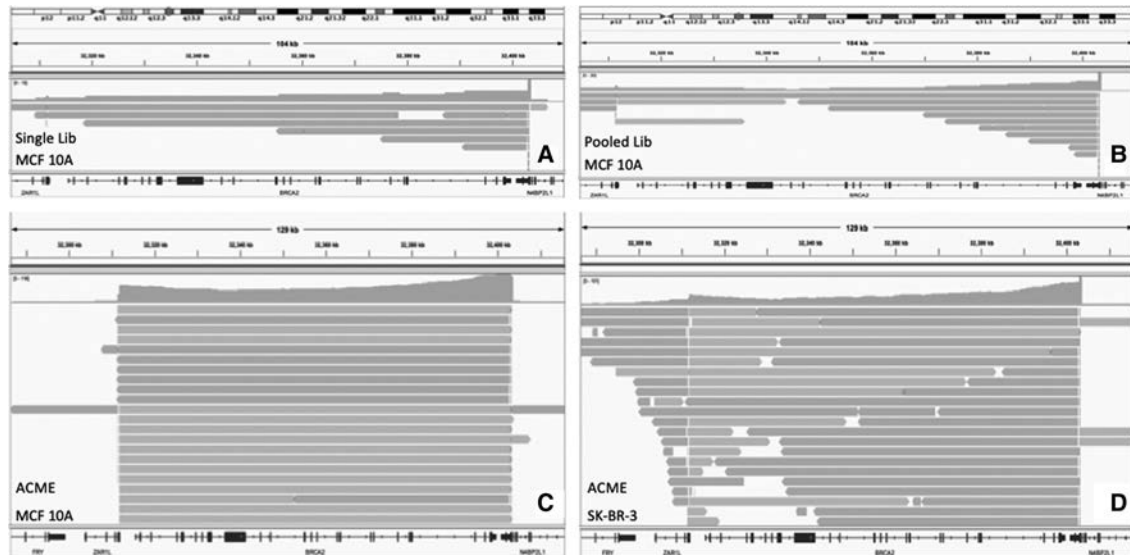
**Figure 3.** Reads that mapped to the *BRCA2* gene in different Cas9-mediated targeting libraries prepared from MCF 10A DNA. Plus strand reads are shown in pink and minus strand reads in blue. (*A*) Single sample library prep. (*B*) Library prepared by pooling together three identical library preps of MCF 10A DNA. (*C,D*) Pooled library prepared from four identical preps of MCF 10A and SK-BR-3 DNA, respectively, with the incorporation of the ACME step to remove non–target DNA. Note that *BRCA2* was part of a panel of genes chosen for targeting in each sample.

In 2020, we will expand our panel to include 35 genes common between the Invitae diagnostic panels for breast, colorectal, and pancreatic cancer. We will also look into further enhancements to increase background depletion and target enrichment using ACME with the ReadUntil-based algorithm UNCALLED (from the Schatz group) and explore multiplexing options to bring down per sample DNA requirements, which will be pertinent when we advance to patient tissues.

## PLANT GENOME ANALYSIS

### Sequencing and Assembly of Tomato Genomes with Oxford Nanopore Technology

S. Goodwin, R. Wappel, W.R. McCombie [in collaboration with S. Soyk, X. Wang, Z. Lemmon, and Z. Lippman, CSHL; F. Sedlazeck, Baylor College of Medicine; M. Alonge and M. Schatz, Johns Hopkins University]

SVs are important for crop improvement and for developing agriculturally important traits. However, resolving these SVs, particularly in a wide and comprehensive manner, has been challenging. We used long-read nanopore sequencing to sequence 100 different tomato genomes to at least 40× coverage. Although this study initially began with the Oxford Nanopore GridION instrument, we moved to the PromethION

as it became available. Over the course of the study we improved PromethION performance, generated ~70 Gb per flow cell, up to 140 Gb. We also improved the N50 of each flow cell from ~15 kb to >30 kb. We determined that there is an inverse relationship between N50 read lengths and yield: Those cells with N50s in the 30-kb range tended to yield less data than those in the 15-kb range, although substantial variability in performance is still a factor. Our collaborators furthered this work by carrying out assemblies and an in-depth analysis of the pan-genome. To facilitate this work, a reference-guided contig ordering and orienting tool dubbed RaGOO was developed (Alonge et al. 2019). Using this tool, 238,490 SVs were identified between the genomes and generated 14 new reference assemblies. The results of this work will be submitted to *Cell* in 2020.

### Exploring the Genomes of Living Fossils

O. Mendivil Ramos, S. Goodwin, M. Kramer, R. Wappel S. Muller, W.R. McCombie [in collaboration with R. Martienssen and C. Alves, CSHL; G Coruzzi, G. Eshel, and V. Sondervan, NYU; D Stevinson, S. Frangos, D. Little, and S. Wilson, NYBG; M. Schatz, JHU; K. Varala, Purdue; S. Kolokotronis, SUNY Downstate School of Public Health]

As part of the NY Plant Genome Consortia, we are sequencing five gymnosperm genomes, which include

two "living fossil" species (i.e., species that have survived an extensive period of drastic climate change without speciation). We are constructing a comprehensive and complementary comparative genomic analysis of four extant lineages of gymnosperms. The analysis is split on each one of the four extant lineages represented by a pair composed of one "living fossil" and its radiated phylogenetic gymnosperm evolutionary sister.

We have deployed ONT long-read sequencing of high-molecular-weight DNA from leaf tissue to elucidate the genomic structure of these very large and complex plant genomes. We have optimized a cetyl trimethylammonium bromide (CTAB) extraction strategy together with needle shearing to produce a high proportion of so-called ultralong fragments (>30 kb) to extend assembly contiguity and overcome complex repeats. We increased accuracy of the long-read sequencing by polishing with the Illumina short sequencing once the assembly is produced.

We have completed Nanopore sequencing of one living fossil genome (*Wollemia nobilis*) with a total coverage of 66×, including some long sequencing reads of >100 kb, with a read length N50 of 35 kb. The second living fossil, *Metasequoia glyptostroboides*, has a 110× total coverage with a read length N50 of 24 kb. Additionally, we are in the process of completing the sequencing of *Araucaria angustifolia*—which achieved a total of ~20× coverage in 2019, with 10× coverage in reads >30 kb. We are also starting to sequence the *Juniperus communis* and *Gnetum gnemom* genomes.

We have started the first draft assembly of the *Wollemia* genome, using the wtbg2 assembly pipeline of the long reads, followed by polishing with the Illumina data. We have performed preliminary analyses of genome size and genome heterozygosity using Genomescope (http://qb.cshl.edu/genomescope/) to anticipate difficulties and paths to optimize the assembly pipeline strategy. Also, transcriptome data has been generated to improve the gene annotation of the ongoing genome assembly. Small RNA sequencing is also under way and will be used along with the methylation data to assess the activity and maintenance of transposable elements in this genome. A parallel effort is ongoing to update the phylogenomic pipeline called PhyloGenious (http://nypg.bio.nyu.edu) based on the publicly available data of 100 RNA-Seq data sets from 20 previously sequenced gymnosperms to identify orthologs and study gene family evolution events among different gymsperms lineages. All of these analyses will be applied and adapted for the remainder of the project. Together with the consortium, we are drafting a manuscript outline to highlight insights gained from this genome sequence of *W. nobillis* in 2020.

## Toward an Encyclopedia of Maize Genomics with the MaizeCODE Project

W.R. McCombie, S. Goodwin, M. Kramer, E. Ghiban
[in collaboration with T. Gingeras, C. Danyko, D. Jackson, R. Martienssen, M. Regulski, D. Micklos, M. Schatz, and D. Ware, CSHL; Ken Birnbaum, NYU; Hank Bass, Florida State]

We previously reported on our collaboration to elucidate the complex genetic architecture of the maize genome using genomic, transcriptomic, and epigenetic assays. In 2019, we continued expression work with the Gingeras group to sequence long, short, and RAMPAGE RNA libraries for multiple tissues of Til11. We also sequenced transcription factor ChIP-seq libraries of transgenic maize lines from the Jackson laboratory. We added additional sequencing coverage to our previous Til11 ONT long-read genome work in order to increase coverage of very long fragments. To further improve assembly quality, we performed PacBio HiFi sequencing of Til11 using ~12-kb fragments. We achieved ~22× genome coverage of HiFi reads with >99% accuracy. Our collaborators in the Schatz lab created a de novo assembly of the Til11 HiFi data using Canu (Koren et al., *Genome Res* **27:** 722 [2017]). The high-quality PacBio data allowed omission of the slow and computationally intensive error correction step. The resulting Til11 assembly was comparable to the previously published B73 genome, with a 2.2-Gb total size, an N50 contig length of >6 Mb, and a maximum contig length of ~30 Mb. This high-quality assembly allowed for meaningful comparison between Til11 and maize varieties B73, NC350, and W22, particularly for structural variants. This data has been submitted to the National Center for Biotechnology Information (NCBI) SRA under the MaizeCODE umbrella project (https://www.ncbi.nlm.nih.gov/bioproject/380952). A manuscript detailing methods to assist the community with access to and analysis of all of the publicly available MaizeCODE data sets via CyVerse has been prepared and will be submitted for publication early next year.

## PUBLICATIONS

Aganezov S, Goodwin S, Sherman R, Sedlazeck FJ, Arun G, Bhatia S, Lee I, Kirsche M, Wappel R, Kramer M, et al. 2019. Comprehensive analysis of structural variants in breast cancer genomes using single molecule sequencing. bioRxiv doi:10.1101/847855

Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol* **20:** 224.

Goodwin S, McCombie WR. 2019. Sequencing complex genomes with PromethION technology in a core setting. *J Biomol Tech* **30:** S36.

McCombie WR, McPherson JD. 2019. Future promises and concerns of ubiquitous next-generation sequencing. *Cold Spring Harb Perspect Med* **9:** a025783.

McCombie WR, McPherson JD, Mardis ER. 2019. Next-generation sequencing technologies. *Cold Spring Harb Perspect Med* **9:** a036798.

# PLANT GENOME ORGANIZATION, EVOLUTION, AND FUNCTION

| D. Ware | K. Chougule | Y.K. Lee | M.K. Tello-Ruiz |
| | A. Diniz | F. Li | P. Van Buren |
| | N. Gladman | Z. Lu | B. Wang |
| | F. Hu | I. Meijs | L. Wang |
| | Y. Jiao | A. Olson | X. Wang |
| | V. Kumar | M. Regulski | S. Wei |
| | S. Kumari | J. Stein | L. Zhang |

The Ware laboratory has two primary goals: (1) understanding plant genome function in agriculturally important crop plants; and (2) development of tools, data sources, and resources for the genomics research community.

## PLANT GENOME RESEARCH

In the last decade, the sequencing and annotation of complete plant genomes has helped us understand plant function and evolution, as well as how to alter economically important traits. Efforts in many disparate disciplines are required to generate reference genomes. The work at the Ware laboratory often starts with wet laboratory scientists who generate the raw sequence data. Next, computational biologists and bioinformaticians kick off a series of computational steps to interpret the raw data. The process of interpretation involves the assembly of raw sequence reads into overlapping segments ("contigs"), which are combined to create a scaffold. This scaffold, in turn, discerns the position, relative order, and orientation of contigs within the chromosomes. The next steps are annotation, the discovery and description of genes and other functional elements, and homologies (evolutionary relationships) with other genomes. This information must be faithfully communicated and visualized in web-based platforms such as Gramene.

All of these activities are rapidly evolving in response to fast-paced improvements in sequencing technologies, algorithms, and data-handling requirements. For example, high-depth and low-cost sequencing of RNA transcripts is providing a vast stream of new evidence that informs genome annotation; this, in turn, has spurred the development of new software for modeling and performing genome annotation. Low-cost sequencing has also made it possible to ask whole new classes of questions, moving beyond the generation of single references for individual species and supporting the development of multispecies representation as a "pan-genome." Ongoing projects within the maize, rice, *Arabidopsis*, sorghum, and grape research communities are now sequencing hundreds or thousands of genotypic backgrounds, chosen from carefully constructed populations, wild populations, and breeding germplasms in each species. Information about genetic variation is helping scientists to understand the genetic basis of phenotypic traits and the origins of domesticated crop species. New technologies are also driving research in epigenetics, the study of heritable variation not attributable to changes in the underlying genome sequence. Epigenetic mechanisms include modification of DNA by methylation and various forms of histone modification, which can cause changes in gene expression and other phenotypes. Both types of modifications can be studied using new sequencing technologies and analytical methods.

### Developing Reference Assemblies for Maize and Sorghum

K. Chougule, B. Wang, S. Wei, A. Olson, Z. Lu
[in collaboration with K. Fengler and V. Llaca, Corteva Agriscience; PIs: C. Hirsch, M. Hufford, K. Dawe, D. Ware]

Maize and sorghum are important feed and fuel crops, as well as model systems in developmental genetics. Complete and accurate reference genomes are imperative for sustained progress in understanding the genetic basis of trait variation and crop improvement. The Ware laboratory has played a leading role in the development and stewardship of the grass genomes such as rice, sorghum, and maize for more than a decade.
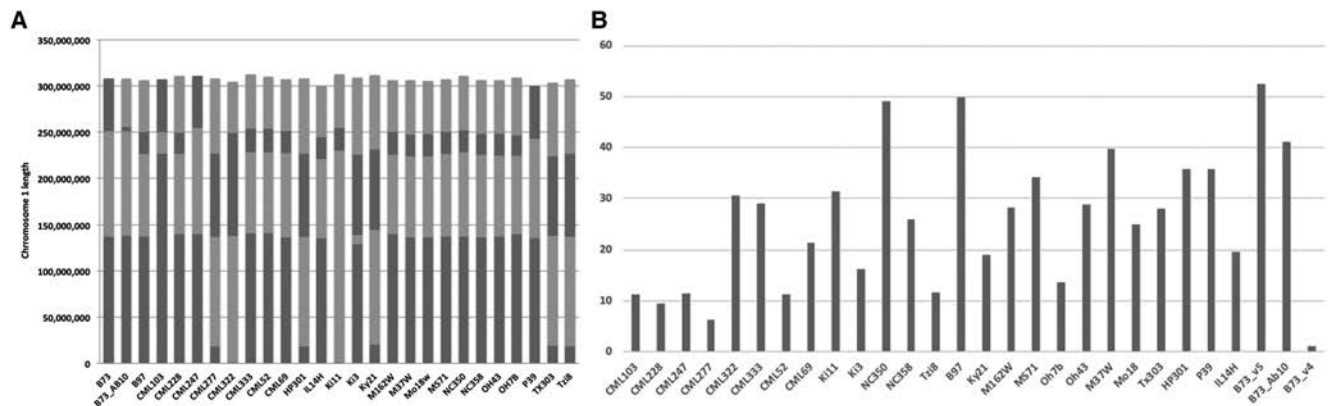
**Figure 1.** Maize NAM founder lines. (*A*) Maize Chromosome1 scaffolds distribution. (*B*) Maize contig N50 distribution.

The 2009 release of the B73 reference sequence was a milestone in plant genomics research because of the unprecedented size and complexity of the maize genome. Through several updates, this foundational resource has remained the principal genome reference for the maize research community. Yet, it continues to be a work in progress, with gaps and misassemblies that have defied available sequencing technologies, especially over the highly repetitive regions that are also the most dynamic and rapidly evolving. Maize is both an important crop and a genetic model system, with high levels of genetic and functional diversity. Gene content can vary by >5% across lines, whereas up to half of the functional genetic information lies outside of coding regions in highly variable and repetitive intergenic space. Characterization of this diversity has been confounded by the reliance on a single reference genome. Therefore, complete and accurate reference genomes from multiple individuals will be needed to characterize the maize pan-genome and to advance the genetic and functional study of this crop. The year 2019 marked a new initiative to generate reference genomes for a core set of 26 highly researched inbreds known as the NAM (Nested Association Mapping) founder lines, which represent a broad cross section of modern maize diversity. Building off of previous success with the B73_v4 reference genome, we devised a strategy for complete and accurate assembly of maize genomes using PacBio SMRT sequencing along with high-resolution BioNano optical maps. Assemblies utilized 60-80X PacBio data; they were polished

with Illumina data and scaffolded using BioNano DLS optical maps. Assembly metrics (N50, number of scaffolds, and assembly size) are substantially improved relative to existing maize references. The contig N50 for the B73_v5 reference is 52Mb, ~50× greater than the previous B73_v4 reference (see Fig. 1).

*Sorghum bicolor*, one of the most important grass crops worldwide, has been known to harbor high genetic diversity. With the available PacBio long-read sequencing and BioNano maps, we constructed a chromosome-level genome assembly for two important sorghum inbred lines, TX2783 and TX436, to discover the large structural variations in sorghum. The final assembly of TX2783 consists of 19 scaffolds with a contig N50 of 25.6 Mb, and the final assembly of TX436 consists of 18 scaffolds with a contig N50 of 20.3 Mb. We identified extensive large structural variations in sorghum genomes BTX623, TX430, TX436, and Rio using TX2783 as a reference. Genome-wide scan of the disease resistance genes (R gene) showed high-level diversity in these five available sorghum genomes.

Although these new technologies have greatly lowered the cost and improved the quality of reference genome resources, challenges remain in annotating these genomes. To this end, a major effort this year has been improving the workflows to support annotating the TEs (transposable elements) and genes, with a target of releasing the annotations in the next year. Current assemblies traverse TEs and provide an opportunity for comprehensive annotation of TEs. Numerous methods exist for annotation of each class of TEs, but their

relative performances have not been systematically compared. Moreover, a comprehensive pipeline is needed to produce a nonredundant library of TEs for species lacking this resource to generate whole-genome TE annotations. Using the most robust programs, we created a comprehensive pipeline called Extensive de-novo TE Annotator (EDTA) that produces a filtered nonredundant TE library for annotation of structurally intact and fragmented elements. EDTA also deconvolutes nested TE insertions frequently found in highly repetitive genomic regions. Using other model species with curated TE libraries (maize and *Drosophila*), EDTA is shown to be robust across both plant and animal species (Ou et al. 2019).

## Variant Phasing and Haplotypic Expression from Long-Read Sequencing

B. Wang, M. Regulski, Y. Jiao, L. Wang, A. Olson, K. Chougule, P. Van Buren, D. Ware [in collaboration with E. Tseng, P. Baybayan, and K. Eng, Pacific Biosciences]

Haplotype phasing maize genetic variants is important for genome interpretation, population genetic analysis, and functional analysis of allelic activity. In 2019, we completed an isoform-level phasing study using two maize inbred lines and their reciprocal crosses, based on single-molecule full-length cDNA sequencing. To phase and analyze transcripts between hybrids and parents, we developed IsoPhase (see Fig. 2). Using this

tool, we validated the majority of single-nucleotide polymorphisms (SNPs) called against matching short-read data from embryo, endosperm, and root tissues; then we identified allele-specific, gene-level, and isoform-level differential expression between the inbred parental lines and hybrid offspring. After phasing 6,907 genes in the reciprocal hybrids, we annotated the SNPs and identified large-effect genes. In addition, we identified parent-of-origin isoforms, distinct novel isoforms in maize parent and hybrid lines, and imprinted genes from different tissues. Finally, we characterized variation in *cis*- and *trans*-regulatory effects. Our study provides measures of haplotypic expression that could increase accuracy in studies of allelic expression (Wang et al. 2020; Zou et al. 2020).

## PLANT GENETICS AND SYSTEMS BIOLOGY

The global challenges confronting agricultural security are falling into sharper relief: declining water for irrigation, surging pest pressures due to longer and hotter growing seasons, degrading arable land, increasing population, and long-term geographical adjustments brought about through climate change. Overcoming these strata of obstacles necessitates nimble and reliable approaches. Predictive genetics of desirable traits in concert with rapid germplasm conversion has become the norm since high-throughput sequencing has reached cost-effectiveness and genome editing and transformation techniques
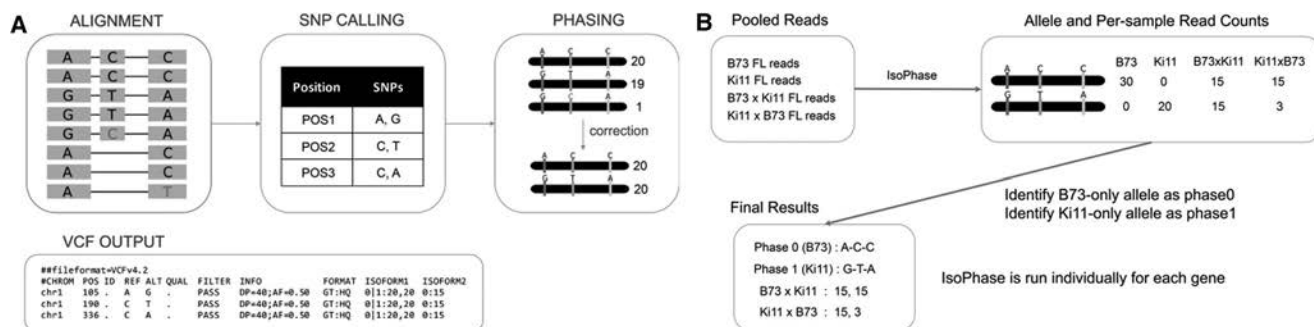


**Figure 2.** IsoPhase workflow. (*A*) For each gene, full-length reads from all 12 samples are aligned to a gene region. SNPs are called individually for each position using Fisher's exact test with the Bonferroni correction, applied with a *p*-value cutoff of 0.01. Only substitution SNPs are called. The full-length reads are then used to reconstruct the haplotypes, and a simple error-correction scheme is applied to obtain the two alleles. (*B*) To determine which allele is derived from B73 versus Ki11, we use the full length count information associated with the homozygous parents: B73 would only express the B73 allele, whereas Ki11 would only express the Ki11 allele.

continually improve; pan-genomes and new crop genomes are available for molecular investigation and comparative genomics, strengthening and accelerating the output of researchers and hastening the fruits of such labor into the hands of producers.

Although the predictive genomic paradigms are still being optimized, they are showing improved reliability depending on the desired trait. However, there are limitations when sought-after changes are focused around heretofore uncharacterized molecular mechanisms and gene regulatory networks acting in narrow spatiotemporal windows. It is crucial for plant science investigators to continue the molecular dissection of pathways controlling beneficial agronomic traits like flower fertility, inflorescence architecture, root formation, microbiome interaction, and nutrient use efficiency. These research arenas have noted quantitative improvements in crop yield and sustainability. Additionally, characterization across numerous plant species can yield a more unified systems biology model that can be effectively applied to numerous agricultural challenges.

## Characterizing Master Regulators of Primary and Lateral Root Development

L. Zhang, A. Olson, Y.K. Lee, C. Hu, D. Ware [in collaboration with C.A. Gaudinier and S. Brady, UC Davis]

Roots are essential plant organs that provide structural support and are primarily responsible for the acquisition of water and certain mineral nutrients. To improve our understanding of the genes controlling root development, we are using a gene-centric approach to characterize upstream regulators of miRNA families known to be involved in root development in the model plant *Arabidopsis thaliana*. Our approach uses a nearly complete root transcription factor (TF) library to screen promoters of root development–related miRNAs and their downstream targets resulting in a regulatory network of protein–DNA interactions, between transcription factors and the promoters. In 2019, we have integrated this with publicly available data, leading to a more complete gene regulatory network (GRN). We further utilize extensive, high-resolution spatiotemporal gene expression data and use predictive models to infer significant interactions and master regulators of root development. These methods were used to prioritize a set of TFs for evaluating their impact on root development in genetically perturbed lines.

## Dissection of Gene Regulatory Networks Associated with Nitrogen Use Efficiency

L. Zhang, A. Olson, V. Kumar, S. Kumari, K. Chougule [in collaboration with A.-M. Bagman, A. Gaudinier, and S. Brady, UC Davis; M. Frank and B. Shen, DuPont-Pioneer, Inc.]

Nitrogen (N) is an essential micronutrient for plants. Maximizing nitrogen use efficiency (NUE) in plants is critical to the increase in crop production and reduction of negative impacts on the environment due to seasonal excess of runoff fertilizer. To explore the GRN that controls these processes, we have used a gene-centric approach to characterize transcription factors that regulate genes involved in nitrogen uptake, transport, and metabolism in the model plant *Arabidopsis*. The outcome of the screen was the identification of 23 novel transcription factors. To prioritize which of these genes are more likely to have an impact on NUE and in turn impact biomass, root development, and time to flower, we developed NECorr (doi: https://doi.org/10.1101/326868), a model that combines network topology and expression data to rank genes and their interactions in a given tissue or condition. Using the prioritized gene list, we identified 26 loss-of-function mutants; amazingly, 25 of the genes perturbed showed a root or flowering time phenotype.

With the demonstration of efficacy in the model, we are now extending it to maize and sorghum—agronomically important species. Using rhAMP Seq, in collaboration with Integrated DNA Technologies (IDT), we have been able to map 15 ARS EMS (ethyl methanesulfonate) pools that harbored deleterious mutations for 16 candidate TFs. After two rounds of cross-breeding with MS8 (wild-type male sterile line) to remove background mutations, these mutant lines will be used to phenotype their fitness impact in the field.

## Developmental Networks Controlling Inflorescence Architecture and Grain Yield in Grasses

Y. Jiao, Y. K. Lee, S. Kumari, N. Gladman [in collaboration with Z. Xin and J. Burke, USDA ARS]

The objective of this work is to integrate genetics and genomics data sets to find molecular networks that influence a variety of agronomic traits to improve the morphology (architecture) of grass inflorescences (flowers), with a specific focus on *Zea mays* and more recently *Sorghum bicolor* [L.] Moench—a top-five global crop in terms of dedicated acreage. Because inflorescences

bear the fruits and grains that we eat, either through direct consumption or via animal feed, the genetic and regulatory factors that govern their development are obviously relevant to important agronomic traits such as grain yield and harvesting capability. Sorghum has recently emerged as a potentially potent bioenergy crop in addition to its role in human consumption in sub-Saharan Africa and other global regions. Because of its drought tolerance and tendency to be grown on marginal land, sorghum can also be used as a predictive genomics model for identifying water and nutrient-use efficiency gene candidates that could be implemented in other broadleaf crops.

The number of grains per panicle is a developmental trait contributing to overall sorghum yield. Sorghum flowers comprise one fertile (sessile) and two sterile (pedicellate) spikelets (see Fig. 3). However, only the sessile spikelet is fertile and will produce seeds. Using a publicly available EMS population, we identified independent multiseeded (*msd*) mutants that manifest both fertile sessile and pedicellate spikelets throughout the inflorescence. A detailed dissection of developmental stages of wild-type and *msd1*, *msd2*, and *msd3* lines revealed that pedicellate spikelets in wild-type do not have floral organs, including ovary, stigma, filament, or anther, whereas the *msd* mutants generate intact floral organs in both sessile and pedicellate spikelet. Using a bulk segregant analysis of F2 individuals, we determined that the *msd1* mutations are located within a TCP (Teosinte branched1/Cincinnata/proliferating cell factor family of genes) transcriptional factor, the *msd2*

mutations in a lipoxygenase (LOX) enzyme, and *msd3* mutations in a fatty acid desaturase (FAD) enzyme. The six causal SNPs found in *msd1* are highly conserved across grass species, as are the lesions in both *msd2* and *msd3* lines. The *TCP* gene was found to be differentially expressed during inflorescence development within a narrow spatiotemporal region. To characterize the gene networks associated with pedicellate spikelet fertility, we generated whole-genome expression profiling data of floral tissues at four different inflorescence development stages in both wild-type and *msd1* and *msd2*. Preliminary analyses suggested all *msd* lines may impact programmed cell death signaling in pedicellate spikelets in wild-type by modulating the jasmonic acid (JA) hormone pathway. To unbiasedly probe for MSD1 regulation targets, DNA affinity purification sequencing (DAPseq) was performed using bacterially expressed GST-MSD1 proteins. The successive peak-calling analysis determined that MSD1 binds near the transcriptional start site of other putatively regulatory genes, including other developmental transcription factors, signaling cascade proteins, and ribosomal proteins; these targets have been confirmed to be acted upon by TCPs in other plant species. The DAPseq analysis showed the direct regulation of JA biosynthetic pathway genes by MSD1. The transcriptomic analysis of *msd2* showed significant similarities to that of the *msd1* backgrounds, but with noticeable differences in assorted regulatory and JA pathway genes indicating more complex methods of regulatory network alteration in the mutants through different downstream effectors and feedback loops (Dampanaboina et al. 2019; Gladman et al. 2019).
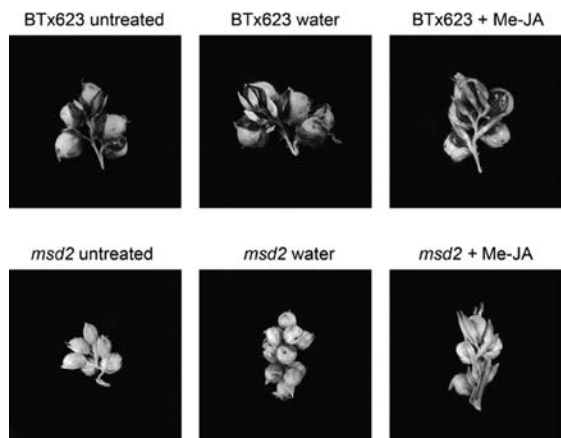


**Figure 3.** The *Multiseeded 2* phenotypic rescue of floral fertility with exogenous application of jasmonic acid during meristem development. Images are of smaller inflorescence sections taken from the larger seed head to more clearly show phenotypic rescue.

## Characterization of a Key Gene Involved in Sorghum Male Fertility

Y. Jiao, D. Ware [in collaboration with Z. Xin and J. Chen, USDA ARS]

Nuclear male sterility (NMS) is important for understanding microspore development and could facilitate the development of a two-line breeding system in sorghum. Several NMS lines and mutants have been identified in sorghum [*Sorghum bicolor* (L.) Moench], but no male-sterile gene has been reported previously. In this study, we characterized a new NMS mutant, *male-sterile 9* (*ms9*), which is distinct from all other reported NMS loci. The *ms9* mutant is stable under a variety of environmental conditions. Homozygous *ms9*

plants produced normal ovaries but small pale-colored anthers that contained no pollen grains. Other than the stable male sterility, the *ms9* mutant has no apparent defect in ovary development or any other developmental processes. Using MutMap, we identified the *Ms9* gene as a plant homeotic domain (PHD)-finger transcription factor similar to *Ms1* in *Arabidopsis* and *Ptc1* in rice. *Ms9* is the first NMS gene identified in sorghum. Thus, the *Ms9* gene and *ms9* mutant provide new genetic tools for studying pollen development and controlling male sterility in sorghum (Chen et al. 2019).

## Characterization of a High-Resolution Gene Expression Atlas Links Dedicated Meristem Genes to Key Architectural Traits in Maize

S. Kumari, D. Ware [in collaboration with M. Timmermans, CSHL]

The shoot apical meristem (SAM) orchestrates the balance between stem cell proliferation and organ initiation essential for postembryonic shoot growth. Meristems show striking diversity in shape and size. How this morphological diversity relates to variation in plant architecture and the molecular circuitries driving it are unclear. By generating a high-resolution gene expression atlas of the vegetative maize shoot apex, we show here that distinct sets of genes govern the regulation and identity of stem cells in maize versus *Arabidopsis*. Cell identities in the maize SAM reflect the combinatorial activity of TFs that drive the preferential, differential expression of individual members within gene families functioning in a plethora of cellular processes. Subfunctionalization thus emerges as a fundamental feature underlying cell identity. Moreover, we show that adult plant characteristics are, to a significant degree, regulated by gene circuitries acting in the SAM, with natural variation modulating agronomically important architectural traits enriched specifically near dynamically expressed SAM genes and the TFs that regulate them. Besides unique mechanisms of maize stem cell regulation, our atlas thus identifies key new targets for crop improvement (Knauer et al. 2019).

## CYBERINFRASTRUCTURE PROJECTS

The success of the Human Genome Project, completed at the start of the twenty-first century, inspired numerous large-scale sequencing projects such as the 1000 Genomes Project to catalog human genetic variations, 1000 Plant Genomes to sequence expressed genes of 1,000 different plant species, 1001 Genomes to discover whole-genome sequence variations in >1,001 strains of the reference plant model *Arabidopsis thaliana*, 100K Pathogen Genomes aimed at 100,000 infectious microbes, and Genome 10K targeting 10,000 vertebrate species. However, deriving biological function and meaningful predictive models from sequences have continued to be a challenge. As a result, the scale and complexity of genomics research have advanced from studying few genomic sequences of an organism to characterizing genome variations, gene expressions, biological pathways, and phenotypes for several thousands of organisms and their complex communities.

This has necessitated the availability of a wide array of computational tools that can process complex and heterogeneous data sets in an interoperable manner and sophisticated workflows that can seamlessly integrate these tools and their results at different stages of the analysis. Moreover, researchers often working collaboratively on large and complex systems need to be able to easily discover and integrate the analyses and results of their peers while simultaneously sharing their own results in a reproducible manner. Sustained efforts to lower the barriers to meet these monumental challenges have come to fruition in the form of a number of cyberinfrastructure platforms that adhere to the FAIR (findable, accessible, interoperable, reusable) guiding principles for research data stewardship. For more than a decade, the Ware laboratory has been invested in the growth and success of several such community initiatives since their inception. The rest of this section briefly discusses the specific projects in which the Ware laboratory is currently engaged and provides guidance and resources, particularly to champion the cause of the plant genomics community.

## Gramene: Comparative Genomic Resource for Plants

K. Chougule, Y. Jiao, V. Kumar, S. Kumari, Y.K. Lee, A. Olson, J. Stein, M.K. Tello-Ruiz, B. Wang, S. Wei, L. Zhang [in collaboration with P. Jaiswal, Oregon State University; P. Kersey, I. Papatheodorou, and R. Petryszak, EMBL-European Bioinformatics Institute; L. Stein, Ontario Institute of Cancer Research; C. Taylor, American Society of Plant Biologists; R. Wing, University of Arizona]

The Gramene project provides online reference resources for plant genomes and curated pathways

to aid functional genomics research in crops and model plant species. Our website (www.gramene.org) facilitates studies of gene function by combining genome and pathway annotation with experimental data and cross-species comparisons. In other words, the data and tools in Gramene enable plant researchers to use knowledge about gene function in one species to predict gene function in other species. Drawing these connections facilitates translational research in plant development and physiology that influences economically important traits—for example, grain development, flowering time, drought tolerance, and resistance to diseases. In 2019, the project accomplished several major milestones, culminating in our 62nd data release (November 2019), which included 67 plant genomes. The Gramene team develops comparative genomics databases in collaboration with the Ensembl Plants project at the European Bioinformatics Institute (EMBL-EBI) and collaborates closely with the EBI's Expression Atlas project to provide manually curated, quality-controlled, and analyzed transcriptomic data. We continue to host genome and pathway annotations via (1) the Ensembl genome browser, and (2) the Plant Reactome pathways portal (Naithani et al. 2020; Howe et al. 2019).

The Gramene project is actively engaging the community through various channels including webinars, presentations, talks, posters, and demonstrations during major community events including Plant Biology, Plant and Animal Genomes, and Biology of Genomes, providing training and the community's feedback on our current tools and user suggestions for new functionality. In 2019, a major focus has been the development of pan-genome resources while closely working with research and breeding initiatives in maize, sorghum, rice, and grapevine communities. Although generating a reference assembly has progressively become easier, there are still major challenges with an accurate prediction of the functional features in the genomes. To this end, a major effort has been in benchmarking and scaling existing algorithms to support improved and consistent gene annotation predictions. These automated predictions are the first step, and there remains a need for human review of the models. We have worked closely with the maize community and university instructors to train curators to help us review and identify problematic gene models and make improvements to the gene annotation pipelines (Tello-Ruiz et al. 2019). An outcome of these efforts has been the establishment of a network of researchers and teaching faculty working on course-based undergraduate research experiences (CUREs). When completed, these efforts will yield new insights into the taxonomic origin of maize genes and patterns of duplication, movement, and loss influenced by genome architecture. The maize V4 assembly and annotations are now available from the main Gramene genome browser. Our efforts in grapevine have included the development of reference genomes, and support for pan-genome resources, which led to the development of a core set of genus-wide genetic markers using the rhAmpSeq, for the breeding community (Zou et al. 2020).

## SciApps: A Cloud-Based Reproducible Workflow Platform

L. Wang, Z. Lu, X. Wang, P. Van Buren [in collaboration with Cold Spring Harbor Laboratory, PIs T. Gingeras, K. Birnbaum, D. Jackson, R. Martienssen, W.R. McCombie, D. Micklos, M. Schatz, D. Ware]

The Ware laboratory has developed SciApps, a cloud-based automated reproducible workflow platform leveraging the CyVerse infrastructure. It is designed to manage distributed and edge computing assets, including both academic and commercial cloud assets. For supporting the FAIR data principles, SciApps assigns a unique workflow ID to each data set analyzed, and the workflow ID is attached as metadata of raw data and derived data like genome browser tracks (Findable); all raw data, derived data, and metadata are retrievable through the unique workflow ID (Accessible); workflows and metadata are available as JSON files and extractable through the use of ontologies and standards (Interoperable/Reusable). In addition to FAIR data principles, SciApps tracks data provenance through analysis steps, recording results in standardized formats and providing access to scripts, runs, and results. As a workflow manager, SciApps provides users with a way to track analysis and ensures reproducibility across clouds via the underlying technology, Docker/Singularity, which is adapted from established community efforts, including BioConda and BioContainer.

For creating a workflow, each analysis job is submitted, recorded, and accessed through the SciApps web portal. Part or all of the series of recorded jobs are saved as reproducible, sharable workflows for future execution with the original/modified inputs

and parameters. The platform is designed to automate the execution of modular Agave apps and support executing workflows on either local or remote clusters or the combination of both. For executing workflows on U.S. supercomputing centers, users need to have a CyVerse account and upload their data to the CyVerse Data Store, a cloud storage capability built on the Integrated Rule-Oriented Data System (iRODS). All intermediate and final results of the analysis workflow are archived back to the CyVerse Data Store so that any step in the workflow can be easily reproduced.

Besides derived data, each SciApps workflow also captures the complete computational metadata to reproduce the entire analysis with one click and provides direct links to input data and associated experimental metadata residing inside the CyVerse Data Store. Therefore, SciApps workflows have been used to share the complete analysis and input data among users or to the public (e.g., for hosting and releasing the MaizeCODE data sets).

MaizeCODE, a project for an initial analysis of functional elements in the maize genome, has assayed five tissues of four maize genomes (B73, NC350, W22, TIL11) for RNA-seq, ChIP-seq, Rampage, small RNA, and MNase (outside collaboration). MaizeCODE is committed to open-access and reproducible science based on the FAIR data principles, supporting various ways to access data: First, all raw data are available through the CyVerse Data Store, and the user can bulk download all data sets through iCommands (command line) or CyberDuck (GUI). Second, all ground-level analysis results of MaizeCODE data are available through SciApps (https://www.sciapps.org/?page_id=dataWorkflows&data_item=MaizeCODE) for supporting downstream analyses. For example, differential expression analysis between tissues can be done in a few minutes because all gene quantification results are staged and chained together by the SciApps workflows. Third, peaks and signals from the ground-level analysis are available on a Genome Browser, JBrowser. To increase data discoverability, by calling CyVerse API, SciApps workflow IDs are added as metadata of the raw data in CyVerse Data Store during the automated analysis process (through SciApps API) so that the users can retrieve the entire analysis when browsing through the CyVerse Data Store or Data Commons. On SciApps.org, the workflow diagram provides a direct link to the raw data via CyVerse Data Commons, so that users can check the experimental

metadata associated with each data set or workflow. On the Genome Browser side, SciApps workflow IDs are attached to each browser track for bringing up the specific workflow in a web-based iframe when users click on the name of the track. In this way, the user can check all parameters that have been used to generate the track data, quality control (QC) report, and metadata of the raw data or reproduce the entire analysis on the cloud with a click.

In addition, all raw data has been submitted to the NCBI Sequence Read Archive (SRA) using the SRA submission pipeline from the CyVerse Discovery Environment (DE). Through the submission process, all experimental metadata are stored in the iRODS-based Data Store of CyVerse and used for automating the ground-level analysis on SciApps.org. SciApps organizes both replicates (and controls if available) of each assay as one experiment (or a workflow with the unique ID), which represents an entity that chains raw data, analysis results, experimental metadata, and computational metadata together. The ground-level analysis includes QC, alignment to the reference genome, filtering, quantification (e.g., for gene expression), and peak calling (if needed). In summary, SciApps provides both a graphical user interface (GUI) and a RESTful API for users to check QC results, process new data, and reproduce existing analysis on the cloud.

## KBase: Department of Energy Systems Biology Knowledgebase

V. Kumar, S. Kumari, Z. Lu [in collaboration with DOE National Laboratories and led by PI A. Arkin, Lawrence Berkeley National Laboratory (LBNL); co-PIs C. Henry, Argonne National Laboratory (ANL) and R. Cottingham of Oak Ridge National Laboratory (ORNL); as Plants Science Lead for KBase, D. Ware continues to informally serve as a co-PI on the project]

The Systems Biology Knowledgebase (KBase, http://kbase.us) is a free, open-source platform for systems biology of plants, microbes, and their communities at scales ranging from the biomolecular to the ecological. The users can collaboratively generate, test, compare, and share hypotheses about biological functions, perform large analyses on scalable computing infrastructure, and finally combine experimental evidence and conclusions to model plant and microbial physiology and community dynamics. The KBase platform has extensible analytical capabilities that currently include (meta) genome assembly, annotation, comparative genomics,

transcriptomics, and metabolic modeling; a web-based user interface that supports building, sharing, and publishing reproducible and well-annotated analyses with integrated data; and a software development kit that enables the community to add functionality to the system. The Ware laboratory has been engaged in the development and maintenance of a number of analysis tools and data resources that enable the plant science community to gain insight into the evolution of genes and genomes, profile transcriptomes, perform genome functional modeling with metabolic networks, and identify differential expression between tissues, developmental stages, environmental conditions, and genetic backgrounds. These capabilities are directly relevant to important DOE research targets such as optimizing biomass production in biofuel feedstocks.

KBase was conceived from the beginning as a knowledgebase that would bring together relevant computational systems biology tools and data for microbes, plants, and interactions between the two. The KBase team rolled out in 2019 new platform functionalities such as feeds, research teams represented as organizations, and a few initial apps in the emerging knowledge engine and relation engine. Organizations provide a means by which KBase users can self-organize to find or create collaborative groups with similar interests to share data, narratives, and tools. Feeds provide users with useful updates about the system and teams they are associated with and data/narratives they have shared with other users. In addition to the platform development, the KBase team embarked upon a number of new scientific capabilities related to experimental and sample design, metagenome annotation design, design and implementation of taxonomy and updated ontology systems, and update of ModelSEED biochemistry database in KBase. The Ware laboratory team is engaged in supporting these design, implementation, and testing efforts particularly in the areas of sample/experiment, taxonomy, and ontology as well as expression profiling and review of public data sets to support compendia and the knowledge engine. These features will ultimately evolve into knowledge-discovery features, enabling KBase to propose new hypotheses by making connections across the system.

KBase is actively engaging the external community to help us improve our tools and workflows for plant science, including support for large-scale reads upload and analysis, plant genome annotation, functional genomic clustering and enrichment, physiological modeling and variation, and trait-based modeling analysis. The Ware laboratory has actively engaged the community through various channels such as webinars, presentations, talks, posters, and demonstrations during major community events including Plant Biology, Plant and Animal Genomes, and Biology of Genomes. The Ware laboratory team also hosted a functional genomics workshop at the Joint Genome Institute (JGI) annual user meeting. We welcome the community's feedback on our current tools and your suggestions on what new functionality we should add, and invite you to share your plant science Narratives with the community.

## PUBLICATIONS

Chen JP, Jiao YP, Laza H, Payton P, Ware D, Xin ZG. 2019. Identification of the first nuclear male sterility gene (*Male-sterile 9*) in sorghum. *Plant Genome* **12:** 1–12.

Dampanaboina L, Jiao Y, Chen J, Gladman N, Chopra R, Burow G, Hayes C, Christensen SA, Burke J, Ware D, Xin Z. 2019. Sorghum *MSD3* encodes an ω-3 fatty acid desaturase that increases grain number by reducing jasmonic acid levels. *Int J Mol Sci* **20:** 5359.

Gladman N, Jiao Y, Lee YK, Zhang L, Chopra R, Regulski M, Burow G, Hayes C, Christensen SA, Dampanaboina LZ, et al. 2019. Fertility of pedicellate spikelets in sorghum is controlled by a jasmonic acid regulatory module. *Int J Mol Sci* **20:** 495.

Howe KL, Contreras-Moreira B, De Silva N, Maslen G, Akanni W, Allen J, Alvarez-Jarreta J, Barba M, Bolser DM, Cambell L, et al. 2019. Ensembl Genomes 2020—enabling non-vertebrate genomic research. *Nucleic Acids Res* **49:** D884–D891.

Knauer S, Javelle M, Li L, Li X, Ma X, Wimalanathan K, Kumari S, Johnston R, Leiboff S, Meeley R, et al. 2019. A high-resolution gene expression atlas links dedicated meristem genes to key architectural traits. *Genome Res* **29:** 1962–1973

Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* **20:** 275.

Tello-Ruiz MK, Marco CF, Hsu FM, Khangura RS, Qiao P, Sapkota S, Stitzer MC, Wasikowski R, Wu H, Zhan J, et al. 2019. Double triage to identify poorly annotated genes in maize: the missing link in community curation. *PLoS ONE* **14:** e0224086.

Wang B, Kumar V, Olson A, Ware D. 2019. Reviving the transcriptome studies: an insight into the emergence of single-molecule transcriptome sequencing. *Front Genet* **10:** 384.

*In Press*

Naithani S, Gupta P, Preece J, D'Eustachio P, Elser JL, Garg P, Dikeman DA, Kiff J, Cook J, Olson A, et al. 2020. Plant Reactome: a knowledgebase and resource for comparative pathway analysis. *Nucleic Acids Res* **48:** D1093–D1103.

Wang B, Tseng E, Baybayan P, Eng K, Regulski M, Jiao Y, Wang L, Olson A, Chougule K, Van Buren P, Ware D. 2020. Variant phasing and haplotypic expression from long-read sequencing in maize. *Commun Biol* **3:** 78.

Zou C, Karn A, Reisch B, et al. Haplotyping the *Vitis* collinear core genome with rhAmpSeq improves marker transferability in a diverse genus. *Nat Commun* **11:** 413.

# QUANTITATIVE BIOLOGY

Human development requires the regulated activity of thousands of genes in hundreds of distinct cell types throughout life. One requirement for this process is that each cell must contain an intact, functional genome free from mutations. One type of mutation can arise from the activation of transposable elements (TEs). These viral-like parasites lay dormant within our genomes but have the capacity to hop into new genomic locations, causing mutations as they break the surrounding DNA sequence. Mounting evidence has implicated transposon activity in a host of human diseases, with particular evidence for TE activation in neurodegenerative diseases: amyotrophic lateral sclerosis (ALS) and fronto-temporal dementia (FTD). This is the focus of the research in **Molly Gale Hammell**'s laboratory.

**Ivan Iossifov** focuses on the development of new methods and tools for genomic sequence analysis and for building and using molecular networks, and he applies them to specific biomedical problems. He studies the genetics of common diseases in humans using two main tools: next-generation sequencing and molecular networks representing functional relationships among genetic loci. These approaches in combination enable the kind of large-scale studies necessary for furthering our understanding of the complex etiology of disorders such as autism, bipolar disorder, and cancer.

**Justin Kinney** completed his Ph.D. in Physics at Princeton University in 2008 and began his term as a Quantitative Biology Fellow in 2010. His research focuses on developing next-generation DNA sequencing as a tool for dissecting the structure and function of large macromolecular complexes. Of particular interest to his laboratory is the biophysical basis of transcriptional regulation—how simple interactions between proteins and DNA allow promoters and enhancers to modulate genes in response to physiological signals. In 2010, Kinney and colleagues published a paper showing Sort-Seq, a novel sequencing-based method that can measure the functional activity of hundreds of thousands of slightly mutated versions of a specific DNA sequence of interest. Using a novel information-theoretic analysis of the resulting data, Kinney et al. were able to quantitatively measure, in living cells, the protein–DNA and protein–protein interactions controlling mRNA transcription at a chosen promoter. Kinney continues to develop this approach using a combination of theory, computation, and experiment. From a biological standpoint, Sort-Seq allows researchers to investigate important but previously inaccessible biological systems. Kinney's laboratory is currently using Sort-Seq to address open problems in transcriptional regulation, DNA replication, and immunology. These experiments also present new challenges for the field of machine learning, and a substantial fraction of Kinney's efforts are devoted to addressing the theoretical and computational problems relevant to the analysis of Sort-Seq data.

**Alexander Krasnitz** and colleagues develop mathematical and statistical tools to investigate population structure of cells comprising a malignant tumor and to reconstruct evolutionary processes leading to that structure. These tools are designed to make optimal use of emerging molecular technologies—chief among them high-throughput genomic profiling of multiple individual cells harvested from a tumor. By analyzing these profiles, Krasnitz derives novel molecular measures of malignancy, such as the number of aggressive clones in a tumor, the invasive capacity of each clone, and the amount of cancer-related genetic alteration sustained by clonal cells. Krasnitz and colleagues collaborate closely with clinical oncologists to explore the utility of such measures for earlier detection of cancer, more accurate patient outcome prediction and risk assessment, and better-informed choice of treatment options.

There is increasing evidence that rare and unique mutations have a significant role in the etiology of many diseases such as autism, congenital heart disease, and cancer. **Dan Levy**'s group develops algorithms to identify these mutations from large, high-throughput data sets comprising thousands of nuclear families. After earlier working with high-resolution comparative genome hybridization (CGH) arrays, Levy's group now uses targeted sequence data. Levy has developed methods for identifying de novo mutations (i.e., those seen in a child but not in his or her parents) by simultaneously genotyping the entire family; the team is currently focused on building algorithms to detect copy number variants and multiscale genomic rearrangements. Although their copy number methods are based on "read" density, there are classes of mutations that require analysis at the level of the read. Thus, they are developing algorithms to identify insertions, deletions, inversions, transpositions, and other complex events. Other projects in the Levy laboratory include analysis of single-cell RNA, phylogenetic reconstruction from sparse data sets, and disentangling haplotypes from sperm and subgenomic sequence data.

**David McCandlish**'s laboratory develops computational and mathematical tools to analyze and exploit data from high-throughput functional assays. The current focus of the laboratory is on analyzing data from so-called "deep mutational scanning" experiments. These experiments simultaneously determine, for a single protein, the functional effects of thousands of mutations. By aggregating information across the proteins assayed using this technique, they seek to develop data-driven insights into basic protein biology, improved models of molecular evolution, and more accurate methods for predicting the functional effects of mutations in human genome sequences.

Critically, these data also show that the functional effects of mutations often depend on which other mutations are present in the sequence. The members of the McCandlish laboratory are developing new techniques in statistics and machine learning to infer and interpret the complex patterns of genetic interaction observed in these experiments. Their ultimate goal is to be able to model these sequence–function relationships with sufficient accuracy to guide the construction of a new generation of designed enzymes and drugs and to be able to predict the evolution of antigenic and drug resistance phenotypes in rapidly evolving microbial pathogens.

The thymus generates and selects a highly variable yet specific T-cell repertoire that discriminates between self and nonself antigens. Within the thymus, medullary thymic epithelial cells express a diverse set of antigens, representing essentially all tissues of the body. This phenomenon, termed promiscuous gene expression, imposes central T-cell self-tolerance by enabling peripheral antigens to be continuously accessible to developing T cells.

Ultimately, understanding the physiological mechanisms that lead to self-tolerance will be crucial in understanding autoimmunity and autoimmune diseases. Despite extensive work on the molecular basis of promiscuous gene expression, many questions concerning both single-cell and tissue-level organization of antigen expression are unanswered. For instance, is promiscuous gene expression organized in a spatially or temporally restricted manner, or a combination of both? How do thymic epithelial cells maintain their integrity despite expressing peripheral proteins that may interfere with epithelial-specific pathways and roles? **Hannah Meyer**'s research group seeks to answer these questions by combining genomics and mathematical modeling. They investigate the spatial and temporal organization of promiscuous gene expression in the thymus and how antigen distribution in the thymus affects epithelial–T-cell interaction and migration. Analogously, on the single-cell level, they conduct research on the spatiotemporal organization of promiscuous gene expression to understand cellular integrity mechanisms in thymic epithelial cells.

Modern genomic technologies make it relatively easy to generate rich data sets describing genome sequences, RNA expression, chromatin states, and many other aspects of the storage, transmission, and expression of genetic information. For many problems in genetics today, the limiting

step is no longer in data generation, but in integrating, interpreting, and understanding the available data. Addressing these challenges requires expertise both in the practical arts of data analysis and in the theoretical underpinnings of statistics, computer science, genetics, and evolutionary biology. **Adam Siepel**'s group focuses on a diverse collection of research questions in this interdisciplinary area. Over the years, their research has touched on topics including the identification of recombinant strains of HIV, the discovery of new human genes, the characterization of conserved regulatory elements in mammalian genomes, and the estimation of the times in early human history when major population groups first diverged. A general theme in their work is the development of precise mathematical models for the complex processes by which genomes evolve over time and the use of these models, together with techniques from computer science and statistics, both to peer into the past and to address questions of practical importance for human health. Recently, they have increasingly concentrated on research at the interface of population genomics and phylogenetics, with a particular focus on humans and the great apes. They also have an active research program in computational modeling and analysis of transcriptional regulation in mammals and *Drosophila*, in close collaboration with Professor John Lis at Cornell University.

# THE CONTRIBUTION OF TRANSPOSABLE ELEMENTS TO NEURODEGENERATIVE DISEASE

M. Gale Hammell    T.L. Forcier  Y. Jin        K. O'Neill    R. Shaw    C. Wunderlich
                   Y. Hao        C. Marshall   N. Rozhkov    O. Tam     M-K. Yip

Human development requires the regulated activity of thousands of genes in hundreds of distinct cell types throughout life. One requirement for this process is that each cell must contain an intact, functional genome free from mutations. One type of mutation can arise from the activation of transposable elements (TEs). These viral-like parasites lie dormant within our genomes but have the capacity to hop into new genomic locations, causing mutations as they break the surrounding DNA sequence. Mounting evidence has implicated transposon activity in a host of human diseases, with particular evidence for TE activation in neurodegenerative diseases, such as Alzheimer's disease (AD), amyotrophic lateral sclerosis (ALS), and frontotemporal dementia (FTD).

## Characterization of Genes That Control Transposable Element Activity

C. Marshall, K. O'Neill, N. Rozhkov, R. Shaw, O. Tam

Human cells devote extensive resources to controlling the activity of TEs to protect the integrity of the genome and to prevent the immune system from mistakenly recognizing TE sequences as pathogenic. Although several genes that control TE activity have been described, the full list of genes that regulate TEs is known to be incomplete. One aim of the Gale Hammell laboratory is to find new genes that contribute to the silencing of transposable elements. The most recently described is an RNA-binding protein called TDP-43. TDP-43 has long been known for its role in regulating the processing of host gene mRNA in neurons. Mutations in TDP-43 have been associated with a variety of neurodegenerative diseases including ALS, frontotemporal lobar degeneration (FTLD), and AD. However, the normal function of TDP-43 in neuronal development and maintenance has not been fully characterized and few of its mRNA targets have been definitively associated with the neurodegenerative

diseases that result from loss of TDP-43 function. In collaboration with the Dubnau laboratory at Stony Brook University, the group has explored the novel hypothesis that TDP-43 normally plays a large and hitherto uncharacterized role in regulating the expression of TEs. In previous publications, members of the Gale Hammell laboratory have shown that TDP-43 binds widely to RNA transcripts from TEs in the fly and mouse, and that TDP-43 binding to TEs is lost in human patients diagnosed with FTLD, a disease characterized by dysfunction of TDP-43 protein. In the latest work from our laboratory, we show that TDP-43 directly binds TE sequences in human neuronal-like cells, and that loss of this binding leads to elevated expression of a wide variety of TE sequences (Tam et al. 2019a). Although this firmly places TDP-43 in the list of genes that contribute to regulation of transposable element activity in human cells, it is not known how TDP-43 normally causes these TEs to be silenced. Ongoing work is focused on identifying the partners of TDP-43 protein that directly silence TE loci, preventing their transcription and degrading any transcripts produced from these loci.

## Transposable Elements Mark a Subtype of Amyotrophic Lateral Sclerosis (ALS)

N. Rozhkov, R. Shaw, O. Tam

Given the identification of TDP-43 protein as a regulator of TE activity, our laboratory next sought to determine whether activation of TEs would be seen in the tissues of patients with ALS and FTD, two neurodegenerative diseases associated with TDP-43 dysfunction. In collaboration with a large ALS patient sequencing consortium under way at the New York Genome Center (NYGC), members of the Gale Hammell laboratory have integrated targeted genotyping data with deep expression profiling for hundreds of ALS patients and controls. Analysis of these patient

profiles has revealed that elevated TE expression does occur in the cortical regions for a substantial fraction of ALS patients (Tam et al. 2019a). Moreover, the patients with the highest levels of TE activity were also the most likely to show large aggregates of TDP-43 protein in cortical tissues, strongly linking TDP-43 dysfunction to TE activity.

Given the fact that TDP-43 pathology and elevated TE activity were only seen in a subset of ALS patients, our laboratory next sought to understand why the ALS patient samples seemed to display such heterogeneity. We developed sophisticated machine learning methods based on non-negative matrix factorization (NMF) algorithms to determine why some ALS patient samples would display high levels of TE activity, whereas other patient samples seemed to display other molecular alterations not associated with the activity of TEs. The results of this study demonstrated that
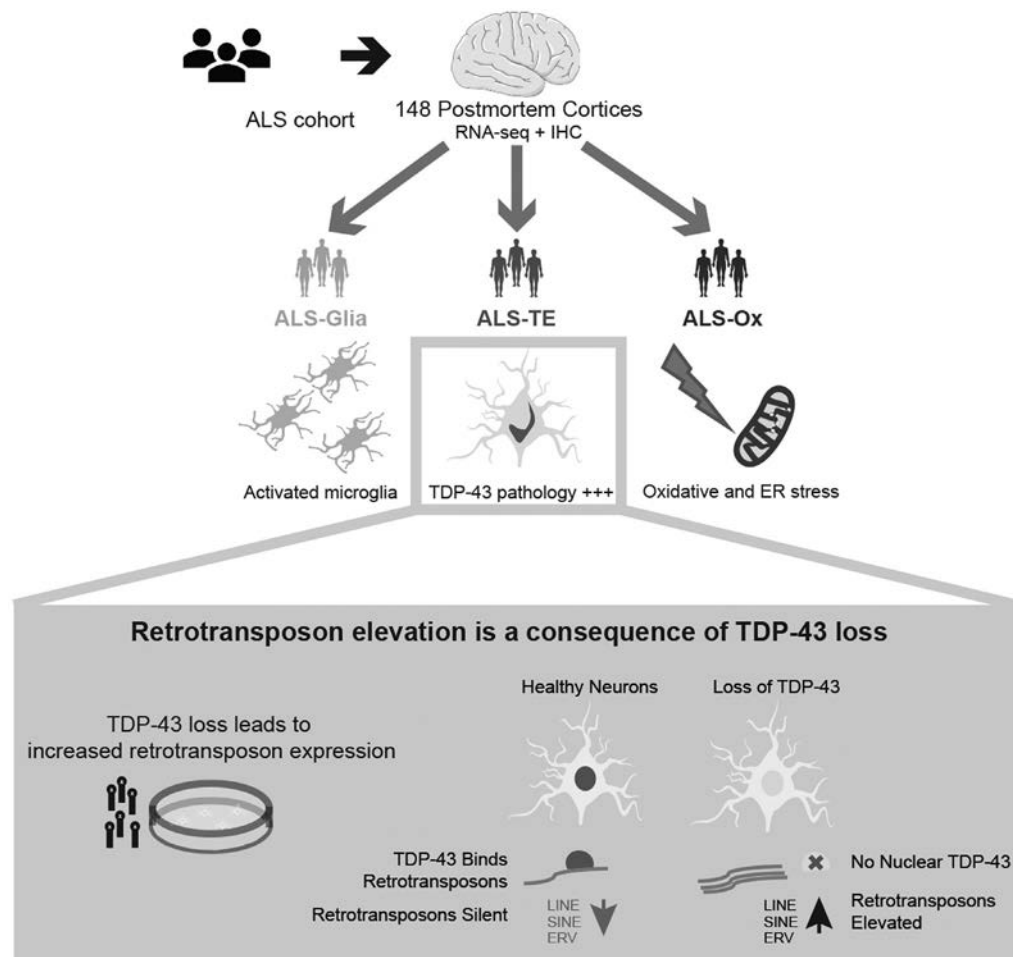


**Figure 1.** Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease characterized by the progressive loss of motor neurons. Although several pathogenic mutations have been identified, the vast majority of ALS cases have no family history of disease. Thus, for most ALS cases, the disease may be a product of multiple pathways contributing to varying degrees in each patient. Using machine-learning algorithms, we stratify the transcriptomes of 148 ALS postmortem cortex samples into three distinct molecular subtypes. The largest cluster, identified in 61% of patient samples, displays hallmarks of oxidative and proteotoxic stress. Another 19% of the samples shows predominant signatures of glial activation. Finally, a third group (20%) exhibits high levels of retrotransposon expression and signatures of TDP-43 dysfunction.

ALS samples could be categorized into three different subtypes (Fig. 1) characterized by three distinct types of cellular dysfunction. The first subtype (ALS-TE) was characterized by elevated TE activity, representing 20% of the ALS patient samples. A second subtype, ALS-Glia, representing an additional 20% of ALS patient samples, showed strong signatures of neuroinflammation as characterized by the activation of glial cells (astrocytes and microglia) that act as immune-like cells in the central nervous system. A final subtype, ALS-Ox, represented the remaining 60% of all ALS samples and showed strong markers of oxidative stress as well as failures in the autophagic systems that normally degrade misfolded proteins. Further details on the NYGC ALS Consortium study and ALS subtypes can be found in Tam et al. (2019a). The laboratory was recently awarded a large grant from the Chan Zuckerberg Initiative to follow up on these results in a larger set of ALS patient tissues, and to see if related processed are occurring in other neurodegenerative diseases (as reviewed in Tam et al. 2019b).

## Transposon Genomics: Statistically Rigorous Algorithms for TE Data Analysis

T. Forcier, Y. Jin, K. O'Neill, O. Tam, C. Wunderlich

Transposable elements are viral-like sequences in our genomes that have been historically difficult to study because of their highly repetitive nature. Millions of copies of TE sequences are scattered throughout the chromosomes, including within other gene sequences. Although most of these copies are nonfunctional, thousands of TEs retain the ability to mobilize and create new genetic mutations. The difficulty lies in differentiating these active TEs from the millions of other harmless copies with nearly identical sequences. This presents both a technical challenge for experimentally isolating TE-derived sequences from the genomes of cells as well as a computational challenge for determining where each sequenced TE copy originates in the genome of a particular sample. Members of the Gale Hammell laboratory have recently developed a suite of novel statistical inference methods to solve the computational challenge of analyzing TE activity in genomic sequencing studies such as TEtranscripts for RNA-Seq data, TEsmall for small-RNA-Seq, TEpeaks for ChIP-seq, and TEsingle for single-cell RNA-Seq and nuc-seq data. All these algorithms, which together form the TEtoolkit, use maximum likelihood frameworks to statistically infer the correct originating locus of sequencing reads that map ambiguously to many related genomic regions. These algorithms have been used to examine the basic mechanisms of controlling TE expression and activity within the Gale Hammell laboratory, and also in several collaborative works, including a recent study on Piwi-interacting RNAs that silence TEs in early development (Stein et al. 2019). Together, these efforts will provide the computational infrastructure with which to determine the extent of TE activity in human development and disease.

## PUBLICATIONS

Anaparthy N, Ho YJ, Martelotto L, Hammell M, Hicks J. 2019. Single-cell applications of next-generation sequencing. *Cold Spring Harb Perspect Med* **9:** a026898. doi:10.1101/cshperspect.a026898

Stein CB, Genzor P, Mitra S, Elchert AR, Ipsaro JJ, Benner L, Sobti S, Su Y, Hammell M, Joshua-Tor L, Haase AD. 2019. Decoding the 5′ nucleotide bias of PIWI-interacting RNAs. *Nat Commun* **10:** 828. doi:10.1038/s41467-019-08803-z

Tam OH, Rozhkov NV, Shaw R, Kim D, Hubbard I, Fennessey S, Propp N, Fagegaltier D, Harris BT, Ostrow LW, et al. 2019a. Postmortem cortex samples identify distinct molecular subtypes of ALS: retrotransposon activation, oxidative stress, and activated glia. *Cell Rep* **29:** 1164–1177. doi:10.1016/j.celrep.2019.09.066

Tam OH, Ostrow LW, Gale Hammell M. 2019b. Diseases of the nERVous system: retrotransposon activity in neurodegenerative disease. *Mob DNA* **10:** 32. doi:10.1186/s13100-019-0176-1

# GENETIC VARIANTS LINKED TO AUTISM TRAITS

**I. Iossifov**  Y-h. Lee   A. Muñoz
              S. Marks   B. Yamrom

In 2019, the bulk of our work was in analyzing the large data set of whole-genome sequencing (WGS) data generated from ~2,400 of the Simons Simplex Collection (SSC) families and ~900 families from the Autism Genetic Resource Exchange (AGRE), a collection of families with multiple children with autism. We also started the analysis of the whole-exome sequencing (WES) data from the growing SPARK collection: data for ~5,000 of the SPARK families have been released and SPARK is expected to grow to ~50,000 families in a couple of years. In addition, we started a pilot project to explore the potential of RNA sequencing in family collections like SSC. These data are a rich resource that we use in numerous projects.

Below are listed abstracts of four projects that are in submission or near-submission status and that show our current efforts in studying the role of de novo noncoding variants, rare structural rearrangements, and common variants in autism's etiology.

## De Novo Disruption of Introns Contributes to Autism

A. Muñoz, B. Yamrom, Y-h. Lee, S. Marks, I. Iossifov
[in collaboration with S. Yoon, P. Andrews, Z. Wang,
D. Levy, and M. Wigler, CSHL; C. Reeves and L. Winterkorn,
New York Genome Center; A.M. Krieger, A. Buja,
and K. Pradhan, Albert Einstein College of Medicine,
Montefiore Medical Center; K.K. Baldwin, Scripps
Research Institute]

Autism arises in high- and low-risk families. De novo mutation contributes to autism incidence in low-risk families as there is a higher incidence in the affected of the simplex families than in their unaffected siblings. But the true rate in low-risk families cannot be determined solely from simplex families, which are a mixture of low and high risk; and the rate of de novo mutation in nearly pure populations of high-risk families, the multiplex families, has not previously been rigorously determined. Moreover, rates of de novo mutation have been underestimated from studies based on low-resolution microarrays and WES. Here, we report

on findings from WGS of both simplex families from the SSC and multiplex families from AGRE. After removing the multiplex samples with excessive genomic drift, we find that the contribution of de novo mutation in multiplex is significantly smaller than the contribution in simplex. We use WGS to provide high-resolution CNV profiles, to analyze more than coding regions, and revise upward the rate in simplex caused by an excess of de novo events targeting introns. Based on this study, we now estimate that de novo events contribute to 60%–80% of cases of autism arising in low-risk families.

## MUMdex: MUM-Based Structural Variation Detection

I. Iossifov, S. Marks [in collaboration with P.A. Andrews,
J. Kendall, Z. Wang, D. Levy, and M. Wigler, CSHL;
L. Muthuswam, New York Genome Center]

Standard genome sequence alignment tools primarily designed to find one alignment per read have difficulty detecting inversion, translocation, and large insertion and deletion events. Moreover, dedicated split read alignment methods that depend only on the reference genome may misidentify or find too many potential split read alignments because of flaws in the reference genome.

We introduce MUMdex, a maximal unique match (MUM)-based genomic analysis software package consisting of a sequence aligner to the reference genome, a storage-indexing format, and analysis software. Discordant reference alignments of MUMs are especially suitable for identifying inversion, translocation, and large indel differences in unique regions. Extracted population databases are used as filters for flaws in the reference genome. We describe the concepts underlying MUM-based analysis, the software implementation, and its usage.

We show via simulation that the MUMdex aligner and alignment format are able to correctly detect and record genomic events. We characterize alignment

performance and output file sizes for human whole-genome data and compare to Bowtie 2 and the BAM format. Preliminary results show the practicality of the analysis approach by detecting de novo mutation candidates in human whole-genome DNA sequence data from 510 families. We provide a population database of events from these families for use by others.

## A Platform for Access and Analysis of Genetic Variants in Phenotype-Rich Family Collections

Y-h. Lee, B. Yamrom, S. Marks, I. Iossifov [in collaboration with M. Cokol, Axcella, Boston; A. Nenkova, University of Pennsylvania; L. Chorbadjiev, SeqPipe Ltd., Sofia]

WES, a technique that enables the inexpensive identification of genetic variants in the gene-encoding regions of the genomes of thousands of people, is quickly transforming human genetics. Particularly successful are the numerous studies that used WES in large collections of families to study the genetic architectures of human disorders with strong detrimental effect in fecundity, including autism, intellectual disability, schizophrenia, epilepsy, and congenital heart disease. These studies identified large numbers of genetic variants segregating in the families or arising de novo in children, gathered detailed phenotypic measurements of the studied individuals, and used the complex data sets to develop models of genotype and phenotype relationships.

There is an enormous amount of work that needs to follow the early success in the genetics of such complex disorders to develop effective treatment and early diagnostic strategies. A variety of future research projects will study in detail the effects of hundreds of genetic variants and genes at molecular, cellular, and organismic levels. Such projects will greatly benefit from the accumulated family WES data sets, but their large size and complex structure create a major obstacle to their efficient use. Here, we present the GPF (genotype and phenotype in families) system, which manages such data sets and has an intuitive interface that makes it possible for the wider scientific community to benefit from the new collections.

## RNA-Seq of SSC

We are finalizing our analysis of the whole-genome data from ~2,400 of the SSC families. The major

result of that effort is the estimate of the contribution of the de novo noncoding variants. Specifically, we observed a significantly increased rate of de novo intronic variants in affected children compared with their unaffected siblings when we restrict the rate observation to the autism genes previously implicated by WES. The increase in the rate is consistent with a contribution of ~5% of de novo intronic indels to the autism diagnosis in SSC (see the De Novo Disruption of Introns Contributes to Autism section above). We do not observe similar increase in the rate of de novo intronic substitutions, but it is expected that the size of the study is insufficient to detect that signal given the much higher rate of background noise for substitutions. Nevertheless, we also expect that de novo intronic substitutions have a contribution, and we guess that is likely of similar magnitude to the contribution of the de novo intronic indels. As others have reported an increased rate of de novo mutation in affected versus unaffected children within the control regions of the intergenic space, we expect that the contribution of noncoding de novo mutation is close to 15%, perhaps only slightly less than the contribution from de novo coding mutation.

Despite the large contribution of the noncoding variants, we have no good purely analytic method to distinguish the specific causal sequence variation from the many random ones. We proposed to address that through study of the RNA. We expect that for the majority of the causal de novo noncoding variants, the immediate effect would be on the expression of nearby genes, and such changes in expression can be detected through RNA-Seq by comparing the expression of the affected gene allele to that of the unaffected allele, a method called allele-specific expression (ASE). In the last year, we initiated pilot experiments to test the feasibility of this approach. This is a collaborative effort including several groups at CSHL (Wigler and Levy laboratories), several groups at the New York Genome Center (Hemali Phatnani's and Tuuli Lappalainen's groups and Tom Maniatis), and Kristen Baldwin from the Scripps Research Institute.

We have access to Epstein–Barr virus (EBV) immortalized lymphoblastoids (LCLs) for all the individuals of the SSC. One of our pilot studies addressed the question of whether LCLs are a good resource for studying ASE. In collaboration with Kristen Baldwin, we transformed an LCL into an induced pluripotent stem cell (iPSC) and further down neuronal lineages.

We then generated and sequenced RNA libraries from the original LCLs, the iPSCs, and the derived neurons. The analysis of these data is ongoing, but we have already made a few useful observations. First, the transformation process worked successfully. Second, the derived neurons express nearly 90% of the autism genes identified by exome sequencing, whereas the LCLs express ~70% of these genes. Third, when a gene is expressed in both cell types, the ASE is preserved.

In an additional pilot study we performed RNA-Seq from LCLs of six of the SSC families, two of which had identical affected twins and an unaffected sibling, and four of which had one affected and one unaffected child. Among the goals of this pilot study were tuning our bench protocols and analytical tools and estimating the noise sources (such as *trans*-regulatory and epigenetic effects) that would decrease the power of detecting ASE. It appears from the pilot study that such noise sources are not negligible, but are manageable.

## PUBLICATIONS

Jansen S, van der Werf IM, Innes AM, Afenjar A, Agrawal PB, Anderson IJ, Atwal PS, van Binsbergen E, van den Boogaard MJ, Castiglia L, et al. 2019. De novo variants in *FBXO11* cause a syndromic form of intellectual disability with behavioral problems and dysmorphisms. *Eur J Hum Genet* **27:** 738–746. doi:10.1038/s41431-018-0292-2

*In Press*

Venkatasubramani JP, Subramanyam P, Pal R, Reddy BK, Srinivasan DJ, Chattarji S, Iossifov I, Klann E, Bhattacharya A. 2020. N-terminal variant Asp14Asn of the human p70 S6 kinase 1 enhances translational signaling causing different effects in developing and mature neuronal cells. *Neurobiol Learn Mem* **171:** 107203. doi:10.1016/j.nlm.2020.107203

# GENE REGULATION, BIOPHYSICS, AND MACHINE LEARNING

**J.B. Kinney**  A. Ayaz  A. Posfai
W-C. Chen  A. Tareen
T.L. Forcier  M.S. Wong

Dr. Kinney's research career began in theoretical physics, but early in graduate school he was drawn to biology by the immense variety of open problems and by the possibility of testing theoretical ideas with simple experiments. At first, he pursued dry laboratory research focused on machine learning methods for analyzing large but noisy biological data sets. Then, in his last year of graduate school, he became captivated by the possibility of using ultra-high-throughput DNA sequencing to quantitatively study the biophysical mechanisms of gene regulation. To pursue this vision, he proposed and carried out wet laboratory experiments that culminated in Sort-Seq, the first massively parallel reporter assay (MPRA) for studies in living cells. As an independent investigator, he has continued pursuing a tightly knit combination of experiment, computation, and theory focused on using MPRAs to quantitatively study sequence–function relationships and to decipher their underlying mechanisms.

## Biophysical Models of *cis*-Regulation as Interpretable Neural Networks

The adoption of deep learning techniques in genomics has been hindered by the difficulty of mechanistically interpreting the models that these techniques produce. In recent years, a variety of post hoc attribution methods have been proposed for addressing this neural network interpretability problem in the context of gene regulation. In this work (Tareen and Kinney 2019b), we describe a complementary way of approaching this problem. Our strategy is based on the observation that two large classes of biophysical models of *cis*-regulatory mechanisms can be expressed as deep neural networks in which nodes and weights have explicit physiochemical interpretations. We

demonstrate how such biophysical networks can be rapidly inferred, using modern deep learning frameworks, from the data produced by certain types of MPRAs. These results, which are briefly illustrated in Figure 1, suggest a scalable strategy for using MPRAs to systematically characterize the biophysical basis of gene regulation in a wide range of biological contexts. They also highlight gene regulation as a promising venue for the development of scientifically interpretable approaches to deep learning.

## Logomaker: Beautiful Sequence Logos in Python

Sequence logos are visually compelling ways of illustrating the biological properties of DNA, RNA, and protein sequences. However, it has been difficult to generate and customize such logos within the Python programming environment. To address this need, our laboratory developed a software package called Logomaker (Tareen and Kinney 2019a). Logomaker is a flexible Python API capable of creating publication-quality sequence logos. Logomaker can produce both standard and highly customized logos from either a matrix-like array of numbers or a multiple-sequence alignment (see Fig. 2 for some examples). Logos are rendered as native matplotlib objects that are easy to stylize and incorporate into multipanel figures. Logomaker thus fills a major need in the Python community for flexible logo-generating software. Indeed, Logomaker has already been used to generate logos for multiple preprints and publications. Logomaker is thoroughly tested, has minimal dependencies, and can be installed from PyPI by executing "pip install logomaker" at the command line. A step-by-step tutorial on how to use Logomaker, as well as comprehensive documentation, are available at http://logomaker.readthedocs.io.
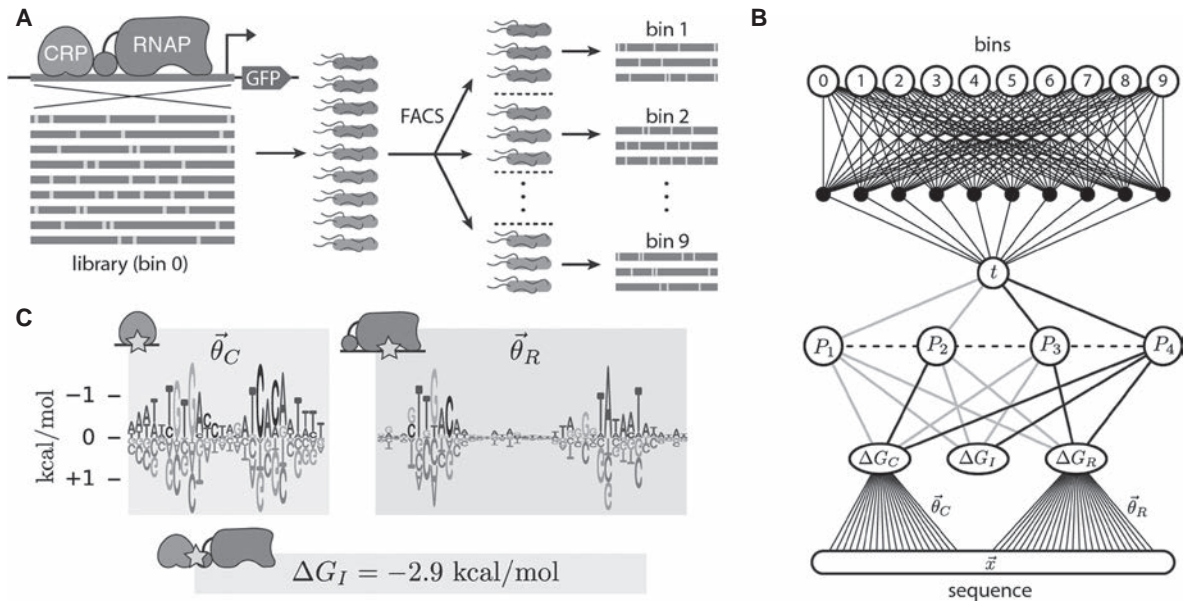
**Figure 1.** Biophysical models of *cis*-regulation as interpretable neural networks (Tareen and Kinney 2019b). (*A*) Illustration of a massively parallel reporter assay (MPRA) previously performed by Kinney and colleagues on the *Escherichia coli lac* promoter. (*B*) A biophysical model of *lac* promoter activity as a function of promoter DNA sequence, formulated as a deep neural network. (*C*) The values of biophysically meaningful parameters were inferred by fitting this neural network to the massively parallel reporter assay (MPRA) data from *A*.
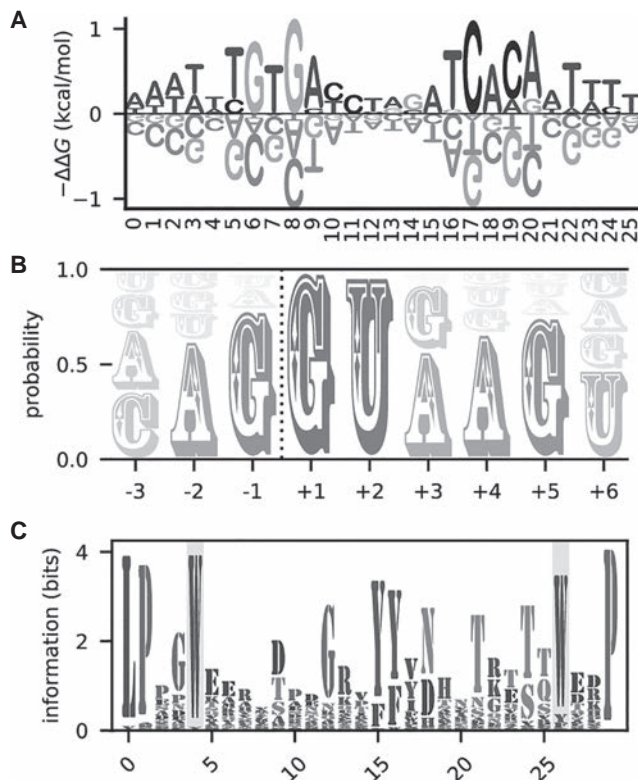


**Figure 2.** Logomaker: beautiful sequence logos in Python (Tareen and Kinney 2019a). Shown are three examples of sequence logos created using Logomaker. (*A*) An energy logo for the transcription factor CRP. (*B*) A probability logo representing the composition of 5′ splice sites in the human genome. (*C*) An information logo representing a multiple sequence alignment of WW domains.

## PUBLICATIONS

Adams RM, Kinney JB, Walczak AM, Mora T. 2019. Epistasis in a fitness landscape defined by antibody-antigen binding free energy. *Cell Syst* **8:** 1–8. doi:10.1016/j.cels.2018.12.004

Barnes SL, Belliveau NM, Ireland WT, Kinney JB, Phillips R. 2019. Mapping DNA sequence to transcription factor binding energy in vivo. *PLoS Comp Biol* **15:** e1006226. doi:10.1371/journal.pcbi.1006226

Kinney JB, McCandlish DM. 2019. Massively parallel assays and quantitative sequence–function relationships. *Annu Rev Genom Hum Genet* **20:** 99–127. doi:10.1146/annurev-genom-083118-014845

Tareen A, Kinney JB. 2019a. Logomaker: beautiful sequence logos in Python. *Bioinformatics* **36:** 2272–2274. doi:10.1093/bioinformatics/btz921

Tareen A, Kinney JB. 2019b. Biophysical models of *cis*-regulation as interpretable neural networks. bioRxiv doi:10.1101/835942

*In Press*

Ireland W, Beeler S, Flores-Bautista E, Belliveau N, Sweredoski M, Moradian A, Kinney JB, Phillips R. 2020. Deciphering the regulatory genome of *Escherichia coli*, one hundred promoters at a time. bioRxiv doi:10.1101/2020.01.18.910323

# COMPUTATIONAL GENOMICS OF CANCER

**A. Krasnitz**    P. Belleau    N. Ranade

The bulk of our research belongs to the field of computational cancer biology. Our choice of research goals within this field is dictated by (1) the potential impact of our work on clinical research and practice at present and in the foreseeable future; (2) the need to maximize the utility of emerging molecular technologies and research platforms in cancer biology; and (3) the opportunity to bring to bear quantitative techniques developed in other areas of science such as computational physics, applied mathematics, and computer science. These goals include (1) examination of intratumor genomic heterogeneity, its origin in cancer evolution, and its predictive value for aggressive and invasive potential of cancer; (2) reducing the complexity of genomic data for better interpretability while retaining their biological content; (3) derivation of clinically relevant molecular subtypes of the disease; and (4) design of predictive models for response to pharmacological interventions. As pursuit of these goals often reveals the inadequacy of existing, and necessitates the development of novel, computational tools, toolmaking is an important component of our activity.

## Integrated Computational Pipeline for Single-Cell Genomics

Investigation of single-cell genomes and transcriptomes is the focus of massive research effort worldwide. In application to cancer, this line of research has revealed the genomic complexity of the disease and the presence of multiple genealogically related cell populations in a tumor. Detailed knowledge of the clonal structure of a cancer potentially is of high clinical value: Multiplicity of clones or of lesions in most advanced clones is a possible measure of progression; spatial pattern of clone dispersal in a tumor may signal elevated propensity to invade; and lesions observed in individual clones but not in the bulk tissue may point to targets for therapy. DNA copy number profiling of cells from sparse sequencing is an accurate, economically feasible technological approach to the study of

cancer subpopulation structure. Novel multiplex sequencing techniques, developed by the Wigler laboratory at CSHL among others, permit simultaneous sequencing of thousands of single-cell DNA specimens and their subsequent copy number profiling at high resolution. Optimal use of this data form for robust reconstruction of cancer cell phylogenies is a challenging computational problem calling for new and robust informatic and statistical tools.

We responded to this challenge by developing and publishing a computational pipeline for single-cell genomics. The pipeline, which has now been released for public use, accomplishes three major tasks: estimation of integer-valued copy number (CN) profiles of individual cells, starting from cell-specific genome sequencing read data; collective analysis of multiple single-cell CN profiles in order to infer clonal structure of the cell populations; and graphical rendition of the output, complemented by nongenomic elements such as histological slide images of the tissues from which the cells originate.

For the last task, we provide Single-Cell Genome Viewer (SCGV), a visualization interface for a collection of single-cell genomes, alongside the clonal structure it represents and additional, not necessarily genomic, relevant information. Some of the key functionalities of SCGV are illustrated in Figure 1, for a collection of cells harvested from a radical-prostatectomy specimen. The surgery was performed to treat prostate cancer, and the surgical specimen scored Gleason 9 on histopathological evaluation, indicating advanced disease. Cells for genomic analysis originate from a number of areas in the prostate, with varying degrees of involvement in cancer. In the opening screen, single-cell genomes are displayed as columns of a heatmap, with the chromosomes concatenated from 1 through Y. Copy number gains and losses are encoded in red and blue colors, respectively, with darker colors corresponding to greater deviations from copy number 2, a norm for diploid cells. The cells are arranged horizontally as leaves of a phylogenetic tree. The two tracks immediately above the heatmap are used to indicate clonal and subclonal

cell populations present in the data. Additional tracks are located below the heatmap and are used to display user-supplied cell-specific metadata. One of these is reserved to display the anatomic location of origin (sector) for each cell, if available. Other tracks are configurable by the user and can be used to display any categorical or numerical metadata associated with the cells. For example, a track may be used to indicate the cell type as determined by flow-cytometric analysis. Cells can be reordered by the values encoded in any of the tracks. A particularly useful reordering is by sector, making it easy to see the sectors in which clonal populations reside (Fig. 1B). A subset of the data, and a corresponding subtree, may be displayed for any chosen value in any of the tracks—in particular, for

any given sector (Fig. 1C). An additional function of SCGV facilitates display of images associated with the sector—in particular, those of hematoxylin and eosin (H&E)-stained tissue slides (Fig. 1D).

In designing the pipeline as described, we sought to ensure that data analysis, interpretation, and visualization do not become a rate-limiting factor in this development nor an obstacle to adoption of single-cell DNA technologies. We therefore provided a complete, extensively documented, and easily deployable solution for all steps in the analysis of sparse single-cell genomic data. Our tools were initially conceived to meet the needs of our in-house single-cell technology development and applications to cancer. They were later adapted for the use of a broader research community,
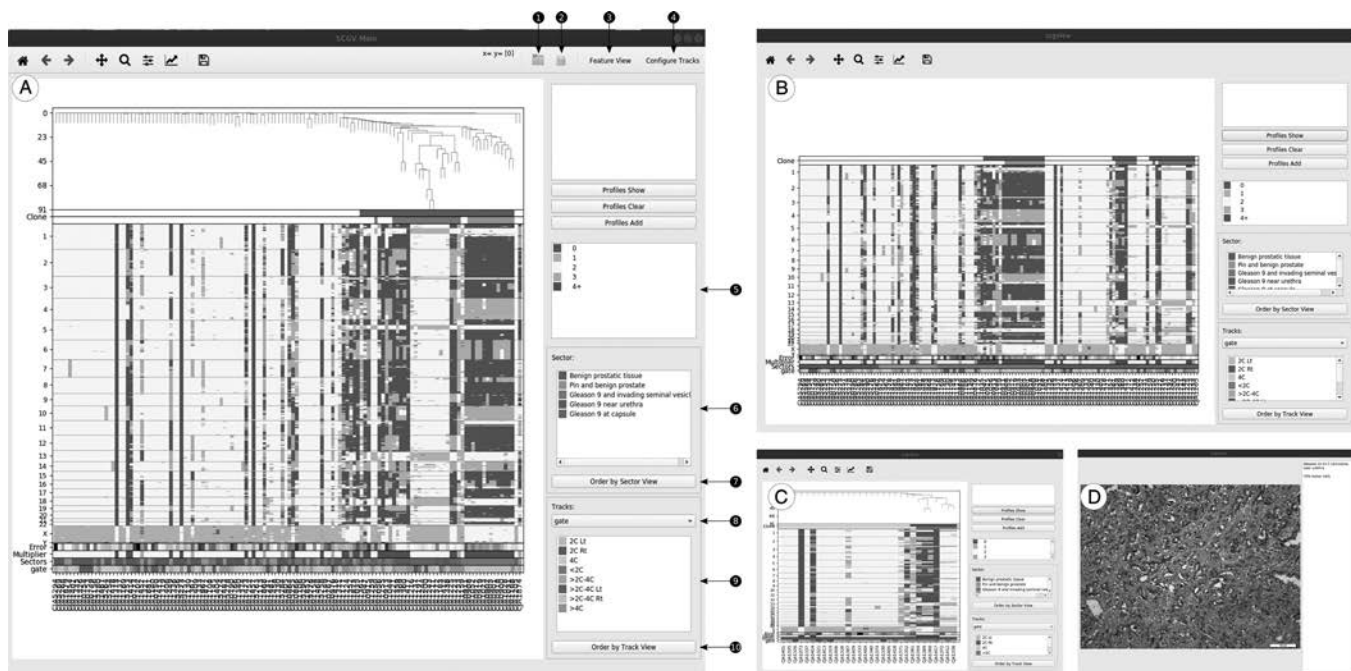


**Figure 1.** Functionalities of Single-Cell Genome Viewer (SCGV), illustrated for a case of 130 cells harvested from multiple locations of a surgical specimen from a radical prostatectomy. (*A*) Opening view of SCGV, with the genomic data and single-cell metadata loaded from (1) a directory or (2) a compressed archive. Copy number (CN) profiles of individual cells are shown as columns of a matrix, with genomic coordinates as rows. Copy number value is color-coded (side panel 9). The phylogenetic structure of the sample is visualized as a tree, shown in the upper portion of the view. The cells belonging to clones, and subclones are indicated in color in the two tracks immediately under the tree. Further tracks, in the lower portion of the view, display, for each cell, the quality-control measures of how close the copy number profile is to being integer-valued and the anatomic location (sector, color code shown in side panel 10) of origin for the cell. Additional tracks may be configured for the available single-cell metadata (4), such as the cell sorter gate for the cell, and the color code for the tracks displayed (12, 13). Cells may be reordered by sector (11), with the result shown in *B*, or by categorical values in any of the configurable tracks (14). Any sector from side panel 10 may be selected to display it separately, as shown in *C*. Annotated pathology slide from a sector may be displayed in a separate view, as shown in *D*.

including users of commercial single-cell genomics platforms such as those provided by 10x Genomics. Their subsequent evolution will be shaped, in large measure, by community feedback.

Novel single-cell molecular protocols such as BAGs developed by the Wigler laboratory at CSHL facilitate high-quality, affordable genomic and transcriptional profiling of many thousands of cells in a single experiment, with other modalities such as methylome and multi-omic profiling to follow soon. Both learning methods of computational biology and their software implementation must be adapted to cope with massive and heterogeneous data thereby produced. With this necessity in mind, we work to exploit richer information content and superior quality of the data from the new protocols for more complete and accurate inference. Two examples illustrate our research interest in this area. One is integration of heterogeneous molecular data, as necessary for understanding the relationship between clonal cell populations of the tumor and their associated stroma. We seek to determine which elements of microenvironment, as defined by their mRNA expression profiles, tend to localize with a given tumor clone, as specified by a shared pattern of SCNV. If the tumor clone in question metastasizes, we must examine which of its associated microenvironmental elements migrate with it to the metastatic site. As another example, BAG-seq samples from a cell genome at a predefined set of positions and provides a unique identifier for each DNA fragment it assesses. For us, these features are an opportunity to derive integer-valued single-cell copy number profiles with statistical rigor, by solving a Bayesian inference problem, rather than follow the current practice of estimating them heuristically. Work on this new inference algorithm is in progress. Our computational pipeline for single-cell genomics, especially its visual component, will evolve to keep up with the technology and algorithm development.

## Organoid Models of Pancreatic Cancer: Linking Genomics, Pharmacology, and Clinic

Pancreatic cancer is one of the least curable malignancies, often diagnosed at an advanced stage and with <30% of patients surviving the first year following diagnosis. Nearly all patients require pharmacological treatment, most commonly chemotherapy. Given the unfavorable prognosis, and the narrow time window it leaves for therapeutic intervention, it is vital that the pharmacological regimen be optimized individually for each patient, accounting for the molecular profile of the disease.

With the ultimate goal of improving patient stratification, we collaborated with the Tuveson laboratory at CSHL, which recently has achieved a major breakthrough in the ability to consistently generate realistic, three-dimensional in vitro models of pancreatic cancer, termed patient-derived organoids (PDOs). These cultures can be grown either from surgical specimens or from fine-needle tissue biopsies, and both from primary tumors and metastatic lesions of pancreatic ductal adenocarcinoma (PDAC). Once grown, PDOs provide a platform for massively parallel screening of pharmacological agents. The goal of our collaboration was to examine the correspondence between the molecular properties of the PDOs and their PDAC tissues of origin and to determine the relevance of drug responses in the PDOs to those observed in PDAC in clinical settings.

To this end, we performed extensive molecular profiling of the PDO cultures in the library, including whole-exome sequencing (WES), RNA-Seq, and deep DNA sequencing for a panel of 500 PDAC-associated genes. In order to examine how faithfully PDOs reflect the somatic mutation spectrum of the tumor of origin, we performed whole-genome sequencing (WGS) of the PDOs, the matching primary PDAC tissue from a surgical specimen, and matching normal tissue from the donor for 20 patient cases. We observed a high degree of concordance for all classes of somatic variants (substitutions, insertions, deletions) occurring at high variant allele frequencies (VAFs) in the PDOs. At the same time, the PDOs made it possible to discover variants at low VAFs when those variants were rendered undetectable by the low neoplastic cellularity in the primary tumors. A broader examination of all genomic data from the PDO library demonstrated the presence in the PDO genomes of mutations in the same driver genes, and at highly similar overall frequencies, as has been previously observed in PDAC patient cohorts. An important example is that of *KRAS*, which was found mutated in 96% of PDOs in the library, highly similar to the rate of >90% found in PDAC patient data.

We further determined that PDO cultures recapitulate molecular phenotypes of PDAC as found in the human hosts. In particular, most PDOs in our library can be assigned, with a high degree of confidence,

either to the previously described classical or basal molecular subtypes based on their mRNA expression profiles. Accordingly, unsupervised clustering of PDO cultures by mRNA expression robustly produces a partition into two classes, one of which is dominated by the classical, and the other by the basal profiles. Thus, PDO cultures retain phenotypic diversity of the disease in the human hosts—unlike PDAC cell lines, which overwhelmingly are basal.

To better understand how organoid sensitivities to chemotherapeutic agents correspond to patient sensitivities, we conducted a screen of PDO responses to five standard-of-care cytotoxic agents for PDAC, using the area under the drug response curve (AUC) as a measure of response. Next, we addressed the questions of (1) whether transcriptional profiles of PDOs are predictive of the therapeutic responses observed in this screen and (2) whether these transcriptional signatures of response are applicable to PDAC patient outcomes. To this end, we identified, separately for each of the five agents, a small subset of genes whose mRNA expression most strongly correlated with the AUC values for the agent in the PDO library. For each PDO, we then summed the $z$-scores of the genes in the drug-specific subset, to form a single predictive molecular sensitivity score (MSS) for each PDO's sensitivity to each agent. We found these predictive scores to be strongly correlated with the corresponding AUC values.

To see how well our PDO-derived MSSs would apply to human patients, we used the gemcitabine-sensitivity genes defined in the PDOs to compute MSSs from 95 patients who received adjuvant treatment containing gemcitabine, one of the five agents in the PDO therapeutic screen. For this analysis, we used RNA-Seq data from resected PDAC specimens. The 50% of the patients with the highest MSS values experienced significantly longer median PFS than the low-scoring 50% of the cohort (608 vs. 442 days, Cox regression $p = 0.046$). Moreover, the MSS values were not associated with survival benefit in the 38 patients who received no adjuvant treatment, demonstrating that the score is predictive only in the setting of the chemotherapy used to calculate it.

Finally, we examined the MSS performance using molecular and clinical data from the COMPASS trial, for which patients with advanced, unresectable PDAC were recruited, and each was assigned a chemotherapy regimen (m-FOLFIRINOX or gemcitabine-paclitaxel).

We focused on the 41 patients on the trial who received FOLFIRINOX. We calculated MSS values for oxaliplatin, a component of FOLFIRINOX, for each patient. Patient MSS values were then compared with the corresponding tumor responses to treatment as measured by changes in the tumor volume. Patients whose MSS predicted they would be sensitive to oxaliplatin exhibited significantly better response to FOLFIRINOX than those deemed nonsensitive ($r = -0.4$, $p = 0.008$). In summary, our results indicate the prognostic utility of PDO-derived MSS for cytotoxic treatment outcomes in patients with advanced, as well as with localized, PDAC.

In our ongoing study of PDAC genomics and pharmacology we seek to (1) achieve deeper understanding of PDAC genomics by exploiting nearly 100% cellularity of the PDOs; (2) establish correspondence between molecular classes of PDAC in PDOs and in patients; (3) increase the pool of potential molecular predictors of sensitivity to pharmacological treatments; (4) design better predictive models for drug sensitivity; and (5) design simple and affordable clinical assays predictive of patient response to therapy.

In keeping with these goals, we recently broadened our characterization of the PDO library to include SCNV analysis. To compare the SCNV spectrum of PDAC in PDOs and in PDAC specimens directly from patients, we collaborated with the COMPASS clinical trial team to obtain PDAC whole-genome sequence data from diagnostic needle biopsies. CORE analysis of the combined data set revealed an SCNV landscape dominated by features present in profiles from both sources, confirming that PDOs are a faithful model of the disease.

In parallel, to boost the predictive power of PDO-based models for drug sensitivity, we moved on from the simple scoring approach in our published work to systematic model building by machine learning, using random forest (RF) regression for mRNA expression data in PDOs. Our results point to strong (on the order of Spearman = 0.6) correlations between the RF-predicted and observed values of the AUC, for all major components of the standard-of-care cytotoxic treatments for PDAC. Importantly, we discovered that a small (on the order of 25 per cytotoxic agent) set of genes whose expression levels underlie our RF models can with equal success predict patient response to therapies containing these agents. Based on these encouraging results, we are currently designing a 100-gene nanostring assay, whose

ability to predict anti-PDAC drug sensitivity will be tested in an ongoing clinical study.

## PUBLICATIONS

Sun G, Krasnitz A. 2019. Statistically supported identification of tumor subtypes. *Methods Mol Biol* **1878:** 209–216. doi:10.1007/978-1-4939-8868-6_12

Sun G, Krasnitz A. 2019. CORE: a software tool for delineating regions of recurrent DNA copy number alteration in cancer. *Methods Mol Biol* **1878:** 85–93. doi:10.1007/978-1-4939-8868-6_4

Krasnitz A, ed. 2019. *Cancer bioinformatics*, 1st ed. Humana, New York.

*In Press*

Chorbadjiev L, Kendall J, Alexander J, Zhygulin V, Song J, Wigler M, Krasnitz A. 2020. An integrated computational pipeline for single-cell genomic profiling. *JCO Clin Cancer Inform* **4:** 464–471. doi:10.1200/CCI.19.00

Moffitt A, Spector M, Andrews P, Kendall J, Alexander J, Stepansky A, Ma B, Kolitz J, Chiorazzi N, Allen S, et al. 2020. Multiplex accurate sensitive quantitation (MASQ) with application to minimal residual disease in acute myeloid leukemia. *Nucleic Acid Res* **48:** e40. doi:10.1093/nar/gkaa090

Li S, Kendall J, Park S, Wang Z, Alexander J, Moffitt A, Ranade N, Danyko C, Gegenhuber B, Fischer S, et al. 2020. Copolymerization of single-cell nucleic acids into balls of acrylamide gel. *Genome Res* **30:** 49–61. doi:10.1101/gr.253047.119

# COMPUTATIONAL GENETICS

**D. Levy**   A. Moffitt

The data-rich environment at CSHL generates a wealth of opportunities for the application of mathematics and computation to further our understanding of biology and genetics. The primary activities of our laboratory are algorithm and protocol development, data analysis, and genetic theory, with a focus on sporadic human diseases such as autism, congenital heart disease, and cancer.

## Quantitative-Sensitive Detection

Measuring genomic variants at very low frequency is important in many applications, but especially relevant in measuring residual disease in cancer. Present-day short-read sequencers generate hundreds of millions of high-quality sequence reads with error rates <1 per 100 bases. These errors determine the lower limits of variant detection. Using standard sequencing, it is impossible to distinguish a variant frequency <1% from machine error. We would like to measure variants at frequencies of 0.0001%, or one part in a million.

Together with Michael Wigler and Zihua Wang, we developed a protocol that adds a unique sequence identifier (or "varietal tag") to the initial template molecule. Because sequence error is sparse and independent of the template molecule, reads with the same varietal tag are unlikely to have the same sequence errors. Taking a sequence consensus from all reads with the same tag corrects for sparse error. We found that systematic errors have a sequence context-specific signature and by analyzing multiple high-depth tagged experiments, we are able to model the error of consensus reads. Some sequence contexts (such as CpG) have a high background rate that has nothing to do with machine error. However, other contexts are very stable, with background error rates for consensus sequences lower than one part in a million. The MASQ (multiplex accurate sensitive quantitation) protocol (Moffitt et al. 2020) enables the simultaneous measurement of variants from up to 50 loci and hundreds

of thousands of templates per locus per reaction. We developed an informatics pipeline that selects variants with low background error rates, determines the optimal reagents for the protocol, and designs the necessary sequencing primers. The MASQ protocol has a specificity and sensitivity beyond any technology presently available for measuring residual disease.

With Mona Spector, Alex Krasnitz, and Joan Alexander, we applied MASQ to measure residual disease in patients treated for acute myeloid leukemia (AML). Using MASQ, we measured tumor-specific variants in blood samples taken after treatment. In some patients, the residual load was detectable at levels observable by traditional sequencing. In other patients, we were able to measure rates as low as one part in 100,000. Our study determined that the frequencies of tumor variants are equivalent in peripheral blood and bone marrow.

We are presently testing the utility of MASQ in the context of solid tumors, in which circulating tumor cells and cell-free tumor DNA provide targets for detection. Circulating tumor cells occur in blood at very low counts relative to the background blood cells, about one part in a billion. To fall within the range of detection for MASQ requires a 1,000-fold enrichment of tumor cells over the background blood elements. To achieve this aim, we are using strategies based on enrichment for epithelial cells, depletion of normal blood elements, or some combination of the two.

The second blood compartment that may harbor a signal from solid tumors is cell-free DNA in the plasma. Cell-free DNA is typically short in length, which requires adapting the MASQ protocol. We are testing new tagging approaches that eliminate the need for restriction sites. The other challenge of cell-free DNA is that there are typically few templates from either the tumor or the background. To still obtain a sensitivity of one part in a million when there are only a few thousand DNA copies per position requires some combination of (1) assaying more positions and (2) increasing the number of reactions. At present, 50 loci is not a hard limit for MASQ and it may scale into the hundreds. Obtaining more

reactions requires either more plasma or another source of cell-free DNA, such as urine.

## Mutational Sequencing (muSeq)

The latest third-generation sequencing platforms, like PacBio and Oxford Nanopore, generate long-read information that is important in high-quality genome assemblies. Long reads are especially useful when the genome studied is diploid and heterozygous variants are too far apart to phase by short reads. Unfortunately, compared with the present generation of short-read sequencers, these long-read platforms are expensive and error-prone.

We previously proposed a method for turning short-read sequencers into virtual long-read sequencers by embedding a unique molecular identity throughout each template molecule by random mutation. Theoretical computations suggested that we could then count templates by counting unique mutation patterns and that we could assemble very long templates by connecting reads with overlapping mutation patterns.

Recently, we implemented this idea in practice, using incomplete bisulfite conversion as the mechanism for introducing mutations. Loosely speaking, sodium bisulfite converts a C to T in a DNA template and, by tuning our rate of conversion to 50%, we label each template molecule with a unique and dense mutational signature of C-to-T conversions. Clustering reads with the same conversion pattern enabled accurate count and long-range assembly of initial template molecules from short-read sequence data. Using a PstI representation, we showed that muSeq improves copy number measurement and significantly reduces sporadic sequencing error. Using a cDNA library, we showed long-range assembly of template molecules up to 4 kb in length.

However, both counting and cDNA assembly required first mapping reads to a reference genome. This limits the utility of muSeq to well-sequencing organisms, and even in those cases, we would like to avoid reference bias when determining insertion/deletion polymorphism, splice junctions, and assembling complex genomic regions rich in polymorphism, such as the HLA locus.

For these reasons, we are liberating muSeq from the reference genome. Together with Siran Li, we developed a protocol and informatics for targeted and phased de novo muSeq assembly. At present, we can generate haplotype-level assemblies for target regions up to 5 kb in length. In place of mapping to a reference, the mutated template molecules are assembled using a de Bruijn graph. We developed custom assembly methods to augment this graph with read-pair information, resulting in many mutated template assemblies that span the full length of the target region. In lieu of a reference genome, we use an unconverted sequence library to "correct" mutations in the template assemblies. We then partition the corrected templates into two (or more) haplotypes. Finally, we compute a consensus over each haplotype, correcting any residual mutations and PCR errors.

The present protocol and informatics can assemble hundreds of mutated template molecules and generate high-quality haplotype-phased assemblies for regions up to 5 kb. The protocol is simple, robust, and, requiring fewer than a million Illumina reads, very cheap. Although changes in technology for long-read sequencers may alter the equation, at present de novo muSeq offers an accessible and inexpensive alternative for targeted applications.

Recently, we have developed an improved protocol that can target regions 10 kb and longer. The new method uses the random incorporation of methylcytosine into the templates followed by complete deamination of unmethylated bases by an enzymatic deaminase. The longer templates require a more robust informatics presently under development.

*In Press*

Moffitt AB, Spector MS, Andrews P, Kendall J, Alexander J, Stepansky A, Ma BC, Kolitz J, Chiorazzi N, Allen SL, et al. 2020. Multiplex accurate sensitive quantitation (MASQ) with application to minimal residual disease in acute myeloid leukemia. *Nucleic Acids Res* **48:** e40. doi:10.1093/nar/gkaa090

# PREDICTING EFFECTS OF MUTATIONS FROM HIGH-THROUGHPUT DATA

**D. McCandlish**    A. Posfai    J. Zhou

Understanding the relationship between the DNA sequence of an organism's genome and the measurable characteristics of that organism is one of the fundamental goals of biology. Recent progress in high-throughput experimental techniques now allows us to measure the effects on a cellular or molecular level of thousands to millions of changes to the DNA sequence in a single experiment. In the McCandlish laboratory, we are focused on developing new computational and mathematical techniques for making sense of this wealth of data. Our ultimate goals are to be able to predict the pathogenicity of mutations observed in human genome sequences, understand the evolution of drug resistance and immune escape, and help to construct highly optimized enzymes for biotechnology applications.

An important challenge in predicting the effects of mutations is that the effect of any given mutation may depend on which other mutations are already present, a phenomenon known as genetic interaction or epistasis. Our group is particularly interested in developing techniques to quantify and better understand the form and causes of these genetic interactions, with the dual goals of improving our ability to predict the effects of combinations of mutations and to understand the influence that these interactions have on the process of biological evolution.

The effort to develop a comprehensive quantitative understanding of the relationship between biological sequences and their degree of functionality within the organism is a field of inquiry emerging at the intersection of high-throughput genetics, machine learning, synthetic biology, medical genetics, and evolution. This makes an interdisciplinary approach essential, and the sustained interest in these questions from several research groups is a strength of the Quantitative Biology group at CSHL. Together with Justin Kinney, this past year, we published a synthetic review (Kinney and McCandlish 2019), which highlights both recent advances and ongoing challenges in this exciting new area.

## Inference of Genetic Interactions

Because of the large number of possible mutations to any given gene and the far larger number of possible combinations of these mutations, the form that genetic interactions can take can be extremely complicated. This past year, we introduced a completely new approach for how to model and understand complex genetic interactions, which we have applied to analyze experimental data from protein–protein and protein–DNA binding assays (Zhou and McCandlish 2020). We call this approach "minimum epistasis interpolation" because it seeks to model phenotypic observations in a manner that requires the smallest possible amount of epistasis without making assumptions about the form that this epistasis takes. This differs from previous models of epistasis, which typically assume that epistasis takes some specific form such as only allowing interactions between pairs of mutations or which depend on already having a mechanistic model for the system in question.

The basic setting for our method is that we are given phenotypic observations for some set of genotypes, and we want to make predictions of the phenotypes for genotypes that we have not yet observed. To do this, we infer the set of predictions that matches our observed data wherever available but which otherwise makes the phenotypic effects of mutations as consistent as possible across adjacent genetic backgrounds. The end result is a model that can provide a complicated fit in which data are abundant, but which behaves in a simple and well controlled manner in regions of sequence space in which data are sparse or absent. Moreover, the mathematical solution to this problem has many special properties that allow us to compute predictions even when the number of possible genotypes is astronomically large. Similarly, these mathematical properties also allow us to provide mathematical guarantees that our predictions behave sensibly even though the number of genotypes is too large to allow for manually assessing the reasonableness of our predictions.

Although our minimum epistasis interpolation method represents a substantial advance in our ability to model complex genetic interactions, in our ongoing work, postdoc Juannan Zhou has developed a further generalization that we call "empirical variance component regression." Intuitively, whereas minimum epistasis interpolation attempts to produce a highly conservative solution that minimizes the extent of the genetic interactions required to explain the data, empirical variance component regression attempts to make predictions that match the type and extent of epistasis present in the original observations. It thus has the potential to provide more accurate predictions at the cost that we can provide fewer mathematical guarantees for model behavior.

To test the phenotypic predictions from this new modeling framework, we have been collaborating with Justin Kinney, Adrian Krainer, and their joint postdoc Mandy Wong to apply these methods to data from a high-throughput splicing assay that they recently developed. We have now successfully validated the predictions of our model with low-throughput experiments and are investigating the biophysical basis of certain genetic interactions inferred by our method.

Our empirical variance component regression method formally falls within a broad class of statistical models known as Gaussian process models. Whereas historically these models have been limited for computational reasons to analyzing data sets containing, at the most, a few thousand observations, by exploiting certain mathematical features of biological sequence space, we have been applying our method to much larger data sets containing hundreds of thousands of observations. Together with Wei-Chia Chen from Justin Kinney's group, we have been working to develop practical open-source software to allow widespread implementation of these new methods. Because Gaussian process models have also received renewed attention in the neuroscience community for modeling neural recording data, David McCandlish and Tatiana Engel taught an advanced course on Gaussian processes in fall 2019 titled "Gaussian Processes for Biological Data Analysis," with participants from both quantitative biology and neuroscience.

Although the above methods are designed to model complex genetic interactions that occur between specific combinations of mutations, we have also been developing methods to model another common form of genetic interaction known as nonspecific epistasis.

Genetic interactions of this form arise commonly in biology because even if mutations have independent effects at the molecular level, genetic interactions can appear when we look at the cellular or organismal level because of the nonlinear mapping between molecular traits and larger-scale organismal traits that is induced by the regulatory architecture, cellular physiology, developmental processes, etc. Previously, the McCandlish laboratory introduced a particular class of models for analyzing nonspecific epistasis, which we call "global epistasis" models. In an ongoing collaboration with Ammar Tareen, a postdoc in the Kinney laboratory, we have been working to better characterize the relationship between global epistasis models and several other models that have been proposed to capture nonspecific epistasis by (1) developing mathematical results to show the formal relationships between these models and (2) implementing this broader class of models within a unified neural network framework to allow practical performance comparisons.

Finally, postdoc Anna Posfai (joint with the Justin Kinney laboratory), who is a mathematician by training, has been working on techniques to allow better interpretation of the results of the above modeling efforts. In particular, it is frequently the case in models fit to data from high-throughput experiments that many different sets of model parameters will give identical predictions. She is developing mathematical techniques to resolve this problem to unambiguously identify the specific positions in a genetic sequence that have the largest influence and/or strongest interactions.

## Probabilistic Models of Genetic Sequence Diversity

Often in computational biology we are presented with a collection of sequences that have some known function, such as homologous protein sequences from different organisms or a collection of known binding sites for a transcription factor of interest. Because we know that many sequences that we have not yet observed are likely to have functionality comparable to these observed sequences, to be able to better identify and understand this set of possible functional sequences, we often want to estimate the probability distribution from which these observed sequences are drawn. In a new collaboration with Justin Kinney and Kinney laboratory postdoc Wei-Chia Chen, we have been

working to address this problem by combining techniques in Bayesian nonparametric density estimation developed in the Kinney laboratory with the graph-theoretic perspective on biological sequence space that underlies much of the approach in the McCandlish laboratory. This past year, we developed and implemented this new technique for flexible Bayesian estimation of complex distributions over sequence space. As a first application, we are now using these techniques to analyze the collection of observed human 5′ splice sites.

## Influence of Mutational Biases on Molecular Adaptation

For many applications such as predicting immune escape and drug resistance mutations, we are interested not only in which mutations can potentially confer these phenotypic changes, but also which mutations are most likely to contribute to the emergence of resistance in a clinical or natural setting. In this context, an important fact is that different mutations occur at substantially different rates. For instance, in many genomes, including those of mammals and birds, mutations occur at highly elevated rates at sites where a cytosine nucleotide is found immediately 5′ to a guanine nucleotide, known as CpG sites. One might hypothesize that if a mutation at such a site is beneficial, it might be particularly likely to contribute to adaptation because of its high rate of occurrence. In a long-standing collaboration with Arlin Stoltzfus at the National Institute of Standards and Technology, we have been working to document the prevalence of

this influence of mutational biases on the outcome of molecular adaptation in nature. This year we collaborated with Jay Storz at the University of Nebraska to show that mutations at CpG sites are highly enriched among the set of mutations that confer increased oxygen affinity to hemoglobin proteins of high-altitude birds, which is a well studied system for understanding molecular adaptation. We also began a new collaboration with Joshua Payne at ETH Zurich to incorporate more detailed, species-specific mutational patterns into our approach, by asking whether the spectrum of observed adaptive substitutions in several different species mirrors differences in their mutational spectrum as measured through mutation accumulation experiments. If the idiosyncratic features of species-specific mutational spectra are reflected in collections of known adaptive substitutions from those species, this will provide additional evidence that mutational biases are important in molecular adaptation beyond our current observation that collections of adaptive substitutions reflect widespread mutational biases.

### PUBLICATIONS

Kinney JB, McCandlish DM. 2019. Massively parallel assays and quantitative sequence–function relationships. *Ann Rev Genomics Hum Gene* **20:** 99–127. doi:10.1146/annurev-genom-083118-014845

Storz JF, Natarajan C, Witt CC, McCandlish DM, Stoltzfus A. 2019. The role of mutation bias in adaptive molecular evolution: insights from convergent changes in protein function. *Phil Trans R Soc Lond B Biol Sci* **374:** 20180238. doi:10.1098/rstb.2018.0238

*In Press*

Zhou J, McCandlish DM. 2020. Minimum epistasis interpolation for sequence–function relationships. *Nat Commun* **11:** 1782. doi:10.1038/s41467-020-15512-5

# POPULATION GENETICS AND TRANSCRIPTIONAL REGULATION

**A. Siepel**  A. Blumberg  Y. Huang  Z. Mo  A. Scheben
         N. Dukler  M. Hubisz  A. Purugganan  A. Xue
         H. Hejase  E. Hutton  R. Ramani  Y. Zhao

For the past several years, our research has focused on two major areas: human population genetics and transcriptional regulation. The research in population genetics is performed either with publicly available genomic sequence data or with a variety of collaborators, whereas most of the work on transcriptional regulation is performed with our collaborators Charles Danko and John Lis at Cornell. We also have smaller collaborative projects on topics ranging from prediction of the fitness consequences of mutations in rice, maize, and other crops (with Michael Purugganan, New York University, and Ed Buckler, Cornell University), to statistical analysis of CRISPR screens (with Chris Vakoc, CSHL), to studying the speciation process of recently diverged *Sporophila* songbirds (with John "Irby" Lovett, Cornell, and Ilan Gronau, Herzliya Interdisciplinary Center, Israel). We focus on theoretical and computational research and do not generate our own data, but we often work closely with experimental collaborators on projects that have substantial experimental as well as computational components. We are broadly interested in molecular evolution, population genetics, and gene regulation, as well as in machine learning, probabilistic modeling, and Bayesian statistics, and our research projects cut a broad swath across these diverse areas. Our research group is highly interdisciplinary, with members trained in computer science, mathematics, physics, genetics, and biochemistry, among other areas. The size of the group is stable at present, with a Ph.D. student (Ziyi Mo) and a new postdoctoral associate (Armin Scheben) having recently joined.

Below we describe recent progress in three main research areas.

## Reconstruction of Demographic History from Complete Genome Sequences

Our research group has a long-standing interest in reconstructing the demographic history of complex, structured populations from DNA sequence data. We developed the first method for inference of human population sizes, divergence times, and gene flow between populations that explicitly models the genealogical relationships among individuals and is efficient enough for genome-wide use. We used this method, *G-PhoCS,* to estimate the origin of one of the earliest branching extant human populations, the San hunter–gatherers of Southern Africa, and more recently in analyses on dogs, wild canids, and six species of *Sporophila* birds. Another method developed by our group, *ARGweaver,* generalizes *G-PhoCS* by capturing the manner in which recombination alters genealogies along the genome sequence. In a joint analysis, we applied *G-PhoCS* and *ARGweaver* to detect significant evidence of gene flow from modern humans into the Altai Neanderthal genome sequence, in the opposite direction and much earlier than previously reported. In addition, we used *ARGweaver* to date this human-to-Neanderthal introgression event at ~100,000 years ago, suggesting an earlier migration of modern humans out of Africa than indicated by most current estimates.

Our work on Neanderthal–human interbreeding suggests that a major advantage of our *ARGweaver* method for ancestral recombination graph (ARG) inference is that it is especially powerful for identifying very early introgression events. However, the *ARGweaver* method is limited in that it naïvely assumes a prior distribution based on a single randomly mating population of constant size. Melissa Hubisz in the group recently introduced *ARGweaver-D*, an extension of the *ARGweaver* algorithm that makes use of a user-defined demographic model, including population divergence times, population sizes, and migration events (Hubisz et al. 2019). Like *ARGweaver*, *ARGweaver-D* is a Bayesian method that samples trees from the posterior distribution to account for the uncertainty. Given genome sequence data from a collection of individuals across multiple closely related populations or subspecies, *ARGweaver-D* can infer trees describing the genetic relationships between

these individuals at every location along the genome, conditional on the demographic model. We used this method to show that ~3% of Neanderthal DNA—and possibly as much as 6%—came from modern humans who mated with Neanderthals more than 200,000 years ago. We also predicted that 1% of the Denisovan genome was introgressed from an unsequenced, but highly diverged, archaic hominin ancestor. About 15% of these "super-archaic" regions—comprising at least ~4 Mb—were, in turn, introgressed into modern humans and continue to exist in the genomes of people alive today.

In a collaborative study led by Hussein Hejase, we have also applied *ARGweaver* to identify introgression and selective sweeps in birds (Hejase et al. 2020a). We examined genomic "islands" of elevated differentiation through the lens of recently obtained genome sequence data for five species of southern capuchino seedeaters, finch-like birds from South America that have undergone a species radiation during the last approximately 50,000 generations. By applying newly developed statistical methods for ancestral recombination graph inference and machine-learning methods for the prediction of selective sweeps, we showed that the striking islands of differentiation in these birds appear to be generally associated with relatively recent, species-specific selective sweeps, most of which are predicted to be "soft" sweeps acting on standing genetic variation. Many of these sweeps coincide with genes associated with melanin-based variation in plumage, suggesting a prominent role for sexual selection. At the same time, a few loci also show indications of possible selection against gene flow. These observations shed new light on the complex manner in which natural selection shapes genome sequences during speciation.

## Analysis of Natural Selection on Regulatory Sequences in the Human Genome

We also have a long-standing interest in characterizing the influence of natural selection on DNA sequences, particularly in noncoding regions of the genome. A few years ago, we developed a probabilistic model and inference method, called INSIGHT, which makes use of joint patterns of divergence and polymorphism to shed light on recent natural selection. We have used INSIGHT to show that natural selection has profoundly influenced transcription factor binding sites

across the genome during the past five million years of evolution, with major contributions both to adaptive changes in humans and to weakly deleterious variants currently segregating in human populations. Afterward, we realized that the INSIGHT model could also be used to produce "fitness consequences" (*fitCons)* scores across the entire human genome. Using high-throughput data from the ENCODE project, we partitioned the genome into classes of sites having characteristic functional genomic "fingerprints" in a given cell type and then used INSIGHT to calculate a *fitCons* score for each fingerprint (Gulko et al., *Nat Genet* **47**: 276 [2015]). Finally, we plotted these scores along the genome sequence. These *fitCons* scores turn out to be remarkably powerful for identifying unannotated regulatory elements in the human genome. We later developed an alternative approach that bypasses the need for clustering genomic sites and instead assumes a linear-logistic relationship between genomic features and the parameters of the INSIGHT model (Huang et al., *Nat Genet* **49**: 618 [2017]). This method, called LINSIGHT, is extremely fast and scalable, enabling it to exploit the "Big Data" available in modern genomics.

More recently, we devised a powerful alternative approach to the fitCons clustering problem, which both scales well and avoids the linearity assumptions of LINSIGHT. This new algorithm, called *fitCons2*, builds a decision tree by repeatedly splitting classes of genomic sites in a manner that is guaranteed to increase a global measure of the 'information' associated with natural selection. This approach allows us to consider dozens of genomic features both individually and in complex combinations. We have now applied *fitCons2* to all the data from Roadmap Epigenomics, considering nine epigenomics features across 115 cell types. A side benefit of this approach is that it allows us to measure the genome-wide "information" about function associated with these epigenomic features (Gulko and Siepel 2019). We found that several epigenomic features yield more information in combination than they do individually. In addition, we found that the entropy in human genetic variation predominantly reflects a balance between mutation and neutral drift. Our cell type–specific *fitCons2* scores reveal relationships among cell types and suggest that ~8% of nucleotide sites are constrained by natural selection.

In a study led by Adrian Platts, we recently collaborated with Michael Purugganan's laboratory

at New York University to produce the first fitCons map for a plant, in rice (*Oryza sativa*) (Joly-Lopes et al. 2020). We inferred fitCons scores (ρ) for 246 inferred genome classes derived from nine functional genomic and epigenomic data sets, including chromatin accessibility, messenger RNA/small RNA transcription, DNA methylation, histone modifications, and engaged RNA polymerase activity. These were integrated with genome-wide polymorphism and divergence data from 1,477 rice accessions and 11 reference genome sequences in the Oryzeae. We found ρ to be multimodal, with ~9% of the rice genome falling into classes in which more than half of the bases would probably have a fitness consequence if mutated. Around 2% of the rice genome showed evidence of weak negative selection, frequently at candidate regulatory sites, including a novel set of 1,000 potentially active enhancer elements. This fitCons map provides perspective on the evolutionary forces associated with genome diversity, aids in genome annotation and can guide crop breeding programs.

In addition, Yi-fei Huang in the group has been developing a deep-learning method that both allows for arbitrarily complex relationships among genomic features and makes use of population genetic theory to estimate allele-specific selection coefficients at every nucleotide in the human genome. This approach, called "linear allele-specific selection inference" (LASSIE), unifies methods for deleterious variant prediction with methods for inferring distributions of fitness effects (Huang and Siepel 2019). We applied LASSIE to 51 high-coverage genome sequences annotated with 33 genomic features and constructed a map of allele-specific selection coefficients across protein-coding sequences in the human genome (Huang and Siepel, bioRxiv doi:10.1101/441337 [2018]). This map is generally consistent with previous inferences of bulk distribution of fitness effects, but reveals pervasive weak negative selection against synonymous mutations. In addition, the estimated selection coefficients are highly predictive of inherited pathogenic variants and cancer-driver mutations, outperforming state-of-the-art variant prioritization methods. By constraining our estimated model with ultra-high-coverage ExAC exome-sequencing data, we identified 1,118 genes under unusually strong negative selection, which tend to be exclusively expressed in the central nervous system or associated with autism spectrum disorder, as well as 773 genes under unusually weak

selection, which tend to be associated with metabolism. This combination of classical population genetic theory with modern machine-learning and large-scale genomic data is a powerful paradigm for the study of both human evolution and disease.

Finally, Elizabeth Hutton in the group recently introduced a new probabilistic model and inference framework, analysis of CRISPR-based essentiality (ACE), designed to test for differential signatures of essentiality between cell lines (Hutton and Siepel 2019). High-throughput knockout screens using CRISPR-Cas9 are now a widespread method for evaluating the essentiality of genes in different cell types. ACE estimates the essentiality of each gene using a flexible likelihood framework based on the CRISPR-Cas9 experimental process and the observed sequencing counts. In addition, it can identify which genes are essential only in a specified subset of samples by directly contrasting the likelihood of competing hypotheses—whether a gene has a constant or differential essentiality between samples. We showed, using simulations, that this approach improves the accuracy of essentiality predictions compared with other methods, and is especially useful for the identification of weaker signals of essentiality. ACE performance was further validated on publicly available CRISPR screen data to distinguish between essential and nonessential genes. Notably, we were able to identify otherwise overlooked candidates for genotype-specific essentiality. Overall, ACE provides an improved quantification of essentiality specific to cancer subtypes, and a robust probabilistic framework to identify genes of interest.

## Transcriptional Regulation and Its Evolution in Primates

For several years, our research program in transcriptional regulation has focused on developing new methods for interpreting the rich nascent RNA sequencing data generated using the powerful GRO-seq (global run-on and sequencing) protocol or its higher-resolution successor, PRO-seq. These methods isolate and sequence newly transcribed RNAs, revealing genome-wide locations of engaged polymerases. It has gradually become clear that an unanticipated benefit of both GRO-seq and PRO-seq is that they are uniquely well suited for detecting so-called enhancer RNAs (or eRNAs), and consequently, for identifying

active enhancers and other regulatory elements in mammalian cells.

In our latest work in this area, Amit Blumberg has developed a method for estimating relative RNA half-lives based on PRO-seq and RNA-Seq. The rate at which RNA molecules are degraded is a key determinant of cellular RNA concentrations, yet approaches for measuring RNA half-lives are generally labor-intensive, limited in sensitivity, and/or disruptive to normal cellular processes. This method treats PRO-seq as a measure of transcription rate and RNA-Seq as a measure of RNA concentration, and estimates the rate of RNA degradation required for steady state equilibrium (Blumberg et al. 2019). We show that this approach can be used to assay relative RNA half-lives genome-wide, with reasonable accuracy and good sensitivity for both coding and noncoding transcription units.

Noah Dukler in the group has developed a more formal modeling approach to address the problem of identifying gains and losses of regulatory elements across an entire genome. Evolutionary changes in gene expression are often driven by gains and losses of *cis*-regulatory elements (CREs). The dynamics of CRE evolution can be examined using multi-species epigenomic data, but so far, such analyses have generally been descriptive and model-free. Noah has implemented a probabilistic framework for the evolution of CREs that operates directly on raw chromatin immunoprecipitation and sequencing (ChIP-seq) data and fully considers the phylogenetic relationships among species (Dukler et al. 2020). This framework includes a phylogenetic hidden Markov model, called epiPhyloHMM, for identifying the locations of multiple aligned regulatory elements, and a combined phylogenetic and generalized linear model, called phyloGLM, for accounting for the influence of a rich set of genomic features in describing their evolutionary dynamics. Noah applied these methods to previously published ChIP-seq data for the H3K4me3 and H3K27ac histone modifications in liver tissue from nine mammals. Noah found that enhancers are gained and lost during mammalian evolution at

about twice the rate of promoters, and that turnover rates are negatively correlated with DNA sequence conservation, expression level, and tissue breadth, and positively correlated with distance from the transcription start site, consistent with previous findings. In addition, Noah found that the predicted dosage sensitivity of target genes positively correlates with DNA sequence constraint in CREs, but not with turnover rates, perhaps owing to differences in the effect sizes of the relevant mutations. Altogether, this probabilistic modeling framework enables a variety of powerful new analyses.

## PUBLICATIONS

Blumberg A, Zhao Y, Huang Y, Dukler N, Rice EJ, Krumholz K, Danko CG, Siepel A. 2019. Characterizing RNA stability genome-wide through combined analysis of PRO-seq and RNA-Seq data. bioRxiv 690644. doi:10.1101/690644

Gulko B, Siepel A. 2019. An evolutionary framework for measuring epigenomic information and estimating cell-type-specific fitness consequences. *Nat Genet* **51:** 335–342. doi:10.1038/s41588-018-0300-z

Huang Y-F, Siepel A. 2019. Estimation of allele-specific fitness effects across human protein-coding sequences and implications for disease. *Genome Res* **29:** 1310–1321. doi:10.1101/gr.245522.118

Hubisz MJ, Williams AL, Siepel A. 2019. Mapping gene flow between ancient hominins through demography-aware inference of the ancestral recombination graph. bioRxiv doi:10.1101/687368

Hutton E, Siepel A. 2019. ACE: A probabilistic model of characterizing gene-level essentiality in CRISPR screens. bioRxiv doi:10.1101/868919

*In Press*

Dukler N, Huang Y-F, Siepel A. 2020. Phylogenetic modeling of regulatory element turnover based on epigenomic data. *Mol Biol Evol* **7:** 2137–2152. doi:10.1093/molbev/msaa073

Hejase HA, Salman-Minkov A, Campagna L, Hubisz MJ, Lovett IJ, Gronau I, Siepel A. 2020a. Genomic islands of differentiation in a rapid avian radiation have been driven by recent selective sweeps. bioRxiv 977694. doi:10.1101/2020.03.07.977694

Hejase HA, Dukler N, Siepel A. 2020b. From summary statistics to gene trees: methods for inferring positive selection. *Trends Genet* **36:** 243–258. doi:10.1016/j.tig.2019.12.008

Hubisz MJ, Siepel AC. 2020. Inference of ancestral recombination graphs using ARGweaver. In *Statistical population genomics* (ed. Dutheil JY), pp. 231–266. Humana, New York.

Joly-Lopes Z, Platts AE, Gulko B, Choi JY, Groen SC, Zhong X, Siepel A, Purugganan MD. 2020. An inferred fitness consequence map of the rice genome. *Nat Plants* **6:** 119–130. doi:10.1038/s41477-019-0589-3

# COLD SPRING HARBOR LABORATORY FELLOWS

In 1986, Cold Spring Harbor Laboratory began a Fellows Program to encourage independent research by outstanding young scientists who, during their graduate studies, displayed exceptional promise of becoming leading scientists of the future. The purpose of this program is to provide an opportunity for these young scientists to work independently at the Laboratory for a period of three to five years on projects of their own choosing. Fellows are provided with a salary, research support, and technical assistance so that they can accomplish their goals free of distraction. The interactions among research groups at the Laboratory and the program of courses and meetings on diverse topics in biology contribute to a research environment that is ideal for scientific innovation by these Fellows.

The CSHL Fellows Program has been tremendously successful and has served as a paradigm for several analogous programs at other institutions, most recently a Fellows Program sponsored by the National Institutes of Health.

The success of the program is apparent from the list of distinguished alumni. Carol Greider—recipient of the 2009 Nobel Prize in Physiology or Medicine for her work on telomerase and telomere function—joined the Fellows Program in 1998. After completing her Fellow training, Carol was on the CSHL faculty for nine years, and she is currently the Daniel Nathans Professor and Director of Molecular Biology and Genetics at Johns Hopkins University School of Medicine. The first CSHL Fellow, Adrian Krainer (1986), is currently a Professor at the Laboratory, as are Chris Vakoc (2008) and Florin Albeanu (2008), currently holding Professor and Associate Professor positions at CSHL, respectively. Scott Lowe (1995) is a Howard Hughes Medical Institute (HHMI) Investigator. After nearly 15 years on the CSHL faculty, he took on a Professorship at Memorial Sloan Kettering Cancer Center. Marja Timmermans (1998) was a member of the CSHL faculty for more than 17 years and recently accepted the Humboldt Professorship at the University of Tübingen. Eric Richards (1989) currently is the Vice President of Research and Senior Scientist at the Boyce Thompson Institute for Plant Research at Cornell University. David Barford (1991) is a Fellow of the Royal Society and Professor of Molecular Biology at the Institute of Cancer Research in London. Ueli Grossniklaus (1994) is Professor at the Institute of Plant Biology, University of Zürich, Switzerland. Térence Strick (2000) left at the end of his fellowship to become a Group Leader at the Institut Jacques Monod in Paris. Lee Henry (2000) joined HHMI's Janelia Farm in Ashburn and joined a project headed by Thomas Südhof. Ira Hall (2004) is a Professor at Yale University and the Director of the Yale Center for Genomic Health. Patrick Paddison, who had joined the Fellows Program in 2004, currently is an Associate Member at the Fred Hutchinson Cancer Research Center in Seattle, Washington.

**Lingbo Zhang** has been a Fellow at the Laboratory since 2013. He joined us from Harvey Lodish's laboratory at the Whitehead Institute of the Massachusetts Institute of Technology, where he studied the regulation of erythroid progenitor cell self-renewal. As a CSHL Fellow, Lingbo is conducting genetic and small-molecule screens to discover novel regulators of normal and aberrant stem cell biology. **Jason Sheltzer** has been a CSHL Fellow since 2015 after completing his graduate work in Angelika Amon's laboratory at MIT. His research focuses on studies of aneuploidy and how it impacts cancer progression. **Semir Beyaz** has been a CSHL Fellow since 2017 after completing his graduate work with Stuart Orkin at Harvard University. He studies how dietary fat intake alters intestinal stem cells, the immune system, and cancer. **Hannah Meyer** became a CSHL Fellow in 2019. She combines genomics and mathematical modeling to study how the immune system discriminates self from non-self in order to effectively fight infection.

# ENVIRONMENT–GENE INTERACTIONS IN REGENERATION, IMMUNITY, AND CANCER

**S. Beyaz**    C. Chung    A. Nizam
                I. Ergin     K. Papciak
                O. Eskiocak  B. Yueh
                H. Mou

Cells respond and adapt to the signals that they receive from their environment. Environmental factors such as nutrients affect cellular states by altering cell state–specific gene expression or metabolic programs. The Beyaz laboratory investigates the causal cellular and molecular mechanisms that link nutrition to organismal health and disease. For example, diets that lead to obesity, such as high-fat diets (HFDs), are significant environmental risk factors that influence cancer incidence and progression in several tissues. Our studies interrogate the functional consequences of diets and diet-induced physiological alterations on stem cell regeneration, immunity, and cancer. We dissect cell-intrinsic and cell-extrinsic mechanisms to uncover mechanistic links that can be therapeutically exploited. The current focus in the laboratory is to decipher the causal mechanisms of environment–gene interactions in two different organs: the intestine and the uterus. We study how dietary and metabolic alterations influence stem cell activity, immunity, and risk of cancer by interrogating the alterations in gene expression, epigenetic state, metabolism, and microbiome. We also develop innovative models to determine the significance of interactions between diverse cell types including stem cells and immune cells within these organs in modulating cancer risk. We ultimately aim to build a comparative blueprint of environment–gene interactions in all relevant tissue types in response to diet and obesity.

## Dietary Regulation of Stem Cell Regeneration

Intestinal stem cells (ISCs) drive the rapid renewal of the intestinal epithelium and remodel intestinal composition in response to diet-induced cues. We previously reported that the fatty acid constituents of a pro-obesity HFD increase ISC and progenitor cell function by activating the lipid-sensing transcription factor PPAR-δ. We found that activation of PPAR-δ increased the vulnerability of stem cells and progenitors to undergo oncogenic transformation. This study revealed a causal stem cell–intrinsic mechanism that links diet-induced obesity to increased intestinal tumor formation.

To explore how diverse dietary fatty acids influence ISC activity in an unbiased manner, we developed an ex vivo fatty acid screening assay in intestinal organoids, which are three-dimensional (3D) epithelial structures grown in a laboratory dish using defined factors. These organoids maintain the structures of normal intestine including stem cells and differentiated cells in culture and therefore allow a reliable comparison of functional and phenotypic changes in epithelial cells in response to diverse fatty acids. Our phenotypic screen in both mouse and human intestinal organoids using all dietary fatty acids revealed that omega-6 fatty acid treatment led to increased stem cell activity and reduced differentiation. Mechanistically, we found that omega-6 fatty acids are metabolized to lipid mediators in ISCs, and signaling through their receptor is both necessary and sufficient to mediate the stem cell–enhancing effects of omega-6 fatty acids. We developed a novel omega-6-rich rodent diet that led to increased abundance of omega-6 and lipid mediators in the mouse intestine. Omega-6-rich diet enhanced stem cell activity and regeneration in the intestine upon injury. Finally, we uncovered the epigenetic mechanisms involving transcription factors that modulate the enhanced ISC state in response to omega-6 fatty acids. We are currently working on a manuscript to report these discoveries.

We performed similar studies in other epithelial organoids including prostate, breast, and endometrium organoids. We found that endometrial organoids exhibit functional and molecular alterations in response

to fatty acids that are similar to intestinal organoids. Our ongoing studies aim to comparatively assess the molecular mechanisms that govern the tissue specificity of fatty acid–induced alterations in epithelial stem cell activity.

## Diet, Metabolism, and Microbiome in Regulating Cancer Immunity

The intestinal epithelium serves as the interface between the dietary intake of nutrients, commensal microbes, and immune cells. Intestinal tumorigenesis in this dynamic interface is significantly influenced by the cross talk between cancer cells, immune cells, and microbes. Although cancers develop several strategies to evade the immune system, little is known about how diet-induced obesity impacts cancer-immune recognition throughout intestinal tumorigenesis. T cells of the adaptive immune system recognize antigens in the context of major histocompatibility complex (MHC) molecules and play a critical role in cancer-immune surveillance. MHC-II-mediated activation of CD4+ T cells can engage multiple mechanisms that contribute to immunity against tumors. Although MHC-II expression and function are usually considered to be restricted to professional antigen-presenting cells like dendritic cells, several studies demonstrated that intestinal epithelial cells express high levels of MHC-II and are able to capture, process, and present antigens to CD4+ T cells. We find that at steady state, ISCs express high levels of MHC-II protein on their cell surface, which is significantly down-regulated in response to HFD-induced obesity. Downregulation of immune recognition molecules is one of the key strategies that cancer cells use to evade immune-mediated clearance. Indeed, several human cancers down-regulate MHC-II expression to evade antitumor-immune responses, and lower MHC-II expression in tumors correlates with poor survival. To test whether HFD-mediated down-regulation of MHC-II in ISCs enhances intestinal tumorigenesis, we used an in vivo orthotopic syngeneic colon transplantation assay in mice that we recently pioneered. Using Lgr5-Cre APC$^{L/L}$ mice that have been on a purified control diet or HFD, we sorted MHC-II+ APC-null or MHC-II− APC-null premalignant ISCs by flow cytometry after deletion of the tumor

suppressor gene *Apc* by tamoxifen administration. We then transplanted these premalignant cells into the distal colon of syngeneic immune-competent or immune-deficient hosts. We found that reduced MHC-II expression in premalignant ISCs leads to increased tumor initiation rate in vivo in immune-competent hosts but not immune-deficient hosts. Moreover, we demonstrated that ISC-specific genetic ablation of MHC-II in engineered *Apc*-mediated intestinal tumor models led to increased tumor burden in a cell-autonomous manner.

The intestinal microbiome plays a significant role in regulating intestinal immunity. Because dietary perturbations are among the major external factors shaping the intestinal microbiome, we asked whether HFD-induced alterations in the microbiome influence MHC-II expression in ISCs and the intestinal epithelium. Consistent with previous findings, HFD-induced obesity led to microbial dysbiosis with reduced bacterial diversity. To determine whether the microbiome is involved in regulation of epithelial MHC-II levels, we treated mice with broad-spectrum antibiotics, which ablated bacterial diversity and massively altered community composition. Notably, antibiotic treatment was accompanied by decreased MHC-II expression in ISCs and the intestinal epithelium, comparable to that observed in HFDs. Among the bacterial genera most strongly ablated under HFD conditions and most strongly correlating with MHC-II levels was *Helicobacter*. Indeed, mice harboring *Helicobacter* species had significantly higher MHC-II expression in ISCs compared to mice lacking these species. Mechanistically, we found that pattern-recognition receptor and JAK/STAT signaling regulate MHC-II expression in ISCs. A manuscript reporting this discovery is currently in revision at *Cell Stem Cell*. Our ongoing studies aim to determine the significance of epithelial MHC-II expression in the context of tumor progression and responsiveness to cancer immunotherapy.

To further explore how diet-induced obesity impacts immunity and contributes to cancer risk in the intestine, we performed single-cell RNA sequencing of intestinal immune cells in response to diet-induced obesity. We found that an HFD leads to dampened cytotoxicity in intestinal T cells. To test whether fatty acid constituents of the HFD are sufficient to drive the impaired cytotoxicity, we pretreated T cells with diverse fatty acids and assayed for cytotoxicity

and effector molecule production ex vivo. Interestingly, fatty acid–treated T cells exhibited defective tumor-killing capacity and effector molecule production. We found that the transcriptional programs regulating T-cell activation, proliferation, and cytotoxicity were down-regulated in response to fatty acid treatment. We used genetic loss-of-function and gain-of-function mouse models to assess necessity and sufficiency of key metabolic and transcriptional mechanisms underlying the defects in T-cell activation in response to fatty acids. Our ongoing studies are investigating the transcriptional, epigenetic, and signaling pathways linking fatty acid metabolism to T-cell function.

## Integrative Analysis of Cancer Risk in Response to Diet and Obesity

Our data suggest that a lard-based HFD leads to increased tumorigenicity in the intestine through several orthogonal mechanisms involving tumor-initiating stem cells, immune cells, and microbiomes. To identify how diverse dietary inputs and patterns perturb the cellular and molecular networks associated with cancer risk, we developed novel dietary models and assessed the interactions between intestinal epithelial cells (including stem cells), immune cells, and microbes over time. We performed gene expression, metagenomics, and metabolomics analyses and integrated the data to discover multidimensional features and define potential mediators that contribute to intestinal tumorigenesis. Moreover, we are performing comparative analyses of different tissue types that exhibit increased cancer risk in the context of obesity. Our goal is to build a unifying model for understanding how diet and obesity influence cancer risk at the molecular, cellular, tissue, and organismal level. In this project, we are collaborating with the Meyer laboratory at CSHL and the Mason laboratory at Cornell on developing innovative analytical models to integrate these multidimensional data.

## Development of Human Organoid Models to Study Endometrial Cancer

Endometrial cancer is one of the most common gynecologic malignancies and is strongly associated with obesity. However, little is known about the causal cellular and molecular mechanisms that contribute to endometrial cancer risk. There are several types of endometrial cancers that are classified based on histological features. Type 1 endometrial cancers involve low-grade endometrioid endometrial cancers. These are the most common endometrial cancers and are strongly associated with obesity. Type 2 endometrial cancers involve high-grade serous endometrial cancer, clear cell endometrial cancer, and uterine carcinosarcoma. In collaboration with Gary Goldberg and Marina Frimer from Northwell Health, we have successfully established a biobank at CSHL for patient-derived organoid models for all endometrial cancer types as well as normal endometrium. We have characterized these organoids using histology, genomics, transcriptomics, epigenomics, and pharmacological tools. Moreover, we optimized coculture conditions for growing endometrial organoids with immune cells to model and study tumor–immune interactions. We aim to use these models to better understand endometrial tumorigenesis and define mechanisms that can be exploited therapeutically.

### PUBLICATIONS

Luo Y, Yang Y, Liu M, Wang D, Wang F, Bi Y, Ji J, Li S, Liu Y, Chen R, et al. 2019. Oncogenic KRAS reduces expression of FGF21 in acinar cells to promote pancreatic tumorigenesis in mice on a high-fat diet. *Gastroenterology* **157:** 1413–1428.e11. doi:10.1053/j.gastro.2019.07.030

*In Press*

Ordovas-Montanes J, Beyaz S. Rakoff-Nahoum S, Shalek AK. 2020. Distribution and storage of inflammatory memory in barrier tissues. *Nat Rev Immunol* **20:** 308–320.

Xu M, Ergin I, Beyaz S. 2020. Advances in diagnostic procedures and their applications in the era of cancer immunotherapy. *Adv Exp Med Biol* **1244:** 37–50.

# GENOMICS APPROACHES TO STUDYING ORGAN FUNCTION

**H.V. Meyer**    A. Garipcani    K. Papciak

The Meyer laboratory combines genomics and mathematical modeling to understand the mechanisms of healthy and pathological organ function. The main focus of our research lies in studying the processes of T-cell education in the thymus. We develop experimental and computational approaches to elucidate the interactions of T cells and thymic epithelial cells that drive self-tolerance and generate diversity in the immune system.

Prior to Dr. Meyer starting her research group at CSHL in March 2019, she completed her Ph.D. at the European Bioinformatics Institute in Cambridge, UK, where she used quantitative genetics to study the function of the human heart in healthy and disease conditions. In the past year, the Meyer laboratory finished a long-standing collaboration project with biomechanical engineers and clinical cardiologists to understand the function of the complex muscular network that lines the inner surface of the heart (Meyer et al. 2020). In the following, we describe our main findings of this collaborative effort and continue with current and ongoing projects on T-cell education.

## Fractal Properties of Heart Muscle Are Crucial to Cardiovascular Function

H.V. Meyer [in collaboration with E. Birney, EMBL-EBI, Cambridge, UK; D. O'Regan, Imperial College London]

The chambers of the mature human heart are covered by a complex network of muscular strands, the myocardial trabeculae. They were first described by the early human anatomists and are remarkably well conserved in vertebrate evolution. Their role in facilitating oxygenation of the developing fetal heart has been established, but their genetic architecture, their physiological function in the adult heart, and their potential role in common diseases are unknown.

To determine the genetic architecture of myocardial trabeculae, we performed a genome-wide association study using fractal analysis of trabecular morphology as an image-derived phenotype in 18,096 UK Biobank participants. We found 16 associations with trabecular complexity in loci related to hemodynamic phenotypes and regulation of cytoskeletal arborization, gene expression variation in cardiac tissues, and cardiac development chromatin annotation (Fig. 1). Knockout models of loci-associated genes in medaka fish showed a marked decrease in myocardial trabecular complexity.

To understand the influence of myocardial trabeculation on cardiac function, we used a biomechanical simulation of the heart in a hemodynamic circuit and derived cardiac parameters (such as stroke volume and work) as a function of trabecular complexity. Analogous to the cardiac function parameters estimated in these simulations, we measured the same parameters in the participants of the UK Biobank. Both experiments show concordant results, suggesting a causal relationship between increased trabecular complexity and ventricular performance.

Finally, through genetic association studies with cardiac disease phenotypes and Mendelian randomization, we find a causal relationship between trabecular morphology and cardiovascular disease risk, with increased trabeculation leading to reduced risk of disease.

The triangulation of theoretical models, observational data, and genomics is persuasive evidence that trabeculae are not simply vestigial features of development but are unexpected determinants of cardiac performance in adult hearts. Understanding the pathways that regulate the development of such complex biological structures provides a foundation for exploring new causal mechanisms in common cardiovascular diseases.

## Transcriptional Regulation in the Thymus

A. Garipcani, K. Papciak, H.V. Meyer

The thymus generates and selects a highly variable yet specific T-cell repertoire that discriminates between self and nonself antigens. Within the thymus, T-cell progenitors known as thymocytes interact with
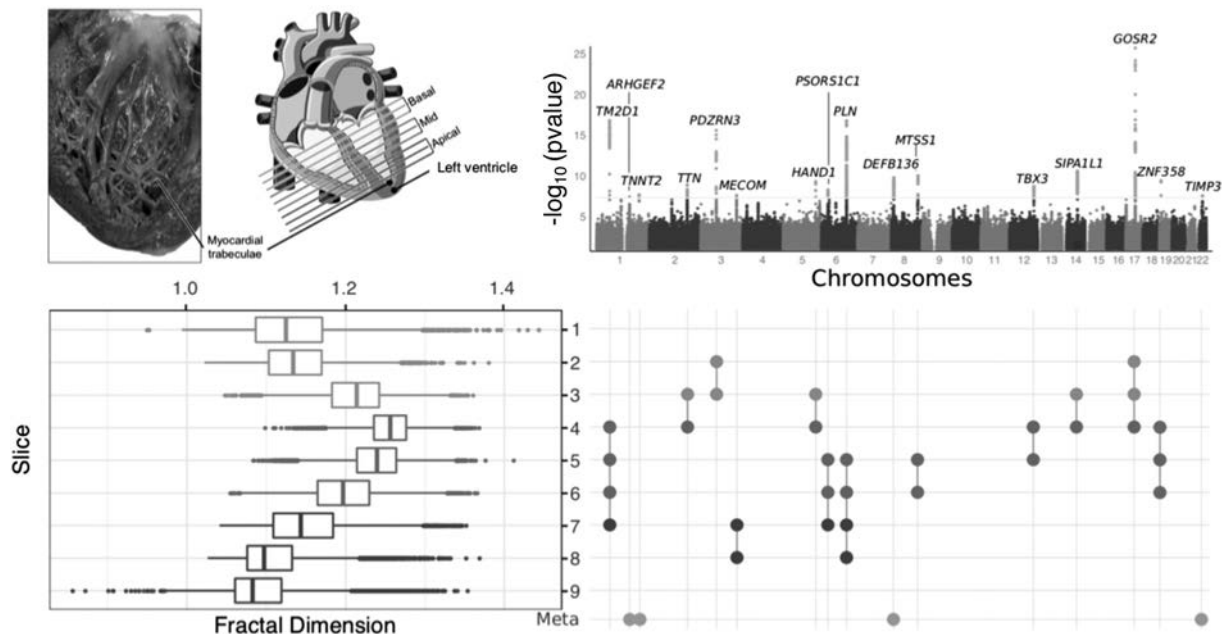
**Figure 1.** Complexity (*lower left*) of muscular network on the inner surface of the heart muscle (*upper left*) is associated with 16 independent loci across the genome (*upper right*), which are specific for distinct regions in the ventricle (*lower right*).

thymic epithelial cells to select for T cells with functional T-cell receptors (positive selection) and against T cells that bind strongly to self-antigens (negative selection). Thymic epithelial cells in the cortical area of the thymus mediate positive selection. The negative selection process that induces central tolerance largely depends on antigen presentation in the thymic medulla. Medullary thymic epithelial cells (mTECs) express nearly all peripheral antigens, representing essentially all tissues of the body. This phenomenon is termed promiscuous gene expression. Developing T cells that react to these peripheral self-antigens are removed to ensure that no autoimmune responses occur.

Collectively, mTECs express the majority of protein coding genes, yet each antigen is expressed by only a small fraction of mTECs. The origin of this heterogeneous expression is not fully understood. Whereas some studies have found coexpression groups of cells that express the same antigens, others have described the expression patterns across cells as purely stochastic. Conclusive evidence is still missing. As patterns of coexpression or lack thereof have implications for the screening and migration process of thymocytes during selection, adequate methods are needed to investigate the regulation of gene expression on a single-cell and tissue-wide level.

## Thymic Transcriptional Landscape at Epitope Resolution

Antigens or even small antigen epitopes missing in the thymus can result in autoimmunity by allowing autoreactive T cells to escape into the periphery. To know which antigens are available for T-cell education, we need to understand the precise transcriptional landscape of promiscuous gene expression in mTECs.

We have collected human thymus samples and generated RNA sequencing data from mTECs, which allows us to analyze their genome-wide transcription start sites (TSSs). Analogously, we processed TSS coordinates from 51 human tissues from the FANTOM5 consortium. Based on this data, we can generate comprehensive antigen epitope maps of thymus and compare them to epitope maps of the periphery (FANTOM5 data), which will allow us to find differentially missing epitopes in the thymus.

This data also allows us to study the features of differentially missing epitopes by comparing the annotation of thymus-derived and peripheral reference start sites to features of annotated genes (e.g., regions around consensus TSSs, exons, introns, and antisense usage of the promoter). Preliminary results suggest differentially shifted TSSs in mTECs compared to peripheral tissues. In particular, we found that TSSs in

mTECs are often found in intronic regions or even in the antisense direction. This shift in TSSs is likely to lead to missing epitopes for tolerance induction and is a first step in detecting risk epitopes of autoimmunity.

### Tissue-Wide Gene Expression Patterns

Tissue-wide gene regulation patterns and spatially restricted T-cell migration could exert a large influence on the self-antigen encounters a given thymocyte can have. Therefore, we will study the regulation of promiscuous gene expression in mTECs on a tissue-wide level. Thymic epithelial cells comprise only ~1% of cells in the thymus, which makes them a rare cell population to study. To capture spatially resolved coregulation profiles of mTECs, we need to be able to capture sufficient numbers from known locations.

K. Papciak (research assistant) and A. Garipcani (postdoctoral fellow) joined my laboratory to develop and establish experimental protocols that will allow us to assess gene expression of single mTECs across continuous small compartments of the thymus. In parallel, we are developing computational models to compile these continuous data sets into a three-dimensional tissue-wide expression profile. We will use these profiles to infer intracellular and intercellular coregulation and study their implications on the migration of T cells during selection.

*In Press*

Meyer HV, Dawes TJW, Serrani M, Bai W, Tokarczuk P, Cai J, de Marvao A, Henry A, Lumbers RT, Gierten J, et al. 2020. Genetic and functional insights into the fractal structure of the heart. *Nature* **584:** 589–594.

# GENETIC APPROACHES TO STUDY CHROMOSOME DOSAGE IMBALANCES AND ANTICANCER DRUG SPECIFICITY

**J. Sheltzer**   D. Adebambo   A. Lakhani   E. Sausville   J. Smith
                  V. Girish      K. Long       K. Schukken    P. Suri
                  K. John        D. Lukow      D. Shen        A. Vasudevan

The Sheltzer laboratory applies a variety of techniques, including chromosome engineering, CRISPR mutagenesis, and single-cell analysis, to address fundamental questions in cancer biology. We are particularly interested in exploring the role of chromosome dosage imbalances in cancer development and progression. Additionally, we are using CRISPR-based genetic approaches to interrogate the specificity of various anticancer drugs.

## Genetic Investigation of Cancer Drug Targets

D. Lukow, K. John, K. Long, E. Sausville, D. Shen, J. Smith

Substantial progress has been made in the treatment of certain malignancies by targeting cancer "addictions" or genetic dependencies that encode proteins required for the survival and/or proliferation of cancer cells. Therapeutic agents that block the function of a cancer dependency—like the kinase inhibitor lapatinib in Her2+ breast cancer—can trigger apoptosis and durable tumor regression. Discovering and characterizing druggable cancer dependencies is a key goal of preclinical research.

We are using CRISPR-Cas9 to identify genetic addictions in different cancer types. While screening cancer cell lines, we discovered that several genes previously reported to be both cancer-essential and the target of anticancer drugs are actually dispensable for cancer growth. For instance, we found that PIM1, a putative "addiction" in multiple cancer types, could be eliminated using CRISPR without any detectable loss in cancer cell fitness. Additionally, we demonstrated that SGI-1776, a small-molecule inhibitor of PIM1 undergoing clinical trials, continued to kill PIM1-knockout cancer cells with no decrease in potency. This indicated that an anticancer agent had entered clinical trials in human patients because of flawed preclinical data and based on an incorrect understanding of that drug's mechanism of action. Remarkably, we uncovered similar results for 10 different drugs and six drug targets.

These findings left us in a strange position; we had discovered 10 drugs that had potent anticancer activity but whose putative mechanisms of action were incorrect. We therefore set out to see whether we could uncover what protein(s) these drugs actually targeted. To accomplish this, we performed a genomic suppressor screen and attempted to isolate cancer cells that had acquired a resistance-granting mutation in whatever protein a drug truly bound to. Using this approach, we were able to discover that the mischaracterized anticancer drug OTS964 actually functions as a potent and specific inhibitor of the cyclin-dependent kinase CDK11. We are currently working to expand this approach to the other drugs that we have studied, and we are working to elucidate the function of CDK11 in normal cells and in cancer.

## Chromosomal Instability and Aneuploidy in Cancer

D. Lukow, A. Vasudevan, K. Schukken, D. Adebambo, A. Lakhani, V. Girish, P. Suri

Human cancers exhibit a diverse array of genomic gains and losses that alter the dosage of hundreds or thousands of genes at once. The prevalence of aneuploidy in cancer—first noted >100 years ago—has led to a widespread belief that genomic imbalances play a crucial role in tumor development. Indeed, in the early twentieth century, Theodor Boveri speculated that abnormal karyotypes altered the balance between pro- and antiproliferative cellular signals and were therefore sufficient to induce transformation. "Boveri's hypothesis" has motivated decades of research into the origins and consequences of aneuploidy, but the precise relationship between abnormal karyotypes and tumorigenesis remains unclear.

We are developing novel models of aneuploidy to explore the impact of genome dosage alterations on tumor development and progression. Using a variety of techniques, including CRISPR-Cas9, microcell-mediated chromosome transfer, and small-molecule mitotic accelerants, we are changing chromosome copy number in human cells. We can then study how these aneuploidies impact a number of cancer-related phenotypes, including metastasis, chemotherapy resistance, and cell cycle progression.

We have characterized a series of isogenic colon cancer cell lines that harbor single extra chromosomes and found that these aneuploidies exhibit significant tumor-suppressive properties (Sheltzer et al., *Cancer Cell* **31:** 1 [2017]). Additionally, we have worked to understand how these aneuploidies affect metastasis. We found that adding a single extra copy of chromosome 5 caused a partial epithelial–mesenchymal transition. These cells strongly down-regulated E-cadherin, Epcam, and Claudin-7 and exhibited increased motility and invasive behavior. These patterns were not observed in cells harboring several additional aneuploidies, suggesting that the phenotype is caused by the increased dosage of a gene or genes found on chromosome 5. We are applying a variety of genetic approaches to identify these key factors.

More generally, we hypothesize that aneuploidy is commonly detrimental under "normal" growth conditions. That is, when a cell is grown in rich media with an adequate supply of nutrients and growth factors, aneuploidy is disfavored. However, in stressful environments, unique karyotypes may exist that confer an environment-specific growth advantage. To test this, we have treated cancer cells with Mps1 inhibitors to generate populations of cells with random aneuploidies. We have observed that pretreatment with Mps1 inhibitors speeds the evolution of drug resistance in cells exposed to various chemotherapy agents. In the case of one drug, vemurafenib, this resistance consistently co-occurs along with the gain of chromosomes 11 and 18. We speculate that similar aneuploidy patterns may exist for other drugs or environments—that is, an "optimal" karyotype can be found to maximize growth potential in each condition. We believe that these results may explain the close association that we have previously documented between aneuploidy and poor prognosis in cancer.

## PUBLICATIONS

Giuliano CJ, Lin A, Girish V, Sheltzer JM. 2019. Generating single cell–derived knockout clones in mammalian cells with CRISPR/Cas9. *Curr Protoc Mol Biol* **128:** e100.

Lin A, Giuliano CJ, Palladino A, John KM, Abramowicz C, Yuan ML, Sausville EL, Lukow DA, Luwei L, Chait AR, et al. 2019. Off-target toxicity is a common mechanism-of-action of cancer drugs undergoing clinical trials. *Sci Transl Med* **11:** eaaw8412.

Vasudevan A, Baruah PS, Smith JC, Wang Z, Sayles NM, Andrews P, Kendall J, Chunduri NK, Levy D, Wigler M, et al. 2019. Single-chromosome aneuploidies can function as metastasis suppressors and metastasis promoters. bioRxiv doi:10.1101/590547

# UNDERSTANDING AND TARGETING HEMATOPOIETIC STEM AND PROGENITOR CELLS

**L. Zhang**  B. Li        S. Mishra      D. Tsang Hu
          G. Mazzeo    A. Thakkar     E. Wong
          W. Mei       G. Trivedi     L. Zhu

The research in our laboratory addresses a central question in blood cell formation—a process also known as hematopoiesis—which is how self-renewal and differentiation are properly balanced in the hematopoietic stem and progenitor cell (HSPC) population. The balance is crucial because dysregulation of this process causes various hematopoietic diseases and malignancies. The Zhang laboratory investigates how growth signals and nutrients coordinate to regulate this process. We utilize both CRISPR-Cas9 functional genomic and chemical genomic approaches to identify novel self-renewal pathways and metabolic vulnerabilities, and we aim to develop novel therapeutic strategies for hematopoietic diseases and malignancies.

The Zhang laboratory has identified muscarinic acetylcholine receptor M4 (CHRM4) as a novel regulator of early erythroid progenitor proliferation and differentiation and a novel drug target for treatment-refractory myelodysplastic syndrome (MDS) (Trivedi et al. 2019). We have demonstrated that the CHRM4-CREB pathway regulates proliferation and differentiation of early erythroid progenitor both in vitro and in vivo. In the *Mx1-Cre Srsf2*$^{P95H/+}$ genetically engineered MDS mouse model that captures essential pathological features of human MDS, we have demonstrated that CHRM4 antagonists completely corrected anemia of MDS. In collaboration with medicinal chemists at Northwell Health System, we have developed CHRM4 selective antagonists with improved drug metabolism and pharmacokinetic (DMPK) properties. Through the collaboration, we have characterized drug efficacy, DMPK, and toxicology properties for lead compounds. We are working on these investigational new drug (IND)-enabling studies and preparing clinical trials for treatment-refractory MDS at Northwell Health System. The novel therapy will not only benefit MDS patients, but also cancer patients who are undergoing chemotherapy and radiation therapy.

Our laboratory has also identified pyridoxal kinase (PDXK) and the vitamin B$_6$ pathway as novel selective metabolic dependencies in leukemia and novel drug targets for the treatment of acute myeloid leukemia (AML) with minimal bone marrow suppression side effects (Chen et al. 2020). We performed a focused CRISPR-Cas9 screen and found that PDXK, an enzyme that catalyzes the formation of the bioactive form of vitamin B$_6$ pyridoxal phosphate (PLP), was selectively required for leukemia cell proliferation relative to normal HSPCs both in vitro and in vivo. We have further established a liquid chromatography–mass spectrometry (LC–MS) method to monitor intracellular PLP and found that PLP-dependent generation of polyamine and PLP-dependent-glutamic-oxaloacetic transaminase 2 (GOT2) pathway, which produces nucleotides, selectively support AML cell proliferation. Our work has identified the vitamin B$_6$ pathway as a pharmacologically actionable dependency in AML. We are collaborating with medicinal chemists at Memorial Sloan Kettering Cancer Center to modify PDXK inhibitors for better DMPK properties, and we expect to extend this work to collaboration with the Tri-Institutional Therapeutics Discovery Institute and Takeda Pharmaceutical Company to provide lead compounds for clinical development.

## Discovering and Defining, for the First Time, the "Hematopoietic Reflex" and Targeting It as a Novel Therapeutic Strategy for Hematological Malignancy

MDS is a form of lethal hematopoietic malignancy that is characterized by pancytopenia resulting from progressive bone marrow failure. Clinically, therapeutic options for MDS are very limited. Only ~20% of MDS patients benefit from standard erythropoietin (EPO) treatment, and many of the initial responders do not

have a long-term response. One major advance in MDS treatment in recent years is the approval of lenalidomide as a novel drug to specifically treat a subgroup of MDS with chromosome 5q deletion, which only accounts for ~5% of the total MDS population. The only option for patients who do not respond to EPO and lenalidomide is red blood cell (RBC) transfusion, but transfusion exposes patients to insufficient correction of anemia, alloimmunization, and organ failure secondary to iron overload. Therefore, there is an unmet and urgent demand for patients and clinicians to have novel therapeutics to treat these refractory MDSs.

My laboratory has discovered CHRM4 as a novel regulator of early erythroid progenitor self-renewal. CHRM4 is expressed in early erythroid progenitors and is up-regulated during erythroid differentiation and down-regulated during self-renewal. In the primary early erythroid progenitor culture system, genetic down-regulation of CHRM4 or pharmacologic inhibition of it using nanomolar concentrations of its selective antagonists promoted early erythroid progenitor proliferation and differentiation. Mechanistically, we have showed that CREB, a downstream effector of CHRM4, preferentially binds to and triggers the up-regulation of genes important for the maintenance of early erythroid progenitor status. These genes include the transcription factor GATA2, haploinsufficiency of which causes pediatric MDS, and ZFP36L2, an RNA-binding protein that we previously identified as a crucial molecular switch balancing early erythroid progenitor self-renewal and differentiation (Zhang et al., *Nature*

**499**: 92 [2013]). Our laboratory has uncovered muscarinic acetylcholine receptor as an important mediator connecting neuronal activity to HSPC self-renewal capacity, and our research discovered this phenomenon. My laboratory for the first time defined and named this phenomenon as the "hematopoietic reflex," which represents the previously unreported regulation of HSPC self-renewal capacity by neuronal activity (Fig. 1).

We have further demonstrated that pharmacological inhibition of CHRM4 completely corrected bone marrow failure of MDS in genetically engineered MDS mouse models and cultured primary human patient samples of MDS. Strikingly, injection of CHRM4 antagonists completely corrected anemias of MDS in vivo in the genetically accurate MDS mouse model *Mx1-Cre Srsf2P95H/WT* that faithfully recapitulates the essential pathological phenotypes of MDS. The treatment has extended survival of MDS mice to the life span of control wild-type mice. We have demonstrated that similar to human MDS patients, MDS mouse models exhibit abnormal reduction of early erythroid progenitor and elevation of plasma EPO levels in comparison to wild-type control mice. Importantly, injection of CHRM4 antagonists corrected both early erythroid progenitor deficiency and the abnormal plasma EPO levels of MDS mice to levels comparable to wild-type control mice. These results further demonstrated that targeting CHRM4 overcame early erythroid progenitor deficiency and EPO refractoriness.

Our laboratory is currently translating these promising preclinical discoveries into clinical developments
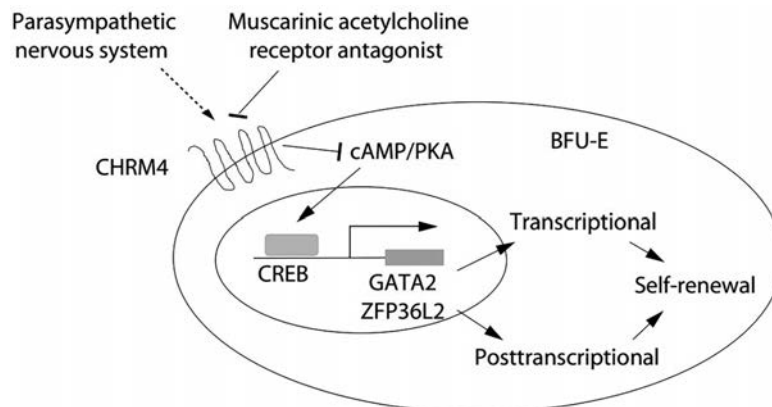


**Figure 1.** Muscarinic acetylcholine receptor is an important mediator connecting neuronal activity to hematopoietic stem and progenitor cell (HSPC) self-renewal capacity. For the first time, our work discovered and defined the "hematopoietic reflex," which represents the previously never-reported regulation of HSPC self-renewal capacity by neuronal activity.

for EPO-refractory MDS. Having demonstrated that the selective antagonists corrected anemia of MDS in the Srsf2P95H genetic background, we will further assess the efficacy of the inhibition in multiple murine genetic models of MDS carrying different genetic abnormalities and identify genetic markers of response to the antagonists using primary MDS patient samples. To maximize therapeutic efficacies and reduce nonspecific side effects of these antagonists, we will modify their structures to optimize DMPK properties. Through collaboration with Northwell Health System, we are currently developing highly selective antagonists of CHRM4 as lead compounds for clinical trials to treat MDS. We have received awards from the National Institutes of Health (NIH), Northwell Health System, and private foundations to support these promising preclinical developments.

## Identifying and Targeting Vitamin B$_6$ Addiction in Acute Myeloid Leukemia

AML is one of the most devastating forms of blood cancer; it affects around one million people and results in 147,000 deaths per year worldwide. AML is characterized by the abnormal production of myeloid lineage of blood cells and the rapid growth of abnormal leukemia blasts in bone marrow and peripheral blood. The symptoms of AML include shortness of breath, bleeding, and increased risk of infection. AML is a very aggressive cancer; it progresses rapidly, and becomes fatal within weeks to months. Currently, therapeutic options for AML are very limited. Only ~35% of AML patients <60 years old and 10% >60 years old benefit from standard chemotherapy. Older AML patients who cannot tolerate intensive chemotherapy only have a survival of 5–10 months. Thus, there is an urgent and unmet medical need to develop novel therapeutics for AML patients.

Cancer cells undergo metabolic reprogramming to support their abnormal proliferation. However, because of our limited understanding of molecular details and lack of validated drug targets, designing therapeutic strategies to exploit aberrant metabolism has proven challenging. The recent success of development of isocitrate dehydrogenase (IDH) inhibitors as novel agents to treat AML highlights the potential of treating AML through targeting dysregulated metabolism in cancer cells. However, the vast majority of oncogenic drivers alter cellular metabolism through indirect mechanisms and, as such, the metabolic regulators that causally contribute to cancer initiation and maintenance are not obvious and cannot be inferred from genomic analyses alone. More importantly, normal stem and progenitor cells in adult tissues such as HSPCs also undergo regular expansion and proliferation, therefore sharing similar metabolic requirements to cancer cells. Consequently, one major side effect of most standard chemotherapies is the depletion of normal HSPCs resulting from targeting of shared metabolic features. Therefore, design of novel therapeutics selectively targeting leukemic cells is challenging.

Through analyzing gene expression profiles from normal HSPCs and leukemic cells, we uncovered that the global metabolic gene expression signature is capable of discriminating these two cell types. Importantly, the metabolic differences among different subtypes of acute myeloid leukemia cells carrying different genetic abnormalities are significantly smaller than the differences between them and HSPCs. This suggests that there are common metabolic vulnerabilities shared by multiple subtypes of acute leukemia, which potentially serve as drug targets to block their proliferation, while exhibiting no or minimal side effects on HSPCs. We identified approximately 300 metabolic genes that follow this pattern, in which they are highly expressed across multiple subtypes of leukemia in comparison to HSPCs. To identify metabolic vulnerabilities in AML not harboring IDH mutations, we performed a focused CRISPR-Cas9 "drop-out" screen using a single-guide RNA (sgRNA) library targeting metabolic genes highly expressed in AML cells.

We have determined PDXK—an enzyme that produces bioactive forms of vitamin B$_6$ PLP—to be preferentially required for AML cell proliferation compared to many other cell types. PDXK kinase activity is required for both PLP production and proliferation of AML cells, and pharmacological suppression of PLP with the antituberculosis drug isoniazid or PDXK with its direct inhibitor 4′-O-methylpyridoxine recapitulated the effects of PDXK genetic disruption. PLP is a cofactor for many enzymes involved in cell proliferation and, accordingly, PDXK disruption reduced intracellular concentrations of key metabolites needed for cell division, and disruption of the PLP-dependent enzymes ornithine decarboxylase (ODC1) or GOT2 selectively inhibited AML cell
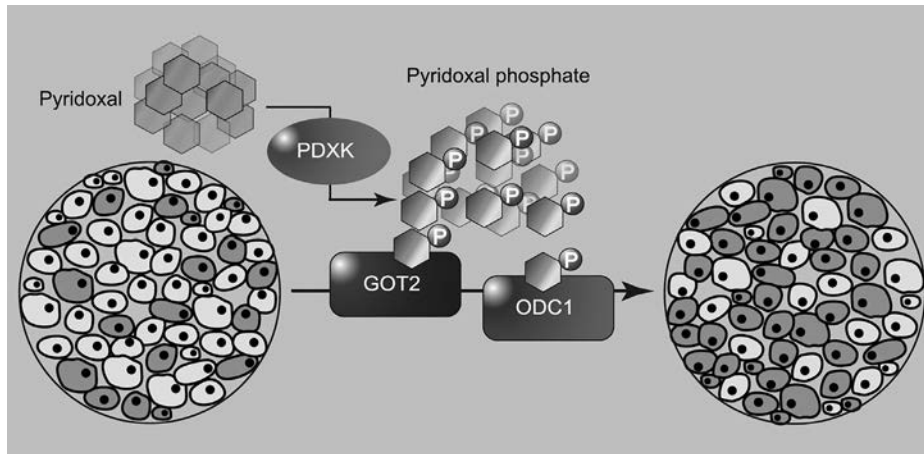
**Figure 2.** Vitamin B$_6$ addiction in acute myeloid leukemia (AML). Vitamin B$_6$–dependent generation of polyamine and the vitamin B$_6$–dependent glutamic-oxaloacetic transaminase 2 (GOT2) pathway, which produces nucleotides, selectively support AML cell proliferation. Our work has identified the vitamin B$_6$ pathway as a pharmacologically actionable dependency in AML and highlighted the importance of studying vitamin pathways in cancer and targeting them as novel cancer therapeutics.

proliferation in a manner that could be partially rescued by addition of downstream products. Our work identified the vitamin B$_6$ pathway as a pharmacologically actionable dependency in AML. We revealed that leukemia cells are addicted to the vitamin B$_6$ pathway, such that its inhibition selectively impairs their proliferation compared to other normal and cancer cell types (Fig. 2).

Our results support the emerging view that essential vitamins—for example, vitamin D in pancreatic cancer or vitamin C in colorectal cancer and leukemia—can ultimately play decisive roles in cancer cell proliferation and maintenance, and align with recent epidemiological studies hinting that vitamin B$_6$ is not chemopreventive but instead may increase cancer risk. Of note, ODC1 has previously been identified as a target in colon and several other cancers, and one biochemical outcome of GOT2 inhibition is a reduction in asparagine, which can also be achieved in leukemia by the FDA-approved drug asparaginase. These observations imply that the vitamin B$_6$ pathway coordinates multiple activities that are critical for cancer maintenance and that PDXK inhibitors, by simultaneously attenuating these pathways, are important antileukemia agents. Together, our work emphasizes the importance of studying vitamin pathways in cancer and targeting them as novel cancer therapeutics. We are collaborating with medicinal chemists at Memorial Sloan Kettering Cancer Center to modify

PDXK inhibitors for better DMPK properties, and we expect to extend this work to collaboration with the Tri-Institutional Therapeutics Discovery Institute and Takeda Pharmaceutical Company to provide lead compounds for clinical development.

## Functional Genomic Screen to Identify Regulatory Network of Early Erythroid Progenitor Self-Renewal

Erythropoiesis is a multistage developmental process that results in erythrocyte production. EPO is a crucial regulator of this process, triggering survival of the late erythroid progenitor, the colony-forming unit erythroid (CFU-E), and has thus been widely used for the treatment of anemias. However, many anemias associated with cancer chemotherapy and radiation therapy are refractory to EPO treatment. The refractoriness of these pathologic conditions is partially due to insufficiency and loss of the burst-forming unit erythroid (BFU-E), an early erythroid progenitor that is unresponsive to EPO. Therefore, targeting BFU-E self-renewal is crucial for sustained erythropoiesis and treatment of anemias in these conditions. A systematic and functional analysis to better understand molecular mechanisms underlying BFU-E self-renewal holds promise in the development of novel therapeutics for EPO-refractory

anemias associated with cancer chemotherapy and radiation therapy.

It has been shown that gene transactivation of the glucocorticoid receptor (GR) is essential for erythroid lineage expansion, and we have previously identified RNA-binding protein ZFP36L2 as a direct GR target gene essential for this process. ZFP36L2 binds to mRNAs highly expressed during erythroid differentiation and negatively regulates their expression. However, whereas several essential GR downstream target genes and cofactors have been identified, a systematic understanding of how the GR regulates BFU-E self-renewal is still missing. We have employed glucocorticoid-induced gene-expression profiling, GR chromatin immunoprecipitation sequencing (ChIP-seq), and shRNA functional genomic screening to provide a systematic and functional understanding of molecular mechanisms underlying BFU-E self-renewal. We combined glucocorticoid-induced differential gene expression profiling, GR ChIP-seq, and shRNA functional genomic screening and identified 43 genes as direct targets of the GR and 21 genes as novel regulators of BFU-E self-renewal. We further validated several identified genes as negative regulators of BFU-E self-renewal. Our results not only provide a systematic and functional understanding of the BFU-E self-renewal mechanism, but also serve as a resource for future research into novel BFU-E self-renewal regulators.

In summary, the research in the Zhang laboratory centers on HSPCs. We investigate how self-renewal and differentiation are properly balanced in the HSPC population and how dysregulation of this process causes diseases. Our laboratory investigates how growth signals and nutrients coordinate to regulate this process and identifies novel self-renewal pathways and metabolic vulnerabilities to target hematological diseases and malignancies. Our work for the first time discovered and defined the "hematopoietic reflex" as a novel concept, which represents the regulation of HSPC self-renewal capacity by neuronal activity. Our work also emphasized the importance of studying and targeting vitamin pathways in cancer. Through both genetic and chemical functional genomics, our research identified receptor CHRM4 and metabolic enzyme PDXK as novel druggable targets to treat hematological malignancies MDS and AML. We are currently collaborating with medicinal chemists to further modify DMPK properties of small-molecule inhibitors to better drug these targets. Through collaboration with hospitals and clinicians, we are testing our lead compounds in primary human disease samples of MDS and AML and planning to move lead compounds into clinical trials to treat these lethal diseases. Our research has been supported by CSHL, NIH, Northwell Health System, and private foundations through multiple awards.

## PUBLICATION

Trivedi G, Inoue D, Chen C, Bitner L, Chung YR, Taylor J, Gönen M, Wess J, Abdel-Wahab O, Zhang L. 2019. Muscarinic acetylcholine receptor regulates self-renewal of early erythroid progenitor. *Sci Transl Med* **11:** eaaw3781.

*In Press*

Al-Abed Y, Zhang L. 2020. *Compositions and methods for treating anemia*. Patent filed.
Chen C, Galeev R, Flygare J, Zhang L. 2020. Functional genomics systematically identify novel regulators in early erythroid progenitor self-renewal (in prep.).
Chen C-C, Li B, Millman S, Chen C, Li X, Morris J IV, Mayle A, Ho Y-J, Loizou E, Liu H, et al. 2020. Vitamin $B_6$ addiction in acute myeloid leukemia. *Cancer Cell* **37:** 71–84.
Li B, Lowe S, Zhang L. 2020. Vitamin $B_6$ pathway synergizes with a gene to target acute myeloid leukemia (in prep.).
Lowe S, Zhang L, Chen C. 2020. *Compositions and methods for treating acute myeloid leukemia*. Patent filed.
Mishra S, Millman S, Lowe S, Zhang L. 2020. Metabolic mechanistic insights and therapeutic targets in acute myeloid leukemia (in prep.).
Trivedi G, Abdel-Wahab O, Zhang L. 2020. A receptor is a cell surface marker that defines "hematopoietic reflex" (in prep.).

# AUTHOR INDEX

WATSON SCHOOL OF
BIOLOGICAL SCIENCES

# WATSON SCHOOL OF BIOLOGICAL SCIENCES

## ADMINISTRATION

**Alexander Gann, Ph.D.**, WSBS Professor and Dean
**Monn Monn Myat, Ph.D.**, Associate Dean
**Alyson Kass-Eisler, Ph.D.**, Postdoctoral Program Officer and Curriculum Director
**Kimberly Creteur, M.Ed., M.S.Ed.**, Admissions and Recruitment Manager
**Kimberley Graham,** Administrative Assistant

### EXECUTIVE COMMITTEE

**Chair**
Alexander Gann

**Members**
Camila dos Santos
Jesse Gillis
John Inglis
Bo Li
David L. Spector (Director of Research)
Anthony Zador

**Student Representatives**
Diogo Maia e Silva, WSBS
Zhezhen Yu, SBU

**Secretary**
Monn Monn Myat

### ADMISSIONS COMMITTEE

**Chair**
Adrian Krainer

**Members**
Alexander Gann (*ex officio*)
Mikala Egeblad
Leemor Joshua-Tor
Bo Li
Zachary Lippman
W. Richard McCombie
Stephen Shea
Adam Siepel
Christopher Vakoc
Linda Van Aelst

**Secretary**
Kimberly Creteur

### QUALIFYING EXAM COMMITTEE

**Chair**
Linda Van Aelst

**Members**
Molly Hammell
W. Richard McCombie
Stephen Shea
Christopher Vakoc

### EXTERNAL ADVISORY COMMITTEE

**Keith Yamamoto (Chair)**
Executive Vice Dean, School of Medicine
University of California, San Francisco

**Victor Corces**
Professor, Department of Biology
Emory University
Professor, Howard Hughes Medical Institute

**Gail Mandel**
Senior Scientist, Vollum Institute
Oregon Health and Science University
Investigator, Howard Hughes Medical Institute

**Barbara Meyer**
Professor of Genetics and Development
University of California, Berkeley
Investigator, Howard Hughes Medical Institute

**Frank Solomon**
Professor, Department of Biology and Center for Cancer
   Research
Massachusetts Institute of Technology

# WATSON SCHOOL OF BIOLOGICAL SCIENCES DEAN'S REPORT

On May 19, 2019, we celebrated the Watson School's 16th graduation ceremony. Seven students were awarded Ph.D. degrees: Talitha Forcier from the Entering Class of 2012, Giorgia Battistoni, Lital Chartarifsky, Sanchari Ghosh, Michael Gutbrod, and Laura Maiorino from the Entering Class of 2013, and Sashank Pisupati from the Entering Class of 2014. Honorary degrees were awarded to Drs. Lilian Clark and Robert Tjian, who also gave the commencement address.

Dr. Lilian Clark received a bachelor's degree from University of Glasgow and her Ph.D. from the University of St. Andrews. She pursued postdoctoral research at Memorial Sloan Kettering Cancer Center, the Dana-Farber Cancer Institute, and the Imperial Cancer Research Fund in London. Dr. Clark took a position as administration manager for research at the ICRF's London Research Institute, while working toward her M.B.A. degree from the University of Westminster. In 1999, she was appointed assistant dean of the newly established Watson School of Biological Sciences at CSHL and was subsequently promoted to associate dean, and ultimately dean. Invaluably, she set in place at the outset many of the procedures and guidelines that are still in use at the School today and are critical in ensuring it functions as a dynamic and cutting-edge accredited educational institution. In August 2007 Dr. Clark returned to the United Kingdom to take up the position of Executive Director for Science at Cancer Research UK and remained there until her retirement at the end of 2010. In retirement she is still actively involved in cancer research as a member of Cancer Research UK's Early Diagnosis and Advisory Group and as a member of research teams at Guy's Hospital and Imperial College in London.

Dr. Robert Tjian received his bachelor's degree in biochemistry from UC Berkeley and his Ph.D. from Harvard. He completed a postdoctoral fellowship at CSHL with James Watson before joining the UC Berkeley faculty. Dr. Tjian studies the biochemical process involved in controlling how genes are turned on and off, key steps in the process of decoding the human genome and regulating how genetic information is transcribed and translated into the thousands of protein biomolecules that keep cells, tissues, and organisms alive. He is a member of the National Academy of Sciences and has received many awards honoring his scientific contributions. He was named an HHMI investigator in 1987 and California Scientist of the Year in 1994. He has served in various scientific leadership positions including President of the Howard Hughes Medical Institute, one of the largest private nonprofit organizations that supports basic research. In 1992 he cofounded a biotech company, Tularik Inc., focused on controlling gene expression by using small chemical drugs to tackle diseases such as cancer, diabetes, inflammation, and HIV infection. Dr. Tjian cofounded The Column Group (TCG), which has successfully launched more than 20 companies to address medical needs in diabetes, cancer, heart disease, neurodegeneration, obesity, and regenerative medicine.

### 2019 WSBS DOCTORAL RECIPIENTS

| Student | Thesis advisor | Academic mentor | Current position |
|---|---|---|---|
| Giorgia Battistoni | Christopher Hammell | Gregory Hannon | Research Associate, Cancer Research UK (Advisor: Gregory Hannon) |
| Emilis Bruzas | Mikala Egeblad | Alea Mills | Medical Director, Biolumina, New York |

| | | | |
|---|---|---|---|
| **2019 WSBS DOCTORAL RECIPIENTS** (*continued*) | | | |
| **Student** | **Thesis advisor** | **Academic mentor** | **Current position** |
| Sanchari Ghosh | Anthony Zador | Josh Dubnau | Graduate Student Post, CSHL |
| Jacqueline Giovanniello | Bo Li | Bruce Stillman | Postdoctoral Fellow, University of California, Los Angeles (Advisor: Kate Wassum) |
| Kristina Grigaityte | Mickey Atwal | John Inglis | Graduate Student Post, CSHL |
| Michael Gutbrod | Robert Martienssen | Zachary Lippman | Postdoctoral Fellow, Massachusetts Institute of Technology (Advisor: Manolis Kellis) |
| Laura Maiorino | Mikala Egeblad | Nicholas Tonks | Postdoctoral Fellow, Koch Institute, Massachusetts Institute of Technology (Advisor: Darrell Irvine) |
| Sashank Pisupati | Anne Churchland | Stephen Shea | Postdoctoral Fellow, Princeton University (Advisor: Yael Niv) |
| Jue Xiang Wang | Hiro Furukawa | Mikala Egeblad | Consultant, Boston Consulting Group, New York |
| Georgi Yordanov | David Tuveson | Leemor Joshua-Tor | Equity Research Associate, Cowen and Company, New York |



2019 Graduates: (*Left to right*) Laura Maiorino, Michael Gutbrod, CSHL President and CEO Bruce Stillman, Giorgia Battistoni, CSHL Chairman Marilyn Simons, WSBS Dean Alexander Gann, Lital Chartarifsky-Lynn, Sanchari Ghosh, Talitha Forcier, Sashank Pisupati, honorary degree recipient Lilian Clark, and honorary degree recipient Robert Tjian.

## 2019 THESIS DISSERTATION DEFENSES

### ENTERING CLASS OF 2013

**Laura Maiorino,** February 12, 2019

*Identification of cancer-cell intrinsic and host-mediated cellular mechanisms of metastasis through intravital imaging*

**Thesis Examining Committee**

| | |
|---|---|
| Chairperson: | **Douglas Fearon** |
| Research Mentor: | **Mikala Egeblad** |
| Academic Mentor: | **Nicholas Tonks** |
| Committee Member: | **Linda Van Aelst** |
| External Examiner: | **Andrew Ewald** |
| | *The Johns Hopkins University* |
| | *School of Medicine* |

**Michael Gutbrod,** April 25, 2019

*Chromosomal functions of RNA interference and small RNAs in early mammalian development*

**Thesis Examining Committee**

| | |
|---|---|
| Chairperson: | **David Spector** |
| Research Mentor: | **Robert Martienssen** |
| Academic Mentor: | **Zachary Lippman** |
| Committee Member: | **Robert Martienssen** |
| External Examiner: | **Geneviève Almouzni** |
| | *Institut Curie, France* |

**Sanchari Ghosh,** May 7, 2019

*Corticostriatal plasticity underlying learning and reversal of auditory-motor associations*

**Thesis Examining Committee**

| | |
|---|---|
| Chairperson: | **Stephen Shea** |
| Research Mentor: | **Anthony Zador** |
| Academic Mentor: | **Josh Dubnau** |
| Committee Member: | **Adam Kepecs** |
| External Examiner: | **Maria Geffen** |
| | *University of Pennsylvania* |

**Giorgia Battistoni,** May 17, 2019

*A novel perspective on heterogeneity during development and epigenetic reprogramming of the mouse male germline*

**Thesis Examining Committee**

| | |
|---|---|
| Chair: | **Linda Van Aelst** |
| Research Mentor: | **Gregory Hannon** |
| Academic Mentor: | **Christopher Hammell** |
| Committee Member: | **Robert Martienssen** |
| External Examiner: | **Samuel Aparicio** |
| | *BC Cancer Research Centre, Canada* |

**Georgi Yordanov,** July 3, 2019

*Myc and Spdef: two transcription factors hijacked by pancreatic cancer*

**Thesis Examining Committee**

| | |
|---|---|
| Chair: | **Douglas Fearon** |
| Research Mentor: | **David Tuveson** |
| Academic Mentor: | **Leemor Joshua-Tor** |
| Committee Member: | **Adrian Krainer** |
| External Examiner: | **Gerard Evan** |
| | *University of Cambridge* |

### ENTERING CLASS OF 2014

**Sashank Pisupati,** May 10, 2019

*A value-based explanation for lapses in perceptual decisions*

**Thesis Examining Committee**

| | |
|---|---|
| Chair: | **Bo Li** |
| Research Mentor: | **Anne Churchland** |
| Academic Mentor: | **Stephen Shea** |
| Committee Member: | **Anthony Zador** |
| External Examiner: | **Joshua Gold** |
| | *University of Pennsylvania* |
| | *School of Medicine* |

**Emilis Bruzas,** July 22, 2019

*The role of immunity in breast cancer dormancy and recurrence*

**Thesis Examining Committee**

| | |
|---|---|
| Chair: | **David Spector** |
| Research Mentor: | **Mikala Egeblad** |
| Academic Mentor: | **Alea Mills** |
| Committee Member: | **Douglas Fearon** |
| Committee Member: | **Linda Van Aelst** |
| External Examiner: | **Alana Welm** |
| | *University of Utah School of Medicine* |

**Jacqueline Giovanniello,** October 4, 2019

*Characterizing amygdala-pallidal circuitry and its role in behavioral dysfunction in a 16p11.2 microdeletion model of ASD*

**Thesis Examining Committee**

| | |
|---|---|
| Chair: | **Stephen Shea** |
| Research Mentor: | **Bo Li** |
| Academic Mentor: | **Bruce Stillman** |
| Committee Member: | **Anne Churchland** |
| External Examiner: | **Rebecca Shanksy** |
| | *Northeastern University* |

**Jue Xiang Wang,** October 30, 2019

*Structural basis of subtype-specific competitive antagonism for GluN2C/2D-containing NMDA receptors*

**Thesis Examining Committee**

| | |
|---|---|
| Chair: | **Linda Van Aelst** |
| Research Mentor: | **Hiro Furukawa** |
| Academic Mentor: | **Mikala Egeblad** |
| Committee Member: | **Leemor Joshua-Tor** |
| Committee Member: | **Bo Li** |
| External Examiner: | **Crina Nimigean** |
| | *Weill Cornell Medical College* |

### ENTERING CLASS OF 2015

**Kristina Grigaityte,** October 24, 2019

*Comprehensive sequencing analyses of high-throughput single T cells in humans*

**Thesis Examining Committee**

| | |
|---|---|
| Chair: | **Justin Kinney** |
| Research Mentor: | **Mickey Atwal** |
| Academic Mentor: | **John Inglis** |
| Committee Member: | **Douglas Fearon** |
| External Examiner: | **Aleksandra Walczak** |
| | *Ecole Normale Supérieure* |

## DOCTORAL THESIS RESEARCH

| Student | Academic mentor | Research mentor | Thesis research |
| --- | --- | --- | --- |
| **ENTERING CLASS OF 2014** | | | |
| **Hamza Giaffar**<br>*Robert and Theresa Lindsay Fellow* | Jan Witkowski | Alexei Koulakov | The primacy model of olfactory coding. |
| **Elizabeth Hutton**<br>*Elisabeth Sloan Livingston Fellow* | Molly Hammell | Adam Siepel | Functional variant prediction in noncoding regions. |
| **Colin Stoneking**<br>*NIH Predoctoral Trainee* | Zachary Lippman | Anthony Zador | Neuronal mechanisms enabling decision-making to be learned. |
| **Anqi Zhang**<br>*Starr Centennial Scholar* | Bo Li | Anthony Zador | From corticostriatal plasticity to a common pathway. |
| **ENTERING CLASS OF 2015** | | | |
| **Benjamin Berube**<br>*National Science Foundation Fellow*<br>*NIH Predoctoral Trainee*<br>*Elisabeth Sloan Livingston Fellow* | Zachary Lippman | Robert Martienssen | A single-cell assessment of germline epigenetic heterogeneity. |
| **Matt Lee**<br>*David H. Koch Fellow* | Nicholas Tonks | Lloyd Trotman | The road to metastasis: defining the initial stages of prostate cancer progression. |
| **Katarina Meze**<br>*Leslie C. Quick, Jr. Fellow* | Jay Lee | Leemor Joshua-Tor | Structural and functional studies of RNA regulatory mechanisms mediated by Lin28. |
| **Alexandra Nowlan**<br>*Genentech Fellow*<br>*George A. and Marjorie H. Anderson Fellow* | Jessica Tollkuhn | Stephen Shea | Multisensory experience dependent plasticity: network dynamics in auditory processing following parturition. |
| **Sofya Polyanskaya**<br>*Starr Centennial Scholar* | Alexander Krasnitz | Christopher Vakoc | Identification of fusion oncoprotein co-dependencies in cancer. |
| **Ngoc Tran**<br>*Samuel Freeman Fellow* | Leemor Joshua-Tor | Alexei Koulakov | Predicting bioactivity using a data-driven representation of three-dimensional chemical structures. |
| **ENTERING CLASS OF 2016** | | | |
| **Brianna Bibel**<br>*National Science Foundation Fellow* | Hiro Furukawa | Leemor Joshua-Tor | Structural and functional studies of phosphorylation-mediated regulation of the RNAi effector Argonaute. |
| **Alberto Corona**<br>*NIH Predoctoral Trainee*<br>*Hearst Foundation Fellow* | David Jackson | Stephen Shea | Identification of neural circuitry underlying paternal behaviors. |
| **David Johnson**<br>*Gilliam Fellow*<br>*National Science Foundation Fellow*<br>*Hearst Foundation Scholar* | Zachary Lippman | Alea Mills | Elucidating the role of BRPF1 in human glioblastoma multiforme. |
| **Christopher Krasniak**<br>*NIH Predoctoral Trainee* | Jan Witkowski | Anthony Zador | The role of cholinergic input to visual cortex in mouse spatial visual attention. |
| **Shaina Lu**<br>*Edward & Martha Gerry Fellow* | Leemor Joshua-Tor | Anthony Zador | Development of a high-throughput pipeline to study the relationship of neuron projections and gene expression underlying mouse models of neuropsychiatric disorders. |
| **Kathryn O'Neill**<br>*National Science Foundation Fellow*<br>*NIH Predoctoral Trainee* | Camila dos Santos | Molly Hammell | Investigations into TDP-43 mediated effects on sRNA biology. |
| **Luqun Shen**<br>*Edward & Martha Gerry Fellow* | David Stewart | Stephen Shea | Predicting bioactivity using a data-driven representation of three-dimensional chemical structures. |
| **Olya Spassibojko**<br>*David & Fanny Luke Fellow* | Jessica Tollkuhn | Ullas Pedmale | Molecular determinants controlling cryptochrome light signal transduction. |
| **Martyna Sroka**<br>*George A. and Marjorie H. Anderson Fellow* | Molly Hammell | Christopher Vakoc | Molecular dissection of the PAX3-FOXO1 fusion oncoprotein pathway in rhabdomyosarcoma. |

## DOCTORAL THESIS RESEARCH (*continued*)

| Student | Academic mentor | Research mentor | Thesis research |
|---|---|---|---|
| **Ran Yan**<br>*George A. and Marjorie H.*<br>*Anderson Fellow* | David Tuveson | Douglas Fearon | Identification of endogenous antigen-specific T cells in pancreatic cancer metastasis. |
| **Chengxiang (Charlie) Yuan**<br>*A\*STAR Fellow* | Nicholas Tonks | Jay Lee | Linking the cell cycle and developmental fate specification. |

### ENTERING CLASS OF 2017

| Student | Academic mentor | Research mentor | Thesis research |
|---|---|---|---|
| **Lyndsey Aguirre**<br>*National Science Foundation Fellow*<br>*Hearst Foundation Fellow* | Ullas Pedmale | Zachary Lippman | Decoding *cis*-regulatory control of quantitative trait variation in tomato. |
| **Sara Boyle**<br>*NIH Predoctoral Trainee* | Jessica Tollkuhn | Bo Li | Can the central amygdala's interaction with midbrain dopamine areas control motivated behavior? |
| **Jordan (Bruno) Gegenhuber**<br>*Charles A. Dana Fellow*<br>*John & Amy Phelan Fellow*<br>*NIH Predoctoral Trainee* | John Inglis | Jessica Tollkuhn | Gene-regulatory mechanisms underlying brain organization by perinatal estradiol. |
| **Benjamin Harris**<br>*Crick-Clay Fellow* | W. Richard McCombie | Jesse Gillis | Robust analysis of single-cell RNA sequencing data to study development and emerging model organisms. |
| **Yuzhao (Richard) Hu**<br>*George A. and Marjorie H.*<br>*Anderson Fellow* | Justin Kinney | Ullas Pedmale | Characterizing cryptochrome 2 downstream targets in *Arabidopsis.* |
| **Dennis Maharjan**<br>*John & Amy Phelan Fellow* | Florin Albeanu | Anthony Zador | Role of corticostriatal potentiation in auditory decision-making. |
| **Diogo Maia e Silva**<br>*Robert and Theresa Lindsay Fellow*<br>*Boehringer Ingelheim Fonds Fellow* | Bruce Stillman | Christopher Vakoc | Defining the molecular origins of squamous pancreatic ductal adenocarcinoma. |
| **Cole Wunderlich**<br>*NIH Predoctoral Trainee* | Adam Siepel | Molly Hammell | Exploring transposable element expression in ALS at single-cell resolution. |

### ENTERING CLASS OF 2018

| Student | Academic mentor | Research mentor | Thesis research |
|---|---|---|---|
| **King Hei (Teri) Cheng**<br>*Robert and Theresa Lindsay Fellow* | Adrian Krainer | Robert Martienssen | Investigating the molecular mechanism of RNAi in resolving transcription–replication conflicts. |
| **Danielle Ciren**<br>*Robert and Theresa Lindsay Fellow* | Ullas Pedmale | Zachary Lippman | Revealing regulatory elements and their interactions in the control of gene expression and quantitative traits in plants. |
| **Marie Dussauze**<br>*Florence Gould Fellow*<br>*Annette Kade Fellow* | Stephen Shea | Florin Albeanu | Understanding algorithms and underlying neuronal substrates of sensorimotor integration in closed-loop olfaction. |
| **Ilgin Ergin**<br>*Starr Centennial Scholar* | Thomas Gingeras | Semir Beyaz | Mechanistic dissection of dietary regulation of T-cell function. |
| **Connor Fitzpatrick**<br>*Ainslie Family Fellow* | Robert Martienssen | Christopher Vakoc | Elucidating the mechanism of transcriptional activation by achaete-scute homolog 1. |
| **Amritha Varshini Hanasoge Somasundara**<br>*Bristol-Myers Squibb Fellow* | Leemor Joshua-Tor | Camila dos Santos | Investigating the role of the mammary immune microenvironment in Brca1-associated oncoprotection after pregnancy. |
| **Asad Aziz Lakhani**<br>*Starr Centennial Scholar* | David Tuveson | Jason Sheltzer | Dissecting the role of recurrent hyperploidy in tumorigenesis. |
| **Ziyi Mo**<br>*Gladys and Roland Harriman*<br>*Foundation Fellow* | David McCandlish | Adam Siepel | A flexible deep learning framework for inferring parameters of selection based on the ancestral recombination graph. |
| **Alexa Pagliaro**<br>*Jordan and Thomas A. Saunders III*<br>*Neuroscience Fellow* | John Inglis | Stephen Shea | Parvalbumin-positive interneuron activity during maternal behavior in a mouse model of Rett syndrome. |
| **Jenelys Ruiz**<br>*NIH Predoctoral Trainee* | Molly Hammell | Jessica Tollkuhn | The molecular and neural substrates of social attachment. |
| **Jonathan Werner**<br>*National Science Foundation Fellow* | Adam Siepel | Jesse Gillis | Revealing cellular lineage and sex-specific gene expression through transcriptional analysis of the X-chromosome. |

## Teaching Award

Dr. Mikala Egeblad, the co-lead instructor of the Specialized Disciplines course in Cancer, was chosen by the first-year students for the 15th Winship Herr Award for Excellence in Teaching. Here is what two students said in their nominations: "This year's cancer course was incredibly well-designed. The connections between topics and the topics themselves made for an easy-to-follow and comprehensive introduction to cancer as a scientific field." And, "Mikala's lectures were very well thought out and organized, and pitched at the perfect level for a class filled with people with diverse backgrounds and interests… . The journal discussions were also very well thought out—each journal directly related to one of the lectures that we had just had."



M. Egeblad

## 20th Anniversary Reunion

On May 19 and 20 the School celebrated the 20th anniversary of the first entering class. More than 50 alumni, former Deans, and administrators returned to campus and, together with current students, enjoyed a banquet and mini-symposium commemorating this important milestone in the School's history.

## Faculty Changes

Two new faculty members joined the Watson School in 2019: Saket Navlakha, an associate professor, and Peter Koo, an assistant professor. A new fellow also joined CSHL in 2019, Hannah Meyer.

Saket earned his Ph.D. in Computer Science from the University of Maryland and conducted postdoctoral research at Carnegie Mellon University. Before coming to CSHL he was an associate Professor of Integrative Biology at the Salk Institute for Biological Studies. His laboratory



20th Anniversary reunion

uses computational algorithms to integrate large, noisy, and heterogeneous biological data sets to understand biological function and dysfunction. He has most recently focused on studying neural circuit computation and plant architecture optimization from this perspective.

Peter earned his Ph.D. in Physics at Yale University. Before coming to CSHL he was a post-doctoral fellow in computational biology at Harvard University. Peter develops machine-learning methods rooted in statistical physics that discover generalizable knowledge from large, noisy biological data sets. His applied focus is genomics, in which computational methods are needed to understand how mutations in the genome alter gene regulation and protein function.

Hannah earned her Ph.D. at EMBL-EBI and the University of Cambridge, where she studied epigenomics and statistical genetics. Her lab at CSHL uses genomics and mathematical modeling to understand the spatial and temporal organization of promiscuous gene expression in the thymus and how antigen distribution in the thymus affects epithelial cell–T cell interaction and migration.

Saket and Peter have already participated in WSBS activities, including giving Research Topics talks to the first-year students. Peter also lectured in the Specialized Disciplines Course on Quantitative Biology. We look forward to their growing participation as members of the faculty.

## Admissions 2019

The School received more than 480 applications for the Entering Class of 2019—the largest number to date. We thank the Admissions Committee, which reviewed, interviewed, and selected candidates for our doctoral program. The Admissions Committee for the 2019 entering class comprised Adrian Krainer (Chair), Mikala Egeblad, Leemor Joshua-Tor, Bo Li, Zachary Lippman, W. Richard McCombie, Stephen Shea, Adam Siepel, Christopher Vakoc, Linda Van Aelst, and myself (ex officio).

## Entering Class of 2019

On August 19, 2019, the WSBS welcomed the 21st incoming class consisting of eight new students: Leah Braviner, Patrick Cunniff, Michael Passalacqua, Leonardo Jared Ramirez Sanchez, Nicole Sivetz, Ziqi (Amber) Tang, Shushan Toneyan, and Julia Wang.

---

**ENTERING CLASS OF 2019**

**Leah Braviner,** University of Leeds, B.Sc. (Hons, Industrial) Biochemistry (2019)
**Academic Mentor:** Linda Van Aelst

**Patrick Cunniff,** University of Notre Dame, B.S. in Biochemistry (2019)
**Academic Mentor:** David Jackson

**Michael Passalacqua,** Rice University, B.A. in Biological Sciences (2019)
**Academic Mentor:** Thomas Gingeras

**Leonardo Jared Ramirez Sanchez,** Universidad Nacional Autonoma de Mexico, B.S. in Genomic Sciences (2018)
**Academic Mentor:** Christopher Hammell

**Nicole Sivetz,** Monmouth University, B.S. in Biology, concentration in Molecular Cell Physiology/Chemistry (2019)
**Academic Mentor:** Camila dos Santos

**Ziqi (Amber) Tang,** University of North Carolina, Chapel Hill, B.S. in Biology/Computer Science (2019)
**Academic Mentor:** Alea Mills

**Shushan Toneyan,** University of Bern, M.S. in Bioinformatics and Computational Biology (2019), University of Oxford, B.A. in Biological Sciences (2017)
**Academic Mentor:** David Stewart

**Julia Wang,** Stanford University, B.S. in Computer Science (2019)
**Academic Mentor:** Tatiana Engel

2019 Entering Class: (*Left* to *right*) Shushan Toneyan, Michael Passalacqua, Nicole Sivetz, Julia Wang, Leah Braviner, Amber Tang, Patrick Cunniff, Leonardo Ramirez Sanchez.

## Academic Mentoring

The Watson School takes great pride in the mentoring that it offers its students. One example is our two-tiered mentoring approach, whereby each student chooses both an academic and a research mentor. The academic mentor is a critical advisor during the intensive coursework of the first term, during their rotations, and when identifying a suitable research mentor. Furthermore, the academic mentor continues to follow them throughout their doctoral experience, often serving as important advocates for the students. Entering students select, by mutual agreement, a member of the research or nonresearch faculty to serve as their academic mentor. This program continues to receive much support from the faculty who volunteer to be academic mentors, and it has rightfully become a vital ingredient in our success. The Academic Mentors for the Entering Class of 2019 are:

| STUDENT | MENTOR |
|---|---|
| Leah Braviner | Linda Van Aelst |
| Patrick Cunniff | David Jackson |
| Michael Passalacqua | Thomas Gingeras |
| Leonardo Ramirez Sanchez | Christopher Hammell |
| Nicole Sivetz | Camila dos Santos |
| Amber Tang | Alea Mills |
| Shushan Toneyan | David Stewart |
| Julia Wang | Tatiana Engel |

## Recruiting Efforts

This year we once again focused primarily on targeted visits to graduate fairs and minority conferences to identify prospective students for the School. In addition to these visits, the WSBS sends information to top undergraduate science departments around the world. A multiprogram booklet, incorporating the graduate, undergraduate, and postdoctoral programs, was updated for

this recruitment season. Additionally, e-mails were sent to personalized contacts and an electronic mailing list of more than 50,000 individuals who receive information from the Cold Spring Harbor Laboratory Press or have attended Meetings or Courses at the lab. We are grateful to these departments for sharing this contact list. We received 576 applications for the Entering Class of 2020, nearly 100 more than our previous record number in 2019, and it appears that many outstanding candidates have once again applied to the program.

### WATSON SCHOOL OF BIOLOGICAL SCIENCES 2019 RECRUITMENT SCHEDULE

| Event | Location | Date |
| --- | --- | --- |
| American Association of Cancer Research, Annual Meeting | Atlanta, Georgia | March 29–April 3 |
| University of California, Davis Biology Undergraduate Scholars Program (BUSP), Information Session | Davis, California | April 5 |
| California Forum for Diversity in Graduate Education Graduate School Fair | Davis, California | April 6 |
| National Conference on Undergraduate Research (NCUR), Annual Conference | Kennesaw, Georgia | April 11–13 |
| CSHL Diversity Recruitment Event, Open House | Cold Spring Harbor, New York | April 13 |
| Caldwell University, 3rd Annual Research and Creative Arts Day | Caldwell, New Jersey | April 24 |
| Cold Spring Harbor Laboratory, Open House | Cold Spring Harbor, New York | June 8 |
| Southeastern Association of Educational Opportunity Program Personnel (SAEOPP) McNair/SSS Scholars Research Conference, Annual Conference | Atlanta, Georgia | June 27–30 |
| NIH Graduate & Professional School Fair, Annual Meeting | Bethesda, Maryland | July 17 |
| University at Buffalo McNair Research Conference, Graduate School Fair | Niagara Falls, New York | July 18–20 |
| Hunter College MARC and MBRS/RISE, Information Session | New York, New York | September 11 |
| Howard University, Information Session | Washington, D.C. | September 17 |
| University of Maryland, Baltimore County, Meyerhoff Scholars Program, Graduate School Fair | Baltimore, Maryland | September 18 |
| Big 10+ Graduate School Expo, Graduate School Fair | West Lafayette, Indiana | September 22–23 |
| Rose-Hulman Institute of Technology, Graduate School Fair | Terre Haute, Indiana | September 23 |
| Norfolk State University, Applied Science Seminar, Information Session | Norfolk, Virginia | September 25 |
| Baylor University McNair Research Conference, Graduate School Fair | Waco, Texas | September 26–27 |
| Atlanta University Center Consortium: Clark Atlanta University, Morehouse College, and Spelman College, Graduate & Professional School Fair | Atlanta, Georgia | October 1 |
| California Forum for Diversity in Graduate Education, Graduate School Fair | Rohnert Park, California | October 12 (canceled) |
| California Polytechnic State University San Luis Obispo, Information Session | San Luis Obispo, California | October 14 |
| American Society for Human Genetics, Annual Meeting | Houston, Texas | October 15–19 |
| University at Albany, State University of New York, Information Session | Albany, New York | October 18 |
| University of Chicago, Information Session | Chicago, Illinois | October 18 |
| Society for Neuroscience Annual Meeting, Graduate School Fair | Chicago, Illinois | October 19–23 |
| Lehman College, Information Session | The Bronx, New York | October 21 |
| SACNAS Community College Day, Resource Fair | Honolulu, Hawaii | October 30 |
| Society for Advancement of Chicanos and Native Americans in Science (SACNAS), National Conference | Honolulu, Hawaii | October 31– November 2 |
| Mount Holyoke College, Information Session | South Hadley, Massachusetts | November 7 |
| Smith College, Information Session | Northampton, Massachusetts | November 8 |
| Williams College, Information Session | Williamstown, Massachusetts | November 8 |
| University of California, Riverside, Information Session | Riverside, California | November 12 |
| San Bernardino Community College, Information Session | San Bernardino, California | November 12 |
| Annual Biomedical Research Conference for Minority Students (ABRCMS), National Conference | Anaheim, California | November 13–16 |
| LaGuardia Community College, Research Symposium | Long Island City, New York | November 21 |
| American Society for Cell Biology, Annual Meeting | Washington, D.C. | December 7–11 |

## Students From Other Institutions

WSBS students account for approximately half of the total graduate student population here at CSHL; the other half constitutes visiting graduate students from other universities who have decided to conduct some or all of their thesis research in CSHL faculty members' laboratories. A large fraction of students are from Stony Brook University (SBU), via a program established between CSHL and SBU more than 40 years ago. Over the years we have built relationships with other institutions around the world, enabling their students to conduct research here at CSHL. Currently we have visiting students from institutions in the Netherlands, Poland, Russia, Spain, and the United States. The Watson School provides a contact person for the students and maintains relationships with the administrators from their home institutions. These students are fully integrated into the CSHL community and receive all the necessary assistance as they navigate the complexities of performing doctoral research away from their home institutions. The following students, listed in the box below, joined us from SBU this year.

| STUDENT | CSHL RESEARCH MENTOR | SBU PROGRAM |
| --- | --- | --- |
| Dexter Adams | Leemor Joshua-Tor | Genetics |
| Apoorva Arora | Adam Kepecs | Neuroscience |
| Katherine Broekman | Jessica Tollkuhn | Neuroscience |
| Charlie Chung | Semir Beyaz | Genetics |
| Craig Marshall | Molly Hammell | Genetics |
| Philip Moresco | Douglas Fearon | Genetics/MSTP |
| Chaoqun Yin | Anne Churchland | Neuroscience |
| Narges Zali | Bruce Stillman | Genetics |

## Graduate Student Symposium

Each year the students participate in two Graduate Student Symposia held at the Laboratory's Genome Research Center in Woodbury: one in May, the other in October/November. Each Symposium consists of senior students giving short talks, whereas coffee breaks and lunch provide opportunities for more informal interactions. The prizes for best talks for the May session were awarded to Martyna Sroka (WSBS, Vakoc lab), and Tzvia Pinkhasov (SBU, Kepecs lab), and for the November session was awarded to Matt Lee (WSBS, Trotman lab). We are grateful to the two student chairs—Kathryn O'Neill (WSBS) and Danilo Segovia (SBU)—for their hard work, and to WSBS's Kim Graham for helping make the Symposium a great success.

## Graduate Student and Postdoctoral Fellow Departures

With each year come not only new arrivals, but also departures. The following graduate students and postdoctoral fellows left the Laboratory during 2019:

POSTDOCTORAL FELLOWS

| | | | |
| --- | --- | --- | --- |
| Jean Albrengues | Chand Parvez Danka Mohammed | Wai Kit Ma | Sebastian Soyk |
| Sara Ballouz | Edgar Demesa Arevalo | Farzaneh Najafi | Sarah Starosta |
| Leah Bank | Jaynee Hart | Ellen Louise Noren Lindback | Yilin Tai |
| Prasamit Baruah | Jianping He | Torben Ott | Yusuke Tarumoto |
| Yael Berstein | Yifei Huang | Sapan Patel | Roger Tseng |
| Giulia Biffi | Anbalagan Jaganathan | Nikolay Rozhkov | Qingyu Wu |
| Amit Blumberg | Yinping Jiao | Subramanian Sankaranarayanan | Fang Xu |
| Irene Casanova Salas | James Jung | Christine Scaduto | |
| Quentin Chevy | Qian Li | Harrison Seidner | |
| Hannes Claeys | Jayon Lihm | Timothy Somerville | |

GRADUATE STUDENTS

| | | | |
|---|---|---|---|
| Naishitha Anaparthy | Jason Carter | Juan Huang | Tzvia Pinkhasov |
| Apoorva Arora | Songhwa Choi | Melissa Hubisz | Sashank Pisupati |
| Giorgia Battistoni | Angeliki Field-Pollatou | Abhay Kanodia | Colin Stoneking |
| Chloe Bizingre | Jacqueline Giovanniello | Navyateja Korimerla | Jue Xiang Wang |
| Emilis Bruzas | Michael Gutbrod | Laura Maiorino | |

## Executive Committee

The School's Executive Committee, in its monthly meetings, provides year-round direction for the School and its students through its invaluable policy recommendations. I wish to thank faculty members Camila dos Santos, Jesse Gillis, John Inglis, Bo Li, David Spector, and Anthony Zador for their service in 2019. I would also like to thank the student representatives Diogo Maia e Silva (WSBS) and Zhezhen Yu (SBU), who contributed to discussions and provided useful suggestions and feedback from their colleagues.

## The Watson School Continues to Benefit from Generous Benefactors

We are extremely grateful for the generous donors, whose one-time gifts or continued support made our 2019 programs possible, including the Annenberg Foundation, Lita Annenberg Hazen Foundation, Bristol-Myers Squibb Corporation, Mr. and Mrs. Landon Clay, Lester Crown, the Dana Foundation, Henriette and Norris Darrell, the Samuel Freeman Charitable Trust, the William Stamps Farish Fund, Mr. and Mrs. Alan Goldberg, Florence Gould Foundation, Gonzalo Río Arronte Foundation, William Randolph Hearst Foundation, Annette Kade Fund, Dr. James Karam, Mr. David H. Koch, Mr. and Mrs. Robert D. Lindsay and Family, Mr. and Mrs. David Luke III, Mr. and Mrs. William R. Miller, Estate of Edward L. Palmer, Mr. and Mrs. John C. Phelan, the Quick Family, Mr. and Mrs. Thomas A. Saunders III, Estate of Benjamin V. Siegel, Estate of Elisabeth Sloan Livingston, Joan Smith and Jason Sheltzer, the Starr Foundation, the Roy J. Zuckerberg Family Foundation, the Ainslie Foundation, and anonymous donors.

We are also grateful for our endowed lectureships: the John P. and Rita M. Cleary Visiting Lectureship, the George W. Cutting Lectureship, the William Stamps Farish Lectureship, the Martha F. Gerry Visiting Lectureship, the Edward H. Gerry Visiting Lectureship, the Edward H. and Martha F. Gerry Lectureship, the Susan T. and Charles E. Harris Visiting Lectureship, the Klingenstein Lectureship, the Mary D. Lindsay Lectureship, the Pfizer Lectureship, the George B. Rathmann Lectureship, the Seraph Foundation Visiting Lectureship, the Sigi Ziering Lectureship, the Daniel E. Koshland Visiting Lectureship, the Michel David-Weill Visiting Lectureship, and the Fairchild Martindale Visiting Lectureship.

We are also very fortunate to hold a prestigious Ruth L. Krischstein National Research Service Award Predoctoral Training Grant from the National Institutes of Health, National Institute of General Medical Sciences, which was competitively renewed for an additional five years in 2017.

## Student and Alumni Achievements

To date, 115 students have received their Ph.D. degree from the WSBS. Thirty graduates have now secured faculty positions: Twelve of them have already been promoted to associate professor, and two are full professors. Our graduates have also moved into influential positions in administration, publishing, consulting, industry, and medicine. In 2019, Emilis Bruzas was appointed as a medical director at Biolumina, Shu-Ling Chiu started her laboratory as an assistant professor at Academia Sinica in Taiwan, Charles Underwood joined the Max Planck Institute for Plant Breeding Research as a faculty member (group leader), Jue Xiang Wang is a consultant at Boston Consulting Group, Georgi Yordanov joined Cowen and Company as an equity research associate, and Petr Znamenskiy is an assistant professor at the Crick Institute.

In 2019, our current students and alumni were successful in receiving the following prestigious awards and fellowships:

- WSBS student **Lyndsey Aguirre** was awarded a National Science Foundation Graduate Research Fellowship.
- WSBS student **Diogo Maia e Silva** was awarded a predoctoral fellowship from the Boehringer Ingelheim Fonds.
- WSBS graduate **Kristen Delevich** received a NARSAD Young Investigator Award from the Brain and Behavior Foundation.
- WSBS graduate **Nilgun Tasdemir** received an NIH K99/R00 Pathway to Independence Award.

---

### 2019 WSBS STUDENT (CURRENT OR PREVIOUS) PUBLICATIONS

Anaparthy N, **Ho YJ**, Martelotto L, Hammell M, Hicks J. 2019. Single-cell applications of next-generation sequencing. *Cold Spring Harb Perspect Med* **9:** a026898.

Caiado F, **Maia-Silva D**, Jardim C, Schmolka N, Carvalho T, Reforço C, Faria R, Kolundzija B, Simões AE, Baubec T, et al. 2019. Lineage tracing of acute myeloid leukemia reveals the impact of hypomethylating agents on chemoresistance selection. *Nat Commun* **10:** 4986.

Carter JA, Preall JB, **Grigaityt K**, Goldfless SJ, Jeffery E, Briggs AW, Vigneault F, Atwal GS. 2019. Single T cell sequencing demonstrates the functional role of αβ TCR pairing in cell lineage and antigen specificity. *Front Immunol* **10:** 1516.

Chae H*, **Kepple DR**\*, Bast WG, Murthy VN, Koulakov AA, Albeanu DF. 2019. Mosaic representations of odors in the input and output layers of the mouse olfactory bulb. *Nat Neurosci* **22:** 1306–1317.

Chen X, Sun YC, Zhan H, **Kebschull JM**, Fischer S, Matho K, Huang ZJ, Gillis J, Zador AM. 2019. High-throughput mapping of long-range neuronal projection using in situ sequencing. *Cell* **179:** 772–786.

**Delás MJ**, Jackson BT, Kovacevic T, Vangelisti S, Munera Maravilla E, Wild SA, Stork EM, Erard N, Knott SRV, Hannon GJ. 2019. lncRNA *Spehd* regulates hematopoietic stem and progenitor cells and is required for multilineage differentiation. *Cell Rep* **27:** 719–729.

**Gegenhuber B**, Tollkuhn J. 2019. Signatures of sex: sex differences in gene expression in the vertebrate brain. *Wiley Interdiscip Rev Dev Biol* **20:** e348.

**Gegenhuber B**, Tollkuhn J. 2019. Sex differences in the epigenome: a cause or consequence of sexual differentiation of the brain? *Genes (Basel)* **10:** E432.

Humpton TJ, Alagesan B, DeNicola GM, Lu D, **Yordanov GN**, Leonhardt CS, Yao MA, Alagesan P, Zaatari MN, Park Y, et al. 2019. Oncogenic KRAS induces NIX-mediated mitophagy to promote pancreatic cancer. *Cancer Discov* **9:** 1268–1287.

**Kepple DR**, **Giaffar H**, Rinberg D, Koulakov AA. 2019. Deconstructing odorant identity via primacy in dual networks. *Neural Comput* **14:** 1–28.

**Maiorino L**, Egeblad M. 2019. Tumours pick the path to cancer inflammation. *Nat Cell Biol* **21:** 1055–1057.

Ott T*, **Masset P**\*, Kepecs A. 2019. The neurobiology of confidence: from beliefs to neurons. *Cold Spring Harb Symp Quant Biol* **83:** 9–16.

**Wang JX**, Furukawa H. 2019. Dissecting diverse functions of NMDA receptors by structural biology. *Curr Opin Struct Biol* **54:** 34–42.

*Authors contributed equally to the work. **Boldface indicates Watson School student.**

---

### WSBS GRADUATES IN FACULTY POSITIONS (IN ORDER OF COMPLETION)

| Graduate | Current Position |
| --- | --- |
| Amy Caudy | Associate Professor, University of Toronto, Canada |
| Ira Hall | Associate Professor, Washington University, St. Louis, Missouri |
| Niraj Tolia | Chief, Host–Pathogen Interactions and Structural Vaccinology Section Laboratory of Malaria Immunology and Vaccinology, National Institutes of Health |
| Patrick Paddison | Associate Member, Fred Hutchinson Cancer Research Center, Seattle, Washington |

*(continued)*

## WSBS GRADUATES IN FACULTY POSITIONS (IN ORDER OF COMPLETION) (*continued*)

| Graduate | Current Position |
| --- | --- |
| Elizabeth Bartom (nee Thomas) | Assistant Professor, Northwestern University, IL |
| Michelle Heck (nee Cilia) | Research Molecular Biologist, U.S. Department of Agriculture-Agricultural Research Service and Adjunct Associate Professor, Cornell University, Ithaca, New York |
| Zachary Lippman | Professor, Cold Spring Harbor Laboratory and Investigator, Howard Hughes Medical Institute |
| Ji-Joon Song | Associate Professor, Korea Advanced Institute of Science and Technology (KAIST), South Korea |
| Elena Ezhkova | Associate Professor, Mount Sinai School of Medicine, New York, New York |
| Masafumi Muratani | Professor, University of Tsukuba, Japan |
| Marco Mangone | Associate Professor, Arizona State University, Phoenix |
| Elizabeth Murchison | Reader, Cambridge University, United Kingdom |
| Hiroki Asari | Group leader, EMBL Monterotondo, Rome |
| François Bolduc | Associate Professor, University of Alberta, Edmonton, Alberta, Canada |
| Wei Wei | Associate Professor, University of Chicago, Illinois |
| Christopher Harvey | Associate Professor, Harvard University, Boston, Massachusetts |
| Tomas Hromadka | Group Leader, Institute of Neuroimmunology, Slovak Academy of Sciences, Slovakia |
| Monica Dus | Assistant Professor, University of Michigan, Ann Arbor |
| Shu-Ling Chiu | Assistant Professor, Academia Sinica in Taiwan |
| Daniel Chitwood | Assistant Professor, Michigan State University, East Lansing |
| Jeremy Wilsuz | Assistant Professor, University of Pennsylvania, Philadelphia |
| Oliver Fregoso | Assistant Professor, University of California, Los Angeles |
| Amy Leung | Assistant Research Professor, City of Hope's Beckman Research Institute, Duarte, California |
| Hiroshi Makino | Assistant Professor, Nanyang Technological University, Singapore |
| Katherine McJunkin | Stadtman Tenure Track Investigator, National Institutes of Health, Bethesda, Maryland |
| Yaniv Erlich | Associate Professor, New York Genome Center, Columbia University, New York, New York and Chief Science Officer, MyHeritage, Or Yehuda, Israel |
| Petr Znamenskiy | Assistant Professor, The Crick Institute, London, United Kingdom |
| Michael Pautler | Research Scientist, Vineland Research and Innovation Centre, Vineland Station, Ontario, Canada |
| Wee Siong Goh | Junior Principal Investigator/GIS Fellow, Genome Institute of Singapore |
| Charles Underwood | Group Leader, Max Planck Institute for Plant Breeding Research, Cologne, Germany |

## WSBS GRADUATES IN NONACADEMIC POSITIONS (IN ORDER OF COMPLETION)

| Graduate | Current Position |
| --- | --- |
| Emiliano Rial-Verde | Vice President, Food & Ingredients Strategy, Bunge Limited, New York |
| Ahmet M. Denli | Associate Editor, *Genome Research*, Cold Spring Harbor Laboratory |
| Rebecca Ewald | International Business Leader, Ventana Medical Systems/Roche, Tucson, Arizona |
| Catherine Seiler (nee Cormier) | Senior Manager Biosample Operations, Kaleido Biosciences, Lexington, Massachusetts |
| Beth Chen | Operations Manager at Homer Scientific Holdings Inc., Bellevue, Washington |
| Darren Burgess | Senior Editor, *Nature Reviews Genetics*, United Kingdom |
| Rebecca Bish-Cornelissen | Scientific Director, The Mark Foundation for Cancer Research, New York, New York |
| Angelique Girard | Director of Finance and Administration, Amplitude Studios, Paris, France |
| Allison Blum | Account Supervisor, LifeSci Public Relations, LLC, New York, New York |
| Keisha John | Director of Diversity Programs, University of Virginia, Charlottesville |
| Colin Malone | Co-Founder & Head of Biology at VNV NewCo, New York, New York |
| Oliver Tam | Computational Science Analyst, Cold Spring Harbor Laboratory |
| Amy Rappaport | Senior Scientist, Gristone Oncology, Emeryville, California |
| Frederick Rollins | Engagement Manager, LEK Consulting, Boston, Massachusetts |

**WSBS GRADUATES IN NONACADEMIC POSITIONS (IN ORDER OF COMPLETION)** (*continued*)

| Graduate | Current Position |
| --- | --- |
| Patrick Finigan | Senior Specialist, Regulatory Affairs CMC Biologics, Merck, Bridgewater, New Jersey |
| Elizabeth Nakasone | Heme-Onc Fellow, Fred Hutchinson Cancer Research Center, Seattle, Washington |
| Maria Pineda | Co-Founder, CEO, Envisagenics, New York, New York |
| Felix Schlesinger | Bioinformatics Scientist, Illumina, Inc., San Diego, California |
| Paloma Guzzardo | Associate Manager, R&D Planning, Strategy and Operations, Regeneron Pharmaceuticals, Eastview, New York |
| Saya Ebbesen | Associate Director, Medical + Scientific Strategy at BluPrint Oncology, London, United Kingdom |
| Joshua Sanders | Founder and C.E.O., Sanworks, L.L.C., Stony Brook, New York |
| Zinaida Perova | Bioimaging Data Scientist, European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom |
| Katie Liberatore | Research Scientist, Calyxt, Minneapolis, Minnesota |
| Kaja Wasik | Co-Founder, Gencove & Variant Bio, New York, New York |
| Stephane Castel | Co-Founder at Variant Bio and Senior Research Fellow at New York Genome Center, New York, New York |
| Mitchell Bekritsky | Senior Bioinformatics Scientist, Illumina, Inc., Cambridge, United Kingdom |
| Sang-Geol Koh | Scientist and Entrepreneur, {Mind}, South Korea |
| Susann Weissmueller | Strategic Partnering Associate, Roche, Switzerland |
| Ian Peikon | Venture Partner, Lux Capital, New York, New York |
| Cinthya Zepeda Mendoza | Laboratory Genetics and Genomics Fellow, Mayo Clinic, Rochester, Minnesota |
| Jack Walleshauser | Research Scientist at Plexxikon Inc., Berkeley, California |
| Colleen Carlston | Medical student, University of California, San Francisco, California |
| Lisa Krug | Scientist, Kallyope, New York, New York |
| Robert Aboukhalil | Bioinformatics Software Engineer, Invitae, San Francisco, California |
| Tyler Garvin | Head of Operations, Stockwell AI Inc., Oakland, California |
| Anja Hohmann | Senior Scientist, KSQ Therapeutics, Boston, Massachusetts |
| Matt Koh | Natural Language Processing Research Scientist, Bloomberg LP, New York, New York |
| Brittany Cazakoff | Law student, Stanford University, California |
| Annabel Romero Hernandez | Associate Scientist, Regeneron Pharmaceuticals, Tarrytown, New York |
| Maria Nattestad | Scientific Visualization Lead, DNA Nexus and Founder, OMGenomics, California |
| Onyekachi Odoemene | Senior Machine Learning Researcher, The Johns Hopkins University, Baltimore, Maryland |
| Daniel Kepple | Senior Machine Learning Engineer, Samsung Artificial Intelligence Center, New York |
| Lital Chartarifsky | Biotechnology and Business Development Coordinator, Cold Spring Harbor Laboratory |
| Georgi Yordanov | Equity Research Associate, Cowen and Company, New York, New York |
| Emilis Bruzas | Medical Director, Biolumina, New York, New York |
| Jue Xiang Wang | Consultant, Boston Consulting Group, New York, New York |

**WSBS GRADUATES IN POSTDOCTORAL POSITIONS (IN ORDER OF COMPLETION)**

| Graduate | Current Position |
| --- | --- |
| Charles Kopec | Associate Professional Specialist, Princeton University, Princeton, New Jersey (Advisor: Dr. Carlos Brody) |
| Claudia Feierstein | Research Associate, Champalimaud Neuroscience Programme, Lisbon, Portugal |
| Gowan Tervo | Research Specialist, Janelia Farms Research Campus, Ashburn, Virginia |
| Shraddha Pai | Postdoctoral Fellow, Centre for Addiction and Mental Health, Toronto, Ontario, Canada (Advisor: Gary Bader) |
| Galen Collins | Postdoctoral Fellow, Harvard Medical School, Boston, Massachusetts (Advisor: Dr. Alfred Goldberg) |
| David Simpson | Postdoctoral Fellow, Stanford University, California (Advisor: Dr. Alejandro Sweet-Cordero) |
| Claudio Scuoppo | Instructor, Columbia University, New York, New York (Advisor: Riccardo Dalla-Favera) |

## WSBS GRADUATES IN POSTDOCTORAL POSITIONS (IN ORDER OF COMPLETION) (*continued*)

| Graduate | Current Position |
| --- | --- |
| Kyle Honegger | Data Scientist, Ann & Robert H. Lurie Children's Hospital of Chicago, Illinois |
| Zhenxun Wang | Research Fellow, Genome Institute of Singapore (Advisor: Wai Leong Tam) |
| Eyal Gruntman | Research Scientist, Janelia Farm, Ashburn, Virginia (Advisor: Michael Reiser) |
| Megan Hogan (nee Bodnar) | Senior Research Scientist, NYU Langone Medical Center, New York |
| Yevgeniy Plavskin | Postdoctoral Fellow, New York University, New York (Advisor: Mark Siegal) |
| Hassana Oyibo | Postdoctoral Fellow, Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland (Advisor: Georg Keller) |
| Marek Kudla | Postdoctoral Fellow, University of California, Berkeley (Advisor: Nick Ingolia) |
| Melanie Eckersley-Maslin | Postdoctoral Fellow, Cambridge University, United Kingdom (Advisor: Wolf Reik) |
| Dario Bressan | Research Associate, Cancer Research UK, Cambridge, United Kingdom (Advisor: Gregory Hannon) |
| Philippe Batut | Postdoctoral Fellow, Princeton University, New Jersey (Advisor: Michael Levine) |
| Ozlem Aksoy (nee Mert) | Postdoctoral Fellow, University of California, San Francisco (Advisor: William Weiss) |
| Nilgun Tasdemir | Postdoctoral Associate, University of Pittsburgh, Pennsylvania (Advisors: Steffi Osterreich and Nancy Davidson) |
| Elvin Wagenblast | Postdoctoral Fellow, University of Toronto, Ontario, Canada (Advisor: John Dick) |
| Kristen Delevich | Postdoctoral Fellow, University of California, Berkeley (Advisor: Linda Wilbrecht) |
| Silvia Fenoglio | Postdoctoral Fellow, Massachusetts Institute of Technology, Cambridge (Advisor: Michael Hemann) |
| Arkarup Bandyopadhyay | Postdoctoral Fellow, New York University, New York (Advisor: Michael Long) |
| Justus Kebschull | Postdoctoral Fellow, Stanford University, California (Advisor: Liqun Luo) |
| Joaquina Delas Vives | Postdoctoral Fellow, Francis Crick Institute, London, United Kingdom (Advisor: James Briscoe) |
| Abram Santana | Postdoctoral Fellow, Harvard University, Boston, Massachusetts (Advisor: Joan Brugge) |
| Fred Marbach | Postdoctoral Fellow, Sainsbury Wellcome Centre, United Kingdom (Advisor: Marcus Stephenson-Jones) |
| Yu-Jui (Ray) Ho | Computational Biologist, Memorial Sloan Kettering Cancer Center, New York, New York (Advisor: Scott Lowe) |
| Paul Masset | Postdoctoral Fellow, Harvard University, Boston, Massachusetts (Advisor: Venkatesh Murthy and Naoshige Uchida) |
| Talitha Forcier | Postdoctoral Fellow, Cold Spring Harbor Laboratory (Advisor: Molly Hammell) |
| Laura Maiorino | Postdoctoral Fellow, Koch Institute, Massachusetts Institute of Technology, Cambridge (Advisor: Darrell Irvine) |
| Michael Gutbrod | Postdoctoral Fellow, Massachusetts Institute of Technology, Cambridge (Advisor: Manolis Kellis) |
| Giorgia Battistoni | Research Associate, Cancer Research UK, Cambridge, United Kingdom (Advisor: Gregory Hannon) |
| Sashank Pisupati | Postdoctoral Fellow, Princeton University, New Jersey (Advisor: Yael Niv) |
| Sanchari Ghosh | Graduate Student, Post, Cold Spring Harbor Laboratory (Advisor: Anthony Zador) |
| Jacqueline Giovanniello | Postdoctoral Fellow, University of California, Los Angeles (Advisor: Kate Wassum) |

Prizes for the best posters by a postdoctoral fellow and by a graduate student were awarded at the Laboratory's annual In-House Symposium held in November 2019. The poster session provides a forum for the postdoctoral fellows and students to show off their research and gives the entire scientific community a chance to come together and share ideas beyond the walls of their individual laboratories. This year, Martyna Sroka, a WSBS student from Chris Vakoc's laboratory, won the graduate student prize. The postdoctoral prize was shared by Andrea Moffitt from Dan Levy's laboratory and Johanna Syrjanen from Hiro Furukawa's laboratory.

**Alexander Gann**
*WSBS Professor and Dean*

# SPRING CURRICULUM

## TOPICS IN BIOLOGY

Each year, invited instructors offer week-long courses at the Banbury Conference Center exploring specialized topics outside the expertise of the Cold Spring Harbor Laboratory faculty. These courses include morning and evening lectures, as well as afternoon sessions during which students read assigned papers or work on problem sets and presentations. In Spring 2019 there were two courses: Immunology and Physical Biology of the Cell.

## Immunology

| | |
|---|---|
| **February 17–23** | **Attended by the entering classes of 2015 and 2017** |
| INSTRUCTOR | **Hidde Pleogh,** Boston Children's Hospital |
| VISITING LECTURERS | **Marco Colonna,** Washington University, St. Louis |
| TEACHING FELLOWS | **Djenet Bousbaine,** Harvard University<br>**Carlos Donado,** Harvard University<br>**Charles Evavold,** Harvard University |

Immunology focuses on understanding the mechanisms by which multicellular organisms defend themselves against external threats of microbial aggression and internal threats associated with genetic instability and cellular transformation. The course focused on the innate immune system and the adaptive immune system. Innate immunity defends against microbes by recognizing evolutionarily conserved molecular patterns; the adaptive immune system has enormous flexibility in molecular recognition but can also target itself to cause autoimmune diseases.

## Physical Biology of the Cell

| | |
|---|---|
| **March 25–30** | **Attended by the entering classes of 2016 and 2018.** |
| INSTRUCTOR | **Rob Phillips,** Caltech |
| VISITING LECTURERS | **Jané Kondev,** Brandeis University<br>**Hernan Garcia,** University of California, Berkeley<br>**Sanjoy Mahajan,** Massachusetts Institute of Technology |
| TEACHING FELLOWS | **Shahrzad Yazdi,** Massachusetts Institute of Technology |

The aim of this course was to provide a hands-on experience in the use of quantitative models as a way to view biological problems. The students began with "order of magnitude biology," showing how simple estimates can be exploited in biology. They showed how to construct simple models of

a variety of different biological problems, primarily using the tools of statistical mechanics. One of the key themes of the course was to show how physical biology unites and organizes topics in a fundamentally different way, often revealing that topics that are nearby in physical biology seem unrelated when viewed from the vantage point of molecular or cell biology. The instructors guided the students from start to finish on several modeling case studies.

# SPECIAL COURSES

## Optical Methods

| | |
|---|---|
| **January 7–10** | **Attended by the entering class of 2018** |
| INSTRUCTOR | **Florin Albeanu,** CSHL |
| TEACHING FELLOWS | **Walter Bast,** CSHL |
| | **Priyanka Gupta,** CSHL |

Optical imaging techniques are widely used in all areas of modern biological research. The aim for this course was to give students an introduction into widely used basic and advanced optical methods. Given the experimental nature of the topic, a central aim of the course was to offer students a practical hands-on experience. This included both the use of commercially available systems and, more importantly, a primer on custom building and adapting optical setups to address specific biological needs.

## Teaching Experience at the Dolan DNA Learning Center

**Entering Class of 2018**

| | |
|---|---|
| DIRECTOR | **David A. Micklos** |
| INSTRUCTORS | **Amanda McBrie**n (Lead) |
| | **Cristina Fernandez-Marco** |
| | **Elna Gottlieb** |
| | **Erin McKechnie** |
| | **Bruce Nash** |
| | **Sharon Pepenella** |

As science plays an increasing role in society, there is also an increasing need for biologists to educate nonscientists of all ages about biology. The doctoral program offers its students the opportunity to teach in the Laboratory's Dolan DNA Learning Center, where students teach laboratory courses to high school and middle school students. In so doing, they learn how to communicate with nonbiologists and to inspire and educate creative young minds. The teaching

module entailed pairs of students teaching one morning or afternoon a week for 12 weeks. In the initial weeks, the Dolan DNA Learning Center instructors taught the Watson School students the didactic process—it was not until the fifth week that the graduate students taught on their own. At the end of the 12 weeks, the students were very excited about their teaching experience.

# Laboratory Rotations

### Entering Class of 2018

The most important element of a doctoral education is learning to perform independent research. After the fall term courses, students participate in laboratory rotations; these provide students and faculty the opportunity to get to know each other and to explore possibilities for doctoral thesis research. At the end of each rotation, students make short presentations of their studies to the other students and their rotation advisors. These talks give students an opportunity to share their laboratory experiences and to practice giving scientific presentations. This year, 22 WSBS faculty members served as rotation mentors, some mentoring more than one student.

| ROTATION MENTORS | | | |
|---|---|---|---|
| Semir Beyaz | Doug Fearon | Justin Kinney | Jason Sheltzer |
| Anne Churchland | Jesse Gillis | Alex Koulakov | Adam Siepel |
| Alex Dobin | Josh Huang | Zachary Lippman | Jessica Tollkuhn |
| Camila dos Santos | David Jackson | Robert Martienssen | Chris Vakoc |
| Mikala Egeblad | Leemor Joshua-Tor | Ullas Pedmale | |
| Tatiana Engel | Adam Kepecs | Steve Shea | |

# FALL CURRICULUM

**Entering Class of 2019**

The students started the semester by attending boot camps in Molecular, Cellular, and Quantitative Biology to introduce them to the techniques and terminology that they encounter in subsequent courses. The Molecular and Cellular Biology boot camp featured seven lectures from faculty members Hiro Furukawa, Dick McCombie, and Lloyd Trotman and Associate Dean Monn Monn Myat and Microscope Facility Manager Ericka Wee. The Quantitative Biology boot camp lectures were given by Justin Kinney.

## CORE COURSES

## The Leslie C. Quick, Jr. Core Course on Scientific Reasoning and Logic

| INSTRUCTORS | Linda Van Aelst (Lead) | Leemor Joshua-Tor |
| --- | --- | --- |
| | Alexander Gann | Jessica Tollkuhn |
| | Christopher Hammell | |
| | | |
| GUEST LECTURERS | Hiro Furukawa | Robert Martienssen |
| | Adrian Krainer | Christopher Vakoc |
| | Bo Li | |
| | | |
| TEACHING ASSISTANTS | Sonali Bhattacharjee | Matt Jaremko |
| | Dhananjay Huilgol | Brian Kinney |

In this core course, which forms the heart of the curriculum, students (1) acquire a broad base of knowledge about the biological sciences, (2) learn the scientific method, and (3) learn how to think critically. The initial four to five modules are on a different general theme; in each, students read an assigned set of research articles and, at the end of the module, provided written answers to a problem set that guides them through several of the articles.

Twice weekly, students attend lectures related to the module's topic that include concepts and fundamental information as well as experimental methods. The students meet among themselves to discuss the assigned papers not covered by the problem set. Each week, students spend an evening discussing the assigned articles with faculty. In the final module of the course, students participated in a mock study section in which real National Institutes of Health R01 grants are reviewed and critiqued. This allows students to evaluate the questions before the answers are known, evaluate routes toward discovery before knowing where they will end, and make critical judgments about how to proceed in the face of an uncertain outcome.

In 2019, the module topics for this course were as follows:

| Topic | Instructor(s) |
| --- | --- |
| Gene Expression | Alex Gann |
| Gene Regulatory Logic and the Construction of Multicellular Organisms: Insights from Flies, Plants, and Worms | Christopher Hammell |
| The Brain: Wiring, Plasticity, and Maladaptation | Jessica Tollkuhn |
| Macromolecular Structure and Function | Leemor Joshua-Tor |
| Study Section | Linda Van Aelst |

# The Darrell Core Course on Scientific Exposition and Ethics

The 2019 Scientific Exposition and Ethics (SEE) core course was again separated into three distinct sections covering writing, oral communication, and ethics. As usual, the course hosted expert guest lecturers who covered topics including scientific presentations, ethical and legal responsibilities of scientists, and DNA profiling and postconviction appeals.

INSTRUCTORS
**David Jackson** (Lead)
**Sydney Gary**
**Rebecca Leshan**
**Jason Sheltzer**

GUEST LECTURERS
**Diane Esposito**
**Molly Hammell**
**Jackie Jansen**
**Alyson Kass-Eisler**
**Charla Lambert**
**Ullas Pedmale**
**Rachel Rubino**
**Richard Sever**

VISITING LECTURERS
**Keith Baggerly,** MD Anderson Cancer Center (retired)
**Susan Friedman,** The Innocence Project
**Radha Ganesan,** Alan Alda Center for Communicating Science
**Kendra Sirak,** Harvard University
**Anna Wexler,** University of Pennsylvania

This core course offered instruction in the fundamental elements of scientific exposition—writing skills and public speaking—and ethics. The ability to communicate effectively and to appreciate the intricacies of ethical issues are essential skills for biologists; both subjects were taught in a series of example-based lectures and discussion groups. Writing skills included the fundamentals of modern scientific English and the organization and preparation of papers, research abstracts, and grant applications. Oral presentation skills were taught by instructors with different modes of presentation. Together with instructors, students critiqued formal seminar presentations at the Laboratory. Instruction and discussions about ethics included the ethical implications of biological discovery on society as well as the nature and boundaries of ethical behavior of scientists and their rights and responsibilities. A primary objective of the course was that students consider exposition and ethics integral aspects of scientific research.

## Research Topics

ORGANIZERS    **Kimberley Graham**
              **Alyson Kass-Eisler**

This core course provided students with an in-depth introduction to the fields of research that the Laboratory scientists investigate. Students and faculty attended a weekly Research Topics seminar at which faculty members and CSHL fellows presented their current research topics and methods of investigation each Wednesday evening over dinner. The students learned how to approach important problems in biology. These seminars, together with the annual fall Laboratory In-House Symposium, provided students with a basis for selecting laboratories in which to do rotations.

## SPECIALIZED DISCIPLINES COURSES

The students in the Entering Class of 2019 took a total of four Specialized Disciplines courses this fall: Quantitative Biology, Genetics and Genomics, Cancer, and Systems Neuroscience.

## Quantitative Biology

**Throughout the semester**

INSTRUCTOR    **Justin Kinney** (Lead)

GUEST LECTURERS    **Alex Dobin**
                   **Tatiana Engel**
                   **Peter Koo**
                   **David McCandlish**
                   **Adam Siepel**

TEACHING ASSISTANTS    **Wei-Chia Chen**
                       **Hussein Hijazi**

Quantitative reasoning is a powerful tool for uncovering and characterizing biological principles, ranging from the molecular scale all the way to the ecological. With the advent of high-throughput technologies in genomics and neuroscience, it has become increasingly necessary for biological researchers to be able to analyze and interpret large data sets and frame biological hypotheses quantitatively. To this end, this course aimed to equip the students with a working knowledge of standard statistics and Python programming, as well as provide exposure to more advanced topics in machine learning, genomics, population genetics, neuroscience, and biophysics.

# Genetics and Genomics

**September 5–27**

| | |
|---|---|
| INSTRUCTOR | **Ullas Pedmale** (Lead) |
| GUEST LECTURER | **Zachary Lippman** |
| VISITING LECTURER | **James Birchler,** University of Missouri |
| TEACHING ASSISTANT | **Benjamin Roche** |

This course placed modern genetics and genomics into the context of classical genetics. History, technique, and perspective of genetic inference were described around four levels of analysis: forward genetics, natural genetic variation, gene interaction, and genomics. Emphasis was on integrating classical with modern questions of genetic analysis: How are genes mapped and "cloned"? How do gene mutations help to define biological processes? How are more complex traits genetically dissected into their component parts? What concepts and techniques are used to organize genes into pathways and networks? What defines a gene, and what gene variation exists in natural populations? What are the functional consequences of gene variation, and how is it detected? How are genomes organized and coordinately regulated? How can genomic information be cataloged, organized, and mined? These questions and concepts were fleshed out using examples from the literature.

# Cancer

**October 4–26**

| | |
|---|---|
| INSTRUCTORS | **Mikala Egeblad** (Co-lead) |
| | **Christopher Vakoc** (Co-lead) |
| GUEST LECTURERS | **Semir Beyaz** |
| | **Tobias Janowitz** |
| | **Jason Sheltzer** |
| | **David Tuveson** |

Cancer represents an increasing cause of morbidity and mortality throughout the world as health advances continue to extend the life spans of our populations. Our basic understanding of cancer has increased considerably since 1971, when United States President Richard Nixon initiated the "War on Cancer." Specific hypotheses developed from our knowledge of cancer biology are being tested in increasingly complex model systems ranging from cell culture to genetically engineered mouse models, and such investigations should prove invaluable in discovering new methodologies for the detection, management, and treatment of cancer in humans. Importantly, our ability to translate our knowledge of cancer biology into a health benefit for patients is now starting to take form.

   At the conclusion of this course, students should be able to elaborate an understanding of cancer as a pathobiological process that invades our bodies without offering any known benefit to the host, discuss how cancer progresses, and contemplate how to expand on the methods currently used to treat cancer. Students will also be able to design tractable methods to investigate

fundamental aspects of cancer biology and will be familiar with translational approaches to defeating cancer. Topics covered in this course included hallmarks of cancer, tumor progression, the cancer genome, microenvironment, tumor immunology, metastasis, and approaches to treating cancer, including targeted therapy.

## Systems Neuroscience

**September 23–November 15**

| INSTRUCTORS | **Stephen Shea** (Lead) |
| | **Florin Albeanu** |

| TEACHING ASSISTANT | **Priyanka Gupta** |

This course provided an overview of key aspects of neuroscience, with a focus on learning and plasticity from its cellular basis, through development, to systems and behavior. Both experimental and theoretical viewpoints were explored. There were three main components to the class: lectures, a problem set, and paper presentations.

# POSTDOCTORAL PROGRAM

PROGRAM DIRECTOR **Nicholas Tonks**

PROGRAM ADMINISTRATOR **Alyson Kass-Eisler**

An important measure of our postdoctoral program's success is the ability of postdoctoral fellows to secure positions after they complete their training. Recently our fellows accepted positions at Allen Brain Institute; Biogen; Case Western Reserve University; Chinese Academy of Agricultural Sciences; Fudan University, China; Garvan Institute, Australia; Inari Agriculture, Belgium; Icahn School of Medicine at Mount Sinai; Jefferies, NY; Kyoto University, Japan; Penn State University; Repare Therapeutics; Texas Tech University; The Johns Hopkins University; Novartis; University of Cambridge; University of Lausanne, Switzerland; and the U.S. Department of Agriculture.

## Postdoctoral Liaison Committee

The Postdoctoral Liaison Committee (PDLC), which is an elected group of postdoctoral fellows who communicate information and ideas between the administration and the postdoctoral community, continues to enhance CSHL's postdoctoral experience. The PDLC is essentially the voice of the community and holds regular meetings, and an annual Town Hall, with Dr. Bruce Stillman, CSHL President. The current PDLC members are Oliver Artz, Debarati Ghosh, Min Yao, and Sophia Zebell. The PDLC hosted a retreat on September 20 at the Banbury Conference Center. The agenda included research talks from CSHL postdocs; a mock chalk talk; a Keynote address from Dr. Carol Mason, Columbia University; a Poster Session; an Adobe Illustrator Workshop; and a barbecue. This year the retreat coincided with the annual National Postdoc Appreciation Week.

CSHL endeavors to prepare postdocs to be competitive for the scarce number of jobs available. It is increasingly becoming CSHL's role to introduce the diversity of career opportunities available and to provide the tools postdocs need to prepare for these positions. As a result, a number of events were organized with the assistance of the PDLC and career development groups.

## Bioscience Enterprise Club

The Bioscience Enterprise Club (BEC) disseminates information about nonacademic careers to the CSHL postdoc community. Topic areas include biotechnology, intellectual property, law, regulatory affairs, and venture capitalism. They hosted a number of guest speakers recently: Bell Labs, Google, and beyond: a conversation with Peter Weinberger; Sue Klapholz, Vice-President of Nutrition & Health at Impossible Foods Inc.; Nancy Ilaya, the Scientific Director of L'Oreal Scientific; and Duncan Bull, Life Sciences & Chemistry Patent Attorney at Kilburn & Strode. The biannual Beyond the Bench Symposium, which highlights speakers experienced in diverse science-based careers outside academia, took place this past summer. Trainees were encouraged to learn more about various companies, organizations, and career trajectories that are open to those with a science Ph.D. The keynote lectures were given by George Yancopoulos, Regeneron, and Puja Sapra, Pfizer.

BEC also hosts in-person visits to local companies. This year CSHL trainees visited Google headquarters, Regeneron's Science to Medicine Forum, and the Alexandria LaunchLabs.

### Academic Career Training

New career development programs, spearheaded by Jackie Jansen, Research Operations Project Manager, have been developed. These include courses, lectures, and workshops on scientific enrichment, career exploration, and transferable skills like leadership, mentorship, and communication. The first sessions in the career development course have included Research Statements and Cover Letters; Teaching and Diversity Statements; Finding Job Listings and Organizing Your Search; a CV workshop; a workshop on Individualized Development Plans (IDPs) and networking; and a Communicating with Your Mentor Workshop. The last two workshops were presented by Keith Micoli from New York University.

A workshop on applications, held specifically for postdocs who are applying for academic jobs this year, included a CSHL faculty committee to review individual research statement and cover letters. We also hosted a mock chalk talk, in which a postdoc on the academic job market practiced the chalk talk component of their interview in front of a CSHL faculty who acted as a mock hiring committee.

### Social Events

The Laboratory pays special attention to the social needs of the postdoctoral community. The vast majority of CSHL postdocs are not from Long Island and do not have a built-in social network. A postdoc barbecue was held at the beginning of summer and was a huge success. Monthly get-togethers (coffee and cookies in the cold months, and ice cream socials in the summer) get postdocs out of the lab for a few minutes to network and socialize. This year, we also organized a bowling night in Melville.

### The Science Alliance

All CSHL postdoctoral fellows and graduate students are enrolled in a special initiative of the New York Academy of Science (NYAS), the Science Alliance. The Science Alliance for graduate students and postdoctoral fellows is a consortium of universities, teaching hospitals, and independent research facilities in the New York City metro area that have formed a partnership with NYAS. The Alliance's aim is to provide career and professional development monitoring for postdoctoral fellows and graduate students in science and engineering, through a series of live events and a dedicated web portal. In addition, the Science Alliance offers graduate students and postdoctoral fellows the opportunity to network with their peers across institutions and with key leaders in academia and industry.

The Science Alliance meetings, courses and workshops this year discussed the following topics: Introductory Coding for Researchers; Using Informational Interviews to Find Your Dream Job; Build Your All Star Resume and LinkedIn Profile; Thinking about Teaching and Scientists Teaching Science Online Course; Grantsmanship for Students and Postdocs: Pathways to Individual Fellowship; Science Alliance Leadership Training (SALT); Risky Business—The Future of Biopharmaceutical Innovation; Success Stories: Overcoming the Barriers of an International Scientist in the USA; Communicating Science to Policy Makers; Cultural Awareness Workshop: Navigating Diversity in the Lab; Conveying Science through Art: A Public Engagement Workshop; Cultivating Your Network to Advance Your Career; Building Your Personal Brand; Empowering Your Journey: A Career Planning Workshop; Science Riot: A Night of Stand-Up Comedy; Communicating Science on Wikipedia; A Recruiter's Perspective—"Why We Need People in STEM"; Science Beyond Borders—How to Get Involved in International Science Policy; and Insights On Effective Lab Management. The biannual "What Can You Be with a Ph.D.," the largest career symposium in the country, took place in October.

# UNDERGRADUATE RESEARCH PROGRAM

PROGRAM DIRECTORS        **Jesse Gillis**
        **Christopher Hammell**

PROGRAM ADMINISTRATOR    **Kimberly Creteur**

Established 60 years ago, the CSHL Undergraduate Research Program (URP) provides undergraduates from around the world with hands-on undergraduate research training in biology. The 10-week program begins the first week of June. Several activities are implemented to ensure that URP participants transition smoothly into the Laboratory community and research. For example, during the first week the students attend various orientations and receive a guided historical tour of campus and all the facilities and resources available to them. The participants in URP work, live, eat, and play among CSHL scientists and have a very busy academic and social calendar throughout the remaining 9 weeks of the summer. The students receive training in scientific research, science communication, career preparation, and bioinformatics and computational biology, all while interacting socially with fellow program participants and members of the CSHL community in formal and informal activities. Some of the 2019 activities included Director's Tea, talks from program alumni, dinner with Dr. and Mrs. Stillman, volleyball games, designing the URP T-shirt, a Broadway show, scavenger hunt, and the ever-famous URP versus PI volleyball match and barbecue.

The students' scientific development is the most important component of the program. At the beginning of the summer, each URP writes an abstract and presents a talk on his or her proposed research. The URP participants work alongside scientists and become increasingly independent throughout the summer. Concluding the program in August, each URP student prepares a final report and presents their results in a 15-minute talk at the URP Symposium. As in previous years, the program directors and faculty mentors were highly impressed with the accomplishments of the URP students.



2019 Undergraduate Research Program participants

The following 18 students, selected from 864 applicants, took part in the 2019 program:

**Dominik Aylard**
Advisor: **Camila dos Santos**
Funding: **William Townsend Porter Foundation Scholar**
Aging and NKT cell inactivity decrease breast cancer prevention in advanced-age pregnancies.

**Alison Bashford**
Advisor: **Stephen Shea**
Funding: **30th Anniversary URP Scholar**
Instinct and altruism in pup-retrieving mice.

**Nathan Castro Pacheco**
Advisor: **Alex Dobin**
Funding: **William Shakespeare Fellowship**
Single-cell transcript isoform abundance estimation using an expectation maximization maximum likelihood (EM-ML) algorithm.

**Andrew Claros**
Advisor: **Florin Albeanu**
Funding: **Burroughs Wellcome Fellowship**
Cortical feedback from the olfactory cortex affects firing of mitral cells.

**Emma Cravo**
Advisor: **Anne Churchland**
Funding: **Dorcas Cummings Scholar**
The role of the parietal and frontal cortex during sensory-guided decision-making.

**Faniya Doswell**
Advisor: **Pavel Osten**
Funding: **Former URP Fund Scholar**
Comparative mapping of neuron populations in prairie voles and mice.

**Ahmet Doymaz**
Advisor: **Leemor Joshua-Tor**
Funding: **James D. Watson Fellow**
Structural study of exonuclease Dis3l2's RNA-unwinding activity.

**Jasmin Fleuranvil**
Advisor: **Lloyd Trotman**
Funding: **Garfield Fellowship**
The role of Axl as a putative regulator of migration and morphology in prostate cancer.

**Tara Gallagher**
Advisor: **Thomas Gingeras**
Funding: **University of Notre Dame URP Scholar**
A study of the role of RNase 1 in the processing of RNA in extracellular vesicles.

**Nathaniel Garry**
Advisor: **Adrian Krainer**
Funding: **Joan Redmond Read Fellowship**
SRSF3-regulated alternative splicing and nonsense-mediated mRNA decay in cancer.

**Owen Hughes**
Advisor: **Tatiana Engel**
Funding: **Katya H. Davey Fellowship**
Toward inference of nonstationary Langevin dynamics from spike data.

**Mackenzie Litz**
Advisor: **Alexei Koulakov**
Funding: **Libby Fellowship**
Understanding the organization of the nervous system: identifying patterns in neuronal responses to 3D molecular structure in the accessory olfactory system.

**Sarah Mantell**
Advisor: **Adam Kepecs**
Funding: **Von Stade Fellowship**
An investigation of the inverse comorbidity between neurodegenerative disorders and cancers.

**Blake Nelson**
Advisor: **David Spector**
Funding: **Alfred L. Goldberg Fellowship**
Examining the expression of MALAT1 long noncoding RNA in human breast tumor organoids.

**Samantha Rothberg**
Advisor: **Doreen Ware**
Funding: **30th Anniversary URP Scholar**
The effect of phosphorus regulatory genes on root system architecture in *Arabidopsis*.

**Charlotte Simpson**
Advisor: **Mikala Egeblad**
Funding: **Robert H.P. Olney Fellow**
The major signalling molecules involved in classical and nonclassical NETosis.

**Abraham Steinberger**
Advisor: **David Jackson**
Funding: **Former URP Fund Scholar**
RAMOSA3 and its potential RNA-binding protein interactors.

**Yin Yuan**
Advisor: **Christopher Vakoc**
Funding: **James D. Watson Undergraduate Scholar**
Defining critical residues of the POU homeodomain transcription factor OCT-11 that sustain tuft cell lung cancer growth.

# NORTHWELL MEDICAL RESEARCH PROGRAM (MRP)

PROGRAM ADMINISTRATOR          **Jessica Gotterer**

Through the CSHL and Northwell Health affiliation, a summer internship program has been created to give first-year Hofstra Northwell School of Medicine students with basic research experience the opportunity to spend a summer working in a CSHL lab and attending relevant seminars. To date, students have been offered positions in labs focusing on neuroscience, cancer, and the genetics of human disease. Students commit 8–10 weeks (roughly July–September) during the summer following their first year of course work to full-time research in a CSHL lab. The students work with their CSHL host PI to design a research project and present their work at the annual "Scholarship Day" at Hofstra University the following Fall.

The following students took part in the 2019 program:

| STUDENT | CSHL MENTOR |
|---|---|
| Elizabeth Beals | Semir Beyaz |
| Alan Gao | Anthony Zador |
| Emily Navlen Krasnow | Linda Van Aelst |
| Dina Moumin | Semir Beyaz/Marina Frimer (Northwell Health) |
| Andrew Seidner | Lingbo Zhang |

# PARTNERS FOR THE FUTURE

PROGRAM DIRECTOR                    **David Jackson**

PROGRAM ADMINISTRATOR              **Lynn Carmen**

The Partners for the Future Program, established in 1990, provides an opportunity for talented Long Island high school students to have hands-on experience in biomedical research at Cold Spring Harbor Laboratory. This highly competitive program is open to Long Island high school students in their junior year. Each high school science chairperson may nominate up to three students. The top candidates are interviewed by CSHL scientists. Students selected to the program are paired with a scientist mentor and spend a minimum of 10 hours per week, September through March of their senior year, conducting original research. At the conclusion, the students present their projects to an enthusiastic audience of the students, scientific mentors and colleagues, CSHL administrators, parents, and teachers. Although the students learn a great deal about modern biology and state-of-the-art research techniques, the main advantage of the program is that they are exposed to day-to-day life in a lab. Interacting with scientists and support staff, the students learn to define and pursue a research goal while solving problems that may occur along the way.

The following 2019–2020 Partners for the Future were chosen from among 70 nominations:

| NAME | HIGH SCHOOL | LABORATORY | MENTOR |
|------|-------------|------------|--------|
| William Borges | Roslyn | Lloyd Trotman | Matt Lee |
| Lily Coffin | Farmingdale | Hannah Meyer | Hannah Meyer |
| Ronit Dhulia | Syosset | Jason Sheltzer | Jason Sheltzer |
| Erhumwunse Eghafona | Walter G. O'Connel Copiague | Douglas Fearon | Min Yao |
| Segal Gupta | Hicksville | Josh Huang | Hemanth Mohan |
| Sophia Jang | Syosset | Ullas Pedmale | Oliver Artz |
| Ruth Lee | Commack | Robert Martienssen | Hyun Soo Kim |
| Marcus Llorente | Island Trees | Christopher Vakoc | Diogo Maia e Silva |
| Sonia Sarju | Sanford H. Calhoun | Tatiana Engel | James Roach/Yanliang Shi |
| Elyse Schetty | Cold Spring Harbor | Anthony Zador | Shina Lu |
| Sarah Shao | Manhasset | Semir Beyaz | Semir Beyaz |
| Allison Taub | Plainview Old Bethpage JFK | Doreen Ware | Lifang Zhang |
| Z'Dhanne Williams | Jackson Amityville Memorial | David Jackson | Fang Xu |
| Jiacheng Yin | The Stony Brook School | Adrian Krainer | Kuan-Ting Lin |

MEETINGS & COURSES
PROGRAM

# ACADEMIC AFFAIRS

The Cold Spring Harbor Laboratory (CSHL) Meetings & Courses program attracts scientists from all over the world to communicate, learn, and critique the latest ideas and approaches in the biological sciences. The Meetings & Courses program attracted strong attendance in 2019, with 7,500 meeting participants, several hundred remote participants, and almost 1,500 course participants (trainees, teaching, and support faculty). The Cold Spring Harbor Asia program included 15 conferences and two summer schools, attracting almost 2,700 participants, bringing the anticipated year-end total for both the United States– and China-based programs to almost 12,000.

The Laboratory held 28 academic meetings this year, which brought together scientists from around the world to discuss their latest research, and several additional ancillary conferences. The spring meeting season culminated in the 84th Cold Spring Harbor Symposium, which focused on RNA Control and Regulation and addressed how central the RNA molecule is in so many areas of the biology of the cell, and increasingly how we think about manipulating RNA in a variety of ways to treat disease. The symposium attracted more than 400 participants, including notable scientists such as David Bartel, Robert Darnell, Caroline Dean, Jennifer Doudna, Gideon Dreyfuss, Susan Gasser, Adrian Krainer, Ruth Lehmann, James Manley, Christine Mayr, Joshua Mendell, Roy Parker, Phillip Sharp, Joan Steitz, and Feng Zhang, to name but a few. Sadly, MRC scientist Kiyoshi Nagai was due to speak at the symposium, but he fell ill and passed away several months later. Dissemination includes the proceedings of the symposium, published each year by the CSHL Press, and videotaped interviews with leading speakers conducted by editors and journalists attending the symposium now available on our Leading Strand YouTube channel. The symposium therefore reaches a much wider audience nationally and internationally than can possibly attend.

CSHL meetings are unique in that organizers are encouraged to select talks from abstracts that are submitted on the basis of scientific merit, ensuring that the latest findings will be presented and that young scientists will have the chance to describe their work. This year saw the continuation of many successful annual and biennial meetings as well as the introduction of several new meetings, including Systems Immunology, Microbiome, and Zebrafish Neural Circuits and Behavior. The CSHL Genentech Center Conferences on the History of Molecular Biology and Biotechnology this year addressed Yeast Research: Origins, Insights, Breakthroughs. The program featured many notable speakers, including Angelika Amon, Rochelle Esposito, Gerald Fink, Daniel Gottschling, Alan Hinnebusch, Nancy Kleckner, Douglas Koshland, Edward Marcotte, Kim Nasmyth, Maynard Olson, Rodney Rothstein, Randy Schekman, and many more. Partial support for individual meetings is provided by grants from the National Institutes of Health, National Science Foundation, other foundations, and companies. Core support for the meetings program is provided by the Laboratory's Corporate Sponsor Program.

The Courses program covers a diverse array of topics in molecular biology, neurobiology, structural studies, and bioinformatics. Instructors update their course curricula annually, invite new speakers who bring a fresh perspective, and introduce new techniques and experimental approaches based on student feedback and progress in the field. New techniques—for example, genome editing using tools such as CRISPR or super-resolution microscopy—are introduced as methodologies develop and evolve. We strongly encourage each course to include the latest technical and conceptual developments in their respective fields. Instructors, lecturers, and assistants come from universities, medical schools, research institutes, and companies around the world to teach at CSHL. Their excellence and dedication make the course program work uniquely well. We would especially like to thank Drs. Harmen Bussemaker, Tamara Caspary,

James Chappell, Carl Cohen, Suzanne Cohen, Sean Cutler, Alex Keene, Michael Orger, Lucy Palmer, Uta Paszkowski, and Elcin Unal, whose exemplary teaching and leadership of their respective courses have benefitted so many young scientists. We were also saddened to hear of the sudden passing of James Taylor in spring 2020, who contributed so much, including training many computational biologists through the Computational Genomics course at CSHL for many years. Students include advanced graduate students, postdoctoral trainees, principal investigators, and senior scientists from around the world.

Grants from a variety of sources support the courses. The core support grants provided through the Helmsley Charitable Trust and Howard Hughes Medical Institute are critical to our course program. The courses are further supported by multiple awards from the National Institutes of Health and the National Science Foundation, and additional support for individual courses is provided by various foundations and other sources. The courses also depend on equipment and reagents that are loaned or donated by a large number of companies—partnerships that are invaluable in ensuring that the courses offer training in the latest technologies (see below).

Now in its 10th year of operation, the Cold Spring Harbor Asia (CSH Asia) program, under the direction of Dr. Maoyen Chi, is headquartered at the Suzhou Dushu Lake Conference Center in the Suzhou Innovation Park high-technology suburb. In 2019, 15 scientific conferences and two summer schools were held in Suzhou. CSH Asia's scientific program is designed for scientists from the Asia/Pacific region, who make up more than 75% of attendance, and include symposia and meetings, training workshops, and occasional Banbury-style discussion meetings. This program is described in more detail in a separate Annual Report.

Special events included the fifth Double Helix Day winter event on Insights into the Human Condition, including excellent talks by Sarah Brosnan, Evan Eichler, Evelina Fedorenko, and Svante Pääbo. We also hosted several bioentrepreneur networking events, as well as numerous local area one- and two-day retreats, including several affiliated with the Feinstein Institute and Northwell Health. Although distinct from our regular academic program, these events attract significant numbers of leaders and individuals associated with biomedicine and bio-business from the tri-state area and beyond. Finally, two special conferences were arranged to celebrate important birthdays for scientific luminaries and old friends of CSHL, Mike Botchan and Harold Varmus, which were enjoyed by all the attendees.

The Meetings & Courses program staff comprises a diverse team of talented professionals who handle the complexities of database design, programming, web and multimedia design, educational grants management, marketing and recruitment, conference and course administration, audiovisual and digital design services, and other activities. We said goodbye to Nicholas Moore and Samuel Stewart after many years of devoted service to CSHL. We also welcomed several new staff in 2019, who are already bringing a high level of professionalism to their positions.

**David Stewart**
*Executive Director,*
*Meetings & Courses Program*
*President, Cold Spring Harbor Asia*

**Academic guidance**
**Terri Grodzicker**
*Dean of Academic Affairs*

## CSH ASIA SUMMARY OF CONFERENCES

| Dates | Title | Organizers |
| --- | --- | --- |
| April 8–12 | Bacterial Infection and Host Defense | Elizabeth Hartland, Kenya Honda, Nina Salama, Feng Shao, Jörg Vogel |
| April 15–19 | Francis Crick Symposium: Transforming Neurosciences: Questions and Experiments | Hailan Hu, Maiken Nedergaaard, John O'Keefe, Hee-sup Shin, Alcino Silva |
| May 6–10 | Membrane Proteins: From Physiology to Pharmacology | Nancy Carrasco, Baoliang Song, Chris Tate, Nieng Yan |
| June 8–12 | Scientific and Technical Advances in Cancer Immunology | Ira Mellman, Zemin Zhang |
| September 2–6 | Cross-Scale Biological Structure: From Macromolecular Complexes and Organelles to Cells and Tissues | Manfred Auer, Masahide Kikkawa, Hongwei Wang, Peijun Zhang |
| September 16–20 | Neurobiology of Behavior and Neuropsychiatric Disorders | Anthony Grace, Minmin Luo, Christian Luscher, Lan Ma, Masashi Yanagisawa |
| September 23–27 | Stem Cells, Aging, and Rejuvenation | Seung-Jae Lee, Hao Li, John Sedivy, Zhou Songyang, Yousin Suh |
| October 7–11 | NF-κB, JAK-STAT, and MAPK: Intercrossing Signaling Pathways in Health, Disease, and Therapy | Yinon Ben-Neriah, Eugene Chinn, Sankar Ghosh, Bing Su |
| October 14–18 | Cilia and Centrosomes | Fanni Gergely, Hiroshi Hamada, Gert Jansen, Guangshuo Ou, Meng-Fu Bryan Tsou |
| October 21–25 | Synthetic Biology | Junbiao Dai, Jay Keasling, Akihiko Kondo, Beatrix Suess |
| October 28– November 1 | Chemical Biology and Drug Discovery | Haian Fu, Yan-Mei Li, Rolf Müller, Minoru Yoshida |
| November 11–15 | Mitochondria and Metabolism in Health and Disease | Paolo Bernardi, Nika Danial, Naotada Ishihara, Xiaodong Wang |
| December 9–12 | Liver, Biology, Diseases, and Cancer [Awaji, Japan] | Gen-Sheng Feng, Stuart Forbes, Lijian Hui, Atsushi Miyajima, Takahiro Ochiya |
| December 9–13 | Kinase and Phosphatase Signaling | Tzu-Ching Meng, Reiko Sugiura, Tony Tiganis, Nicholas Tonks |

# 84TH COLD SPRING HARBOR LABORATORY SYMPOSIUM ON QUANTITATIVE BIOLOGY

## RNA Control and Regulation

**May 29–June 3**      408 Participants

ARRANGED BY      **Terri Grodzicker, David Stewart, and Bruce Stillman**, Cold Spring Harbor Laboratory

The Cold Spring Harbor Symposia on Quantitative Biology series is now in its 84th year, having been initiated by then-director Reginald Harris back in 1933—when the symposium lasted a full month! The Cold Spring Harbor symposia bring together scientists to present and evaluate new data and ideas in rapidly moving areas of biological research. Each year, a topic is chosen at a stage in which general and intensive scrutiny and review is warranted. Many previous Cold Spring Harbor symposia have addressed different aspects of RNA biology, including Nucleic Acids and Nucleoproteins in 1947, Transcription of Genetic Material in 1970, Mechanisms of Transcription in 1998, The Ribosome in 2001, and Regulatory RNAs in 2006. The enormous progress in the field in the past 15 years led us to conclude that the time was past due for another symposium squarely focused on RNA. In fact, this is only the second symposium in its entire history to include RNA in the title of the meeting, which we think conveys how central this molecule is in so many areas of the biology of the cell, and increasingly to how we think about treating disease. Topics addressed at the 2019 symposium included RNA-based structures; RNA modifications; nuclear localization of RNA; quality control and editing; RNA and gene regulation; cotranscriptional splicing; intron/exon boundaries; alternative polyadenylation; transposon control; small noncoding RNAs; long noncoding RNAs; RNA and development; membrane-less organelles; phase separation; RNA-based diseases; and novel RNA functions.

The symposium attracted more than 400 participants and provided an extraordinary five-day synthesis of current understanding in the field. Opening night talks setting the scene for later sessions included Roy Parker (HHMI/University of Colorado Boulder) on RNP granules in health and disease, Christine Mayr (Memorial Sloan Kettering Cancer Center) on the



Symposium Picnic

A. Mangilet, P. Sharp



C. Dean, C. Pikaard

regulation of 3′ UTR–mediated protein–protein interactions, Joshua Mendell (University of Texas Southwestern Medical Center) on the regulation and function of noncoding RNAs in mammalian physiology and disease, and Adrian Krainer (Cold Spring Harbor Laboratory), who addressed targeted modulation of splicing or NMD for disease therapy. Jennifer Doudna (HHMI/University of California, Berkeley) delivered a fascinating Dorcas Cummings Lecture on "Editing the Code of Life" for the Laboratory's friends and neighbors. Rising to the challenging task of condensing more than 50 talks over the prior five days, Gideon Dreyfuss (HHMI/University of Pennsylvania School of Medicine) provided a masterly summary of the state of the field at the conclusion of the symposium. Interviews with leading scientists captured during the symposium provide a snapshot of the state of current research and are available on the CSHL Leading Strand channel (https://www.youtube.com/user/LeadingStrand).

## PROGRAM

**Introduction**
B. Stillman, *Cold Spring Harbor Laboratory*

**Global Analyses and Structures**
*Chairperson:* L. Joshua-Tor, *HHMI/Cold Spring Harbor Laboratory*

**Regulation of RNA Functions**
*Chairperson:* D. Bartel, *HHMI/Whitehead Institute, Massachusetts Institute of Technology, Cambridge*

**Gene Regulation**
*Chairperson:* R. Darnell, *HHMI/The Rockefeller University, New York Genome Center, New York*

**RNA and Disease**
*Chairperson:* S. Gasser, *Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland*

**Small RNAs and Defense Systems**
*Chairperson:* P. Sharp, *Massachusetts Institute of Technology, Cambridge*



J. Steitz, T. Nilsen



D. Stewart, M.A. Jensen

J. Doudna


R. Singer, R. Lehmann

**Developmental Regulation**
*Chairperson:* C. Dean, *John Innes Centre, Norwich, United Kingdom*

**Dorcas Cummings Lecture**
J. Doudna, *HHMI/University of California, Berkeley*

**Condensates and Phase Separation**
*Chairperson:* J. Steitz, *HHMI/Yale University, New Haven, Connecticut*

**Chromatin and RNA**
*Chairperson:* N. Proudfoot, *University of Oxford, United Kingdom*

**Summary**
G. Dreyfuss, *HHMI/University of Pennsylvania School of Medicine, Philadelphia*


A. Virtanen, A. Stenlund


C. Arraiano, B. Jalloh


S. McKnight, S. Gasser

# MEETINGS

## Double Helix Day: Insights into the Human Condition

**February 26**      79 Participants

ARRANGED BY      **David Stewart and Jan Witkowski,** Cold Spring Harbor Laboratory

This special annual celebration ("Double Helix Day") is intended to co-incide with the actual date that James Watson and Francis Crick discovered the double-helix structure of deoxyribonucleic acid (February 28, 1953) in Cambridge, England.

Each year, a theme related to DNA science is explored through a series of 35 talks aimed at a broad scientific audience drawn from throughout the CSHL community. This year, the invited speakers worked at opposite ends of the range of studies providing insights into the Human Condition: molecular investigations of our origins and studies exploring the neuroscience of behaviors that characterize human beings.

The program concluded with an early evening reception and dinner.



S. Pääbo

### PROGRAM

*Chairperson:* S. Brosnan, *Georgia State University, Atlanta*: The evolutionary foundations of human cooperation.
*Chairperson:* E. Eichler, *University of Washington, Seattle:* Dynamic mutations and the genes that make us human.
*Chairperson:* E. Fedorenko, *Massachusetts Institute of Technology, Cambridge:* The language system in the human mind and brain.

*Chairperson:* S. Pääbo, *Max Planck Institute for Evolutionary Anthropology, Germany:* A Neanderthal perspective on human origins.



J. Simons, T. Janowitz



S. Brosnan



E. Fedorenko

# Systems Immunology

**March 13–16**    252 Participants

ARRANGED BY    **Kathryn Miller-Jensen,** Yale University, New Haven, Connecticut
**Harinder Singh,** University of Pittsburgh, Pennsylvania
**Sarah Teichmann,** Wellcome Sanger Institute, United Kingdom
**John Tsang,** National Institute of Allergy & Infectious Diseases, Bethesda, Maryland

This inaugural meeting hosted 252 participants, of whom 39.6% were female, 26.8% were graduate students, and 20.4% were postdocs. Notably, 25 corporate scientists and five journal staff were also among the attendees. This meeting was designed to bring together scientists working at the interface of experimental immunology and computational and systems biology. Recent advances in single-cell genomic and proteomic profiling as well as spatial and time-resolved imaging and their coupling to computational approaches have made possible analyses of immune responses at unprecedented resolution and across various scales of organization. These experimental and computational breakthroughs, along with rational design of immune



H. Singh, K. Miller-Jensen

cells—notably CAR T cells—are ushering in the dawn of systems immunology. Hence, the meeting was intended to nucleate and foster the nascent community of systems immunologists. The scientific program featured sessions on single-cell analyses of genomic and signaling states, modeling of intra- and intercellular information processing, high-resolution cellular and molecular profiling of immune responses and engineering of immune cells with therapeutically beneficial capabilities, and systems-level dissection of human immunity at different scales. The format of the meeting provided equal time for the oral presentations regardless of whether the speaker was invited or selected from among submitted abstracts. On the basis of the vigorous discussions spawned at the oral and poster presentations and the informal feedback received, the meeting appeared to be a resounding success. We expect the next meeting to be held in April 2021 to sustain and nurture this vital and



A. Nguyen, J. Suurmon



R. Ahmad, H. Machado

M. Chavez, K. Abadie



Y. Koguchi, A. Joglekar

rapidly evolving field. Major fundamental as well as translational advances in this field are anticipated in the coming years.

This meeting was funded in part by the National Institute of Allergy and Infectious Diseases, a branch of the National Institutes of Health, and by Merck.

**PROGRAM**

**Single-Cell Analysis of Genomic and Signaling States**
*Chairperson*: A. Cvejic, *University of Cambridge, United Kingdom*

**Modeling and Regulatory Networks**
*Chairperson:* H. Singh, *University of Pittsburgh, Pennsylvania*

**Immunoreceptors**
*Chairperson:* M. Davis, *HHMI/Stanford University School of Medicine, California*

**Cellular Dynamics, Interactions, and Communication**
*Chairperson:* K. Miller-Jensen, *Yale University, New Haven, Connecticut*



C. Bargmann, N. Chevrier

**Systems and Synthetic Immunity: The Future**
*Chairperson:* A. Regev, *Broad Institute of MIT and Harvard, Cambridge, Massachusetts*

**Human Immunology**
*Chairperson:* J. Tsang, *National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland*

# Network Biology

In cells and organisms, genomic information is translated into phenotypes by complex and highly dynamic molecular networks formed by proteins, nucleic acids, and small molecules. A systems-level understanding of biological systems, as well as the design of rational biotechnological or pharmaceutical interventions in humans, crops, and microbes, hinges on our knowledge of these networks.

As molecular networks remain largely incomplete, an important goal of biological network science is to experimentally map or computationally infer the wiring of cells. A second major goal is the mechanistic characterization of smaller network modules, translating large-scale network connectivity into molecular mechanisms as a basis for development of quantitative predictive models. Finally, network science also aims to develop and apply statistical tools to extract insights from known biological networks to identify disease-causing genes and modules, identify targets for intervention, and decipher the fundamental principles that underlie biological systems and their evolution. These goals have been aided by the rapid advances in high-throughput techniques, synthetic biology, and organism editing that have led to an explosion not just in the quantity, but in types of data available. Two key goals of this meeting remain to bridge these three aspects of network biology and to cover a diversity of biological systems, from humans and model organisms to plants and microbes, while continuing to highlight new experimental and computational opportunities and approaches.



M. Springer, B. Lehner, A-R. Carvunis, P. Falter-Braun



K. Salehi, A.F. Roth, A. Khalil



L. Dressler, M. Colic

315

C. Myers, P. Beltrao



Z. Wunderlich, V. Periwal

The most recent meeting successfully achieved these goals, and this gathering continues to serve as the main international meeting for the network biology community.

We continued the open panel discussions, which were highly successful for community reflection in the 2015 and 2017 meetings. This year's panel discussed the structural challenges women are still facing in science. The discussion was greatly aided by a questionnaire on gender perception bias that meeting participants completed prior to the meeting. This highly interactive session involved panelists and plenty of audience participation. The "meet the PI" lunch was replaced by a chalk talk session on the first evening in which all participants were assigned to 10 parallel groups and everybody presented their science to the group. This innovative format was intended to facilitate and catalyze networking and put young scientists in direct contact with senior scientists and PIs. These two elements were again praised by attendees and resulted in continued discussions throughout the meeting.

The scientific program opened on March 19 in the evening. There were 23 invited presentations and 20 short talks selected from submitted abstracts, all of them outstanding and many given by postdocs and Ph.D. students, with good gender balance among the presenters (35% women presenters). The talks covered a wide range of concepts spanning dynamics in signaling networks to the development of new methods to map the complete human protein interaction network, and addressed diverse questions from plant and animal development to cancer and other human diseases. This highlights how systems biology brings together people from different fields of biology. New areas and areas of growth for the community included machine learning, deep mutagenesis, translation, and the microbiome. These presentations showcased recent advances and also the open questions, exciting opportunities, and technological challenges that remain. The presentations were followed by dynamic and lively discussions.

Two Keynote Addresses inspired junior scientists by presenting a historical perspective on how the field has developed over the past 10–20 years and what the future may hold. Fritz Roth opened the meeting with an overview of his laboratory's efforts to systematically map functional consequences of genetic perturbations using a suite of new experimental technologies. Trey Ideker closed the meeting with a demonstration of how biology-informed neural networks can lead to transparent deep-learning approaches in which causative and implicated features and processes can be identified by inspection of the network to complement successful precisions and classifications with biological insights.



D. Layman, I. Overton

**PROGRAM**

**Opening Keynote Address Speaker**
F. Roth, *University of Toronto, Canada*

**Round Tables/Chalk Talks**

**Regulatory Networks**
*Chairperson:* S. Gaudet, *Dana-Farber Cancer Institute/ Harvard Medical School, Boston, Massachusetts*

**Networks in Differentiation**
*Chairperson:* G. Coruzzi, *New York University, New York*

**Networks of Translation**
*Chairperson:* M. Lee, *University of Massachusetts Medical School, Worcester*

**Lightning Talks**
*Chairperson:* B. Lehner, *Centre for Genomic Regulation, Barcelona, Spain*

**Genetic Networks**
*Chairperson:* A-R. Carvunis, *University of Pittsburgh, Pennsylvania*

**Computational Methods for Network Biology**
*Chairperson:* A-R. Carvunis, *University of Pittsburgh, Pennsylvania*

**Integrative Networks**
*Chairperson:* M. Springer, *Harvard Medical School, Boston, Massachusetts*

**Panel Discussion: Women in Network Science**
*Chairperson:* J. Roecklein-Canfield, *Simmons College, Boston, Massachusetts*

**Network Evolution**
*Chairperson:* B. Lehner, *Centre for Genomic Regulation, Barcelona, Spain*

**Microbiome Networks**
*Chairperson:* P. Falter-Braun, *Helmholtz Zentrum Münich/ Ludwig-Maximilians University, Neuherberg, Germany*

**Protein Networks**
*Chairperson:* M. Springer, *Harvard Medical School, Boston, Massachusetts*

**Signaling Networks**
*Chairperson:* P. Beltrao, *Embl-Ebi, Hinxton, United Kingdom*

**Disease Networks**
*Chairperson:* T. Ideker, *University of California, San Diego*

**Closing Keynote Address Speaker**
T. Ideker, *University of California, San Diego*

# RNA and Oligonucleotide Therapeutics

**March 27–30**     121 Participants

ARRANGED BY     **Annemieke Aartsma-Rus,** Leiden University Medical Center, the Netherlands
**Masad Damha,** McGill University, Montreal, Quebec, Canada
**Laura Sepp-Lorenzino,** Vertex Pharmaceuticals, Boston, Massachusetts
**Matthew Stanton,** Generation Bio, Cambridge, Massachusetts

This fifth conference focused on development of nucleic acids as drugs, covering different nucleic acid modalities (siRNA, mRNA, RNaseH, splice modulation, CRISPR-Cas9) and involving different aspects of nucleic acid therapy development (i.e., chemistry, delivery, and preclinical and clinical studies). Currently, seven nucleic acid therapies have received regulatory approval, most recently the first siRNA patisaran (ONPATTRO®) for the treatment of aTTR-induced amyloidosis. The meeting opened with a Keynote Address by Alnylam's Dr. Muthiah Manoharan ("Mano") outlining the road toward this landmark approval. Other sessions showed a diversity of nucleic acid modalities being developed preclinically and clinically. There were also sessions on important aspects such as safety, chemistry, and delivery and on newly arising nucleic acid technologies.



M. Damha, L. Sepp-Lorenzino, A. Aartsma-Rus, M. Stanton

The participants came from 11 companies and from universities and research institutions from the United States and abroad. The seven scientific sessions featured 37 platform talks, 20 posters, and a panel discussion and included 128 registered attendees. Animated and insightful exchanges during the sessions continued throughout the sessions and during breaks and social events. Most participants indicated the meeting was excellent and expressed interest in attending the next edition. The next RNA and Oligonucleotide Therapeutics meeting will take place in March 2021.

The topic of the meeting attracted funding from several companies. The main sponsors were Alnylam Pharmaceuticals, AXOlabs, Levin Biosciences, Generation Bio, Intella Therapeutics, Ionis, Moderna, Sarepta Therapeutics, uniQure, Vertex Pharmaceuticals, and Wolfe Laboratories.

**Keynote Address Speaker**
M. Manoharan, *Alnylam Pharmaceuticals, Cambridge, Massachusetts*

**NAT Pipeline: Preclinical Programs**
*Chairpersons:* A. Aarstma-Rus, *Leiden University Medical Center, the Netherlands*; A. Krainer, *Cold Spring Harbor Laboratory*

**NAT Pipeline: Clinical Programs**
*Chairpersons:* L. Sepp-Lorenzino, *Vertex Pharmaceuticals Inc., Boston, Massachusetts*; R. MacLeod, *Ionis Pharmaceuticals, Carlsbad, California*



P. Smith, S. Henry

A. Levin, A. Krainer



K. Neuman, J. Brown

**Safety**

*Chairperson:* P. Smith, *Alnylam Pharmaceuticals, Cambridge, Massachusetts*

**NAT Chemistry**

*Chairpersons:* M. Damha, *McGill University, Montreal, Quebec, Canada*; M. Manoharan, *Alnylam Pharmaceuticals, Cambridge, Massachusetts*

**NAT Delivery**

*Chairpersons:* M. Stanton, *Generation Bio, Cambridge, Massachusetts*; S. Dowdy, *University of California, San Diego*

**Emerging NATs**

*Chairpersons:* M. Stanton, *Generation Bio, Cambridge, Massachusetts*; L. Sepp-Lorenzino, *Vertex Pharmaceuticals Inc., Boston, Massachusetts*



J. Scharner, Y.J. Kim



J. Lane, F. Hillebrand

# Blood–Brain Barrier

**April 3–6**     166 Participants

ARRANGED BY     **Dritan Agalliu,** Columbia University, New York
                     **Robert Bell,** Pfizer, Inc., Cambridge, Massachusetts
                     **Chenghua Gu,** Harvard Medical School, Boston, Massachusetts

This meeting highlighted advances in the cellular and molecular aspects of blood–brain barrier (BBB) development, function, and disease. The conference was well attended, particularly by young scientists (graduate students and postdoctoral fellows), who also had a prominent role in giving oral presentations. This made for a lively discussion of unpublished data, giving foresight into the future of BBB research. Moreover, women scientists were well represented at the meeting. There were 57.9% female attendees, and many of them were session chairs and oral presenters.



D. Agalliu, C. Gu, R. Bell

As in years past, the conference had a substantial representation of genetic model organisms, live imaging, and powerful molecular genetics highlighted in studying mammalian and nonmammalian BBB physiology and development. A major focus of the conference was the emerging field of molecular mechanisms underlying BBB permeability, with a focus on molecular mechanisms of transcytosis and transport functions. BBB modulation for drug transport, and a balanced focus on the BBB in disease and therapeutic development, were also discussed during the oral and poster presentations.

The Keynote Address, Specialization and Functional Heterogeneity of Vascular Cells' Disease, was given by Ralf Adams.

Topics addressed by the conference included BBB development, the functional BBB, emerging BBB model systems, molecular BBB transport, the aging and diseased BBB, and functional interactions at the neurovascular unit. As is traditional at Cold Spring Harbor meetings, selection of material for oral and poster presentation was made by the organizers and individual session chairs on the basis of scientific merit.

This meeting was supported with funds by the National Institute of Neurological Disorders and Stroke of the National Institutes of Health, and by Pfizer.



C. Collier

## PROGRAM

**Keynote Address Speaker**
R.H. Adams, *Max Planck Institute for Molecular Biomedicine, Muenster, Germany*

**Development and Maturation of CNS Barriers**
*Chairpersons:* S. Liebner, *Goethe University Clinic, Frankfurt, Germany*; S. Nicoli, *Yale University, New Haven, Connecticut*
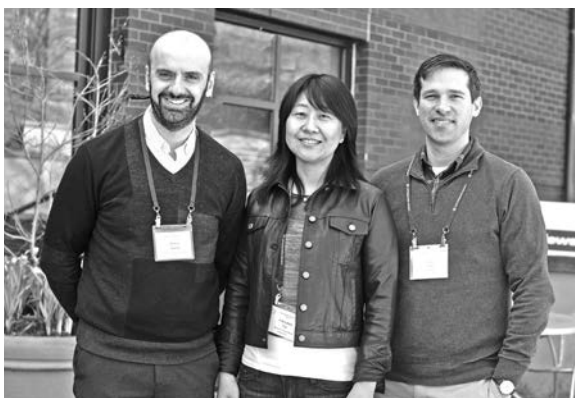
S. Zhang, L. Drewes



T. Quigley, S. Axelrod

**Model Systems to Study CNS Barriers**
*Chairpersons:* A. Sehgal, *University of Pennsylvania Perelman School of Medicine, Philadelphia*; E. Lippmann, *Vanderbilt University, Nashville, Tennessee*

**Special Lecture and Panel Discussion**
*Chairperson:* A. Cimpian, *New York University, New York*

**Imaging and Function of the BBB in Health and Disease**
*Chairpersons:* A. Shih, *Seattle Children's Research Institute, Washington*; H. de Vries, *VU Medical Center, Amsterdam, the Netherlands*

**Cell Biology of the BBB**
*Chairpersons:* M. Taylor, *University of Wisconsin, Madison*; S.E. Lutz, *University of Illinois, Chicago*

**The Gut-Brain Axis and Neuroimmune Interactions**
*Chairpersons:* J. Kipnis, *University of Virginia, Charlottesville*; J.R. Huh, *Harvard Medical School, Boston, Massachusetts*

**Gene Therapy and Biotherapeutic Transport Across the BBB**
*Chairpersons:* B. Obermeier, *Biogen, Cambridge, Massachusetts*; A. Ben-Zvi, *Hebrew University of Jerusalem, Israel*



S. Lutz, D. Reich



R. Daneman, E. Crouch

# Cellular Dynamics and Models

**April 9–12**        82 Participants

ARRANGED BY        **James Faeder,** University of Pittsburgh, Pennsylvania
**Dyche Mullins,** University of California, San Francisco
**Ed Munro,** University of Chicago, Seattle
**Susanne Rafelski,** Allen Institute for Cell Biology, Seattle, Washington

Veterans Ed Munro and James Faeder, together with newcomers Susanne Rafelski and Dyche Mullins, maintained the historical focus of the meeting on quantitative approaches to cell biology—specifically computational modeling, quantitative measurement, and mathematical analysis. Within this framework, the biological questions address spanned a wide range, from optimizing cancer chemotherapy regimens, to describing the physical properties of living matter, to understanding how cellular hydrodynamics governs the carbon cycle of the ocean.

The blueprint for the scientific program contained one Keynote talk; eight themed sessions with longer, *invited,* and shorter, *selected,* talks; two "Tradecraft" sessions devoted to cutting-edge tools for computational and quan-



D. Mullins, E. Munro, S. Rafelski, J. Faeder

titative cell biology; and two afternoon poster sessions. The Tradecraft sessions were designed to replace the "Tutorials" of previous years, and although the focus of these Tradecraft sessions was squarely on tools (e.g., software platforms, microfluidic devices, and algorithms), speakers were encouraged to present an overview rather than a "how to" session. For attendees wanting more detail, the Tradecraft session speakers were encouraged to present posters and demonstrations during the afternoon poster sessions. The organizers added a second set of poster sessions to the meeting this year to facilitate discussion and productive one-on-one interactions. The Tradecraft and poster sessions were very well received, and the organizers agreed that this was a successful approach that should be continued in the future.

For the 2021 meeting, Carlos Lopez (Vanderbilt University) will replace James Faeder as co-organizer, and Hana el Samad (University of California, San Francisco) will replace Ed Munro.

V. Periwal, M. Blinov



P. Maheshwari, D. Deritei

## PROGRAM

**Dynamic Control of Cell Fate**
*Chairpersons:* D. Weitz, *Harvard University, Cambridge, Massachusetts*; E. Read, *University of California, Irvine*

**Dynamic Control of Biological Information Flow**
*Chairpersons:* H. El-Samad, *University of California, San Francisco*; C. Lopez, *Vanderbilt University, Nashville, Tennessee*

**Dynamics of Force Generation and Dissipation**
*Chairpersons:* M. Murrell, *Yale University, New Haven, Connecticut*; S. Eaton, *Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany*

**Keynote Address Speaker**
J. Theriot, *University of Washington, Seattle*

**And Now for Something Completely Different**
*Chairperson:* M. Prakash, *Stanford University, California*

**The Self-Organized Cytoplasm**
*Chairpersons:* S. Weber, *McGill University, Montreal, Quebec, Canada*; L. Holt, *New York University, New York*

**Dynamic Interplay of Molecular Assemblies, Cell Shape, and Cellular Behavior**
*Chairpersons:* P. Rangamani, *University of California, San Diego*; E. Garner, *Harvard University, Cambridge, Massachusetts*

**Evolutionary Cell Biology**
*Chairperson:* N. Elde, *University of Utah, Salt Lake City*

**Dynamics of Gene Expression**
*Chairperson:* A. Zidovska, *New York University, New York*; A. Boettiger, *Stanford University, California*



D. Lew, D. Ghose



Cocktails before banquet

# Ubiquitins, Autophagy, and Disease

**April 23–27**     206 Participants

ARRANGED BY     **Anne Bertolotti,** MRC LMB, United Kingdom
**Michael Rape,** University of California, Berkeley
**Cynthia Wolberger,** Johns Hopkins University School of Medicine, Baltimore, Maryland

This is the eighth Ubiquitin meeting (previous title: Ubiquitin Family meeting) following the successful inauguration of this series in 2003. The ubiquitin field has seen dramatic changes over the last years: Following groundbreaking enzymatic and structural work, recent years have brought new insight into the physiology of ubiquitin-dependent signaling in development and disease, particularly its function in autophagy. Moreover, the first small-molecule modulators of the ubiquitin system have recently entered the clinic. As highlighted by its new title Ubiquitins, Autophagy, and Disease, this meeting paid tribute to the evolution of the ubiquitin field by focusing more on autophagy as well as biomedical applications that alter the activity of ubiquitin-dependent signaling. In addition, as in previous years, this meeting discussed fundamental questions underlying ubiquitin-dependent signaling, including how specificity of ubiquitin or ubiquitin-like conjugation is established; how the many ubiquitylation enzymes control crucial signaling events in protein quality control, cell cycle control, or development; or how aberrant ubiquitylation contributes to disease. These questions are being elegantly addressed using structural biology, sophisticated kinetic studies, cutting-edge library screening, and quantitative proteomics technologies by an increasing number of investigators in the field. The ubiquitin field thus brings together researchers from very different areas of biology, which fostered deeply informed and creative discussion throughout the meeting.



A. Bertolotti, C. Wolberger, M. Rape

Two highlights of this 2019 ubiquitin meeting were its exciting Keynote Addresses. The opening Keynote lecture was delivered by Jonathan Weissman, an HHMI investigator who spearheaded genetic discovery of new signaling enzymes and thus discovered many critical enzymes and components of ribosome quality control. The second Keynote lecture was delivered by Richard Youle, a senior investigator at NIH who has produced groundbreaking work in identifying the molecular origins of a familial form of Parkinson disease. The 206 participants in this meeting then witnessed many exciting talks presenting unpublished studies that underscored the rapid pace of discovery and the complexity of the ubiquitin field. To name but a few examples, presented were the bacterial conservation of ribosome quality control; new regulation of ubiquitin and polyglutamine modification by pathogenic bacteria; new mechanisms of ubiquitin-dependent cell cycle control in stem cells; new insights into the complex interplay of autophagy receptors, their substrates, and the autophagosome machinery; and breathtaking structural studies revealing modes of chromatin regulation or protein degradation. The meeting also attracted several members of the budding community of ubiquitin researchers in biomedical companies that witnessed talks by leaders of pharmaceutical companies or academic centers that are now deeply involved in developing ubiquitin-directed drugs. The very active discussions were continued in the poster

A. Simonsen, H. Vodermaier



V. Weerasekar, A. Asmar

sessions, which were introduced by a series of lightning talks and allowed many graduate students and postdoctoral researchers to highlight their important work, including first insights into molecular mechanisms of protein quality control. The collaborative and interdisciplinary nature of this field, now deeply relevant for drug discovery, was particularly obvious throughout this newly designed meeting

This meeting was sponsored in part by Kymera Therapeutics, Inc.

## PROGRAM

**Quality Control**
*Chairpersons:* R. Hampton, *University of California, San Diego*; B. Schulman, *St. Jude Children's Research Hospital, Memphis, Tennessee*

**Keynote Address Speaker**
J. Weissman, *University of California, San Francisco*

**Physiology/Screens**
*Chairpersons:* I. Dikic, *Goethe University Frankfurt, Germany*; J.W. Harper, *Harvard Medical School, Boston, Massachusetts*

**Lightning Talks (Poster I)**

**Drug Discovery/Disease**
*Chairpersons:* R. Deshaies, *Amgen, Inc., Thousand Oaks, California*; E. Fischer, *Harvard Medical School, Boston, Massachusetts*

**Autophagy I**
*Chairpersons:* E. Holzbaur, *University of Pennsylvania, Philadelphia*; A. Simonsen, *University of Oslo, Norway*



S. Gates, E. Blythe



M. Metzger, E. Rusilowicz

**Ubiquitin in the Nucleus**
*Chairpersons:* A. Smogorzewska, *The Rockefeller University, New York*; N. Thomä, *Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland*

**Lightning Talks (Poster II)**

**Disease**
*Chairpersons:* I. Wertz, *Genentech, South San Francisco, California*; F. Randow, *MRC, Cambridge, United Kingdom*

**Autophagy II**
*Chairpersons:* M. Hochstrasser, *Yale University, New Haven, Connecticut*; R. Klevit, *University of Washington School of Medicine, Seattle*

**Keynote Address Speaker**
R. Youle, *National Institutes of Health, Bethesda, Maryland*

**Degradation**
*Chairpersons:* K. Walters, *National Cancer Institute, Frederick, Maryland*; T. Ravid, *Hebrew University of Jerusalem, Israel*



C. Sitron, T. Mayor

# Telomeres and Telomerase

**April 30–May 4**      246 Participants

ARRANGED BY      **Steven Artandi,** Stanford University, California
                **Julia Promisel Cooper,** National Cancer Institute, Bethesda, Maryland
                **Jan Karlseder,** The Salk Institute, La Jolla, California

This was the 11th Telomeres and Telomerase conference, held every two years from 1999 onward. It consisted of eight sessions of talks and two poster sessions. As for every one of these meetings, the format was to invite two chairpersons per session, who were a mix of established scientists in the field and younger scientists who had already made their mark as independent investigators. Session chairs were given the choice of giving a 12-minute presentation themselves or having a member of his/her lab give a talk. The remainder of the talks (also 12 minutes) were chosen from submitted abstracts, allowing as many presentations as possible. These presentations were primarily made by graduate students and postdoctoral fellows, a large fraction of whom presented the 130 posters and 76 talks.



S. Artandi, J.P. Cooper, J. Karlseder

The talks and posters covered diverse aspects of telomere and telomerase biology, including regulation of telomerase expression and activity, telomerase biogenesis and structure, telomere replication, mechanisms of ALT, mechanisms of telomere protection, telomere protein functions at telomeres and throughout the genome, telomere shortening and mechanisms of senescence and aging, and the role of telomeres in human health and disease.

The quality and novelty of scientific content was very high throughout the conference in both the talks and the posters. Most of the presented data were unpublished and extensively discussed in an open fashion. Formal and informal discussions were lively and informative. The conference was judged to be highly successful based on verbal and email communications to the organizers. There is strong enthusiasm for another meeting on the same topic in 2021.

## PROGRAM

**Replication Regulation**
*Chairpersons:* T. de Lange, *The Rockefeller University, New York*; E. Hendrickson, *University of Minnesota, Minneapolis*

**Telomerase Biogenesis and Structure**
*Chairpersons:* T. Cech, *HHMI/University of Colorado, Boulder*; J. Chen, *Arizona State University, Tempe*

**Telomere Dysfunction and Genomic Integrity**
*Chairpersons:* V. Lundblad, *Salk Institute for Biological Sciences, La Jolla, California*; D. Pellman, *Dana-Farber Cancer Institute, Boston, Massachusetts*



A. Ries, J. Hentschel

V. Lundblad, D. Wuttke



D. Rhodes, T. Cech

**Telomerase Function and Regulation**
*Chairpersons:* P. Opresko, *University of Pittsburgh, Pennsylvania*;
C. Azzalin, *Instituto de Medicina Molecular, Lisboa, Portugal*

**Telomere Protection I**
*Chairpersons:* S. Smith, *New York University School of
Medicine, New York*; K. Tomita, *Brunel University London,
United Kingdom*

**ALT**
*Chairpersons:* R. Reddel, *Children's Medical Research
Institute, Westmead, Australia*; A. Decottignies, *Univerisité
Catholique de Louvain, Brussels, Belgium*

**Telomeropathies and Aging**
*Chairpersons:* C.W. Greider, *Johns Hopkins University,
Baltimore, Maryland*; C. Garcia, *Columbia University
Medical Center, New York*

**Telomere Protection II**
*Chairpersons:* M. Kupiec, *Tel Aviv University, Israel*;
R. O'Sullivan, *University of Pittsburgh, Pennsylvania*



Nicholls Biondi Poster Session



V. Geli, T. de Lange

# The Biology of Genomes

**May 7–11**     485 Participants

ARRANGED BY     **Matthew Hurles,** Wellcome Trust Sanger Institute, United Kingdom
**Elaine Ostrander,** National Human Genome Research Institute, Bethesda, Maryland
**Dana Pe'er,** Sloan Kettering Institute, New York, New York
**Jay Shendure,** University of Washington, Seattle

This 32nd annual meeting hosted 485 participants, with more than 300 abstracts presented. The work presented spanned a broad array of topics related to genomics, including functional characterization of genetic variants, single-cell biology, the development of new computational methods, the analysis of human population history, the application of genomics to study the etiology of both rare and common human diseases, and the advancement of genome editing technologies. The scope of the genome sciences continues to grow, and the sophistication of both experimental and computational methods continues to deepen. As we approach the 30th anniversary of the BOG meeting, the talks continue to be fresh, with the usual fierce competition for speaking slots. The meeting featured



D. Pe'er, J. Shendure, M. Hurles

talks on a wide variety of topics (listed below under Program). There was a reasonable balance of genders in terms of speakers and a strong focus on younger graduate students and postdoc presentations. However, the balance in both gender and training levels in terms of the questions solicited at the end of each talk is something that we can improve upon. To this end, midway through the meeting, we urged better balance to the audience and also piloted an "anonymous" web-based interface for submitting questions. Both measures helped to broaden the diversity of those who ask questions.

The specific talks spanned a wide range of topics: for example, including a "calling card" assay for transcription-factor-binding coupled to single-cell RNA-Seq (Arnav Moudgil), extensive GTeX-based maps of postzygotic mutations (Nicole Rockweiler), the description of blood-based epigenetic signatures associated with



D. Schwartz, R. Waterson



B. Stillman, H. Yang

R. Christ, S. Batra



R. Mitra, M. Zody

chromatin modifier–caused developmental disorders (Julie Handsaker), population genomic analyses of selection in U.S. beef cattle (Troy Rowan), evidence for an extremely early origin for some developmental enhancers (Emily Wong), investigations of the relationship between 3D genome architecture and GWAS signal in the context of pancreatic islets and type II diabetes (Inês Cebola), and quantification of the heritability of the gut microbiome (Ran Blekhman). Three poster sessions allowed for comprehensive discussions of abstracts that did not make the talks and were well attended. Indeed, all sessions were well attended, stretching the capacity of the CSHL facilities. For her Keynote presentation, Molly Przeworski focused on the determinants of the human germline mutation rate, and in his Keynote, Robert Waterston discussed the integration of the Sulston lineage of *C. elegans* and single-cell RNA-Seq to yield new general principles about worm embryonic development.

The ELSI (Ethical, Legal, and Social Implications) panel, moderated by Nicole Lockhart, focused on "The Use of Genetic Genealogy for Forensic Investigations." Panelists included Thomas F. Callaghan (FBI Laboratory); James W. Hazel (Vanderbilt University Medical Center); Sara Huston Katsanis (The Duke Initiative for Science & Society); and Ellen McRae (Greytak, Parabon NanoLabs).

This meeting was funded in part by the National Human Genome Research Institute, a branch of the National Institutes of Health, by Oxford Nanopore, and by PacBio.



D. Thybert, T. Keane



C. Robles, R. Walker

## PROGRAM

**Functional Genomics**

*Chairpersons:* A. Raj, *University of Pennsylvania, Philadelphia*; H. Kilpinen, *University College London, United Kingdom*

**Computational Genomics**

*Chairpersons:* E. Eskin, *University of California, Los Angeles*; B. Engelhardt, *Princeton University, New Jersey*

**Cancer and Medical Genomics**

*Chairpersons:* S. Kathiresan, *Massachusetts General Hospital, Boston*; C. Wu, *Dana-Farber Cancer Institute/Harvard Medical School, Boston, Massachusetts*

**Population Genomics**

*Chairpersons:* H. Martin, *Wellcome Trust Sanger Institute, Hinxton, United Kingdom*; S. McCarroll, *Broad Institute of MIT and Harvard, Cambridge, Massachusetts*

**ELSI Panel and Discussion: The Use of Genetic Genealogy for Forensic Investigations**

*Chairperson:* N.C. Lockhart, *NIH/National Human Genome Research Institute, Bethesda, Maryland*

**Evolutionary Non-Human Genomics**

*Chairpersons:* K. Lohmueller, *University of California, Los Angeles*; J. Thornton, *University of Chicago, Illinois*

**Genome Engineering and Editing**

*Chairpersons:* C. Bock, *Austrian Academy of Sciences, Vienna*; F. Isaacs, *Yale University, New Haven, Connecticut*

**Guest Speakers**

*Chairpersons:* M. Przeworski, *Columbia University, New York*; R. Waterston, *University of Washington, Seattle*

**Complex Traits and Microbiome**

*Chairpersons:* G. Sella, *Columbia University, New York*; J. Xavier, *Memorial Sloan Kettering Cancer Center, New York*

# Mechanisms of Metabolic Signaling

**May 14–18**      189 Participants

ARRANGED BY      **Mitchell Lazar,** University of Pennsylvania, Perelman School of Medicine, Philadelphia
**Susanne Mandrup,** University of Southern Denmark, Odense
**Jared Rutter,** HHMI/University of Utah School of Medicine, Salt Lake City

This meeting followed a highly successful Metabolic Signaling and Disease meeting two years earlier. The main goal of this meeting was to bring together researchers from diverse fields to explore how principles of cellular metabolism manifest in different cell types; how metabolic regulation underlies the functions of specialized tissues; and how these differences impact both normal physiology and diseases such as diabetes and cancer.

The 24 invited speakers were leaders in the various aspects of metabolic research from all over the world, and the majority had not been part of the 2015 or 2017 programs. The meeting opened with an inspiring Keynote Address by Eileen White, which beautifully integrated mechanistic metabolism studies with an understanding of the physiological impacts. A total of 10 sessions followed, eight of which featured oral presentations, all highlighting unpublished research, and focused on key areas in the field of metabolism. Emphasis included, but was not limited to, genomic and epigenomic mechanisms, signaling pathways, lipid flux and storage, and mitochondrial function, with an accent on comparing and contrasting normal and pathologic metabolic states. Short talks were chosen from abstracts to increase the exposure of younger investigators and to highlight hot topics that complemented and extended the exciting



J. Rutter, S. Mandrup, M. Lazar



Posters at Nicholls Biondi Hall

A. Regmi, G. Varga



A. Parrish, P. Strzyz

program. There was ample time for discussion, which was very lively and often spilled over to other venues including the cafeteria and social events designed to encourage interactions between trainees, young investigators, and senior faculty. In all, there were 44 talks by speakers from Canada, Europe, and Asia in addition to the United States; 14 of the talks were given by women.

In addition to the eight oral sessions, there were two lively poster sessions, featuring a total of 97 posters, that were extremely well attended and presented. All 10 sessions were characterized by open and wide-ranging discussions, and the meeting provided a unique forum for the exploration of the commonalities and differences in metabolic principles and details across different laboratories, systems, and diseases. All attendees gained in-depth exposure to the remarkable cell, organ, and disease specificity of metabolic flux and its regulation. Indeed, a great success of the meeting was the interactive nature of the meeting, in which stimulating questions and discussion led to new concepts and future collaborations.



This meeting was funded in part by the National Institute of Diabetes and Digestive and Kidney Diseases, a branch of the National Institutes of Health.

M. Montgomery, M. Green



S. Barritt, T. Kadowaki



J. Bass, C. Hepler

## PROGRAM

**Keynote Address Speaker**
*Chairperson:* E. White, *Rutgers Cancer Institute of New Jersey, New Brunswick*

**Integrated Physiology**
*Chairperson:* K. Funai, *University of Utah, Salt Lake City*

**Circadian Regulation**
*Chairperson:* Z. Gerhart-Hines, *University of Copenhagen, Denmark*

**Emerging Technologies**
*Chairperson:* J. Estall, *IRCM, Montreal, Canada*

**Genetic and Epigenomic Programming**
*Chairperson:* C. Cummins, *University of Toronto, Canada*

**Mitochondria and Cancer Metabolism**
*Chairperson:* A. Walker, *University of Massachusetts Medical School, Worcester*

**Fat Storage and Mobilization: Putting Lipids in Their Place**
*Chairperson:* P. Gordts, *University of California, San Diego*

**Neurocontrol/Immunometabolism**
*Chairperson:* T. Unterman, *University of Illinois, Chicago*

# Retroviruses

**May 20–25**          408 Participants

ARRANGED BY          **Heinrich Gottlinger,** University of Massachusetts Medical School, Worcester
                     **Clare Jolly,** University College London, United Kingdom

This meeting, considered the best on the basic biology of retroviruses including HIV, brings together scientists from around the world, fosters friendships and collaborations, and has frequently been the venue where major scientific breakthroughs have first been announced.

Groundbreaking presentations this year included the first structure of a SERINC family member, cryoelectron microscopy of HIV-1 interaction with co-factors, multiple presentations on the role of the capsid in nuclear import, the mechanism of antiviral activity by ZAP, and numerous talks on viral evasion and antagonism of innate immunity.



R. Sutton, H. Gottlinger

This year's Keynote speakers were Drs. Bryan Cullen and Jeremy Luban. Dr. Cullen presented an overview of his major and influential original discoveries, from his early work on transcriptional interference in retroviruses, the identification of HIV-1 Rev as a nuclear export factor, the groundbreaking insight that HIV-1 differs fundamentally from other retroviruses in that it can infect nondividing cells, all the way to his recent work on retroviral epitranscriptomics.

Dr. Luban gave an interesting and captivating talk about his career to date, including his early days in the inspirational Ralph Steinman's laboratory that sparked his desire for a career in scientific research. Luban covered his numerous contributions to understanding basic HIV-1 biology, HIV–dendritic cell interactions, and vaccine work, which highlighted the breadth of his extensive research.

The meeting maintained the overall arrangement of 13 sessions. Sessions were grouped with an emphasis on the viral life cycle as has become convention in the past few years, rather than specific proteins. This resulted in more diverse and stimulating sessions that ensured strong attendance at all the sessions and illuminating questions and answers.



Lunch on Blackford Lawn



S. Hughes, H. Levin, E. Freed

V. Simon, C. Carter



A. Zamborlini, J-C. Paillart

The meeting retained two new and very well-attended additional lunchtime sessions:

(1) The popular live recording of the podcast This Week in Virology (TWiV) by Dr. Racaniello and featuring the Keynote speaker Bryan Cullen, his past mentorees Michael Malim and Paul Bieniasz, and Anna Marie Skalka, who has a long-standing connection to the Laboratory having worked as a postdoc here in the 1960s before turning her attention to retroviruses; and (2) a lunchtime panel discussion featuring Drs. Eric Cohen, Julia Garcia Prado, Richard Sutton, and ViiV head of discovery Mark Cockett on gene editing and curing HIV, which allowed the participants of the meeting to discuss recent advances and challenges for HIV Cure research and was a very well attended and popular session.

The oral sessions were once again notable for their mix of presentations by new and more established researchers, which is an important feature of the meeting. Session chairs, presenters, and attendees reflected the diversity of the meeting participants, maintained a gender balance, and showcased the international flavor of the retrovirology field.

This meeting was funded in part by the National Institute of Allergy and Infectious Diseases, a branch of the National Institutes of Health.

## PROGRAM

**Env, Entry, and Restriction Factors**
*Chairpersons:* S-L. Liu, *Ohio State University, Columbus*; W. Mothes, *Yale University, New Haven, Connecticut*

**Late Events and Assembly**
*Chairpersons:* A. Ono, *University of Michigan Medical School, Ann Arbor*; P. Zhang, *University of Oxford, United Kingdom*

**Lunchtime Panel Discussion: Gene Editing and Curing HIV**
*Chairperson:* M. Cockett, *Head of Discovery, ViiV*

**Virus Dissemination, and Pathogenesis**
*Chairpersons:* T. Hatziioannou, *The Rockefeller University, New York*; B. Chen, *Icahn School of Medicine at Mount Sinai, New York*

**Keynote Address Speaker**
B. Cullen, *Duke University, Durham, North Carolina*

**Early Post-Entry Events and Restriction Factors**
*Chairpersons:* K. Bishop, *The Francis Crick Institute, London, United Kingdom*; O. Fregoso, *University of California, Los Angeles*

**Nuclear Import and Integration**
*Chairpersons:* P. Cherepanov, *The Francis Crick Institute, London, United Kingdom*; F. Di Nunzio, *Institut Pasteur, Paris, France*

**Viral RNA and RNA Packaging**
*Chairpersons:* S. Kutluay, *Washington University School of Medicine in St. Louis, Missouri*; K. Musier-Forsyth, *Ohio State University, Columbus*

**Eighth Annual Uta von Schwedler Prize for Retrovirology**
Awarded to J. Wang, *University of Minnesota, Minneapolis*

**Lunchtime TWiV Podcast: This Week in Virology**
*Chairperson:* V. Racaniello, *Columbia University, New York*

**Innate Sensing**
*Chairpersons:* S. Neil, *King's College London, United Kingdom*;
   C. Goffinet, *Charite–Universitatsmedizin, Berlin, Germany*

**Keynote Address Speaker**
J. Luban, *University of Massachusetts Medical School, Worcester*

**New Virus–Host Interactions**
*Chairpersons:* M. Malim, *King's College London, United
   Kingdom*; M. OhAinle, *Fred Hutchinson Cancer Research
   Center, Seattle, Washington*

**Thirteenth Annual Andy Kaplan Prize**
Awarded to: M. OhAinle, *Fred Hutchinson Cancer Research
   Center, Seattle, Washington*

**Transcription and Latency**
*Chairpersons:* Y-C. Ho, *Yale University School of Medicine,
   New Haven, Connecticut*; B.M. Peterlin, *University of
   California, San Francisco*

**Ninth Annual Daniel Wolf Prize**
Awarded to: The Best Poster Presentation



O. Cingoz, D. Ferhadian

**Endogenous Retroviruses and Evolution**
*Chairpersons:* W. Johnson, *Boston College, Chestnut Hill,
   Massachusetts*; C. Kozak, *NIAID/National Institutes of
   Health, Bethesda, Maryland*

# Microbiome

The microbiome field has experienced dramatic growth over the past ~15 years. After an initial phase in which microbial enumeration in host-associated and environmental microbiomes predominated, studies aimed at developing a mechanistic understanding of how microbes interact with one another, with hosts, and with the environment are under way. Most microbiome researchers segregate into two groups—those who study host-associated microbiomes and those who study environmental microbiomes. A central goal of this meeting was to provide a forum where individuals from these classically separate communities could interact, share unpublished work, and develop new collaborations to propel the field forward.



A. Bhatt, J. Segre, M. Yassour

    Two of the highlights of this meeting were its exciting Keynote Addresses. The opening Keynote lecture was delivered by Emily Balskus, Professor of Chemistry and Chemical Biology at Harvard University. She has collaboratively contributed to core mechanistic insights about xenobiotic metabolism by gut microbes. Emily spoke to the role of specific microbial enzymes that act in succession to metabolize commonly used but relatively less well understood human medications, such as levodopa (a medication used for Parkinson's disease). The second Keynote lecture was delivered by Eric Pamer, Director of the Duchossois Family Institute and Professor of Medicine at the University of Chicago. Eric presented exquisite mechanistic microbiology work that demonstrated the role of a gut commensal, *Blautia producta*, in contributing to colonization resistance to vancomycin-resistant Enterococcus. Specifically, he showed that an



B. Seelbinder



F. Key, S. Costa

R. Jones, S. Conlan



D. Segre, K. Graubics

antibiotic produced by a specific strain of *B. producta* is sufficient to limit Enterococcus colonization in an animal model.

The 217 participants in this meeting were highly engaged in a series of exciting talks from trainees, early investigators, mid-career investigators, and leaders in the field. Unpublished work shared at this meeting covered a range of topics from ecological network modeling to genome assembly to spatial features of the intestinal and oral microbiomes. The meeting also attracted participation from a variety of researchers from the biotechnology, pharmaceutical, and probiotic industries. Meeting participants were invited to attend two optional industry-sponsored events—a technology-focused presentation from Oxford Nanopore technologies and a multi-industry panel that addressed questions relating to industry–academia collaborations and the future prospects of microbiome-derived and microbiome-targeted therapeutics. Participants from 18 countries registered for the meeting and an encouraging proportion of women (43%) and trainees (44%) attended. More than 100 posters were presented in an extremely well-attended poster session, with topics presented ranging from using advanced microscopy/in situ hybridization approaches to visualize the microbiome of growing sea kelp to understanding the roles of bacteriophage in developing human infants. Several new collaborations were forged during this meeting, and many attendees remarked on how this is a meeting they will plan to attend every time it is offered.

## PROGRAM

**Introductory Session**
*Chairperson:* J. Segre, *National Human Genome Research Institute/NIH, Bethesda, Maryland*

**Keynote Address Speaker**
E.P. Balskus, *Harvard University, Cambridge, Massachusetts*

**Lightning Pitches**
**Metagenomes, Epigenomes, and Cancer**
*Chairperson:* M. Yassour, *Hebrew University of Jerusalem, Israel*

**Mathematical Modeling Microbial Communities Over Space and Time**
*Chairperson:* A. Bhatt, *Stanford University, California*

**Host-Microbiome-Community Assembly**
*Chairperson:* R. Dutton, *University of California, San Diego*

**Synthetic Biology, Bioengineering, and Spatial Studies of the Microbiome**
*Chairperson:* D. Segre, *Boston University, Massachusetts*

**Drugs, Bugs, and Small Molecules**
*Chairperson:* K. Guillemin, *University of Oregon, Eugene*

**Keynote Address Speaker**
E.G. Pamer, *University of Chicago, Illinois*

# Cell Death

**August 13–17**       157 Participants

ARRANGED BY       **David Andrews,** University of Toronto, Canada
**Simone Fulda,** Goethe-Universität Frankfurt, Germany
**Anthony Letai,** Dana-Farber Cancer Institute, Boston, Massachusetts

As usual, this meeting was regarded as the landmark cell death meeting of the year. More than 116 excellent presentations engendered lively discussion centered on topics in the field of cell death (see Program below for list of topics).

The Keynote Lectures were given by Doug Green on "Matters of Life and Death" and by Beth Levine, on "Beyond Cell Death to Autophagy—The Story of Bcl-2 and Beclin 1." Doug Green has been a leader in the Cell Death world for decades, and he is always a scintillating speaker. He informed and entertained us with his latest genetic dissection, this time of the disease autoimmune lymphoproliferative syndrome (ALPS). His most significant finding is that, based



A. Letai, D. Andrews, S. Fulda

on mouse modeling, ALPS is not caused by a lack of apoptosis as had been previously thought. In fact, his work contained the surprising hint that the microbiome may regulate the disease. Beth Levine's work has been instrumental in connecting cell death to autophagy. She has extended this work to find that Beclin 1 may interact with AMPK as well as BCL-2, forging a link to energy metabolism.

As is typical at this meeting, the talks were of excellent quality. Both Daohong Zhou and James Kirkland spoke about strategies to selectively induce apoptosis in senescent cells to reverse aging phenotypes. Brent Stockwell performed another tour de force on ferroptosis, which he discovered. Marcus Peter startled the audience with his explication of a brief RNA sequence that can selectively target survival genes and do so with greater efficiency in cancer cells. We learned from several speakers, including Marcela Maus and Alexandra Pourzia, about how cytotoxic T cells use cancer cell programmed cell death pathways to kill. We also learned from Judy Lieberman



C. Bortner, J. Lemke



P. Meyer, P. Meier

how Gasdermin E can be proteolytically activated, opening up the possibility that there may be many occult programs of cell death that are activated by compartment-specific proteases. Of the oral presentations, 42% were given by women, reflecting well the overall composition of the meeting.

An innovation previously suggested by Sandy Zinkel was incorporated yet again. Before both poster sessions, time in the main session was devoted to lightning talks by poster presenters. For each poster section, 10 presenters gave two-minute talks to encourage poster attendance. This seemed to be successful as the 69 posters were presented in two poster sessions that were crowded and loud.

There is little doubt that attendance suffered somewhat because of the unlucky circumstance of several cell death meetings having been held in proximity to this meeting. Nonetheless, once again, the quality of presentations and attendees was matchless, and the informal interactions at meals and in the bar provide the added value found in no other venue. This meeting continued the tradition of great cell death science in a great setting at Cold Spring Harbor.

J. Weinberg, A. Linkermann

## PROGRAM

**Keynote Address Speakers: Apoptosis and Autophagy**
D. Green, *St. Jude Children's Research Hospital;* B. Levine, *University of Texas Southwestern Medical Center, Dallas*

**Cell Fate and Cell Cycle**
*Chairpersons:* D. Zhou, *University of Florida, Gainesville*; J. Kirkland, *Mayo Clinic, Rochester, Minnesota*

**Toxic Cell Death and Ferroptosis**
*Chairpersons:* M. MacFarlane, *MRC Toxicology Unit, Leicester, United Kingdom*; B. Stockwell, *Columbia University, New York*

**Cancer Cell Death**
*Chairpersons:* C. Dive, *Cancer Research UK Manchester Institute, United Kingdom*; M. Peter, *Northwestern University, Chicago, Illinois*

**BCL-2 Family Proteins**
*Chairpersons:* E. Cheng, *Memorial Sloan Kettering Cancer Center, New York*; A. Strasser, *The Walter and Eliza Hall Institute of Medical Research, Parkville, Australia*



D. Kulms, S. Larisch



D. Green, J. Witkowski

**Death Receptors, IAPs, and Necrosome**

*Chairpersons:* P. Meier, *The Institute of Cancer Research, London, United Kingdom*; P. Vandenabeele, *VIB, Ghent, Belgium*

**Immunogenic Cell Death**

*Chairpersons:* J. Magarian Blander, *Weill Cornell Medicine, New York*; J. Borst, *Leiden University Medical School, the Netherlands*

**Novel Cell Death Regulation I**

*Chairpersons:* B. Kile, *Monash University, Melbourne, Australia*; H. McBride, *McGill University, Montreal, Quebec, Canada*

**Novel Cell Death Regulation II**

*Chairpersons:* K. Ryan, *Cancer Research UK Beatson Institute, Glasgow, United Kingdom*; X. Wang, *National Institute of Biological Sciences, Beijing, China*

**Immunologic Cell Death**

*Chairpersons:* J. Lieberman, *Boston Children's Hospital/ Harvard Medical School, Massachusetts*; M. Maus, *Massachusetts General Hospital, Charlestown*

**Mitochondrial Cell Death and Disease**

*Chairpersons:* M. Falk, *Children's Hospital of Philadelphia, University of Pennsylvania*; S. Tait, *Cancer Research UK Beatson Institute, Glasgow, United Kingdom*

# Eukaryotic mRNA Processing

**August 20–24**        277 Participants

ARRANGED BY        **Alberto Kornblihtt,** University of Buenos Aires, South America
**Jens Lykke-Andersen**, University of California, San Diego
**Karla Neugebauer,** Yale University, New Haven, Connecticut

This 11th Eukaryotic mRNA Processing meeting presented and discussed recent developments in mRNA metabolism in eukaryotes. The meeting consisted of sessions on the topics listed below under Program.

As per tradition for this meeting, oral presentations were selected from submitted abstracts focusing on unpublished work primarily from graduate students, postdoctoral researchers, and junior faculty. This meeting continues to serve a critical role in the field as a place for young researchers to present their research to an international audience and to discuss new and exciting unpublished findings. This year, there were 80 selected talks and 120 posters, with 277 participants in total. This was slightly lower than



J. Lykke-Andersen, K. Neugebauer, A. Kornblihtt

recent years, likely resulting from a competing symposium on RNA Control and Regulation held two months earlier at CSHL.

There were many research highlights presented at the meeting in both the platform talks and poster presentations. In the session on Mechanisms of RNA Splicing, state-of-the-art cryo-EM, single-molecule microscopy, and high-throughput sequencing methods had been used to reveal the structure of a new intermediate and detailed kinetics of the pre-mRNA splicing process. Moreover, new molecular insights into a range of disorders associated with mutations in splicing factors were detailed in this session as well as the session on RNA Processing in Disease. In sessions on Alternative Splicing, there were presentations on mechanisms and networks of alternative splicing, as well as their biological consequences; a particular highlight was the identification of a temperature-regulated splicing factor kinase that might regulate alternative splicing during circadian rhythms and in reptile sex determination.



T.C.T. Lan, E. Guo



W. Shao, Y. Matsuura

343

C. Gooding, J. Quick-Cleveland, J. Talkish



M. Akinyi, J. Staley

The 3′-End Processing session focused on new insights into the mechanism and regulation of mRNA 3′-end processing and on the 3′-end processing of small noncoding RNAs. New sophisticated techniques to globally identify RNA modifications and structures in cells were presented in the Modification and Structure sessions. Sessions on Cotranscriptional RNA Processing focused on the timing and coordination between transcription and RNA processing events, including splicing. In the session on RNA Turnover and Quality control, new insights into mechanisms of cotranslational mRNA quality control pathways were presented, as well as new findings on the central role of deadenylation in mRNA decay. A highlight was a presentation on an unexpected structure formed by poly(A) that provides poly(A) specificity for a central mRNA deadenylase as depicted on the abstract book cover. A session about Viral RNAs focused on how viruses manipulate host and viral mRNAs to the viruses' advantage, and the final session of the meeting on RNA–Protein Interactions covered a number of areas of eukaryotic mRNA processing where RNA–protein complexes have central roles.

Clearly, research into RNA processing continues to bring about new important and surprising insights into cellular processes and disease!

This meeting was funded in part by the National Cancer Institute, a branch of the National Institutes of Health.

## PROGRAM

**Mechanisms of RNA Splicing**
*Chairpersons:* R. Zhao, *University of Colorado, Denver*; A. Hoskins, *University of Wisconsin, Madison*

**Alternative Splicing I: Mechanisms and Networks**
*Chairperson:* K. Lynch, *University of Pennsylvania, Philadelphia*

**Alternative Splicing II: Biological Consequences**
*Chairperson:* M. Ares, *University of California, Santa Cruz*

**3′-End Processing**
*Chairpersons:* Y. Shi, *University of California, Irvine*; B. Tian, *Rutgers University, Newark, New Jersey*

**Gene Regulation by RNA Modifications**
*Chairperson:* W. Gilbert, *Yale University, New Haven, Connecticut*

**RNA Structure in Cells**
*Chairperson:* A. Laederach, *University of North Carolina, Chapel Hill*

**Cotranscriptional RNA Processing I: Dynamics of Cotranscription RNA Processing**
*Chairperson:* J. Beggs, *University of Edinburgh, United Kingdom*

**Cotranscriptional RNA Processing II: Communication between the RNA Processing and Transcriptional Machineries**
*Chairperson:* S. Murphy, *University of Oxford, United Kingdom*

**RNA Turnover and Quality Control**
*Chairpersons:* R. Hogg, *NHLBI/National Institutes of Health, Bethesda, Maryland*; A. Pasquinelli, *University of California, San Diego*

**Viruses and RNA Processing**
*Chairperson:* N. Conrad, *University of Texas Southwestern Medical Center, Dallas*

**RNA Processing in Disease**
*Chairperson:* T. Cooper, *Baylor College of Medicine, Houston, Texas*

**RNA–Protein Interactions/RNP Complexes**
*Chairpersons:* T. Preiss, *Australian National University, Canberra*; B. Graveley, *University of Connecticut Health Center, Farmington*



T. Cooper, P. Boutz

# Mechanisms of Eukaryotic Transcription

**August 27–31**     346 Participants

ARRANGED BY     **Patrick Cramer,** Max Planck Institute, Germany
**Michael Levine,** Princeton University, New Jersey
**Jane Mellor,** University of Oxford, United Kingdom

Transcription is the first step in the expression of the genome, and its regulation has a central role in the growth and development of eukaryotic organisms. Transcriptional responses occur as a consequence of cell signaling, environmental stresses, and developmental cues. The transcription field ranges from biochemistry and structural biology all the way to functional genomics, stem-cell research, and developmental biology. This meeting covered many aspects of the field and brought together a diverse group of scientists, including scientists from related fields who never participated before.

This 16th CSHL meeting on eukaryotic transcription consisted of eight plenary sessions and four poster sessions. It also marked the 50th anniversary of the discovery of the three eukaryotic RNA polymerases by Roeder and Rutter and highlighted this occasion with a special anniversary lecture on the first evening given by Bob Roeder. This lecture was very inclusive and motivating, and it created a collaborative and friendly atmosphere from the very beginning that could be felt throughout the meeting.

The meeting was structured based on the steps in the cycle of transcription, as before, with seven sessions encompassing 10 talks each (the sessions are listed below under Program). Care was taken to include several junior fellows as speakers who were at the postdoc stage or had just started their independent careers as principal investigators, and this was highly appreciated not only by these speakers, but also by the audience. Several new topics were also included via invited speakers or selected speakers, such as the mechanisms of chromatin transcription, histone variants in transcription, single-molecule nanoscopy, enhancers in plant development, zygotic



M. Levine, P. Cramer
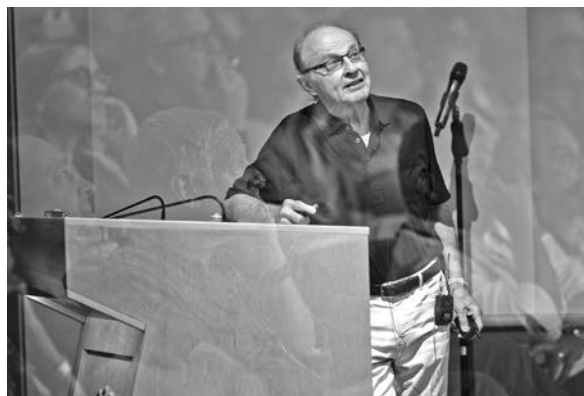


R. Roeder



A. Berk, B. Stillman

M. Ueberschar, Z. Okur



S. Churchman, R. Young

genome activation, single-cell genomics and transcription mechanisms, and transcription and metabolism.

Interspersed with these oral presentations were the four poster sessions (Sessions 1, 3, 6, and 10), in which a wide variety of exciting unpublished transcriptional research was presented. The posters were extremely thoughtful and covered state-of-the-art research. The evening format on the first day, together with a reception, provided a relaxed atmosphere for discussions. Overall, the meeting provided an overview of the state of the art of this large research field and its related areas. The meeting was oversubscribed, demonstrating the strong interest in the field.

This meeting was extremely well received by the participants and will occur again in 2021, then organized by Jane Mellor, Michal Levine, and Eva Nogales.

This meeting was funded in part by Arima Genomics.

## PROGRAM

**Anniversary Opening Lecture**
*Chairperson:* R.G. Roeder, *The Rockefeller University, New York*

**Initiation Mechanisms**
*Chairperson:* F. Robert*, IRCM, Montréal, Quebec, Canada*

**Elongation Mechanisms**
*Chairperson:* D. Luse, *Cleveland Clinic Lerner Research Institute, Ohio*

**Chromatin Transcription and Condensates**
*Chairperson:* D. Slade, *University of Vienna, Austria*

**Enhancers and Transcription Regulation**
*Chairperson:* A. Berk, *University of California, Los Angeles*

**Cotranscriptional RNA Processing and Termination**
*Chairperson:* P. Verrijzer**,** *Erasmus Medical Center, Rotterdam, the Netherlands*

**Transcriptional Regulation, Development, and Medicine**
*Chairperson:* A. Boija, *Whitehead Institute, Cambridge, Massachusetts*



A. Eccleston, K.J. Armache

**Emerging Concepts and Technologies**
*Chairperson:* S. Vos**,** *Max Planck Institute for Biophysical Chemistry, Göttingen, Germany*

# Eukaryotic DNA Replication and Genome Maintenance

**September 3–7**     361 Participants

ARRANGED BY     **Karlene Cimprich,** Stanford University, California
**John Diffley,** The Francis Crick Institute, United Kingdom
**Anne Donaldson,** University of Aberdeen, United Kingdom

This meeting provides an important forum for discussion and exchange of ideas in the field of chromosome replication and genome stability. The strategy adopted several years ago of broadening the conference scope to encompass the intersection of replication processes with genome stability mechanisms continues to work very well. It has expanded the range of the research covered and brought together biochemists, cell biologists, and geneticists in new and meaningful ways. As an indication of the meeting's strength, the number of participants (361), the percentage of female participants (43.5%), and the percentage of students and postdocs (54.9%) have continued to rise steadily over the last several iterations of the meeting. The meeting



K. Cimprich, J. Diffley, A. Donaldson

featured 80 platform talks and more than 200 posters. The poster sessions were packed throughout and distinguished by lively discussion. The Nicholls Biondi poster pavilion has proven to be a success. For the second meeting in a row, both poster sessions were held in the evenings. This was very well received by presenters and viewers alike.

The 10 platform sessions reflected the great breadth of the field, covering a wide range of topics from detailed molecular mechanisms to global regulation, chemical tools, and human disease. Approaches ranged from high-resolution cryo-electron microscopy through to genome-wide analysis of replication origin use and timing. Similarly, talks involved a wide range of model organisms complementing talks involving human cells. Talks drew many stimulating questions from the audience and provoked ongoing discussion continuing through the social parts of the program. The meeting began with two sessions covering the mechanism and regulation of replication initiation, replisome assembly, and replication termination. Highlights in these sessions included cryo-EM



E.A.A. Obara, B. Stillman



N. Shastri, V. Jagannathan

348

M. Sokka, S. Gerbi



D. Baretic, K. McCluskey

structures of the replisome and the helicase loading process, and the continued dissection of the replication termination reaction. Three sessions covered various aspects of processes critical for ensuring stable genome maintenance: Fork Stalling, Recovery, and Generation of Mutation; Responses to Replication and Transcription Stress; and Checkpoint Signaling, Stress, and the Cell Cycle. Among the highlights from these sessions was an increasingly clear view of mechanisms involved in fork protection and the emerging importance of the PrimPol pathway in replication restart. The session on origin mapping included talks describing the use of nanopore sequencing and high-content imaging to monitor replication. There is still a lack of consensus on what constitutes a replication origin in human cells, but bringing these exciting technologies into the field should help resolve this issue. Highlights from the sessions on Chromatin and Replication Timing and on Chemical Tools and Human Disease; ORC and Replication Origins included evolutionary and structural analysis of sequence-specific DNA binding by yeast ORC and further understanding of how replication–transcription conflicts occur and are resolved.

This meeting continues to be the preeminent meeting in the field. The quality of the presentations and discussions indicates this will continue to be the case going forward.

## PROGRAM

**Replication Initiation Mechanisms and Requirements**
*Chairpersons:* P. Zegerman, *University of Cambridge, United Kingdom*; A. Costa, *The Francis Crick Institute, London, United Kingdom*

**Replisome Structure, and Termination**
*Chairpersons:* M. O'Donnell, *The Rockefeller University, New York*; M. Kanemaki, *National Institute of Genetics, Mishima, Japan*

**Fork Stalling, Recovery, and Generation of Mutation**
*Chairpersons:* T. Kunkel, *NIEHS/National Institutes of Health, Research Triangle Park, North Carolina*; A. Vindigni, *Washington University School of Medicine in St. Louis, Missouri*

**Responses to Replication and Transcription Stress**
*Chairpersons:* J. Downs, *The Institute of Cancer Research, London, United Kingdom*; A. Donaldson, *University of Aberdeen, United Kingdom*



D. Remus, D. Boos

**Chemical Tools and Human Disease; ORC and Replication Origin**

*Chairpersons:* J. Morris, *University of Birmingham, United Kingdom*; M.K. Raghuraman, *University of Washington, Seattle*

**Origin Mapping**

*Chairpersons:* A. Dutta*, University of Virginia School of Medicine, Charlottesville*; C. Fox, *University of Wisconsin, Madison*

**Chromatin and Replication Timing**

*Chairpersons:* A. Groth, *University of Copenhagen, Denmark*; C. Sansam, *University of Oklahoma Medical Research Foundation, Oklahoma City*

**Checkpoint Signaling, Stress, and the Cell Cycle**

*Chairpersons:* S. Lambert*, Institut Curie, CNRS, Orsay, France*; A. Nussenzweig, *NCI/National Institutes of Health, Bethesda, Maryland*

# Microbial Pathogenesis and Host Response

**September 10–14**     290 Participants

ARRANGED BY     **Denise Monack,** Stanford University, California
**Anita Sil,** University of California, San Francisco
**Raphael Valdivia,** Duke University Medical Center, Durham, North Carolina

Despite advances in modern healthcare, infectious diseases continue to be major causes of human morbidity and mortality. The evolution of microbial pathogens with humans has resulted in complex interactions that impact the struggle between the infectious invader and the susceptible host. Increased understanding of these interactions with the goal of developing new therapeutics and preventive strategies requires collaborative and interdisciplinary scientific approaches. This meeting brought together a diverse group of international scientists who approach the study of bacterial, parasitic, and fungal pathogens from a broad range of perspectives. Investigators from the disciplines of



A. Sil, R. Valdivia, D. Monack

molecular microbiology, eukaryotic cell biology, immunology, and genomics and those representing academia, scientific publishing, industry, and the public health sector shared recent findings concerning microbial and host aspects of infectious diseases.

The meeting focused on the cross-talk between microbial pathogens and the host, facilitating an increased understanding of host response and defense mechanisms to these invading microbes. Oral sessions were topic-based and included studies of multiple and diverse organisms (see Program below for a list of these sessions). The session on Gastronauts specifically focused on gut bacteria and their influence on health. Speakers for each session were a mixture of established leaders in the field and young investigators. Half of the speakers, including postdoctoral fellows and graduate students, were chosen from submitted abstracts. Dr. Lalita Ramakrishnan, an internationally recognized leader in molecular mechanisms of mycobacterial disease, presented the Keynote Address. Her presentation, "A Zebrafish Guide to the Pathogenesis of



I. Derre, R. Watson



F. Cakar, C. Lim

S. Gilk, I. Derre



M. Byndloss, K. Patras

Tuberculosis," provided an exciting story of how she has integrated mutant screens in zebrafish, human genetics, and mechanistic characterization of human and zebrafish phenotypes to guide therapies against tuberculosis.

The informal atmosphere combined with the broad perspectives of the meeting participants resulted in a free flow of novel and refreshing ideas on pathogenesis and clinical treatment, with the atmosphere of a small meeting. Extensive question and answer sessions followed each oral presentation. The poster sessions were engaging with vibrant discussion and continued during a wine and cheese reception and other social gatherings. We strongly encouraged submission of abstracts by junior researchers in the field, and many young investigators were in attendance. Some of these interactions have already produced fruitful scientific collaborations. This year's meeting also included opportunities for trainees to interact more formally with speakers at Meet the Speaker lunches and was also an informal gathering for trainees and junior faculty who are seeking to network for further training opportunities.

## PROGRAM

**Microbial Evolution and Pathogenesis**
*Chairperson:* H. Malik, *Fred Hutchinson Cancer Research Center, Seattle, Washington*

**Cell Biology of Pathogen Infection I**
*Chairperson:* K. Orth, *HHMI/University of Texas Southwestern Medical Center, Dallas*



E. Bratanis, M. Collin
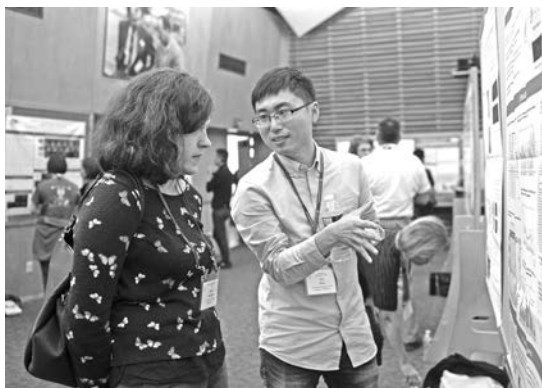


S. Perelman, X. Zhao

**Heterogeneity in Pathogen Populations, Antibiotic Resistance, and Disease Outcomes**
*Chairperson:* D. Weiss, *Emory University School of Medicine, Atlanta, Georgia*

**Regulation of Pathogen Virulence**
*Chairperson:* A. Darwin, *New York University School of Medicine, New York*

**Innate Immunity and Cell Biology during Pathogen Infection**
*Chairperson:* S. Shin, *University of Pennsylvania Perelman School of Medicine, Philadelphia*

**Gastronauts: A Journey through the Gut**
*Chairperson:* V. Sperandio, *University of Texas Southwestern Medical Center, Dallas*

**Keynote Address Speaker**
L. Ramakrishnan, *University of Cambridge, United Kingdom*

# Stem Cell Biology

**September 17–21**     170 Participants

ARRANGED BY     **Anne Grapin-Botton,** University of Copenhagen, Denmark
**Marius Wernig,** Stanford University, California
**Kenneth Zaret,** University of Pennsylvania/Perelman School of Medicine, Philadelphia

This fifth biannual conference hosted 170 participants. Roughly 23% of participants were graduate students and another 23% were postdocs from 20 different countries. Representatives from major publishing houses were present such as *Science, Nature Cell Biology*, and *Development.* The organizers sought to emphasize cross-disciplinary work on cellular and molecular mechanisms of stem-cell and stem-cell-related tissue systems and focus on new speakers for the meeting. The meeting was organized into thematic topics that were updated in the time since the last meeting (see Program below for a list of these topics). Speakers comprised a mix of internationally known leaders in the individual



M. Wernig, A. Grapin-Botton, K. Zaret

disciplines as well as emerging junior researchers, who presented their recent work. In addition, a total of 29 short talks were chosen from abstracts featuring postdocs and students as well as additional early career independent investigators. Main talks were 20–25 minutes, plus five to 10 minutes of discussion, whereas short talks were 10 minutes plus five minutes of discussion. The Keynote Address was given by Luis Parada, who presented exciting unpublished work on understanding the quiescence of glioblastoma stem cells and how that enables the cells to be resistant to chemotherapeutics and radiotherapy. Using the latest single-cell RNA-Seq technology and genetic lineage tracing, diverse new stem-cell populations were described at the meeting, including for lung, tendon, mammary, brain, liver, gut, blood, testis, and skin, above and beyond stem-cell populations that had already been described for such tissues. Numerous stem-cell/organoid models of human disease were presented, including for celiac disease, lung



R. Moses, O. Soudry



L. Antunes, I. Samuelson

M. Mansour, K. Jo


J. Kurian, R. Jothi

damage, viral infection of the brain, autism, hepatocellular carcinoma, and squamous cell carcinoma. Additional breakthrough talks focused on early fate decisions in development. Most of the presented work was unpublished and everyone agreed that the quality of discussion among the participants at the end of each talk was unusually high. There also were excellent discussions during poster sessions, meals, and at the bar, all facilitated by the intimate campus at Cold Spring Harbor. Preliminary (informal) feedback by students, postdocs, speakers, and journal editors was uniformly very positive. Attendees liked the mixed topics of sessions and the fact that most speakers stayed for several days and were approachable throughout the conference.

## PROGRAM

**Keynote Address Speaker**
L. Parada, *Memorial Sloan Kettering Cancer Center, New York*

**Stem-Cell Regeneration**
*Chairperson:* K. Zaret, *University of Pennsylvania Perelman School of Medicine, Philadelphia*

**Stem-Cell Transitions In Vivo**
*Chairperson:* S. Wickström, *University of Helsinki, Finland*

**Organoid for Modeling Stem-Cell Transitions and Disease**
*Chairperson:* G-l. Ming, *University of Pennsylvania, Philadelphia*

**The Stem-Cell Niche**
*Chairperson:* F. Guillemot, *The Francis Crick Institute, London, United Kingdom*

**Pluripotency and Reprogramming**
M. Wernig, *Stanford University School of Medicine, California*

**Stem-Cell Differentiation In Vitro for Modeling Disease**
*Chairperson:* E. Apostolou, *Weill Cornell Medicine, New York*


K. Kaestner, E. Morrisey

**Stem-Cell Epigenetics and Transcription**
*Chairperson:* M.E. Torres-Padilla, *Institute of Epigenetics & Stem Cells, Helmholtz Centre Münich, Germany*

**Tissue Engineering and Stem-Cell Mechanics**
*Chairperson:* A. Grapin-Botton, *University of Copenhagen, Denmark*

# Biology of Cancer: Microenvironment and Metastasis

**September 24–28**    253 Participants

ARRANGED BY    **Scott Lowe,** Sloan Kettering Memorial Cancer Center, New York
**Senthil Muthuswamy,** Princess Margaret Cancer Center, Toronto, Ontario, Canada
**M. Celeste Simon,** University of Pennsylvania Medical School, Philadelphia
**Valerie Weaver,** University of California, San Francisco

Our understanding of cancer is evolving rapidly. We now recognize that cancer is highly heterogeneous at the level of both genetic mutations and cell types within a growing tumor. Furthermore, we are now beginning to appreciate the fact that not only do cancers metastasize to different parts of the body, they can also have systemic effects that produce and influence pathologies in the organism as a whole. This meeting aimed to capture this complexity by considering the cancer problem from an integrated and organismal perspective. It highlighted emerging topics that are actively investigated by cancer researchers worldwide and brought together molecular, cellular, and computational biologists to dis-



M.C. Simon, V. Weaver, S. Muthuswamy, S. Lowe

cuss recent advances on a diverse array of topics such as circulating tumor cells, tumor immunology, cellular plasticity, stress adaptation, metastasis, cancer dormancy, cell metabolism, and systemic effects. All the talks were of extremely high caliber and led to extended and insightful discussion from the audience. The first Keynote Address speaker, Daniel Haber, discussed his elegant and powerful approaches to study circulating tumor cells and his recent finding on the role of ribosomal proteins in regulating function of metastasis-competent circulating tumor cells. In the second Keynote Address, Tyler Jacks presented studies that displayed the power and potential of single-cell sequencing to understand tumor progression in lung cancer and to identify cells with unique plasticity to modulate metastatic progression. Finally, to motivate trainees to stay engaged in cancer research, we incorporated a career mentoring session over lunch. Collectively, participants enjoyed both the diversity and the depth of the topics covered and felt proud to be part of this community.

Caricaturist at work



B. Becklund, F. Sulzmaier

B. Smith, N. Novaresi



D. Ingber, E. Williams

## PROGRAM

**Microenvironment and Matrix**
*Chairpersons:* V. Weaver, *University of California, San Francisco*; D. Ingber, *Wyss Institute at Harvard University, Boston, Massachusetts*; M. Swartz, *University of Chicago, Illinois*

**Phenotypic and Genotypic Heterogeneity**
*Chairpersons:* M. Brown, *Dana-Farber Cancer Institute, Boston, Massachusetts*; J. Aguirre-Ghiso, *Icahn School of Medicine at Mount Sinai, New York*

**Keynote Address Speaker**
D.A. Haber, *HHMI/Massachusetts General Hospital, Charlestown*

**Metabolism**
*Chairpersons:* R. DeBerardinis, *HHMI/University of Texas Southwestern Medical Center, Dallas*; J. Debnath, *University of California, San Francisco*

**Drug Resistance**
*Chairpersons:* C. Der, *University of North Carolina, Chapel Hill*; D. Tuveson, *Cold Spring Harbor Laboratory*

**Plasticity and Hypoxia**
*Chairpersons:* S. Berger, *University of Pennsylvania, Philadelphia*; I. Macara, *Vanderbilt University School of Medicine, Nashville, Tennessee*

**Metastasis and Dormancy**
*Chairpersons:* S. Morrison, *HHMI/University of Texas Southwestern Medical Center, Dallas*; P. Friedl, *Radboud University Medical Centre, Nijmegen, the Netherlands*

**Keynote Address Speaker**
T. Jacks, *MIT/Cambridge, Massachusetts*

**Senescence and Signaling**
*Chairpersons:* M. Park, *McGill University, Montreal, Quebec, Canada*; M. Serrano, *Institute for Research in Biomedicine (IRB), Barcelona, Spain*

**Tumor Immunology**
*Chairpersons:* A. Montovano, *Humanitas University, Milan, Italy*; P. Sharma, *MD Anderson Cancer Center, Houston, Texas*



D. Yang, M. Oudin



X.(C.) Sun, R. DeBerardinis

# Neurobiology of *Drosophila*

**October 1–5**       451 Participants

ARRANGED BY       **Kate O'Connor-Giles,** Brown University, Providence, Rhode Island
**Scott Waddell,** University of Oxford, United Kingdom

This meeting provided a forum for the discussion of new discoveries, techniques, and advances in *Drosophila* neurobiology. Eleven sessions ran in series over four days with alternating platform and poster presentations. Alternating between these more and less formal presentations provided excellent opportunities for attendees to meet one another and engage in one-on-one discussions. These interactions facilitated collaborations, exchange of reagents (e.g., antibodies, clones, mutants, and other stocks), methods (genetic, physiological, and optical), and ideas between both new and established investigators. These interactions are especially important for new scientists immersing in the field and building their professional networks.



K. O'Connor-Giles, S. Waddell

The eight platform session topics were chosen to reflect the areas in which cutting-edge advances are being made (see Program below for a list of these topics). Session chairs and the meeting organizers selected presenters for these platform sessions from submitted abstracts, whereas the remaining abstracts were presented as posters. The vast majority of the speakers were graduate students and postdoctoral fellows, and approximately one-half were female or members of historically underrepresented groups. Many people commented positively on the diversity of presentations and balance of research investigating the nervous system at different levels. The Seymour Benzer Lecture was presented by 2017 Nobel laureates Michael Young and Michael Rosbash. They presented engaging accounts of their seminal and ongoing research into the molecular mechanisms of the circadian clock in *Drosophila* and humans. The Elkins Award Memorial Lecture is presented at each meeting by a graduate student whose dissertation exemplifies the finest work in our field. This year, the Elkins Lecture was presented by Dr. Stephen X. Zhang, who trained as a graduate student with Dr. Michael Crickmore at Harvard, and who presented his exceptional work on neural circuit mechanisms underlying motivated courtship behavior.



J. Jepson, S. Lowe



A. Rothenfluh, D. Stewart

F.M. Tenedini, L. Solomon



B. van Alphen, E. de Tredern

The Neurobiology of *Drosophila* meeting was well attended (oversubscribed, in fact), with presentations spanning the breadth of modern neurobiology. The many opportunities for interaction and career development fostered by this meeting are sure to enhance this vibrant field.

This meeting was funded in part by the National Institute of Neurological Disorders and Stroke, a branch of the National Institutes of Health, and by WellGenetics.

## PROGRAM

**Higher Brain Function and Behavioral Plasticity**
*Chairperson:* K. Kaun, *Brown University, Providence, Rhode Island*

**Nervous System Development**
*Chairperson:* G. Tavosanis, *German Center for Neurodegenerative Diseases, Bonn, Germany*

**Sensory Systems**
*Chairperson:* L. Prieto-Godino, *The Francis Crick Institute, London, United Kingdom*

**Elkins Memorial Lecture Award Recipient**
S.X. Zhang, *Harvard Medical School/Boston Children's Hospital, Massachusetts*

**Neuronal Cell Biology**
*Chairperson:* J. Wildonger, *University of Wisconsin, Madison*

**Technological Innovations**
*Chairperson:* K. Branson, *HHMI/Janelia Research Campus, Ashburn, Virginia*

**Seymour Benzer Lecture**
*Chairpersons:* M. Rosbash, *Brandeis University, Waltham, Massachusetts*; M. Young, *The Rockefeller University, New York*

**Synaptic Transmission and Plasticity**
*Chairperson:* D. Dickman, *University of Southern California, Los Angeles*

**Regulation of Homeostatic Behaviors**
*Chairperson:* A. Seeds, *University of Puerto Rico, San Juan*

**Neurological Disease and Injury**
*Chairperson:* C. Collins, *University of Michigan, Ann Arbor*



O. Akin, T. Reis



W. Ja, M. Dus

# Genome Engineering: Frontiers in CRISPR/CAS

**October 10–13**    414 Participants

ARRANGED BY    **Jennifer Doudna,** HHMI/University of California, Berkeley
**Maria Jasin,** Memorial Sloan Kettering Cancer Center, New York
**J. Keith Joung,** Massachusetts General Hospital, Boston
**David R. Liu,** Broad Institute of Harvard and MIT, Cambridge, Massachusetts
**Jonathan Weissman,** HHMI/University of California, San Francisco

This meeting was the fifth consecutive conference in the series held at Cold Spring Harbor. The robust attendance continued again this year (>400) and it was planned to hold the meeting again in 2020. The meeting continued to be unusual in the number of corporate attendees (>100); there were also two workshops hosted by companies (Cell Microsystems and IDT). A mouse engineering workshop was held again during the meeting, bringing together many core heads from a number of institutions. A number of journalists were present, representing *Cell, Nature Protocols, The CRISPR Journal,* and *GenomeWeb.* Noted author and journalist Walter Isaacson, who is writing a book about gene editing, was also present and interacted with speakers and attendees.



J. Doudna, W. Isaacson

A new feature was an introductory talk to set the stage regarding recent genome engineering advances and controversies. This talk was presented by Fyodor Urnov from the IGI at Berkeley and was entitled "Our Genome-Edited World: A Data-Based Preview." Comments from attendees suggested that the talk was well received. An introductory talk should be considered at subsequent meetings pending identification of suitable individuals to present such a talk.

The goal of this meeting has been to bring together researchers working in diverse fields to stimulate discussions and ideas to further exploit CRISPR-Cas9 and related technologies for biological discovery, organismal engineering, and medical applications. Session titles are listed below under Program. More than 20 speakers were invited to cover these diverse topics and another almost 30



T.J. Gradick, S. Maragh



J. Weissman, S. Weissman

A. Cereseto, M. Durham



R. Turk, M. Sellman

speakers were chosen from submitted abstracts. Speakers represented institutions from the United States and abroad (Lithuania, Scotland, Denmark, Germany, United Kingdom, Holland, and Italy), with a few representatives from industry. Speakers chosen from submitted abstracts, laboratory heads/ staff scientists, and postdoctoral fellows included graduate students and lab heads and even a post-bac. Approximately 150 posters were presented in two sessions, complementing the oral presentations.

Talks throughout the meeting used molecular, cell, and computational biology in diverse model organisms, as well as economically important and some unconventional organisms. Much of the data presented was unpublished or only very recently published. Highlights included the introduction of prime editing by co-organizer David Liu, programmable transposition by Sam Sternberg, a relatively young faculty member at Columbia, and CRISPR-Cas systems from diverse ecosystems by Jill Banfield, a senior faculty member from the University of California, Berkeley, who was new to the meeting.

This meeting was funded in part by support from Agilent Technologies, by Benchling, by BEX Co., Ltd., by Cell Microsystems, by Integrated DNA Technologies, by MilliporeSigma, by Synthego, and by Thermo Fisher Scientific.

## PROGRAM

**Introductory Talk/CRISPR Biology I**
*Chairpersons:* E. Sontheimer, *University of Massachusetts Medical School, Worcester*; F. Urnov, *University of California, Berkeley*

**CRISPR Biology II/Technology I**
*Chairpersons:* E. Bier, *University of California, San Diego*; S. Sternberg, *Columbia University, New York*

**Cell and Embryo Engineering**
*Chairpersons:* D. Egli, *Columbia University, New York*; K. Niakan, *The Francis Crick Institute, London, United Kingdom*

**DNA Repair/Plants**
*Chairpersons:* B. Adamson, *Princeton University, New Jersey*; D. Ramsden, *University of North Carolina, Chapel Hill*



T.J. Cradick, A. Kernytsky

**Gene Therapy**
*Chairpersons:* M. Porteus, *Stanford University, California*; C. Gersbach, *Duke University, Durham, North Carolina*

**Technology II**
*Chairpersons:* W. Wei, *Peking University, Beijing, China*; M. Bassik, *Stanford University, California*

# Yeast Research: Origins, Insights, Breakthroughs

**October 23–26**     185 Participants

ARRANGED BY     **David Botstein,** Calico Life Sciences, South San Francisco, California
**Rochelle Esposito,** University of Chicago, Illinois
**Gerald Fink,** Whitehead Institute for Biomedical Research/MIT, Cambridge, Massachusetts
**Ludmila (Mila) Pollock,** Cold Spring Harbor Laboratory

Yeast has been an important model eukaryote in modern genetics, cell, and developmental biology. It has been a major research organism from the 1940s to the present because it is a simple eukaryote that can be handled like bacteria and has a short life cycle, a well-known biochemistry, a stable haplophase and diplophase, and a normal mitosis and meiosis. Early yeast studies focused primarily on isolating genetic markers and making chromosome maps, tetrad and polyploid analysis, gene fine structure and defining the basic properties of gene conversion and reciprocal recombination, identifying tRNA suppressors, mating-type alleles and MAT switching, mutants affecting mitochondrial



G. Fink, F. Winston

function, and genes regulating galactose fermentation. These investigations laid the foundation for rapid growth of the field, further stimulated by introduction of the CSHL course on yeast genetics in the 1970s.

Subsequent studies pioneered the analysis of mitosis, meiosis, and growth controls. These included identification of the first cell division cycle genes; DNA synthesis and progression controls; cohesion and condensin regulators; loci involved in meiotic initiation, recombination, centromere cohesion, and spore formation; centromere and telomere structure/function, actin cytoskeleton development; and spindle assembly. Yeast research also had a seminal role in

K. Nasmyth, A. Amon
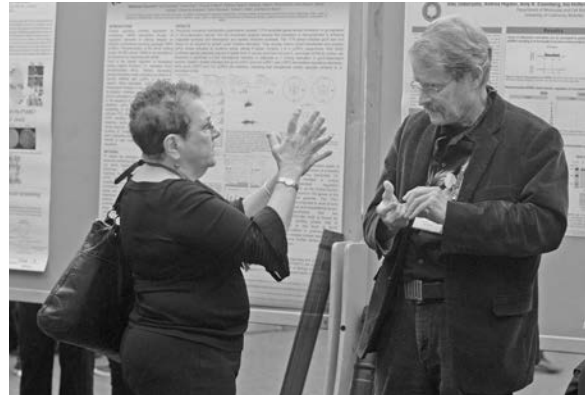


A. Hopper, J. Strathern

defining genes and mechanisms in RNA processing, gene silencing, protein trafficking and degradation, signaling, autophagy, and development of technologies used for recombinant DNA, gene replacement, and whole-genome transcriptional analysis. Yeast work set key paradigms for higher eukaryotes. With its relatively small genome (~6,000 genes) and flexible life cycle, it continues to provide groundbreaking insights in defining gene interactions affecting survival, aging, and human disease via its unique ability to allow assays in both haploids and diploids by two-hybrid, multiple-deletion, and suppressor analysis. To date, seven yeast investigators have received Nobel Prizes and eight have received Breakthrough prizes for their cutting-edge work; a number of them attended the meeting.

The main topics focused on key areas in which yeast research has made critical contributions (see Program below for a list of these topics). These sessions encompassed 63 lectures by the very top investigators in the field of yeast research, as well as approximately 90 poster presentations on a broad range of topics. The panel discussion elicited lively exchanges and included key representatives from academia and industry. In attendance were approximately 200 scientists, historians, scholars, and science journalists. In summary, this meeting was an exceptional symposium and a huge success, worthy of the high reputation of Cold Spring Harbor Laboratory Symposia. All talks are now available on the specially dedicated website at http://library.cshl.edu/Meetings/History-of-Science.

This meeting was funded in part by Zymergen, Inc.



J. Frydman, S. Liebman



P. Hieter, M. Dunham

K. Struhl, A. Hinnebusch



D. Gresham, A. Johnson

## PROGRAM

**Early Influential Yeast Centers**
*Chairperson:* G. Fink, *Whitehead Institute for Biomedical Research/MIT, Cambridge, Massachusetts*

**Mitosis, Meiosis, and Growth Controls**
*Chairpersons:* J. Pringle, *Stanford University School of Medicine, California*; R. Esposito, *University of Chicago, Illinois*

**RNA Synthesis, Processing, Translation, and Regulation**
*Chairperson:* C. Guthrie, *University of California, San Francisco*

**Gene Expression and Silencing**
*Chairpersons:* M. Johnston, *University of Colorado, Aurora*; M. Grunstein, *University of California, Los Angeles*

**Protein Transport, Autophagy, Degradation, and Signaling**
*Chairpersons:* N. Segev, *University of Illinois, Chicago*; J. Thorner, *University of California, Berkeley*

**Mitochondria, Metabolism and Aging**
*Chairperson:* T. Fox, *Cornell University, Ithaca, New York*

**Panel Discussion: Medical and Industrial Uses of Yeast**

**DNA Replication, Recombination, and Repair**
*Chairpersons:* J. Campbell, *California Institute of Technology, Pasadena*; T. Petes, *Duke University, Durham, North Carolina*

**Genomics and Evolution**
*Chairpersons:* R. Davis, *Stanford University, California*; E. Louis, *University of Leicester, United Kingdom*



D. Gottschling, S. Emr



R. Wickner, M. Olson

# Genome Informatics

**November 6–9**  350 Participants

ARRANGED BY  **Nicholas Loman,** University of Birmingham, United Kingdom
**Alicia Oshlack,** Murdoch Children's Research Institute, Australia
**Melissa Wilson,** Arizona State University, Tucson

This 15th Cold Spring Harbor Laboratory/Wellcome Trust conference continues to demonstrate a robust attendance, both from the United States and from abroad. It attracted more than 345 registrants presenting 236 abstracts, offering a snapshot of the latest developments in the field. There were 12 invited talks and two Keynote Addresses. The remaining 35 talks were all selected for presentation by session chairs from openly submitted abstracts; 189 posters were presented, of which 28 student posters were also given as one-minute lightning talks.

This year, abstracts covered a wide variety of genomic analyses, with a special emphasis on innovations in genetic variant discovery, data visualization, and new insights gained from integrating large-scale genomics data



M. Wilson, A. Oshlack, N. Loman

sets. There was also an emphasis on new developments in metagenomics and single-cell analysis. The sessions in this conference are listed below under Program.

The first Keynote Address, delivered by Dr. Jonathan Pritchard, focused on new developments and approaches to understand the genetics of complex traits. The second Keynote Address, delivered by Dr. Dana Pe'er, discussed new data and analysis approaches in single-cell genomics. Both talks were extremely well received and led to extensive follow-up discussions.

This meeting was very actively discussed on Twitter (using hashtag #gi2019), with the social media policy being opt-out (and virtually all talks being tweetable). More than 2000 messages were broadcast by users around the world to discuss and debate the ideas presented.

This meeting was funded in part by the National Human Genome Research Institute, a branch of the National Institutes of Health.



I. Friedberg, P. Melsted



G. Wheeler, H. Li

S. Brown, V. Tomic



Z. Sun, N. Zhou

## PROGRAM

**Genome Structure and Function**

*Chairpersons:* K. Le Roch, *University of California, Riverside*; E. Eichler, *University of Washington, Seattle*

**Sequencing Algorithms, Variant Discovery, and Genome Assembly**

*Chairpersons:* B. Langmead, *Johns Hopkins University, Baltimore, Maryland*; H. Li, *Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts*

**Lightning Talks**
**Transcriptomics**

*Chairpersons:* A. Conesa, *University of Florida, Gainesville*; R. Patro, *University of Maryland, College Park*

**Microbial and Metagenomics**

*Chairpersons:* A. Dilthey, *Heinrich Heine University Düsseldorf, Germany*; L. Cowley, *University of Bath, United Kingdom*

**Evolution and Phylogenetics**

*Chairpersons:* W. Hanage, *Harvard T.H. Chan School of Public Health, Boston, Massachusetts*; J. Kelso, *Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*

**Keynote Address Speakers**

J. Pritchard, *Stanford University, California*; D. Pe'er, *Sloan Kettering Institute, New York*

**Personal and Medical Genomics**

*Chairpersons:* S. Ramachandran, *Brown University, Providence, Rhode Island*; E. Worthey, *University of Alabama, School of Medicine, Birmingham*



T. Ballinger, S. Barreira



J. Chin, S. Koren

# Single-Cell Analyses

**November 13–16**    235 Participants

ARRANGED BY    **Nancy Allbritton,** University of Washington, Seattle
**Scott Fraser,** University of Southern California
**Junhyong Kim,** University of Pennsylvania, Philadelphia

The goal of this fifth workshop was to bring together scientists who analyze and engineer single cells using a wide variety of experimental paradigms to discuss the progress that is being made. More than 200 scientists convened with two Keynote Address talks, 30 oral presentations, and 75 posters. This year, we celebrated the 10-year anniversary of the first CSHL Single-Cell Analyses meeting in 2009, organized by Sunney Xie and Jim Eberwine. The two organizers were invited as the Keynote Address speakers for the 2009 meeting, which was the first meeting on single-cell biology in the field, and both Sunney and Jim presented progress of the field for the last 10 years. Participants in the meeting presented work addressing many different areas (see Program below for list of topics). Applications of single-cell technologies were presented in a session for "Single Cells in Development." A session was organized with two-minute flash talks, which was very popular and covered a wide range of topics that were presented in detail in the poster sessions.



J. Musser, J. Kim

As in previous years, much of the meeting concentrated on covering new technologies and methods, such as methods to sample contents from a single cell while maintaining viability, novel integrated imaging methods, and advances in single-cell proteomics, as well as multimodal measurements from single cells. We also had presentations of new single-cell modeling approaches, including modeling of cell differentiation as vector fields on a manifold. Remarkable advances in high-throughput methods were presented with single-cell-level analysis of developmental processes ranging from translational applications in kidney development, to evolution of chordate development, to sponge neural systems. The results presented in the meeting continued to demonstrate



Single-Cell Analyses group photo

S. Higgins, J. Quinn



S. Mohammadi, A. Bayegan

the explosive impact of single-cell biology on normal human health, disease states, and insight into evolutionary mechanics. All participants remarked on the breadth of the meeting and how refreshing the meeting was in its concentration on novel developments, rather than single topics such as cell atlas projects as is prevalent in other meetings. There was consensus that single-cell biology is now an established research area with continued growth, and there was also support for holding meetings at annual intervals.

This meeting was funded in part by Bio-Techne, by Cell Microsystems, by SeqGeq, by Takara Bio, and by 10x Genomics.

## PROGRAM

**Keynote Address Speaker**
X.S. Xie, *Peking University, Beijing, China/Harvard University, Cambridge*

**Spatial Single Cell Biology**
*Chairperson:* L. Keren, *Stanford University, California*

**Epigenome, Informatics, and Applications**
*Chairperson:* J. Kim, *University of Pennsylvania, Philadelphia*

**Flash Talks**
*Chairperson:* J. Kim, *University of Pennsylvania, Philadelphia*

**Photonics and Imaging for Single-Cell Biology**
*Chairperson:* H. Garcia, *University of California, Berkeley*

**Single Cells in Development**
*Chairperson:* E. van Nimwegen*, University of Basel, Switzerland*

**Keynote Address Speaker**
J. Eberwine, *University of Pennsylvania, Philadelphia*

**Physical/Chemical Single-Cell Biology**
*Chairperson:* S. Fraser, *University of Southern California, Los Angeles*



D. Mitroi, M. Ford



J. Güler, G. John-Schuster

# Zebrafish Neural Circuits and Behavior

**November 20–23**  187 Participants

ARRANGED BY  **Martha Bagnall,** Washington University, St. Louis
**Michael Granato,** University of Pennsylvania, Philadelphia
**Mark Masino,** University of Minnesota, Minneapolis
**Ethan Scott,** University of Queensland, Australia

DEPUTY ORGANIZERS  **Rainer Friedrich,** Friedrich Miescher Institute, Switzerland
**David Schoppik,** New York University School of Medicine, New York

This inaugural meeting brought together an international group of neuroscientists using the zebrafish as a model system to understand fundamental problems in neurobiology. The community had come together in smaller venues annually over the past decade to share the latest findings, disseminate innovative technologies, and begin new collaborations. This year, in keeping with the tradition of the best CSHL meetings, speakers were selected from submitted abstracts, allowing the meeting to highlight the most promising ideas, and organized into sessions that reflected the state of the field. As a result, there were nine sessions featuring short talks on topics varying from development and glia to neuropsychiatric disease and translation, as well



M. Masino, E. Scott, M. Bagnall, D. Schoppik, R. Friedrich, M. Granato

as a core set of sessions focused on more traditional approaches to understand sensory systems and motor control and behavior. Sessions were chaired by faculty members charged with facilitating discussion and introducing speakers. Across the meeting, gender and rank were well balanced: six of nine session chairs and 20/43 speakers were female, and 29/43 speakers were either graduate students or postdoctoral fellows. The poster sessions were lively: Highlights ranged from the latest in unpublished design improvements of microscopes for in vivo imaging to advances in the molecular underpinnings of complex behaviors in health and disease. Ample time was allotted for discussion at both social events and later in the evenings, where trainees and faculty mingled. A number of new



S. Wahlstrom-Helgren, J. Montgomery



M. Granato, W. Driever

369

A. Vaziri, M. Ahrens


B. Sanghera, J. Kanwal

collaborations were proposed among laboratories that had not worked together, and reagents were shared, moving the field forward. There was considerable enthusiasm to hold this meeting on a biannual basis, beginning in November 2022.

## PROGRAM

**Development and Glia**
*Chairperson:* L. Sheets*, Washington University School of Medicine in St. Louis, Missouri*

**Sensorimotor Circuits**
*Chairperson:* G. Downes, *University of Massachusetts, Amherst*

**New Tools: Anatomy, Imaging**
*Chairperson:* F. Kubo, *National Institute of Genetics, Mishima, Japan*

**Social Behaviors and Sleep**
*Chairperson:* H. Okamoto, *RIKEN Center for Brain Science, Wako, Japan*

**Learning, Memory, Neuromodulation**
*Chairperson:* C. Wee, *A\*STAR (Agency for Science, Technology & Research), Singapore*

**Motor Control and Behavior**
*Chairperson:* D. Prober, *California Institute of Technology, Pasadena*

**Neuropsychiatric Disease and Translation**
*Chairperson:* E. Hoffman, *Yale University, New Haven, Connecticut*


Y. Tanimoto, I. Jeong

**Community Session**
  **Sensory Systems**
*Chairperson:* K. Kindt, *National Institute on Deafness and Other Communication Disorders/NIH, Bethesda, Maryland*

**New Tools: Genetics**
*Chairperson:* M. Halpern, *Carnegie Institute of Washington, Baltimore, Maryland*

# Plant Genomes, Systems Biology and Engineering

**December 4–7**     133 Participants

ARRANGED BY      **Zachary Lippman,** Cold Spring Harbor Laboratory
**Jane Parker,** Max-Planck Institute for Plant Breeding Research, Germany
**Seung (Sue) Rhee,** Carnegie Institution for Science, Washington, D.C.

This meeting demonstrated the power of genome-enabled plant biology in a broad spectrum of areas ranging from environmental adaptation to metabolic network modeling and crop improvement. Participants and presenters were a healthy mix of junior and established scientists, staying true to the vision of the first meeting in 1997 and the spirit of Cold Spring Harbor Laboratory. A breadth and depth of exciting new, unpublished work toward discovery of fundamental principles and direct applications in agriculture was presented.



J. Parker, Z. Lippman, S.(S.) Rhee

A major theme across sessions was of that of multifaceted interactions of genes, proteins, and metabolites within and across organisms, as well as between organisms and their environments. For example, some talks discussed the importance of genetic interactions discovered through pan-genome sequencing and phenotypic studies, revealing many combinations of alleles that alter fitness, including the discovery of major immune receptor interactions. Other talks described the importance of microbiome in the plant rhizosphere, and the interactions between soil, minerals, and plants in plant mineral nutrition.

Another prominent thread in the meeting related to engineering plants to improve the health of people and the planet. Examples included installing alternative synthetic photorespiratory pathways to increase photosynthetic efficiency, discovering and engineering colchicine biosynthetic pathways, and developing synthetic biology tool kits to alter root architecture. Additional topics included engineering efficient plant–rhizobia interactions as well as developing tools to enable synthetic biology on nonmodel plants. These new approaches are enabling the possibility of

L. Feiz, K. Joshi



T. Liu, C. Huang

engineering plants that could dramatically change our bio-economy. Descriptions of state-of-the-art, high-throughput technologies and genome biology were complemented nicely by presentations on genetic, molecular, and spatial dissections of biological processes in model and crop species, including important new insights into plant defense mechanisms.

The Keynote Addresses were presented by Dr. Hailing Jin and Dr. Detlef Weigel. Dr. Jin's talk focused on the cross-kingdom communications via small RNAs that regulate gene expression. The work she described has huge implications in protecting crops against pathogens sustainably by using RNAcides. Dr. Weigel presented work on epistatic interactions of genes at the genomic scale and their implications for plant adaptation. The work described has implications for plant engineering aimed at mitigating species extinction due to global warming.

## PROGRAM

**Keynote Address Speaker**
D. Weigel, *Max Planck Institute for Developmental Biology, Tübingen, Germany.*

**Crop Biology and Trait Enhancement**
*Chairperson:* J. Vogel, *DOE Joint Genome Institute, Walnut Creek, California*

**Genomes and Epigenomes**
*Chairperson:* M. Gehring, *Whitehead Institute for Biomedical Research, Cambridge, Massachusetts*

**Biodiversity and Environmental Adaptation**
*Chairperson:* U. Krämer, *Ruhr University Bochum, Germany*



D. Weigel, K. Krasileva



D. Jackson, P. Borrill, W. Busch

**Plants and Microbes**
*Chairperson:* J-M. Zhou, *Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China*

**Metabolic Circuits**
*Chairperson:* D. Ort, *University of Illinois, Urbana*

**Development: Modules to Networks**
*Chairperson:* M. Gifford, *University of Warwick, Coventry, United Kingdom*

**Keynote Address Speaker**
*Chairperson:* H. Jin, *University of California, Riverside*

**Frontier Technologies and Synthetic Biology**
*Chairperson:* J. Haseloff, *University of Cambridge, United Kingdom*

# Development and 3D Modeling of the Human Brain

**December 9–12**      200 Participants

ARRANGED BY      **Guo-li Ming,** University of Pennsylvania/Perelman School of Medicine, Philadelphia
**Sergiu Pasca,** Stanford University, California

The goal of this meeting was to bring together leaders in the field of human brain development, evolution, stem-cell biology, and neurological diseases to discuss progress in these fields and how they are informing 3D modeling of the human brain using human brain organoids. Participants convened with two Keynote talks, 33 oral presentations, and 42 posters. The presentations were intentionally diverse in the goals of the studies and the range of techniques used in this emerging field. The meeting started with a Keynote Address on the early development of preimplantation embryos using novel, engineered in vitro models based on self-organization and molecular mechanisms that define the positional identity of different types of cells within these engineered early embryo models. The first session covered



S. Pasca, G-l. Ming

recent advances in how we construct and deconstruct the human brain using 3D brain organoid models. The second session focused on the evolution aspect and how we can use organoids as a model to study brain evolution from great apes to humans. The second Keynote Address discussed how patient or disease-relevant organoids can be developed, scaled up, and applied in collaborative studies among different research groups. Additionally, presentations covered different types of brain-region-specific organoids and assembloids and how they can be applied to model different diseases in the nervous system. Brain tumor organoids, such as glioblastoma organoids, were also highlighted in several talks. Finally, transplantation of various human cell types and organoids into the rodent models was discussed.

This meeting was funded in part by STEMCELL Technologies and by System1 Biosciences.



Brain group

A. López Tobón, L. Gonzalez Cano



A. Berdenis van Berkelom, A. Krontira

## PROGRAM

**Keynote Address Speaker**
M. Zernicka-Goetz, *University of Cambridge, United Kingdom*

**Human Brain Development and Disease**
*Chairperson:* H. Bateup, *University of California, Berkeley*

**Human Brain Evolution**
*Chairperson:* N. Sestan, *Yale School of Medicine, New Haven, Connecticut*

**Keynote Address Speaker**
S. Temple, *Neural Stem Cell Institute, Rensselaer, New York*

**Brain Models, Development, and Disease I**
*Chairperson:* B. Treutlein, *Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*

**Brain Models, Development, and Disease II**
*Chairperson:* O. Reiner, *Weizmann Institute of Science, Rehovot, Israel*

**Brain Models and Disease**
*Chairperson:* J. Knoblich, *Institute of Molecular Biology, Vienna, Austria*



J. Knoblich, R. Jaenisch



P. Vanderhaeghen, F. Gage

# POSTGRADUATE COURSES

## High-Throughput Biology: From Sequence to Networks

**March 11–17**

INSTRUCTORS

**M. Brazas,** Ontario Institute for Cancer Research, Toronto, Canada
**F. Ouellette,** Bioinformatics, Montreal, Quebec, Canada

TEACHING ASSISTANTS

**K. Cotto,** Washington University in St. Louis, Missouri
**H. Gibling,** Ontario Institute for Cancer Research, Toronto, Canada
**P. Ronning,** McDonnell Genome Institute/Washington University in St. Louis, Missouri

With the introduction of next-generation sequencing platforms, it is now feasible to use high-throughput sequencing approaches to address many research questions. Now more than ever, it is crucial to know what bioinformatics tools and resources are available, and it is necessary to develop informatic skills to analyze high-throughput data using those tools. The Canadian Bioinformatics Workshops (CBW), in collaboration with Cold Spring Harbor Laboratory, has developed a comprehensive seven-day course covering key bioinformatics concepts and tools required to analyze DNA- and RNA-sequence reads using a reference genome. This course combined the materials and concepts from three established CBW workshops.

The course began with the workflow involved in moving from platform images to sequence generation, after which participants gained practical skills for evaluating sequence read quality, mapped reads to a reference genome, and analyzed sequence reads for variation and expression

level. The course concluded with pathway and network analysis on the resultant "gene" list. Participants gained experience in cloud computing and data visualization tools. All class exercises were self-contained units that included example data (e.g., Illumina paired-end data) and detailed instructions for installing all required bioinformatics tools.

This course was supported with funds provided by Helmsley Charitable Trust and the Howard Hughes Medical Institute and Computational Resources supported by Amazon Web Services.

## LECTURERS

**M. Bourgey,** McGill University, Montreal, Quebec, Canada

**S. Goodwin,** Cold Spring Harbor Laboratory

**M. Griffith,** McDonnell Genome Institute/Washington University in St. Louis, Missouri

**R. Haw,** Ontario Institute for Cancer Research, Toronto, Canada

**M. Hoffman,** Princess Margaret Cancer Centre/University of Toronto, Ontario, Canada

**Z. Lu,** University Health Network, Thornhill, Ontario, Canada

**C. Miller,** Washington University School of Medicine in St. Louis, Missouri

**Q. Morris,** University of Toronto, Ontario, Canada

**J. Reimand,** Ontario Institute for Cancer Research, Toronto, Canada

**J. Simpson,** Ontario Institute for Cancer Research, Toronto, Canada

**V. Voisin,** University of Toronto, Ontario, Canada

## PARTICIPANTS

Biesterveld, B., B.S., University of Michigan, Ann Arbor

Blair, J., B.A., Northwestern University, Chicago, Illinois

Cincotta, S., B.A./M.A., University of California, San Francisco

Ding, L., B.S., St. Jude Children's Research Hospital, Memphis, Tennessee

Edwards, D., B.A., Duke University, Durham, North Carolina

Gopinath, S., B.E., Icahn School of Medicine at Mount Sinai, New York

Hidalgo-Bravo, A., B.S., National Institute of Rehabilitation, Mexico City, Mexico

Lancaster, H., B.A., Arizona State University, Tempe

Li, T., M.S., Washington University in St. Louis, Missouri

Liu, Z., B.S., National Institutes of Health, Bethesda, Maryland

Lu, X., B.A., University of Notre Dame, Indiana

McNeil, M., B.Sc., University of Otago, Dunedin, New Zealand

Mitchell, R., B.S., USDA Boerne, Texas

Muir, J., B.Sc., McGill University, Montreal, Quebec, Canada

Paropkari, A., B.S., University of California, Merced

Porrett, P., B.A., University of Pennsylvania, Philadelphia

Prum, S., B.A., U.S. Naval Medical Research Unit No.2, Phnom Penh, Cambodia

Ramzan, F., B.Sc., The University of Auckland, New Zealand

Ruzicka, B., B.S., McLean Hospital/Harvard Medical School, Belmont, Massachusetts

Saha, K., B.S., M.S., University of California, San Diego, La Jolla

Singh, K., B.S., National Institutes of Health, Bethesda, Maryland

Singh, S., B.A., University of Virginia, Charlottesville

Thomas, K., B.S., St. Jude Children's Research Hospital, Memphis, Tennessee

Yang, S., B.S., The Jackson Laboratory, Bar Harbor, Maine

## SEMINARS

Bourgey, M., McGill University, Montreal, Quebec, Canada: Module 3: Genome alignment. Module 4: Small-variant calling and annotation. Module 5: Structural variant calling.

Cotto, K., Washington University in St. Louis, Missouri: Module 11: Isoform discovery and alternative expression.

Gibling, H., Ontario Institute for Cancer Research, Toronto, Canada: Module 2: Data visualization.

Goodwin, S., Cold Spring Harbor Laboratory: Module 1: Introduction to HT sequencing.

Griffith, M., Washington University School of Medicine in St. Louis, Missouri: Module 7: Introduction to RNA sequencing and analysis. Module 10: Reference free alignment.

Haw, R., Ontario Institute for Cancer Research, Toronto, Canada: Module 14: Network visualization and analysis with cytoscape and reactome. Module 14 Lab: De novo subnetwork clustering analysis: Reactome.

Hoffman, M., Princess Margaret Cancer Centre, Toronto, Ontario, Canada: Module 17: Gene regulation network analysis.

Lu, Z., University Health Network, Thornhill, Ontario, Canada: Module 3: Connecting to the cloud.

Miller, C., Washington University School of Medicine in St. Louis, Missouri: Module 8: RNA-Seq alignment and visualization. Module 9: Expression and differential expression.

Morris, Q., University of Toronto, Ontario, Canada: Module 16: Gene function prediction.

Reimand, J., Ontario Institute for Cancer Research, Toronto, Canada: Module 12: Introduction to pathway and network analysis. Module 13: Finding overrepresented pathways. Module 13 Lab: Enrichment-based analysis: performing ORA.

Simpson, J., Ontario Institute for Cancer Research, Toronto, Canada: Module 6: De novo assembly.

Voisin, V., University of Toronto, Ontario, Canada: Module 15: More depth on network and pathway + cytoscape/enrichment map. Module 15 Lab: Cytoscape/enrichment map. Module 16 Lab: geneMANIA. Module 17 Lab: iRegulon. Optional: Integrated assignment.

# Cryo-Electron Microscopy

**March 14–27**

INSTRUCTORS  **J. Kollman,** University of Washington, Seattle
       **G. Lander,** The Scripps Research Institute, La Jolla, California
       **M. Ohi,** University of Michigan, Ann Arbor

CO-INSTRUCTOR **M. Vos,** Thermo Fisher Scientific, Eindhoven, the Netherlands

ASSISTANTS   **A. Erwin,** University of Michigan, Ann Arbor
       **A. Hernandez,** The Scripps Research Institute, La Jolla, California
       **M. Johnson**, University of Washington, Seattle

Cryo-electron microscopy (cryo-EM) is a rapidly developing technique in structural biology wherein the biological sample of interest is flash frozen under cryogenic conditions. The utility of cryo-electron microscopy stems from the fact that it allows specimens to be observed under "near-to-native" conditions without the need for staining or fixation. This is in contrast to X-ray crystallography, which requires crystallizing the specimen, which can be a long and challenging process, which often involves the introduction of biomolecules into nonphysiological environments that can occasionally lead to functionally irrelevant conformational changes. Cryo-EM is now routinely applied to study the structures of viruses, ribosomes, ion channels, transcription and splicing machinery, and many other protein and nucleoprotein complexes. The spiraling number of publications that incorporate cryo-EM methodologies is evidence of this technique's importance

to the structural community: Since 2017, single-particle cryo-EM has been used to solve the structures of more than 1800 molecules, nearly half of which are resolved to >4 Å resolution. The resolution of single-particle cryo-EM maps is improving steadily, with recent improvements in processing methodologies yielding structures at >2 Å resolution. This powerful technique additionally enables researchers to study the conformational landscape of a biological specimen from a single flash-frozen sample, in order to deduce the mechanism by which it works.

The course covered the theory, practice, and application of single-particle cryo-EM. Participants in the course learned how to perform all steps involved in solving high-resolution cryo-EM structures, including sample prep, microscope alignment, data collection, image processing, and model building. Students had supervised access to CSHL's Titan Krios and K2 direct electron detector. This hands-on course included lectures by leading experts who discussed practical and conceptual approaches to structure determination using these techniques and covered a wide range of state-of-the-art applications of cryo-EM in the biological sciences.

## PARTICIPANTS

Benning, F., Ph.D., Massachusetts General Hospital, Boston
Eren, E., Ph.D., National Institutes of Health, Bethesda, Maryland
Hoelz, A., Ph.D., California Institute of Technology, Pasadena
Krishna Kumar, K., Ph.D., Stanford University, California

Meze, K., M.S., Cold Spring Harbor Laboratory
Takagi, Y., Ph.D., Indiana University School of Medicine, Indianapolis
Willcox, S., M.S., University of North Carolina, Chapel Hill
Wu, R., M.S., Purdue University, West Lafayette, Indiana
Zhang, C., Ph.D., University of Pittsburgh, Philadelphia

## SEMINARS

Adams, P., Lawrence Berkeley Laboratory, California: Model validation. Introduction to refinement in Phenix.
Bai, X., UT Southwestern Medical Center, Dallas: Cryo-EM image processing 3: RELION.
Carragher, B., New York Structural Biology Center, New York: Discussion: EM in industry vs. academia.
Cheng, A., New York Structural Biology Center, New York: Automated data acquisition with Leginon.
Cianfrocco, M., University of Michigan, Ann Arbor: Cloud computing for cryo-EM.
DiMaio, F., University of Washington, Seattle: Introduction to Rosetta.
Eng, E., New York Structural Biology Center, New York: How to run an efficient cryo-EM facility.
Erwin, A., University of Michigan, Ann Arbor: Intro to atomic modeling of cryo-EM densities in Coot.
Hernandez, A., The Scripps Research Institute, La Jolla, California: Intro to atomic modeling of cryo-EM densities in Coot.
Johnson, M., University of Washington, Seattle: Intro to atomic modeling of cryo-EM densities in Coot.
Kollman, J., University of Washington, Seattle: Make continuous carbon-negative stain grids.

Lander, G., The Scripps Research Institute, La Jolla, California: Course overview: Structure determination of biological macromolecules using cryo-EM.
Nogales, E., University of California, Berkeley: Single-particle cryo-EM: a historical perspective.
Ohi, M., University of Michigan, Ann Arbor: Why use electrons? Introduction to cryo-EM workflow. Introduction to sample prep for cryo-EM. Automated data acquisition: successes and failures.
Potter, C., New York Structural Biology Center, New York: Discussion: EM in industry vs. academia.
Rohou, A., Genentech, San Francisco, California: Discussion: EM in industry vs. academia.
Rosenthal, P., The Francis Crick Institute, London, United Kingdom: Cryo-EM validation. Future directions of cryo-EM.
Russo, C., MRC Laboratory of Molecular Biology, Cambridge, United Kingdom: Fundamentals of image formation. Detectors: history and breakthroughs.
Vos, M., Thermo Fisher Scientific, Eindhoven, the Netherlands: Detectors: history and breakthroughs.

# Workshop on Leadership in Bioscience

**March 22–25**

INSTRUCTORS  **C. Cohen,** Science Management Associates, Newton, Massachusetts
       **S. Cohen,** Science Management Associates, Newton, Massachusetts

In this highly interactive 3.5-day workshop, students developed the skills necessary to lead and interact effectively with others, in both one-on-one and group settings. Participants gained a solid experience-based foundation in managing others, negotiating win/win outcomes, running effective meetings, selecting the best team members, and setting goals with mentees, direct reports, and teams. The workshop focused on techniques, situations, and challenges that relate specifically to leading and managing in the scientific workplace. It emphasized learning by doing and involved role playing, giving and receiving feedback, and group problem solving. Much of the learning was peer-to-peer.

The workshop helped participants identify areas in which they need guidance and growth, as well as how to capitalize on areas of strength. Participants had the opportunity to share their experiences and challenges with others and to receive feedback and guidance from others with experience in leading scientists in a variety of settings. At the end of the course, participants linked through a unique online community in which they can continue learning from one another and from the course instructors.

Key focus areas of the workshop included:

- Recognizing and understanding leadership in a science setting.
- Using negotiation as a tool in scientific discussions and problem solving.
- Identifying and resolving conflicts in the laboratory.
- Dealing with difficult people and situations.
- Communicating your ideas and plans in a way that engages others.
- Leading effective and productive meetings.
- Becoming effective citizen scientists.
- Hiring and retaining a team.

The workshop was targeted to life scientists making, or recently having made, the transition to leadership or managerial positions. Many of the situations discussed were from the perspective of independent investigators running their own laboratories. As such, relatively new investigators (e.g., less than three years) are particularly encouraged to apply, as are senior postdoctoral scholars on the cusp of tenure-track research positions.

This course was supported with funds provided by the National Institute of General Medical Sciences.

## PARTICIPANTS

Aldana, B., B.Sc., University of Copenhagen, Denmark

Antes, A., B.S., Washington University School of Medicine in St. Louis, Missouri

Chambwe, N., B.S., Institute for Systems Biology, Seattle, Washington

Chan, L., B.S., Massachusetts Institute of Technology, Cambridge

de Hoon, M., M.Sc., IKEN, Yokohama-shi, Japan

Folmes, C., B.Sc., Mayo Clinic, Scottsdale, Arizona

Goldberg, E., B.S., Yale University, New Haven, Connecticut

Gomez, J., B.Sc., Vanderbilt University Medical Center, Nashville, Tennessee

Hanna, J., B.S., St. Jude Children's Research Hospital, Memphis, Tennessee

Hevener, K., B.S., University of Tennessee, Health Science Center, Memphis

Hilliard, T., B.S., University of Notre Dame, South Bend, Indiana

Hook, J., B.S., Columbia University Medical Center, New York

Jain, P., B.Tech., Children's Hospital of Philadelphia, Pennsylvania

Kurshan, P., B.S., Stanford University, Palo Alto, California

Lee, D., B.A., California Institute of Technology, Pasadena

Lewis, C., B.Sc., Whitehead Institute for Biomedical Research, Cambridge, Massachusetts

Memili, E., D.V.M., Mississippi State University, Mississippi State

Natoli, S., B.S., University of California, Berkeley

Okoye-Okafor, U., B.A., Albert Einstein College of Medicine, Bronx, New York

Rieder, L., B.S., Emory University, Atlanta, Georgia

Schindler, A., B.S., VA Puget Sound/University of Washington, Seattle

Sjulson, L., B.A., Albert Einstein College of Medicine, Bronx, New York

Steinmetz, N., B.S.E., University of Washington, Seattle

Vogel Ciernia, A., B.S., University of California, Davis

Wilkins, O., B.Sc., McGill University, Montreal, Quebec, Canada

## SEMINARS

Cohen C., Cohen S., Science Management Associates, Newton, Massachusetts: Session1: Who we are. Session 2: Leadership challenges: case study overview. Session 3: Difficult conversations and interactions: fundamentals of negotiation. Session 4: Running productive scientific team and project meetings. Session 5: Case study analysis, Part 1. Session 6: Hiring and retaining your science team; Interviewing, selecting, and orienting. Session 7: Projecting leadership. Session 8: Case study analysis, Part 2. Session 9: Managing your science team: mentoring, feedback, and more. Session 10: Concluding group discussion.

# Quantitative Imaging: From Acquisition to Analysis

| | |
|---|---|
| INSTRUCTORS | **H. Elliott,** Harvard Medical School, Boston, Massachusetts |
| | **T. Lambert,** Harvard Medical School, Boston, Massachusetts |
| | **J. Waters,** Harvard Medical School, Boston, Massachusetts |
| | |
| CO-INSTRUCTORS | **F. Jug,** Max Planck Institute CBG, Dresden, Germany |
| | **S. Manley,** Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland |
| | |
| TEACHING ASSISTANTS | **G.E. Campbell,** Harvard Medical School, Boston, Massachusetts |
| | **M. Cicconet,** Harvard Medical School, Boston, Massachusetts |
| | **S. Hirsch,** Columbia University, New York, New York |
| | **J. Hornick,** Northwestern University, Evanston, Illinois |
| | **A. Payne-Tobin Jost,** Harvard Medical School, Boston, Massachusetts |
| | **T.C. Rao,** University of Alabama, Birmingham |

Combining careful image acquisition with rigorous computational analysis allows extraction of quantitative data from light microscopy images that is far more informative and reproducible than what can be seen by eye. This course focused on advanced quantitative fluorescence microscopy techniques used for imaging a range of biological specimens, from tissues to cells to single molecules. The course was designed for quantitative cell and molecular biologists, biophysicists, and bioengineers.

We provided a thorough treatment of the complete process of quantitative imaging, from the photons emitted from the sample to the extraction of biologically meaningful measurements from digital images. Material was covered in lectures, discussion groups, and hands-on quantitative exercises using commercial microscopes and open-source image analysis tools.

Concepts covered included:

- Wide-field fluorescence microscopy.
- Laser scanning and spinning-disk confocal microscopy.
- CCD, EM-CCD, and sCMOS cameras.
- Total internal fluorescence microscopy (TIRF).
- Light sheet microscopy.
- Super-resolution microscopy (structured illumination, STED, and localization microscopy).
- Imaging and analyzing ratiometric "biosensors" (including FRET).
- Fluorescent proteins and live-sample imaging.
- Image processing (filtering, denoising, corrections, and deconvolution).
- Image segmentation.
- Quantitative shape and intensity measurements.
- Object detection and tracking.
- Machine learning.
- Designing and troubleshooting quantitative imaging experiments.

The course also included a series of seminars from guest speakers who applied the methods we discussed.

## PARTICIPANTS

Boot, M., M.S., Ph.D., Yale School of Medicine, New Haven, Connecticut
Chmiel, T.A., Ph.D., University of Chicago, Illinois
Deng, L., Ph.D., University of Washington, Seattle
Eaton, D.S., Ph.D., Harvard Medical School, Boston, Massachusetts
Haley, J.S., Ph.D., University of California, San Diego, La Jolla
Harnagel, A.E., The Rockefeller University, New York, New York
Hobson, C.M., Ph.D., University of North Carolina, Chapel Hill
Jamali, N., Ph.D., Mount Sinai School of Medicine, New York
Molina, R.S., Ph.D., Montana State University, Bozeman

Ojeda Naharros, I., Ph.D., University of California, San Francisco
Otopalik, A.G., Ph.D., Columbia University, New York, New York
Peloggia de Castro, J., Ph.D., Stowers Institute for Medical Research, Kansas City, Missouri
Smith, M.J., Ph.D., New York University Langone Medical Center, New York
Stephens, A.D., Ph.D., Northwestern University, Evanston, Illinois
Sundararajan, K., Ph.D., Stanford School of Medicine, California
Weiss, B.G., M.D., Ph.D., UT Southwestern Medical Center, Dallas

## SEMINARS

Canman, J.C., Columbia University, New York, New York: Cytokinetic diversity and the mechanisms that promote robust cell division.
Elliott, H., Harvard Medical School, Boston, Massachusetts: Digital imaging: cameras. Basics of image processing and digital microscopy: resolution, SNR, and diffraction-limited objects. Image processing 2: image corrections and advanced filtering. Image segmentation and morphometry. Image correlation methods: co-localization, registration, and stitching. Machine learning in bioimage analysis.

3D image analysis and deconvolution. Image time series analysis: tracking, photo-bleach correction, and FRAP analysis.

Jug, F., Max Planck Institute CBG, Dresden, Germany: Applied machine learning. Content-aware image restoration.

Lambert, T., Harvard Medical School, Boston, Massachusetts: Confocal microscopy theory and hardware. Light sheet microscopy: Part 2. Super-resolution microscopy II: patterned illumination.

Lavis, L.D., Janelia Farm Research Campus, Ashburn, Virginia: Designing brighter fluorophores for advanced live-cell imaging and beyond.

Manley, S., Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland: Super-resolution. Expanding horizons with large field-of-view, automated localization microscopy.

Payne-Tobin Jost, A., Harvard Medical School, Boston, Massachusetts: Fluorescence microscopy.

Shaner, N., The Scintillon Institute, San Diego, California: Fluorescent proteins.

Shroff, H., National Institutes of Health, Bethesda, Maryland: Biological imaging at high spatiotemporal resolution.

Waters, J., Harvard Medical School, Boston, Massachusetts: Introduction to quantitative microscopy. Objective lenses. Koehler illumination, and image formation. Phase, darkfield, and DIC microscopy. Quantifying fluorescence: image acquisition and controls. Live cell imaging. Total internal reflection fluorescence (TIRF) microscopy. Live confocal microscopy and intensity measurements over time. Multi-photon microscopy. Limitations on quantitative imaging of thick samples. Light sheet microscopy: Part 1.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics and rigor in the biosciences.

# Expression, Purification, and Analysis of Proteins and Protein Complexes

**April 3–12**

INSTRUCTORS     **A. Courey,** University of California, Los Angeles
                **M. Marr,** Brandeis University, Waltham, Massachusetts
                **S. Nechaev,** University of North Dakota School of Medicine, Grand Forks


ASSISTANTS      **N. Clark,** Brandeis University, Waltham, Massachusetts
                **S. Gartland,** Brandeis University, Waltham, Massachusetts
                **M. Harris,** Brandeis University, Waltham, Massachusetts
                **M. Jamal,** University of California, Los Angeles
                **D. Parrello,** University of North Dakota School of Medicine, Grand Forks
                **J. Rigal,** Brandeis University, Waltham, Massachusetts
                **T. Yau,** University of California, Los Angeles

This course was for scientists, including graduate students, postdoctoral scholars, staff scientists, and principal investigators, who wanted a rigorous introduction to expression and purification of proteins, as well as analysis of protein structure and function.

Through hands-on experience in the lab as well as extensive lecture and discussion, each student became familiar with key approaches in expression, purification, and analysis of soluble and membrane proteins and protein complexes from both natural sources and overexpression systems. The emphasis of the course was on the following:

1. Approaches in protein expression: Choosing the best bacterial or eukaryotic expression system tailored for the particular protein and experimental problem; determining how to optimize expression; and understanding protein tagging: the advantages and pitfalls of various affinity and solubility tags.

2. Approaches in protein purification: Choosing the best strategy for a given protein including solubilization; bulk fractionation; liquid chromatography, including conventional methods (ion exchange, size exclusion, reverse phase, etc.) and affinity methods (e.g., MAC, DNA affinity, immunoaffinity, etc.), as well as FPLC/HPLC.

3. Approaches in protein analysis: Introduction to common approaches for characterization of proteins including binding assays; activity assays; mass spectroscopy to identify protein interaction partners; and posttranslational modifications.

In addition to purification, students also gained exposure to fundamental analytical approaches such as mass spectroscopy and protein structure determination (e.g., X-ray crystallography, cryo-EM, etc.).

## PARTICIPANTS

Castro, A., B.Sc., University of Puerto Rico Rio Piedras, San Juan

Kmezik, C., M.S., Chalmers University of Technology, Goteborg, Sweden

Liang, H., Ph.D., Texas Tech University Health Sciences Center, Lubbock

Ouyang, M., Ph.D., Cold Spring Harbor Laboratory

Park, J., B.S., University of Maryland, Baltimore

Rodriguez Rios, J., B.S., University of Puerto Rico Rio Piedras, San Juan

Rojo, R., Ph.D., MD Anderson Cancer Center, Houston, Texas

Sowaileh, M., Ph.D., St. Jude Children's Research Hospital, Memphis, Tennessee

Termini, C., Ph.D., University of California, Los Angeles

Voth, S., B.S., University of South Alabama College of Medicine, Mobile

## SEMINARS

Courey, A., University of California, Los Angeles: System-wide analyses of Groucho and SUMO in *Drosophila*.

Jarvis, J., University of Wyoming, Laramie: Recombinant protein production in the baculovirus–insect cell system.

Love, J., Expression Technologies, Newark, California: High-throughput purification of membrane proteins.

Marr, M., Brandeis University, Waltham, Massachusetts: Introduction to protein purification. Controlling gene expression in response to stress.

Nechaev, S., University of North Dakota School of Medicine, Grand Forks: Stable Pol II pausing is retained during gene activation to provide a platform for regulation.

Pappin, D., Cold Spring Harbor Laboratory: Introduction to mass spectrometry of proteins. Quantitative approaches to mass spectrometry of proteins.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, rigor, and reproducibility.

# Cell and Developmental Biology of *Xenopus*

INSTRUCTORS     **C. Chang,** University of Alabama, Birmingham
    **L. Davidson,** University of Pittsburgh, Pennsylvania

CO-INSTRUCTORS     **M. Khokha,** Yale University, New Haven, Connecticut
    **K. Liu,** King's College London, United Kingdom

ASSISTANTS     **C. Exner,** University of California, San Francisco
    **R. Huebner,** The University of Texas, Austin
    **M. Lane,** Yale University, New Haven, Connecticut
    **E. Mis,** Yale University, New Haven, Connecticut
    **H. Rankin Willsey,** University of California, San Francisco
    **R. Stephenson,** University of Michigan, Ann Arbor

In vivo animal models are an important tool for the understanding of human development and disease. Studies using the frog *Xenopus* have made remarkable contributions to our understanding of fundamental processes such as cell cycle regulation, transcription, translation, and many other topics. *Xenopus* are remarkable for studying development and disease, including birth defects, cancer, and stem cell biology. Because *Xenopus* are easy to raise, producing many thousands of eggs per day, these frogs have emerged as a premiere model for the understanding of human biology from the fundamental building blocks to the whole organism.

The recent development of CRISPR-Cas9 technology has made it easy to target genes of interest using *Xenopus*. This course was designed with that in mind. Our goal was for students to design a set of experiments focusing on their genes or biological interests. Prior to starting the course, students

were expected to choose gene(s) of interest, and the instructors generated sgRNAs targeting these genes. These were either the students' own genes or chosen from a bank provided by the instructors.

During the course, the students analyzed phenotypes generated from CRISPR-Cas9-based gene depletion and learned the diverse array of techniques available in *Xenopus*. In previous courses, we have guided students in the ablation of a wide variety of genes and helped them design suitable assays for their biological interests. Most recently, students have targeted autism genes, thyroid genes, and immune modulators, several of which have already led to publications. Approaches covered included microinjection and molecular manipulations such as CRISPR-Cas9 knockouts, antisense morpholino-based depletions, transgenics, and mRNA overexpression. In addition, students combined these techniques with explant and transplant methods to simplify or test tissue-level interactions. Additional methods included mRNA in situ hybridization and protein immunohistochemistry as well as basic bioinformatic techniques for gene comparison and functional analysis. Biochemical approaches such as proteomics and mass spectrometry and biomechanical concepts were discussed.

Finally, to visualize subcellular and intercellular activities, we introduced a variety of sample preparation and imaging methods including time-lapse, fluorescent imaging, optical coherence tomography, and confocal microscopy. These were facilitated by state-of-the-art equipment from Nikon, Leica, Thorlabs, and Bruker.

## PARTICIPANTS

Angermeier, A., B.S., University of Alabama, Birmingham

Bayarri Olmos, R., M.S., University of Copenhagen, Denmark

Cadart, C., Ph.D., University of California, Berkeley

Cervino, A., M.S., IFIBYNE (UBA–CONICET), Buenos Aires, Argentina

Chae, S., B.S., Ulsan National Institute of Science & Technology, South Korea

Colwell, M., B.S., University of Minnesota, St. Paul

Guerin, D., B.S., University of Nevada, Las Vegas

Klein, H., M.S., The Hebrew University, Jerusalem, Israel

Lynch, D., M.S., King's College London, United Kingdom

Nanos, V., M.S., Tufts University, Boston, Massachusetts

Parasyraki, E., M.Sc., Institute of Molecular Biology, Mainz, Germany

Phong, C., Ph.D., Stanford School of Medicine, California

Pillai, E., M.Sc., University of Cambridge, United Kingdom

Royle, S., B.A., Harvard University, Cambridge, Massachusetts

Truchado Garcia, M., Ph.D., University of California, Berkeley

## SEMINARS

Cha, S-W., University of Central Missouri, Warrensburg: How to make a long gut tube.

Chang, C., University of Alabama, Birmingham: Embryonic induction and signaling, a walk of a century following Spemann.

Conlon, F., University of North Carolina, Chapel Hill: Proteomic approaches to *Xenopus* biology.

Davidson, L., University of Pittsburgh, Pennsylvania: Leveraging *Xenopus* mechanics and morphogenesis.

Heald, R., University of California, Berkeley: Experiments you can only do with frogs.

Keller, R., University of Virginia, Charlottesville: Early morphogenesis of *Xenopus*.

Kelley, D., Columbia University, New York, New York: *Xenopus* in space and time.

Khokha, M., Yale University, New Haven, Connecticut: Patient-driven gene discovery: oxygen, mitochondria, and *Xenopus* power.

Mayor, R., University College London, United Kingdom Mechanisms of neural crest migration.

Nascone-Yoder, N., North Carolina State University, Cary: Gut coiling morphogenesis.

Wills, A., University of Washington, Seattle: Interrogating mechanisms of regeneration in *Xenopus*.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, rigor, and reproducibility.

Woolner, S., University of Manchester, United Kingdom: Using *Xenopus* to investigate how mechanical force regulates cell division.

# Workshop on Pancreatic Cancer

INSTRUCTORS    **H. Crawford,** University of Michigan, Ann Arbor
               **T. Hollingsworth,** University of Nebraska Medical Center, Omaha
               **M. Pasca di Magliano,** University of Michigan, Ann Arbor

ASSISTANTS     **G. Biffi,** Cold Spring Harbor Laboratory
               **C. Halbrook,** University of Michigan, Ann Arbor

Pancreatic cancer is one of the deadliest cancers: Tumors are often diagnosed at advanced stages of the disease and metastasize rapidly. This one-week discussion course provided a comprehensive overview of clinical and biological aspects of pancreatic cancer with special emphasis on disease diagnosis and management, molecular pathways involved in tumor development and progression, mechanism-based therapeutic strategies, advanced research tools, and ethical concerns. Attendees were able to interact with senior investigators on a one-to-one basis in an informal environment.

Topics included:

- Organ overview: Anatomy and physiology.
- Clinical aspects of pancreatic cancer: Diagnosis and treatment.
- Molecular genetics of pancreatic cancer: Gene signatures and predisposition syndromes.
- Pathobiology of pancreatic cancer: Pathways, cell of origin, and tumor microenvironment.
- Tools and techniques: Mouse models, imaging, genomics, proteomics, metabolomics, and bioinformatics.
- Therapeutics: Target identification and validation, preclinical studies, and clinical trial design.
- Resources: Biobanks and funding strategies.

PARTICIPANTS

Arner, E., B.A., UT Southwestern Medical Center, Dallas

Chougoni, K.K., M.S., Virginia Commonwealth University, Richmond

Cornwell, A., B.S., Roswell Park Graduate Division at University of Buffalo, New York

Decker, A., M.S., Columbia University Irving Medical Center, New York, New York

Ding, L., Ph.D., Mayo Clinic, Rochester, Minnesota

Dixit, A., Ph.D., University of Minnesota, Minneapolis

Ferguson, L., Ph.D., University of California, San Diego, La Jolla

Galenkamp, K., Ph.D., Sanford Burnham Presbyterian Medical Discovery Institute, La Jolla, California

Hanna, S., Ph.D., Weill Cornell Medicine, New York, New York

Kim, P., M.S., University of Michigan Rogel Cancer Center, Ann Arbor

Kim, S.E., Ph.D., University of California, San Francisco

Kudelka, M., M.D./Ph.D., Emory University, Atlanta, Georgia

Lu, L., Ph.D., Ohio State University, Columbus

Matiash, K., B.S., University of Cincinnati, Ohio

Menjivar, R., B.S., University of Michigan, Ann Arbor

Nadella, S., M.D., Georgetown University, Washington, D.C.

Pal Choudhuri, S., Ph.D., UT Southwestern Medical Center, Dallas

Pita Grisanti, V., M.S., Ohio State University, Columbus

Raoof, M., M.S., City of Hope Cancer Center, Duarte, California

Robert, M., M.D., Yale University School of Medicine, New Haven, Connecticut

Seilstad, C., B.S., University of Nebraska Medical Center, Omaha

Somerville, T., Ph.D., Cold Spring Harbor Laboratory

Zarmer, S., B.S., University of North Carolina, Chapel Hill

SEMINARS

Aguirre, A., Dana-Farber Cancer Institute, Boston, Massachusetts: Traditional and emerging treatments for pancreatic cancer.

Al-Hawary, M., University of Michigan, Ann Arbor: Radiology in the detection of pancreatic cancer.

Bar-Sagi, D., New York University Langone Medical Center, New York: Ras and PDA.

Beatty, G., University of Pennsylvania/Perelman School of Medicine, Philadelphia: Inflammation and pancreatic cancer.

Crawford, H., University of Michigan, Ann Arbor: Cell plasticity in transformation and early neoplasia.

Cruz-Monserrate, Z., Ohio State University, Dublin: Environmental influences on pancreatic cancer.

Cukierman, E., Fox Chase Cancer Center/Temple Health, Philadelphia, Pennsylvania: TME: fibroblasts extracellular matrix.

DeNardo, D., Washington University in St. Louis, Missouri: Immunotherapy in pancreatic cancer.

Der, C., University of North Carolina, Chapel Hill: Targeting Ras.

Herman, J., University of Texas MD Anderson Cancer Center, Houston: The role of radiation in pancreas cancer.

Hollingsworth, M.A., University of Nebraska Medical Center, Omaha: Early detection of pancreatic cancer.

Kelly, K., University of Virginia, Charlottesville; VanBrocklin, H., University of California, San Francisco: Molecular imaging preclinically.

Lyssiotis, C., University of Michigan, Ann Arbor: Pancreatic cancer metabolism.

Maitra, A., University of Texas MD Anderson Cancer Center, Houston: Histological and molecular precursor neoplasms, pancreatitis, PDA, genetic progression series.

McDonald, O., Vanderbilt University, Nashville, Tennessee: Epigenetics of pancreatic cancer.

Olive, K., Columbia University, New York, New York: PDA models: mice.

O'Reilly, E., Memorial Sloan Kettering Cancer Center, New York, New York: Clinical trials in pancreas cancer.

Pasca di Magliano, M., University of Michigan, Ann Arbor: Inflammation in PDA evolution.

Perera, R., University of California, San Francisco: Autophagy in pancreatic cancer.

Petersen, G., Mayo Clinic, Rochester, Minnesota: Risk factors for pancreatic cancer.

Rhim, A., University of Pennsylvania, Philadelphia: Gastroenterology and pancreas cancer.

Saif, W., Northwell Health, Lake Success, New York: Ways for scientists to collaborate with physicians to conduct clinical trials for pancreatic cancer patients.

Sherman, M., Oregon Health & Science University, Portland: Metabolic and gene-regulatory functions of the pancreatic tumor microenvironment.

Simeone, D., New York University Langone Medical Center, New York: Surgery for pancreatic cancer.

Stanger, B., University of Pennsylvania, Philadelphia: Cell plasticity II: subtypes and metastasis.

Wood, L., Johns Hopkins School of Medicine, Baltimore, Maryland: Pancreatic cancer genomics.

OPEN SYMPOSIUM

Aguirre, A., Dana-Farber Cancer Institute, Boston, Massachusetts

Bar-Sagi, D., New York University Langone Medical Center, New York

Crawford, H., University of Michigan, Ann Arbor

Cruz-Monserrate, Z., Ohio State University, Dublin

Cukierman, E., Fox Chase Cancer Center/Temple Health, Philadelphia, Pennsylvania

DeNardo, D., Washington University in St. Louis, Missouri

Der, C., University of North Carolina, Chapel Hill

Fearon, D., Cold Spring Harbor Laboratory

Ferguson, L., University of California, San Diego, La Jolla

Hollingsworth, M.A., University of Nebraska Medical Center, Omaha

Kelly, K., University of Virginia, Charlottesville

McDonald, O., Vanderbilt University, Nashville, Tennessee

Olive, K., Columbia University, New York, New York

Pasca di Magliano, M., University of Michigan, Ann Arbor

Perera, R., University of California, San Francisco

Rhim, A., University of Pennsylvania, Philadelphia

Rustgi, A., Columbia University, New York, New York

Sherman, M., Oregon Health & Science University, Portland

Stanger, B., University of Pennsylvania, Philadelphia

Tuveson, D., Cold Spring Harbor Laboratory

# Advanced Bacterial Genetics

**June 4–24**

INSTRUCTORS    **L. Bossi,** Institute of Integrative Biology of the Cell (I2BC), Paris, France
    **A. Camilli,** Tufts University Medical School, Boston, Massachusetts
    **A. Grundling,** Imperial College London, United Kingdom

ASSISTANTS    **R. Balbontin Soria,** Instituto Gulbenkian de Ciência, Oeiras, Portugal
    **J. Bourgeois,** Tufts University School of Medicine, Boston, Massachusetts
    **N. Figueroa-Bossi,** Institute of Integrative Biology of the Cell (I2BC), Paris, France
    **M. Zeden,** Imperial College London, United Kingdom

This course presented logic and methods used in the genetic dissection of complex biological processes in diverse bacteria.

Laboratory methods included:

- Classical and cutting-edge mutagenesis using transposons, allelic exchange, and TargeTron.
- Recombineering with single- and double-stranded DNA.
- CRISPR-Cas genome editing, genome sequencing, and assembly.
- Mapping mutations using genetic and physical techniques.
- Modern approaches to the generation and analysis of targeted gene disruptions and reporter gene fusions.
- Fluorescence microscopy.

Key components of the course were the use of sophisticated genetic methods in the analysis of model bacteria (including *Escherichia coli, Salmonella, Staphylococcus aureus,* and *Vibrio cholerae*),

and the use of the wealth of new genomic sequence information to motivate these methods. Invited lecturers presented various genetic approaches to study bacterial mechanisms of metabolism, development, and pathogenesis.

## PARTICIPANTS

Bedree, J., B.S., University of California, Los Angeles

Bhave, D., M.S., Max Planck Institute for Evolutionary Biology, Germany

Creamer, K., B.A., University of California, San Diego, La Jolla

Dempsey, J., B.S., University of Washington, Seattle

Gao, B., Ph.D., South China Institute of Oceanology, Guangzhou, China

Getz, L., B.Sc., Dalhousie University, Halifax, Nova Scotia, Canada

Graniczkowska, K., B.S., University of Kentucky, Lexington

Fehler, A., MSc, University of Copenhagen, Denmark

Guardiola Flores, K., B.S., University of Michigan, Ann Arbor

Hennessey-Wesen, M., B.A., Institute of Science and Technology, Klosterneuberg, Austria

Kim, N., B.A., Stony Brook School of Medicine, Stony Brook, New York

Lopez, J., B.A., Washington University in St. Louis, Missouri

Riley, K., Ph.D., Rollins College, Winter Park, Florida

Siegel, S., Ph.D., UT Southwestern, Dallas

Silva, A., M.S., Universidade Nova de Lisboa, Portugal

Thomas, A., B.S., University of Chicago, Illinois

## SEMINARS

Dalia, A., Indiana University, Bloomington: Molecular dissection of natural transformation and exploiting it as a genetic tool to study *Vibrio* species.

Gross, C., University of California, San Francisco: Systems level analyses in bacteria.

Kearns, D., Indiana University, Bloomington: A chalk talk about *Bacillus* genetics.

Lesser, C., Massachusetts General Hospital/Harvard Medical School, Boston: Leveraging bacterial secretion systems to develop therapeutic designer probiotics.

Maloy, S., San Diego State University, California: How do new pathogens evolve?

Ng, D.K-L., Tufts University, Boston, Massachusetts: Chemical signaling pathways in *Vibrio cholerae*.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, rigor, and reproducibility.

# Ion Channels in Synaptic and Neural Circuit Physiology

**June 4–24**

INSTRUCTORS
**C. Schmidt-Hieber,** Pasteur Institute, Paris, France
**A. Scimemi,** University at Albany, New York
**N. Wanaverbecq,** Aix Marseille University, Marseille, France

CO-INSTRUCTORS
**J. Grundemann,** University of Basel, Switzerland
**A. Lampert,** RWTH Aachen University, Germany

ASSISTANTS
**R. Bott,** RWTH Aachen University, Germany
**R. Gomez-Ocadiz,** Institute Pasteur, Paris, France
**J. McCauley,** University at Albany, New York
**M. Petroccione,** University at Albany, New York
**J. Trachtenberg,** University of California, Los Angeles

Ion channels are the fundamental building blocks of excitability in the nervous system. The primary goal of this course was to demonstrate, through lectures and laboratory work, the different biophysical properties of ion channels that enable neurons to perform unique physiological functions in a variety of neural systems.

Areas of particular interest included (1) voltage- and ligand-gated ion channels at central and peripheral synapses, (2) synaptic integration and plasticity, (3) neural circuit function in vitro and in vivo, and (4) optogenetic strategies for circuit manipulation. A typical day consisted of morning lectures followed by hands-on laboratory practical sessions in the afternoon and evening with guest lecturers available to give one-on-one practical advice.

The laboratory component of the course introduced students to state-of-the-art electro-physiological approaches for the study of ion channels in their native environments. The course provided students with hands-on experience in using patch-clamp electrophysiology to examine single-channel activity in cultured cells, ion channel biophysics in acutely dissociated neurons and synaptic integration, plasticity and circuit dynamics in in vitro slice and in vivo preparations. Different recording configurations were used (e.g., cell-attached, whole-cell dendritic and somatic patch and voltage- and current-clamp configurations), and the advantages and limitations of each method were discussed in relation to specific scientific questions. The course also provided practical experience in cellular and circuit manipulation techniques (i.e., pharmacological, electrophysiological, and optogenetic) both in vitro and in vivo.

## PARTICIPANTS

Bryant, S., Ph.D., Central Michigan University, Mount Pleasant

Jones, S., M.Sci., University College London, United Kingdom

Kang, S.K., M.S., Northwestern University, Chicago, Illinois

Kast, R., Ph.D., Massachusetts Institute of Technology, Cambridge

Keles, M., Ph.D., Johns Hopkins University, Baltimore, Maryland

Ko, M., B.S., Purdue University, West Lafayette, Indiana

Pavon Arocas, O., M.Sc., University College London, United Kingdom

Santiago, C., Ph.D., Harvard Medical School, Boston, Massachusetts

Sempou, E., Ph.D., Yale University, New Haven, Connecticut

Voufo, C., B.S., University of California, Berkeley

Wang, Y., B.S., The Scripps Research Institute, La Jolla, California

Xu, R., B.S., Harvard University, Cambridge, Massachusetts

## SEMINARS

Bean, B., Harvard Medical School, Boston, Massachusetts: Control of intrinsic excitability.

Beeton, C., Baylor College of Medicine, Houston, Texas: Potassium channel: phenotype and function in health and disease.

Bolton, M., Max Planck Florida Institute for Neuroscience, Jupiter: ChR2 biophysics.

Branco, T., University College London, United Kingdom: Synaptic integration in single neurons.

Cohen, J., The Johns Hopkins University, Baltimore, Maryland: In vivo neurophysiology.

Diamond, J., National Institutes of Health, Bethesda, Maryland: Diverse dendritic signaling in the retina.

Dudman, J., Howard Hughes Medical Institute, Ashburn, Virginia: Optogenetic tools and approaches for circuit dissection.

Duguid, I., University of Edinburgh, United Kingdom: Thalamocortical control of goal-directed motor behavior.

Gasparani, S., Louisiana State University, New Orleans: HCN channels.

Kammermeier, P., University of Rochester Medical Center, New York: Using channel modulation to examine mGluR function in neurons.

Laezza, F., The University of Texas Medical Branch, Galveston: Voltage-gated $Na^+$ channels: physiology and pharmacology.

Overstreet-Waldiche, L., University of Alabama, Birmingham: Synaptic inhibition.

Plested, A., Humboldt University, Berlin, Germany: Single channels and glutamate receptors.

Sjostrom, J., McGill University, Montreal, Quebec, Canada: Unconventional NMDA receptor signaling at central synapses.

Spruston, N., Howard Hughes Medical Institute, Ashburn, Virginia: Deconstructing memory circuits at cell type-specific resolution. A crash course in neuronal excitability: how it works and why it (still) matters.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, rigor, and reproducibility.

Xu-Friedman, M., University of Buffalo, New York: Regulation of auditory nerve synaptic function by activity.

# Mouse Development, Stem Cells, and Cancer

**June 5–24**

INSTRUCTORS    **T. Caspary,** Emory University School of Medicine, Atlanta, Georgia
**C. Forsberg,** University of California, Santa Cruz

CO-INSTRUCTORS    **D. Laird,** University of San Francisco, California
**F. Mariani,** University of South California School of Medicine, Altadena

ASSISTANTS    **S. Cincotta,** University of San Francisco, California
**T. Cool,** University of California, Santa Cruz
**E. Gigante,** Emory University School of Medicine, Atlanta, Georgia
**R. Jaszczak,** University of San Francisco, California
**K. Piotrowska Nitsche,** Emory University School of Medicine. Atlanta, Georgia
**M. Serowoky,** University of South California School of Medicine, Altadena
**S. Smith-Berdan,** University of California, San Francisco
**B. Sozen Kaya,** University of Cambridge, United Kingdom

This intensive lecture and laboratory course was designed for scientists interested in using mouse models to study mammalian development, stem cells, and cancer. The lecture portion of the course, taught by leaders in the field, provided the conceptual basis for contemporary research in embryogenesis; organogenesis in development and disease; embryonic, adult, and induced pluripotent stem cells; and cancer biology.

The laboratory and workshop portions of the course provided hands-on introduction to engineering of mouse models, stem-cell technologies, and tissue analyses. Experimental techniques included genome editing by CRISPR-Cas9, zygote microinjection, isolation and culture/manipulation of pre- and postimplantation embryos, embryo transfer, embryo electroporation,

roller bottle culture, sperm cryopreservation, in vitro fertilization, culture of mouse embryonic stem cells and fibroblasts, synthetic embryo generation, vibratome and cryosectioning, in situ RNA hybridization, immunostaining, FACS sorting, analysis of hematopoietic stem cells, skeletal preparation, organ explant culture and fluorescent imaging, and live time-lapse microscopy.

## PARTICIPANTS

Alam, S., M.S., McGill University, Montreal, Quebec, Canada

Bondarenko, V., M.S., EMBL, Heidelberg, Germany

Bush, M., B.A., St. Jude Graduate School of Biomedical Sciences, Memphis, Tennessee

Davenport, M., B.S., University of Alabama, Birmingham

Frost, E., B.Sc., University of Newcastle, Callaghan, New South Wales, Australia

Kurlovich, J., M.S., University of Göttingen, Germany

Lopez, A., Ph.D., Baylor College of Medicine, Houston, Texas

Madhavan, M., B.Tech., Michigan State University, East Lansing

Mathew, S., Ph.D., National Centre for Biological Sciences, Bangalore, India

Murugapoopathy, V., Ms.C., McGill University, Montreal, Quebec, Canada

Olbrich, T., M.D., National Cancer Institute, Bethesda, Maryland

Shadle, S., Ph.D., Howard Hughes Medical Institute, Salt Lake City, Utah

Sharma, T., Ph.D., UT Southwestern, Children's Research Institute, Dallas

Sundaramurthy, V., B.Tech., Baylor College of Medicine, Houston, Texas

## SEMINARS

Caspary, T., Emory University School of Medicine. Atlanta, Georgia: What the mouse mutants teach us about neural patterning.

Caspary, T., Emory University School of Medicine, Atlanta, Georgia; Forsberg, C., University of California, Santa Cruz: Overview of mouse development.

Cooper, K., University of California, San Diego, La Jolla: Genetic complexity in vertebrate limb development and evolution.

Dive, C., Cancer Research UK Manchester Institute, Macclesfield, United Kingdom: Patient circulating tumor cell–derived mouse models that allow study of small cell lung cancer from diagnosis to disease progression.

Forsberg, C., University of California, Santa Cruz: Blood: it does a body good.

Justice, M., Hospital for Sick Children, Toronto, Ontario, Canada: Mighty mouse: the history, application, and future of forward genetic mutagenesis for modeling human disease.

Laird, D., University of San Francisco, California: Lineage and fate in the germ line.

Lewandoski, M., National Institutes of Health, Frederick, Maryland: Mouse genetic technologies.

Mager, J., University of Massachusetts, Amherst: Microinjection tools and techniques: ever evolving tricks of the trade.

Mariani, F., University of Southern California School of Medicine, Altadena: Skeletal stem cells: from bench to bedside.

McNeill, H., Washington University in St. Louis, Missouri: Fat cadherins in mammalian kidney development.

Reeves, M., University of California, San Francisco: Mouse models of cancer and tumor.

Rodriguez, T., Imperial College London, United Kingdom: Cell competition: a two-edged sword that can shape embryonic development or promote tumor expansion.

Sander, M., University of California, San Diego, La Jolla: Tracking cells in the pancreas: insights into cell plasticity and regeneration.

Shen, M., Columbia University, New York, New York: Prostate development, stem cells, and cancer.

Soriano, P., Icahn School of Medicine at Mount Sinai, New York: FGF signaling pathways in craniofacial development.

Tam, P., University of Sydney School of Medicine, New South Wales, Australia: Building a body plan: lineage allocation and embryonic patterning.

Trainor, P., Stowers Institute for Medical Research, Kansas City, Missouri: Neural crest cells and their fundamental roles in development, evolution, and disease.

Vokes, S., The University of Texas, Austin: Transcriptional interpretation of Hedgehog signaling.

Yamanaka, Y., McGill University, Montreal, Quebec, Canada: Epithelial morphogenesis in preimplantation development and ovarian cancer.

Zernicka-Goetz, M., University of Cambridge, United Kingdom: Building the mouse and human embryo in vivo and in vitro.

# Metabolomics

INSTRUCTORS    **A. Caudy,** University of Toronto, Ontario, Canada
**J. Cross,** Memorial Sloan Kettering Cancer Center, New York, New York
**A. Rosebrock,** Stony Brook University, Stony Brook, New York

ASSISTANTS    **I. Abramovich,** Technion Institute of Technology, Haifa, Israel
**D. Sumpton,** Cancer Research Beatson Institute, Glasgow, United Kingdom

Advances in genomics, transcriptomics, and proteomics have enabled both broad and deep analysis of genomes and their encoded proteins. Metabolomics focused on measuring the biochemical contents of cells, tissues, and organisms. Biochemical phenotypes represent a unique view into the dynamic state of biological systems and are relevant to a range of fields, from model organisms to patients, from bioprocess to bedside.

This course combined theoretical and practical training including hands-on experience with a range of cutting-edge approaches to interrogate biochemical state. Mass spectrometry (MS) is currently the most powerful and flexible approach in the metabolomics toolbox. Students became proficient in the generation and analysis of both gas-chromatography (GC) and liquid-chromatography (LC) mass spectrometry data. New biochemistry awaited discovery, even in well characterized systems. Participants learned how to quantitate known metabolites in complex biological samples and discovered and characterized unknown compounds using LC- and GC-mass spectrometry. Metabolomics made use of many complementary tools in addition to mass spectrometry.

Students gained hands-on experience in measuring metabolic state in live cells using the Agilent Seahorse XF platform; determining reaction kinetics for purified enzymes using in vitro assays; and engaging in friendly competition to develop practical chromatographic separations.

The course integrated practical lab sessions, hands-on data analysis, and lecture-based learning. Students had the opportunity to interact with instructors and TAs as well as a diverse panel of field-leading guest speakers who presented both formal talks and a nuts-and-bolts view of metabolomics in their labs.

The objectives for students were to (1) become proficient in quantitative and qualitative analysis of GC- and LC-MS data using currently available vendor and open-source tools; (2) understand the use cases and limitations of currently available metabolomics instrumentation and be able to identify the right approach for a given question; (3) learn key factors in experimental design and sample preparation that enabled collection of interpretable and actionable metabolomics data; and (4) gain the core knowledge and vocabulary required to fruitfully interact with other researchers in the metabolomics field.

Past laboratory exercises have included:

- Development of mass spectral transitions for targeted metabolite analysis using a triple-quadrupole LC-MS.
- Full scan/untargeted analysis of gene knockout/drug treatment, including follow-up characterization of significantly changed metabolites by tandem mass spectrometry and other methods for identification.
- Identification of significantly changed metabolites and pathways in perturbed biological systems including the development of a targeted methods for new analytes by triple-quadrupole LC-MS.
- Determination of analytical specifications of merit (LOD/LOQ/IDL), linearity, and steps necessary for method validation.
- Measurement of metabolite flux using stable isotope labeling (kinetic flux profiling).
- Determination of reaction kinetics using enzyme assays.
- Determination of glycolytic and respiratory rate by Seahorse XF analysis.
- Development of optimized chromatographic and ionization conditions for separation of both focused and multi-analyte analyses.
- Identification of metabolic constituents of complex fermentation products.

## PARTICIPANTS

Andersen, J., B.Sc./M.S., University of Copenhagen, Denmark

Barritt, S., Ph.D., Harvard Medical School, Boston, Massachusetts

Cano, A., M.S., Luxembourg Institute of Health, Luxembourg City

Cui, J., Ph.D., University of Washington, Seattle

Cui, X., Ph.D., Columbia University Medical Center, New York

Forni, F., Ph.D., Yale University, New Haven, Connecticut

Godbole, A., Ph.D., University of Massachusetts Medical School, Worcester

Grasset, E., Ph.D., Icahn School of Medicine at Mount Sinai, New York

Guan, K., Ph.D., University of New South Wales, School of Medical Science, Sydney, Australia

Hitchings, R., Ph.D., Albert Einstein College of Medicine, Bronx, New York

Howie, R., Ph.D., Georgia Institute of Technology, Atlanta

Jaisinghani, N., Ph.D., Stony Brook University, Stony Brook, New York

Lee, N., Ph.D., University of Massachusetts Medical School, Worcester

Mishra, P., Ph.D., Center for Prostate Disease Research, Bethesda, Maryland

Ward, K., Pharm.D., University of Michigan College of Pharmacy, Ann Arbor

Wright, H., Ph.D., Fred Hutchinson Cancer Research Center, Seattle, Washington

Wu, W., B.A., New York University School of Medicine, New York

SEMINARS

Amador-Noguez, D., University of Wisconsin, Madison: Bile acid transformation by the human gut microbiome.

Clasquin, M., Pfizer, Boston, Massachusetts: Understanding disease through unlabeled and isotope tracer–based metabolomics and lipidomics.

Evans, A., Metabolon, Morrisville, North Carolina: Precision metabolomics: a single technology for understanding human health.

Fan, T., University of Kentucky, Lexington: Exploring human tumor metabolism in situ and in preclinical models using stable isotope–resolved metabolomics (SIRM).

Gottlieb, E., Technion Institute of Technology, Haifa, Israel: Targeting metabolic vulnerabilities of cancer.

Keshari, K., Memorial Sloan Kettering Cancer Center, New York, New York: Interrogating cancer metabolism using hyperpolarized magnetic resonance.

Kind, T., University of California Genome Center, Davis: State-of-the-art approaches for compound identification.

Rhee, K., Weill Cornell Medical College, New York, New York: Mining metabolic dark matter.

Rost, H., University of Toronto, Ontario, Canada: Developing the tools for the personalized medicine revolution: using mass spectrometry for longitudinal molecular profiling.

Vander Heiden, M., Massachusetts Institute of Technology, Cambridge: Understanding metabolic limitations of cell proliferation.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, rigor, and reproducibility.

# Vision: A Platform for Linking Circuits, Behavior, and Perception

**June 12–26**

INSTRUCTORS      **F. Briggs,** University of Rochester Medical Center, New York
                                   **J. Carroll,** Medical College of Wisconsin Eye Institute, Milwaukee

ASSISTANT          **L. Salay,** Stanford University School of Medicine, California

The purpose of this course was to bring together students and faculty for in-depth and high-level discussions of modern approaches for probing how specific cell types and circuits give rise to defined categories of perception and action. It was also designed to address novel strategies aimed at overcoming diseases that compromise sensory function.

The visual system is the most widely studied sensory modality. Recently, three major shifts have occurred in the field of neuroscience. First, because of the large array of genetic techniques available in mice and the relative ease of imaging and recording from the cortex of small rodents, the mouse visual system has become a premiere venue for attacking the fundamental unresolved question of how specific cells and circuits relate to visual performance at the receptive field and whole-animal level. Second, genetic and viral methods have evolved to the point where neurophysiologists can directly probe the role of defined circuits in species such as macaque monkeys, thus bridging the mechanism–cognition gap. Third, the field of visual neuroscience is rapidly paving the way for widespread clinical application of stem-cell, gene therapy, and prosthetic devices to restore sensory function in humans.

The time is ripe to build on the classic paradigms and discoveries of visual system structure, function, and disease, in order to achieve a deep, mechanistic understanding of how receptive fields are organized and filter sensory information, how that information is handled at progressively

higher levels of neural processing, and how different circuits can induce defined categories of percepts and behaviors in the healthy and diseased brain.

## PARTICIPANTS

Cooler, S., M.S., Northwestern University, Chicago, Illinois
De Jesus-Cortes, H., Ph.D., Massachusetts Institute of Technology, Cambridge
Freschl, J., B.A., University of Massachusetts, Boston
Himmelberg, M., B.Psych., University of York, United Kingdom
Jalligampala, A., Ph.D., University of Louisville, Kentucky
Johnson, K., B.S., The George Washington University, Washington, D.C.
Jolly, J., B.Sc., University of Oxford, United Kingdom
Lanfranchi, F., M.S., California Institute of Technologies, Pasadena
Miller, A., B.S., University of Chicago, Illinois
Murphy, A., B.S., University of Rochester, New York
Patel, D., B.S., Medical College of Wisconsin, Milwaukee
Patterson, S., B.S., University of Washington, Seattle
Pons, C., Ph.D., SUNY College of Optometry, New York

Rasmussen, R., M.Sc., Aarhus University, Denmark
Sabbagh, U., B.S., Virginia Tech, Roanoke
Scalabrino, M., Ph.D., Duke University, Durham, North Carolina
Seidl, S., Ph.D., University of California, Davis
Shah, S., M.S., University of Rochester Medical Center, New York
Skyberg, R., Ph.D., University of Virginia, Charlottesville
Smith, J., B.A., Vanderbilt University, Nashville, Tennessee
Szatko, K., M.S., Max Planck Institute for Biological Cybernetics, Tübingen, Germany
Varadarajan, S., Ph.D., Stanford University, California
Wheatcroft, T., M.Sci., University College London, United Kingdom
Woertz, E., M.D./Ph.D., Medical College of Wisconsin, Milwaukee

## SEMINARS

Araj, H., Johns Hopkins University, Baltimore, Maryland; Huberman, A., Stanford University School of Medicine, Palo Alto, California; Rose, S., Foundation Fighting Blindness, Columbia, Maryland; Wright, C., NEI/National Institutes of Health, Bethesda Maryland: Career development day.
Boye, S., University of Florida, Gainesville; Bridge, H., University of Oxford, Great Britain: Restoring vision.
Briggs, F., University of Rochester Medical Center, New York; Carroll, J., Medical College of Wisconsin Eye Institute, Milwaukee: Visual system, overview.
Chen, C., Boston Children's Hospital, Harvard Medical School, Massachusetts; Bickford, M., University of Louisville, Kentucky: Focus on the thalamus.
Curcio, C., University of Alabama, Birmingham; Horton, J., University of California, San Francisco: Diseases of the eye and brain.

Field, G., Duke University School of Medicine, Durham, North Carolina; Hirsch, J., University of Southern California, Los Angeles: Precortical computations.
Maunsell, J., University of Chicago, Illinois; Krauzlis, R., NEI/National Institutes of Health, Bethesda, Maryland: Attention and goal-directed behavior.
Movshon, J.A., New York University, New York; Connor, E., Johns Hopkins University, Baltimore, Maryland: What does the extrastriate cortex do?
Usrey, W.M., University of California, Davis; Sherman, S., The University of Chicago, Illinois: Visual system big picture.
Witkowski, J., Cold Spring Harbor Laboratory: Ethics seminar.

# Statistical Methods for Functional Genomics

**June 28–July 11**

| | |
|---|---|
| INSTRUCTORS | **H. Bussemaker,** Columbia University, New York, New York |
| | **S. Davis,** National Institutes of Health, Columbia, Maryland |
| | **M. Love,** University of North Carolina, Chapel Hill |
| | |
| CO-INSTRUCTORS | **K. Korthauer,** Dana-Farber Cancer Institute, Boston, Massachusetts |
| | **H.T. Rube,** Columbia University, New York, New York |
| | |
| ASSISTANT | **X. Li,** Columbia University, New York, New York |

Over the past decade, high-throughput assays have become pervasive in biological research because of both rapid technological advances and decreases in overall cost. To properly analyze the large data sets generated by such assays and thus make meaningful biological inferences, both experimental and computational biologists must understand the fundamental statistical principles underlying analysis methods. This course was designed to build competence in statistical methods for analyzing high-throughput data in genomics and molecular biology.

Topics included:

- The R environment for statistical computing and graphics.
- Introduction to Bioconductor.
- Review of basic statistical theory and hypothesis testing.
- Experimental design, quality control, and normalization.
- High-throughput sequencing technologies.
- Expression profiling using RNA-Seq and microarrays.
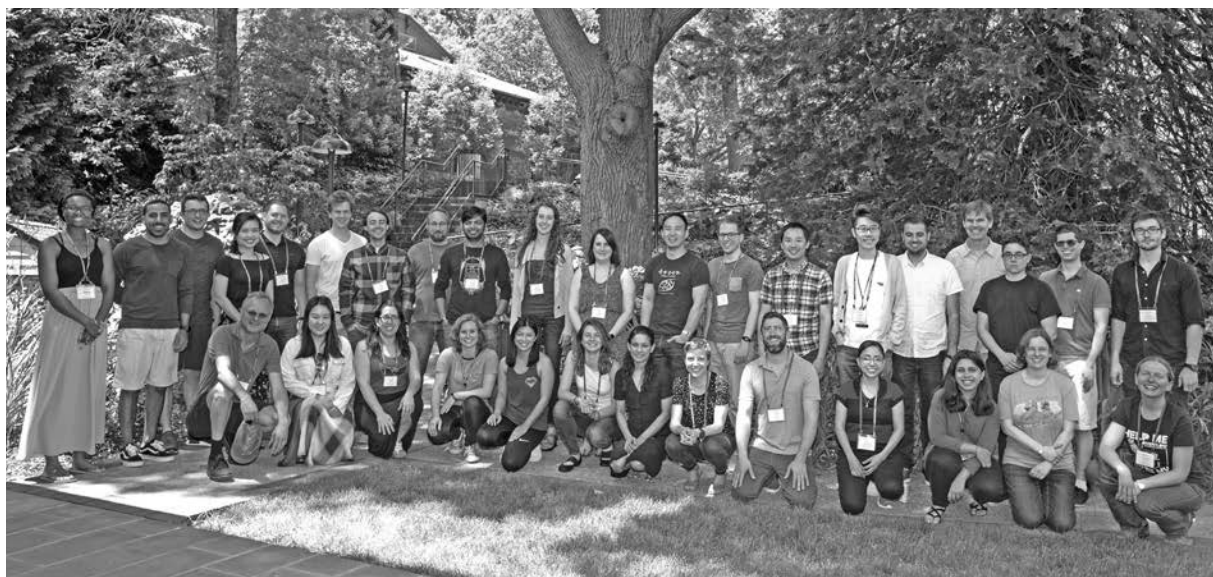
- In vivo protein binding using ChIP-seq.
- High-resolution chromatin footprinting using DNase-seq.
- DNA methylation profiling analysis.
- Integrative analysis of data from parallel assays.
- Representations of DNA-binding specificity and motif discovery algorithms.
- Predictive modeling of gene regulatory networks using machine learning.
- Analysis of posttranscriptional regulation, RNA-binding proteins, and microRNAs.

Detailed lectures and presentations by instructors and guest speakers were combined with hands-on computer tutorials. The methods covered in the lectures were applied to example high-throughput data sets.

This course was supported with funds provided by the National Institute of General Medical Sciences.

## PARTICIPANTS

Begnis, M., M.Sc., EPFL, Switzerland

Bhattacharya, D., B.S., Cornell University, New York, New York

Bott, A., B.S., University of Utah, Salt Lake City

Buchmuller, B., B.Sc., Technical University of Dortmund, Germany

Chatterjee, A., B.S., University of Otago, Dunedin, New Zealand,

Gaine, M., Bsc.HON., The University of Iowa, Iowa City

Ghezzi, A., B.S., University of Puerto Rico, Rio Piedras

Gomez, N., B.S., The Rockefeller University, New York, New York

Greenstein, R., B.A., University of California, San Francisco

Herzner, A-M., Diplom., Genentech, Inc., California

Lopez-Fuentes, E., B.S., University of California, San Francisco

MacPherson, R., B.S., Clemson University, South Carolina

Maron, S., B.S., Memorial Sloan Kettering Cancer Center, New York, New York

McDiarmid, T., B.Sc., University of British Columbia, Vancouver, Canada

Medina-Feliciano, J., B.S., University of Puerto Rico Rio Piedras

Oh, S., B.S., City University of New York/SPH, New York

Oomen, M., B.S., University of Massachusetts Medical School, Worcester

Pacheco, N., B.S., Inova Translational Medicine Institute, Falls Church, Virginia

Shrestha, P., B.S., Cold Spring Harbor Laboratory

Silvester, J., B.A., Harvard Medical School, Boston, Massachusetts

Tsuda, S., B.S., The Scripps Research Institute, Jupiter, Florida

Wade, A., B.A., University of California, Davis

Wang, Z., B.Sc., Memorial Sloan Kettering Cancer Center, New York, New York

Yang, S., B.S., The Jackson Laboratory, Bar Harbor, Maine

Zamuner, F., B.S., Johns Hopkins School of Medicine, Baltimore, Maryland

## SEMINARS

Adamson, B., University of California, San Francisco: CRISPR-based functional genomics performed with single-cell resolution.

Bussemaker, H., Columbia University, New York, New York: Normal distribution and multiple testing. Scoring differential expression. Gene Ontology scoring. Mapping the genetic determinants of TF activity.

Davis, S., National Institute of Health, Columbia, Maryland: Introduction to R. Bioconductor overview. Bioconductor II.

Korthauer, K., Dana-Farber Cancer Institute, Boston, Massachusetts: Single-cell RNA-Seq.

Lappalainen, T., New York Genome Center/Columbia University, New York: Introduction to eQTLs and allele-specific expression.

Love, M., University of North Carolina, Chapel Hill, Carrboro: Basics of RNA-Seq analysis. RNA-Seq isoform level analysis.

Mohlke, K., University of North Carolina, Chapel Hill: ATAC-seq.

Patro, R., Stony Brook University, Stony Brook, New York: Quantifying transcript abundance.

Patro, R., Stony Brook University, Stony Brook, New York; Love, M., University of North Carolina, Chapel Hill, Carrboro: Quantification and QC lecture/lab (Salmon, Tximport, MultiQC).

Rube, T., Columbia University, New York, New York: Bayes' theorem. Student's t-distribution demystified. Dimensional reduction. Basics of linear regression. Basics of ChIP-seq analysis.

# Advanced Techniques in Molecular Neuroscience

INSTRUCTORS      **C. Lai,** Indiana University, Bloomington
**J. LoTurco,** University of Connecticut, Storrs
**A. Schaefer,** Icahn School of Medicine at Mount Sinai, New York

ASSISTANTS      **A. Badimon,** Icahn School of Medicine at Mount Sinai, New York
**A. Battison,** University of Connecticut, Storrs
**A. Chan,** Mount Sinai School of Medicine, New York
**M. Kaye Duff,** Mount Sinai School of Medicine, New York
**N. Khatri,** Indiana University, Carmel
**E. Perez,** Indiana University, Bloomington
**H. Strasburger,** Icahn School of Medicine at Mount Sinai, New York
**J. Sullivan,** Icahn School of Medicine at Mount Sinai, New York
**S. Veugelen,** Icahn School of Medicine at Mount Sinai, New York

This laboratory and lecture course was designed to provide neuroscientists at all levels with a conceptual and practical understanding of several of the most advanced techniques in molecular neuroscience. The course curriculum was divided into three sections: an extensive and up-to-date set of laboratory exercises, daily lectures covering the theoretical and practical aspects of the various methods used in the laboratory, and a series of evening research seminars. The informal and interactive evening lectures were given by leading molecular neuroscientists and served to illustrate the ways in which the various experimental approaches have been used to advance specific areas of neurobiology. In this year's course, the laboratory portion included topics such

as an introduction to the design and use of animal virus vectors in neurobiology; the use of CRISPR genome editing and RNAi approaches for regulating the expression of specific genes in neurons; practical exercises in gene delivery systems including mammalian cell infection and transfection and electroporation techniques for targeted gene transfer in vivo; an introduction to overall strategies, use, and design of BAC transgenic vectors; real-time RT-PCR analyses; assays of chromatin and chromatin structure in neurons, and mRNA isolation from specified neural subtypes by TRAP.

This course was supported with funds provided by the National Institute of Mental Health and partial scholarship support was provided by the Regeneron Scholars Fund.

## PARTICIPANTS

Chen, S., M.D./Ph.D., Indiana University School of Medicine, Indianapolis

Cheng, S., M.D., University of California, Los Angeles

De La Torre, A., B.S., Dartmouth College, Hanover, New Hampshire

Delgado Garcia, L.M., M.Sc., UNIFESP, Sao Paulo, Brazil

Edokpolor, K., B.A., Emory University, Atlanta, Georgia

Fernandez, V., Ph.D., Washington University School of Medicine in St. Louis, Missouri

Itaman, S., B.S., Stony Brook University, Stony Brook, New York

Lin, T-C., Ph.D., German Center for Neurodegenerative Diseases, Germany

Lu, Y., Ph.D., Yale University, New Haven, Connecticut

Maciel Camargo, C., Ph.D., University of California, Santa Barbara

Morgunova, A., B.A., McGill University/Douglas Mental Health Institute, Montreal, Quebec, Canada

Paisley, C., B.S., Duke University, Durham, North Carolina

Reid, C., B.S., Harvard University, Boston, Massachusetts

Seaks, C., B.S., University of Kentucky, Lexington

Shekhar, K., Ph.D., Broad Institute of MIT and Harvard, Cambridge, Massachusetts

Yuan, C., Ph.D., Weill Cornell Medical College, New York

## SEMINARS

Darnell, R., The Rockefeller University/HHMI, New York: RNA genomics in health and disease.

Desplan, C., New York University, New York: The generation of neural diversity.

Kenny, P., Icahn School of Medicine at Mount Sinai, New York: microRNA regulation of drug craving.

Liddelow, S., New York University Langone Medical Center, New York: Methods to study physiological and pathological roles of glia.

Haas, K., University of British Columbia, Vancouver, Canada: Single-cell electroporation for in vivo neuronal transfection.

Harvey, B., NIDA/National Institutes of Health, Bethesda, Maryland: Genome editing in the adult rat brain.

Harwell, C., Harvard Medical School, Boston, Massachusetts: Generating neural diversity in the forebrain.

Nestler, E., Icahn School of Medicine at Mount Sinai, New York: Transcriptional and epigenetic mechanisms of depression.

Schafer, D., University of Massachusetts Medical School, Worcester: Studying microglial function and dysfunction within neural circuits in health and disease.

Schmidt, E., The Rockefeller University, New York, New York: Molecular phenotyping of discrete cell types using the translating affinity purification (TRAP) approach.

Silver, D., Duke University Medical Center, Durham, North Carolina: Ex vivo and in vitro techniques in cortical development.

Tollkuhn, J., Cold Spring Harbor Laboratory: Transcription factor occupancy profiles from limited neuronal populations: the promise of CUT&RUN.

Treweek, J., University of Southern California, Los Angeles: Transcriptional and morphological profiling of complex cellular niches with tissue clearing.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, rigor, and reproducibility.

# Single-Cell Analysis

**June 28–July 13**

INSTRUCTORS

**D. Chenoweth,** University of Pennsylvania, Merion Station
**M. McConnell,** University of Virginia School of Medicine, Charlottesville
**G. Yeo,** University of California, San Diego, La Jolla

COURSE TEACHING ASSISTANTS

**N. Ahmed,** University of California, San Diego
**M. Haakenson,** University of Virginia School of Medicine, Charlottesville
**R. Lackner,** University of Pennsylvania, Philadelphia
**S. Shaffer,** University of Pennsylvania, Philadelphia
**Y. Shi,** University of Virginia, Charlottesville
**Y. Song,** University of California, San Diego
**D. Wu,** University of Pennsylvania, Philadelphia
**B. Yee,** University of California, San Diego, La Jolla

The goal of this two-week course was to familiarize students with cutting-edge technologies for characterization of single cells. Modules of the course were taught by scientists with expertise in distinct areas of single-cell analysis. Topics that were covered included quantitative single-cell analysis by RNA-Seq, genomic DNA analysis, proteomics, and metabolomics. Multiple nucleic amplification methodologies including droplet-based RNA-Seq, MALBAC, and MDA were used. In addition, students were instructed in basic bioinformatic analysis of next-generation sequencing data.

Topics covered in the course included:

- Single-cell genome, transcriptome, and proteome measurement.
- Introductory next-generation sequencing data analysis.
- Photoactivatable single-cell probes.
- Single-cell mass spectrometry/soft X-ray tomography.

## PARTICIPANTS

Balasooriya, G., Ph.D., Cold Spring Harbor Laboratory

Brandt, L., B.A./M.S., MRC Laboratory of Molecular Biology, Cambridge, United Kingdom

Chege, P., Ph.D., University of Melbourne, Parkville, Victoria, Australia

Ding, K., B.S., University of Pittsburgh School of Medicine, Pennsylvania

Dona, R., B.S., Albert Einstein College of Medicine, Bronx, New York

Gardner, E., Ph.D., Weill Cornell Medicine, New York, New York

He, J., Ph.D., University of Pittsburgh, Pennsylvania

Hopkins, J., B.S., Tufts University, Boston, Massachusetts

Lazaris, C., M.S./Ph.D., University of Southern California, Los Angeles

LeBon, L., Ph.D., Calico Life Sciences, San Francisco, California

Lehmann, V., B.A., Icahn School of Medicine at Mount Sinai, New York

Mika, K., Ph.D., University of Chicago, Illinois

Narvaes del Pilar, O., B.S., University of Texas MD Anderson Graduate School of Biomedical Science, Houston

Oprescu, S., B.S., Purdue University, West Lafayette, Indiana

Smoklin, R., B.S., Memorial Sloan Kettering Cancer Center, New York

Sulka, K., B.A., Sackler School of Graduate Biomedical Sciences, Boston, Massachusetts

Svoboda, M., B.A., Dartmouth College, Hanover, New Hampshire

Vaishnavi, A., Ph.D., University of Utah, Salt Lake City

## SEMINARS

Allbritton, N., University of North Carolina, Chapel Hill: Cell separation based on complex phenotypes.

Herr, A., University of California, Berkeley: Protein analysis.

Knouse, K., Whitehead Institute/Massachusetts Institute of Technology, Cambridge: Prevalence and prevention of large-scale somatic copy number alterations.

Larabell, C., University of California, San Francisco: Single-cell CT scans.

Raj, A., University of Pennsylvania, Philadelphia: Single-cell analysis in cancer.

Rubakhin, S., University of Illinois, Urbana-Champaign: Single-cell mass spectrometry: fundamentals and applications.

Schultz, C., Oregon Health & Science University, Portland: Novel imaging tools for single-cell analysis.

Sims, P., Columbia University, New York, New York: New tools for single-cell transcriptomics in cancer.

Sweedler, J., University of Illinois, Urbana: The cell by cell chemical characterization of the brain using mass spectrometry.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, rigor, and reproducibility.

# *Drosophila* Neurobiology: Genes, Circuits, and Behavior

**June 28–July 18**

<table>
<tr><td>INSTRUCTORS</td><td><strong>C.A. Frank,</strong> University of Iowa, Iowa City<br><strong>E. Heckscher,</strong> University of Chicago, Illinois<br><strong>A. Keene,</strong> Florida Atlantic University, Jupiter</td></tr>
<tr><td>ASSISTANTS</td><td><strong>E. Brown,</strong> Florida Atlantic University, Jupiter<br><strong>M. Greaney,</strong> University of Chicago, Illinois<br><strong>K. Lembke,</strong> University of Iowa, Iowa City<br><strong>B. Mallik,</strong> University of Iowa, Iowa City<br><strong>Y-W. Wang,</strong> University of Chicago, Illinois</td></tr>
</table>

This laboratory/lecture course was intended for researchers at all levels from beginning graduate students through established primary investigators who want to use *Drosophila* as an experimental system for nervous system investigation. This three-week course was designed to introduce students to a wide variety of topics and techniques, including the latest approaches for studying nervous system development, activity, and connectivity, as well as complex behaviors and disease models. Daily research seminars presented comprehensive overviews of specific subfields of nervous system development or function or focused on state-of-the-art techniques and approaches in *Drosophila* neuroscience. Expert guest lecturers discussed their findings and approaches and brought along their own assays and techniques for students to learn in the laboratory part of the course. The hands-on portion of the course was centered on inquiry-based projects, using the different morphological and physiological measurements and behavioral paradigms learned at the course. This included molecular – genetic analyses, immunocytochemistry, recording of activity using electrophysiology and genetically encoded calcium indicators, optogenetic and thermogenetic

control of neural activity, and numerous quantitative behavioral measures. Collectively, the course provided a comprehensive and practical introduction to modern experimental methods for studying the neural basis of behavior in *Drosophila*.

## PARTICIPANTS

Beigaite, E., B.S., University of York, United Kingdom

Dopp, J., M.Sc., VIB-KU Leuven Center for Brain & Disease Research, Belgium

Fernandez Chiappe, F., B.S., Biomedicine Research Institute of Buenos Aires, Argentina

Gonzalez Segarra, A., B.S., University of California, Berkeley

Hernandez, R., B.A., Florida Atlantic University, Jupiter

Li, X., Ph.D., Princeton University, New Jersey

M'Angale, P., Ph.D., University of Massachusetts Medical School, Worcester

Nemtsova, Y., B.A., Brown University, Providence, Rhode Island

Niesman, P., B.S., Yale University, New Haven, Connecticut

Shin, M., Ph.D., University of Virginia, Charlottesville

Silva Moeller, V., B.S., Universidad de Valparaiso, Chile

Smoyer, C., Ph.D., University of California, Davis

Turkcan, M., B.Sc., Columbia University, New York, New York

Vaikakkara Chithran, A., B.Tech., University of British Columbia, Vancouver, Canada

## SEMINARS

Ashley, J., and Carrillo, R., University of Chicago, Illinois: Larval dissection and immunostaining.

Collins, C., University of Michigan, Ann Arbor: Injury and stress response signaling.

Dacks, A., West Virginia University, Morgantown: Neuromodulation: what flies can learn from the crunchies and the squishies.

Desplan, C., New York University, New York: The generation of neural diversity.

Frank, C.A., University of Iowa, Iowa City: Homeostatic plasticity and extensions to NMJ physiology.

Frank, C.A., University of Iowa, Iowa City; Carrillo, R., University of Chicago, Illinois: Neurophysiology and synaptic neurotransmission.

Heckscher, E., University of Chicago, Illinois: Development and function of neural circuits in the motor system.

Jayaraman, V., Howard Hughes Medical Institute, Ashburn, Virginia: Abstract internal representations and attractor dynamics in the fly brain.

Keene, A., Florida Atlantic University, Jupiter: Genetic dissection of sleep–metabolism interactions.

Keene, A., Florida Atlantic University, Jupiter; Inglis, J., BioRxiv, Woodbury, New York: Scientific rigor and reproducibility in *Drosophila* neurobiology research. BioRxiv.

Kohwi, M., Columbia University, New York, New York: Neuroblast development.

Lazar, A., Columbia University, New York, New York; Younger, M., The Rockefeller University, New York, New York: Building the functional map of the fruit fly brain.

Levine, J., University of Toronto, Mississauga, Ontario, Canada: Social networks weather or not?

Louis, M., University of California, Santa Barbara: Algorithms controlling sensory navigation in the *Drosophila.*

Mosca, T., Thomas Jefferson University, Philadelphia, Pennsylvania: *The microscopy tolkiy for the modern Drosophilist.*

Muraro, N., Biomedicine Research Institute of Buenos Aires, Argentina: Electrophysiology in central neurons of *Drosophila.* Circadian rhythms and electrophysiology of clock neurons.

Nagel, K., New York University School of Medicine, New York: Algorithms and circuits for olfactory navigation.

Reis, T., University of Colorado Medical School, Aurora: Neuronal control of energy balance.

Reiser, M., Howard Hughes Medical Institute, Ashburn, Virginia: Vision in flies.

Stahl, B., Florida Atlantic University, Jupiter: Effects of glia on behavior and neurodegeneration. Wallerian degeneration.

Tomchik, S., The Scripps Research Institute, Jupiter, Florida: Central brain Ca-imaging.

Tracey, D., Indiana University, Bloomington: Nociception in *Drosophila.*

Vecsey, C., Skidmore College, Saratoga Springs, New York: Sleeping on the fly.

Volkan, P., Duke University, Durham, North Carolina: David Featherstone Memorial Lecture: How development and environment shape the brain: lessons from the fly olfactory system.

Wildonger, J., University of Wisconsin, Madison: How to build a neuron: microtubules and molecular motors.

# Frontiers and Techniques in Plant Science

INSTRUCTORS    **S. Cutler,** University of California, Riverside
**J. Dinneny,** Stanford University, California
**J. Law,** The Salk Institute, La Jolla, California
**U. Paszkowski,** University of Cambridge, United Kingdom

ASSISTANTS    **A. Kolbeck,** Universite de Lausanne, Switzerland
**N.C. Sanden,** University of Copenhagen, Denmark
**M. Sorkin,** Danforth Plant Science Center, St. Louis, Missouri

This course provided an intensive overview of topics in genomics, genetics, physiology, biochemistry, development, and evolution, and hands-on experiences in molecular, imaging, computational, and high-throughput approaches to understanding plant biology. It emphasized recent results from model organisms including *Arabidopsis*, maize, and tomato, as well as a variety of other plants, and provided an introduction to current methods used in basic and applied plant biology, both theoretically and practically. The seminar series included plant morphology and anatomy, development, evolution, light, and circadian biology, hormones, small RNAs and epigenetic inheritance, biotic and abiotic interactions, plant biochemistry, crop domestication, and applications addressing current agronomic problems. Speakers provided expert overviews of their fields, followed by in-depth discussions of their own work. The laboratory sessions provided exposure to cutting-edge experimental and computational techniques currently used in plant research. These included approaches for studying plant development, regulatory networks,

transient gene expression, cell type–specific gene expression analysis, computational large-scale data analysis, and applications of fluorescent proteins including live imaging, genome editing, and chromatin immunoprecipitation.

Students gained hands-on experience in computational tools and environments for genome assembly, plant imaging and image analysis, design and use of fluorescent sensors, transcriptomics, identification of quantitative trait loci, mapping by sequencing, mathematical modeling of development and hormone action, purification of cell type–specific nuclei (INTACT), and high-throughput cloning.

## PARTICIPANTS

Borowsky, A., Ph.D., University of California, Riverside

Bull, T., B.S., The Genome Center/University of California, Davis

Chiu, H., Ph.D., University of Cambridge, United Kingdom

Gomez De La Cruz, D., M.Sc., The Sainsbury Laboratory, Norwich, United Kingdom

Gregory, J., M.S., Rutgers/The State University of New Jersey, New Brunswick

Hull, R., M.Bioch., Max Planck Institute for Plant Breeding Research, Germany

Jores, T., Ph.D., University of Washington, Seattle

Joshi, S., M.S., University of Kentucky, Lexington

Kang, X., Ph.D., University of Illinois, Urbana-Champaign

Meda, S., M.Sc., University of Tübingen, Germany

Ragland, C., Ph.D., Stanford University, California

Riaz, N., M.S., Dartmouth College, Hanover, New Hampshire

Serra Serra, N., M.S., Gregor Mendel Institute, Vienna, Austria

Shih, K., Ph.D., Stanford University, California

Weiss, T., B.S., University of Minnesota, St. Paul

Yasmin, F., M.S., University of North Carolina, Charlotte

## SEMINARS

Ainsworth, L., United States Department of Agriculture/Agricultural Research Service Urbana, Illinois: Stress responses in crops: a view from the field.

Bailey-Serres, J., University of California, Riverside; Deal, R., Emory University, Atlanta, Georgia: Plant TRAP and INTACT workshop. The INTACT method for cell type–specific nuclei.

Balcombe, D., University of Cambridge, United Kingdom: RNA silencing.

Bergmann, D., Stanford University, California: Making a difference.

Dolan, L., University of Oxford, United Kingdom: Root history: evolution of the earliest land plant rooting systems.

Dong, X., Duke University, Durham, North Carolina: Regulation of plant immune responses.

Haswell, L., Washington University in St. Louis, Missouri: Plants under pressure: tools for studying mechanobiology.

Hibberd, J., University of Cambridge, United Kingdom: Understanding and engineering photosynthesis.

Jander, G., Boyce Thompson Institute for Plant Research, Ithaca, New York: *Arabidopsis* quantitative trait mapping.

Johnson, M., Brown University, Barrington, Rhode Island: Plant reproduction.

Jones, A., University of Cambridge, United Kingdom: Hormone biosensing.

Kellogg, T., Donald Danforth Plant Science Center, St. Louis, Missouri: Conservation and diversity of grass abscission zones.

Martienssen, R., Cold Spring Harbor Laboratory: The epigenetic inheritance of transposons, chromosomes, and small RNA.

Nodine, M., Gregor Mendel Institute of Molecular Plant Biology, Austria: Zygotic genome activation and early plant embryogenesis. Small RNA functions in *Arabidopsis*.

Oldroyd, G., University of Cambridge, United Kingdom: Reducing agricultural reliance on inorganic fertilizers through beneficial associations between plants and microbes.

Paszkowski, U., University of Cambridge, United Kingdom: Arbuscular mycorrhizal symbiosis.

Pedmale, U., Cold Spring Harbor Laboratory: Plant photosensory perception and signaling.

Provart, N., University of Toronto, Ontario, Canada: Raising the BAR for hypothesis generation in plant biology using open big data.

Sinha, N., University of California, Davis: The arms race between tomato and cuscuta.

Surrige, C., SpringerNature, Cambridge, United Kingdom: Publishing without tears.

Voytas, D., University of Minnesota, Saint Paul: Editing the plant genome.

Wilkins, O., McGill University, Montreal, Quebec, Canada: Regulatory networks.

# Neural Data Analysis

**July 13–26**

INSTRUCTORS     **M. Pachitariu,** HHMI/Janelia, Ashburn, Virginia
    **M. Reimers,** Michigan State University, East Lansing
    **P. Wallisch,** New York University, New York

CO-INSTRUCTORS     **K. Bonnen,** New York University, New York
    **M. Moore,** Michigan State University, East Lansing
    **M. Savestani,** Max Planck Florida Institute of Neuroscience, Jupiter
    **C. Stringer,** HHMI/Janelia, Ashburn, Virginia

Today's technologies enable neuroscientists to gather data in previously unimagined quantities. This necessitates—and allows for—the development of new analysis methods to address dynamic systems function of brain networks.

This course was designed to help neuroscience practitioners to develop the conceptual and practical capabilities to meet the challenges posed by the analysis of these hard-won and large data sets. We emphasized statistical issues such as the preprocessing of data, sampling biases, estimation methods, and hypothesis testing, as well as data wrangling (in MATLAB). We worked with data from a variety of recording technologies including multi-electrode array recordings, local field potentials, and EEG, as well as two-photon and wide-field optical imaging.

## PARTICIPANTS

Akrouh, A., B.A., Columbia University, New York

Faulkner, A., B.S., University of Michigan, Ann Arbor

Foustoukos, G., B.S./M.Sc., Ecole Polytechnique Federale de Lausanne, Switzerland

Gross, I-M., B.A., Max Planck Institute of Neurobiology, Germany

Gruzdeva, A., B.S., National Research Center "Kurchatov Institute," Russia

Ianni, G., B.A., The Rockefeller University, New York, New York

Jaepel, J., B.S./M.S., Max Planck Florida Institute for Neuroscience, Jupiter

Lebedeva, A., Spec., University College London, United Kingdom

Li, P., B.S., The Salk Institute for Biological Studies, La Jolla, California

Marachlian, E., B.S., ISERM u1024, France

Mimica, B., M.S., Norwegian University of Science and Technology, Norway

Mohan, H., B.E., Cold Spring Harbor Laboratory

Oryshchuk, A., B.S./M.S., Ecole Polytechnique Federale de Lausanne, Switzerland

Reinhard, K., B.S./M.S., KU Leuven, IMEC, VIB, Belgium

Singh, Alvarado, J., B.S., Duke University, North Carolina

Smith, C., B.A., Baylor College of Medicine, Texas

Solyga, M., B.Sc., Basel University, Switzerland

Spencer-Segal, J., B.A., University of Michigan, Ann Arbor

Stagkourakis, S., B.S., Caltech, Pasadena, California

Stern, M., B.S., Emory University School of Medicine, Georgia

Sun, Y., B.S., University of California, San Francisco

Xia, F., B.Sc., University of California, San Francisco

## SEMINARS

Babadi, B., University of Maryland, College Park: Methods and limitations of inferring networks from spiking data. Inferring networks from calcium data (practical).

Bassett, D., University of Pennsylvania, Philadelphia: Using graph theory to understand interaction patterns in neural data.

Churchland, A., Cold Spring Harbor Laboratory: International Brain Laboratory.

Cohen, M., University of Amsterdam, the Netherlands: Analog signals II (LFP); linear algebra + least squares. Analog signals II (LFP); eigendecomposition and source separation. Analog signals II (LFP); space curves.

Giovannucci, A., Simons Center for Data Analysis, New York: CaImAn: an open source toolbox for large-scale calcium imaging data analysis on standalone machines.

Iyer, V.: New parallel computing MATLAB functions for big data in neuroscience.

Kleinfeld, D., University of California, San Diego, La Jolla: Electrons, photons, and viruses to reverse engineer sensorimotor dynamics. The good, the bad, and the ugly of emerging directions in neurotechnology.

Kording, K., Northwestern University, Evanston, Illinois: Intro to spike analysis and machine learning. Approaches to decoding population data: classification methods.

Kording, K., Northwestern University, Evanston, Illinois; Reimers, M., Michigan State University, East Lansing: Debate: How to make neuroscience better.

Mohajerani, M., University of Lethbridge, Alberta, Canada: Wide-field optical imaging: characteristics of different imaging modalities and indicators. Research questions that can be addressed with wide-field optical methods. Quantitative analysis toolbox for characterization of spatiotemporal dynamics in mesoscale optical imaging of brain activity.

Pachitariu, M., University College London, United Kingdom: What you can find out by recording 10K neurons. Finding cells in calcium images: the Suite2P approach. Using Suite2P (practical). High-density electrophysiology: spike sorting and analyzing neuropixels data.

Pillow, J., Princeton University, New Jersey: Spikes II—ENCODING—model-based spike analysis.

Reimers, M., Michigan State University, East Lansing: Spike sorting: an unsolvable problem. The landscape of dimension reduction methods. Visualizing state space: dimensionality reduction methods.

Reimers, M., Michigan State University, East Lansing; Wallisch, P., New York University, New York: Introduction and approach to neural data analysis. Synthesis.

Smith, M., University of Pittsburgh, Pennsylvania: Analog signals I (LFP); time and frequency domain analysis. Spikes III: array analysis (spike count correlations).

Steinmetz, N., University of Washington, Seattle: Analog signals I (LFP); TF domain methods, continued.

Wallisch, P., New York University, New York: Analysis of "big" neural data. Of mice, monkeys, and (wo)men.

Wallisch, P., New York University, New York; Steinmetz, N., University of Washington, Seattle: How to write a paper.

# Synthetic Biology

**July 23–August 5**

INSTRUCTORS   **J. Chappell,** Rice University, Houston, Texas
              **E. Franco,** University of California, Los Angeles
              **P. Romero,** University of Wisconsin, Madison
              **M. Smanski,** University of Minnesota, St. Paul
              **O. Venturelli,** University of Wisconsin, Madison

ASSISTANTS    **R. Clark,** University of Wisconsin, Madison
              **C. Cuba Samaniego,** Massachusetts Institute of Technology, Cambridge
              **R. Hon Hsu,** University of Wisconsin, Madison
              **B. Liu,** Rice University, Houston, Texas
              **A. Sychla,** University of Minnesota, Falcon Heights

Cells are the world's most sophisticated chemists, and their ability to adapt to changing environments offers enormous potential for solving modern engineering challenges. Nonetheless, biological systems are noisy, massively interconnected, and nonlinear, and they have not evolved to be easily engineered. The grand challenge of synthetic biology is to reconcile the desire for a predictable, formalized biological design process with the inherent "squishiness" of biology.

   This course focused on how the complexity of biological systems can be combined with traditional engineering approaches to result in new design principles for synthetic biology. The centerpiece of the course was an immersive laboratory experience in which students worked in teams to learn the practical and theoretical underpinnings of synthetic biology research. Broadly,

the course explored how cellular regulation (transcriptional, translational, posttranslational, and epigenetic) can be used to engineer cells that accomplish well-defined goals.

Laboratory modules covered the following areas: Microfluidics for high-throughput characterization of biological systems, cell-free transcription and translation systems to characterize genetic circuits and RNA regulators, modeling gene expression using ordinary differential equations, DNA assembly and design of expression cassettes, and computational modeling of genetic circuits and microbial communities.

Students first learned essential synthetic biology techniques in a four-day "boot camp" at the beginning of the course. Following the boot camp, they rotated through research projects in select areas. Students also interacted closely with a panel of internationally recognized speakers who collectively provided a broad overview of synthetic biology applications, including renewable chemical production and therapeutics, state-of-the-art techniques, case studies in human practices, and socially responsible innovation.

## PARTICIPANTS

Buckley, S., Ph.D., Boston University/Harvard Medical School, Boston

Capel, P., M.Chem., University of Warwick, Coventry, United Kingdom

Castano-Uruena, J., B.S., University of Minnesota, St. Paul

D'Ambrosio, V., M.Sc., Technical University of Denmark, Kongens Lyngby, Denmark

Erickson, S., B.A., University of Minnesota, Falcon Heights

Hindley, J., Ph.D., Imperial College London, United Kingdom

Klocke, M., B.A., Bourns College of Engineering/University of California, Riverside

Lebovich, M., M.S., University of Massachusetts, Amherst

Markakiou, S., M.Sc., Chr. Hansen A/S, Hörsholm, Denmark

Nguyen, M., B.S., Aarhus University, Denmark

Pruefer, F., Ph.D., Grupo Maxwerk & Maxwerk Biotech, Mexico City, Mexico

Rutter, J., M.Eng., University College London, United Kingdom

Semkiv, M., Ph.D., Institute of Cell Biology, National Academy of Sciences of Ukraine, Lviv, Ukraine

Shieh, P., Ph.D., Massachusetts Institute of Technology, Cambridge

Suarez Heredia, R., B.Eng., University College London, United Kingdom

Yu, M., Ph.D., Toyota Technological Institute, Chicago, Illinois

## SEMINARS

Andrews, L., University of Massachusetts, Amherst: Programming sequential logic in microbial consortia and probiotic bacteria.

Arkin, A., University of California, Berkeley: High-throughput genetics for discovering and designing complex phenotypes and manipulation of communities.

Beisel, C., Helmholtz Institute for RNA-based Infection Research, Würzburg, Germany: From CRISPR biology to versatile technologies.

Ceroni, F., Imperial College London, United Kingdom: Characterization of host–construct interactions and cellular burden for synthetic biology.

Ellington, A., University of Texas, Austin: The tenuous balance between systems and synthetic biology.

Haynes, K., Emory University, Atlanta, Georgia: Chromatin epigenetic engineering in triple negative breast cancer.

Khammash, M., ETH Zürich, Basel, Switzerland: Theory and applications of genetic control systems.

Schulman, R., Johns Hopkins University, Baltimore, Maryland: Synthetic biology for dynamic materials.

Weiss, R., Massachusettes Institute of Technology, Cambridge: Mammalian synthetic biology.

Zhao, H., University of Illinois, Urbana-Champaign: Discovery of novel natural products via synthetic biology.

# Chromatin, Epigenetics, and Gene Expression

**July 23–August 11**

INSTRUCTORS    **K. Adelman,** Harvard Medical School, Boston, Massachusetts
**A. Johnson,** University of Colorado School of Medicine, Aurora
**M. Mendillo,** Northwestern University School of Medicine, Chicago, Illinois

ASSISTANTS    **E. Kaye,** Harvard Medical School, Boston, Massachusetts
**B. Martin,** Harvard Medical School, Boston, Massachusetts
**R. Smith,** Northwestern University, Chicago, Illinois
**A. Zukowski,** University of Colorado, Aurora

This course was designed for students, postdocs, and principal investigators who have recently ventured into the exciting area of gene regulation. Emphasis was placed on exposing students to a broad array of methodologies to study gene regulation, chromatin structure, and dynamics, including both state-of-the-art and well-developed methods.

Students performed widely used techniques such as chromatin immunoprecipitation (ChIP) coupled with sequencing (ChIP-seq), reporter assays of enhancer activity, and RNA expression analysis. They applied a basic pipeline to analyze sequencing results and discussed current informatics strategies. They learned about state-of-the-art genetic perturbation strategies. They performed two of these methods to reduce or eliminate the expression of a gene of interest: RNA interference (RNAi), and CRISPR-Cas9 targeted disruption. Furthermore, students compared how each method affects gene expression and function.

In this course, students also learned how to assemble recombinant chromatin with modified histones and test specificity of chromatin "reader" proteins and enzymes that modify chromatin. Quantitative methods were used to analyze activity and selectivity for specific substrates.

Given the broad biological roles for DNA-binding transcription factors, and emerging roles of noncoding RNAs in transcription regulation, electrophoretic mobility shift assays (EMSAs) are again becoming widely used for assessing transcription factor binding to regulatory DNA or RNA elements. Students learned how to perform and interpret EMSA experiments, using quantitative gel-based methods.

This course also provided the basic concepts behind different methods to analyze the chromatin architecture of the genome. Moreover, we discussed the computational methods required to analyze data concerning three-dimensional chromatin architecture.

Experience with basic recombinant DNA and molecular biology techniques was a prerequisite for admission to this course. Lectures by the instructors covered the current state of the gene expression and epigenetics fields, theoretical aspects of the methodology, and broader issues regarding strategies for investigating the regulation of gene expression in eukaryotes. Emphasis was placed on advantages and limitations of specific techniques, and data interpretation. Each evening, an invited speaker who is an expert in the field presented their work and interacted with students. The students were encouraged and expected to actively participate in these discussions, and to take advantage of the many opportunities to network and receive input on their projects and future plans.

## PARTICIPANTS

Albero Gallego, R., Ph.D., Columbia University, New York, New York
Choi, M.H., M.S., University of Bergen, Norway
Costa, E., B.S., Memorial Sloan Kettering Cancer Center, New York, New York
Divakaran, A., B.S., University of Minnesota, Minneapolis
Dunham, N., University of Virginia, Charlottesville
Fitzpatrick, C., B.S., Cold Spring Harbor Laboratory/WSBS
Guerra Resendez, R.S., B.S., Rice University, Houston, Texas
Honnell, V., B.S., St. Jude Children's Research Hospital, Memphis, Tennessee

Lewis, R., B.S., ETH Zürich, Switzerland
Miao, L., Ph.D., Yale University, New Haven, Connecticut
Neofytou, C., M.S., Karolinska Institutet, Solna Stockholm, Sweden
Reddi, P., Ph.D., University of Illinois, Urbana-Champaign
Rodier, J-A., Ph.D., Princeton Neuroscience Institute, New Jersey
Rodriguez, A., M.D./Ph.D., University of Arkansas Medical Sciences, Little Rock
Talley, M.J., B.S., Cincinnati Children's Hospital Medical Center, Ohio
Vaid, R., M.Sc., Stockholm University, Sweden

## SEMINARS

Arndt, K., University of Pittsburgh, Pennsylvania: Analysis of proteins at the interface of chromatin and transcription.
Buratowski, S., Harvard Medical School, Boston, Massachusetts: Coupling transcription with RNA processing and chromatin.
Di Croce, L., ICREA and Center for Genomic Regulation, Spain: Gene regulation dynamics mediated by Polycomb and MLL complexes.
Henikoff, S., Fred Hutchinson Cancer Research Center, Seattle, Washington: Genome-wide mapping of protein − DNA interaction dynamics.

Johnson, T., University of California, Los Angeles: Chromatin modification, RNA processing, and the coordinated control of gene expression.
Kingston, R., Massachusetts General Hospital/Harvard Medical School, Boston: Chromatin compaction and phase separation as epigenetic mechanisms in Polycomb-group function.
Levine, M., Princeton University, New Jersey: Visualization of enhancer−promoter communication in living *Drosophila* embryos.
Lis, J., Cornell University, Ithaca, New York: Probing mechanisms of transcription regulation.

Marti-Renom, M., CNAG-CRG, Barcelona, Spain: Structure determination of genomes and genomic domains by satisfaction of spatial restraints.

Spector, D., Cold Spring Harbor Laboratory: Long noncoding RNAs: basic biology to therapeutic targets.

Stark, A., Research Institute of Molecular Pathology, Vienna, Austria: Decoding transcriptional regulation.

Shilatfard, A., Northwestern University Feinberg School of Medicine, Chicago, Illinois: Principles of epigenetics and chromatin in development and human disease.

Taatjes, D., University of Colorado, Boulder: Understanding transcription regulation through kinase inhibition and biochemical reconstitution.

Tahiliani, M., New York University School of Medicine, New York: TET enzymes: expanding the epigenetic landscape.

Tyler, J., Weill Cornell Medical College, New York, New York: Chromatin assembly and disassembly dynamics.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, rigor, and reproducibility.

# Imaging Structure and Function in the Nervous System

**July 23–August 12**

INSTRUCTORS      **F. Albeanu,** Cold Spring Harbor Laboratory
                 **S. Dieudonné,** Ecole Normale Supérieure, Paris, France
                 **B. Judkewitz,** Charité and Humboldt University, Berlin, Germany
                 **M. Orger,** Champalimaud Foundation, Lisbon, Portugal
                 **L. Palmer,** University of Melbourne, Victoria, Australia
                 **P. Tsai,** University of California, San Diego

CO-INSTRUCTOR    **J. Donovan,** Max Planck Institute of Neurobiology, Martinsreid, Germany

ASSISTANTS       **A. Bandyopadhyay,** New York University Langone Medical Center, New York
                 **C. Berlage,** Charité University Medicine, Berlin, Germany
                 **L. Godenzini,** Florey Institute of Neuroscience & Mental Health, Parkville, Victoria, Australia
                 **A. Groneberg,** Champalimaud Foundation, Lisbon, Portugal
                 **A. La Chioma,** Max Planck Institute of Neurobiology, Martinsreid, Germany
                 **B. Mathieu,** Ecole Normale Supérieure, Paris, France
                 **S. Musall,** Cold Spring Harbor Laboratory
                 **B. Pichler,** Independent NeuroScience Services, Lewes, East Sussex, United Kingdom
                 **C. Whitmire,** Max Delbrück Center for Molecular Medicine, Berlin, Germany

Advances in light microscopy and digital image processing and the development of a variety of powerful fluorescent probes present expanding opportunities for investigating the nervous system, from synaptic spines to networks in the brain. This intensive laboratory and lecture course provided participants with the theoretical and practical knowledge to use emerging imaging technologies. The

primary focus of the course was on in vivo applications of light microscopy, particularly functional imaging with genetically encoded calcium indicators. Methods taught included multiphoton and light-sheet microscopy and combination of imaging with optogenetics. Lectures by leading experts progressed through basic concepts to presentation of cutting-edge methods. Students learned the fundamentals of optics and microscopy, as well as the use of different types of cameras, laserscanning systems, in vivo preparations, and image processing and analysis software. A strong emphasis was placed on building exercises that allowed students to develop an understanding of basic principles, while also introducing them to a variety of state-of-the-art commercial systems.

## PARTICIPANTS

Athalye, V., Ph.D., Columbia University, New York, New York
Erskine, A., Ph.D., University of Southern California, Los Angeles
Fekir, S., B.S., Brown University, Providence, Rhode Island
Gauthier, J., Ph.D., Princeton University, New Jersey
Hansen, E., Ph.D., INSERM U1024, Paris, France
Namboodiri, V.M.K., Ph.D., University of Washington, Seattle
Kaelberer, M., Ph.D., Duke University, Durham, North Carolina
Li, S., B.S., University of Pennsylvania, Philadelphia
Salinas, A., Ph.D., National Institute on Alcohol Abuse and Alcoholism, Bethesda, Maryland

Schroeder, A., Ph.D., Max Planck Institute for Brain Research, Frankfurt, Germany
Schuman, B., B.S., New York University School of Medicine, New York
Siciliano, C., Ph.D., Massachusetts Institute of Technology, Cambridge
Son, S., Ph.D., Penn State University, Hershey, Pennsylvania
Tombaz, T., Ph.D., Norwegian University of Science and Technology, Norway
Wang, Y., B.S., Emory University, Atlanta, Georgia

## SEMINARS

Campbell, R., University of Tokyo, Bunkyo-ku, Japan: Genetically encoded indicators for imaging of neural activity.
Denk, W., Max Planck Institute of Neurobiology, Martinsried, Germany: Block face EM/connectomics.
Donovan, J., Max Planck Institute of Neurobiology, Martinsried, Germany: Principles of 2P microscopy.
Emiliani, V., Vision Institut, Paris, France: Wave front shaping for optical brain stimulation.
Haas, K., University of British Columbia, Vancouver, Canada: Electroportative transfection for in vivo imaging of neuronal growth and activity.
Hillman, E., Columbia University, New York, New York: Light sheet imaging and SCAPE microscopy.
Icaruso, M.F., The Francis Crick Institute, London, United Kingdom: Functional organization of synaptic connectivity in primary visual cortex.
Ji, N., University of California, Berkeley: Adaptive optical microscopy. Super-resolution microscopy.
Lichtman, J., Harvard University, Cambridge, Massachusetts: Part 1: Principles and practice of confocal microscopy. Part 2: Applications of confocal and other imaging approaches to connectomics.
Lin, M., Stanford University, California: The need for speed: voltage imaging in the nervous system.
Mace, E., Institut de la Vision, INSERM/CNRS/UPMC, Paris, France: Whole-brain functional ultrasound imaging (fUSi).

Mertz, J., Boston University, Massachusetts: Phase contrast.
Peterka, D., Columbia University, New York, New York: Fluorescence, Jablonski diagrams, filters.
Podgorski, K., Howard Hughes Medical Institute, Ashburn, Virginia: Computational microscopes for in vivo imaging.
Portugues, R., Max Planck Institute of Neurobiology, Martinsreid, Germany: Lightsheet and lightfield.
Stringer, C., HHMI/Janelia research Campus, Ashburn, Virginia: Computational processing of two-photon calcium imaging data.
Tsai, P., University of California, San Diego: Incoherent emission point source, PSF/Airy pattern, aberrations.
Ventalon, C., European Neuroscience Institute at Paris Descartes University, France: Functional fluorescence imaging and targeted photoactivation in freely behaving rodents.
Waters, J., Allen Institute for Brain Science, Seattle, Washington: Calcium indicators.
Waters, J., Harvard Medical School, Boston, Massachusetts: Noise and detectors.
Witkowski, J., Cold Spring Harbor Laboratory: Ethics, rigor, and reproducibility.
Xu, C., Cornell University, Ithaca, New York: Imaging deeper and faster using multiphoton.

# Yeast Genetics and Genomics

**July 23–August 12**

INSTRUCTORS       **G. Brown,** University of Toronto, Ontario, Canada
**G. Lang,** Lehigh University, Bethlehem, Pennsylvania
**E. Ünal,** University of California, Berkeley

ASSISTANTS       **J. Goodman,** University of California, Berkeley
**X. Saayman,** Oxford University, United Kingdom
**R. Vignogna,** Lehigh University, Coopersburg, Pennsylvania

This course is a modern and intensive laboratory course that teaches students the full repertoire of genetic and genomic approaches needed to dissect complex problems using the yeast *Saccharomyces cerevisiae.* Both classical and modern approaches were emphasized, including the isolation and characterization of mutants, tetrad analysis, and complementation. Synthetic biology was explored through CRISPR-Cas9-directed engineering of heterologous biosynthetic pathways in yeast. Students learned genome-based methods of analysis facilitated by the *Saccharomyces* Genome Database, yeast genome sequences, the gene deletion collection, and other genomic resources available to the community. Molecular genetic techniques, including yeast transformation, gene replacement by PCR, construction and analysis of gene fusions, and generation of mutations, were also emphasized.

Students combined classical approaches with whole-genome sequencing to gain experience in identifying and interpreting genetic interactions, including suppression and synthetic lethality.

Students performed genome-scale screens using the synthetic genetic array (SGA) methodology. Students were immersed in yeast genomics and performed and interpreted experiments using colony arrays and whole-genome sequencing. Computational methods for data analysis were introduced. Students gained first-hand experience in modern cytological approaches such as epitope tagging and imaging yeast cells using fluorescence microscopy with GFP–protein fusions and fluorescent indicators for different subcellular structures and organelles. Lectures on fundamental aspects of yeast genetics and genomics were presented along with seminars given by prominent experts in the field on topics of current interest.

## PARTICIPANTS

Duvernoy, M-C., Ph.D., University of California, Berkeley

Eerlings, R., B.S., KU Leuven, Belgium

Goike, J., M.Sc., University of Texas, Austin

Hunt-Isaak, I., B.A., Harvard University, Cambridge, Massachusetts

Ivanova, E., M.S., Institute of Molecular Biology, Mainz, Germany

Kerscher, O., Ph.D., The College of William & Mary, Williamsburg, Virginia

Lee, C., Ph.D., Fred Hutchinson Cancer Research Center, Seattle, Washington

Leynaud-Kieffer, L., M.S., Lawrence Berkeley National Laboratory, Emeryville, California

Miller, A., B.S., Cornell University, Ithaca, New York

Monen, J., Ph.D., Ramapo College of New Jersey, Mahwah

Musaev, D., B.Sc., Yale University, New Haven, Connecticut

Ó Cinnéide, E., B.Sc., University College, Dublin, Ireland

Phillips, M., Ph.D., Oregon State University, Corvallis

Reith, P., M.S., Chalmers University of Technology, Gothenburg, Sweden

Roussou, R., M.S., Graduate School of Quantitative Biosciences, Munich, Germany

Scopel Ferreira da Costa, E., M.S., University of Georgia, Athens

## SEMINARS

Berchowitz, L., Columbia University Medical Center, New York, New York: Meiosis-specific translational repression of retrotransposons.

Berman, J., Tel Aviv University, Israel: Drug responses in pathogenic yeasts.

Bochman, M., Indiana University, Bloomington: The Hrq1 helicase stimulates Pso2 nuclease activity during DNA interstrand cross-link repair.

Cobb, J., University of Calgary, Alberta, Canada: Early events in DNA double-strand break repair pathway choice.

Denic, V., Harvard University, Cambridge, Massachusetts: Defining the genetic landscape of mammalian autophagy.

Dunham, M., University of Washington, Seattle: Using yeast and deep mutational scanning to predict the consequences of human genetic variation.

Hoyt, M., Impossible Foods, Redwood City, California: Soybean leghemoglobin production using *Pichia pastoris.*

Keeney, S., Memorial Sloan Kettering Cancer Center/ HHMI, New York, New York: How do little chromosomes know how big they are?

Lackner, L., Northwestern University, Evanston, Illinois: Shaping the dynamic mitochondrial network.

Magwene, P., Duke University, Durham, North Carolina: Mapping QTL across time and environments: function valued traits in yeast.

Nash, R., Stanford University, Palo Alto, California: Navigating data at SGD with YeastMine.

Sadhu, M., National Human Genome Research Institute, Bethesda, Maryland: Studying molecular biology and evolution with high-throughput genome engineering.

Tyler, J., Weill Cornell Medical College, New York, New York: Chromatin assembly and disassembly dynamics.

Zalckvar, E., Weizmann Institute of Science, Rehovot, Israel: Peroxisystem: harnessing high-content screens to study peroxisome biology.

# Workshop on Autism Spectrum Disorders

**July 29–August 4**

INSTRUCTORS     **J. McPartland,** Yale University, New Haven, Connecticut
                          **S. Pasca,** Stanford University, California
                          **J. Veenstra-Vander Weele,** Columbia University, New York, New York

ASSISTANT       **A. Julta,** Columbia University/New York State Psychiatric Institute, New York, New York

Autism spectrum disorders (ASDs) are developmental disorders with complex phenotypes defined by a triad of symptoms that include disrupted social abilities, verbal and nonverbal communication skills, and restricted interests with repetitive behaviors. Co-occurring neurological and medical conditions often occur in this disorder. The underlying etiology remains a mystery, but ASD is one of the most highly heritable of neuropsychiatric disorders.

This workshop examined dimensions of ASD on various levels, including sessions on characteristics of the clinical syndrome, the neuropathology, imaging, and cognitive neuroscience studies that implicate circuits and systems involved in ASD, the current state of findings from human genetics, concepts regarding the developmental neurobiological basis, the use of experimental models, and current etiological theories and hypotheses of ASD.

In addition to learning about the most recent research in these areas, the course explored and debated controversial topics and challenges of basic assumptions in the field. An exceptional faculty with diverse interests brought the most up-to-date results and theories to the students, making this workshop a valuable resource for young researchers starting out in this fast-moving

and expansive field. Not only did it help them build the foundation for their future research, it also introduced them to many potential collaborators working to understand ASD from different disciplines.

This workshop was supported by the Nancy Lurie Marks Family Foundation.

## PARTICIPANTS

Ahmed, N., B.S., UT Southwestern Medical Center, Dallas

Awad, P., B.S., Boston Children's Hospital, Massachusetts

Balaan, C., B.A., University of Hawaii, Manoa

Behesti, H., B.Sc., The Rockefeller University, New York, New York

Blok, L., B.Sc., Donders Institute, Radboudumc, the Netherlands

Burket, J., B.S., Eastern Virginia Medical School, Norfolk

Cohen, A., B.A., Boston Children's Hospital, Massachusetts

Cristian, F-B., B.S., Institute of Human Genetics Heidelberg, Germany

Davies, J., B.Sc., University of Exeter, United Kingdom

Dias, C., B.A., Boston Children's Hospital, Massachusetts

Eisenberg, C., B.A., Rutgers University, New Brunswick, New Jersey

Fagbayi, Y., B.Sc., University of Lagos, Nigeria

Godoy, P., B.A., Universidade de Sao Paulo (USP), Brazil

Hu, R., B.A., Shanghai Institutes for Biological Sciences, China

Itskovich, E., B.Sc., Stanford University, California

Kolodny, T., B.A., University of Washington, Seattle

Kundu, S., B.S., Stanford University, California

Mazon Cabrera, R., B.S., Hasselt University, Belgium

McDiarmid, T., B.Sc., University of British Columbia, Vancouver, Canada

Ortiz, A., B.A., UT Southwestern, Dallas

Su, J., B.S., Donald and Barbara Zucker School of Medicine, New York, New York

Xia, Q., B.S., University of Alabama, Birmingham

Xu, J., B.S., California Institute of Technology, Pasadena

## SEMINARS

Allen, N., Salk Institute for Biological Studies, La Jolla, California: Glia biology in ASD.

Champagne, F., Columbia University, New York, New York: Epigenetics in ASD.

Fallin, M.D., Johns Hopkins University, Baltimore, Maryland: Epidemiology of ASD. Environmental risk factors.

Feng, G., Massachusetts Institute of Technology, Cambridge: Synaptic function in ASD.

Geschwind, D., University of California, Los Angeles: Introduction to human genetics and genetic findings in ASD, Part 1. Genetic findings in ASD, Part 2.

Kasari, C., University of California, Los Angeles: Outcome measures in ASD. Outcome measures in ASD behavioral treatments in ASD.

McPartland, J., Yale University, New Haven, Connecticut: Electrophysiology methods and findings in ASD.

Murphy, D., King's College London, United Kingdom: Neuroimaging methods, Part 1. Neuroimaging methods and findings, Part 2.

Pasca, S., Stanford University, California: Human cellular models of disease (iPS cells and organoids).

Platt, M., University of Pennsylvania, Philadelphia: Primate models of ASD.

Powell, C., UT Southwestern, Dallas: Introduction to mouse behavior.

Sahin, M., Harvard University, Cambridge, Massachusetts: Tuberous sclerosis and related syndromes associated with ASD.

Schumann, C., University of California/Davis MIND Institute, Sacramento: Neuropathology challenges and findings.

Sestan, N., Yale University, New Haven, Connecticut: Introduction neural development in ASD, Part 1. Neurodevelopment in ASD, Part 2.

Shatz, C., Stanford University, California: Critical periods/ plasticity mechanisms.

Singer, A., President of Autism Science Foundation, New York, New York: Parent/sibling perspective.

Spence, S., Boston Children's Hospital, Massachusetts: Neurological/medical perspective in ASD. Discussion of clinical assessments in ASD.

Tager-Flusber, H., Boston University Medical School, Massachusetts: Psychological models of ASD.

Veenstra-Vanderweele, J., Columbia University, New York, New York: Pharmacology: bench to bedside.

Volkmar, F., Yale University Medical School, New Haven, Connecticut: Historical perspective and clinical presentation in ASD.

# Neuroscience of Addiction

INSTRUCTORS      **D. Belin,** University of Cambridge, United Kingdom
**C. Evans,** University of California, Los Angeles
**M. Von Zastrow,** University of California, San Francisco

Drug addiction is the most costly neuropsychiatric disorder faced by our nation. Acute and repeated exposure to drugs produces neuroadaption and long-term memory of the experience, but the cellular and molecular processes involved are only partially understood.

The primary objective of this workshop was to provide an intense dialog of the fundamentals, state-of-the-art advances, and major gaps in the cell and molecular biology of drug addiction. Targeted to new or experienced investigators, the workshop combined formal presentations and informal discussions to convey the merits and excitement of cellular and molecular approaches to drug addiction research. With the advent of genomics and proteomics, an extraordinary opportunity now exists to develop comprehensive models of neuroadaptive processes fundamental to addiction, withdrawal, craving, and relapse to drug use and to brain function.

A range of disciplines and topics were represented, including:

- Noninvasive brain imaging to identify drug targets and adaptive processes.
- Neuroadaptive processes at the molecular and cellular level.
- Neural networks and their modulation.
- Relevance of genotype to susceptibility and drug response.
- Tolerance and adaptation at the cellular level.
- Approaches to exploiting the daunting volume generated by neuroinformatics.

This workshop provided an integrated view of current and novel research on neuroadaptive responses to addiction, fostered discussion on collaboration and integration, provided critical information needed to construct a model of addiction as a disease, and novel molecular targets for biological treatments. Beyond the plane of scientific endeavor, the information is vital for formulating public policy and for enlightening the public on the neurobiological consequences of drug use and addiction.

The workshop was designed to generate interest in this level of analysis, open conduits for collaborations and present novel routes to investigating the neurobiology of addictive drugs.

## PARTICIPANTS

Armenta-Resendiz, M., B.S., Medical University of South Carolina, Charleston

Barbee, B., B.S., Emory University, Atlanta, Georgia

Gong, S., B.S., University of Colorado, Anschutz Medical Campus, Aurora

Guillaumin, A., B.S., Uppsala University, Sweden

Hodebourg, R., B.S., Medical University of South Carolina, Charleston

Jimenez Chavez, C., B.A., University of California, Santa Barbara

Jones, J., B.A., Medical University of South Carolina, Charleston

Kruyer, A., B.S., Medical University of South Carolina, Charleston

Lewitus, V., B.S., George Mason University, Fairfax, Virginia

Martin, A., B.S., Indiana University, Bloomington

Marti-Prats, L., B.S., University of Cambridge, United Kingdom

Minnig, M., B.S., Boston University School of Medicine, Massachusetts

Muthusamy, A., B.A., Caltech, California

Nal, R., B.S., Medical University of South Carolina, Charleston

Nalberczak-Skóra, M., B.A., Nencki Institute of Experimental Biology, Poland

Nentwig, T., B.A., Medical University of South Carolina, Charleston

Okhuarobo, A., B.S., Scripps Research, La Jolla, California

Ornelas, L., B.A., Bowles Center for Alcohol Studies, North Carolina

Palombo, P., B.S., Federal University of São Paulo (UNIFESP), Brazil

Pérez-Cardona, E., B.S., University of Puerto Rico, Medical Sciences Campus

Przybysz, K., B.A., Binghamton University, New York

Sambo, D., B.S., National Institute on Alcohol Abuse and Alcoholism, Baltimore, Maryland

Sequeira, M., B.S., Emory University, Atlanta, Georgia

## SEMINARS

Alvarez, V., National Institutes of Health, Bethesda, Maryland: Synaptic and circuit mechanisms underlying a well-known vulnerability for addiction. Actions of drugs of abuse on striatal dopamine: dogma and beyond.

Belin, D., University of Cambridge, United Kingdom: Neural substrates of the inter-individual vulnerability to develop compulsive drug seeking habits.

Cheer, J., University of Maryland School of Medicine, Baltimore: Endogenous cannabinoids and the pursuit of reward.

Evans, C., University of California, Los Angeles: Opioids in pain and addiction.

Goldman, D., National Institute on Alcohol Abuse and Alcoholism/LNG, Rockville, Maryland: Addictions: oligogenic, omnigenic, or somewhere between?

Hurd, Y., Icahn School of Medicine at Mount Sinai, New York: Human brain stuff.

Kalivas, P., Medical University of South Carolina, Charleston: Using the neurobiology of will power to cure addiction.

Kenny, P., Icahn School of Medicine at Mount Sinai, New York: The habenula links nicotine addiction to tobacco-related diseases.

Kober, H., Yale University, New Haven, Connecticut: Human neuroscience of addiction.

Koob, G., National Institute on Alcohol Abuse and Alcoholism, Rockville, Maryland: Hedonic allostasis/stress.

Manglik, A., University of California, San Francisco: Mechanistic basis of addictive drug action.

Marisela Morales, F., National Institute on Drug Abuse (NIDA), Baltimore, Maryland: Neuronal diversity and cotransmission.

Nestler, E., Icahn School of Medicine at Mount Sinai, New York: Transcriptional and epigenetic mechanisms of drug addiction.

Otis, J., Medical University of South Carolina, Charleston: Innovative technologies for understanding the neural circuitry of addictive behaviors.

Picciotto, M., Yale University, Guilford, Connecticut: Molecular basis of nicotine addiction.

von Zastrow, M., University of California, San Francisco: Drug actions viewed from a lonely neuron's perspective.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, reproducibility, and rigor.

# Proteomics

INSTRUCTORS  **R. Chalkley,** University of California, San Francisco
**G. Knudsen,** Alaunus Biosciences, San Francisco, California
**D. Pappin,** Cold Spring Harbor Laboratory

PART-TIME INSTRUCTORS  **E. Soderblom,** Duke Center for Genomic and Computational Biology, Apex, North Carolina
**J. Thompson,** Duke Center for Genomic and Computational Biology, Apex, North Carolina

ASSISTANTS  **Z. Darula,** Biological Research Centre of the Hungarian, Szeged, Hungary
**J. Maynard,** University of California, San Francisco
**E. Soderblom,** Duke University, Durham, North Carolina

This intensive laboratory and lecture course focused on cutting-edge proteomic approaches and technologies. Students gained practical experience isolating, purifying, and identifying protein complexes: Sample preparation with in-solution digestion was performed, and then the students were trained using high-sensitivity nano HPLC coupled with nanospray-ESI and tandem mass spectrometry analysis. Different search engines and bioinformatic approaches were introduced for data evaluation. Students were shown how to recognize unexpected posttranslational modifications.

Diverse techniques for PTM peptide enrichment, including affinity chromatography for phosphopeptides and immunoenrichment of GlyGly-Lys, methyl-Arg, and phospho-Tyr peptides, and the characterization of the resulting complex mixtures, including site assignments, were performed. For shotgun proteomic analysis sections, students used label-free and covalent isotopic-labeling quantitative approaches to profile changes in protein complexes and whole proteomes. In a section focused on targeted proteomics, students learned to analyze and process shotgun proteomic data for the development of SRM/PRM assays that accurately identify and quantify targeted proteins. Students designed transitions for selected peptides and performed SRM/PRM analyses. They learned to process and interpret the acquired data to measure changing quantities of targeted peptides in a variety of biological samples, and specifically spent significant time using Skyline for both MS1 and MS2 data analysis. For all sections of the course, a strong emphasis was placed on data analysis. There was opportunity to discuss and provide feedback on individual research projects, and students had the opportunity to learn to process their own data (acquired outside the course) in Skyline if so desired.

An industrial lecture series was delivered by drug discovery scientists. The students received in-depth knowledge about chemoproteomics techniques routinely used in industry, and how they are used to profile compounds and potential protein targets. This was followed by laboratory work on in-lysate affinity enrichment techniques and a deep dive into data analysis. This broadened students' vision toward chemoproteomics application in drug discovery programs.

A series of outside lecturers discussed various proteomics topics including: de novo sequence analysis, intact protein analysis, advanced mass spectrometry methods, glycosylation, and functional proteomics. The aim of the course was to provide each student with the fundamental knowledge and hands-on experience necessary for performing and analyzing proteomic experiments. The overall goal was to train students to identify new opportunities and applications for proteomic approaches in their biological research.

This course was supported with funds provided by the National Institute of Child Health & Human Development, and partial scholarship support was provided by the Regeneron Scholars Fund.

## PARTICIPANTS

Akaniro-Ejim, N., M.Sc., University of Nottingham, United Kingdom

Albin, J., M.D./Ph.D., Massachusetts General Hospital, Boston

Almada, A., Ph.D., Harvard University, Cambridge, Massachusetts

Chan, A., B.S., MIT/Whitehead Institute for Biological Research, Cambridge, Massachusetts

Cheng, R., M.S., Stanford University, California

Kundinger, S., B.S., Emory University, Atlanta, Georgia

Lee, C., Ph.D., University of Texas, Austin

Lee, J., A.L.M., Mayo Clinic, Jacksonville, Florida

Moen, J., B.S., Yale University, West Haven, Connecticut

Rai, S., Ph.D., Massachusetts General Hospital, Charlestown

Sasaki, S., Ph.D., UFABC, Campo, Brazil

Sinha, N., Ph.D., Johns Hopkins Medical Institute, Baltimore, Maryland

Stebbings, K., Ph.D. MD Anderson Cancer Center, Houston, Texas

Trinidad, C., B.S., University of Kansas Medical Center, Lawrence

Zhao, L., Ph.D., Central Michigan University, Mount Pleasant

## SEMINARS

Anderson, L., NHMFL/Florida State University, Tallahassee: High-throughput intact protein characterization.

Clauser, K., Broad Institute of MIT and Harvard, Cambridge, Massachusetts: Manual de novo peptide MS/MS interpretation for evaluating database search results.

Darula, Z., Biological Research Centre of the Hungarian Academy of Sciences, Hungary: Phosphopeptide enrichment using metal-ion affinity chromatography.

Farnsworth, C., Cell Signaling Technology, Concord, Massachusetts: PTMScan® technology: an antibody-based proteomics discovery platform.

Hendricks, A., Kawatkar, A., Pachl, F., and Tomlinson, R., AstraZeneca, Waltham, Massachusetts: Chemical proteomics. Chemical biology and proteomics in drug discovery.

Searle, B., Institute for Systems Biology, Seattle, Washington: Data independent acquisition.

Thompson, W., Duke Center for Genomic and Computational Biology, Apex, North Carolina: Label-free Quant, AUC vs. spectral counting, tools for data interpretation.

Trnka, M., University of California, San Francisco: Cross-linking mass spectrometry for exploring the structure and interactions of protein complexes.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, rigor, and reproducibility.

# Antibody Engineering, Phage Display and Immune Repertoire Analysis

**October 11–22**

INSTRUCTOR      **G. Silverman,** New York University School of Medicine, New York

CO-INSTRUCTORS   **C. Rader,** Scripps Research Institute, Jupiter, Florida
                            **G. Veggiani,** University of Toronto, Ontario, Canada

Advances in the generation and selection of antibodies from combinatorial libraries allow for the rapid production of antibodies from immune and nonimmune sources. This intensive laboratory/ lecture course focused on the construction of combinatorial antibody libraries expressed on the surface of filamentous phage and the subsequent selection of desired antibodies from the library. Students learned the theoretical and practical aspects of constructing combinatorial libraries from immune and nonimmune sources, as well as the construction of synthetic antibody libraries. Antigen-specific recombinant monoclonal antibodies were selected from the library. Production, purification, and characterization of antibody fragments expressed in *Escherichia coli* was covered.

The lecture series, presented by course faculty and a number of invited speakers, emphasized theory and practice of antibody display technologies, expression of antibodies in *E. coli* and mammalian cells, antibody structure and function, bacterial display of antibodies and other ligand-binding domains, the immunobiology of the antibody response, and the use of monoclonal antibodies for therapy including the design of chimeric antigen receptor T cells. Principles and protocols for generation and analysis of immune repertoires determined by next-generation sequencing were discussed.

This course was supported with funds provided by the National Institute of General Medical Sciences, and partial scholarship support was provided by the Regeneron Scholars Fund.

## PARTICIPANTS

Antony, S., M.S., The Jackson Laboratory, Bar Harbor, Maine

Dahlsson Leitao, C., M.S., KTH Royal Institute of Technology, Stockholm, Sweden

Dongyan, Z., M.S., The University of Hong Kong, Pokfulam

Florez, C., B.S., United States Military Academy, West Point, New York

Hjelm, L., M.S., KTH Royal Institute of Technology, Stockholm, Sweden

Karlander, M., M.S., KTH Royal Institute of Technology, Stockholm, Sweden

Le Guezennec, X., Ph.D., Institute of Molecular and Cell Biology, Singapore

Llauger Iriarte, G., Ph.D., National Institute of Agricultural Technology, Hurlingham, Argentina

Mack, T., M.S., Boehringer Ingelheim Pharmaceutical, Inc., Ridgefield, Connecticut

Reddy, R., M.S., University of Alberta, Edmonton, Canada

Sachidanandam, R., Ph.D., Mount Sinai School of Medicine, New York

Saud, Z., Ph.D., Avacta Life Sciences Ltd., Cambridge, United Kingdom

Scheffel, J., M.S., KTH Royal Institute of Technology, Stockholm, Sweden

Yao, M., Ph.D., Cold Spring Harbor Laboratory

Zhao, L., Ph.D., Harvard Medical School, Boston, Massachusetts

## SEMINARS

Dekosky, B., Kansas University, Lawrence: Interpreting native antibody function on a repertoire scale.

Dreier, B., University of Zürich, Switzerland: In vitro evolution of proteins by ribosome display.

Koide, S., New York University, New York: Design of exquisite specificity in synthetic binding proteins.

Pohl, M.A., Tri-Institutional Therapeutics Discovery Institute, New York, New York: Antibody discovery at the Tri-D TDI.

Rader, C., Scripps Research Institute, Jupiter, Florida: From phage display to cancer immunotherapy.

Sidhu, S., University of Toronto, Ontario, Canada: From systems biology to systems biologics.

Siegel, D., University of Pennsylvania, Philadelphia: Phage display of autoantibodies and immune pathogenesis.

Silverman, G., New York University School of Medicine, New York: Overview of phage display for antibody and epitope selection.

Stahl, S., KTH Royal Swedish Institute, Stockholm, Sweden: Generating affibody molecules for medical applications.

# Workshop on Cereal Genomics

**October 15–21**

INSTRUCTORS       **S. Hake,** USDA/University of California Berkeley Plant Gene Expression Center
                               **D. Jackson,** Cold Spring Harbor Laboratory
                               **D. Ware,** USDA/Cold Spring Harbor Laboratory

This one-week workshop enabled participants to take advantage of emerging genetic tools and the completed cereal genome sequences of most of the major cereal crops. The workshop featured morning and evening lectures with afternoon lab exercises, including hands-on laboratory work and computer sessions in comparative anatomy, GWAS, next-generation expression analysis, whole-genome sequencing assembly, emerging model systems, genome editing, and phenomics. The faculty (instructors and invited lecturers) were/are active researchers in cereal genetics and genomics who have made significant contributions to the field, ensuring that the latest techniques and ideas were presented. The course was structured to provide time for informal discussions and exchange with instructors.

Topics included:

- Comparative anatomy and phylogeny.
- Cereal genomes, assembly, annotation, and synteny.
- Genetics and databases.
- Quantitative trait locus mapping and genome-wide association studies.
- Genome wide expression analyses.
- Reverse genetics and genome editing.
- Phenomics.

## PARTICIPANTS

Cantos, C., M.S., Penn State University, University Park, Pennsylvania

Choi, J., Ph.D., University of Cambridge, United Kingdom

Giri, A., Ph.D., Cornell University, Ithaca, New York

Haider, Y., B.S., Arcadia Biosciences, Davis, California

Iohannes, S., B.S., Sant'Anna School of Advanced Studies, Pisa, Italy

Iwuala, E., B.S., M.S., University of Lagos, Nigeria

Kenchanmane Raju, S.K., Ph.D., Michigan State University, East Lansing

Lee, H-S., Ph.D., Department of Agronomy, Daejeon, South Korea

Meehan, C., Ph.D., University of Warwick, Coventry, United Kingdom

Obi, Q., M.S., International Institute of Tropical Agriculture, Ibadan, Nigeria

Pardo, J., B.S., Michigan State University, East Lansing

Poudel, P., M.S., Oklahoma State University, Stillwater

Provencher, C., M.S., Inari Agriculture, Cambridge, Massachusetts

Raquid, R., M.S., International Rice Research Institute, Los Baños, Laguna, Philippines

Scheben, A.P., Ph.D., Cold Spring Harbor Laboratory

Suarez, S., M.S., Oklahoma State University, Stillwater

## SEMINARS

Buckler, E., USDA-ARS, Ithaca, New York: Nucleotide-level natural variation and genetic architecture. QTL mapping and prediction. Machine learning the central dogma of molecular biology. Learning from evolution and molecular biology for a new agriculture.

Devos, K., University of Georgia, Augusta: Genetic analysis of nonmodel plants: challenges and successes.

Dubcovsky, J., University of California, Davis: Resources for wheat functional genomics.

Eveland, A., Donald Danforth Plant Science Center, St. Louis, Missouri: Extracting biological insights through multiomics data integration.

Kellogg, E., Donald Danforth Plant Science Center, St. Louis, Missouri: Introduction to the grass family.

Leiboff, S., University of California, Berkeley: Hands-on RNA-Seq: from maize sequence to developmental inference.

O'Connor, D., Pairwise Plants, Research Triangle Park, North Carolina: Enabling comparative biology with comparative genomics tools.

Paszkowski, U., University of Cambridge, United Kingdom: Molecular genomics of arbuscular mycorrhizal symbiosis in cereals.

Topp, C., Donald Danforth Plant Science Center, St. Louis, Missouri: Plant phenomics.

Yang, B., University of Missouri, Columbia: Reverse genetics and genome editing.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, rigor, and reproducibility.

# Programming for Biology

INSTRUCTORS　　**S. Prochnik,** Intrexon, Inc., San Francisco, California
　　　　　　　　**S. Robb,** Stowers Institute for Medical Research, Kansas City, Missouri

ASSISTANTS　　　**J. Bredeson,** University of California, Berkeley
　　　　　　　　**D. Diaz,** Stowers Institute for Medical Research, Kansas City, Missouri
　　　　　　　　**K. Gotting,** The University of Wisconsin, Madison
　　　　　　　　**J. Orkin,** Universitat Pompeu Fabra, Barcelona, Spain
　　　　　　　　**M. Waas,** University of Nebraska Medical Center, Omaha
　　　　　　　　**S. Webb,** University of Calgary, Alberta, Canada

More often than not, today's biologist is studying data that are too complex or numerous to be analyzed without a computer, and only boilerplate analysis can be performed with existing tools. Questions specific to the data set require novel analysis pipelines to be designed and written in computer code. Designed for lab biologists with little or no programming experience, this course gave students the bioinformatics and scripting skills necessary to exploit this abundance of biological data. The only prerequisite for the course was a strong commitment to learning basic UNIX and a scripting language.

This year, we offered the course in Python, an easy-to-learn scripting language with a growing code base and community of users. The course began with one week of introductory coding, continued with practical topics in bioinformatics, with plenty of coding examples, and ended with a group coding project. Formal instruction was provided on every topic by the instructors, teaching assistants, and invited experts. Students solved problem sets covering common scenarios in the acquisition, validation, analysis, and visualization of biological data. They learned how to

design, construct, and run powerful and extensible analysis pipelines in a straightforward manner. Final group projects were chosen from ideas proposed by students and were guided by faculty. Students were provided with a library of Python reference print and e-books that they were able to bring home with them.

The primary focus of this course was to provide students with practical programming experience, rather than to present a detailed description of the algorithms used in computational biology.

## PARTICIPANTS

Barrows, D., Ph.D., The Rockefeller University, New York

Cardinault, M.A., Yucatan Center for Scientific Research (CICY), Conkal, Mexico

Cirillo, L., Ph.D., Institute for Cancer Research, London, United Kingdom

Crespo, E., B.A., Central Michigan University, Mount Pleasant

Donmez, O., B.S., Cincinnati Children's Hospital, Ohio

Fisher, N., B.S., Vanderbilt University, Nashville, Tennessee

Geck, R., B.S., Beth Israel Deaconess Medical Center, Boston, Massachusetts

Harding, K., M.Sc., Intrexon, South San Francisco, California

Hennessey, R., Ph.D., National Cancer Institute, Gaithersburg, Maryland

Janiak, M., Ph.D., University of Calgary, Alberta, Canada

Kimura, J., B.S., Harvard University, Cambridge, Massachusetts

Kojima, M., Ph.D., Yale University School of Medicine, New Haven, Connecticut

Konkel, Z., B.S., Ohio State University, Columbus

Kozinova, M., M.D., Fox Chase Cancer Center, Philadelphia, Pennsylvania

Lance-Byrne, A., B.A., University of California, Santa Cruz

Mann, M., B.S., Cornell University, Ithaca, New York

Massilani, D., Ph.D., Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Mavian, C., Ph.D., University of Florida, Gainesville

Saraceno, C., B.S., University of Kentucky, Lexington

Scott, K., M.S., Ohio State University, Columbus

Tomlinson, B., B.S., University of South Florida, Tampa

Underhill, H., Ph.D., University of Utah, Salt Lake City

Werner, J., Ph.D., Wisconsin Lutheran College, Milwaukee

## SEMINARS

Bredeson, J., University of California, Berkeley: Next-generation sequencing: file types and analysis.

Cain, S., Ontario Institute for Cancer Research, Toronto, Ontario, Canada: Managing genomic data: GMOD.

Dobin, A., Cold Spring Harbor Laboratory: Aligners.

Haas, B., Broad Institute, Cambridge, Massachusetts: RNA-seq and transcript assembly.

Marques-Bonet, T., Universitat Pompeu Fabra, Barcelona, Spain: Genome variation.

Mi, H., University of Southern California, Los Angeles: Gene function annotation and gene set analysis.

Pearson, W., University of Virginia, Charlottesville: Sequence similarity Part 1: homology and alignments. Sequence similarity Part 2: protein domains and distant relatives.

Perera, A., Stowers Institute for Medical Research, Kansas City, Missouri: Next-generation sequencing: technologies.

Prochnik, S., Joint Genome Institute, Walnut Creek, California: Unix 1: Unix overview, the basics, advanced Unix, Unix cheat sheet. Python I: overview, running Python, syntax, data types, and variables. Python III: sequences, strings. Python VI: I/O and files. Python IX: exceptions. Python X: functions. Python XI: BioPython. Pandas and dataframes.

Prochnik, S., Joint Genome Institute, Walnut Creek, California; Robb, S., Stowers Institute for Medical Research, Kansas City, Missouri: Ethics and reproducibility in research.

Robb, S., Stowers Institute for Medical Research, Kansas City, Missouri: Unix 2: text editors, Git for beginners. Python II: operators, truth, logic, numbers. Python IV: lists, tuples, loops. Python V: dictionaries, sets. Python VII: regular expressions. Python VIII: data structures.

Triant, D., University of Virginia, Charlottesville: Genome assembly, Parts 1 and 2.

# X-Ray Methods in Structural Biology

INSTRUCTORS     **P. Adams,** Lawrence Berkeley Laboratory, California
**J. Newman,** CSIRO, Parkville, Victoria, Australia
**A. Perrakis,** Netherlands Cancer Institute, Amsterdam
**J. Pflugrath,** Rigaku Americas Texas, The Woodlands

ASSISTANT     **C. Fan,** California Institute of Technology, Pasadena
**T. Heidebrecht**, Netherlands Cancer Institute, Amsterdam
**D. Liebschner**, Lawrence Berkeley National Laboratory, California
**T. Peat**, CSIRO, Parkville, Victoria, Australia

X-ray crystallography has been the cornerstone of structural biology for half a century and remains the technique of choice for atomic resolution understanding of macromolecules and for structure guided drug discovery. This intense course combined laboratory and computational instruction to train course participants in the major techniques used to determine three-dimensional structures. It was designed for scientists with a working knowledge of protein structure and function, but who were new to macromolecular crystallography or who wished to increase their in-depth knowledge of macromolecular crystallography.

Topics covered included:

- Basic diffraction theory.
- Crystallization (proteins, nucleic acids, complexes, and membrane proteins).
- Synchrotron X-ray sources and optics.
- Data collection and processing.

- Structure solution by experimental phasing methods (SAD, MAD, MIR, and others) and molecular replacement.
- Electron density maps improvement
- Model building and refinement.
- Structure validation.
- Coordinate deposition.
- Structure presentation.

Participants had extensive hands-on training in well equipped labs in how to crystallize multiple proteins and determine their crystal structures by several methods, while they learned through lectures on theory and methods. Both basic and advanced subjects were covered during lectures, which were given by leaders in the field. Informal discussions behind the techniques were frequent and students were expected to pose questions to be answered in interactive sessions.

## PARTICIPANTS

Adams, M., B.A., Cornell University, Ithaca, New York

Aljedani, S., Ph.D., Fred Hutchinson Cancer Research Center, Seattle, Washington

Beenken, A., M.D./Ph.D., Columbia University Medical Center, New York, New York

Chan, S., Ph.D., National Jewish Health, Denver, Colorado

Chopra, A., M.S., Rutgers University, New Brunswick, New Jersey

Davarinejad, H., M.S., University of Ottawa, Ontario, Canada

Kayode, O., Ph.D., National Cancer Institute, Frederick, Maryland

Kumar, H., Ph.D., National Institutes of Health, Bethesda, Maryland

Monteiro Araújo dos Santos, T., Ph.D., Harvard University, Cambridge, Massachusetts

Nguyen, H.A., B.S., Emory University, Atlanta, Georgia

Olatunji, S., Ph.D., Trinity College Dublin, Dublin, Ireland

Panneels, V., Ph.D., Paul Scherrer Institute, Switzerland

Pitt, A., B.S., National Institute of Diabetes and Digestive Kidney Diseases, Bethesda, Maryland

Rodrigues, A., Ph.D., Lawrence Berkeley National Laboratory, California

Uzuncayir, S., M.S., Lund University, Sweden

Velez, G., B.S., Stanford University, Palo Alto, California

## SEMINARS

Adams, P., Lawrence Berkeley Laboratory, California: Waves, vectors, and complex numbers. Fundamental diffraction, Bragg's law, diffraction pattern, reciprocal space. Fourier transforms, electron density. What is Phenix? Structure refinement. Low-resolution structure refinement.

Borek, D., UT Southwestern Medical Center, Dallas: Anisotropy, order–disorder and radiation damage in practice. X-ray data processing, scaling and merging.

Caffrey, M., Trinity College Dublin, Ireland: Membrane protein crystallization using lipidic systems.

Conway, J., University of Pittsburgh, Philadelphia: Introduction to cryo-EM.

Emsley, P., MRC Laboratory of Molecular Biology, Cambridge, United Kingdom: Macromolecular model building and refinement. Map interpretation and model building.

Furey, W., University of Pittsburgh, Pennsylvania: Experimental phasing theory: an overview and history.

Gilliland, G., Janssen Research & Development, LLC, Charleston, South Carolina: Maximizing crystallization success through seeding.

Hendrickson, W., Columbia University, New York, New York: MAD and SAD phasing, history, and future.

Holton, J., University of California, San Francisco: Space groups, beamline basics, radiation damage, data collection, and processing.

Kleywegt, G., European Bioinformatics Institute, Cambridge, United Kingdom: Structure validation.

McPherson, A., University of California, Irvine: What is a crystal: thought you should know. Crystallization of macromolecules: theory and growth experiments.

Newman, J., CSIRO, Parkville, Victoria, Australia: Introduction to crystallization and data collection practicals. Crystallization practice: how to make crystals. Extreme approaches to crystallization.

Noinaj, N., Purdue University, West Lafayette, Indiana: Membrane proteins: production and crystallization. Membrane protein crystallization tips and tricks.

Perrakis, A., Netherlands Cancer Institute, Amsterdam: Overview of crystallography. Symmetry, periodicity, unit cells, space groups, and lattices. Introduction to the phase problem: phasing methods and Patterson space. Basic phasing theory: isomorphous replacement and anomalous scattering, substructures, phase improvement. What is CCP4? Phasing strategies: what to do and when. Automated model building in ARP/wARP. Water molecules and side chains: decision-making. Automated crystallization, data collection, and processing. Automated model rebuilding and validation in PDB_REDO.

Pflugrath, J., Rigaku Americas Texas, The Woodlands: Introduction to crystallization and data collection practicals. Crystallization practice: how to prepare crystals for data collection. Data collection and processing. Understanding data processing and scaling statistics.

Read, R., University of Cambridge, United Kingdom: Molecular replacement.

Richardson, J., Duke University, Durham, North Carolina: Validation with MolProbity. Presentation of structures: history and perspectives.

Smith, C., Stanford University, Menlo Park, California: Serial femtosecond crystallography methods at XFELs and synchrotrons.

Terwilliger, T., Los Alamos National Laboratory, New Mexico: Automated experimental phasing. Automated model building and other things.

Thorn, A., University of Hamburg, Würzburg, Germany: Twinning in crystallography. Phasing with SHELXC/D/E.

Williams, W., Duke University, Durham, North Carolina: Validation with MolProbity.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, rigor, and reproducibility.

# Advanced Sequencing Technologies and Applications

**November 5–17**

INSTRUCTORS **M. Griffith,** Washington University School of Medicine in St. Louis, Missouri
      **O. Griffith,** Washington University School of Medicine in St. Louis, Missouri
      **E. Mardis,** Nationwide Children's Hospital Research Institute, Columbus, Ohio
      **W.R. McCombie,** Cold Spring Harbor Laboratory
      **A. Quinlan,** University of Utah, Salt Lake City

ASSISTANTS  **J. Belyeu,** University of Utah, Salt Lake City
      **E. Bogenschutz,** University of Utah, Salt Lake City
      **M. Cormier,** University of Utah, Salt Lake City
      **K. Cotto,** Washington University in St. Louis, Missouri
      **F. Gomez,** Washington University School of Medicine in St Louis, Missouri
      **S. Goodwin,** Cold Spring Harbor Laboratory
      **S. Iyer,** Cold Spring Harbor Laboratory/Stony Brook University
      **J. Preall,** Cold Spring Harbor Laboratory
      **C. Regan,** Cold Spring Harbor Laboratory
      **R. Wappel,** Cold Spring Harbor Laboratory
      **H. Xia,** Washington University in St. Louis, Missouri

Over the last decade, massively parallel DNA sequencing has markedly impacted the practice of modern biology and is being utilized in the practice of medicine. The constant improvement of these platforms means that costs and data generation timelines have been reduced by orders of magnitude, facilitating investigators to conceptualize and perform sequencing-based projects that heretofore were time-, cost-, and sample-number-prohibitive. Furthermore, the application of these technologies to answer questions previously not experimentally approachable is broadening their impact and application. However, data analysis remains a complex and often vexing challenge, especially as data volumes increase.

This intensive two-week course explored use and applications of massively parallel sequencing technologies, with a focus on data analysis and bioinformatics. Students were instructed in the detailed operation of several platforms, including library construction procedures, general data processing, and in-depth data analysis. A diverse range of the types of biological questions enabled by massively parallel sequencing technologies were explored including DNA resequencing of known cancer genes, de novo DNA sequencing and assembly of genomes, RNA sequencing, and others that were tailored to the student's research areas of interest.

Cloud-based computing was also explored. Guest lecturers highlighted unique applications of these disruptive technologies. We encouraged applicants from a diversity of scientific backgrounds, including molecular evolution, development, neuroscience, medicine, cancer, plant biology, and microbiology.

## PARTICIPANTS

Barefoot, M., B.S., Georgetown University Medical Center, Washington, D.C.
Crespo, N., B.S., University of Puerto Rico Medical Sciences Campus, San Juan.
Dadey, R., B.S., University of Pittsburgh, Pennsylvania
Everman, E., Ph.D., University of Kansas, Lawrence
Gomez-Arroyo, J., Ph.D., University of Cincinnati, Ohio
Gordovez, F.J., B.S., National Institute of Mental Health, Bethesda, Maryland
Karidas, P., Ph.D., Harvard University, Cambridge, Massachusetts
Khan, A., M.D., Memorial Sloan Kettering Cancer Center, New York, New York
Lim, T.Y., M.Sc., Columbia University, New York, New York
Linhoff, M., Ph.D., Oregon Health & Science University, Portland

O'Toole, B., B.S., Fordham University, Bronx, New York
Petersen, U.S., M.S., Research group, Odense M, Denmark
Reames, C., B.S., University of Massachusetts Medical School, Worcester
Reich, S., D.V.M., M.S., University of Missouri, Columbia,
Robles-Oteiza, C., B.S.E., Yale University, New Haven, Connecticut
Sayal, K., M.D., University of Oxford, United Kingdom
Tarui, T., M.D., Tufts Medical Center/Tufts University School of Medicine, Boston, Massachusetts
van der Pol, Y., M.Sc., Amsterdam University Medical Center, the Netherlands
Weyhrauch, D., M.D., University of Utah, Salt Lake City
Yen, E-R., M.S., MD Anderson Cancer Center, Houston, Texas

## SEMINARS

Chakravarti, A., New York University Langone, New York: Genetic regulatory control of cardiac diseases.
Hodges, E., Vanderbilt University School of Medicine, Nashville, Tennessee: Bridging DNA methylation dynamics and patient data to understand the noncoding genome.
Lareau, C., Harvard University, Brookline, Massachusetts: Single-cell genomics: scATAC-seq.
Mardis, E., Nationwide Children's Hospital, Columbus, Ohio: Overview of next-generation short read sequencing technologies.
Marth, G., University of Utah, Salt Lake City: Single-cell cancer genomics: somatic mutation detection and clonal heterogeneity.
Martienssen, R., Cold Spring Harbor Laboratory: The *Arabidopsis* retrovirome and its regulation by small RNA.
McCombie, W.R., Cold Spring Harbor Laboratory: PacBIO sequencing.

Miga, K., University of California, Santa Cruz: Complete "telomere-to-telomere" assemblies of human chromosomes.
Petti, A., Washington University School of Medicine in Saint Louis, Missouri: Single-cell RNA-sequencing data analysis.
Preall, J., Cold Spring Harbor Laboratory: Single-cell sequencing technology and applications.
Scacheri, P., Case Western Reserve University, Cleveland, Ohio: The application of epigenomic profiling strategies to study cancer and other common diseases.
Siepel, A., Cold Spring Harbor Laboratory: An overview of bioinformatics.
Smibert, P., New York Genome Center, New York: CITE seq and other topics.
Witkowski, J., Cold Spring Harbor Laboratory: Ethics in genetics and genomics research.

# Scientific Writing Retreat

**November 13–17**

INSTRUCTORS       **C. Lambert,** Cold Spring Harbor Laboratory
                             **S. Matheson,** *Cell Reports*, Cambridge, Massachusetts

WRITING COACHES   **L. Connell,** *Genes & Development,* Senior Editor, Cold Spring Harbor Laboratory
                             **S. Gary,** Cold Spring Harbor Laboratory
                             **E. Gaskell,** Broad Institute of Harvard and MIT, Somerville, Massachusetts
                             **J. Jansen,** Cold Spring Harbor Laboratory
                             **K. Kelly,** Cold Spring Harbor Laboratory
                             **F. Maderspacher,** *Current Biology,* London, United Kingdom
                             **J. Rubin,** Columbia University, New York, New York

The goal of this retreat was to have participants progress significantly on writing projects while improving their professional communication skills. The retreat included a mix of formal sessions and less-structured writing time.

    The formal sessions covered:

- Publication writing for scientific journals from the perspectives of *Cell Press* and *Cold Spring Harbor Press.*
- Writing clearly and conversationally about research in ways that engage diverse audiences, a skill particularly useful when developing lay summaries for NIH and NSF proposals.
- Style tips and considerations for clear professional writing in all forms.

The less-structured sessions of the retreat included small writing groups and dedicated individual writing time. For the small group sessions, participants were preassigned to groups of three to four people for the purpose of soliciting peer feedback on writing samples they submitted ahead of time. For the individual writing sessions, coaches were on hand to work with participants one-on-one. As with all CSHL meetings and courses, participants were required to respect the confidentiality of any unpublished research they may have read during the retreat.

This course was supported with funds provided by the NIH National Institute of General Medical Sciences.

## PARTICIPANTS

Altindis, E., Ph.D., Boston College, Chestnut Hill, Massachusetts
Balcioglu, A., Ph.D., Massachusetts Institute of Technology, Cambridge
Beaudoin, J-D., Ph.D., Yale University, New Haven, Connecticut
Brown, C., Ph.D., University of Massachusetts Medical School, Worcester
Du, J., Ph.D., West Virginia University, Morgantown
Ghose, P., Ph.D., The Rockefeller University, New York
Jin, S.C., Ph.D., The Rockefeller University, New York
King, J., Ph.D., Ursinus College, Collegeville, Pennsylvania
Lazari, C., M.S./Ph.D., University of Southern California, Los Angeles
Lopez Del Amo, V., Ph.D., University of California, San Diego, La Jolla

Lu, M., Ph.D., Yale University School of Medicine, New Haven, Connecticut
Luo, H., Ph.D., Peking University, Beijing, China
Matoo, O., Ph.D., University of Nebraska, Lincoln
Perry-Hauser, N., Ph.D., Columbia University, New York
Roberts, J., Ph.D., University of Kentucky, Lexington
Sapkota, D., Ph.D., Washington University in St. Louis, Missouri
Sugimura, R., Ph.D., Kyoto University, Japan
Tabuchi, M., Ph.D., Johns Hopkins University, Baltimore, Maryland
Tucci, S., Ph.D., Princeton University, New Jersey
van Solingen, C., Ph.D., New York University Langone Medical Center, New York

## SEMINARS

Jansen, J., Cold Spring Harbor Laboratory; Gaskell, E., Broad Institute of Harvard and MIT, Somerville, Massachusetts: Grant writing and grantsmanship.
Lambert, C., Cold Spring Harbor Laboratory; Matheson, S., *Cell Press,* Cambridge, Massachusetts: Session on top 10 tips. Lay summaries and writing for nonexpert audiences.

Matheson, S., *Cell Press,* Cambridge, Massachusetts; Connell, L., Senior Editor, Cold Spring Harbor Press; Maderspacher, F., *Current Biology,* London, United Kingdom: Publications and manuscripts.

# Computational Genomics

INSTRUCTORS    **D. Hawkins,** University of Washington, Seattle
**W. Pearson**, University of Virginia, Charlottesville
**J. Taylor**, Johns Hopkins University, Baltimore, Maryland

ASSISTANTS    **P. DeFord,** Johns Hopkins University, Baltimore, Maryland
**M. Heydarian**, Johns Hopkins University, Baltimore, Maryland
**E. Overbey**, University of Washington, Seattle
**O. Sabik**, University of Virginia, Charlottesville

This course presented a comprehensive overview of the theory and practice of computational methods for the identification and characterization of functional elements from DNA sequence data. The course focused on approaches for extracting the maximum amount of information from protein and DNA sequence similarity through sequence database searches, statistical analysis, and multiple sequence alignment.

Additional topics included:

- Alignment and analysis of "next-generation" sequencing data, with applications from metagenomic, RNA-seq, and ChIP-seq experiments.
- The Galaxy environment for high-throughput analysis.
- Regulatory element and motif identification from conserved signals in aligned and unaligned sequences.

- Integration of genetic and sequence information in biological databases.
- Genome Browsers and Genome Features.

The course combined lectures with hands-on exercises; students were encouraged to pose challenging sequence analysis problems using their own data. The course was designed for biologists seeking advanced training in biological sequence and genome analysis, computational biology core resource directors and staff and for individuals in other disciplines (e.g., computer science) who wish to survey current research problems in biological sequence analysis. Advanced programming skills were not required.

The primary focus of this course was the theory and practice of algorithms in computational biology, with the goals of using current methods more effectively for biological discovery and developing new algorithms.

This course was supported with funds provided by the National Human Genome Research Institute.

## PARTICIPANTS

Alexander, M., Ph.D., Yale University, New Haven, Connecticut
Choi, J.K., Ph.D., NEI/National Institutes of Health, Bethesda, Maryland
Collins, M., Ph.D., University of Minnesota, Minneapolis
Conte, M., Ph.D., Columbia University, New York, New York
Feliciano, P., Ph.D., Simons Foundation, New York, New York
Huang, X., Ph.D., National Institutes of Health, Bethesda, Maryland
Jin, S.C., Ph.D., The Rockefeller University, New York, New York
Kim, B., Ph.D., University of Pennsylvania, Philadelphia
Kim, W., Ph.D., University of Pennsylvania, Philadelphia
Kitavi, M., Ph.D., International Potato Center, Nairobi, Kenya
Kwon, J., M.S./Ph.D., Georgetown University, Washington, D.C.

Lopez Soto, E., Ph.D., Brown University, Providence, Rhode Island
Milioli, H.H., Ph.D., Garvan Institute of Medical Research, Sydney, New South Wales, Australia
Peters, S., B.A., AbSci, Vancouver, Washington
Robinson, D., B.Sc., Ottawa Hospital Research Institute, Ontario, Canada
Rodrigues, M., Ph.D., UT Southwestern Medical Center, Dallas
Sakers, K., Ph.D., Duke University Medical Center, Durham, North Carolina
Seraphin, M., Ph.D., University of Florida, Gainesville
Shin, B., Ph.D., California Institute of Technology, Pasadena
Toum, L., Ph.D., ITANOA, Buenos Aires, Argentina
Wang, L., Ph.D., Columbia University Medical Center, New York
Zheng, W., Ph.D., National Cancer Institute Frederick, Maryland.

## SEMINARS

Hawkins, D., University of Washington, Seattle: Regulatory genomics. Chromatin states 1: analysis of histone modifications. Chromatin states 2: overlapping data sets.
Henikoff, S., Fred Hutchinson Cancer Research Center, Seattle, Washington: Genome-wide profiling of the chromatin landscape.
Leek, J., Johns Hopkins University, Baltimore, Maryland: Statistics I: experiments, data, and visualization. Statistics II: models, experimental design, batch effects.
Mahony, S., Penn State University, University Park, Pennsylvania: Genomics of gene regulation 1: analyzing protein–DNA-binding interactions. Genomics of gene regulation 2: characterizing transcription factor–binding dynamics.
Mills, L., University of Minnesota, Minneapolis: Accessing public sequence data sets.

Pearson, W., University of Virginia, Charlottesville: Protein evolution and sequence similarity searching. Practical sequence similarity searching. PSSMs and HMMs: customized scoring matrices.
Sabik, O., University of Virginia, Charlottesville: R and RStudio introduction. Gene lists to pathways.
Taylor, J., Johns Hopkins University, Baltimore, Maryland: Sequencing technologies: new genomics. Genome assembly I: long reads. Genome assembly II: short reads. Galaxy for high-throughput analysis. RNA sequencing technologies. Short read alignment and genome variation. Discussion: single-cell recommendations. Probing higher-dimension chromatin structure.
Wilson Sayres, M., Arizona State University, Tempe: Sex bias in reference-based alignments.

# The Genome Access Course

INSTRUCTORS      **D. Fagegaltier,** New York Genome Center. New York
**E. Hodges,** Vanderbilt University School of Medicine, Nashville, Tennessee
**B. King,** University of Maine, Orono
**S. Munger,** The Jackson Laboratory, Farmington, Connecticut

INVITED SPEAKERS    **C. Baker** (November), The Jackson Laboratory, Ellsworth, Maine
**S. Goodwin** (November), Cold Spring Harbor Laboratory
**P. Sims** (May), Columbia University, New York, New York

The Genome Access Course (TGAC) is an intensive two-day introduction to bioinformatics offered multiple times each year. The course is broken into modules that are each designed to give a broad overview of a given topic, with ample time for examples chosen by the instructors. Each module features a brief lecture describing the theory, methods, and tools, followed by a set of worked examples that students complete. Students are encouraged to engage instructors during the course with specific tasks or problems that pertain to their own research.

    The core of TGAC is the analysis of sequence information framed in the context of completed genome sequences. Featured resources and examples primarily come from mammalian species, but concepts can be applied to any species. TGAC has been offered continuously since 2002 and has evolved over that time to meet the needs of scientists venturing into the analysis of large sequencing data sets. In 2019, the topics covered by TGAC included Genome Browsers, Sequence, Gene and Protein Resources, De Novo Analysis of Sequences, Sequence Variation, Comparative Genome Analysis, and Functional Genomic Elements and High-Throughput Sequence Data.

    Two iterations of TGAC took place in 2019, in May and November.

## May 5–7 (Cold Spring Harbor Laboratory)

25 PARTICIPANTS

Baer, R., Columbia University Medical Center, New York, New York

Bryll, A., University of Massachusetts Medical School, Worcester

Carvalho, Jr., J.R., Columbia University Medical Center, New York, New York

Fitzgerald, M., Children's Hospital of Philadelphia, Pennsylvania

Futcher, B., Stony Brook University, Stony Brook, New York

Gathungu, G., Stony Brook University Hospital, Stony Brook, New York

Gonzalez-Vicente, A., Cleveland Clinic/Lerner Research Institute, Cleveland, Ohio

Growe, J., Intellia Therapeutics, Cambridge, Massachusetts

Hughes, C., University of California, Irvine

Kelley, L., Syracuse University, New York

Khokhar, E., The Jackson Laboratory, Bar Harbor, Maine

Kim, E., Harvard Medical School, Boston, Massachusetts

Kirmaier, A., Boston College, Chestnut Hill, Massachusetts

McCoy, M., Syracuse University, New York

Mincer, S., Stony Brook University, Stony Brook, New York

Nguyen, D., National Institutes of Health, Bethesda, Maryland

Park, J., Dicerna Pharmaceuticals, Cambridge, Massachusetts

Pavia, M., National Institutes of Health, Bethesda, Maryland

Punnette, D., UK Department for International Trade, New York

Rieder, L., Emory University, Atlanta, Georgia

Salimi, S., University of Maryland, Baltimore

Smith, J., University of Massachusetts Medical School, Worcester

Weinberg, R., Massachusetts Institute of Technology, Cambridge

Yusufova, N., Weill Cornell Medicine, New York, New York

Zhou, H., Amgen Inc., South San Francisco, California

## November 11–13 (Cold Spring Harbor Laboratory)

37 PARTICIPANTS

Abston, K., University of Rochester Medical Center, Rochester, New York

Barlev, A., Feinstein Institutes for Medical Research, Manhasset, New York

Beckner, M., Kent State University, Willoughby Hills, Ohio

Benson, M., Beth Israel Deaconess Medical Center/Harvard Medical Center, Boston, Massachusetts

Collins, A., Columbia University Medical Center, New York, New York

Comstra Skye, H., Emory University, Atlanta, Georgia

Dudley, E., Ohio State University, Columbus

Ferland, J-M., Icahn School of Medicine at Mount Sinai, New York

Gray, T., Georgia State University, Atlanta

Han, W., Dicerna Pharmaceuticals, Cambridge, Massachusetts

Hunkele, A., The Estee Lauder Companies, Inc., Melville, New York

Keller, R., Memorial Sloan Kettering Cancer Center, New York, New York

Kennedy, A., Dana-Farber/Boston Children's, Massachusetts

Kitt, M., Case Western Reserve University, Cleveland, Ohio

Koenig, S., Ohio State University, Columbus

Kong, Y.W., Massachusetts Institute of Technology, Cambridge

Marnin, L., University of Maryland, Baltimore

Martino, J., Columbia University Medical Center, New York, New York

Mitchell, C., Columbia University Medical Center, New York, New York

Mohan, K., East Tennessee State University, Johnson City

Park, K.Y., University of Wisconsin, Madison

Park, W., Memorial Sloan Kettering Cancer Center, New York, New York

Patel, N., Intellia Therapeutics, Cambridge, Massachusetts

Pruvost, M., Advanced Science Research Center, New York, New York

Quintus, N., University of Arizona, Phoenix

Rana, M., Feinstein Institute for Medical Research, Manhasset, New York

Sadek, J., Weill Cornell Medicine, New York, New York

Santiago-Sanchez, G., University of Puerto Rico/Medical Sciences Campus, Guaynabo

Shastri, A., Albert Einstein College of Medicine, Bronx, New York

Sucharski, H., Ohio State University, Columbus

Thakkar, P., Weill Cornell Medicine, New York, New York

Tharin, S., Stanford University Medical School, California

Tintos-Hernandez, A., Center for Mitochondrial & Epigenomic Medicine, Philadelphia, Pennsylvania

Wang, H., Mayo Clinic Health System, Mankato, Minnesota

Wright, J., Simons Foundation, New York, New York

Yeh, S-Y., Icahn School of Medicine at Mount Sinai, New York

Yoon, J.S., Cystic Fibrosis Foundation, Lexington, Massachusetts

# SEMINARS

## INVITED SPEAKERS PROGRAM ("CSHL SEMINAR SERIES")

Each year, Cold Spring Harbor Laboratory invites speakers from outside the institution to present their findings. These weekly seminars keep the CSHL staff current on the latest scientific developments and broaden their perspectives. Graduate students and postdoctoral fellows meet with the seminar speakers for lunch immediately after the seminar, providing an opportunity for the exchange of ideas in an informal setting.

| Speaker | Title | Host |
|---|---|---|
| **January** | | |
| Elaine Mardis, Ph.D., Co-Executive Director, Institute for Genomic Medicine, Nationwide Children's Hospital | NGS-based exploration of CNS disease | Jason Sheltzer |
| Roy Parker, Ph.D., Investigator and Cech-Leinwand Endowed Chair of Biochemistry, HHMI and University of Colorado Boulder | RNP granules in health and disease | Leemor Joshua-Tor |
| Li-Huei Tsai, Ph.D., Professor and Director, Picower Institute for Learning and Memory at MIT | Network level approaches to studying neurodegeneration | Jessica Tollkuhn |
| **February** | | |
| Kristin Baldwin, Ph.D., Professor, The Scripps Research Institute | Reprogramming development to define mechanisms of cellular diversity and disease | Mike Wigler |
| Charles Swanton, M.B.Ph.D., F.R.C.P., F.Med. Sci., F.R.S., Royal Society Napier Professor, Cancer Research UK Chief Clinician, Chair, Personalised Cancer Medicine, University College London, Director, CRUK UCL/ Manchester Lung Cancer Centre of Excellence | Chromosomal instability and genome plasticity in cancer evolution, immune evasion, and metastasis | Richard Sever |
| Keiko Torii, Ph.D., Professor and Investigator, HHMI and Department of Biology, University of Washington | Cellular decision-making during stomatal patterning and differentiation | Ullas Pedmale |
| Ana Carrizosa Anderson, Ph.D., Associate Professor, Harvard Medical School | Using genomics to understand the CD8$^+$ T cell landscape in cancer | CSHL WiSE *McClintock Lecture |
| **March** | | |
| Albert J.R. Heck, Ph.D., Chair, Biomolecular Mass Spectrometry & Proteomics, Utrecht University | The diverse and expanding role of mass spectrometry in structural and molecular biology | Leemor Joshua-Tor |
| Valerie Weaver, Ph.D., Professor, Departments of Surgery, Radiation Oncology, and Bioengineering and Therapeutic Sciences, UCSF School of Medicine | Forcing tumor risk and progression | CSHL Postdocs |
| Kang Shen, Ph.D., Professor, Stanford University, HHMI Investigator | How genetic and developmental programs instruct neuronal function in *C. elegans* | Linda Van Aelst |
| Tomas Kirchhausen, Ph.D., Springer Family Chair, Boston Children's Hospital, Professor, Cell Biology, Professor, Pediatrics, Harvard Medical School | Subcellular dynamics from molecules to multicellular organisms | Lloyd Trotman |

| Speaker | Title | Host |
|---|---|---|
| **April** | | |
| David D. Ginty, Ph.D., Department of Neurobiology, Harvard Medical School, Investigator, Howard Hughes Medical Institute | Functional organization of the mammalian tactile sensory system: a view from the periphery | Bruce Stillman |
| **October** | | |
| William G. Kaelin, Jr., M.D., Investigator, Howard Hughes Medical Institute and Sidney Farber Professor of Medicine, Dana-Farber Cancer Institute *Winner of the 2019 Nobel Prize* | The von Hippel–Lindau tumor suppressor protein: insights into oxygen sensing, cancer metabolism, and drugging the undruggable | Jason Sheltzer |
| Nieng Yan, Ph.D., Shirley M. Tilghman Professor of Molecular Biology, Princeton University | How is electrical signal generated? Structural and mechanistic investigations of Nav channels | Hiro Furukawa |
| Geoff Wahl, Ph.D., Professor, The Salk Institute for Biological Studies, California | Deconstructing cancer from a developmental perspective | Chris Vakoc |
| **November** | | |
| Nenad Sestan, M.D., Ph.D. , Professor, Department of Neuroscience and Kavli Institute for Neuroscience, Yale School of Medicine | Building the human brain: molecular logic of neural circuit formation and evolution | Josh Huang |
| Helen Mayberg, M.D., Professor, Director, Center of Advanced Circuit Therapeutics, Icahn School of Medicine at Mount Sinai | Rethinking depression and its treatment: insights from studies of deep brain stimulation | Adam Kepecs *The William R. Miller Lecture |
| Emily Bernstein, Ph.D., Professor, Icahn School of Medicine at Mount Sinai | Chromatin aberrations in cancer | Camila dos Santos |
| **December** | | |
| Angelika Amon, Ph.D. | RAD21 is a driver of chromosome 8 gain in Ewing sarcoma to enhance DNA repair | CSHL WiSE |
| Tracy Johnson, Ph.D., Maria Rowena Ross Chair, Cell Biology and Biochemistry, HHMI Professor, Molecular, Cell, and Developmental Biology, UCLA | | CSHL DIAS |
| Xavier Darzacq, Ph.D., Associate Professor of Genetics, Genomics, and Development, UC Berkeley | | Graduate Students |

# CSHL IN-HOUSE SEMINAR PROGRAM

Cold Spring Harbor Laboratory's In-House Seminars were initiated to provide a semiformal avenue for communication among the various research groups at the Laboratory. The seminars also afford a necessary opportunity for graduate students and postgraduate staff to develop their skills in organizing, presenting, and defending their research.

| Speaker | Title |
| --- | --- |
| **January** | |
| Mikala Egeblad | Caught in the act: visualizing and targeting tumor–stroma interactions in metastasis |
| Steve Shea | Neural circuitry of parental care in mice |
| **February** | |
| Grinu Mathew | Defining new vulnerabilities of PTEN-null, AKT-Off lethal prostate cancer |
| Ivan Iossifov | Genetic variants linked to autism traits |
| Alexei Koulakov | Neural networks with motivation |
| **March** | |
| Jean Albrengues | Neutrophil extracellular traps promote and maintain lung fibrosis |
| Semir Beyaz | Dietary regulation of cancer and immunity |
| Josh Huang | Exploring the neural circuits of dexterity and object manipulation |
| Longwen Huang | High-throughput mapping of mesoscale connectomes in individual mice |
| Doug Fearon | The immune suppressive pathway of cancer-associated fibroblasts |
| **April** | |
| Christopher Hammell | The digital and analog control of animal development |
| **October** | |
| Jesse Gillis | The transcriptional legacy of developmental stochasticity in armadillos |
| Mohammad Rahman | NMD as a potential mediator of oncogenesis in spliceosomal-mutant MDS |
| **November** | |
| Christopher Vakoc | Aberrant transcriptional regulation in pancreatic cancer |
| Yixin (Elaine) Hu | Co-evolution of DNA replication origins and gene silencing |
| John Inglis | Preprints in biology and medicine: progress and prospects |
| **December** | |
| Peter Koo | Interpretable deep learning for regulatory genomics |
| Adam Kepecs | Neurobiology of confidence (and cancer) |

# BANBURY CENTER
# EXECUTIVE DIRECTOR'S REPORT

Since 1978, Cold Spring Harbor Laboratory's Banbury Center has convened impactful discussion meetings that allow small groups of experts to debate important issues and inspire new thinking. Meetings are organized around issues and challenges in the biosciences that benefit from the Center's unique style of discussion: emerging issues in need of strategy, established fields in need of review, controversial subjects calling for compromise or consensus, and areas in which diverse stakeholders/sectors need to engage or collaborate.

## Activities

It was another productive year at Banbury, with more than 50 events using the estate, including traditional Banbury meetings, Meetings and Courses Program (Workshop on Leadership in Bioscience; Workshop on Pancreatic Cancer; Vision: A Platform for Linking Circuits, Perception, and Behavior; Neural Data Science; Workshop on Autism Spectrum Disorders; Neuroscience of Addiction; Scientific Writing Retreat); Watson School of Biological Sciences courses (Immunology, Physiology of the Cell); and laboratory retreats.

The Center welcomed 475 expert participants in 2019—72% participating in their first Banbury meeting. Women represented 40% of total participants as well as 37% of meeting organizers. Experts were drawn from academia (76%), industry (9%), not-for-profit organizations (8%), U.S. and foreign governments (5%), and publishing or writing (1%). We continue to have global reach, with experts representing five continents, 20 nations, and 32 U.S. states. International attendees constituted 24% of total Banbury meeting participants.

Banbury meetings in 2019 spanned six thematic areas: cancer, neuroscience, technology, public health, plant biology, and science policy; many individual meeting topics touched on more than one of these themes. More than 60% centered on developing strategies for emerging fields or innovating in existing fields, whereas more than a quarter aimed to bridge divides across sectors, disciplines, and communities, and just more than 10% tackled challenging policy issues.

### Cancer

Cancer is a common theme for Banbury convenings, bringing together groups to examine the latest research as well as emerging concepts. In March, experts in *Cancer Fibroblasts and Therapies* examined roles for these cells in tumor progression and response to treatment. Also in March, *Cancer Immunotherapy: Where to Go Next* participants reviewed the current state of immunotherapy agents, with discussions of new approaches to translational research in this area. The following month, clinicians and researchers met to ask, *Glioblastoma: Why Is Impactful Science So Hard to Translate?* During their two and a half days, the group explored major obstacles to translating the "bench" to the "bedside." We ended our 2019 cancer meetings, and the meetings season as a whole, with December's *The Nervous System in Cancer*. This truly interdisciplinary meeting convened neuroscientists, cancer biologists, clinicians, academia, and industry to chart the course for a new field: cancer neuroscience.

### Neuroscience

In neuroscience, four Banbury meetings tackled questions spanning foundational research to issues in translation, practice, and policy. Our first meeting of the 2019 season, *Computational*

*Psychiatry*, used a series of breakout groups to develop recommendations for improving the use of computational psychiatry methods to address clinical problems. Further targeting mental health practice, our *Bridging the Research-to-Practice Chasm in Digital Mental Health* meeting convened a diverse group of stakeholders to identify the major obstacles to implementation of digital tools in U.S. mental health care and to agree to recommendations to overcome these challenges. Shifting further into translational research, March's *Integrated Control of Feeding and Energy Balance by Hypothalamic and Hindbrain Circuits* explored the intersection of brain systems involved in weight regulation, taste, and illness and mapped out important next steps for research. Finally, in October, experts gathered for *CaMKII and Its Role as a Self-Tuning Structural Protein at the Synapse*, sharing their latest research and inspiring new ideas and collaborations.

### Technology

Two 2019 meetings centered on technology development, with very different applications. In March, we reconvened a 2018 group to continue discussions of *DNA for Digital Storage*. Experts drawn from diverse disciplines and sectors, who may not have a consistent place to connect, were able to discuss progress in the field as well as new or persisting challenges. At September's *Liquid Biopsies* meeting, the focus was on this rapidly developing technology for diagnosing cancer and monitoring treatment. Companies developing this technology, as well as translational researchers and clinicians, discussed new research and future strategies.

### Public Health

Banbury's history convening discussions around public health issues continued with two meetings. In May, a highly international group convened at the Center to discuss *Intermediate Indicators for Impact: The Art and Science of Effective Definition and Use of Prevention Indicators in the HIV Response*. The meeting brought together experts experienced with collecting and analyzing data, those responsible for using the analyses and knowledge, as well as those affected by resulting programs to consider opportunities for the future of the global HIV response. Similarly considering future outlooks, November's *Microbiology of the Built Environment* meeting reviewed ongoing work studying microbes living on and inside built structures that can be linked to human health, focusing on strategies to ensure long-term momentum for this intersectional field.

### Plants

Further drawing from the microbiology theme, but shifting from humans and buildings to plants, April's *The Plant Microbiota* meeting convened plant and microbial scientists to review the latest advances in plant–microbial interactions research, as well as implications for plant biotechnology and food security.

### Science Policy

With an eye toward science policy, two meetings centered on the growing movement of science and health products into the hands of consumers and the ethical challenges that have resulted. October's *Reconceptualizing the Challenges of Direct-to-Consumer Health Products* built on a 2018 meeting that focused on DTC neuroscience, but expanded to include a broader suite of products. The group considered issues of safety and regulatory challenges, as well as how these products are transforming the physician–patient relationship. That same month, the Center welcomed expert stakeholders to explore *Emerging Issues of Privacy, Trust, and Societal Benefit from Consumer Genomics*. The discussions of the use of genetic genealogy databases by law enforcement investigating violent crimes were especially timely, following recent apprehension of the Golden State Killer in

Rita Allen Foundation Scholars Symposium participants walk to the conference room.

part through use of one such database, and also coinciding with release of a Department of Justice Interim Policy on forensic genetic genealogy.

### *Collaborations with Foundations*

Finally, we were delighted to collaborate with three excellent foundations for events aiming to celebrate and support researchers, and to develop strategy. In August, the 2019 *Rita Allen Foundation Scholars Symposium* was held at Banbury, celebrating the current scholars' research and inspiring new connections. The *Boehringer Ingelheim Fellows Retreat* returned in September for several days of training in all aspects of scientific communication. The Lustgarten Foundation also returned to the Conference Room for their 2019 Scientific Meeting, providing an opportunity for the Scientific Advisory Board, as well as Foundation-supported investigators, to discuss research and strategy.

## Outcomes and Impact

It can be difficult to quantify the impact that Banbury meetings have on science and society. Often, Banbury meetings inspire new ideas that lead to discovery or new directions, instigate new connections that build to productive partnerships, or permit sensitive discussions that build understanding and consensus. As we attempt to better track the short-, medium-, and long-term impacts of the Center, we have begun a coordinated effort to follow up meetings at regular intervals enquiring about quantifiable outputs and outcomes (e.g., new collaborations, papers, funding, facilities, policy changes, and fields) as well as individual anecdotes and testimonies as to how past participants view the impact of a specific meeting on the field. Although still in the process of collecting these data, we have compiled recent policy outcomes and new publications[*] resulting from Banbury Center meetings. Among these are changes to FDA and CDC guidelines on Lyme disease diagnostics, which resulted from discussions at a 2016 Banbury meeting and are poised to have major impact in Lyme detection.

## Support

Funding continues to be a major hurdle in supporting Banbury meetings, as topics often lie at new intersections of science and technology or deal with delicate ethical or policy issues. We are

---

[*]See lists at end of Director's Report.

ever grateful to the organizations and individuals that provide the financial support to enable Banbury to convene global leaders. In 2019, Banbury secured financial support from not-for-profit organizations (54%) and the private sector (26%). The CSHL Corporate Sponsor Program remains a critical resource for cutting-edge meetings and contributed 19% of funding for Banbury meetings.

## The Team

The Center is successful thanks to a team of professionals who ensure that the estate and programs are running at a high level. Michelle Corbeaux expertly manages the Center's finances, working closely with Development's Michael Gurtowski and Cat Donaldson to manage a quality Corporate Sponsor Program. In 2019, we lost Jasmine Breeland to graduate school after serving as Banbury's inaugural Communications and Special Projects Coordinator. We were lucky to welcome Allison Eichler, who took the reins and has ensured continuity in Banbury's communications. Basia Polakowski continues to oversee our three residence buildings, ensuring our guests feel welcome and comfortable, along with our housekeepers, Miriam and Maria, supervised by Claudia Schmid and Patricia McAdams. The Culinary Services team, led by Jim Hope and overseen by Christina DeDora, keeps guests well fed, while Bill Dickerson and the entire Audiovisual staff led by Ed Campodonico ensure technology supports rather than distracts. Finally, Jose Peña-Corvera, Paulo Krizanovski, and Juan Colocho skillfully maintain 55 acres of impeccable grounds, and the entire Facilities team quite literally keeps us running.

**Rebecca Leshan**
*Executive Director*

## 2019 Policy Changes Resulting from Banbury Meetings

Mead P, Petersen J, Hinckley A. 2019. Updated CDC recommendation for serologic diagnosis of Lyme disease. *MMWR Morb Mortal Wkly Rep* **68:** 703. doi:10.15585/mmwr.mm6832a4external icon

U.S. Food and Drug Administration (July 29, 2019). FDA clears new indications for existing Lyme disease tests that may help streamline diagnoses [Press Release]. Retrieved from www.fda.gov/news-events/press-announcements/fda-clears-new-indications-existing-lyme-disease-tests-may-help-streamline-diagnoses

## 2019 Publications Resulting from Banbury Meetings

Amaravadi RK, Kimmelman AC, Debnath J. 2019. Targeting autophagy in cancer: recent advances and future directions. *Cancer Discovery* **9:** 1167–1181. doi:10.1158/2159-8290.CD-19-0292

Greider CW, Sheltzer JM, Cantalupo NC, Copeland WB, Dasgupta N, Hopkins N, Jansen JM, Joshua-Tor L, McDowell GS, Metcalf JL, et al. 2019. Increasing gender diversity in the STEM research workforce. *Science* **366:** 692–695. doi:10.1126/science.aaz0649

Jamieson KH, McNutt M, Kiermer V, Sever R. 2019. Signaling the trustworthiness of science. *Proc Natl Acad Sci* **116:** 19231–19236. doi:10.1073/pnas.1913039116

Mirabelli CK, Nusse R, Tuveson DA, Williams BO. 2019. Perspectives on the role of Wnt biology in cancer. *Sci Signal* **12:** eaay4494. doi:10.1126/scisignal.aay4494

Njølstad PR, Andreassen OA, Brunak S, Børglum AD, Dillner J, Esko T, Franks PW, Freimer N, Groop L, Heimer H, et al. 2019. Roadmap for a precision-medicine initiative in the Nordic region. *Nat Genet* **51:** 924–930. doi:10.1038/s41588-019-0391-1

Ward H, et al., Guest Eds. 2019. Special Issue: Maximizing the impact of HIV prevention technologies in sub-Saharan Africa [Supplement]. *J Int AIDS Soc* **22**.

Wexler A, Reiner PB. 2019. Oversight of direct-to-consumer neurotechnologies. *Science* **363:** 234–235. doi:10.1126/science.aav0223

Wurtzel ET, Vickers CE, Hanson AD, Millar AH, Cooper M, Voss-Fels KP, Nikel PI, Erb TJ. 2019. Revolutionizing agriculture with synthetic biology. *Nat Plants* **5:** 1207–1210. doi:10.1038/s41477-019-0539-0

*In Press*

Browning M, Carter CS, Chatham C, Den Ouden H, Gillen CM, Baker JT, Chekroud AM, Cools R, Dayan P, Gold J, et al. 2020. Realizing the clinical potential of computational psychiatry: report from the Banbury Center Meeting, February 2019. *Biol Psychiatry* **88:** e5–e10. doi:10.31234/osf.io/5qbxp

Gomes-Solecki M, Arnaboldi PA, Backenson PB, Benach JL, Cooper CL, Dattwyler RJ, Diuk-Wasser M, Fikrig E, Hovius JW, Laegreid W, et al. 2020. Protective immunity and new vaccines for Lyme disease. *Clinical Infectious Diseases* **70:** 1768–1773. doi:10.1093/cid/ciz872

Monje M, Borniger JC, D'Silva NJ, Deneen B, Dirks PB, Fatahi F, Frenette PS, Garzia L, Gutmann DH, Hanahan D, et al. 2020. Roadmap for the emerging field of cancer neuroscience. *Cell* **181:** 219–222. doi:10.1016/j.cell.2020.03.034

Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM, Fearon D, Greten FR, Hingorani SR, Hunter T, et al. 2020. A framework for advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer* **20:** 174–186. doi:10.1038/s41568-019-0238-1

## BANBURY CENTER MEETINGS

| Dates | Title | Organizer(s) |
|-------|-------|--------------|
| February 3–6 | Computational Psychiatry | M. Browning, M. Frank, Q. Huys, M. Paulus |
| March 3–5 | DNA for Digital Storage II | E. Birney, Y. Erlich, N. Goldman, J-F. Lutz |
| March 10–13 | Cancer Fibroblasts and Therapies | C. Jørgensen, E. Puré, D. Tuveson |
| March 16–19 | Cancer Immunotherapy: Where to Go Next | I. Mellman, M. Merad |
| March 31–April 3 | Integrated Control of Feeding and Energy Balance by Hypothalamic and Hindbrain Circuits | L. Heisler, M. Myers |
| April 7–10 | Glioblastoma: Why Is Impactful Science So Hard to Translate? | P. Dirks, E. Maher, W. Weiss |
| April 14–17 | The Plant Microbiota | J. Dangl, P. Schulze-Lefert, J. Vorholt |
| May 12–15 | Intermediate Indicators for Impact: The Art and Science of Effective Definition and Use of Prevention Indicators in the HIV Response | C. Holmes, N. Kilonzo, M. Mahy |
| August 13–15 | Rita Allen Foundation Scholars Symposium | E. Christopherson |
| September 8–11 | Liquid Biopsies | L. Diaz, V. Velculescu |
| September 13–18 | Communicating Science—Boehringer Ingelheim Fellows Retreat | K. Achenbach, S. Schedler, C. Walther |
| October 6–8 | Bridging the Research-to-Practice Chasm in Digital Mental Health | P. Areán, D. Mohr |
| October 14–16 | Reconceptualizing the Challenges of Direct-to-Consumer Health Products | T. Caulfield, L. Turner, A. Wexler |
| October 19–22 | Emerging Issues of Privacy, Trust, and Societal Benefit from Consumer Genomics | Y. Erlich, A.L. McGuire |
| October 27–30 | CaMKII and Its Role as a Self-Tuning Structural Protein at the Synapse | Y. Hayashi, J. Hell |
| November 3–6 | Microbiology of the Built Environment | J. Green, R Kolter, J. Peccia |
| November 10–12 | Lustgarten Foundation Scientific Advisory Board Meeting | K. Kaplan, D. Tuveson, R. Vizza |
| December 10–13 | The Nervous System in Cancer | S. Knox, M.L. Monje, T. Wang |

# BANBURY CENTER MEETINGS

## Computational Psychiatry

**February 3–6**

ARRANGED BY      **M. Browning,** University of Oxford, United Kingdom
**M. Frank,** Brown University, Providence, Rhode Island
**Q. Huys,** University College London, United Kingdom
**M. Paulus,** Laureate Institute for Brain Research, Tulsa, Oklahoma

FUNDED BY      **The Society of Biological Psychiatry; The William K. Warren Foundation; The Carney Institute for Brain Science at Brown University**

Although there has been a great deal of progress in the field of computational psychiatry over recent years, much of this progress has occurred in addressing foundational mechanisms of brain and behavior, and their alterations in patient populations, but without having a direct impact on treatment. The goal of this meeting was to bridge that gap by facilitating the process by which computational models can be used to address real-world clinical questions in psychiatry.

**Welcoming Remarks:** **R. Leshan,** Director, Banbury Center, Cold Spring Harbor Laboratory

**Introduction to Meeting, Goals, Outputs, Structure:** **M. Browning,** University of Oxford, United Kingdom
**M. Frank,** Brown University, Providence, Rhode Island
**Q. Huys,** University College London, United Kingdom
**M. Paulus,** Laureate Institute for Brain Research, Tulsa, Oklahoma
**A. Churchland,** Cold Spring Harbor Laboratory

D. Pizzagalli, R. Goldstein



Q. Huys, C. Carter, C. Hartley, M. Paulus

**SESSION 1A:** Breakout Groups I

**Group 1: Questions and Study Designs**

C. Chatham, F. Hoffmann–La Roche Ltd., Basel, Switzerland

A. Chekroud, Spring Health/Yale University, New York

J. Gold, University of Maryland School of Medicine, Baltimore

Q. Huys, University College London, United Kingdom

J. Krystal, Yale School of Medicine, New Haven, Connecticut

M. Phillips, University of Pittsburgh, Pennsylvania

A. Powers, Yale University School of Medicine, New Haven, Connecticut

**Group 2: Infrastructure and IT**

J. Baker, Harvard University, Belmont, Massachusetts

C. Carter, University of California, Davis, Sacramento

K. Enno Stephan, University of Zürich & ETH Zürich, Switzerland

M. Ferrante, National Institute of Mental Health, Rockville, Maryland

R. Goldstein, Icahn School of Medicine at Mount Sinai, New York

J. Mourao-Miranda, University College London, United Kingdom

M. Paulus, Laureate Institute for Brain Research, Tulsa, Oklahoma

**Group 3: Optimizing the Task**

R. Cools, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands

P. Dayan, Max Planck Institute for Biological Cybernetics, Tübingen, Germany

H. Den Ouden, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands

M. Frank, Brown University, Providence, Rhode Island

C. Hartley, New York University, New York

J. Roiser, University College London, United Kingdom

K. Schmack, Cold Spring Harbor Laboratory

M. Sebold, Charité–Universitätsmedizin Berlin, Germany

**Group 4: Optimizing Selection between Tasks**

M. Browning, University of Oxford, United Kingdom

B. Cuthbert, National Institute of Mental Health, Fitchburg, Wisconsin

C. Gillan, Trinity College Dublin, Ireland

A. Kepecs, Cold Spring Harbor Laboratory

R. Lawson, University of Cambridge, United Kingdom

D. Pizzagalli, Harvard University, Belmont, Massachusetts

D. Rindskopf, The Graduate Center, CUNY, New York

D. Schiller, Icahn School of Medicine at Mount Sinai, New York



A. Churchland, P. Dayan, Q. Huys, M. Browning, M. Paulus



K. Schmack, J. Krystal

**SESSION 1B:** Feedback Session 1

**All Participants**

**SESSION 2:** Breakout Groups II

**Group 3, Plus Participants from Groups 1, 2**

R. Cools, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands

P. Dayan, Max Planck Institute for Biological Cybernetics, Tübingen, Germany

H. Den Ouden, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands

M. Frank, Brown University, Providence, Rhode Island

C. Hartley, New York University, New York

J. Roiser, University College London, United Kingdom

K. Schmack, Cold Spring Harbor Laboratory

M. Sebold, Charité–Universitätsmedizin Berlin, Germany

**Group 4, Plus Participants from Groups 1, 2**

M. Browning, University of Oxford, United Kingdom

B. Cuthbert, National Institute of Mental Health, Fitchburg, Wisconsin

C. Gillan, Trinity College Dublin, Ireland

A. Kepecs, Cold Spring Harbor Laboratory

R. Lawson, University of Cambridge, United Kingdom

D. Pizzagalli, Harvard University, Belmont, Massachusetts

D. Rindskopf, The Graduate Center, CUNY, New York

D. Schiller, Icahn School of Medicine at Mount Sinai, New York

**SESSION 3A:** Breakout Groups III

**Group 1, Plus Participants from Groups 3, 4**

C. Chatham, F. Hoffmann–La Roche Ltd., Basel, Switzerland

A. Chekroud, Spring Health/Yale University, New York

J. Gold, University of Maryland School of Medicine, Baltimore

Q. Huys, University College London, United Kingdom

J. Krystal, Yale School of Medicine, New Haven, Connecticut

M. Phillips, University of Pittsburgh, Pennsylvania

A. Powers, Yale University School of Medicine, New Haven, Connecticut

**Group 2, Plus Participants from Groups 3, 4**

J. Baker, Harvard University, Belmont, Massachusetts

C. Carter, University of California, Davis, Sacramento

M. Ferrante, National Institute of Mental Health, Rockville, Maryland

R. Goldstein, Icahn School of Medicine at Mount Sinai, New York

J. Mourao-Miranda, University College London, United Kingdom

M. Paulus, Laureate Institute for Brain Research, Tulsa, Oklahoma

K.E. Stephan, University of Zürich/ETH Zürich, Switzerland

**SESSION 3B:** Feedback Session 2

**All Participants**

**SESSION 4:** Breakout Groups IV

**Group 1, Plus Participants from Groups 2, 4**

C. Chatham, F. Hoffmann–La Roche Ltd., Basel, Switzerland

A. Chekroud, Spring Health & Yale University, New York

J. Gold, University of Maryland School of Medicine, Baltimore

Q. Huys, University College London, United Kingdom

J. Krystal, Yale School of Medicine, New Haven, Connecticut

M. Phillips, University of Pittsburgh, Pennsylvania

A. Powers, Yale University School of Medicine, New Haven, Connecticut

**Group 3, Plus Participants from Groups 2, 4**

R. Cools, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands

P. Dayan, Max Planck Institute for Biological Cybernetics, Tübingen, Germany

H. Den Ouden, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands

M. Frank, Brown University, Providence, Rhode Island

C. Hartley, New York University, New York

J. Roiser, University College London, United Kingdom

K. Schmack, Cold Spring Harbor Laboratory

M. Sebold, Charité–Universitätsmedizin Berlin, Germany

**SESSION 5:** Breakout Groups V

**Group 2, Plus Participants from Groups 1, 3**

J. Baker, Harvard University, Belmont, Massachusetts

C. Carter, University of California, Davis, Sacramento

M. Ferrante, National Institute of Mental Health, Rockville, Maryland

R. Goldstein, Icahn School of Medicine at Mount Sinai, New York

J. Mourao-Miranda, University College London, United Kingdom

M. Paulus, Laureate Institute for Brain Research, Tulsa, Oklahoma

K.E. Stephan, University of Zürich/ETH Zürich, Zürich, Switzerland

**Group 4, Plus Participants from Groups 1, 3**

M. Browning, University of Oxford, United Kingdom

B. Cuthbert, National Institute of Mental Health, Fitchburg, Wisconsin

C. Gillan, Trinity College Dublin, Ireland

A. Kepecs, Cold Spring Harbor Laboratory

R. Lawson, University of Cambridge, United Kingdom

D. Pizzagalli, Harvard University, Belmont, Massachusetts

D. Rindskopf, The Graduate Center, CUNY, New York

D. Schiller, Icahn School of Medicine at Mount Sinai, New York

**SESSION 6:** Breakout Groups VI

**Group 1**

C. Chatham, F. Hoffmann–La Roche Ltd., Basel, Switzerland

A. Chekroud, Spring Health/Yale University, New York

J. Gold, University of Maryland School of Medicine, Baltimore

Q. Huys, University College London, United Kingdom

J. Krystal, Yale School of Medicine, New Haven, Connecticut

M. Phillips, University of Pittsburgh, Pennsylvania

A. Powers, Yale University School of Medicine, New Haven, Connecticut

**Group 2**

J. Baker, Harvard University, Belmont, Massachusetts

C. Carter, University of California, Davis, Sacramento

M. Ferrante, National Institute of Mental Health, Rockville, Maryland

R. Goldstein, Icahn School of Medicine at Mount Sinai, New York

J. Mourao-Miranda, University College London, United Kingdom

M. Paulus, Laureate Institute for Brain Research, Tulsa, Oklahoma

K.E. Stephan, University of Zürich/ETH Zürich, Zürich, Switzerland

**Group 3**

R. Cools, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands

P. Dayan, Max Planck Institute for Biological Cybernetics, Tübingen, Germany

H. Den Ouden, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands

M. Frank, Brown University, Providence, Rhode Island

C. Hartley, New York University, New York

J. Roiser, University College London, United Kingdom

K. Schmack, Cold Spring Harbor Laboratory

M. Sebold, Charité–Universitätsmedizin Berlin, Germany

**Group 4**

M. Browning, University of Oxford, United Kingdom

B. Cuthbert, National Institute of Mental Health, Fitchburg, Wisconsin

C. Gillan, Trinity College Dublin, Ireland

A. Kepecs, Cold Spring Harbor Laboratory

R. Lawson, University of Cambridge, United Kingdom

D. Pizzagalli, Harvard University, Belmont, Massachusetts

D. Rindskopf, The Graduate Center, CUNY, New York

D. Schiller, Icahn School of Medicine at Mount Sinai, New York

**SESSION 7:** Feedback Session 3

**Chairperson: M. Paulus,** Laureate Institute for Brain Research, Tulsa, Oklahoma

[*Groups present a summary of their recommendations*]

**SESSION 8:** Review and Finalization of Recommendations

**Chairperson: Q. Huys,** University College London, United Kingdom

**SESSION 9:** Meeting Wrap-Up

**Chairpersons: M. Browning,** University of Oxford, United Kingdom; **M. Frank,** Brown University, Providence, Rhode Island; **Q. Huys,** University College London, United Kingdom; **M. Paulus,** Laureate Institute for Brain Research, Tulsa, Oklahoma

# DNA for Digital Storage II

A spring 2018 Banbury Center meeting convened experts from diverse sectors and disciplines to examine the use of DNA encoding for data storage. This 2019 Banbury meeting revisited the state of the field, including current knowledge (new and residual) limitations and opportunities for research and commercial exploitation.

**Welcoming Remarks:  R. Leshan,** Director, Banbury Center, Cold Spring Harbor Laboratory

**Introduction and Meeting Objectives:  E. Birney,** European Bioinformatics Institute, Hinxton, United Kingdom

SESSION 1: Information: Theoretic Progress

**Chairperson: E. Birney,** European Bioinformatics Institute, Hinxton, United Kingdom

N. Goldman, European Bioinformatics Institute, Hinxton, United Kingdom: DNA storage channel error rates.
K. Strauss, Microsoft, Redmond, Washington: Thoughts and recent advances in DNA data storage.

Z. Yakhini, IDC Herzliya, Herzliya, Israel: Composite DNA alphabets enable less synthesis cycles.
Y. Erlich, Columbia University/MyHeritage, New York, New York: On optimal encoding of files in asynchronous DNA synthesizers.

J. Huang, K. Strauss



E. Miller, E. Zadok, L. Ceze, D. Zielinski

**SESSION 2:** Synthesis Technologies

**Chairperson: D. Zielinski,** Institut Curie, Paris, France

H. Lee, Harvard Medical School, Boston, Massachusetts: Ter-
minator-free enzymatic DNA synthesis.

J-F. Lutz, CNRS/Institut Charles Sadron, Strasbourg, France:
Recent progress on abiotic digital polymers.

P. Cai, University of Manchester, United Kingdom: Synthetic
genomics: from genetic parts to genomes.

R. Nolte, Radboud University, Nijmegen, the Netherlands:
Encoding information into polymers: a supramolecular cata-
lytic approach.

M. Somoza, University of Vienna, Austria: Large-scale photo-
lithographic synthesis of DNA.

**SESSION 3:** Companies

**Chairperson: N. Goldman,** European Bioinformatics Insti-
tute, Hinxton, United Kingdom

B. Bramlett, Twist Bioscience, San Francisco, California:
Considerations for media design.

H. Park, CATALOG, Cambridge, United Kingdom: Com-
mercializing DNA-based data storage.

S. Palluk, Ansa Biotechnologies, Berkeley, California: De
novo DNA synthesis using enzymes.

J. Huang, Nuclera Nucleics, Cambridge, United Kingdom:
Enzyme-mediated DNA printer.

**SESSION 4:** Economics, Ecosystem, and the Storage Industry

**Chairperson: K. Strauss,** Microsoft, Redmond, Washington

E. Miller, University of California, Santa Cruz; E. Zadok,
Stony Brook University, Stony Brook, New York: Glass and
more: recent trends in data storage technologies.

D. Zielinski, Institut Curie, Paris, France: Communicating
advances in DNA storage technology.

**SESSION 5:** Systems

**Chairperson: J-F. Lutz,** CNRS/Institut Charles Sadron, Stras-
bourg, France

R. Grass, ETH Zurich, Switzerland: Integration of DNA into
materials.

L. Ceze, University of Washington, Seattle: End-to-end DNA
data storage systems and near-molecule processing.

**SESSION 6:** Meeting Wrap-Up

**Chairperson: Y. Erlich,** Columbia University/MyHeritage,
New York, New York



N. Goldman, Z. Yakhini

# Cancer Fibroblasts and Therapies

**March 10–13**

ARRANGED BY  **C. Jørgensen,** Cancer Research UK Manchester Institute, United Kingdom
       **E. Puré,** University of Pennsylvania, Philadelphia
       **D. Tuveson,** Cold Spring Harbor Laboratory

FUNDED BY   **The Northwell Health/Cold Spring Harbor Laboratory Affiliation**

Cancer-associated fibroblasts (CAFs) are integral components of carcinomas, where they influence tumor progression and therapeutic response. Recent studies have revealed juxtacrine and paracrine interactions between CAFs and neoplastic cells that promote metabolic adaptation and tissue patterning. In addition, analyses have demonstrated subtypes of CAFs with different roles in the tumor microenvironment, including immune modulation. This meeting convened experts to discuss current understanding of CAF biology, with an emphasis on new approaches to probe the fundamental properties of CAFs and medical applications of CAF targeting.

**Welcoming Remarks: R. Leshan,** Director, Banbury Center, Cold Spring Harbor Laboratory

**Introduction and Meeting Objectives: E. Puré,** University of Pennsylvania, Philadelphia
                  **D. Tuveson,** Cold Spring Harbor Laboratory

**SESSION 1:** Matricellular Regulation and CAFs

**Chairperson: D. Tuveson,** Cold Spring Harbor Laboratory

V. Weaver, University of California, San Francisco: Tumor fibrosis, inflammation, and stromal fibroblast-mediated collagen cross-linking.

R. Hynes, Massachusetts Institute of Technology, Cambridge: Extracellular matrix vulnerabilities in cancer.

E. Cuckierman, Fox Chase Cancer Center, Philadelphia, Pennsylvania: Desmoplastic fibroblasts and self-produced ECMs protect pancreatic adenocarcinoma via metabolic support and innate immunosuppression.

Z. Werb, D. Tuveson



M. Egeblad, R. Hynes, T. Hunter, E. Sahai

**SESSION 2:** CAFs and Metastasis

**Chairperson: T. Tlsty,** University of California, San Francisco

Z. Werb, University of California, San Francisco: How mesenchymal cells promote breast cancer growth and metastasis.

E. Sahai, Francis Crick Institute, London, United Kingdom: The role of cancer-associated fibroblasts in modulating invasion and the immune system.

T. Tlsty, University of California, San Francisco: Identifying and targeting stromal states that contribute to cancer progression.

**SESSION 3:** The Permissive Niche and Tumor Development

**Chairperson: F. Greten,** Georg-Speyer-Haus, Frankfurt, Germany

E. Puré, University of Pennsylvania, Philadelphia: Involvement of a stromagenic switch in establishment of a tumor-hospitable environment.

A. Weeraratna, The Wistar Institute, Philadelphia, Pennsylvania: A wrinkle in TiME: how the aged tumor microenvironment drives melanoma progression.

S. Stewart, Washington University in St. Louis, Missouri: Age-related changes in the tumor microenvironment drives tumorigenesis.

**SESSION 4:** CAFs and Immune Regulation

**Chairperson: R. Evans,** Salk Institute for Biological Studies, La Jolla, California

D. Fearon, Cold Spring Harbor Laboratory: The immune suppressive pathway of cancer-associated fibroblasts.

F. Greten, Georg-Speyer-Haus, Frankfurt, Germany: CAFs and colon cancer.

**SESSION 5:** Fibroblast Heterogeneity

**Chairperson: R. Evans,** Salk Institute for Biological Studies, La Jolla, California

D. Tuveson, Cold Spring Harbor Laboratory: Fibroblast heterogeneity in pancreatic cancer.

F. Watt, King's College London, United Kingdom: Exploiting human skin fibroblast subpopulations to alleviate scarring.

M. Egeblad, Cold Spring Harbor Laboratory: Neutrophils orchestrate the establishment of a fibrotic, metastatic microenvironment.

**SESSION 6:** Reciprocal Interactions between Tumor Cells and CAFs

**Chairperson: A. Weeraratna,** The Wistar Institute, Philadelphia, Pennsylvania

T. Hunter, Salk Institute for Biological Studies, La Jolla, California: LIF, a stromal cell pancreatic cancer driver.

S. Powers, Stony Brook University, Stony Brook, New York: Discovery-driven exploration of breast cancer progression using mouse models and single-cell RNA-seq.

M. Sherman, Oregon Health and Science University, Portland: Determinants and consequences of pancreatic cancer stromal evolution.

**SESSION 7:** Regulatory Networks of Stromal Dysfunction

**Chairperson: M. Sherman,** Oregon Health and Science University, Portland

R. Evans, Salk Institute for Biological Studies, La Jolla, California: Epigenetic control of the stromal response.

R. Scherz-Shouval, Weizmann Institute of Science, Rehovot, Israel: Transcriptional stress networks underlying phenotypic plasticity in the tumor microenvironment.

R. Maki and D. Ramirez, Northwell Health, Lake Success, New York: Soft-tissue sarcoma: when the stroma is the cancer.

**SESSION 8:** Metabolic Regulation and CAFs

**Chairperson: T. Janowitz,** Cold Spring Harbor Laboratory

I. Astsaturov, Fox Chase Cancer Center, Philadelphia, Pennsylvania: CAFs as a source of lipids in pancreatic carcinogenesis.

A. Kimmelman, New York University Langone Medical Center, New York: Metabolic cross talk in pancreatic cancer.

**SESSION 9:** Reprogramming the Tumor Microenvironment

**Chairperson: E. Cuckierman,** Fox Chase Cancer Center, Philadelphia, Pennsylvania

D. DeNardo, Washington University School of Medicine in St. Louis, Missouri: Reprogramming the pancreatic tumor microenvironment to improve responses to therapy.

T. Janowitz, Cold Spring Harbor Laboratory: Interleukin-6-induced metabolic reprogramming in pancreatic cancer.

**SESSION 10:** Enhancing Cancer Treatment

**Chairperson: E. Puré,** University of Pennsylvania, Philadelphia

M. Kolonin, University of Texas, Houston: Fibroblasts from adipose tissue as a drug target in cancer progression to chemo resistance and metastasis.

S. Hingorani, Fred Hutchinson Cancer Research Center, Seattle, Washington: Stromal plasticity and perfidity in pancreas cancer.



D. Fearon, E. Cukierman

R. Jain, Harvard Medical School, Boston, Massachusetts: Re-engineering the tumor microenvironment to enhance cancer treatment: bench to bedside.

**SESSION 11:** Meeting Wrap-Up

**Chairpersons: E. Puré,** University of Pennsylvania, Philadelphia; **D. Tuveson,** Cold Spring Harbor Laboratory

# Cancer Immunotherapy: Where to Go Next

**March 16–19**

ARRANGED BY      **I. Mellman,** Genentech, South San Francisco, California
                      **M. Merad,** Icahn School of Medicine at Mount Sinai, New York

FUNDED BY         **Genentech; Cold Spring Harbor Laboratory;** with additional support from **AbbVie**

This Banbury meeting convened experts to critically review the mechanisms that control tumor response or underlie the lack of response to current immunotherapy agents. Key themes included main regulatory pathways that limit antitumor immunity; clinical benefit of novel immunotherapy agents alone or in combination; and novel approaches to accurately assess clinical and biomarker responses and the fundamental features of cancer immunity.

**Welcoming Remarks: R. Leshan,** Director, Banbury Center, Cold Spring Harbor Laboratory
                               **I. Mellman,** Genentech, South San Francisco, California
                               **M. Merad,** Icahn School of Medicine at Mount Sinai, New York

**SESSION 1:** Resistance Mechanisms to Current Checkpoint Inhibitors

**Chairperson: I. Mellman,** Genentech, South San Francisco, California

R. Ahmed, Emory University, Atlanta, Georgia: T-cell exhaustion and PD-1 therapy.
N. Hacohen, Massachusetts General Hospital, Boston: Drivers and resistors of tumor immunity.

**SESSION 2:** Tumor Antigen Immunity

**Chairperson: D. Pardoll,** Johns Hopkins University, Baltimore, Maryland

D. Pardoll, Johns Hopkins University, Baltimore, Maryland: Analysis of repertoire and function of anti-tumor T-cell responses elicited by checkpoint inhibition.
R. Seder, National Institute of Allergy and Infectious Disease, Bethesda, Maryland: Optimizing neoantigen-specific CD8

L. Zitvogel, M. Merad, N. Hacohen



U. Sahin, U. Weiss

responses by intravenous delivery of a nanoparticle vaccine.

L. Delamarre, Genentech, South San Francisco, California: Neoantigens for personalized cancer immunotherapy.

U. Sahin, University Medical Center, Mainz, Germany: Personalized cancer immunotherapy.

**SESSION 3:** Tumor Microenvironment Modulation of Tumor Immunity

**Chairperson: M. Merad,** Icahn School of Medicine at Mount Sinai, New York

D. Lambrechts, VIB-KU Leuven Center for Cancer Biology, Belgium: Single-cell profiling of the pan-cancer tumor microenvironment during checkpoint immunotherapy.

M. Merad, Icahn School of Medicine at Mount Sinai, New York: Myeloid microenvironment of cancer lesions.

S. Demaria, Weill Cornell Medicine, New York, New York: Activation of the DNA damage response by radiotherapy enhances the expression and cross-presentation of immunogenic mutations to CD8 T cells.

D. Tuveson, Cold Spring Harbor Laboratory: Fibroblast subsets during pancreatic inflammation and cancer.

**SESSION 4:** Immune Cell Engineering

**Chairperson: M. Sadelain,** Memorial Sloan Kettering Cancer Center, New York, New York

M. Sadelain, Memorial Sloan Kettering Cancer Center, New York, New York: New directions in CAR T-cell engineering.

B. Brown, Icahn School of Medicine at Mount Sinai, New York: Antigen and antitumor responses.

**SESSION 5:** Microbiome: Biomarker or Therapeutic Immunity

**Chairperson: L. Zitvogel,** Gustave Roussy, Villejuif, France

G. Trinchieri, National Cancer Institute, Bethesda, Maryland: Microbiota in cancer and cancer therapy.

L. Zitvogel, Gustave Roussy, Villejuif, France: Gut microbiome and tumor immunosurveillance.

**SESSION 6:** Systems Understanding of Therapeutic Immunity

**Chairperson: D. Pe'er,** Sloan Kettering Institute, New York, New York

D. Pe'er, Sloan Kettering Institute, New York, New York: A single-cell lens into immune eco-systems.



R. Ahmed, R. Seder, I. Mellman



B. Brown, A. Kamphorst

J. Heath, Institute for Systems Biology, Seattle, Washington: New single-cell methods and algorithms for cancer immunotherapy applications.

B. Greenbaum, Icahn School of Medicine at Mount Sinai, New York: Quantifying the emergence of non-self in tumors.

**SESSION 7:** Beyond Checkpoint Inhibitors

**Chairperson: M. Merad,** Icahn School of Medicine at Mount Sinai, New York

I. Mellman, Genentech, South San Francisco, California: Cancer immunotherapy beyond checkpoint inhibitors.

**SESSION 8:** Meeting Conclusions and Next Steps

**Chairpersons: M. Merad,** Icahn School of Medicine at Mount Sinai, New York; **I. Mellman,** Genentech, South San Francisco, California

# Integrated Control of Feeding and Energy Balance by Hypothalamic and Hindbrain Circuits

This meeting convened experts to critically review recent findings about the hypothalamic and hindbrain systems that control food intake, their integration, and the mechanisms by which these mediate aversive and nonaversive anorexia and set overall energy balance. In addition to identifying important questions and next steps in research, participants considered how systems that convey information about taste, illness, and other contextual information interact with these circuits to contribute to the acute and long-term control of food intake and body weight.

**Welcoming Remarks:  R. Leshan,** Director, Banbury Center, Cold Spring Harbor Laboratory

**Introduction and Meeting Objectives:  L. Heisler,** University of Aberdeen, United Kingdom
                                          **M. Myers,** University of Michigan, Ann Arbor

**SESSION 1:** Overview and Vagus

**Chairperson: L. Heisler,** University of Aberdeen, United Kingdom

H. Grill, University of Pennsylvania, Philadelphia: Neural mechanisms of feeding inhibition.

I. de Araujo, Icahn School of Medicine at Mount Sinai, New York: The vagus nerve and reward circuits.

S. Appleyard, Washington State University, Pullman: Modulation of the vagal-NTS synapse: changing how the gut talks to the brain.

T. Tan, D. Williams, S. Trapp, R.A. Travagli



D. Sandoval, A. Ferrante

A. Travagli, Pennsylvania State University, Hershey: Brainstem control of GI motility.

R. Seeley, University of Michigan, Ann Arbor: The role of GFRAL neurons in the regulation of energy balance.

**SESSION 2:** The Nucleus Tractus Solitarius

**Chairperson: H. Grill,** University of Pennsylvania, Philadelphia

R. Ritter, Washington State University, Pullman: Central modulation of vagal afferent signaling in hindbrain to control of food intake.

S. Trapp, University College London, United Kingdom: The role of GLP-1-producing brainstem neurons in the control of food intake.

G. D'Agostino, University of Aberdeen, United Kingdom: Caudal brainstem CCKergic circuits: mapping, deconstruction, and functional interrogation.

**SESSION 3:** The Gut/Brain Axis and Taste

**Chairperson: N. Thornberry,** Kallyope Inc., New York, New York

H. Lee, Northwestern University, Evanston, Illinois: Wiring the taste system.

D. Sandoval, University of Michigan, Ann Arbor: The role of the gut/brain axis in the success of bariatric surgery.

P. Di Lorenzo, Binghamton University, New York: Effects of obesity and gastric bypass surgery on the neural code for taste in the nucleus of the solitary tract.

D. Small, Yale University, New Haven, Connecticut: Metabolic drivers of oral sensation and food reinforcement.

**SESSION 4:** Hindbrain Neurons as Drug Targets

**Chairperson: M. Myers,** University of Michigan, Ann Arbor

D. Williams, Florida State University, Tallahassee: Brain GLP-1 integrates satiety, reward, and response to stress.

L. Heisler, University of Aberdeen, United Kingdom: Targeting 5-HT2C receptors in the NTS to influence food intake.

**SESSION 5:** Pharmacotherapies

**Chairperson: A. McElvaine,** American Diabetes Association, Arlington, Virginia

T. Tan, Imperial College London, United Kingdom: Gut hormones in combination: effects on appetite and metabolism.

M. Tschöp, Helmholtz Center, Neuherberg, Germany: Neuroendocrine polypharmacy targeting obesity and diabetes.

**SESSION 6:** Nausea, Aversion, and Links to the Forebrain

**Chairperson: R. Seeley,** University of Michigan, Ann Arbor

S. Luckman, The University of Manchester, United Kingdom: Gut–brain signaling: satiety and aversion.

C. Campos, University of Washington, Seattle: Chronic activation of hypothalamic AgRP neurons does not result in long-term hyperphagia and weight gain.

M. Myers, University of Michigan, Ann Arbor: Aversive and nonaversive hindbrain satiety circuits.

C. Blouet, University of Cambridge, United Kingdom: Leucine engages a hindbrain-to-forebrain neurocircuit to rapidly inhibit appetite and produce physiological satiety.

**SESSION 7:** Hypothalamic Aspects

**Chairperson: C. Blouet,** University of Cambridge, United Kingdom

A. Garfield, Rhythm Pharmaceuticals, Boston, Massachusetts: An MC4R agonist, setmelanotide, for the treatment of rare genetic disorders of obesity.

M. Schwartz, University of Washington, Seattle: Perineuronal nets and the hypothalamic feeding circuits they enmesh.

S. Panda, R. Ritter, I. de Araujo



D. Small, I. de Araujo

**SESSION 8:** Integrative Issues

**Chairperson: D. Sandoval,** University of Michigan, Ann Arbor

A. Ferrante, Columbia University Irving Medical Center, New York, New York: Defense against weight gain.

S. Panda, Salk Institute for Biological Studies, La Jolla, California: Interaction between diet quality and daily eating: fasting rhythms in health and disease.

**SESSION 9:** Meeting Wrap-Up

**Chairpersons: L. Heisler,** University of Aberdeen, United Kingdom; **M. Myers,** University of Michigan, Ann Arbor

# Glioblastoma: Why Is Impactful Science So Hard to Translate?

Glioblastoma (GBM) is the most common primary brain tumor and among the most lethal of cancers. Although scientific understanding and impact have been formidable over the past decade, the clinical translation of these insights remains disappointing. This meeting convened experts to discuss our current understanding of GBM biology and therapy, with an emphasis on identifying bottlenecks limiting the ability to successfully translate basic science discoveries into clinical care and on developing approaches to improve this bench-to-bedside transition.

**Welcoming Remarks:**  **R. Leshan,** Director, Banbury Center, Cold Spring Harbor Laboratory
                                **P. Dirks,** The Hospital for Sick Children, Toronto, Canada
                                **E. Maher,** UT Southwestern Medical Center, Dallas
                                **W. Weiss,** University of California, San Francisco

**SESSION 1:** Genomics and Epigenomics

**Chairperson: R. Verhaak,** The Jackson Laboratory, Farmington, Connecticut

R. Verhaak, The Jackson Laboratory, Farmington, Connecticut: Adaptive changes of glioma during treatment.

P. Mischel, University of California, San Diego/Ludwig Institute for Cancer Research, La Jolla: In cancer, as in real estate, location matters: the role of extrachromosomal oncogene amplification in glioblastoma.

M. Suvà, Massachusetts General Hospital, Charlestown: Deciphering human gliomas by single-cell genomics.

B. Bernstein, J. Phillips, A. Mills, O. Becher



M. Suvà, E. Maher

J. Costello, University of California, San Francisco: GABP, a key to tumor cell immortality in TERT promoter mutant tumors.

B. Bernstein, Massachusetts General Hospital, Boston: Epigenetic mechanisms and therapeutic opportunities in glioblastoma.

**SESSION 2:** Pediatric Glioblastoma and DIPG

**Chairperson: N. Jabado,** McGill University Health Centre, Montreal, Quebec, Canada

M. Taylor, The Hospital for Sick Children, Toronto, Ontario, Canada: Childhood cerebellar tumors mirror conserved fetal transcriptional programs.

N. Jabado, McGill University Health Centre, Montreal, Quebec, Canada: Exploiting vulnerabilities generated by histone mutation.

M. Filbin, Dana-Farber Cancer Institute, Boston, Massachusetts: Resolving the developmental origins of pediatric high-grade gliomas in single cells.

M. Monje, Stanford University, California: Neuronal activity drives high-grade glioma growth.

O. Becher, Northwestern University, Chicago, Illinois: Learning from genetic mouse modeling of diffuse intrinsic pontine glioma.

**SESSION 3:** Stem Cells and Glioblastoma

**Chairperson: P. Dirks,** The Hospital for Sick Children, Toronto, Ontario, Canada

J. Rich, University of California, San Diego, La Jolla: Brain tumor stem cells.

S. Pollard, Cancer Research UK Edinburgh, Edinburgh, United Kingdom: Regulation of neural stem-cell self-renewal mechanisms in glioblastoma.

P. Dirks, The Hospital for Sick Children, Toronto, Ontario, Canada (on behalf of Robert Bachoo): Transcriptional networks drive gliomagenesis.

A. Demopoulos, Northwell Health, Lake Success, New York: Role of subventricular zone in gliomagenesis.

**SESSION 4:** Microenvironment

**Chairperson: W. Weiss,** University of California, San Francisco

J. Phillips, University of California, San Francisco: GBM heterogeneity and extracellular regulation of oncogenic signaling.

S. Parrinello, University College London, United Kingdom: Microenvironmental regulation of glioblastoma invasion.



M. Monje, N. Jabado, S. Pollard, M. Filbin, G. Bergers, E. Maher



M. Gilbert, W. Kaelin

M. Symons, Feinstein Institute for Medical Research, Manhasset, New York: Overcoming glioblastoma intratumor heterogeneity and therapeutic resistance using nanoparticle-mediated delivery of miR-34a.

G. Bergers, VIB-KU Leuven Center for Cancer Biology, Belgium: Studying and targeting the tumor microenvironment in intra-heterogeneous glioblastoma.

**SESSION 5:** Modeling

**Chairperson: L. Parada,** Memorial Sloan Kettering Cancer Center, New York, New York

L. Parada, Memorial Sloan Kettering Cancer Center, New York, New York: Mouse models of GBM: cancer stem cells and therapeutic opportunities.

E. Holland, University of Washington, Seattle: Big data and mouse models.

W. Kaelin, Dana-Farber Cancer Institute/HHMI, Boston, Massachusetts: Targeting IDH mutant gliomas.

Y. Li, University of Toronto, Ontario, Canada: Modeling human neural development and diseases in neurons and brain organoids.

**SESSION 6:** Clinical

**Chairperson: E. Maher,** UT Southwestern Medical Center, Dallas

M. Schulder, Zucker School of Medicine at Hofstra/Northwell, Lake Success, New York: Intraoperative imaging and the impact of surgical resection in patients with glioblastoma.

J. Boockvar, Lenox Hill Hospital/Zucker School of Medicine at Hofstra/Northwell, New York: New strategies to overcome the blood–brain barrier to deliver chemotherapeutics in human GBM.

D. Haas-Kogan, Dana-Farber Cancer Institute, Boston, Massachusetts: Challenges to rational incorporation of novel agents into multimodality therapy of pediatric gliomas.

M. Gilbert, National Cancer Institute, Bethesda, Maryland: Challenges in implementing correlative biology in brain tumor clinical trials.

**SESSION 7:** Meeting Wrap-Up

**Chairpersons: P. Dirks,** The Hospital for Sick Children, Toronto, Ontario, Canada; **E. Maher,** UT Southwestern Medical Center, Dallas; **W. Weiss,** University of California, San Francisco

# The Plant Microbiota

**April 14–17**

ARRANGED BY      **J. Dangl,** University of North Carolina/HHMI, Chapel Hill
                                 **P. Schulze-Lefert,** Max Planck Institute for Plant Breeding Research, Cologne, Germany
                                 **J. Vorholt,** ETH Zürich, Switzerland

Growing interest in the plant microbiota is being spurred by basic, curiosity-driven research that seeks to understand the principles underlying microbiota assembly and the impact of the microbiota on the plant host. There is also an increasing awareness that knowledge gained can be harnessed for rational bioprospecting and the discovery of agriculturally useful molecules, genes, and inoculants in plant-associated microbes. This Banbury meeting gathered plant and microbial scientists to discuss the latest advances in (1) microbiota assembly and the plant innate immune system, (2) host colonization and mechanisms of interbacterial communication, (3) metabolic interdependence of the plant host and its associated microbes, (4) nutrient mobilization and nutritional functions provided by root-associated microbes, (5) microbial interkingdom interactions and microbiota homeostasis, (6) invasion and persistence in microbial consortia, and (7) commensal functions in pathogen protection.

**Welcoming Remarks:  R. Leshan,** Director, Banbury Center, Cold Spring Harbor Laboratory

**Introduction and Meeting Objectives:  P. Schulze-Lefert,** Max Planck Institute for Plant Breeding Research, Cologne, Germany

D. Weigel, J. Vorholt, E. Kemen



S.Y. He, P. Schulze-Lefert

**SESSION 1:** Big Picture Microbiomes

**Chairperson: J. Vorholt,** ETH Zürich, Switzerland

S. Tringe, DOE Joint Genome Institute, Walnut Creek, California: Sequence-based interrogation of plant microbiomes.

V. Sunderesan, University of California, Davis: Host–microbiome interactions in rice roots.

M. Wagner, University of Kansas, Lawrence: Patterns and consequences of breeding-induced microbiome variation in maize.

S. Kembel, University of Quebec at Montreal, Canada: The biogeography of phyllosphere plant–microbe associations.

D. Weigel, Max Planck Institute, Tübingen, Germany: Coexistence of *Arabidopsis thaliana* and *Pseudomonas foliar* pathogens over different temporal and spatial scales.

**SESSION 2:** Phyllosphere

**Chairperson: S. Tringe,** DOE Joint Genome Institute, Walnut Creek, California

J. Vorholt, ETH Zürich, Switzerland: The leaf microbiota: disassembling and rebuilding to explore plant microbe interactions.

E. Kemen, University of Tübingen, Germany: Understanding dynamics in leaf microbial communities.

B. Wolfe, Tufts University, Medford, Massachusetts: Linking patterns with processes in the Napa cabbage phyllosphere.

**SESSION 3:** Abiotic Stress and the Microbiome

**Chairperson: E. Kemen,** University of Tübingen, Germany

S. Lebeis, University of Tennessee, Knoxville: Under pressure: surviving selective pressure in the root-soil interface.

D. Coleman-Derr, USDA/ARS, Albany, California: The effects of drought and development on the root microbiome.

A. Shade, Michigan State University, East Lansing: Resuscitation and recruitment of rhizosphere microbiota during plant stress.

**SESSION 4:** Metabolic Control of Microbiome Function

**Chairperson: S. Hacquard,** Max Planck Institute for Plant Breeding Research, Cologne, Germany

K. Schläppi, University of Bern, Switzerland: Plant secondary metabolites drive rhizosphere microbiome traits.



S. Tringe, K. Schläppi



P. Teixiera, V. Sundaresan

**SESSION 5:** Plant Immune System and the Microbiome

**Chairperson: S. Lebeis,** University of Tennessee, Knoxville

P. Teixeira, University of North Carolina, Chapel Hill: Colonization of *Arabidopsis* roots by a bacterial microbiome is associated with suppression of a specific sector of the plant immune response.

C. Haney, University of British Columbia, Vancouver, Canada: Mechanisms in microbial regulation of plant growth and defense.

S. Yang He, Michigan State University/HHMI, East Lansing: Genetic control of dysbiosis in *Arabidopsis.*

P. Schulze-Lefert, Max Planck Institute for Plant Breeding Research, Cologne, Germany: Strain-specific interference of root commensals with the plant innate immune system.

C. Pieterse, Utrecht University, the Netherlands: The root microbiome and plant immunity: an IRONic love story.

**SESSION 6:** Multi-Kingdom Interactions

**Chairperson: P. Schulze-Lefert,** Max Planck Institute for Plant Breeding Research, Cologne, Germany

B. Koskella, University of California, Berkeley: The role of phyllosphere bacteria and bacteriophage viruses in shaping pathogen colonization and disease of host plants.

L. Kinkel, University of Minnesota, Saint Paul: Cross-kingdom microbial interactions in plant microbiome systems.

S. Hacquard, Max Planck Institute for Plant Breeding Research, Cologne, Germany: Structural and functional architectures of multi-kingdom microbial consortia colonizing plant roots.

**SESSION 7:** General Discussion and Meeting Wrap-Up

**Chairpersons: J. Vorholt,** ETH Zürich, Switzerland; **Paul Schulze-Lefert,** Max Planck Institute for Plant Breeding Research, Cologne, Germany

# Intermediate Indicators for Impact: The Art and Science of Effective Definition and Use of Prevention Indicators in the HIV Response

This workshop brought together those experienced with collecting and analyzing data with those responsible for using the knowledge generated, as well as those affected by the programs, to explore how indicators and metrics can be optimized to impact the HIV epidemic over the coming decades.

**Welcoming Remarks:  R. Leshan,** Director, Banbury Center, Cold Spring Harbor Laboratory

**Introduction and Meeting Objectives:  C. Holmes,** Georgetown University, Washington, D.C.
                                             **N. Kilonzo,** National AIDS Control Council of Kenya, Nairobi
                                           **M. Mahy,** UNAIDS, Geneva, Switzerland

**SESSION 1:** Setting the Stage: What Should We be Measuring? Overview of Indicator Development and Use

**Chairpersons: G. Dallabetta,** Bill & Melinda Gates Foundation, Washington, D.C.; **M. Morrison,** Bill & Melinda Gates Foundation, Seattle, Washington

C. Benedikt, UNAIDS, Geneva, Switzerland: HIV prevention: setting the stage on our understanding of prevention.

N. Kilonzo, National AIDS Control Council of Kenya, Nairobi: Overview of development and use of prevention indicators, country perspective.

P. Bhattacharjee, B. Rice, M. Marston



I. Benach, C. Benedikt, M. Mahy, M. Khasiani

M. Mahy, UNAIDS, Geneva, Switzerland: Indicator development, global stakeholder perspective.

**SESSION 2:** Current State of M&E in Prevention

**Chairpersons: G. Dallabetta and M. Morrison,** Bill & Melinda Gates Foundation, Seattle, Washington

B. Rice, London School of Hygiene & Tropical Medicine, London, United Kingdom: What does the prevention M&E landscape look like?

J. DePasse, Boston Consulting Group, Seattle Washington: HIV prevention indicators: variability between global stakeholders and countries.

J. Zhao, The Global Fund, Geneva, Switzerland: Measurement of HIV programs supported by The Global Fund.

G. Garnett, Bill & Melinda Gates Foundation, Seattle, Washington: Prevention cascade: use of cascade in action.

**SESSION 3:** Learnings from Local Experiences with Indicators

**Chairperson: N. Kilonzo,** National AIDS Control Council of Kenya, Nairobi

T. Kalua, Malawi Ministry of Health, Lilongwe: Case study: local challenges in indicator development.

M. Khasiani, National AIDS Control Council of Kenya, Nairobi: The complementing role of non-health facility indicators in HIV prevention: experiences from community and public sector reporting.

**BREAKOUT SESSION I:** What Are the Enablers to Overcome the Challenges of Data Collection and Valid Responses for HIV Prevention Indicators?

**Group 1 Facilitator: S. Johnson,** Genesis Analytics, Johannesburg, South Africa: Sexual behavior, patterns of condom use.

**Group 2 Facilitator: P. Bhattacharjee,** University of Manitoba, Nairobi, Kenya: Key populations (including size estimates).

**SESSION 4:** Unique Considerations for Prevention Metrics in HIV

**Chairperson: G. Garnett,** Bill & Melinda Gates Foundation, Seattle, Washington



R. Baggaley, N. Hasan, J. DePasse



J. Zhao, C. Holmes, M. Morrison

R. Baggaley, World Health Organization, Geneva, Switzerland: Monitoring prevention efforts in other fields: unique challenges of HIV prevention metrics.

J. Eaton, Imperial College London, United Kingdom; L. Johnson, University of Cape Town, South Africa: Modeling impact of new indicators in HIV.

M. Warren, AVAC, New York, New York: Unique considerations for prevention metrics in HIV: complexity with data collection, analysis, utilization, etc.

**SESSION 5:** Evidence behind Current Metrics in HIV Prevention

**Chairperson: G. Garnett,** Bill & Melinda Gates Foundation, Seattle, Washington

C. Holmes, Georgetown University, Washington, D.C.; B. Rice, London School of Hygiene & Tropical Medicine, United Kingdom: Understanding the evidence behind current HIV prevention metrics.

M. Marston, London School of Hygiene & Tropical Medicine, United Kingdom: Prevention metrics: highlights/findings from the alpha network.

**WORKING SESSION:** Metrics Prioritization

**Facilitators: C. Holmes,** Georgetown University, Washington, D.C.; **N. Kilonzo,** National AIDS Control Council of Kenya, Nairobi; **M. Mahy,** UNAIDS, Geneva, Switzerland

**SESSION 6:** Emerging Metrics and Methods for HIV Prevention

**Chairperson: J. Eaton,** Imperial College London, United Kingdom

N. Hasen, Population Services International, Washington, D.C.: Measuring, monitoring, and tracking sexual networks as a key to HIV prevention metrics.

C. Ryan, Centers for Disease Control and Prevention, Eswatini, Africa: New horizons in HIV prevention: use and potential directions of recency testing.

T. Smith, Cooper/Smith, Washington, D.C.: Novel data streams and triangulation methods for identifying risk, location target populations over time and space, and differentiating HIV prevention and care.

**BREAKOUT SESSION II:** Recommendations and Socialization

**Group 1 Facilitators: C. Holmes,** Georgetown University, Washington, D.C.; **N. Kilonzo,** National AIDS Control Council of Kenya, Nairobi; **M. Mahy,** UNAIDS, Geneva, Switzerland: Global stakeholder engagement.

**Group 2 Facilitators: C. Holmes,** Georgetown University, Washington, D.C.; **N. Kilonzo,** National AIDS Control Council of Kenya, Nairobi; **M. Mahy**, UNAIDS, Geneva, Switzerland: Country stakeholder engagement.

**Group 3 Facilitators: C. Holmes,** Georgetown University, Washington, D.C.; **N. Kilonzo,** National AIDS Control Council of Kenya, Nairobi; **M. Mahy,** UNAIDS, Geneva, Switzerland: Academic future directions.

**SESSION 7:** Closing and Next Steps

**Chairpersons: C. Holmes,** Georgetown University, Washington, D.C.; **N. Kilonzo,** National AIDS Control Council of Kenya, Nairobi; **M. Mahy,** UNAIDS, Geneva, Switzerland

# Rita Allen Foundation Scholars Symposium

Banbury was pleased to welcome the Rita Allen Foundation for their first annual meeting of Rita Allen Foundation Scholars in 2019, which included current and former Scholars and other scientific leaders. The symposium generated lively discussions on cancer, neuroscience, immunology, and pain research, as well as on diversity and inclusion in the sciences and Rita Allen Foundation's work to connect science and society.

**Welcoming Remarks:** **E. Christopherson,** President and CEO, Rita Allen Foundation, Princeton, New Jersey; **B. Stillman,** President and CEO, Cold Spring Harbor Laboratory and Rita Allen Scholar, New York

**Keynote Address: From Mice to Molecules:** **The Genetics,** Development, and Function of Tails **H. Hoekstra,** Harvard University, Cambridge, Massachusetts

**Introduced by E. Gracheva,** Yale University School of Medicine, New Haven, Connecticut

**FLASH TALKS**

**Chairperson: R. Sharif-Naeini,** McGill University, Montreal, Quebec, Canada
M. Banghart, University of California, San Diego, La Jolla
M. Burton, University of Texas at Dallas, Richardson

J. Clowney, University of Michigan, Ann Arbor
P. Grace, University of Texas MD Anderson Cancer Center, Houston
P. Greer, University of Massachusetts Medical School, Worcester

H. Lai, UT Southwestern Medical Center, Dallas
V. Luca, Moffitt Cancer Center, Tampa, Florida
J. McCall, Washington University in St. Louis, Missouri
L. O'Connell, Stanford University, California
J. Parker, California Institute of Technology, Pasadena

**Keynote Address:** The Will and the Ways
S. Zárate, Howard Hughes Medical Institute, Washington, D.C.

**Introduced by D. Fiedler,** Leibniz-Institute for Molecular Pharmacology, Berlin, Germany

**FLASH TALKS AND SCHOLAR TALKS**

**Chairperson: R. Seal,** University of Pittsburgh, Pennsylvania
C. Paulsen, Yale University, New Haven, Connecticut
V. Tawfik, Stanford University, California
L. Zhao, The Rockefeller University, New York, New York
M. Boyce, Duke University, Durham, North Carolina: Cell signaling through protein glycosylation.
M. Hammell, Cold Spring Harbor Laboratory: Retrotransposon reactivation in neurodegenerative diseases.
K. Baumbauer, University of Kansas Medical Center, Kansas City: Persistence of pain following spinal cord injury may be established in the acute phase of injury.

**Keynote Address:** Co-evolution of DNA Replication Origin Specificity and Transcriptional Gene Silencing.
B. Stillman, Cold Spring Harbor Laboratory

Introduced by M. Boyce, Duke University, Durham, North Carolina

**SCHOLAR TALKS**

**Chairperson: E. Gracheva,** Yale University School of Medicine, New Haven, Connecticut

D. Fiedler, Leibniz-Institute for Molecular Pharmacology, Berlin, Germany: Inositol pyrophosphate signaling revealed with chemical tools.
K. Hanlon, Presbyterian College School of Pharmacy, Clinton, South Carolina: Neuroprotective role of macrophages within dorsal root ganglia.
R. Daneman, University of California, San Diego, La Jolla: Blood–brain barrier regulation of brain function and behavior.
K. Meyer, Duke University School of Medicine, Durham, North Carolina: Detecting mRNA methylation and its role in gene expression regulation.

**SCHOLAR TALKS**

**Chairperson: W. Greenleaf,** Stanford University, California
E. Gracheva, Yale University School of Medicine, New Haven, Connecticut: Molecular adaptations to the unique lifestyle in mammalian hibernators.
Y. Kozorovitskiy, Northwestern University, Evanston, Illinois: A neuromodulatory meta-plasticity hypothesis for rapidly acting antidepressant effects.
B. Li, University of North Carolina, Chapel Hill: Discovering bacterial metabolites that modulate host biology.



K. Foley, L. O'Connell

H. Hoekstra, S. Zárate



P. Grace, M. Burton, M. Banghart

## SCHOLAR TALKS

**Chairperson: M. Boyce,** Duke University, Durham, North
Carolina

R. Sharif-Naeini, McGill University, Montreal, Quebec,
Canada: TACAN: a novel ion channel necessary for pain
sensing.

J. Wilusz, University of Pennsylvania, Philadelphia: Unexpected
mechanisms that control the outputs of protein-coding genes.

## SCHOLAR TALKS

**Chairperson: R. Seal,** University of Pittsburgh, Pennsylvania

L. Ding, Columbia University, New York, New York: Under-
standing the fetal liver hematopoietic stem-cell niche.

S. Davidson, University of Cincinnati, Ohio: Using viable
human neural tissues to improve validity for pain medicine.

K. Schlacher, University of Texas MD Anderson Cancer
Center, Houston: DNA replication instability in disease,
cancer, and inflammation.

A. Khoutorsky, McGill University, Montreal, Quebec,
Canada: Remodeling of spinal extracellular matrix
modulates the development of pain hypersensitivity.

**Keynote Address:** How to Succeed in Your Academic Career.
K. Davidson, Northwell Health, New York

**Introduced by R. Sharif-Naeini,** Montreal, Quebec, Canada

## CLOSING REMARKS

**E. Christopher,** Rita Allen Foundation, Princeton, New Jersey

# Liquid Biopsies

Liquid biopsies, the analysis of cells and nucleic acids from blood samples, are poised to create major impact for cancer care. Before liquid biopsies can become commonplace in the clinic, however, several issues will need resolution. This Banbury discussion meeting convened experts and stakeholders best positioned to tackle these challenges in order to stimulate collaboration and identify potential solutions.

**Welcoming Remarks:** **R. Leshan,** Director, Banbury Center, Cold Spring Harbor Laboratory
                               **M. Cleary,** CEO, The Mark Foundation for Cancer Research, New York, New York

**Introduction and Meeting Objectives:** **L. Diaz,** Memorial Sloan Kettering Cancer Center, New York, New York
                                               **V. Velculescu,** Johns Hopkins University, Baltimore, Maryland

**SESSION 1:** Approaches and Technology

**Chairperson: V. Velculescu,** Johns Hopkins University, Baltimore, Maryland

M. Murtaza, Translational Genomics Research Institute, Phoenix, Arizona: Improving accuracy and precision for liquid biopsies using personalized targeted digital sequencing.

L. Diaz, M. Cleary, V. Velculescu, E. Heitzer, C.J. Lin



V. Anagnostou, G. Meijer

A. Moffitt, Cold Spring Harbor Laboratory: Sensitive and quantitative detection of patient-specific cancer variants in blood.

K-T. Varley, Huntsman Cancer Institute, University of Utah, Salt Lake City: Targeted sequencing of mutations and methylation in ctDNA using patch capture.

D. Landau, Weill Cornell Medicine/New York Genome Center, New York: Genome-wide mutational integration for ultrasensitive ctDNA detection.

E. Heitzer, Medical University of Graz, Graz, Austria: Moving beyond DNA sequence: nucleosome occupancy profiling of plasma DNA.

R. Scharpf, Johns Hopkins University, Baltimore, Maryland: Modeling fragmentation patterns of cell-free DNA.

**SESSION 2:** Circulating Tumor Cells and Other Sample Sources

**Chairperson: R. Bish,** The Mark Foundation for Cancer Research, New York

C. Alix-Panabières, University Medical Centre of Montpellier, Montpellier, France: Functional study of circulating tumor cells.

K. Pantel, University Medical Center Hamburg-Eppendorf, Hamburg, Germany: Clinical application of circulating tumor cells as liquid biopsy in cancer patients.

M. Cleary and Ryan Schoenfeld, The Mark Foundation for Cancer Research, New York, New York: Announcement of liquid biopsies RFP.

**SESSION 3:** Biomarker Applications

**Chairperson: L. Diaz,** Memorial Sloan Kettering Cancer Center, New York, New York

J. Phallen, Johns Hopkins University School of Medicine, Baltimore, Maryland: Circulating tumor DNA as a biomarker for cancer.

R. Fijneman, Netherlands Cancer Institute, Amsterdam, the Netherlands: Circulating tumor DNA as a biomarker in colorectal cancer: turning research into care.

C. Andersen, Aarhus University Hospital, Denmark: Investigations of the potential clinical utility of minimally invasive circulating tumor DNA analysis in the management of colorectal cancer.

**SESSION 4:** Treatment Monitoring

**Chairperson: R. Schoenfeld,** The Mark Foundation for Cancer Research, New York, New York

V. Anagnostou, Johns Hopkins University School of Medicine, Baltimore, Maryland: Liquid biopsy approaches for rapid determination of response to immune checkpoint blockade.

M. Sausen, Bristol-Myers Squibb, Pennington, New Jersey: Noninvasive detection of microsatellite instability and high tumor mutation burden in cancer patients treated with PD-1 blockade.

A. Leal, Johns Hopkins University, Baltimore, Maryland: Matched white blood cell and cell-free DNA analyses for the detection of minimal residual disease in patients with cancer.



K-T. Varley

**SESSION 5:** Clinical Translation

**Chairperson: N. Dracopoli,** Delfi Diagnostics, Washington, D.C.

M. Berger, Memorial Sloan Kettering Cancer Center, New York, New York: Clinical ctDNA profiling to guide treatment selection.

E. Carpenter, University of Pennsylvania School of Medicine, Philadelphia: Clinical use of liquid biopsy for the management of advanced non-small-cell lung cancer.

J. Lin, Freenome, South San Francisco, California: Use of tumor-informed ctDNA for minimal residual disease testing and monitoring in lung, breast, colorectal, and bladder cancers.

H. Jørgen Nielsen, University of Copenhagen/Hvidovre Hospital, Copenhagen, Denmark: Development of blood-based cancer screening concepts.

G. Meijer, Netherlands Cancer Institute, Amsterdam, the Netherlands: ctDNA on the way to implementation in the Netherlands.

# Communicating Science—Boehringer Ingelheim Fellows Retreat

**September 13–18**

ARRANGED BY  **K. Achenbach,** Boehringer Ingelheim Fonds, Mainz, Germany
**S. Schedler,** Boehringer Ingelheim Fonds, Mainz, Germany
**C. Walther,** Boehringer Ingelheim Fonds, Mainz, Germany

FUNDED BY  **Boehringer Ingelheim Fonds**

The Boehringer Ingelheim Fonds has an international fellowship program supporting outstanding Ph.D. students. Among the opportunities provided to fellows is rigorous training in communication through an annual retreat. It was a great pleasure to have them return in 2019 for interactive instruction in matters such as oral presentations and writing papers; this year's retreat marked the 11th such visit to Banbury.

**Opening Remarks:  C. Walther,** Boehringer Ingelheim Fonds, Mainz, Germany

**Introduction to the Bottom Line:  A. Katsnelson,** Science writer and editor, Northampton, Massachusetts: Ground rules for writing to be read and understood.

**W. Tansey,** Vanderbilt University, Nashville, Tennessee: Preparing and delivering a scientific talk.

**Deadline Writing Assignment 1: PowerPoint Presentations**

**A. Katsnelson,** Science writer and editor, Northampton, Massachusetts: Discussion and questions on writing assignment 1.

**Deadline Writing Assignment 2: PowerPoint Presentations**

**A. Katsnelson,** Science writer and editor, Northampton, Massachusetts: On cover letters and how to deal with editor/reviewer comments; Discussion of writing assignment 2.

**T. Janowitz,** Cold Spring Harbor Laboratory: Bench to bedside and back again: reflections of a physician scientist.

**M. Krzywinski,** BC Cancer Agency, Vancouver, British Columbia, Canada: Design of scientific concept and data figures.

**C. Walther,** Boehringer Ingelheim Fonds, Mainz, Germany: All about BIF.

# Bridging the Research-to-Practice Chasm in Digital Mental Health

**October 6−8**

ARRANGED BY
**P. Areán,** University of Washington, Seattle
**D. Mohr,** Northwestern University, Chicago, Illinois

FUNDED BY
**Cold Spring Harbor Laboratory Corporate Sponsor Program; Microsoft**

Digital mental health (DMH) interventions using web and mobile technologies have consistently demonstrated effectiveness in more than 100 randomized controlled trials conducted over two decades. This Banbury meeting established an interdisciplinary work group to define the path toward successful, sustainable DMH implementation. Participants (1) outlined the grand challenges facing digital mental health implementation; (2) identified short- (one to three years) and mid-range (three to five years) goals that can move us toward sustainable implementation; and (3) identified immediate tasks (6–12 months) that participants agreed to that will move the field of digital mental health forward.

**Welcoming Remarks:  R. Leshan,** Director, Banbury Center, Cold Spring Harbor Laboratory

**Introduction and Meeting Objectives:  P. Areán,** University of Washington, Seattle
**D. Mohr,** Northwestern University, Chicago, Illinois

J. Torous, M. De Choudhury



C. Hartwell, N. Leibowitz

**SESSION 1:** Perspectives I: Healthcare Systems

**Chairperson: P. Areán,** University of Washington, Seattle

H. Harbin, consultant, Baltimore, Maryland: Expanding access to digital behavioral interventions.

A. Bertagnolli, Optum Behavioral Health, San Francisco, California: Digital mental health: implementation challenges.

T. Histon, Kaiser Permanente, Oakland, California: Project Chamai: digital mental health tools to support emotional wellness needs.

M. Cunningham-Hill, Northeast Business Group on Health, New York: Leveraging digital mental health to increase access in the employer space.

F. Azocar, Optum, San Francisco, California: An assessment tool for the evaluation of iCBT programs for use in a managed behavioral health organization.

**SESSION 2:** Perspectives II: Public Health

**Chairperson: D. Mohr,** Northwestern University, Chicago, Illinois

N. Titov, Mindspot Clinic, Macquarie University, Sydney, New South Wales, Australia: Delivering digital mental health services across Australia: challenges, lessons, and opportunities.

S. Schueller, University of California, Irvine: Implementation of digital mental health across California: the Help@Hand Project.

T. Nguyen, Mental Health America, Alexandria, Virginia: Early digital interventions for individuals with untreated mental health.

**SESSION 3:** Perspectives III: Company Perspectives

**Chairperson: P. Areán,** University of Washington, Seattle

N. Leibowitz, Talkspace, New York, New York: Digital mental health startup challenges: from research to implementation.

D. Richards, SilverCloud Health, Dublin, Ireland: Implementing iCBT into routine care: the SilverCloud experience.

C. Hartwell, Bridge Builders Collaborative, St. Paul, Minnesota: The investment landscape in mental wellness.

**SESSION 4:** Perspectives IV: Users

**Chairperson: D. Mohr,** Northwestern University, Chicago, Illinois

M. Czerwinski, Microsoft Research, Redmond, Washington: Using technology for health and wellbeing.

P. Areán, University of Washington, Seattle: Use of human-centered design in development of technology tools for research and practice.

D. Mohr, Northwestern University, Chicago, Illinois: Design for care managers and health systems.

T. Choudhury, Cornell University/HealthRhythms, Ithaca, New York: Closing the sensing-to-intervention loop for behavioral health.

**SESSION 5:** Harnessing Technological Affordances

**Chairperson: P. Areán,** University of Washington, Seattle

M. De Choudhury, Georgia Institute of Technology, Atlanta: Challenges and opportunities of social media.



T. Histon, D. Richards, P. Chrisp

J. Torous, Harvard Medical School, Boston, Massachusetts: Evaluation and regulation of mental health apps.

C. Nebeker, University of California, San Diego, La Jolla: Ethical, legal/regulatory, and social implications of digital mental health.

P. Chrisp, National Institute for Health and Care Excellence, Manchester, United Kingdom: Digital health technologies and the evidence ecosystem.

**SESSION 6:** Synthesis, Planning, and Next Steps

**Chairpersons: P. Areán,** University of Washington, Seattle; **D. Mohr,** Northwestern University, Chicago, Illinois

# Reconceptualizing the Challenges of Direct-to-Consumer Health Products

Health products and services are increasingly moving from the realm of medical professionals into the domain of consumers. To date, questions about safe and responsible marketing have largely remained within individual professional domains. Yet, it may be beneficial to conceptualize these questions as part of a larger social phenomenon. This Banbury meeting brought together an inter-disciplinary group of physicians, bioethicists, legal scholars, health and science policy researchers, and communications scholars to rethink the challenges of DTC health products and services.

**Welcoming Remarks:** **R. Leshan,** Director, Banbury Center, Cold Spring Harbor Laboratory

**Introduction and Meeting Objectives:** **A. Wexler,** University of Pennsylvania, Philadelphia
**L. Turner,** University of Minnesota, Minneapolis
**T. Caulfield,** University of Alberta, Edmonton, Canada

**SESSION 1:** Historical Background

**Chairperson: L. Turner,** University of Minnesota, Minneapolis

S. Woloshin, Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, New Hampshire: The rise of DTC medical marketing and services.

B. Southwell, A.M. Navar, E. Suhay, C. Funk



T. Caulfield, R. Redberg

**SESSION 2:** DTC Health Technologies: Information

**Chairperson: S. Joffe,** University of Pennsylvania, Philadelphia

C. Bloss, University of California, San Diego, La Jolla: The rise of direct-to-consumer genomics and implications for the future.

R. Redberg, University of California, San Francisco: DTC heart rate monitors.

M. Kyweluk, University of Pennsylvania, Philadelphia: An ethnography of direct-to-consumer ovarian reserve testing.

**SESSION 3:** DTC Health Technologies: Interventions

**Chairperson: P. Zettler,** Ohio State University, Columbus

L. Turner, University of Minnesota, Minneapolis: Direct-to-consumer marketing of unproven stem cell interventions: ethical concerns and regulatory responses.

V. Manchaiah, Lamar University, Beaumont, Texas: Direct-to-consumer hearing devices: an overview.

A. Wexler, University of Pennsylvania, Philadelphia: Studying users of direct-to-consumer neurotechnology and orthodontics.

**SESSION 4:** Public Understanding of Science

**Chairperson: M. Kyweluk,** University of Pennsylvania, Philadelphia

C. Funk, Pew Research Center, Washington, D.C.: Public perspectives on medical science.

B. Southwell, RTI International, Research Triangle Park, North Carolina: Public understanding of risk in direct-to-consumer advertising.

A. Marie Navar, Duke University, Durham, North Carolina: The dangerous intersection of DTC marketing and medical misinformation: lessons from vaccines and statins.

E. Suhay, American University, Washington, D.C.; Opportunities and risks in commercial science communication with the public.

**SESSION 5:** Government Regulation

**Chairperson: A. Wexler,** University of Pennsylvania, Philadelphia

P. Zettler, Ohio State University, Columbus: FDA and DTC health technologies.

**SESSION 6:** Alternatives to Government Regulation

**Chairperson: T. Caulfield,** University of Alberta, Edmonton, Canada

L. Brett, National Advertising Division, New York, New York: Self-regulation and advertising for direct-to-consumer health products.

B. Patten, Truth in Advertising, Madison, Connecticut: Going viral: deceptive marketing in the health and wellness industry.

A. Zarzeczny, University of Regina, Saskatchewan, Canada: Direct-to-consumer health markets and the roles of physicians: can professional regulation keep up?

**SESSION 7:** Impact on Patient–Physician Relationships

**Chairperson: L. Turner,** University of Minnesota, Minneapolis

S. Joffe, University of Pennsylvania, Philadelphia: The changing nature of autonomy in contemporary health care.

A. Levine, Georgia Institute of Technology, Atlanta: Rethinking authority and trust in the global DTC health marketplace.

**SESSION 8:** Issues on the Horizon

**Chairperson: B. Patten,** Truth in Advertising, Madison, Connecticut

S. Woloshin, B. Southwell



B. Patten, E. Suhay

J. Snyder, Simon Fraser University, Burnaby, British Columbia, Canada: Crowd control: crowdfunding for unproven DTC medical intervention.

T. Caulfield, University of Alberta, Edmonton, Canada: Marketing the microbiome: DTC and gut hype.

**SESSION 9:** Discussion, Meeting Wrap-Up, and Next Steps

**Chairpersons: A. Wexler,** University of Pennsylvania, Philadelphia; **T. Caulfield,** University of Alberta, Edmonton, Canada; **L. Turner,** University of Minnesota, Minneapolis

# Emerging Issues of Privacy, Trust, and Societal Benefit from Consumer Genomics

**October 19–22**

ARRANGED BY      **Y. Erlich,** Columbia University/MyHeritage, New York, New York
                             **A.L. McGuire,** Baylor College of Medicine

FUNDED BY         **Cold Spring Harbor Laboratory Corporate Sponsor Program; MyHeritage**

**Welcoming Remarks: R. Leshan,** Director, Banbury Center, Cold Spring Harbor Laboratory

**Introduction and Meeting Objectives: A.L. McGuire,** Baylor College of Medicine
                                                     **Y. Erlich,** Columbia University/MyHeritage

**SESSION 1:** Stakeholder Perspectives I: Companies

**Chairperson: A. McGuire,** Baylor College of Medicine

Y. Erlich, Columbia University/MyHeritage
C. Rogers, GEDmatch
C. Ball, Ancestry DNA, LLC
S. Kahn, LunaPBC
S. Elson, 23andMe
C. Bormans, Gene by Gene

**SESSION 2:** Stakeholder Perspectives II: Law Enforcement

**Chairperson: T. Callaghan,** Federal Bureau of Investigation

T. Hunt, U.S. Department of Justice
S. Kramer, Federal Bureau of Investigation
J. Shimp, Federal Bureau of Investigation
B. Budowle, Center for Human Identification
C. Fitzpatrick, Identifinders International
L. Napolitano, Florida Department of Law Enforcement
C.C. Moore, DNA Detectives/Parabon Nanolabs

**SESSION 3:** Stakeholder Perspectives III: Consumers

**Chairperson: L. Lyman Rodriguez,** Geisinger National Precision Health

C. Guerrini, Baylor College of Medicine

Y. Erlich, S. Elson



A. McGuire, C. Guerrini

B. Bettinger, The Genetic Genealogist
V. Potkin, Innocence Project
V. Eidelman, American Civil Liberties Union

L. Lyman Rodriguez, Geisinger National Precision Health
T. Callaghan, Federal Bureau of Investigation
N. Ram, University of Maryland Carey School of Law
E. Murphy, New York University
L. Brody, National Human Genome Research Institute
M. Fullerton, University of Washington School of Medicine

**SESSION 4:** When Things Go Wrong

**Chairperson: B. Wible,** Science

I. Rawlins, U.S. Navy
P. Ney, University of Washington
Y. Erlich, Columbia University/MyHeritage
E. Tromer, Columbia University/Tel Aviv University

**SESSION 6:** Synthesis and Next Steps

**Chairpersons: Y. Erlich,** Columbia University/MyHeritage;
 **A. McGuire,** Baylor College of Medicine

**SESSION 5:** Policy and Ethics

**Chairperson: Y. Erlich,** Columbia University/MyHeritage



E. Murphy, T. Callaghan, V. Potkin



B. Wible, S. Kramer, E. Tromer

# CaMKII and Its Role as a Self-Tuning Structural Protein at the Synapse

The Ca²⁺/calmodulin-dependent protein kinase CaMKII is the most abundant noncytoskeletal protein at the synapse. Despite a number of studies on CaMKII function, many unexplained findings and open questions remain. This Banbury meeting convened experts to stimulate discussion, seeded new ideas, and facilitated collaboration in ways that will accelerate new discoveries about CaMKII regulation at both the basic and translational levels.

**Welcoming Remarks:  R. Leshan,** Director, Banbury Center, Cold Spring Harbor Laboratory

**Introduction and Meeting Objectives:  Y. Hayashi,** Kyoto University, Japan
                                          **J. Hell,** University of California, Davis

**SESSION 1:** The Bigger Picture

**Chairperson: Y. Hayashi,** Kyoto University, Japan

M. Kennedy, California Institute of Technology, Pasadena: CaMKII as a central element in the biochemical network regulating excitatory synapses.

R. Yasuda, K. Zito



R. Tsien, J. Hell

H. Schulman, Stanford University School of Medicine/Panorama Research Institute, Palo Alto, California: CaMKII structure/function informs genetic and other diseases.

R. Nicoll, University of California, San Francisco: Is CaMKII a memory molecule?

R. Tsien, New York University Neuroscience Institute, New York: Function of CaMKII in Hebbian and homeostatic plasticity.

**SESSION 2:** Cell Biology

**Chairperson: H. Bito,** University of Tokyo, Japan

L. Griffith, Brandeis University, Waltham, Massachusetts: Local translation of CaMKII in *Drosophila*.

R. Colbran, Vanderbilt University, Nashville, Tennessee: Roles of CaMKII-associated proteins in CaMKII signaling.

P. De Koninck, Université Laval, Quebec, Canada: Dendritic signaling by CaMKII supporting synaptic plasticity.

R. Yasuda, Max Planck Florida Institute for Neuroscience, Jupiter: Mechanisms of CaMKII activation in single dendritic spines.

**SESSION 3:** Role in Synaptic Plasticity I

**Chairperson: J. Hell,** University of California, Davis

K. Zito, University of California, Davis: Signaling through NMDAR and CaMKII in spine structural plasticity.

U. Bayer, University of Colorado Anschutz Medical Campus, Aurora: The CaMKII/DAPK1 competition in the LTP/LTD decision and beyond.

H. Murakoshi, National Institute for Physiological Science, Okazaki, Japan: Optogenetic induction of synaptic plasticity at single synapses by photoactivatable CaMKII.

**SESSION 4:** In Vivo Perspectives

**Chairperson: K. Roche,** National Institute of Neurological Disorders and Stroke, Bethesda, Maryland

G. van Woerden, Erasmus Medical Center, Rotterdam, the Netherlands: Novel insight in the developmental role of CAMK2A in learning, memory and plasticity, and neurodevelopmental disorders.

Y. Hayashi, Kyoto University, Japan: Structural role of CaMKII.

M. Stratton, University of Massachusetts, Amherst: The mechanism of CaMKII regulation: from equilibrium to oscillations.

E. Grandi, University of California, Davis: Modeling CaMKII signaling in the heart.

**SESSION 5:** Role in Synaptic Plasticity II

**Chairperson: K. Zito,** University of California, Davis

T. Hosokawa, Kyoto University, Japan: Reconstitution of synaptic long-term potentiation in vitro.

M. Zhang, Hong Kong University of Science and Technology, China: Activity-dependent formation of CaMKII and PSD condensates via phase separation.

K. Roche, National Institute of Neurological Disorders and Stroke, Bethesda, Maryland: Regulation of synaptic NMDA receptors by CaMKII phosphorylation.



L. Griffith

M. Dell'Acqua, University of Colorado School of Medicine, Aurora: Cross-talk between CaMKII, PKA and calcineurin signaling during NMDA receptor-dependent LTD.

J. Hell, University of California, Davis: Postsynaptic signaling by CaMKII.

**SESSION 6:** Structure

**Chairperson: M. Stratton,** University of Massachusetts, Amherst

J. Kuriyan, University of California, Berkeley: Phosphorylation control in CaMKII.

S. Vogel, National Institutes of Health, Rockville, Maryland: Using Förster resonance energy transfer (FRET) spectroscopy and fluorescence correlation spectroscopy (FCS) to study conformational changes associated with Venus-tagged CaMKII activation and T-site interactions.

**SESSION 7:** Meeting Wrap-Up

**Chairpersons: Y. Hayashi,** Kyoto University, Japan; **J. Hell,** University of California, Davis

# Microbiology of the Built Environment

**November 3–6**

ARRANGED BY      **J. Green,** University of Oregon/Phylagen, Inc., Eugene
                 **R. Kolter,** Harvard Medical School, Boston, Massachusetts
                 **J. Peccia,** Yale University, New Haven, Connecticut

FUNDED BY        **Alfred P. Sloan Foundation**

In recent years, microbiome research has grown rapidly as the mutualisms, antagonisms, and beneficial or pathogenic effects of these communities are revealed and linked to human health consequences. This meeting convened experts to examine the critical, cross-sector, and cross-disciplinary issues associated with the microbiology of the built environment, as well as the underlying challenges of long-term momentum for the field and strategies for continued progress.

**Welcoming Remarks:  R. Leshan,** Director, Banbury Center, Cold Spring Harbor

**Introduction and Meeting Objectives:  J. Ausubel,** The Rockefeller University, New York, New York
**J. Green,** University of Oregon/Phylagen, Inc., Eugene

**SESSION 1:** Health Impacts

**Chairperson: V. Loftness,** Carnegie Mellon University, Pittsburgh, Pennsylvania

V. Loftness, Carnegie Mellon University, Pittsburgh, Pennsylvania: The importance of access to nature for human health and sustainability (and the unknowns relative to microbiomes).

A. Prussin, K. Bibby, F. Ling, C. Fernandez Marco, J. Siegel, L. Marr



R. Colwell, G. Anderson

E. Matsui, University of Texas, Austin: Animals and their allergens and microbes: implications for asthma disparities.

J. Portnoy, Children's Mercy Hospitals and Clinics, Kansas City, Missouri: Indoor allergen assessment and interventions to control asthma.

A. Hoisington, Air Force Institute of Technology, Wright-Patterson AFB, Ohio: The built environment and mental health.

J. Handelsman, University of Wisconsin, Madison: Novel approaches to sourcing antibiotics and resistance genes from the environment.

**SESSION 2:** Approaches for the Study of Indoor Microbes I

**Chairperson: G. Andersen,** Lawrence Berkeley National Laboratory, California

R. Knight, University of California, San Diego, La Jolla: Integrative omics approaches for linking built environment microbiology to human health.

G. Andersen, Lawrence Berkeley National Laboratory, California: RNA as a proxy for microbial activity in the indoor environment.

**SESSION 3:** Water and Microbes in Built Environments

**Chairperson: F. Ling,** Washington University in St. Louis, Missouri

F. Ling, Washington University in St. Louis, Missouri: Process-based and data-driven approaches to understand water microbiome dynamics.

A. Pruden, Virginia Tech, Blacksburg: Waterborne microbiomes: bridging the gap between environmental and public health.

R. Colwell, University of Maryland, College Park: Providing healthy water in urban environments.

K. Bibby, University of Notre Dame, Indiana: Bringing viral indicators indoors.

**SESSION 4:** Fungi/Molds in Built Environments

**Chairperson: D. Betancourt,** Environmental Protection Agency, Research Triangle Park, North Carolina

D. Betancourt, Environmental Protection Agency, Research Triangle Park, North Carolina: An EPA pilot study characterizing fungal and bacterial populations at home with flooding events at the Martin Pena Channel Community, San Juan, Puerto Rico.

J. Peccia, Yale University, New Haven, Connecticut: Advances in using fungal ecology to diagnose water-damaged homes.

T. Reponen, University of Cincinnati, Ohio: Measured versus observed mold and the development of children's asthma.

B. Sothern, Microecologies, Inc., New York, New York: Mold allergic diseases and asthma: assessing validity and usefulness of indoor air sampling as a tool for health-protective advice to occupants.

**SESSION 5:** Microbial Exposure and Control

**Chairperson: W. Bahnfleth,** Penn State University, University Park, Pennsylvania

W. Bahnfleth, Penn State University, University Park, Pennsylvania: Optical radiation for control of microbes in air and on surfaces.

E. Hartmann, Northwestern University, Evanston, Illinois: The impact of antimicrobial chemicals on microbial viability and antibiotic resistance.

W. Nazaroff, University of California, Berkeley: Indoor bioaerosol dynamics: microbial exposure consequences.

A.J. Prussin, Virginia Tech, Blacksburg: Viral aerosols in the built environment.

J. Siegel, University of Toronto, Ontario, Canada: Intentionally manipulating the indoor microbiome: challenges and opportunities.

R. Kolter, J. Green



T. Reponen, J. Portnoy

**SESSION 6:** Approaches for the Study of Indoor Microbes II

**Chairperson: L. Marr,** Virginia Tech, Blacksburg

L. Marr, Virginia Tech, Blacksburg: Humidity, microbiology, microchemistry, and exposure in droplets and aerosols.

G. Mainelis, Rutgers University, New Brunswick, New Jersey: Challenges and successes in bioaerosol sampling to investigate microbiology of the built environment.

K. Van Den Wylmelenberg, University of Oregon, Eugene: Design the unseen.

**SESSION 7:** Meeting Wrap-Up and Next Steps

**Chairpersons: P. Olsiewski,** Alfred P. Sloan Foundation, New York, New York; **R. Kolter,** Harvard Medical School, Boston, Massachusetts; **J. Ausubel,** The Rockefeller University, New York, New York

# Lustgarten Foundation Scientific Advisory Board Meeting

Banbury was pleased to welcome back the Lustgarten Foundation for their 2019 Scientific Meeting, which provided an opportunity for the Scientific Advisory Board, as well as Foundation-supported investigators, to discuss research and strategy, evaluate performance, provide feedback for improvement, strengthen collaboration, and identify new ideas to bolster progress in the field.

**J. Abbruzzese,** Duke Cancer Institute, Durham, North Carolina

**S. DeGarabedian,** The Lustgarten Foundation, Woodbury, New York

**R. Evans,** Salk Institute for Biological Studies, La Jolla, California

**D. Fearon,** Cold Spring Harbor Laboratory

**F. Froeling,** Cold Spring Harbor Laboratory

**C. Fuchs,** Yale School of Medicine, New Haven, Connecticut

**T. Hunter,** Salk Institute for Biological Studies, La Jolla, California

**T. Jacks,** Koch Institute at MIT, Cambridge, Massachusetts

**E. Jaffee,** Johns Hopkins School of Medicine, Baltimore, Maryland

**K. Kaplan,** The Lustgarten Foundation for Pancreatic Cancer Research, Woodbury, New York

**D. Kelsen,** Memorial Sloan Kettering Cancer Center, New York, New York

**R. Mayer,** Dana-Farber Cancer Institute/Harvard University, Boston, Massachusetts

**M. Muzumdar,** Yale Cancer Center, New Haven, Connecticut

**J. O'Brien,** The Lustgarten Foundation, Woodbury, New York

**S. Park,** Johns Hopkins University, Baltimore, Maryland

**A. Parker,** Thrive Earlier Detection, Cambridge, Massachusetts

**D. Plenker,** Cold Spring Harbor Laboratory

**E. Sawey,** Lustgarten Foundation, Woodbury, New York

**D. Tuveson,** Cold Spring Harbor Laboratory

**M. Vander Heiden,** Massachusetts Institute of Technology, Cambridge

**R. Vizza,** The Lustgarten Foundation for Pancreatic Cancer Research, Woodbury, New York

**B. Vogelstein,** Sidney Kimmel Cancer Center at Johns Hopkins University, Baltimore, Maryland

**B. Wolpin,** Dana-Farber Cancer Institute, Boston, Massachusetts

**A. Yuille,** Johns Hopkins University, Baltimore, Maryland

# The Nervous System in Cancer

**December 10–13**

ARRANGED BY      **S. Knox,** University of California, San Francisco
                          **M.L. Monje,** Stanford University, California
                          **T. Wang,** Columbia University, New York, New York

FUNDED BY         **Cygnal Therapeutics**

The nervous system critically modulates development, homeostasis, and plasticity. A similarly powerful role for neural regulation of the cancer microenvironment is emerging. Neurons promote the growth of cancers in many tissue types. Parallel mechanisms shared in development and cancer suggest that neural modulation of the tumor microenvironment may prove a universal theme, although the mechanistic details of such modulation remain to be discovered for many malignancies. This meeting convened experts to discuss both local and systemic cross talk between the nervous system and cancer, and the emerging principles of cancer neuroscience.

**Welcoming Remarks: R. Leshan,** Director, Banbury Center, Cold Spring Harbor Laboratory

**Introduction and Meeting Objectives: S. Knox,** University of California, San Francisco
                                              **M. Monje,** Stanford University, California
                                              **T. Wang,** Columbia University, New York, New York

**An Exoneural Platform in Drug Discovery: J. Hurov,** Cygnal Therapeutics, Cambridge, Massachusetts

B. Deneen, S. Hervey-Jumper



P. Frenette, L. Trotman

**SESSION 1:** Neural Regulation of Development, Plasticity, and Regeneration

**Chairperson: S. Knox,** University of California, San Francisco

A. Lloyd, University College London, United Kingdom: Links between peripheral nerve regeneration and cancer.

H. Hondermarck, The University of Newcastle, Callaghan, New South Wales, Australia: Nerve dependence: from regeneration to cancer.

S. Knox, University of California, San Francisco: Neuronal control of glandular development, regeneration, and cancer.

R. Segal, Harvard University, Boston, Massachusetts: Cancer, chemotherapy, and nerves.

M. Taylor, University of Toronto, Ontario, Canada: Identifying the lineages of origin for childhood brain cancers.

**SESSION 2:** Neural–Immune Interactions

**Chairperson: D. Gutmann,** Washington University in St. Louis, Missouri

K. Tracey, The Feinstein Institute for Medical Research, Manhasset, New York: Neural signaling regulating immunity.

X. Sun, University of California, San Diego, La Jolla: Neural control of lung function and pathogenesis.

E. Sloan, Monash University, Parkville, Victoria, Australia: Neural remodeling of the tumor microenvironment.

D. Gutmann, Washington University in St. Louis, Missouri: Low-grade glioma ecosystem biology.

**SESSION 3:** Neural Regulation of Primary Brain Cancers

**Chairperson: M. Monje,** Stanford University, California

M. Monje, Stanford University, California: Neuronal activity regulates the proliferation of normal and malignant glia.

L. Garzia, McGill University, Montreal, Quebec, Canada: Electrical activity regulates medulloblastoma pathogenesis.

P. Dirks, The Hospital for Sick Children, Toronto, Ontario, Canada: Neuromodulators and the control of brain tumor cell proliferation.

F. Winkler, Universität Heidelberg, Germany: Glioma: a brain in the brain.

**SESSION 4:** Neural Regulation of Endo-/Ectodermal Cancers

**Chairperson: T. Wang,** Columbia University, New York, New York

T. Wang, Columbia University, New York, New York: The role of nerves in gastric cancer.

F. Fattahi, University of California, San Francisco: Derivation of peripheral nervous system lineages from human pluripotent stem.

J. Saloman, University of Pittsburgh, Pennsylvania: Neuroplasticity and neuroimmune interactions in pancreatic tumorigenesis.

R. White, Columbia University Medical Center, New York, New York: Targeting the parasympathetic nervous system in pancreatic adenocarcinoma.

N. D'Silva, University of Michigan, Ann Arbor: Nerves and cancer: a dynamic interaction.

**SESSION 5:** Neural Regulation of Metastasis

**Chairperson: H. Hondermarck,** The University of Newcastle, Callaghan, New South Wales, Australia

P. Frenette, Albert Einstein College of Medicine, Bronx, New York: Nerves at the forefront of hematopoietic stem cell migration, aging, and cancer.

D. Hanahan, Ecole Polytechnique Fédérale de Lausanne, Switzerland: Glutamate-stimulated NMDAR signaling drives cancer invasion and brain metastasis.

**SESSION 6:** Cancer Regulation of Neuronal Activity

**Chairperson: B. Deneen,** Baylor College of Medicine, Houston, Texas

M. Monje (*back to camera*), N. D'Silva



A. Lantermann, P. Frenette

B. Deneen, Baylor College of Medicine, Houston, Texas: Merging functional genomics and the neuroscience of brain tumors.

C. Magnon, INSERM, Fontenay-aux-Roses, France: Role of the central nervous system in the development of cancer.

S. Hervey-Jumper, University of California, San Francisco: Glioma-neuron synapses enriched within intratumoral functional connectivity network hubs influences language plasticity in adult IDH-wt glioblastoma.

J. Borniger, Stanford University, California: Distal modulation of subcortical neural activity by cancer in the periphery.

**SESSION 7:** Meeting Summary, Wrap-Up, and Next Steps

**Chairperson: C. Jhappen,** National Cancer Institute, Rockville, Maryland

DNA LEARNING CENTER

# DNA LEARNING CENTER
# EXECUTIVE DIRECTOR'S REPORT

**ADMINISTRATION**

Lindsay Barone
Lauren Correri
Mary Lamont
Valerie Meszaros
David Micklos
Collette Riccardi

**INSTRUCTION**

Lina Bader
Louise Bodt
Justin Burke
Megan Capobianco
Elna Carrasco
Heather Cosel-Pieper
Cristina Fernandez-Marco
Jennifer Hackett
Brittany Johnson
Melissa Lee

Allison Mayle
Amanda McBrien
Erin McKechnie
Bruce Nash
Michael Paul
Sharon Pepenella
Jeffry Petracca
Lina Ruiz-Grajales
Xiaoqun Catherine Zhang

**BIOMEDIA**

Cornel Ghiban
Daniel Jacobs
Susan Lauter
Jason Williams
Chun-hua Yang

## Coming Home to Brooklyn

On October 1, we signed a 30-year, no-cost lease with the City University of New York (CUNY) for an 18,000 square-foot space on the campus of the New York City (NYC) College of Technology (City Tech) in downtown Brooklyn. This was preceded by three years of legal wrangling with city and state authorities that control the property and the complicated public bonds through which it is financed. In this effort we had the unflagging support of City Tech President Russ Hotzler, who used his former experience as vice chancellor and president of several CUNY colleges to guide us through the bureaucratic process. The new facility was designed by Centerbrook Architects and Planners, who have done all architectural work for CSHL over the past 50 years—including the expansion of the DNALC in 2001. With renovations beginning in spring 2020, we expect to bring the facility into operation in time for summer camps in 2021.

Starting a new DNALC location at the "Gateway to Brooklyn" also recalls CSHL's little-known historical relationship to Brooklyn. The Biological Laboratory at Cold Spring Harbor was, in fact, founded in 1890 as the first operating unit of the newly incorporated Brooklyn Institute of Arts and Sciences—which later grew to include Brooklyn Museum, Botanical Garden, Children's Museum, and Conservatory of Music. So, in joining its sister institutions along the axis of Flatbush Avenue, the DNALC will be coming home to Brooklyn.

Finding the City Tech location was the culmination of a 13-year search for a space in which to extend the DNALC's model for bioscience enrichment to the students of metropolitan New York. Our vision is to provide a place where all NYC students have the same science learning and research opportunities that have been available to elite Long Island students for several decades. The Brooklyn center will build upon our success with *Harlem DNA Lab*, which has provided laboratory experiences for 33,000 students since its opening in 2008. The Harlem facility demonstrated that we can readily serve a natural constituency of underrepresented minority (URM) students, who compose about two-thirds of public school students in NYC. We tested the feasibility of the Brooklyn



The City Tech temporary lab in the midst of reorganization in preparation for field trips in the fall.

location when, in the spring, we opened a temporary lab in the same building on the City Tech campus. This filled quickly, with 1,287 students doing labs in 2019.

The City Tech property fulfills CSHL's key requirements of educational zoning, high visibility, and ready access. DNALC branding on the building facade will be immediately visible to all traffic coming off the Brooklyn Bridge. Downtown Brooklyn is one of the most accessible parts of New York City, with eight subway lines located within several blocks of City Tech. There will be easy bus drop-off on both Tillary and Adams Streets. A dedicated entrance on Tillary Street will provide direct access to the second floor.

The new facility will occupy the entire second floorplate of City Tech's Pearl Street Building. With six teaching labs and two bioinformatics labs, it will be twice the size of our flagship center in Cold Spring Harbor. The City Tech facility is also more than double the size of the 7,000 square-foot rental space we had originally conceived for NYC. So, the scope of the capital project was further magnified by a "gut" renovation of the space, asbestos abatement, entire HVAC upgrade, and replacement of windows and exterior cladding of the entire second floor.

The enlarged space will have increased operating costs—especially staffing. Early on, we also made the commitment to provide free tuition to at least half of students attending academic-year field trips and to initiate an ambitious research program for CUNY students. These activities will require additional endowment support. Taking all this into account, we have increased our funding goal from $25 million to $30 million—including about $18 million for construction and $12 million for endowment.



Partial view of the City Tech space.

At City Tech, we will continue our proven program of academic-year field trips and summer camps for precollege students. The hands-on laboratory work will be complemented by minds-on bioinformatics exercises, which will engage students with the coding, computational, and data science skills critical for STEM success.

The City Tech location will also provide a proving ground for our work in undergraduate education. Here the emphasis is on course-based undergraduate research experiences (CUREs). As opposed to the traditional undergraduate research model, which places a small number of students with individual faculty mentors, CUREs expand research opportunities to all interested students in the context of for-credit courses. When rigorously implemented, especially in the freshman year, CUREs increase retention in STEM majors and on-time graduation



Approaching the City Tech building from the Brooklyn Bridge.

by ~20%. Most of the DNALC's federal funding over the last two decades has been devoted to developing experiment and computer infrastructure to support CUREs. Two teaching labs in the Brooklyn facility will be used exclusively for research by CUNY students, and we hope to quickly establish one of the country's largest CURE sites—serving up to 900 students per year.

The City Tech facility will provide an ideal setting for our CURE work and also as the designated Genomics Hub of InnovATE*BIO*, the National Biotechnology Education Center. As explained below, this National Science Foundation (NSF) project explicitly supports workforce development and bioscience career pathways in two-year institutions. City Tech is unique among CUNY institutions in having both two-year and four-year programs on the same campus. City Tech has 17,300 students—of whom 62% are Black and Hispanic and 67% are the first in their family to attend college—and is officially designated as a Hispanic Serving Institution (HSI). CUNY is the largest urban university system in the United States and one of the largest producers of African–American doctoral degrees in the natural sciences and engineering.

The downtown Brooklyn location has strong foot traffic, so we intend to develop a substantial business of "drop-in" visitors on the weekends. We want to make the DNALC a prominent travel destination for people interested in learning what DNA can tell them about their health and their place in the human family. The stories told in the 2,700 square-foot exhibit will integrate tightly tailored student experiments and family activities. For example, the COVID-19 pandemic illustrates how knowledge of DNA is used to test for the disease, track its spread, and design rational treatments and vaccines. Our most popular human experiment uses a person's own DNA type to predict their bitter-tasting ability, a direct analog of the pharmacogenetics approach of using a DNA signature to predict a person's response to a particular drug or chemotherapy.

We popularized the first "personal DNA" experiment 20 years ago, so we envision a day when every DNALC visitor has the opportunity to look at his/her DNA and compare it to classmates and world populations to show the shared ancestry of all people. This experiment would articulate with a large-scale interactive map that uses personal DNA data to illustrate the prehistoric migration of humans out of Africa that peopled the Old and New Worlds. We will also show the genetic signatures of historic diasporas of ethnic and religious groups, and the spread of innovations—farming, horse culture, and dairying. All of this would be complemented by replicas of ancient ancestors, including those developed by the DNALC—Ötzi the Iceman, developed from CT-scan data, and the first articulated Neanderthal skeleton. Visitors could extend their DNA journey with a visit to nearby Ellis Island National Monument, which greeted the ancestors of ~40% of Americans alive today. The Tenement Museum and Museum of the City of New York give further depth to the story of how New York became the U.S. melting pot and provide opportunities for detailed exploration of individual ancestry.

## Biotechnology in American High Schools

As part of an early grant from the NSF's Advanced Technological Education (ATE) program, the DNALC conducted a nationwide survey of 4,100 high school biology teachers. This sample took a snapshot of biotechnology/molecular genetics instruction in American high schools in 1998. The survey was designed to compare laboratory instruction and student exposures to six major techniques of biotechnology/molecular genetics that were measured in the original survey: bacterial transformation, DNA restriction analysis, DNA recombination, plasmid isolation, polymerase chain reaction (PCR), and DNA sequencing. With renewed funding from NSF, we repeated this study in 2018, receiving 2,100 responses from high school biology teachers across the country. Here are the highlights of changes we found in biotechnology instruction over the past two decades.

*The number of students exposed to six biotech labs has increased, and teaching these labs has become more mainstream. However, fewer faculty are involved in biotech teaching, and the pace of integrating new labs has slowed.*

- The number of faculty offering biotech labs in AP Biology has decreased (62% to 54%), whereas those offering biotech laboratories in general biology have increased by a similar margin (21% to 28%).

- Although more faculty offer labs on PCR and DNA sequencing today, it is at half the rate that teachers offered the then-novel methods of transformation and restriction analysis in 1998.

| | 1998 ($n = 4{,}100$) | | 2018 ($n = 2{,}100$) | |
|---|---|---|---|---|
| | Teaching (%) | Reported annual student exposures | Teaching (%) | Reported annual student exposures |
| Transformation | 51 | 80,384 | 45 | 106,453 |
| Restriction analysis | 60 | 118,490 | 43 | 135,240 |
| DNA recombination | 32 | 47,666 | 25 | 65,785 |
| Plasmid isolation | 17 | 24,312 | 15 | 31,044 |
| PCR | 12 | 21,576 | 26 | 28,498 |
| DNA sequencing | 16 | 29,398 | 20 | 39,334 |
| Total exposures | | 321,826 | | 406,354 |

*Biotechnology funding and electives have doubled since 1998. However, few schools with biotechnology electives were aligned with the school-to-work movement as advocated by the NSF ATE program, and advanced teaching is concentrated in wealthier districts.*

- Adjusted for inflation, per-teacher funding for biotechnology has increased from $8,236 to $16,651.

- Schools offering laboratory-based biotechnology electives increased from 16% to 35%.

- Although 35% of faculty at schools with biotechnology electives used curriculum materials provided by industry, only 11% used ATE materials and only 22% of these schools had articulation agreements with colleges.

- Sixty-eight percent of schools with biotechnology electives in 2018 were located in zip codes above the U.S. median household income.

*Although 2018 faculty are more academically prepared, they are less involved with professional societies and extracurricular activities.*

- Eighty percent of 2018 teachers had graduate degrees compared to 74% in 1998.

- Significantly fewer 2018 teachers belonged to major professional societies, including NABT, NSTA, and state science teachers' associations.

- Significantly fewer 2018 teachers participated in all types of extracurricular activities with their students. Significantly more 2018 teachers said they did no out-of-class science activities at all.

| | 1998 | 2018 |
|---|---|---|
| After-school student research | 26% | 18% |
| Science fairs/competitions | 36% | 26% |
| Science field trips | 53% | 42% |
| Joint activities with scientists from local universities/institutes | 24% | 16% |
| None of the above | 27% | 41% |
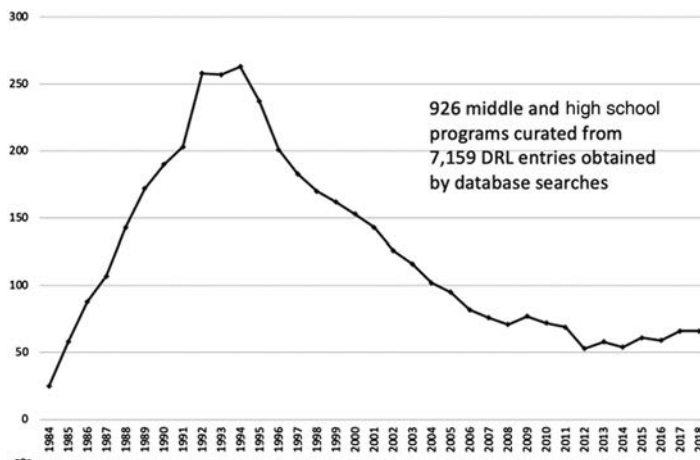
NSF-funded teacher training 1984–2018.

Consistently over the last 20 years, teachers said they valued summer workshops of 5+ days and workshops at professional meetings as the important contributors to innovation in the classroom. However, whereas 65% of 1998 teachers had attended one or more professional meetings in the past year, 60% of 2018 teachers had attended none. Furthermore, 39% of 2018 faculty thought there were fewer opportunities for training at workshops and summer institutes than in the past, compared to 27% who thought there were more. This suggested a disturbing trend of today's biology teachers having less access to the very types of training that can keep them involved and up-to-date. This jived with my own observations and those of others who have done grant-funded training over the last several decades.

Traditionally, the NSF has been the major provider of focused, high-quality training for precollege biology teachers. So, we examined 7,454 entries in the database of education grants made by the NSF, going back to 1982 when the database appears to have been started. Of these, we determined that 948 offered training opportunities of interest to middle and high school biology educators, and we plotted a graph of the years in which they were active. NSF teacher training opportunities peaked in 1994, when 263 training programs were operational. From that point, NSF training dropped precipitously, falling to 170 in 1998 and reaching its nadir in 2012, when there were 53 programs. The number of new programs had recovered somewhat to 66 by 2018. Our database analysis confirmed the subjective feeling among 2018 teachers that there were fewer training opportunities available to them. In fact, 2018 teachers had only 38% of the NSF training opportunities as did the 1998 cohort and only 25% as many as teachers in 1994. This amounts to an abdication of a core principle of the NSF Authorization Act of 1973, which made NSF explicitly responsible for "science education at all levels." It is reason for alarm, at a time when biology is progressing so quickly and pandemic isolation is rendering laboratory instruction nearly impossible. It is a perfect storm to rip apart the hands-on biology instruction that is the bastion of American science learning.

## NSF CyVerse Study of Bioinformatics Education

CyVerse is an NSF-funded cyberinfrastructure for life sciences. The project merges high-performance computing, data storage, and people to solve complex biological problems. DNALC training offered through CyVerse focuses on the "people" component of cyberinfrastructure, equipping educators with bioinformatics and data science teaching skills. As part of the Network for Integrating Bioinformatics into Life Science Education (NIBLSE), we led research on barriers

educators face in teaching bioinformatics. In the largest nationwide study on the topic to date[*], more than 1,200 undergraduate biology faculty revealed that they struggle with teaching bioinformatics primarily because of their own lack of training. URM faculty and faculty at two-year and minority-serving institutions reported increased barriers compared to their peers. Surprisingly, although recent graduates had better training in bioinformatics, they were less likely to teach this topic than senior faculty. Taken together, the findings highlight the need for more professional development and better support for biology faculty as they teach computational topics. Our study also pointed up the need to level the intellectual playing field for URMs. In response to our own studies, we have refocused CyVerse training efforts on longer-duration workshops, including a new, week-long Foundational Open Science Skills (FOSS) course that helps faculty integrate bioinformatics and computational tools into their classroom teaching. We are also redoubling our effort to include URM faculty and to site training at URM-serving institutions.

## DNA Barcoding and Metabarcoding

The DNALC continued its concerted efforts to enable high school and college students to conduct authentic biodiversity research using DNA barcoding. Three programs support high school students. *Barcode Long Island (BLI),* funded by the National Institutes of Health (NIH), involves students in "campaigns" to compare biodiversity across Long Island. The *Urban Barcode Project (UBP),* funded by the Thompson Family Foundation*,* and *Urban Barcode Research Program (UBRP),* funded by matching grants from the Pinkerton Foundation and Simons Science Sandbox, involve students in independent research of biodiversity in NYC. Science teachers are mentors for *BLI* and *UBP* students, whereas scientists from NYC institutions mentor *UBRP* students. A new collaboration with Hudson River Park, funded by the Lounsbery Foundation, piloted student and citizen science involvement in metabarcoding of fish. Major funding from the NSF Improving Undergraduate STEM Education (IUSE) and ATE programs supports the development and dissemination of CUREs for undergraduate students.

Over the year, we improved the biochemical and online resources that support all of these programs, and are making barcoding and metabarcoding accessible to students worldwide. A new, rapid DNA isolation kit makes DNA barcoding more achievable in short classes. A similar fast procedure for microbial and vertebrate metabarcoding also shows promise. After several rounds of testing, we redesigned our sample indexing strategy for metabarcoding—improving sequence quality and taxonomic resolving power, while maintaining the low costs we achieved in previous years. We also developed indexed primers to support invertebrate metabarcoding, expanding the repertoire of research questions we can support.

Improvements to the DNALC *DNA Barcoding 101* website (http://www.dnabarcoding101.org) included support materials for the new isolation method, new PCR primer information, updated DNA staining and sequencing instructions, a table of protocol choices, and guides for taxonomic identification and sample documentation. The Sample Database evolved as a simple means to enter, store, and access all information related to each student sample and barcode sequence. The website now allows users to create independent programs, enabling faculty and independent groups to manage their own DNA barcoding projects.

*DNA Subway*, the bioinformatics gateway developed by the DNALC as part of CyVerse, supports our biodiversity programs and is a popular tool for educators at all levels. In 2019, *DNA Subway* had 36,322 registered users, 63,410 visits (~4% decrease from 2018), and 1.22 million page views

[1]Williams JJ, Drew JC, Galindo-Gonzalez S, Robic S, Dinsdale E, Morgan WR, Triplett EW, Burnette JM III, Donovan SS, Fowlkes ER, et al. 2019. Barriers to integration of bioinformatics into undergraduate life sciences education: a national study of US life sciences faculty uncover significant barriers to integrating bioinformatics into undergraduate instruction. *PLoS ONE* **14:** e0224288. doi:10.1371/journal.pone.0224288

(~3% decrease from 2017). Students created 43,492 projects (~10% increase from 2018) across the five *Subway* lines. The Blue Line supports DNA barcoding sequence analysis, whereas the Purple Line is a custom and approachable interface to the metabarcoding analysis package QIIME2. This year, the Purple Line was updated to improve the speed and quality of analyses, and support for invertebrate sequence analysis was added. Crucially, we solved an analysis bottleneck that thwarted new users—we created a self-contained "wizard" that seamlessly manages and creates metadata files. Design updates gave the Purple Line a cleaner look, mirroring changes to the rest of the site.

"Upstream" of *DNA Subway*, a new laboratory information management system (LIMS) allows faculty to assign indexes to different groups and manage next-generation sequencing (NGS) libraries. Anticipating a nationwide sequencing service for student metabarcoding projects, this tool can manage hundreds of samples submitted by dozens of users—and separate (deconvolute) the millions of sequence reads generated in a single NGS run.

## High School DNA Barcoding Research Programs

The 2019 *BLI* program included 208 students working in 79 teams and representing 25 high schools from Suffolk, Nassau, and Queens counties. Twelve percent of participants were African–American, Latino, or Native American. During the year, 73 students attended seven open laboratory sessions held at the Dolan DNALC, DNALC *West*, Stony Brook University (SBU), or Brookhaven National Laboratory (BNL), while 114 students used borrowed equipment kits. Nine teams (22 students) used high-throughput sequencing to perform metabarcoding to study marine fish, microbiomes from water, or invertebrates. More than 800 samples were processed, resulting in more than 1,200 sequencing reads and 1.2 million NGS reads. We published 239 sequences in GenBank, including two new barcode sequences and 40 with sequence polymorphisms.

Sixty-eight DNA barcoding and metabarcoding projects were presented at the annual *BLI* research symposium on June 4 at CSHL. These included biodiversity studies of plants, invertebrates, fungi, algae, and lichens; microbiome studies of water, excrement, and invertebrates; and eDNA studies of fish. Dr. Semir Beyaz, CSHL Fellow and Donaldson Translational Fellow at CSHL, gave the keynote address on the interplay between diet and microbiome in cancer risk. *BLI* students received awards at numerous competitions, including the Long Island Science and Engineering Fair (LISEF) and NYC Science and Engineering Fair (NYCSEF). One group received the *Brooklyn Friends of Clearwater Award* for increasing awareness of the environment. Three teams from William Floyd High School were invited by their county legislator to present their research on effects of heavy metals on aquatic biodiversity.

The 2019 *UBP* and *UBRP* programs had 161 students working in 61 teams and representing 23 NYC high schools. *UBP* and *UBRP* students made ample use of DNALC resources: 62 students
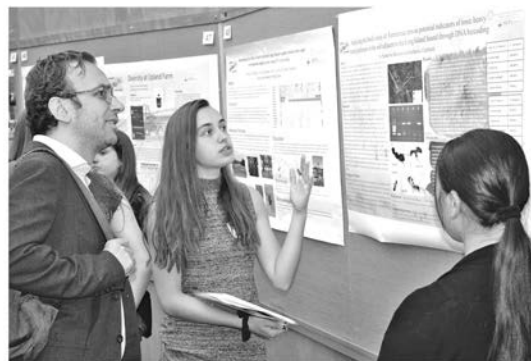


(*Left*) Dr. Semir Beyaz presents the keynote address to *BLI* Symposium attendees and (*right*) talks to *BLI* participants at their poster in Bush Auditorium.
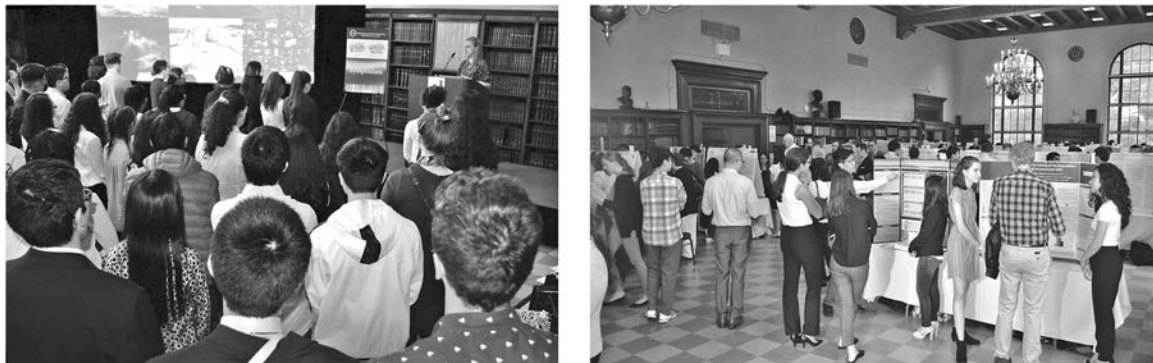
(*Left*) Dr. Claudia Wultsch addresses *UBP* and *UBRP* student researchers who (*right*) shared project results at a poster session.

attended open lab sessions at *Harlem DNA Lab* or *DNALC NYC*, whereas 49 students borrowed equipment. Teams collected and processed more than 1,100 samples for DNA sequencing, resulting in more than 1,150 single sequences and 6.2 million NGS reads. The annual research symposium on May 30 at the New York Academy of Medicine showcased 61 projects and included a keynote speech by Dr. Claudia Wultsch of Hunter College and American Museum of Natural History (AMNH) on the microbiomes of wild carnivores. One *UBRP* team was recognized with an outstanding poster award at the event, for a project that examined the microbiomes from the noses of e-cigarette users and nonusers. The winner for *UBP* showed that earthworm diversity is higher in private—compared to public—locations in Greenpoint, Brooklyn. One *UBP* team submitted a manuscript to the *Journal of Emerging Investigators* on using DNA barcoding to identify plant species and create a phenology trail in Central Park.

This year, 118 students across all three barcoding programs (*BLI, UBP,* and *UBRP*) completed surveys as a part of our ongoing effort to monitor the impact of participation in science research. Participants were asked about their experiences in the programs, how much they had learned, and how they felt about science. The students were overwhelmingly proud of the research they had done (86.4%) and felt that problem-solving approaches learned during their research would be helpful in future science courses (80.9%) and careers (77.3%). Nearly three-quarters (71.3%) said they were more interested in continuing science study and, specifically, biology (74.1%). Overall, our results suggest that DNA barcoding demystifies the process of science research and encourages students to continue on STEM pathways.

### Studying Biodiversity in the Hudson River

A new grant from the Richard Lounsbery Foundation supported a collaboration with the Hudson River Park (HRPK) to systematically sample water from the Hudson River. Using eDNA to identify the creatures living in the water provides a window into the life of this wild space adjacent to the nation's busiest metropolis. Beginning in January, HRPK staff and high school students collected more than 200 water samples. HRPK staff were trained in eDNA processing and supported the students participating in INCLUDES, an intensive summer research program, extracted and amplified DNA for sequencing. More than 250 citizen scientists participating in HRPK summer programs collected and filtered additional water samples. Meanwhile, we led 500 visitors as they isolated DNA from organisms collected in the park during HRPK's sixth annual Submerge Marine Science Festival.

Initial results were promising, with many of the fish expected in the river appearing in the eDNA results. Highlights included identifying dolphin, endangered sturgeon, and shiner DNA. A valuable, but less exciting, finding is that Hudson River vertebrate DNA is dominated by human

and human-associated DNA—from our pets and food. This made it harder to find the fish DNA we were looking for.

A separate small collaboration with the Billion Oyster Project (BOP) aims to help teachers work with their students to collect and identify organisms that populate oyster cages. A Billion Oyster Project teacher joined *Urban Barcode Project* training during the summer to learn how to lead teams in preparation for work at BOP sites. In turn, we presented DNA barcoding at the BOP Annual Science Fair, letting participants know about this new opportunity.

### Barcoding in Undergraduate Classes

We continued to develop, disseminate, and assess DNA barcoding and metabarcoding as "formatted" solutions for CUREs through our $2 million IUSE collaboration with James Madison University (JMU), CUNY City Tech, Bowie State University (BSU), and Austin Community College (ACC). JMU has adapted the DNALC's barcoding curriculum to support a model CURE. Remarkably, this CURE reaches more than 1,700 students per year. Pushing DNA barcoding into introductory classes with many underrepresented minorities, our collaborators at City Tech and BSU implemented DNA barcoding with 24 and 22 freshman students, respectively.

During the summer, JMU hosted a five-day "DNA Barcoding for CUREs" workshop. Twenty-two undergraduate educators (18% underrepresented minority; 77% female, 23% male) representing seven two-year public, six four-year public, and five four-year private undergraduate institutions participated in the workshop designed to prepare them to successfully implement DNA barcoding CUREs. Participants learned by doing, carrying out a biodiversity study of JMU forest habitat; presentations by collaborators on the science and CURE implementations highlighted potential challenges and solutions. In a remarkably fast turnaround, nine workshop participants implemented DNA barcoding CUREs during the fall semester—reaching 242 students, including 24% underrepresented minorities. These students reported similar or better learning and attitudinal effects as students taught by project co-PIs, suggesting that the week-long training format is sufficient to launch an effective DNA barcoding CURE.

In preparation for training in 2020, the DNALC's metabarcoding pipeline was adapted and integrated into an upper-level genomics course at JMU. In this course, 14 junior and senior undergraduates and two graduate students showed that male and female snakes of the same species have different microbiomes. While developing course materials, a JMU student co-authored a manuscript outlining an introduction to command line analysis of NGS data, which is currently in review with the peer-reviewed journal *CourseSource*. BSU also piloted metabarcoding with



(*Left*) NSF IUSE *DNA Barcoding* Workshop participants sample campus biodiversity at JMU and (*right*) explore costs, benefits, and barriers of barcoding CUREs.

30 students in an upper-level molecular biology class, comparing the microbial diversity in different aquatic environments on campus.

## New National Center for Biotechnology Education

In October 2019, the DNALC became a lead institution in the InnovATE*BIO* national biotechnology education center. The center is funded through NSF ATE, which seeks to keep America's workforce competitive. Although the ATE program focuses almost exclusively on two-year colleges, this is the fourth grant that the DNALC has received. Previous grants supported development and dissemination of experiments that illustrate key methods in biotechnology, as well as our survey of high school biotechnology education. Our long-time collaborator Linnea Fletcher, at Austin Community College, leads the project team. Our role is to develop a New York City Genomics Hub to support genome-based experiments in two-year colleges. The ATE hub will be based at the new *DNALC* at City Tech, in Brooklyn.

As part of this national center, we will develop a supply chain model that will dramatically reduce DNA sequencing costs and allow an unprecedented number of students to participate in authentic research. The genomics supply chain will entail laboratory, quality assurance, data science, and "soft" skills that will prepare students for successful careers in biotechnology. In this model, students at community colleges will learn key technical skills as they produce the products and services to support classroom experiments. Biotechnology students will learn while producing the reagents needed to isolate DNA, amplify the barcode region, and prepare it for sequencing. Students will not only produce reagents for their own programs, but also distribute kits for students in other community colleges and high schools. Advanced students will also assist their peers as "clients," supporting them as they learn biochemistry and bioinformatics.

A crucial element of the system will be a student-staffed sequencing service, giving students real-world experience in the world of biological big data. Using approaches developed at the DNALC, the sequencing service will coordinate sequence submissions from classes around the country and stage them for cost-effective sequencing on machines run by students in community college biotechnology programs.

## NSF MaizeCODE

MaizeCODE continued developing data that will become an important resource for breeders and plant scientists. Our pilot study showed how to use MAKER-P quality scores and the alignments between a translated protein sequence and its homologs across species to identify errors in gene predictions[*]. Student curators then corrected the flagged gene models using the Apollo annotation editor and uploaded their corrections as a track on the *Gramene* genome browser.

As part of our outreach efforts on this project, the second and third Maize Annotation Jamborees were held January 10–11, 2019, at the Scripps Institution of Oceanography in San Diego, California, and March 13–14, 2019, at the Biology Department of Washington University in St. Louis, Missouri. We trained PUI faculty and researchers to use our genome annotation pipeline, with the objectives of integrating maize annotation CUREs and establishing a larger community curation effort to improve the *Zea mays* gene models. These efforts will continue via periodic meetings to discuss progress on the partnerships and by

---

[2]Tello-Ruiz MK, Marco CF, Hsu F-M, Khangura RS, Qiao P, Sapkota S, Stitzer MC, Wasikowski R, Wu H, Zhan J, et al. 2019. Double triage to identify poorly annotated genes in maize: the missing link in community curation. *PLoS ONE* **14:** e0224086. doi:10.1371/journal.pone.0224086

Jamboree participants in (*left*) San Diego and (*right*) St. Louis.

providing assistance in developing bioinformatics lessons and wet laboratory resources that can be implemented in the classroom.

We presented the results of our annotation project at the XXVII Plant and Animal Genome Meeting in San Diego, California (January), at the 61st Maize Genetics Conference in St. Louis, Missouri (March), at Middle Tennessee State University, Murfreesboro, Tennessee (November), and the 11th CSHL Plant Genomes System Biology and Engineering Meeting in Cold Spring Harbor, New York (December).

## Licensed Centers

We celebrated the official opening of the *Regeneron DNA Learning Center* in December. Located on Regeneron's Sleepy Hollow campus, this new 4,700 square-foot facility has two teaching laboratories and a large prep laboratory with space specifically designed for assembling footlocker kits. The *Regeneron DNALC* is easily accessible to schools in Westchester, Rockland, and Putnam counties, as well as New Jersey and Connecticut. In our first month of operation in 2019, more than 100 high school students visited for field trips, and reservations for an additional 1,500 students were made for spring 2020.

In 2019, 1,691 students from 42 different schools participated in hands-on molecular biology labs supported by the *DNALC at Notre Dame* (DNALC-ND). Under the leadership of director Dr. Amy Stark, instructional programs included laboratory field trips to the DNALC-ND, in-school instruction, and engagement at regional and state-level science fairs. More than 112 students, including two from Canada, participated in week-long residential and day camps.

## International Partnerships
### DNALC Asia, Suzhou, China

As part of our collaboration with *DNALC Asia* in Suzhou Industrial Park (SIP), we organized and interviewed candidates for a new Education Director after the departure of Jessica Talamas. Finding the right person for this position was key, because we wanted an educator versed in the U.S. style of instruction and prepared for the differences of life in China. We were very fortunate to recruit Dr. John Olson, a New York native who was already a working lecturer at Peking University. After

Dr. John Olson

training here at the DNALC, John started in Suzhou in August and worked to set up barcoding research projects at international schools and local universities.

*DNALC Asia* continued to ramp up its instructional capacity. In September and October, the Center saw 1,584 student visitors for on-site labs (vs. 307 in 2018) and taught 1,072 in local high schools (vs. 375 in 2018). The *DNALC Asia* "Young Biologist" program selects talented students in the life sciences and helps them develop skills in experimental biology. After six months of training, students independently complete a scientific poster, present their material to the public, and then meet face-to-face with expert judges. The 2018–2019 program began in December of 2018 with 70 students, and after two rounds of selection, 20 students were selected to participate in the final presentations in April.

The 2019–2020 program began with 80 students in December. Training for SIP teachers brought in 30 high school and 26 middle school faculty who were trained to bring courses into the classroom. Overall, *DNALC Asia* offered individual courses with a total enrollment of 4,655 students in 2019.

### Beijing 166, China

Aiming to improve biology education at secondary schools in China, the DNALC established a licensed center at Beijing 166 in 2014. Under the collaboration contract, Beijing students and teachers come to New York to attend two- or three-week camps during the summer and winter; additionally, DNALC instructors conduct workshops in Beijing for a total of four weeks in the spring and fall each year. In 2019, 299 BJ 166 students and 20 teachers attended DNALC camps and workshops. In April 2019, DNALC executive director Dave Micklos and international collaboration manager Catherine Zhang traveled with 112 Beijing students to biodiversity hot spots in southern China to collect samples for DNA barcoding research. Despite rainy weather, the students collected and developed DNA barcodes from 244 samples; 57 DNA sequences have been published in GenBank.

The experiments in hands-on biology education and student research the DNALC has conducted at Beijing 166 are being noticed by innovative educators, especially at the international and foreign language schools in China. In December 2019, Dave was invited to visit schools at Shenzhen, Dongguan, and Suzhou, giving talks to a total of 750 educators, parents, and students. Two leading



(*Left*) Dr. Sharon Pepenella (*center*) trained 26 students working on biodiversity and consumer issue–focused projects during a two-week DNA barcoding research course at *DNALC Asia*.

international schools—Tsinglan School and Shen Wai International School—intend to become DNALC partner schools in the near future.

### Nigeria/Bowie State

On Wednesday November 20, a formal license to operate and host the *DNA Learning Center Nigeria* was awarded to Godfrey Okoye University (GOU) in Enugu state, Nigeria. This is the result of a multiyear collaboration between the DNALC and GOU, facilitated through our close collaborator, Dr. George Ude of Bowie State University. DNALC Scholar and NYU graduate Michael Okoro leads



Amanda McBrien leads a Beijing 166 student camp in August.

the project and oversaw the refurbishment of the new center. GOU Vice Chancellor Christian Anieke provided a dedicated building for the DNALC on the new GOU Ugwuomu campus. The DNALC provided the lab design and $50,000 for equipment, including its signature lab table, as well as continued salary support for Michael. Over the next year, the center will begin to offer programs that benefit students and teachers at GOU and colleges throughout Nigeria.

## Dissemination at Professional Meetings

As in previous years, we continued to disseminate our programs at meetings. We presented our Ötzi the Iceman activity and results from our Biotechnology in American High Schools research at NABT. DNALC staff presented DNA barcoding and metabarcoding at the Invertebrates in Education and Conservation Conference, International Plant and Animal Genome Conference (PAG), American Society for Microbiology Conference, Long Island Natural History Conference, Community College Undergraduate Experience Summit, NIH SEPA SciEd Conference, and NSF IUSE PI Conference. Our data science programs were also presented at PAG and BioCodigo de Barras Symposium, and our efforts to democratize science and science education were presented at the ISMB/ECCB Conference.

## Lab Instruction and Outreach

In 2019, 20,358 students attended laboratory field trips at our five facilities: Dolan DNA Learning Center, DNALC *West, Harlem DNA Lab, Regeneron DNALC*, and *DNALC NYC* at City Tech. In-school instruction programs reached 7,728 students and 1,157 students attended weeklong camps, including some three international campers from Mexico and Spain. Footlocker kits were used by 1,758 students, 262 of whom were conducting independent research through *UBP*, *UBRP*, or *BLI*.

A grant from National Grid Foundation paid tuition for field trips and in-school instruction for 693 students from the Central Islip UFSD. An additional 1,105 students from other public school districts received scholarships—including Amityville, Brentwood, Malverne, Roosevelt, Uniondale, Connetquot, William Floyd, Ossining, and Valley Stream.

This year 2,667 (58%) of the students who attended field trips at *Harlem DNA Lab* and *DNALC NYC* at City Tech came from Title I schools that qualified for tuition assistance. The William Townsend Porter Foundation subsidized 20% of student scholarships for students visiting the *Harlem DNA Lab*. An additional 14 students from IS 59 in Queens received sequential lab instruction at DNALC *West* as part of an ongoing collaboration with Northwell Health.

The *Partner Member Program* continued to provide custom science sequences and advanced electives for seven independent schools in the tri-state region.

- Research teams from our newest member, Fontbonne Hall Academy, used DNA barcoding to create novel GenBank entries for flowering plants.
- Grace Church School offered a summer program that included using DNA barcoding to survey biodiversity of the plants and insects found near the school.
- At Marymount School of New York, genetics programs were incorporated as key parts of the biology curriculum, and students in molecular biology continued projects to analyze environmental DNA (eDNA) from NYC parks and the Hudson River.
- Research teams from Sacred Heart Greenwich used DNA barcoding to identify shellfish in food products and confirm identity of sushi products. One team used next-generation sequencing to analyze the effect of different ceramic surfaces on the microbiome.
- Lycée Français de New York continued to refine the eighth-grade forensics elective and offered *Human Genomics* and *Green Genes* camps during the summer.
- The Chapin School implemented genetics programs at several grade levels, including the advanced Molecular Genetics elective.
- St. David's School integrated basic genetics and DNA barcoding programs with existing curricula in grades five and eight.

As part of ongoing local partnerships, eight students from St. Dominic High School received daily instruction by DNALC educators. Students enrolled in the Molecular and Genomic Biology Research course visited the DNALC each afternoon for customized laboratory experiences in DNA barcoding as well as DNA and genome science. DNALC educators also worked with 22 students from Cold Spring Harbor High School's ninth-grade research program to do a survey of the biodiversity of Cold Spring Harbor using DNA barcoding. All students in both classes participated in the *Barcode Long Island* research program.

This year we had 5,041 visitors to the Ötzi the Iceman exhibit, either on their own or as part of a field trip. With the success of the Ötzi exhibit, we focused our attention this year on completing the redesign of our additional exhibit space. Nine *Saturday DNA!* sessions drew 277 participants who learned about DNA isolation, crime scene analysis, genetically modified foods and gel electrophoresis, ancient humanity, the science of the five senses, genetic engineering, and Mendelian inheritance. A microbial masterpiece created at our fall Agar Art session won first place in the American Society for Microbiology (ASM) Agar Art Kids contest. In this workshop,



(*Left*) Cold Spring Harbor High School's research program students spent a day in September collecting barcoding samples in and around Cold Spring Harbor. (*Right*) Then they documented and prepared samples for DNA isolation at the DNALC.

participants learned how laboratory techniques that scientists commonly use to study the living world could also be used to create unique works of art. In addition, DNALC staff presented as part of the *SUBMERGE Science Festival* at Hudson River Park, *Student Conference on Conservation Science* at the American Museum of Natural History, Bronx Center for Science and Mathematics Career Fair, STEM Teachers NYC Expo, *Pine Barrens Discovery Day* at Wertheim National Wildlife Refuge, and *Saturday Science for Students* at the Explorer's Club in Manhattan.

We continued to work with recipients of the *Junior Breakthrough Challenge*, an international competition in which young people submit short videos explaining big scientific ideas. Winners receive scholarships and a new laboratory for their school that is designed and equipped with help from the DNALC. This year we began work with the school of 2018 winner Samay Godika of the National Public School-Koramangala in Bangalore, India. The school building was already being refurbished, making this a perfect time to plan a new lab space! We have provided some designs and will begin working with faculty to purchase equipment.



National Geographic filmed archaeologist Patrick Hunt and DNALC's Lindsay Barone with our Ötzi replica in May for the *Iceman Murder Mystery: Lost in the Ice* documentary.

As part of our ongoing partnership with CSHL Women in Science and Engineering (WiSE), we hosted the third WiSE *Fun with DNA* summer camp. Held on the main campus of CSHL in Delbrück Laboratory, 21 young female science enthusiasts, two of whom received WiSE scholarships to attend, had the opportunity to meet and interact with enthusiastic female role models pursuing careers in the sciences. Each afternoon, the girls participated in WiSE activities on herd immunity, neuroscience, and astrophysics. They also took a "field trip" to Uplands Farm to tour the greenhouses and learn about the rich history of plant research at CSHL.

The DNALC has long been interested in reaching diverse audiences and communities and has made progress through scholarships and our locations in Harlem and Brooklyn. This year, we created a summer camp exclusively to reach URM students underrepresented in the sciences. The Science Technology AND Research Scholars (STARS) program is a two-week research experience designed to support the next generation of minority scientists, doctors, and health professionals. STARS provides students with state-of-the-art laboratory and computer science skills needed to succeed in STEM in college and beyond. Led by DNALC Assistant Director Jason Williams and Middle School Educator Brittany Johnson, this program also involved collaboration with Dr. Carol Carter, a professor at Stony Brook University, Dr. Paul Lichtman, a research coordinator at Adelphi University, and David Johnson, a student from the CSHL graduate school. Together, this team designed a curriculum that exposed students to DNALC laboratory activities, guest lectures from CSHL graduate students and researchers, a tour of the Stony Brook University campus and medical school, and guidance on research opportunities.



The camp attracted 27 students from 15 school districts; 80% had not previously attended DNALC programs. After a Saturday orientation for the students and their families, the first week focused on laboratory work, and the second week focused on bioinformatics and computer coding. The experience concluded with students

Students in the STARS program prep a gel.

(*Left*) Jason Williams leads a lab during week one. (*Right*) STARS students during bioinformatics session in week two.

presenting their work to their families—with the goal of sharpening their science communication skills and educating their families on STEM careers. Several participants have gone on to pursue independent high school research projects. We hope that this program will become a key element of CSHL's growing commitment to generate a "pipeline" of minority students in STEM higher education and careers.

Our collaboration with the CSHL School of Biological Sciences continued with exposing graduate students to skills needed to communicate science to a variety of audiences. As part of their required curriculum, first-year graduate students work with DNALC instructors to complete 12 half-day sessions in which they progress from classroom observation to lesson planning to co-instructing alongside a DNALC staff member to independently leading lab classes. Students learn classroom management skills—including how to quickly assess an audience and customize a presentation accordingly. Graduate students interact equally with both middle- and high-school-aged students during their required rotations, then complete three elective classes in which they implement their new skills.

### *BioMedia* Visitation and Projects

In 2019, 5.2 million visitors accessed our suite of multimedia resources. Google Analytics counted 3.7 million visits to DNALC websites, our YouTube videos received 883,944 views, and the *3D Brain*, *Weed to Wonder*, and *Gene Screen* smartphone/tablet apps were downloaded 590,471 times. In-app purchases of 3D Brain netted $6,175 for the year.

We completed a total redesign of the DNALC.org site, giving it a fresh look, better organization, and easier navigation. We worked with the Public Affairs Department to prepare for a seamless merger with the Cold Spring Harbor Laboratory (www.cshl.edu) website by transitioning the DNALC site domain to https://dnalc.cshl.edu.

The *BioMedia* Group continued to support the educational objectives of the DNALC through web design and programming, print design, photography, videography, and lab classroom layout planning for collaborators around the world. We followed up on our earlier development of *Our Human Inheritance*, a museum exhibit that features Ötzi the Iceman and ancient human ancestors in the main gallery space. Working on content and design with DNALC educators, the *BioMedia* staff completed displays for the rear gallery in December. Similar to the large mural of the Italian Alps in the Ötzi exhibit, the focal point of the new display is a stunning floor-to-ceiling, 30-foot-long image of the universe that serves as a backdrop for a timeline of the history of life on Earth and some of the key developments that have allowed life to flourish on our planet. In an exploration of the processes and outcomes of evolution, the exhibit showcases the evolution of the eye—from
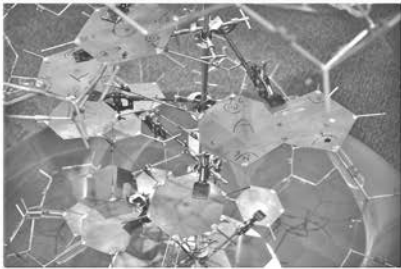
DNA Learning Center exhibits

a simple light detecting eyespot to the compound eye. Additionally, several interesting human evolution stories connect the new space to the existing exhibits on human ancestry. A touchscreen with an interactive chromosome map enables students to explore the human genome one gene at a time. Finally, an interactive human variation wall highlights how traits manifest themselves in different people—holding a literal mirror up to our visitors and allowing them to explore some of their own traits.

## Staff and Interns

During the year, the DNALC staff was strengthened by the addition of Brittany Johnson, Justin Burke, Jennifer Hackett, Ph.D., Louise Bodt, Lina Bader, and Lina Ruiz-Grajales to the education and instruction staff and Daniel Jacobs to the *BioMedia* group.

Brittany Johnson started in January as a middle-school educator. A native of Long Island, Brittany remembers visiting the DNALC as a child and being captivated by the "Mystery of Anastasia." She received a B.A. in biology from Fisk University and a Master of Biological Medical Sciences from Mississippi College. While volunteering in the Central Islip School, Brittany met a DNALC educator who had come to provide an in-school lab. Brittany pitched in with the class, impressed us, and was offered the next vacant position.

Justin Burke joined the DNALC in February as our laboratory technician. He is responsible for testing, assembling, and organizing supplies and reagents for all DNALC instruction, as well as managing our high-school and college intern programs. Justin is a native Long Islander and has been interested in science his whole life. His work at the DNALC is preparing him to return to his studies of biochemistry and cell biology at Stony Brook University.

In June Daniel Jacobs joined our *BioMedia* Group as a programmer charged with maintaining, designing, and updating bioinformatics tools for our popular *DNA Subway* website. Daniel's interest in coding began in high school. He initially studied physics at Adelphi University, but then transferred to Queens College and earned his computer science degree. While there, he became "multilingual" in programming languages and learned Python to help with research on factors affecting the sustainability of world peace.

Jennifer Hackett joined our NYC education team in July. Jenny attended DePauw University, where she developed her passion for science through research in the Science Research Fellows Program, including as part of the team that discovered the gene for frontotemporal dementia. She completed her Ph.D. thesis at Johns Hopkins University School of Medicine, where she studied the role of telomere dysfunction with Nobel laureate and former CSHL researcher Carol Greider. While developing shRNA libraries for genome-wide screens as a postdoc at Harvard Medical School, she volunteered with programs for children through Boston Cares. This motivated her to join the NYC Teaching Fellows and ultimately teach science at the prestigious Dalton School in Manhattan. She also consulted on the creation of BSCS/NIH curriculum supplements and is the author of *Molecular Biology: Concepts for Inquiry*, a high-school textbook and curriculum.

Louise Bodt started in August as an educator and *UBP* manager based out of our City Tech Temporary Lab. A Brooklyn native, Louise participated in science classes and internships at the American Museum of Natural History (AMNH) during high school, which sparked her interest in genetics. She earned a B.A. from Smith College, where she studied molecular biology and worked in a parasitology lab. After three years teaching science at two NYC private schools, she taught at the AMNH while receiving her M.S. in Biology from NYU, where she focused on the population genetics of European starlings.

Lina Bader joined the DNALC as an instructor for *Regeneron DNALC*. After a Bachelor's degree in biology from the University of Pennsylvania and graduate program for education, Lina taught biology at a Philadelphia public school. Although she loved classroom teaching, she was limited in her desire to implement lab teaching in molecular genetics. Being familiar with DNALC resources, Lina jumped at the opportunity to join the DNALC team.

(*Top*, *left* to *right*) Brittany Johnson, Justin Burke, Daniel Jacobs, and Jenny Hackett. (*Bottom*, *left* to *right*) Louise Bodt, Lina Ruiz-Grajales, and Lina Bader.

Lina Ruiz-Grajales joined the DNALC in November as an instructor for *Regeneron DNALC*. After four years as a pharmaceutical chemistry student at the University of Antioquia in Colombia, she moved to New York and enrolled at Purchase College, where she studied the effects of climate on the model plant *Arabidopsis thaliana.* As an Amgen summer scholar at UC Berkeley, she explored the role of the plant microbiomes in protection against plant pathogens. Her strong research background and knowledge of metabarcoding made her a natural to help with our student research programs using DNA barcodes.

We said goodbye to three staff members in 2019: genetics educators Alison Cucco and Pauline McGlone and *Urban Barcode Project* manager Christine Marizzi.

After eight fruitful years at the DNALC, Christine Marizzi accepted a position in September as a lead community scientist at BioBus. Christine started as manager of DNALC *West*, but quickly took on management of our two barcoding projects in New York City: *UBP* and *UBRP*. Christine did it all, from recruiting student researchers and faculty mentors to organizing the annual symposia for 500+ participants. She brought her strong love of citizen science to the DNALC, initiating our collaboration with Genspace and organizing a monumental "agar art" map of Manhattan that garnered worldwide attention.

In the spring Pauline McGlone earned a Master's degree in healthcare administration from Hofstra University and accepted a position as a project associate at NYU Langone. Pauline started her journey at the DNALC as a high-school intern in 2012. While attending college locally, Pauline continued to work as a college intern. She embraced this learning experience, which helped her with college biology courses and required labs. After interning for five years, Pauline came full circle—transitioning to a middle-school educator who taught the labs she had prepped as an intern.

Alison Cucco left the DNALC in the spring to become environmental compliance coordinator for PSE&G. Alison was an educator for the *Harlem DNA Lab* and Partner Member schools. She provided customized in-school instruction, and assisted with student barcoding projects, including a pilot program using environmental DNA to monitor the health of the Hudson River Estuary.

Since the DNALC opened, we have relied on high school and college interns to support our day-to-day operations. An internship offers students the unique opportunity to gain real laboratory or design experience in an educational environment. We gathered an amazing group of interns this year, and said farewell as others left for college.

**High School Interns**

| | |
|---|---|
| Jacqueline Albert, Syosset High School | Sarah Nace, Walt Whitman High School |
| Christopher Catalano, Garden City High School | Jack O'Hara, St. Anthony's High School |
| Christopher Cizmeciyan, Syosset High School | Julia Padro, Grace Church School |
| Kaela Deriggi, St. Anthony's High School | Aveline Roderick, St. Anthony's High School |
| Thomas Kamara, All Hallows High School | Mina Samaras, Plainedge High School |
| Brady Lyons, St. Dominic High School | Samantha Sgrizzi, Huntington High School |
| Ava Maiella, Harborfields High School | Esha Sharma, Syosset High School |
| Ethan McGuinness, Huntington High School | Michael Stabile, Plainedge High School |
| James McKechnie, Northport High School | Nicholas Stabile, Plainedge High School |
| Sonja Michaluk, Hopewell Valley Central High School | Alejandro Wiltshire, St. Mary's High School |

**High School Interns Departing for College**

| | |
|---|---|
| Yusiry Acevedo Nunez, Farmingdale State College | Randy Diaz Arias, University of Rochester |
| Gavin Calabretta, Cornell University | Sibelle O'Donnell, University of Southern California |
| Elijah Calle, University of Buffalo | |

**College Interns**

| | |
|---|---|
| Nadia Alomari, New York City College of Technology | Jillian Maturo, Boston College |
| Gabrielle Blazich, Fordham University | William McBrien, Stony Brook University |
| Taehwan Cha, New York University | Katherine Parra, New York City College of Technology |
| Megan Erhardt, University of New Haven | Joni Sebastiano, Stony Brook University |
| Omotayo Ikuomenisan, Hunter College | Jon Triscari, University of Rochester |
| Isabella Martino, Stony Brook University | |

# Workshops and Visitors

| | |
|---|---|
| January 7–8 | Software Carpentry Workshop, University of Arkansas, Fayetteville, Arkansas |
| January 10–11 | Maize Annotation Jamboree, Scripps Institute of Oceanography, San Diego, California |
| January 11 | The Central Pine Barrens Cooperators Meeting, Hyatt Place Long Island/East End, Riverhead, New York |
| January 12 | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| January 12–16 | International Plant and Animal Genome XXVII Conference 2019, CyVerse Education Sessions: "CyVerse Software, Tools, and Services for Data-Driven Discovery, Data Science, and Education," "Advanced Computational Methods—Machine Learning, Contains, and Clouds," "Biochemical and Bioinformatics Infrastructure to Support Metabarcoding CUREs," San Diego, California |
| January 17 | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| January 19 | *Saturday DNA!* "Enzymes in Action," DNALC |
| January 22 | NIH *Barcode Long Island* Open Lab, DNALC |
| | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| January 25 | Site visit by Armando Barriguete and Hugo Scherer, Mexico DNA Learning Center Development, Mexico City, Mexico |
| January 30 | Site visit by Vision Gifted Chinese Children's School, Shenzhen, China |
| January 31 | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| February 1 | *RNA-Seq With DNA Subway* Webinar, DNALC |
| February 2 | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| February 7 | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| February 8 | *DNA Barcoding* Teacher Workshop, DNALC @ City Tech |
| February 9 | *Saturday DNA!* "BioArt" DNALC |
| | *Ötzi the Iceman* Tour, DNALC |
| February 11–22 | *DNA Science, DNA Barcoding and Research* Workshops, Beijing 166 School, DNALC |
| February 12 | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| February 19–20 | Go Fish - eDNA Teacher Workshop, Pier 84, New York, New York |
| February 19–22 | *Urban Barcode Research Program Conservation Genetics* Workshop, *Harlem DNA Lab* |
| February 25 | Arkansas Bioinformatics Consortium Annual Meeting, "CyVerse Cyberinfrastructure for Research and Education in Genomics and Metagenomics," University of Arkansas for Medical Sciences, Little Rock, Arkansas |
| February 28 | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| March 2 | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |

| | |
|---|---|
| March 8 | Site visit by Passaic County Community College Delegation, Paterson, New Jersey |
| March 9 | NIH *Barcode Long Island* Open Lab, Stony Brook University, Stony Brook, New York |
| March 9 | Intrepid Outreach Day, "*Considering eDNA*," Intrepid Sea, Air & Space Museum, Pier 86, New York |
| March 12 | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| March 13–14 | Maize Annotation Jamboree, Washington University, St. Louis, Missouri |
| March 14–17 | Maize Genetics Conference, "Evaluating Community Curation Approaches for Improving Annotation on Classical Maize Gene Models," Poster Session, Washington University, St. Louis, Missouri |
| March 16 | NIH *Barcode Long Island* Open Lab, Brookhaven National Laboratory, Upton, New York |
| | *Saturday DNA!* "Ancient Ancestry," DNALC |
| | *Ötzi the Iceman* Tour, DNALC |
| March 21 | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| March 26 | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| Mar 27–Apr 14 | *DNA Barcoding and Research* Workshops, Beijing 166 School, DNALC |
| April 2–5 | NSF NEON Diversity in Data Science Conference, "Broadening Participation in Data Science," NEON, Boulder, Colorado |
| April 6 | NIH *Barcode Long Island* Open Lab, Brookhaven National Laboratory, Upton, New York |
| | NIH *Barcode Long Island* Open Lab, DNA Learning Center |
| | *West Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| April 10 | "DNA Barcoding Research: The First Step in a Life of Science," Lecture, DNALC Asia, Suzhou, China |
| April 11 | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| April 13 | NIH *Barcode Long Island* Open Lab, Stony Brook University, Stony Brook, New York |
| | *Saturday DNA!* "WiSE Presents: Get to Know GMOs!" DNALC |
| April 16 | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| April 17 | NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC |
| April 22–26 | *DNA Barcoding* and *Bioinformatics* Training Workshop, DNALC @ City Tech |
| April 24 | NIH SciEd Conference, "Barcoding Long Island," Poster Session, Grand Hyatt Washington Hotel, Washington, D.C. |
| April 27 | NIH *Barcode Long Island* Open Lab, DNALC |
| April 29 | Regeneron Cultivation Event, Regeneron, Tarrytown, New York |
| April 30 | *Urban Barcode Research Program* Update Event, The Irondale Center for Theater, Education, and Outreach, Brooklyn, New York |
| May 8 | NIH *Barcode Long Island—Bioinformatics* Open Lab, DNALC |
| May 9 | *Ötzi the Iceman* Tour, DNALC |
| May 11 | BioCodigo de Barras Symposium, "Proyectos de Investigacion Educativa a Traves de Biocodigos de Barras," National Institute of Genomic Medicine, Ciudad de Mexico, Mexico |
| | *Saturday DNA!* "April Showers Bring May Flowers," DNALC |
| | *Ötzi the Iceman* Tour, DNALC |
| May 15 | City Tech Cultivation Event, DNALC @ City Tech |
| May 18–19 | National Geographic Filming @ DNALC |
| May 20 | "The Slippery Slope of Eugenical Thinking," Cold Spring Harbor High School Science Symposium, Cold Spring Harbor, New York |
| May 21 | Regeneron Cultivation Event, Regeneron, Tarrytown, New York |
| May 29 | "Iceman - Ötzi's Final Days," Cinema Arts Centre, Huntington, New York |
| May 30 | *Urban Barcode Project*/Pinkerton *Urban Barcode Research Program* Symposium, New York Academy of Medicine, New York, New York |
| May 30–31 | NSF CyVerse *Genomics Data Carpentry* Workshop, University of Arizona, Tucson, Arizona |
| June 1 | *Saturday DNA!* "Dust Away Crime: The Truth About Fingerprints," DNALC |
| | *Ötzi the Iceman* Tour, DNALC |
| June 4 | *Barcode Long Island* Student Symposium, CSHL |
| | Introduction to Regeneron Event, Rye Country Day School, Rye, NY |
| June 7 | St. David's School Science Expo, St. David's School New York, New York |
| June 8 | Cold Spring Harbor Laboratory Open House, CSHL |
| | *Ötzi the Iceman* Tour, DNALC |
| June 10–14 | DNA Barcoding for CURES Workshop, James Madison University, Madison, Wisconsin |
| | *Genome Science* Workshop, Lycée Français, New York, New York |
| | *Green Genes* Workshop, Lycée Français, New York, New York |
| June 11 | *Ötzi the Iceman* Tour, DNALC |
| June 13 | *Ötzi the Iceman* Tour, DNALC |
| June 13–July 3 | *Biotechnology* Workshops, Grace Church High School, New York, New York |
| June 18 | *Ötzi the Iceman* Tour, DNALC |

| | |
|---|---|
| June 23–28 | Gordon Research Conferences—Undergraduate Biology Education Research, Bates College, Lewiston, Maine |
| June 24–28 | *BioCoding* Workshop, Toms River High School East, Toms River, New Jersey |
| June 25 | 5th Annual BOP Research Symposium, Governors Island, New York, New York |
| June 30 | CSHL Frontiers in Plant Science Workshop, "240 - Minute R Tutorial," CSHL |
| July 1 | HRPT Collaborator Training, Pier 84, Hudson River Park, New York, New York |
| July 1–5 | *Fun with DNA* Workshop, DNALC |
| | *Genome Science* Workshop, DNALC |
| | *Green Genes* Workshop, DNALC |
| | *World of Enzymes* Workshop, DNALC |
| | Pinkerton *Urban Barcode Research Program Conservation Genetics* Workshop, *Harlem DNA Lab* |
| July 8–12 | *BioCoding* Workshop, DNALC |
| | *DNA Science* Workshop, DNALC |
| | *Forensic Detectives* Workshop, DNALC |
| | *Fun with DNA* Workshop, DNALC |
| | *World of Enzymes* Workshop, DNA Learning Center *West* |
| | *Fun with DNA* Workshop, Portledge School, Locust Valley, New York |
| | *World of Enzymes* Workshop, Toms River High School East, Toms River, New Jersey |
| July 15–19 | *DNA Barcoding* Workshop, DNALC |
| | *BioCoding* Workshop, DNALC |
| | *Green Genes* Workshop, DNALC |
| | *Fun with DNA/World of Enzymes* Workshops, Beijing 166, DNALC (2 sessions) |
| | *Pinkerton Urban Barcode Research Program DNA Barcoding and Bioinformatics* Workshop, *Harlem DNA Lab* |
| | *DNA Science* Workshop, DNA Learning Center *West* |
| | *World of Enzymes* Workshop, Portledge School, Locust Valley, New York |
| July 17 | *Ötzi the Iceman* Tour, DNALC |
| July 17–Aug 5 | *DNA Barcoding* Workshop, DNALC Asia, Suzhou, China |
| July 22 | NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC |
| July 22–23 | "DNA Barcoding: Uncovering Hidden Biodiversity in Your Own Back Yard," K–12 Summer Institute, Kerrville, Texas |
| July 22–26 | *DNA Science* Workshop, DNALC |
| | *World of Enzymes* Workshop, DNALC |
| | *Forensic Detectives* Workshop, Beijing 166, DNALC |
| | *Green Genes* Workshop, Beijing 166, DNALC |
| | *Green Genes* Workshop, DNA Learning Center *West* |
| | Pinkerton *Urban Barcode Research Program Conservation Genetics* Workshop, *Harlem DNA Lab* |
| July 24 | ISMB/ECCB Conference, "Overview of CyVerse Tools & Services: Intro to Data/Metadata Management," "Training, Technology, Togetherness—Promoting Knowledge Exchange in Life Sciences through Communities of Practice," Basel, Switzerland |
| | *Ötzi the Iceman* Tour, DNALC |
| July 26 | "Cyberinferstructure for Scaling Research, Education, and People," Lecture, Swiss Institute of Bioinformatics, Lausanne, Switzerland |
| | Site visit by Emily Zeng and Xiaoli Wu, Shen Wai International School, Shenzhen, China |
| July 29–Aug 2 | *DNA Science* Workshop, DNALC |
| | *Fun with DNA* Workshop, DNALC |
| | *Forensic Detectives* Workshop, Beijing 166, DNALC |
| | *Green Genes* Workshop, Beijing 166, DNALC |
| | *Fun with DNA* Workshop, DNA Learning Center *West* |
| | *Pinkerton Urban Barcode Research Program DNA Barcoding and Bioinformatics* Workshop, *Harlem DNA Lab* |
| July 30–Aug 4 | Invertebrates in Education & Conservation Conference, El Conquistador Hilton, Tucson, Arizona |
| July 31 | Site Visit by Nan Gerson, Bethpage Federal Credit Union, Bethpage, New York |
| August 1 | ASM Conference for Undergraduate Educators (ASMCUE), "Course-based Microbiome Research," Sheraton Tysons Hotel, Tysons, Virginia |
| August 5–9 | *Being Human* Workshop, DNALC |
| | *DNA Science* Workshop, DNALC |
| | *Green Genes* Workshop, DNALC |
| | *World of Enzymes* Workshop, DNALC |
| | *DNA Science* Workshop, DNA Learning Center *West* |
| | *DNA Science* Workshop, DNALC @ City Tech |
| | Pinkerton *Urban Barcode Research Program Conservation Genetics* Workshop, *Harlem DNA Lab* |
| August 12–16 | *DNA Barcoding* Workshop, DNALC |

|  | *DNA Science* Workshop, DNALC |
|---|---|
|  | *Fun with DNA* Workshop, DNALC |
|  | *World of Enzymes* Workshop, DNALC |
|  | *Forensic Detectives* Workshop, DNA Learning Center *West* |
|  | *DNA Barcoding* Workshop, DNALC @ City Tech |
|  | *Pinkerton Urban Barcode Research Program DNA Barcoding and Bioinformatics* Workshop, *Harlem DNA Lab* |
|  | NIH *Barcode Long Island* Teacher Workshop, Hyatt Place East End, Riverhead, New York |
| August 14 | *Ötzi the Iceman* Tour, DNALC |
| August 16 | Site visit by Kim Libertini, Cold Spring Harbor Central School District, Cold Spring Harbor, New York |
| August 19–23 | *Fun with DNA* Workshop, DNALC |
|  | *Forensic Detectives* Workshop, DNALC |
|  | *Genome Science* Workshop, DNALC |
|  | *Green Genes* Workshop, DNALC |
|  | *Fun with DNA* Workshop, DNA Learning Center *West* |
|  | *Fun with DNA* Workshop, DNALC @ City Tech |
|  | *STARS DNA Barcoding* Workshop, CSHL |
|  | *DNA Barcoding* and *Bioinformatics* UBP Teacher Workshop, *Harlem DNA Lab* |
| August 26–30 | *DNA Science* Workshop, DNALC |
|  | *Fun with DNA* Workshop, DNALC |
|  | *Green Genes* Workshop, DNALC |
|  | *World of Enzymes* Workshop, DNALC |
|  | WiSE *Fun with DNA* Workshop, CSHL |
|  | *World of Enzymes* Workshop, DNA Learning Center *West* |
|  | *STARS BioCoding* Workshop, DNALC |
| August 28 | *Ötzi the Iceman* Tour, DNALC |
| September 14 | Pine Barrens Discovery Day, Wertheim National Wildlife Refuge, Shirley, New York |
| September 23–24 | Regeneron Software Carpentry Workshop, "Reproducible Analysis in R," Regeneron, Tarrytown, New York |
| September 23–27 | Week of Science Student Workshops, *"Human Family,"* South Tyrol Museum, Bolzano, Italy |
|  | SUBMERGE Marine Science Festival, "DNA Barcoding for Biodiversity Research," Hudson River Park, New York, New York |
| September 26 | "Urban Barcode Research," Kickoff Event, The Irondale Center for Theater, Education, and Outreach, Brooklyn, New York |
| October 5 | *Saturday DNA!* "Agar Art," DNALC |
| October 21 | *"DNA Restriction Analysis,"* Teacher Workshop, Math for America, New York, New York |
| October 22 | NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC |
| October 23 | Regeneron Training Session "Day of Doing Good," Regeneron, Tarrytown, New York |
|  | Demystifying Science Reimagined at CSHL, "DNA Barcoding: Infrastructure for Student and Citizen Science," CSHL |
| October 24 | "Urban Barcode Research," Kickoff Event, The Irondale Center for Theater, Education, and Outreach, Brooklyn, New York |
| October 25 | Day of Doing Good Science Expo, *"DNA Extraction from Wheat Germ," "Mutant Organisms," and "Diversity of Life,"* Regeneron Pharmaceuticals, Tarrytown, NY |
| Oct 31–Nov 7 | Site visit by Dr. Peter Bickerton, Earlham Trust, Norwich, UK |
| November 1 | Site visit by Michael Maturo, Frank Pusinelli, and David Garten, RXR Realty, LLC, Uniondale, New York |
| November 2 | *Saturday DNA!* "Making Sense of Your Senses," DNALC |
| November 4 | *"DNA Restriction Analysis,"* Teacher Workshop, Math for America, New York, New York |
| November 5 | *"DNA Restriction Analysis," and "Bacterial Transformation,"* Teacher Workshop, Regeneron, Tarrytown, New York |
| November 15 | NABT Conference, "The Last Days of Ötzi," Sheraton Grand, Chicago, Illinois |
| November 16 | NABT Conference, "Biotechnology in American High Schools: Then and Now," Sheraton Grand, Chicago, Illinois |
| November 18 | *"DNA Restriction Analysis,"* Teacher Workshop, Math for America, New York, New York |
| November 18–22 | NSF CyVerse *Train the Trainer* Workshop, Technical University of Graz, Graz, Austria |
| November 21–22 | Community College Undergraduate Experience Summit, "DNA Barcoding: The CURE for Citizen Science," Poster Session, Omni Shoreham Hotel, Washington, DC. |
| December 3 | Invited BD2k Seminar, "Computational Thinking, Learning, and Doing in 21st Century Biology," University of Puerto Rico, San Juan |
| December 3–13 | *Genome Science* Workshop, Beijing 166 School, DNALC |
| December 4 | Tech Night @ Jack Abrams STEM Magnet School, Jack Abrams STEM Magnet School, Huntington Station, NY |
| December 4–5 | CyVerse Workshop, "Computational Tools and Reproducibility Workshop," University of Puerto Rico, San Juan |
| December 6 | Shelter Island Science Fair Judging, Shelter Island, New York |

| December 7 | "Biotechnology in American High Schools and Asian Models of DNALC Practice," Lecture, Shen Wai International School, Shenzhen, China |
| | NIH *Barcode Long Island* Open Lab, DNALC |
| | *Saturday DNA!* "DNA Detectives" DNALC |
| | *Ötzi the Iceman* Tour, DNALC |
| December 8 | "Biotechnology in American High Schools and Asian Models of DNALC Practice," Lecture, Tsinglan School, Dongguan, China |
| December 9 | "The Rules of Life: Thinking Like a Biologist," Lecture, SIP No. 2 Senior High School, Suzhou, China |
| | Site visit by Laura Slatkin, Nest Fragrances, New York, New York |
| December 10 | "The View from Nowhere in Computational Infrastructure," Lecture, University of Scotland, Dundee, U.K. |
| December 12 | Regeneron DNALC Launch Event, Regeneron, Tarrytown, New York |
| December 14 | *Urban Barcode Project* Open Lab, *Harlem DNA Lab* |
| December 18 | Site visit by Nancy Lippman and Carissa Jordan, CSHL Association Directors, Cold Spring Harbor, New York |

## Sites of Major Faculty Workshops

Program Key:    *Middle School*        High School        **College**

| *State* | *Institution* | *Year(s)* |
| --- | --- | --- |
| ALABAMA | University of Alabama, Tuscaloosa | 1987–1990 |
| | **Hudson Alpha Institute, Huntsville** | **2014** |
| ALASKA | **University of Alaska, Anchorage** | **2012** |
| | **University of Alaska, Fairbanks** | 1996 |
| ARIZONA | Arizona State University, Tempe | 2009 |
| | Tuba City High School | 1988 |
| | **University of Arizona, Tucson** | **2011, 2019** |
| | **United States Department of Agriculture, Maricopa** | **2012** |
| ARKANSAS | Henderson State University, Arkadelphia | 1992 |
| | **University of Arkansas, Fayetteville** | **2017, 2019** |
| | **University of Arkansas, Little Rock** | **2012** |
| | **University of Arkansas for Medical Sciences, Little Rock** | **2019** |
| CALIFORNIA | **California State University, Dominguez Hills** | **2009** |
| | **California State University, Fullerton** | **2000** |
| | **California State University, Long Beach** | **2015** |
| | **California Institute of Technology, Pasadena** | **2007** |
| | **Chan Zuckerberg Biohub, San Francisco** | **2018** |
| | Canada College, Redwood City | 1997 |
| | City College of San Francisco | 2006 |
| | **City College of San Francisco** | **2011, 2013** |
| | Contra Costa County Office of Education, Pleasant Hill | 2002, 2009 |
| | **Foothill College, Los Altos Hills** | **1997** |
| | Harbor-UCLA Research & Education Institute, Torrance | 2003 |
| | **Los Angeles Biomedical Research Institute (LA Biomed), Torrance** | **2006** |
| | Laney College, Oakland | 1999 |
| | Lutheran University, Thousand Oaks | 1999 |
| | Oxnard Community College, Oxnard | 2009 |
| | **Pasadena City College** | **2010** |
| | **Pierce College, Los Angeles** | **1998** |
| | Salk Institute for Biological Studies, La Jolla | 2001, 2008 |
| | **San Francisco State University** | 1991 |
| | **San Diego State University** | **2012** |
| | **San Jose State University** | **2005** |
| | Santa Clara University | 2010 |
| | **Scripps Institute, San Diego** | **2019** |
| | **Southwestern College, Chula Vista** | **2014–2015** |
| | **Stanford University, Palo Alto** | **2012** |
| | **University of California, Berkeley** | **2010, 2012** |
| | University of California, Davis | 1986 |
| | **University of California, Davis** | **2012, 2014–2015** |

|  | | |
|---|---|---|
| | **University of California, Long Beach** | **2015** |
| | **University of California, Northridge** | **1993** |
| | University of California, Riverside | 2011 |
| | **University of California, Riverside** | **2012** |
| | **University of California, San Francisco** | **2015** |
| COLORADO | Aspen Science Center | 2006 |
| | Colorado College, Colorado Springs | 1994, 2007 |
| | **Colorado State University, Fort Collins** | **2013, 2018** |
| | **Community College of Denver** | **2014** |
| | **United States Air Force Academy, Colorado Springs** | **1995** |
| | University of Colorado, Denver | 1998, 2009–2010 |
| CONNECTICUT | Choate Rosemary Hall, Wallingford | 1987 |
| | **Jackson Laboratory, Farmington** | **2016** |
| DELAWARE | **University of Delaware, Newark** | **2016** |
| DISTRICT OF COLUMBIA | **Howard University, Washington** | **1992, 1996, 2009–2010** |
| FLORIDA | Armwood Senior High School, Tampa | 1991 |
| | Florida Agricultural & Mechanical University, Tallahassee | 2007–2008 |
| | **Florida Agricultural & Mechanical University, Tallahassee** | **2011** |
| | **Florida SouthWestern State University, Fort Myers** | **2015** |
| | North Miami Beach Senior High School | 1991 |
| | **Seminole State College, Sanford** | **2013–2014** |
| | University of Florida, Gainesville | 1989 |
| | University of Miami School of Medicine | 2000 |
| | University of Western Florida, Pensacola | 1991 |
| GEORGIA | Fernbank Science Center, Atlanta | 1989, 2007 |
| | **Gwinnett Technical College, Lawrenceville** | **2011–2012** |
| | **Morehouse College, Atlanta** | **1991, 1996** |
| | Morehouse College | 1997 |
| | **Spelman College, Atlanta** | **2010** |
| | **University of Georgia, Athens** | **2015** |
| HAWAII | Kamehameha Secondary School, Honolulu | 1990 |
| | **University of Hawaii at Manoa** | **2012** |
| IDAHO | University of Idaho, Moscow | 1994 |
| ILLINOIS | Argonne National Laboratory | 1986–1987 |
| | iBIO Institute/Harold Washington College, Chicago | 2010 |
| | Illinois Institute of Technology, Chicago | 2009 |
| | **Kings College, Chicago** | **2014** |
| | **University of Chicago** | **1992, 1997, 2010** |
| | **University of Southern Illinois, Carbondale** | **2016** |
| INDIANA | Butler University, Indianapolis | 1987 |
| | **Purdue University, West Lafayette** | **2012** |
| IOWA | Drake University, Des Moines | 1987 |
| KANSAS | University of Kansas, Lawrence | 1995 |
| KENTUCKY | **Bluegrass Community & Technical College, Lexington** | **2012–2014** |
| | Murray State University | 1988 |
| | University of Kentucky, Lexington | 1992 |
| | Western Kentucky University, Bowling Green | 1992 |
| LOUISIANA | Bossier Parish Community College | 2009 |
| | Jefferson Parish Public Schools, Harvey | 1990 |
| | John McDonogh High School, New Orleans | 1993 |
| | **Southern University at New Orleans** | **2012** |
| | **University of New Orleans** | **2018** |
| MAINE | Bates College, Lewiston | 1995 |
| | **Southern Maine Community College** | **2012–2013** |
| | Foundation for Blood Research, Scarborough | 2002 |
| MARYLAND | Annapolis Senior High School | 1989 |
| | **Bowie State University** | **2011, 2015** |
| | Frederick Cancer Research Center | 1995 |
| | McDonogh School, Baltimore | 1988 |

|  |  |  |
|---|---|---|
|  | Montgomery County Public Schools | 1990–1992 |
|  | National Center for Biotechnology Information, Bethesda | 2002 |
|  | *St. John's College, Annapolis* | *1991* |
|  | **University of Maryland, School of Medicine, Baltimore** | **1999** |
| MASSACHUSETTS | **Arnold Arboretum of Harvard University, Roslindale** | **2011** |
|  | Beverly High School | 1986 |
|  | Biogen Idec, Cambridge | 2002, 2010 |
|  | **Boston University** | **1994, 1996** |
|  | CityLab, Boston University School of Medicine | 1997 |
|  | Dover-Sherborn High School, Dover | 1989 |
|  | Randolph High School | 1988 |
|  | The Winsor School, Boston | 1987 |
|  | Whitehead Institute for Biomedical Research, Cambridge | 2002 |
| MICHIGAN | Athens High School, Troy | 1989 |
|  | **Schoolcraft College, Livonia** | **2012** |
| MINNESOTA | **American Society of Plant Biologists, Minneapolis** | **2015** |
|  | Minneapolis Community and Technical College, Madison | 2009 |
|  | **Minneapolis Community and Technical College, Madison** | **2013** |
|  | University of Minnesota, St. Paul | 2005 |
|  | **University of Minnesota, St. Paul** | **2010** |
| MISSISSIPPI | Mississippi School for Math & Science, Columbus | 1990–1991 |
|  | **Rust College, Holly Springs** | **2006–2008, 2010** |
| MISSOURI | **St. Louis Science Center** | **2008–2010** |
|  | Stowers Institute for Medical Research, Kansas City | 2002, 2008 |
|  | **University of Missouri, Columbia** | **2012** |
|  | Washington University, St. Louis | 1989 |
|  | **Washington University, St. Louis** | **1997, 2011, 2019** |
| MONTANA | **Montana State University, Bozeman** | **2012** |
| NEBRASKA | **University of Nebraska-Lincoln, Lincoln** | **2014** |
| NEVADA | University of Nevada, Reno | 1992, **2014** |
| NEW HAMPSHIRE | Great Bay Community College, Portsmouth | 2009 |
|  | **New Hampshire Community Technical College, Portsmouth** | **1999** |
|  | St. Paul's School, Concord | 1986–1987 |
| NEW JERSEY | Coriell Institute for Medical Research, Camden | 2003 |
|  | Raritan Valley Community College, Somerville | 2009 |
| NEW MEXICO | Biolink Southwest Regional Meeting, Albuquerque | 2008 |
|  | **Los Alamos National Lab** | **2017** |
|  | **New Mexico State University, Las Cruces** | **2017** |
|  | **Santa Fe Community College, Santa Fe** | **2015** |
| NEW YORK | Albany High School | 1987 |
|  | American Museum of Natural History, New York | 2007, 2015 |
|  | Bronx High School of Science | 1987 |
|  | Brookhaven National Laboratory, Upton | 2015–2018 |
|  | Canisius College, Buffalo | 2007 |
|  | **Canisius College, Buffalo** | **2011** |
|  | City College of New York | 2012 |
|  | Cold Spring Harbor High School | 1985, 1987 |
|  | **Cold Spring Harbor Laboratory** | **2014–2015, 2018–2019** |
|  | **Columbia University, New York** | **1993** |
|  | **Cornell University, Ithaca** | **2005** |
|  | *DeWitt Middle School, Ithaca* | *1991, 1993* |
|  | Dolan DNA Learning Center | 1988–1995, 2001–2004, 2006–2009, 2015–2019 |
|  | **Dolan DNA Learning Center** | **1990, 1992, 1995, 2000–2011** |
|  | *Dolan DNA Learning Center* | *1990–1992* |
|  | DNA Learning Center *West* | 2005 |
|  | DNA Learning Center NYC | 2019 |
|  | Environmental Science Center, Bergen Beach, Brooklyn | 2015–2016 |
|  | *Fostertown School, Newburgh* | *1991* |

| | | |
|---|---|---|
| | *Harlem DNA Lab*, East Harlem | *2008–2009, 2011–2013,* *2016–2019* |
| | **Harlem DNA Lab, East Harlem** | **2015–2016** |
| | Huntington High School | 1986 |
| | Irvington High School | 1986 |
| | John Jay College of Criminal Justice | 2009 |
| | *Junior High School 263, Brooklyn* | *1991* |
| | *Lindenhurst Junior High School* | *1991* |
| | **Math for America** | **2017–2019** |
| | Michel J. Petrides School, Staten Island | 2018 |
| | Mount Sinai School of Medicine, New York | 1997 |
| | **Nassau Community College, Garden City** | **2013** |
| | **New York Botanical Garden, Bronx** | **2013** |
| | New York City Department of Education | 2007, 2012 |
| | New York City Technical College (City Tech) | 2018 |
| | New York Institute of Technology, New York | 2006 |
| | **New York Institute of Technology, New York** | **2006** |
| | *Orchard Park Junior High School* | *1991* |
| | *Plainview-Old Bethpage Middle School* | *1991* |
| | **Regeneron Pharmaceuticals, Inc.** | 2019 |
| | **School of Visual Arts, New York** | **2017** |
| | State University of New York, Purchase | 1989 |
| | State University of New York, Stony Brook | 1987–1990, 2015–2018 |
| | **State University of New York, Stony Broo**k | **2014, 2016** |
| | Stuyvesant High School, New York | 1998–1999 |
| | The Rockefeller University, New York | 2003, 2015–2016 |
| | **The Rockefeller University, New York** | **2010** |
| | *Titusville Middle School, Poughkeepsie* | *1991, 1993* |
| | Trudeau Institute, Saranac Lake | 2001 |
| | Union College, Schenectady | 2004 |
| | **United States Military Academy, West Point** | **1996** |
| | Wheatley School, Old Westbury | 1985 |
| NORTH CAROLINA | CIIT Center for Health Research, Triangle Park | 2003 |
| | **North Carolina Agricultural & Technical State University, Greensboro** | **2006–2007, 2009–2011** |
| | North Carolina School of Science, Durham | 1987 |
| | **North Carolina State University, Raleigh** | **2012, 2018** |
| NORTH DAKOTA | **North Dakota State University, Fargo** | **2012** |
| OHIO | Case Western Reserve University, Cleveland | 1990 |
| | Cleveland Clinic | 1987 |
| | Langston University, Langston | 2008 |
| | North Westerville High School | 1990 |
| | **The Ohio State University, Wooster** | **2016** |
| OKLAHOMA | **Oklahoma City Community College** | **2000** |
| | Oklahoma City Community College | 2006–2007, 2010 |
| | Oklahoma Medical Research Foundation, Oklahoma City | 2001 |
| | Oklahoma School of Science and Math, Oklahoma City | 1994 |
| | Tulsa Community College, Tulsa | 2009 |
| | **Tulsa Community College, Tulsa** | **2012–2014** |
| OREGON | Kaiser Permanente-Center for Health Research, Portland | 2003 |
| | **Linfield College, McMinnville** | **2014** |
| PENNSYLVANIA | Duquesne University, Pittsburgh | 1988 |
| | Germantown Academy | 1988 |
| | Kimmel Cancer Center, Philadelphia | 2008 |
| RHODE ISLAND | Botanical Society of America, Providence | 2010 |
| SOUTH CAROLINA | **Clemson University** | **2004, 2015** |
| | Medical University of South Carolina, Charleston | 1988 |
| | **University of South Carolina, Columbia** | **1988** |
| SOUTH DAKOTA | **South Dakota State University, Brookings** | **2015** |
| TENNESSEE | NABT Professional Development Conference, Memphis | 2008 |

| TEXAS | Austin Community College – Rio Grande Campus | 2000 |
|---|---|---|
| | **Austin Community College – Eastview Campus – Roundrock Campus** | **2007–2009, 2013** |
| | **Austin Community College – Roundrock Campus** | **2012** |
| | **Austin Community College - Austin** | **2018** |
| | **Houston Community College Northwest** | **2009–2010** |
| | J.J. Pearce High School, Richardson | 1990 |
| | K–12 Summer Institute, Kerrville | 2019 |
| | Langham Creek High School, Houston | 1991 |
| | **University of Lone Star College, Kingwood** | **2011** |
| | Midland College | 2008 |
| | Southwest Foundation for Biomedical Research, San Antonio | 2002 |
| | Taft High School, San Antonio | 1991 |
| | **Texas A&M University, College Station, TX** | **2013** |
| | **Texas A&M University, Prairie View, TX** | **2013** |
| | Texas A&M, AG Research and Extension Center, Weslaco | 2007 |
| | **Trinity University, San Antonio** | **1994** |
| | **University of Texas, Austin** | **1999, 2004, 2010, 2012** |
| | **University of Texas, Brownsville** | **2010** |
| UTAH | **Brigham Young University, Provo** | **2012** |
| | University of Utah, Salt Lake City | 1993 |
| | **University of Utah, Salt Lake City** | **1998, 2000** |
| | **Utah Valley State College, Orem** | **2007** |
| VERMONT | University of Vermont, Burlington | 1989 |
| | Champlain Valley Union High School | 1989 |
| VIRGINIA | Eastern Mennonite University, Harrisonburg | 1996 |
| | **James Madison University, Harrisonburg** | **2017** |
| | Jefferson School of Science, Alexandria | 1987 |
| | Mathematics and Science Center, Richmond | 1990 |
| | Mills Godwin Specialty Center, Richmond | 1998 |
| | **Virginia Polytechnic Institute and State University, Blacksburg** | **2005, 2008–2009** |
| WASHINGTON | Fred Hutchinson Cancer Research Center, Seattle | 1999, 2001, 2008 |
| | Shoreline Community College | 2011, 2012 |
| | **University of Washington, Seattle** | **1993, 1998, 2010** |
| WEST VIRGINIA | **Bethany College** | **1989** |
| WISCONSIN | Blood Center of Southeastern Wisconsin, Milwaukee | 2003 |
| | **James Madison University, Madison** | **2019** |
| | **Madison Area Technical College/Madison Area College** | **1999, 2009, 2011–2014** |
| | Marquette University, Milwaukee | 1986–1987 |
| | University of Wisconsin, Madison | 1988–1989 |
| | **University of Wisconsin, Madison** | **2004, 2012** |
| WYOMING | University of Wyoming, Laramie | 1991 |
| PUERTO RICO | **Universidad del Turabo, Gurabo, Puerto Rico** | **2011, 2012, 2014** |
| | University of Puerto Rico, Mayaguez | 1992 |
| | **University of Puerto Rico, Mayaguez** | **1992** |
| | **University of Puerto Rico, Rio Piedras** | **1993** |
| | University of Puerto Rico, Rio Piedras | 1994 |
| | University of Puerto Rico, San Juan | **2019** |

---

| AUSTRALIA | Walter and Eliza Hall Institute and University of Melbourne | 1996 |
|---|---|---|
| | **EMBL/Australian Bioinformatics Resource, University of Melbourne** | **2016** |
| | **University of Western Australia, Perth** | **2018** |
| AUSTRIA | **Vienna Open Lab, Vienna** | **2007, 2012** |
| | **Technical University of Graz** | **2019** |
| CANADA | Red River Community College, Winnipeg, Manitoba | 1989 |
| | **University of Quebec, Montreal** | **2018** |
| CHINA | Beijing No. 166 High School, Beijing | 2013–2019 |
| | **Ho Yu College, Hong Kong** | **2009** |

| | | |
|---|---|---|
| DENMARK | **Faroe Genome Project, Torshavn, Faroe Islands** | **2013** |
| GERMANY | Urania Science Center, Berlin | 2008 |
| IRELAND | **European Conference on Computational Biology/Intelligent** | **2015** |
| | **System for Molecular Biology Conference, Dublin** | |
| | **University College Dublin** | **2018** |
| ITALY | International Institute of Genetics and Biophysics, Naples | 1996 |
| | Porto Conte Research and Training Laboratories, Alghero | 1993 |
| MEXICO | ADN Mexico, Morelia | 2016 |
| | ASPB Plant Biology, Mérida | 2008 |
| | **Langebio/Cinvestav, Irapuato** | **2016** |
| NIGERIA | Godfrye Okoye University, Enugu, Nigeria | 2013, 2018 |
| PANAMA | **University of Panama, Panama City** | **1994** |
| PHILIPPINES | Eastern Visayas Campus, Philippine Science High School, Palo, Leyte | 2017 |
| RUSSIA | Shemyakin Institute of Bioorganic Chemistry, Moscow | 1991 |
| SINGAPORE | National Institute of Education | 2001–2005 |
| | Singapore Science Center | 2013 |
| SOUTH AFRICA | **North-West University, Potchefstroom** | **2016** |
| | **South African Bioinformatics Society, Durban** | **2016** |
| SWEDEN | Kristineberg Marine Research Station, Fiskebäckskil | 1995 |
| | Uppsala University | 2004 |
| THE | **International Chromosome Conference, Amsterdam** | **2007** |
| NETHERLANDS | **Wageningen University and Research Center, Wageningen** | **2014** |
| UNITED KINGDOM | **Earlham Institute, Norwich** | **2018** |
| | **The Genome Analysis Center, Norwich** | **2015** |
| | **University of York, York** | **2017** |
| | **Wellcome Trust Conference Center, Hinxton** | **2012–2013** |
| | **University of Warwick, Coventry** | **2013** |

# PRESS PUBLICATIONS

## Serials

*Genes & Development*, Vol. 33 (www.genesdev.org)

*Genome Research*, Vol. 29 (www.genome.org)

*Learning & Memory*, Vol. 26 (www.learnmem.org)

*RNA*, Vol. 25 (www.rnajournal.org)

*Cold Spring Harbor Symposia in Quantitative Biology*, Vol. 83: *Brains and Behavior: Order and Disorder in the Nervous System*, edited by David Stewart and Bruce Stillman

*Cold Spring Harbor Protocols* (www.cshprotocols.org)

*Cold Spring Harbor Perspectives in Biology* (www.cshperspectives.org)

*Cold Spring Harbor Perspectives in Medicine* (www.perspectivesinmedicine.org)

*Cold Spring Harbor Molecular Case Studies* (www.molecularcasestudies.org)

## Monographs (Topic Collections from *Perspectives in Biology* and *Perspectives in Medicine*)

*Next-Generation Sequencing in Medicine*, edited by W. Richard McCombie, Elaine R. Mardis, James A. Knowles, and John D. McPherson

*Function and Dysfunction of the Cochlea: From Mechanisms to Potential Therapies*, edited by Guy P. Richardson and Christine Petit

*Protein Homeostasis*, Second Edition, edited by Richard I. Morimoto, F. Ulrich Hartl, and Jeffery W. Kelly

*Bioelectronic Medicine*, edited by Valentin A. Pavlov and Kevin J. Tracey

*Engineering Plants for Agriculture*, edited by Pamela C. Ronald

*Calcium Signaling*, Second Edition, edited by Geert Bultynck, Martin D. Bootman, Michael J. Berridge, and Grace E. Stutzmann

*Genetic Counseling: Clinical Practice and Ethical Considerations*, edited by Laura Hercher, Barbara Biesecker, and Jehannine C. Austin

*Metastasis: Mechanism to Therapy*, edited by Jeffrey W. Pollard and Yibin Kang

## Other

*Restriction Enzymes: A History*, by Wil A.M. Loenen

*Conscience and Courage: How Visionary CEO Henri Termeer Built a Biotech Giant and Pioneered the Rare Disease Industry*, by John Hawkins

*The Digital Cell: Cell Biology as a Data Science*, by Stephen J. Royle

CSHL Annual Report 2017, Yearbook Edition

## E-books

*Restriction Enzymes: A History*, by Wil A.M. Loenen

*Function and Dysfunction of the Cochlea: From Mechanisms to Potential Therapies*, edited by Guy P. Richardson and Christine Petit

*Career Opportunities in Biotechnology and Drug Development*, by Toby Freeman

*Protein Homeostasis*, Second Edition, edited by Richard I. Morimoto, F. Ulrich Hartl, and Jeffery W. Kelly

*Bioelectronic Medicine*, edited by Valentin A. Pavlov and Kevin J. Tracey

*Conscience and Courage: How Visionary CEO Henri Termeer Built a Biotech Giant and Pioneered the Rare Disease Industry*, by John Hawkins

*Engineering Plants for Agriculture*, edited by Pamela C. Ronald

*Calcium Signaling*, Second Edition, edited by Geert Bultynck, Martin D. Bootman, Michael J. Berridge, and Grace E. Stutzmann

*Cold Spring Harbor Symposia in Quantitative Biology*, Vol. 83: *Brains and Behavior: Order and Disorder in the Nervous System*, edited by David Stewart and Bruce Stillman

*The Digital Cell: Cell Biology as a Data Science*, by Stephen J. Royle

*Genetic Counseling: Clinical Practice and Ethical Considerations*, edited by Laura Hercher, Barbara Biesecker, and Jehannine C. Austin

*Metastasis: Mechanism to Therapy*, edited by Jeffrey W. Pollard and Yibin Kang

*The PTEN Family*, edited by Charis Eng, Joanne Ngeow, and Vuk Stambolic

## Websites

Cold Spring Harbor Monographs Archive Online (www.cshmonographs.org)

Cold Spring Harbor Symposium on Quantitative Biology Archive (symposium.cshlp.org)

## Services

BioSupplyNet, scientific supply directory (www.biosupplynet.com)

# COLD SPRING HARBOR LABORATORY PRESS EXECUTIVE DIRECTOR'S REPORT

The mission of Cold Spring Harbor Laboratory Press is to provide scientists worldwide with authoritative, affordable, and pertinent information to further their research and aid in their career development. The Laboratory's commitment to scientific communication began with the publication of the proceedings of the first Annual Symposium in 1933 and has proceeded through decades of innovation that produced a monograph series that first defined the contours of molecular biology, lab manuals that empowered scientists everywhere with the emerging techniques of the new biology, successful community-driven journals, and most recently independent, not-for-profit preprint servers through which investigators can share their most recent research with millions of readers at no cost to them or their audience.

The Press publishes nine journals and more than 250 books in print and electronic form. *Genome Research* and *Genes & Development* are broadly based, long-established, and much valued, whereas *RNA* and *Learning & Memory* serve more specialized research communities with equal care. The review journals *CSH Perspectives in Biology*, *CSH Perspectives in Medicine*, and *CSH Protocols* continue to gain readership with valued types of content previously confined to print books, now made available in digital, readily discoverable, and reusable serial form.



The *Genome Research* editorial team and friends of the journal at the Biology of Genomes Meeting, May 9, 2019 at CSHL
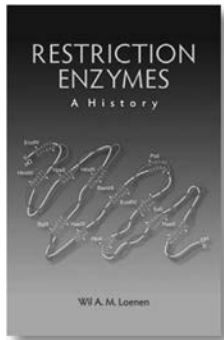
The transition of the Laboratory's journal publishing program to an open-access model was signaled with the 2016 launch of the precision medicine journal *Cold Spring Harbor Molecular Case Studies.* It enables the open sharing of clinical insights that genomic and molecular analysis bring to the understanding and potential treatment of disease, and in its fourth year it continued to gain submissions and readership. An additional open access journal, *Life Science Alliance* (LSA), was launched a year ago, published jointly by CSHL, the European Molecular Biology Organization, and Rockefeller University, and in 2019, monthly usage of its content grew by more than 300%.

Overall, online usage of Press journals remained extensive in 2019, with more than 19 million full-text article downloads worldwide, including more than 5.4 million via the U.S. National Institutes of Health's National Library of Medicine.

In the book-publishing program, 12 new print titles and 13 new e-books were added in 2019.

*Restriction Enzymes*: *A History*, by Wil A.M. Loenen, is the first full-length history of the discovery and commercial development of DNA-cutting enzymes with many uses in molecular biology, genetics, and biotechnology. Funding for the book was made possible by the Genentech Center for the History of Molecular Biology and Biotechnology at Cold Spring Harbor Laboratory. A companion website to the book (www.restrictionenzymes.org) makes an electronic version of the text and ancillary material freely accessible to readers. The biography *Conscience and Courage: How Visionary CEO Henri Termeer Built a Biotech Giant and Pioneered the Rare Disease Industry* was released in early October and rapidly became a best seller. *The Digital Cell: Cell Biology as a Data Science*, by Stephen J. Royle, has been well received with strong sales, complimentary reviews, possible use as a course text, and a Japanese translation agreement.
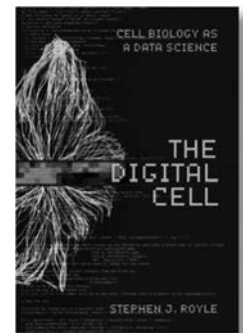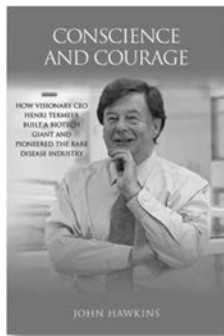
A Portuguese translation of the classic backlist title *The Eighth Day of Creation, Makers of the Revolution in Biology*, Commemorative Edition, by Horace Judson, is in progress and should be completed by the end of 2021. An appreciation of the book's importance on its 40th anniversary was published in December (https://www.molbiolcell.org/doi/full/10.1091/mbc.E19-11-0619).

The strong performance of the book program in 2019 was again supported by our highly effective direct-to-customer marketing and e-commerce. Online sales via the Press website accounted for 19% of all book sales, an exceptional performance in today's retail environment that reflects the reputation of the Press among scientists worldwide. Providing such customers the added value of electronic editions of our titles has also helped drive this success. In 2019, more than 44% of Press website sales included an e-book as a companion to a print edition or as a stand-alone publication.
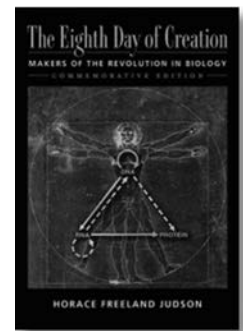
## Staff

During the year, we welcomed Michael Siragusa as a newly appointed Assistant Technology Developer and Sonali Bhattacharjee as the Assistant Editor for *Genes & Development*. Jan Argentine relocated to Texas, where she continued working remotely with the Press as the bioRxiv Screening Lead.

The Press creates publications and services that help scientists succeed while making a positive financial contribution of unrestricted funds to the Laboratory and supporting the institution's worldwide reputation for research and scientific education and communication. As a publishing organization, the Press works closely and successfully with the research community worldwide, an accomplishment made possible by the care, competence, and dedication the entire staff brings to the Press's purpose. I am grateful to them all and in particular to those who provide outstanding leadership in our diverse activities: Assistant Director Richard Sever, journal editors Terri Grodzicker and Hillary Sussman, and departmental directors Wayne Manos, Stephen Nussbaum, Marcie Siconolfi, and Linda Sussman. And as always, I thank Mala Mazzullo for her warm, generous, and efficient presence at the heart of the Office of the Director.

**John R. Inglis, Ph.D.**
*Executive Director and Publisher*

PREPRINT SERVERS

# bioRxiv, THE PREPRINT SERVER FOR BIOLOGY

A preprint is a research manuscript made freely available to the community by its authors before its acceptance for publication by a journal. The Laboratory's preprint server bioRxiv turned 6 years old in November 2019. By the end of 2019 there were 70,000 unique manuscripts on the server. The submission rate continued to increase and by December exceeded 2,500 new manuscripts each month. The platform contained work from more than 369,000 authors in 111 countries. The most prolific institutions were Stanford, Oxford, Cambridge, and Harvard. The largest subject categories were neuroscience (18%), bioinformatics (10%), and microbiology (8%).

Usage of the site also continued to climb, with page views and downloads exceeding 6 million per month by year's end.

To cope with the increasing volume, the screening team continued to expand and additional scientific affiliates were recruited, bringing the total to 126.

Thirty-six journals gave authors the opportunity to post a submitted manuscript simultaneously on bioRxiv. And 169 journals offered authors the opportunity to submit their preprints for editorial consideration directly from bioRxiv. Seventy percent of manuscripts posted to bioRxiv are published in a journal within two years: By December 2019, 27,000 of them had been published in more than 2,200 journals.

A pilot project, Transparent Review in Preprints (TRiP), was announced in November (https://www.cshl.edu/transparent-review-in-preprints). If authors consent, four journals and two independent peer-review services will post the reviews they have commissioned on manuscripts posted to bioRxiv. A growing number of journals have adopted "transparent" peer review, in which the reviews appear alongside a paper published in the journal. TRiP's benefit is that a reader will see the reviews attached to the version of the manuscript that was submitted to the journal, not the final, postreview version published by the journal—making more transparent the manuscript's journey from the authors' hands through the process of peer review.
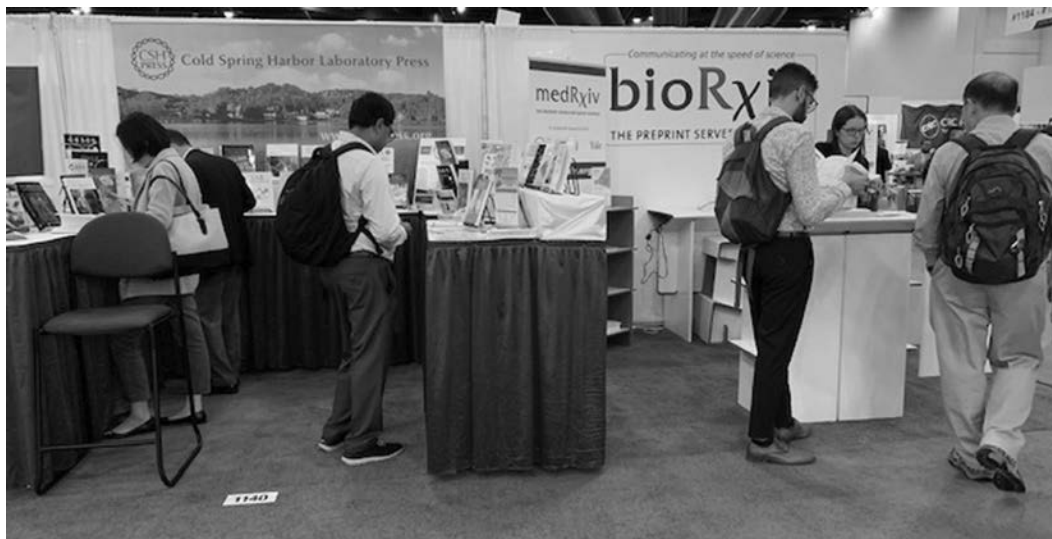


Promotion of bioRxiv, along with medRxiv and CSHL Press, at the American Society of Human Genetics Meeting; October 2019, Houston, TX.

John Inglis and Richard Sever, the co-founders of bioRxiv, and Samantha Hindle, the Content Lead, continue to give invited national and international conference talks, press interviews, panels, and podcasts discussing the rapid adoption of preprints in biology and other disciplines. Booths at major scientific meetings permit discussion with current and prospective authors, and encourage greater awareness of the benefits to science and scientists of early sharing of new research.

bioRxiv is a service of the Laboratory, not a publication of the Press or a product. The work required to maintain a 24/7/365 service is intense and would not be possible without the extraordinary dedication of the bioRxiv team: staff members Richard Sever, Samantha Hindle, Ted Roeder, K.J. Black, Linda Sussman, Jan Argentine, and Inez Sialiano, assisted by an able team of freelance screeners. Their skills and diligence are enabling bioRxiv to transform the way biologists communicate their science.

**John R. Inglis, Ph.D.**
*Co-founder, bioRxiv and medRxiv*

# medRxiv, THE PREPRINT SERVER FOR HEALTH SCIENCES

Two years into the life of bioRxiv, articles by prominent physicians in *The New York Times* and *The Lancet* argued for the benefits of a preprint server for medicine and health sciences. In response, as a pilot, two medically relevant subject areas—epidemiology and clinical trials—were added to bioRxiv. The lessons learned from the pilot fueled further exploration of the concept through discussion between bioRxiv and medical specialists, publishers, and journal editors. These conversations revealed significant anxieties about the consequences of distributing non-peer-reviewed, medically relevant content. Editors were worried that wrong information or inappropriate conclusions uncorrected by peer review—particularly when amplified by the mass media—would confuse the medical profession and the public and might pose health risks. There were also concerns about commercially motivated manipulation and the undermining of pillars of medical communication such as peer-reviewed journals, conferences, and community resources like ClinicalTrials.gov. And in addition, authors were worried about obstacles to having preprinted manuscripts published in journals of their choice.

Nevertheless, there was enthusiasm in many circles for a stand-alone health science preprint server, particularly from the Yale Open Access Data Project and BMJ, the global health information provider that had briefly hosted a preprint server in the 1990s. And policies and procedures were identified that if implemented would mitigate the perceived risks.

As a consequence, medRxiv was launched on June 25, 2019, as a partnership between Cold Spring Harbor Laboratory, Yale University, and BMJ with a management team consisting of Harlan Krumholz and Joe Ross from Yale, Theo Bloom and Claire Rawlinson from BMJ, and John Inglis and Richard Sever from Cold Spring Harbor.

medRxiv offers the potential benefits of all preprint servers—the acceleration of research through free, early sharing of results and community feedback to authors on manuscript improvement. It also enables distribution of less-publishable outputs of clinical research, like protocols, quality innovations, and inconclusive or null trial data.

To increase readers' confidence in the content of medRxiv, authors who submit papers are required to make declarations about competing interests, ethics approval, participant consent, clinical trial registration, and data availability. All manuscripts are subject to a two-stage screening process that includes a first, in-house check that the submission requirements are met, the article type is acceptable (no case reports or reviews), and there is no inappropriate content such as images of human subjects or identifiers of small human populations. In the second stage, principal investigators or their clinical equivalents assess each paper (but do not peer-review it) to provide assurance that the

Promotion of medRxiv, along with bioRxiv, at the Society for Neuroscience Meeting; October 2019, Chicago, IL.

content is research and appropriate for a professional community. They also ask if the paper contains claims or recommendations that if wrong could have negative consequences for health-related behavior. Particular concerns focus on vaccine safety, carcinogenicity of common substances, and nonstandard use of medication. If there is disquiet of this kind, the management team may conclude that a paper is better disseminated after peer review than before.

The medRxiv site and each manuscript on it contain prominent warnings that the work described has not been peer-reviewed and should not be used to guide clinical practice or health-related behavior. Journalists are urged to be responsible in reporting on preprints, emphasizing the absence of peer review.

To address authors' concerns about the fate of their preprinted manuscripts, the medRxiv founders are engaged in conversations with editors and publishers to ensure they have a full understanding of how the project is run. Several journals have already changed their policies and will now consider submitted preprints. Others have gone further, enabling authors of medRxiv preprints to submit those papers directly for review. Twenty-two journals are currently or will soon be available from 12 publishers.

By year's end, approximately six months after launch, medRxiv had received 1,230 submissions and 937 manuscripts had been posted, with a rising monthly rate of submission and usage. The average time from submission to acceptance was 10 days due to an initial abundance of care in screening.

The launch of medRxiv was much anticipated and the six co-founders/management team members received multiple invitations to speak about it at conferences and other events.

It is appropriate that the Laboratory pioneered this initiative, which builds on the transformational effects bioRxiv is having on scientific communication and adds further to the Laboratory's reputation for innovation in this area.

**John R. Inglis, Ph.D.**
*Co-founder, bioRxiv and medRxiv*

FINANCE

# FINANCIAL STATEMENTS

## CONSOLIDATED BALANCE SHEET
### December 31, 2019
(with comparative financial information as of December 31, 2018)

|  | 2019 | 2018 |
|---|---|---|
| **Assets:** | | |
| Cash and cash equivalents | $ 92,612,941 | 121,078,957 |
| Grants receivable | 8,600,047 | 7,730,792 |
| Contributions receivable, net | 136,070,457 | 70,177,263 |
| Investments | 674,813,273 | 535,006,949 |
| Investment in employee residences | 6,966,747 | 6,811,348 |
| Restricted use assets | 2,435,247 | 1,975,292 |
| Other assets | 5,860,267 | 6,186,059 |
| Land, buildings and equipment, net | 255,045,989 | 249,420,636 |
| Total assets | $ 1,182,404,968 | 998,387,296 |
| **Liabilities and net assets:** | | |
| Liabilities: | | |
| Accounts payable and accrued expenses | $ 12,607,178 | 12,948,173 |
| Deferred revenue | 68,131,842 | 81,386,316 |
| Interest rate swap | 33,067,453 | 25,281,037 |
| Bonds payable | 95,873,965 | 95,807,696 |
| Total liabilities | 209,680,438 | 215,423,222 |
| Commitments and contingencies | | |
| Net assets: | | |
| Without donor restrictions | 529,961,474 | 426,827,607 |
| With donor restrictions | 442,763,056 | 356,136,467 |
| Total net assets | 972,724,530 | 782,964,074 |
| Total liabilities and net assets | $ 1,182,404,968 | 998,387,296 |

# CONSOLIDATED STATEMENT OF ACTIVITIES
Year ended December 31, 2019
(with summarized financial information for the year ended December 31, 2018)

| | Without Donor Restrictions | With Donor Restrictions | 2019 Total | 2018 Total |
|---|---|---|---|---|
| **Revenue and other support:** | | | | |
| Public support—contributions and nonfederal grant awards | $ 41,424,618 | 110,540,984 | 151,965,602 | 62,823,243 |
| Federal grant awards | 52,296,078 | – | 52,296,078 | 44,976,944 |
| Indirect cost allowances | 36,391,296 | – | 36,391,296 | 36,554,683 |
| Investment return utilized | 33,925,072 | – | 33,925,072 | 22,641,750 |
| Royalty and license revenue | 14,525,176 | – | 14,525,176 | 17,005,951 |
| Program fees | 8,980,991 | – | 8,980,991 | 7,990,826 |
| Publications sales | 9,803,707 | – | 9,803,707 | 9,564,563 |
| Dining services | 5,154,640 | – | 5,154,640 | 4,898,093 |
| Rooms and apartments | 4,060,359 | – | 4,060,359 | 3,739,034 |
| Miscellaneous | 951,499 | – | 951,499 | 751,769 |
| Net assets released from restrictions | 51,515,978 | (51,515,978) | – | – |
| Total revenue and other support | 259,029,414 | (59,025,006) | 318,054,420 | 210,946,856 |
| **Expenses:** | | | | |
| Research | 117,365,794 | – | 117,365,794 | 106,880,122 |
| Educational programs | 19,782,548 | – | 19,782,548 | 18,950,499 |
| Publications | 9,054,869 | – | 9,054,869 | 8,768,554 |
| Banbury Center conferences | 2,378,243 | – | 2,378,243 | 2,301,320 |
| DNA Learning Center programs | 4,046,329 | – | 4,046,329 | 3,860,017 |
| Watson School of Biological Sciences programs | 3,372,646 | – | 3,372,646 | 3,199,759 |
| General and administrative | 27,747,660 | – | 27,747,660 | 30,841,520 |
| Total expenses | 183,748,089 | – | 183,748,089 | 174,801,791 |
| Excess of revenue and other support over expenses | 75,281,325 | 59,025,006 | 134,306,331 | 36,145,065 |
| **Other changes in net assets:** | | | | |
| Investment return (loss) excluding (including) amount utilized | 35,638,959 | 27,601,583 | 63,240,542 | (47,237,042) |
| Change in fair value of interest rate swap | (7,786,417) | – | (7,786,417) | 6,064,458 |
| Increase (decrease) in net assets | 103,133,867 | 86,626,589 | 189,760,456 | (5,027,519) |
| Net assets at beginning of year | 426,827,607 | 356,136,467 | 782,964,074 | 787,991,593 |
| Net assets at end of year | $ 529,961,474 | 442,763,056 | 972,724,530 | 782,964,074 |

# CONSOLIDATED STATEMENT OF CASH FLOWS
## Year ended December 31, 2019
### (with comparative financial information for the year ended December 31, 2018)

|  | 2019 | 2018 |
|---|---|---|
| Cash flows from operating activities: |  |  |
| Increase (decrease) in net assets | $ 189,760,456 | (5,027,519) |
| Adjustments to reconcile increase (decrease) in net assets to net cash provided by operating activities: |  |  |
| Change in fair value of interest rate swap | 7,786,417 | (6,064,458) |
| Depreciation and amortization | 12,983,279 | 13,140,747 |
| Donated equipment | (39,000) | – |
| Amortization of deferred bond costs | 66,269 | 66,269 |
| Net (appreciation) depreciation in fair value of investments | (89,294,070) | 29,986,414 |
| Contributions restricted for long-term investment | (15,008,418) | (18,810,536) |
| Changes in assets and liabilities: |  |  |
| Grants receivable | (869,255) | 1,217,225 |
| Contributions receivable, net | (89,218,475) | 20,118,950 |
| Restricted use assets | (459,955) | 2,095,278 |
| Other assets | 325,792 | 5,004,346 |
| Accounts payable and accrued expenses | (340,996) | (2,940,671) |
| Deferred revenue | (13,254,474) | 75,784,363 |
| Net cash provided by operating activities | 2,437,570 | 114,570,408 |
| Cash flows from investment activities: |  |  |
| Capital expenditures | (18,569,632) | (27,315,485) |
| Proceeds from sales and maturities of investments | 141,495,081 | 108,718,791 |
| Purchases of investments | (192,007,335) | (141,885,620) |
| Net change in investment in employee residences | (155,399) | (462,742) |
| Net cash used in investment activities | (69,237,285) | (60,945,056) |
| Cash flows from financing activities |  |  |
| Contributions restricted for long-term investment | 249,663 | 900,022 |
| Contributions restricted for investment in capital | 14,758,755 | 17,910,514 |
| Decrease (increase) in contributions receivable | 23,325,281 | (15,535,947) |
| Net cash provided by financing activities | 38,333,699 | 3,274,589 |
| Net (decrease) increase in cash and cash equivalents | (28,466,016) | 56,899,941 |
| Cash and cash equivalents at beginning of year | 121,078,957 | 64,179,016 |
| Cash and cash equivalents at end of year | $ 92,612,941 | 121,078,957 |
| Supplemental disclosure: |  |  |
| Interest paid | $ 3,680,812 | 3,762,449 |
| Purchases of capital expenditures in accounts payable | $ 772,157 | 6,166,487 |

# FINANCIAL SUPPORT OF THE LABORATORY

Cold Spring Harbor Laboratory, Banbury Center, and the Dolan DNA Learning Center receive a substantial portion of funding through grants from the federal government and through grants, capital gifts, and annual contributions from New York state, private foundations, corporations, and individuals. The following section summarizes funding that occurred during 2019.

## GRANTS January 1–December 31, 2019

### COLD SPRING HARBOR LABORATORY

| Grantor | Program/Principal Investigator | Duration of Grant | | 2019 Funding[1] | |
|---|---|---|---|---|---|
| **FEDERAL GRANTS** | | | | | |
| NATIONAL INSTITUTES OF HEALTH | | | | | |
| *Program Project and Center Support* | Drs. Stillman/Egeblad/Krainer/McCombie/ Pappin/D. Spector/Vakoc | 02/20/18 | 01/31/23 | 4,344,544 | |
| | Dr. Tuveson, Cancer Center Core | 08/01/16 | 07/31/21 | 4,403,952 | |
| *Cooperative Research Agreement Support*[2] | Drs. Huang/Gillis/Mitra/Osten/Zador | 09/21/17 | 05/31/22 | 4,502,687 | |
| | Drs. Osten/Albeanu/Mitra | 09/20/17 | 05/31/22 | 1,605,281 | |
| | Dr. Tuveson | 03/06/18 | 02/28/23 | 420,863 | |
| | Dr. Vakoc | 08/01/19 | 07/31/24 | 439,200 | * |
| *Contract Support* | | | | | |
| Leidos Biomedical Research, Inc. - NCI | Drs. Tuveson/D. Spector | 01/09/17 | 07/31/20 | 3,535,800 | |
| *Research Support* | Dr. Albeanu | 02/09/16 | 01/31/21 | 408,000 | |
| | Drs. Albeanu/Koulakov | 09/11/18 | 08/31/23 | 1,083,575 | |
| | Drs. Albeanu/Koulakov | 06/15/19 | 03/31/24 | 894,985 | * |
| | Dr. Churchland | 05/01/18 | 04/30/22 | 480,000 | |
| | Dr. Crow | 04/01/19 | 03/31/21 | 118,421 | * |
| | Drs. Dobin/Gingeras | 08/18/17 | 05/31/22 | 480,000 | |
| | Drs. Engel/Churchland | 09/20/18 | 06/30/21 | 441,600 | |
| | Dr. Furukawa | 03/01/15 | 02/28/21 | 475,206 | |
| | Dr. Furukawa | 04/15/19 | 03/31/24 | 564,879 | * |
| | Dr. Furukawa | 08/01/19 | 03/31/24 | 2,716,377 | * |
| | Dr. Gillis | 09/15/17 | 08/31/21 | 480,000 | |
| | Drs. Gillis/Huang/Lee | 07/13/17 | 05/31/22 | 480,000 | |
| | Dr. Goodwin | 09/11/19 | 08/31/24 | 122,709 | * |
| | Dr. C. Hammell | 03/01/16 | 12/31/20 | 403,200 | |
| | Dr. Huang | 07/01/11 | 07/31/23 | 698,459 | |
| | Drs. Huang/Gillis | 08/01/16 | 07/31/21 | 788,983 | |
| | Dr. Joshua-Tor | 06/10/16 | 03/31/20 | 345,600 | |
| | Dr. Kepecs | 09/15/17 | 12/31/19 | 420,000 | |
| | Dr. Kepecs | 08/15/15 | 05/31/20 | 420,000 | |
| | Dr. Kepecs | 09/15/15 | 07/31/20 | 432,772 | |
| | Dr. Kepecs | 08/02/19 | 12/31/19 | 2,232,506 | * |
| | Dr. Kinney | 09/01/19 | 08/31/24 | 480,000 | * |

[1]Awarded, including direct and indirect costs
[2]Funding amounts include only CSHL's portion of the award
*New or competing renewals or supplements awarded in 2019

| Grantor | Program/Principal Investigator | Duration of Grant | | 2019 Funding[1] | |
|---|---|---|---|---|---|
| | Drs. Koulakov/Li | 09/30/19 | 07/31/24 | 432,000 | * |
| | Dr. Krainer | 07/01/17 | 06/30/22 | 806,400 | |
| | Dr. Lee | 08/19/16 | 05/31/21 | 480,000 | |
| | Drs. Li/Huang | 09/28/15 | 06/30/20 | 660,707 | |
| | Dr. Martienssen | 06/01/17 | 04/30/21 | 422,400 | |
| | Dr. McCandlish | 09/01/19 | 07/31/24 | 480,000 | * |
| | Dr. Navlakha | 11/21/19 | 08/31/21 | 365,001 | * |
| | Dr. Park | 09/15/16 | 08/31/21 | 174,117 | |
| | Dr. Pedmale | 08/04/17 | 07/31/22 | 724,177 | |
| | Dr. Shea | 12/08/15 | 11/30/20 | 467,068 | |
| | Dr. Shea | 06/01/19 | 03/31/24 | 480,000 | * |
| | Dr. Sheltzer | 09/18/15 | 08/31/20 | 480,000 | |
| | Dr. Siepel | 03/01/18 | 02/28/23 | 479,215 | |
| | Dr. D. Spector | 08/20/19 | 07/31/24 | 768,000 | * |
| | Dr. Stenlund | 05/01/15 | 12/31/19 | 504,989 | |
| | Dr. Stillman | 03/01/17 | 02/28/21 | 729,600 | |
| | Drs. Tollkuhn/Gillis | 03/01/18 | 12/31/22 | 480,000 | |
| | Dr. Tonks | 05/14/15 | 04/30/20 | 753,375 | |
| | Drs. Tuveson/M. Hammell/Pappin | 12/07/16 | 11/30/21 | 554,114 | |
| | Dr. Vakoc | 12/18/18 | 11/30/23 | 496,969 | |
| | Dr. Vakoc | 12/02/19 | 11/30/21 | 250,560 | * |
| | Drs. Vakoc/Tuveson | 07/02/19 | 06/30/24 | 413,053 | * |
| | Dr. Van Aelst | 04/01/19 | 01/31/24 | 690,674 | * |
| | Dr. Zador | 09/01/17 | 08/31/20 | 907,874 | |
| *Research Subcontracts* | | | | | |
| NIH/Cornell University Consortium Agreement | Dr. Siepel | 05/01/19 | 02/28/23 | 313,407 | * |
| NIH/Envisagenics, Inc. Consortium Agreement | Dr. Krainer | 04/05/18 | 03/31/20 | 50,198 | |
| NIH/Harvard Medical School Consortium Agreement | Dr. Osten | 07/01/17 | 04/30/22 | 84,000 | |
| NIH/Harvard Medical School Consortium Agreement | Dr. Osten | 09/01/17 | 07/31/21 | 86,400 | |
| NIH/Johns Hopkins University Consortium Agreement | Dr. Joshua-Tor | 03/01/15 | 02/29/20 | 33,264 | |
| NIH/New York Genome Center Consortium Agreement | Drs. Wigler/Iossifov/Levy/Siepel | 01/14/16 | 11/30/19 | 422,972 | |
| NIH/New York University Consortium Agreement | Dr. Koulakov | 09/01/19 | 05/31/24 | 420,692 | * |
| NIH/Oregon Health & Science University Consortium Agreement | Dr. Li | 12/15/17 | 11/30/22 | 412,376 | |
| NIH/The Research Foundation for the State of New York–Stony Brook Consortium Agreement | Dr. M. Hammell | 09/15/17 | 06/30/22 | 92,820 | |
| NIH/University of California San Diego Consortium Agreement | Dr. Gillis | 09/18/19 | 08/31/22 | 178,817 | * |
| NIH/University of California San Diego Consortium Agreement | Dr. Mitra | 09/15/18 | 05/31/23 | 271,645 | |
| NIH/University of Minnesota Consortium Agreement | Dr. dos Santos | 01/01/17 | 12/31/20 | 103,183 | |
| NIH/University of Minnesota Consortium Agreement | Dr. dos Santos | 12/01/19 | 11/30/24 | 37,962 | * |

[1]Awarded, including direct and indirect costs
*New or competing renewals or supplements awarded in 2019

| Grantor | Program/Principal Investigator | Duration of Grant | | 2019 Funding[1] | |
|---|---|---|---|---|---|
| NIH/University of Nebraska Medical Center Consortium Agreement | Drs. Tuveson/Pappin | 05/01/17 | 04/30/22 | 354,100 | |
| NIH/University of Pittsburgh Consortium Agreement | Dr. Mitra | 09/11/18 | 08/31/23 | 93,068 | |
| NIH/University of Texas at Austin Consortium Agreement | Dr. Osten | 09/18/17 | 06/30/22 | 379,200 | |
| *Fellowship/Career Development Support* | Dr. Bravo-Rivera | 06/01/17 | 05/31/20 | 63,746 | |
| | Dr. Fisher | 03/01/18 | 02/28/21 | 60,610 | |
| | Dr. Jaremko | 12/01/18 | 11/30/21 | 61,610 | |
| *Institutional Training Program Support* | Dr. Mills/Cancer Postdoctoral | 08/14/17 | 08/31/21 | 208,147 | |
| | Dr. Gann/Watson School of Biological Sciences | 07/01/17 | 06/30/22 | 333,291 | |
| *Course Support* | Advanced Sequencing Technologies and Applications | 04/10/12 | 06/30/21 | 58,895 | |
| | Advanced Techniques in Molecular Neuroscience | 07/06/15 | 03/31/20 | 105,668 | |
| | Cellular Biology of Addiction | 08/01/16 | 07/31/21 | 40,799 | |
| | Computational and Comparative Genomics | 08/15/17 | 06/30/20 | 67,704 | |
| | Empowering Nextgen: Advanced Biomedical Leadership | 06/01/15 | 02/29/20 | 521,040 | |
| | Eukaryotic Gene Expression | 04/13/17 | 03/31/22 | 99,473 | |
| | Molecular Embryology of the Mouse | 04/13/17 | 03/31/22 | 136,621 | |
| | Programming for Biology | 09/01/17 | 06/30/20 | 83,641 | |
| | Protein Purification and Characterization | 04/13/17 | 03/31/22 | 80,988 | |
| | Proteomics | 08/03/18 | 04/30/23 | 123,254 | |
| | Quantitative Imaging: From Cells to Molecules | 04/01/16 | 03/31/21 | 97,417 | |
| | Vision: A Platform for Linking Circuits, Perception and Behavior | 04/01/19 | 03/31/24 | 30,000 | * |
| | X-Ray Methods in Structural Biology | 09/01/17 | 08/31/22 | 90,745 | |
| *Meeting Support* | Biology of Cancer: Microenvironment and Metastasis | 08/01/19 | 07/31/20 | 7,000 | * |
| | Biology of Genomes | 04/01/18 | 03/31/23 | 60,000 | |
| | Blood Brain Barrier | 03/15/19 | 02/29/20 | 10,000 | * |
| | Eukaryotic DNA Replication and Genome Maintenance | 09/01/19 | 08/31/20 | 10,000 | * |
| | Eukaryotic mRNA Processing | 08/05/19 | 07/31/20 | 6,000 | * |
| | Genome Informatics | 08/01/19 | 07/31/20 | 33,305 | * |
| | Mechanisms of Metabolic Signaling | 02/01/19 | 01/31/20 | 15,000 | * |
| | Microbial Pathogenesis and Host Response | 07/01/19 | 06/30/20 | 7,000 | * |
| | Microbiome | 07/01/19 | 06/30/20 | 12,500 | * |
| | Neurobiology of *Drosophila* | 07/01/19 | 06/30/20 | 20,000 | * |
| | Retroviruses | 04/17/18 | 03/31/21 | 35,000 | |
| | Systems Biology: Global Regulation of Gene Expression | 09/10/19 | 07/31/24 | 29,775 | * |
| | Systems Immunology | 03/01/19 | 02/29/20 | 14,000 | * |
| | Telomeres and Telomerase | 01/15/19 | 12/31/19 | 20,000 | * |
| | Zebrafish Neural Circuits and Behavior | 07/15/19 | 06/30/20 | 10,000 | * |

[1]Awarded, including direct and indirect costs
*New or competing renewals or supplements awarded in 2019

| Grantor | Program/Principal Investigator | Duration of Grant | | 2019 Funding[1] | |
|---|---|---|---|---|---|
| **NATIONAL SCIENCE FOUNDATION** | | | | | |
| *Multiple Project Award Support* | Drs. Jackson/Lippman | 09/01/16 | 08/31/20 | 1,136,483 | |
| | Drs. McCombie/Martienssen | 09/01/18 | 08/31/22 | 961,981 | |
| *Research Support* | Dr. Albeanu | 08/01/17 | 07/31/21 | 221,626 | |
| | Dr. Jackson | 08/01/19 | 07/31/22 | 240,000 | * |
| | Dr. Jackson | 08/01/18 | 07/31/21 | 231,094 | |
| | Drs. Jackson/Gillis | 07/01/18 | 06/30/20 | 147,872 | |
| | Dr. Lippman | 07/01/17 | 06/30/21 | 1,136,700 | |
| | Dr. Pedmale | 09/01/18 | 08/31/21 | 289,542 | |
| | Dr. Ware | 06/01/19 | 05/31/20 | 632,215 | * |
| *Research Subcontracts* | | | | | |
| NSF/Cornell University Consortium Agreement | Dr. Siepel | 09/01/18 | 08/31/22 | 172,629 | |
| NSF/Iowa State University Consortium Agreement | Dr. Jackson | 03/01/13 | 02/28/20 | 12,460 | |
| NSF/New York University Consortium Agreement | Drs. Jackson/Gillis | 11/01/19 | 10/31/23 | 645,305 | * |
| NSF/University of Arizona Consortium Agreement | Drs. Ware/Micklos | 09/01/13 | 08/31/19 | 13,433 | |
| NSF/University of Georgia Consortium Agreement | Dr. Ware | 01/15/18 | 12/31/20 | 328,576 | |
| *Fellowship Support* | L. Aguirre | 06/01/19 | 05/31/22 | 46,000 | * |
| | B. Berube | 06/01/17 | 05/31/20 | 46,000 | |
| | D. Johnson | 06/01/17 | 08/31/19 | 11,500 | |
| | K. O'Neill | 06/01/18 | 05/31/21 | 46,000 | |
| | J. Werner | 09/01/19 | 08/31/21 | 46,000 | * |
| *Course Support* | Advanced Bacterial Genetics | 06/15/17 | 05/31/22 | 90,000 | |
| | *Drosophila* Neurobiology: Genes, Circuits, and Behavior | 07/01/17 | 06/30/20 | 21,597 | |
| | Frontiers and Techniques in Plant Science | 08/01/18 | 07/31/21 | 111,155 | |
| | Synthetic Biology | 09/01/18 | 08/31/21 | 64,175 | |
| | Yeast Genetics and Genomics | 07/01/17 | 06/30/22 | 90,000 | |
| **UNITED STATES DEPARTMENT OF AGRICULTURE** | | | | | |
| *Research Support* | Dr. Jackson | 01/01/18 | 12/31/21 | 118,014 | |
| | Dr. McCombie | 09/10/19 | 09/09/20 | 1,331,926 | * |
| *Meeting Support* | Workshop on Cereal Genomics | 10/01/19 | 09/30/20 | 25,000 | * |
| **UNITED STATES DEPARTMENT OF THE ARMY** | | | | | |
| *Research Support* | Dr. Vakoc | 07/01/19 | 06/30/21 | 331,926 | * |
| *Research Subcontracts* | | | | | |
| DOA/University of Southern California Consortium Agreement | Dr. Churchland | 08/23/16 | 02/22/21 | 86,400 | |
| *Fellowship Support* | Dr. Chen | 04/01/19 | 09/30/20 | 255,160 | * |
| | Dr. Scaduto | 08/15/18 | 08/14/21 | 197,092 | |

[1]Awarded, including direct and indirect costs
*New or competing renewals or supplements awarded in 2019

| Grantor | Program/Principal Investigator | Duration of Grant | | 2019 Funding[1] | |
|---|---|---|---|---|---|
| **UNITED STATES DEPARTMENT OF ENERGY** | | | | | |
| *Research Support* | Dr. Martienssen | 09/15/17 | 09/14/21 | 1,098,044 | |
| *Research Subcontracts* | | | | | |
| DOE/Brookhaven National Laboratory Consortium Agreement | Dr. Ware | 09/01/19 | 05/31/20 | 105,000 | * |
| DOE/Lawrence Berkeley National Laboratory Consortium Agreement | Dr. Ware | 05/18/17 | 09/30/20 | 665,995 | |
| DOE/New York University Consortium Agreement | Dr. McCombie | 08/15/15 | 08/14/20 | 244,033 | |
| **UNITED STATES DEPARTMENT OF THE INTERIOR** | | | | | |
| *Research Subcontracts* | | | | | |
| NIH/ Baylor College of Medicine Consortium Agreement | Dr. Zador | 07/14/19 | 01/14/21 | 101,582 | * |
| **MISCELLANEOUS SOURCES OF FUNDING** | | | | | |
| *Equipment Support* | | | | | |
| F.M. Kirby Foundation, Inc. | Dr. Tuveson | 05/01/19 | 04/30/20 | 100,000 | * |
| The Simons Foundation/CSHL Innovative Center | Dr. Lyons | 07/01/17 | 06/30/20 | 38,004 | |
| The Wasily Family Foundation | Dr. Tuveson | 07/01/19 | 06/30/20 | 50,000 | * |
| *Program Project Support* | | | | | |
| The Simons Foundation / Autism | Dr. Wigler | 01/01/17 | 12/31/20 | 1,773,129 | |
| The Simons Foundation / Cancer | Dr. Wigler | 01/01/17 | 12/31/19 | 2,256,048 | |
| *Research Support* | | | | | |
| Rita Allen Foundation | Dr. dos Santos | 09/01/16 | 08/31/21 | 100,000 | |
| Allergan Sales, LLC | Dr. Furukawa | 01/01/19 | 12/31/21 | 278,500 | * |
| American Association for Cancer Research | Dr. dos Santos | 07/01/18 | 06/30/20 | 75,000 | |
| | Dr. Sheltzer | 07/01/19 | 06/30/20 | 25,000 | * |
| Anonymous | Dr. Tuveson | 11/01/18 | 10/31/21 | 675,000 | |
| Assured Guaranty Corp. | Dr. Plenker | 01/01/19 | 12/31/19 | 10,000 | * |
| Austin's Purpose Corporation | Dr. Furukawa | 12/29/16 | 12/28/20 | 10,000 | |
| R. and E. Bailenson in memory of Jerry Frankel | Dr. Plenker | 01/01/19 | 12/31/19 | 10,000 | * |
| The Breast Cancer Research Foundation | Dr. Wigler | 10/01/19 | 09/30/20 | 250,000 | * |
| Joe W. and Dorothy Dorsett Brown Foundation | Drs. Furukawa/Sheltzer/Van Aelst | 12/11/19 | 12/10/20 | 33,000 | |
| Cedar Hill Foundation | Dr. Fearon | 11/16/19 | 11/15/20 | 70,000 | * |
| CSHL Association Fellowship | Dr. Stillman | 01/01/19 | 12/31/20 | 280,000 | * |
| CSHL Translational Cancer Support | Dr. Atwal | 02/25/17 | 02/24/20 | 305,560 | |
| | Dr. Beyaz | 01/31/19 | 01/30/23 | 282,875 | * |
| | Dr. Beyaz | 07/08/19 | 07/07/20 | 129,999 | * |
| | Dr. Chang | 06/01/18 | 05/31/21 | 267,661 | |
| | Dr. dos Santos | 05/26/17 | 05/25/20 | 288,000 | |
| | Dr. Egeblad | 08/26/19 | 08/25/20 | 97,920 | * |
| | Dr. Evans | 11/01/19 | 10/31/20 | 298,685 | * |
| | Dr. Gillis | 03/01/18 | 02/28/22 | 232,454 | |

[1]Awarded, including direct and indirect costs
*New or competing renewals or supplements awarded in 2019

| Grantor | Program/Principal Investigator | Duration of Grant | | 2019 Funding[1] | |
|---|---|---|---|---|---|
| | Dr. C. Hammell | 03/01/18 | 02/28/21 | 382,941 | |
| | Dr. Janowitz | 08/22/18 | 08/21/23 | 237,168 | |
| | Dr. Joshua-Tor | 03/25/19 | 03/24/21 | 32,728 | * |
| | Dr. Kepecs | 07/01/17 | 06/30/20 | 141,264 | |
| | Dr. Kinney | 02/25/17 | 02/24/21 | 486,029 | |
| | Dr. Krasnitz | 03/01/18 | 02/28/22 | 200,934 | |
| | Dr. Levy | 03/01/18 | 02/28/22 | 18,646 | |
| | Dr. Li | 03/01/18 | 02/28/22 | 201,440 | |
| | Dr. Li | 03/01/18 | 02/28/22 | 192,000 | |
| | Dr. Lyons | 07/01/17 | 06/30/20 | 336,682 | |
| | Dr. Lyons | 04/01/18 | 03/31/21 | 497,086 | |
| | Dr. Mills | 01/31/19 | 01/30/23 | 415,273 | * |
| | Dr. Mills | 01/01/17 | 12/31/20 | 674,058 | |
| | Dr. Mills | 10/25/17 | 10/24/21 | 741,585 | |
| | Dr. Mitra | 07/01/17 | 06/30/20 | 393,934 | |
| | Dr. Osten | 07/21/17 | 07/20/20 | 256,022 | |
| | Dr. Pappin | 03/25/19 | 03/24/21 | 163,353 | * |
| | Dr. Pappin | 05/26/17 | 05/25/20 | 716,022 | |
| | Dr. Preall | 06/01/17 | 05/31/20 | 889,707 | |
| | Dr. Sheltzer | 03/25/19 | 03/24/20 | 494,692 | * |
| | Dr. D. Spector | 12/01/18 | 11/30/21 | 574,264 | |
| | Dr. D. Spector | 06/01/16 | 05/31/21 | 196,798 | |
| | Dr. D. Spector | 04/01/18 | 03/31/21 | 423,596 | |
| | Dr. Tonks | 03/25/19 | 03/24/21 | 981,058 | * |
| | Dr. Trotman | 01/31/19 | 01/30/22 | 433,598 | * |
| | Dr. Trotman | 08/26/19 | 08/25/20 | 373,090 | * |
| | Dr. Tuveson | 04/01/18 | 03/31/21 | 848,068 | |
| | Dr. Vakoc | 05/17/19 | 05/16/22 | 918,852 | * |
| | Dr. Van Aelst | 07/01/18 | 06/30/21 | 421,039 | |
| | Dr. Wigler | 05/17/19 | 05/16/24 | 2,478,274 | * |
| | Dr. Wigler | 03/01/18 | 02/28/22 | 549,148 | |
| | Dr. Yeh | 05/04/19 | 05/03/20 | 431,633 | * |
| | Dr. Yeh | 07/01/17 | 06/30/20 | 29,723 | |
| | Dr. Zhang | 01/31/19 | 01/30/21 | 377,868 | * |
| | Dr. Zhang | 01/13/17 | 01/12/21 | 339,251 | |
| | Dr. Zhang | 05/04/18 | 05/03/20 | 894,668 | |
| Cystic Fibrosis Foundation | Dr. Krainer | 02/01/18 | 01/31/20 | 108,000 | |
| Dr. Mikala Egeblad | Dr. Egeblad | 07/01/19 | 06/30/20 | 1,952 | * |
| The Edward P. Evans Foundation | Dr. Zhang | 07/01/18 | 06/30/20 | 125,000 | |
| George and Diane Fellows | Dr. Vakoc | 01/18/19 | 01/17/24 | 20,000 | * |
| Douglas and Christine Fox | Dr. Furukawa | 06/01/18 | 05/31/22 | 50,000 | |
| The Claire Friedlander Family Foundation | Dr. Beyaz | 03/21/19 | 03/20/20 | 48,000 | * |
| Glen Cove C.A.R.E.S. | Dr. dos Santos | 01/27/19 | 01/26/20 | 5,000 | * |
| Google, Inc. | Dr. Sheltzer | 08/01/19 | 07/31/20 | 4,500 | * |
| Gyeongsang National University/ The Republic of Korea | Dr. Jackson | 01/01/18 | 12/31/19 | 50,858 | |
| Irving Hansen Foundation | Dr. Tonks | 08/01/19 | 07/31/20 | 25,000 | * |
| Jo-Ellen and Ira Hazan | Dr. Tuveson | 12/18/18 | 12/17/20 | 10,000 | |
| Heartfelt Wings Foundation Inc. | Dr. Furukawa | 09/05/18 | 09/04/22 | 250,000 | |
| The Hope Foundation | Dr. Plenker | 01/31/19 | 01/30/20 | 22,800 | * |
| Inari Agriculture, Inc. | Dr. Jackson | 11/01/19 | 10/31/20 | 44,902 | * |
| Indian Institute of Technology Madras | Dr. Mitra | 01/01/15 | 12/31/19 | 28,377 | |

[1]Awarded, including direct and indirect costs
*New or competing renewals or supplements awarded in 2019

| Grantor | Program/Principal Investigator | Duration of Grant | | 2019 Funding[1] | |
|---|---|---|---|---|---|
| Industry-Academic Cooperation Foundation of Wonkwang University in the Republic of Korea | Dr. Lippman | 01/01/18 | 12/31/20 | 52,423 | |
| The Leukemia & Lymphoma Society | Dr. Vakoc | 07/01/15 | 06/30/20 | 110,000 | |
| Long Island Bioscience Hub | Dr. Vaughan | 11/15/17 | 11/14/19 | 20,000 | |
| Dr. Robert Lourie | Dr. Zador | 01/11/19 | 01/10/20 | 500,000 | * |
| The Lustgarten Foundation | Dr. Fearon | 07/01/14 | 06/30/20 | 1,000,000 | |
| | Dr. Tuveson | 09/01/17 | 12/31/22 | 1,000,000 | |
| | Dr. Tuveson | 01/01/19 | 12/31/20 | 820,124 | * |
| | Dr. Tuveson | 07/01/19 | 06/30/20 | 70,000 | * |
| The Mark Foundation for Cancer Research Ltd. | Dr. Egeblad | 10/15/19 | 10/14/20 | 250,000 | * |
| John and Patti Maroney | Dr. Wigler | 12/18/18 | 12/17/20 | 5,000 | |
| The G. Harold and Leila Y. Mathers Charitable Foundation | Dr. Meyer | 12/01/19 | 11/30/22 | 233,237 | * |
| | Dr. Mitra | 01/01/17 | 12/31/20 | 330,000 | |
| Breast Cancer Awareness Day in Memory of Elizabeth McFarland | Dr. Wigler | 01/01/19 | 12/31/19 | 51,475 | * |
| The Don Monti Memorial Research Foundation | Drs. Stillman/Tonks/Vakoc | 01/01/19 | 12/31/19 | 200,000 | * |
| | Drs. Stillman/Tonks/Vakoc | 12/17/19 | 12/16/20 | 300,000 | * |
| Louis Morin Charitable Trust | Drs. Tollkuhn/Martienssen | 12/12/19 | 12/11/20 | 125,000 | * |
| Northwell Health, Inc. | Drs. Tuveson/Freoling | 10/17/19 | 10/16/20 | 5,844 | * |
| Ono Pharmaceutical Co., Ltd. | Dr. Tuveson | 04/01/18 | 03/31/20 | 71,498 | |
| Katie Oppo Research Fund Inc. | Dr. Egeblad | 01/01/19 | 12/31/20 | 75,000 | * |
| The Michelle Paternoster Foundation | Dr. Vakoc | 09/16/18 | 09/15/20 | 50,000 | |
| The Pershing Square Foundation | Dr. dos Santos | 07/01/18 | 06/30/21 | 200,000 | |
| | Dr. Egeblad | 07/01/17 | 06/30/20 | 180,000 | |
| | Dr. Vakoc | 07/01/18 | 06/30/20 | 20,000 | |
| Christina Renna Foundation Inc. | Dr. Vakoc | 09/16/18 | 09/15/20 | 35,000 | |
| Charles and Marie Robertson Foundation | Dr. dos Santos | 01/01/19 | 12/31/19 | 30,000 | * |
| | Dr. Shea | 01/01/19 | 12/31/19 | 20,000 | * |
| The Mary Ruchalski Foundation, Inc. | Dr. Vakoc | 09/16/18 | 09/15/20 | 30,000 | |
| Damon Runyon Cancer Research Foundation | Dr. Sheltzer | 01/01/19 | 12/31/20 | 200,000 | * |
| Schmidt Futures | Dr. Zador | 12/16/19 | 12/15/20 | 750,000 | * |
| Eleanor Schwartz Charitable Foundation | Dr. Churchland | 07/26/18 | 07/25/20 | 100,000 | |
| Barbara & Kristopher Selden | Dr. Furukawa | 02/20/19 | 02/19/20 | 2,623 | * |
| Edith and Alan Seligson | Dr. Koike | 05/23/19 | 05/22/20 | 100,000 | * |
| Dr. Jason Sheltzer and Joan Smith | Dr. Sheltzer | 08/01/19 | 07/31/20 | 7,000 | * |
| The Simons Foundation/CSHL Innovative Center | Dr. dos Santos | 07/01/17 | 04/30/22 | 240,716 | |
| | Dr. Egeblad | 07/01/17 | 04/30/22 | 204,000 | |
| | Dr. Fearon | 07/01/17 | 04/30/22 | 334,141 | |
| | Dr. Janowitz | 07/01/17 | 04/30/22 | 227,654 | |
| | Dr. Kepecs | 07/01/17 | 04/30/22 | 184,016 | |
| | Dr. Krainer | 07/01/17 | 04/30/22 | 328,895 | |
| | Dr. Lyons | 07/01/17 | 06/30/20 | 57,570 | |
| | Dr. Osten | 07/01/17 | 04/30/22 | 188,351 | |
| | Dr. Tonks | 07/01/17 | 06/30/20 | 645,615 | |
| | Dr. Trotman | 07/01/17 | 06/30/20 | 169,655 | |

[1]Awarded, including direct and indirect costs
*New or competing renewals or supplements awarded in 2019

| Grantor | Program/Principal Investigator | Duration of Grant | | 2019 Funding[1] | |
|---|---|---|---|---|---|
| | Dr. Trotman | 07/01/17 | 04/30/22 | 193,260 | |
| | Dr. Tuveson | 07/01/17 | 06/30/20 | 125,000 | |
| | Dr. Tuveson | 07/01/17 | 04/30/22 | 185,491 | |
| | Dr. Vakoc | 07/01/17 | 06/30/20 | 125,000 | |
| | Dr. Vakoc | 07/01/17 | 04/30/22 | 117,043 | |
| | Dr. Van Aelst | 07/01/17 | 04/30/22 | 180,067 | |
| The Simons Foundation | Dr. Churchland | 07/01/17 | 06/30/22 | 141,086 | |
| | Dr. Iossifov | 12/01/19 | 11/30/21 | 257,212 | * |
| | Dr. Zador | 07/01/17 | 06/30/22 | 140,400 | |
| Skyhawk Therapeutics, Inc. | Dr. Krainer | 05/01/19 | 04/30/20 | 200,000 | * |
| Alfred P. Sloan Foundation | Dr. McCandlish | 09/15/19 | 09/14/21 | 35,000 | * |
| Lauri Strauss Leukemia Foundation | Dr. Wei | 05/01/19 | 04/30/20 | 10,000 | * |
| Swim Across America Nassau/ Suffolk | Dr. Egeblad | 12/20/19 | 12/19/20 | 50,000 | * |
| The Thompson Family Foundation | Dr. Tuveson | 03/06/17 | 03/05/21 | 1,387,716 | |
| Three Strohm Sisters Family Foundation | Dr. Egeblad | 12/13/19 | 12/12/20 | 5,000 | * |
| Friends of the TJ Foundation Inc. | Dr. Vakoc | 09/16/18 | 09/15/20 | 50,000 | |
| The Toronto-Dominion Bank | Dr. Tuveson | 11/19/19 | 12/18/22 | 400,000 | * |
| Trisomy 18 Foundation | Dr. Sheltzer | 05/01/19 | 04/30/20 | 5,000 | * |
| U.S. – Israel Binational Agricultural Research and Development Foundation | Dr. Lippman | 12/01/18 | 11/30/21 | 42,000 | |
| | Dr. Shea | 09/01/16 | 08/31/20 | 16,667 | * |
| The University of Texas M.D. Anderson Cancer Center/Chan Zuckerberg Foundation | Dr. Lee | 07/01/19 | 06/30/20 | 161,000 | * |
| Joan & Sanford I. Weill Medical College | Dr. Fearon | 07/01/14 | 12/31/19 | 81,347 | |
| Samuel H. Wolcott | Dr. Fearon | 10/01/19 | 09/30/20 | 50,000 | * |
| Women's Partnership in Science | Drs. Egeblad/Mills/Sordella/Van Aelst/Ware | 01/01/19 | 12/31/19 | 184,627 | * |
| Elisabeth R. Woods Foundation | Dr. Egeblad | 10/26/17 | 10/25/19 | 22,700 | |
| The Bradley Zankel Foundation, Inc. | Dr. Mills | 11/15/19 | 11/14/20 | 20,000 | * |
| Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation | Dr. M. Hammell | 12/01/18 | 11/30/23 | 500,000 | |
| *Fellowship Support* | | | | | |
| Agency for Science, Technology and Research | Watson School of Biological Sciences | 09/01/17 | 08/31/20 | 43,836 | |
| Rita Allen Foundation | Dr. Fearon | 01/01/19 | 12/31/19 | 6,000 | * |
| The American Association of Immunologists, Inc. | Dr. Fein | 02/01/19 | 01/31/20 | 50,376 | * |
| Autism Speaks, Inc. | D. Rupert | 10/01/18 | 09/30/20 | 40,000 | |
| Brain & Behavior Foundation | Dr. Crow | 01/15/18 | 01/14/21 | 35,000 | |
| | Dr. Gschwend | 01/15/18 | 01/14/21 | 70,000 | |
| | Dr. Huilgold | 01/15/18 | 01/14/20 | 35,000 | |
| | Dr. Sturgill | 01/15/18 | 01/14/20 | 35,000 | |
| | Dr. Yu | 01/15/19 | 01/14/21 | 35,000 | * |
| German National Academy of Sciences Leopoldina | Dr. Schmack | 05/01/18 | 04/30/20 | 66,410 | |
| | Dr. Urai | 05/01/19 | 04/30/21 | 56,268 | * |
| German Research Foundation (DFG) | Dr. Dassler Plenker | 06/01/19 | 05/31/21 | 22,837 | * |
| | Dr. Klingbeil | 06/01/18 | 05/31/20 | 48,055 | |

[1]Awarded, including direct and indirect costs
*New or competing renewals or supplements awarded in 2019

| Grantor | Program/Principal Investigator | Duration of Grant | | 2019 Funding[1] | |
|---|---|---|---|---|---|
| | Dr. Plenker | 01/01/19 | 12/31/20 | 46,500 | * |
| | Dr. Scheufele | 10/01/19 | 09/30/21 | 11,749 | * |
| Lola A. Goldring | Dr. Stillman | 10/01/19 | 09/30/20 | 90,000 | * |
| Howard Hughes Medical Institute | Drs. Johnson/Mills | 09/01/19 | 08/31/22 | 50,000 | * |
| Betty and Jim Karam | Watson School of Biological Sciences | 02/21/18 | 12/31/21 | 2,000 | |
| Annette Kade Fund | Watson School of Biological Sciences | 12/27/19 | 12/26/20 | 31,250 | * |
| Manhasset Women's Coalition Against Breast Cancer | Dr. Bhatia | 01/01/19 | 12/31/19 | 100,000 | * |
| The Meier and Linnartz Family Foundation | Dr. Li | 06/01/19 | 05/31/20 | 20,000 | * |
| The Patrina Foundation | Watson School of Biological Sciences/Women in Science and Engineering Initiative | 09/27/17 | 09/26/20 | 1,700 | |
| John and Amy Phelan Foundation | Watson School of Biological Sciences | 01/01/19 | 08/31/22 | 100,000 | * |
| The Swartz Foundation | Dr. Koulakov | 01/01/19 | 12/31/19 | 10,000 | * |
| | Dr. Engel | 01/01/19 | 12/31/19 | 61,500 | * |
| | Dr. Engel | 01/01/19 | 12/31/19 | 61,500 | * |
| Charles H. Revson Foundation, Inc. | Dr. Balasooriya | 08/01/18 | 07/31/20 | 104,744 | |
| | Dr. Furlan | 10/01/19 | 09/30/21 | 105,661 | * |
| | Dr. Syrjanen | 11/01/19 | 10/31/21 | 105,661 | * |
| Dr. Jason Sheltzer and Joan Smith | Watson School of Biological Sciences | 01/01/19 | 12/31/19 | 1,000 | * |
| The Research Foundation for State University of New York, Stony Brook | D. Cheng | 01/16/17 | 01/14/21 | 4,200 | |
| | N. Gallo | 11/16/18 | 11/15/20 | 4,200 | |
| | Y. Kim | 08/01/19 | 07/31/20 | 4,200 | * |
| | R. Raudales | 02/16/19 | 02/15/20 | 4,200 | * |
| University of Southern California | N. Anaparthy | 12/27/16 | 04/02/20 | 10,411 | |

*Training Support*

| | | | | | |
|---|---|---|---|---|---|
| New York State Department of Economic Development | Dr. Mills/CSHL Cancer Gene Discovery & Post Doctoral Research Training Program | 02/01/16 | 01/31/21 | 99,344 | |
| William Townsend Porter Foundation | Undergraduate Research Program | 01/03/19 | 01/02/20 | 11,800 | * |
| University of Notre Dame | Undergraduate Research Program | 04/01/16 | 03/31/21 | 20,000 | |

*Course Support*

| | | | | | |
|---|---|---|---|---|---|
| CSHL Translational Cancer Support | Workshop on Pancreatic Cancer | 05/17/19 | 05/16/20 | 38,400 | * |
| Foundation Fighting Blindness, Inc. | Vision: A Platform for Linking Circuits, Behavior and Perception | 04/01/19 | 03/31/20 | 5,000 | * |
| The Leona M. & Harry B. Helmsley Charitable Trust | Course Program | 02/01/19 | 01/31/20 | 1,210,000 | * |
| Howard Hughes Medical Institute | Course Program | 08/01/15 | 07/31/20 | 500,000 | |
| Estee Lauder Inc. | Course Scholarship Program | 06/26/18 | 06/25/21 | 20,000 | |
| Lustgarten Foundation | Workshop on Pancreatic Cancer | 04/01/19 | 03/31/20 | 20,000 | * |
| The Nancy Lurie Marks Family Foundation | Workshop on Autism Spectrum Disorders | 06/14/19 | 06/13/20 | 38,000 | * |
| Pancreatic Cancer Action Network, Inc. | Workshop on Pancreatic Cancer | 04/01/19 | 03/31/20 | 10,000 | * |
| Regeneron Pharmaceuticals Inc. | Regeneron Scholars Account Funder: Regeneron Pharmaceuticals | 01/01/19 | 12/31/23 | 100,000 | * |
| Society for Neuroscience/International Brain Research Organization | Summer Neuroscience Course | 07/01/19 | 06/30/20 | 21,070 | * |

*Meeting Support*

| | | | | | |
|---|---|---|---|---|---|
| Alnylam Pharmaceuticals Inc. | RNA and Oligonucleotide Therapeutics | 12/21/18 | 12/20/19 | 10,000 | |

[1]Awarded, including direct and indirect costs
*New or competing renewals or supplements awarded in 2019

| Grantor | Program/Principal Investigator | Duration of Grant | | 2019 Funding[1] | |
|---|---|---|---|---|---|
| Amgen Inc. | Blood–Brain Barrier | 11/01/18 | 10/31/20 | 7,500 | |
| Axolabs GmbH | RNA and Oligonucleotide Therapeutics | 12/21/18 | 12/20/19 | 2,500 | |
| Biogen Inc. | Blood–Brain Barrier | 11/01/18 | 10/31/20 | 5,000 | |
| CSHL Translational Cancer Research | Nutrient Signaling | 06/01/18 | 12/31/19 | 447,946 | |
| Dragonfly Therapeutics, Inc. | Harold Varmus 80th Birthday | 09/01/19 | 08/31/20 | 10,000 | * |
| Genentech, Inc. | Harold Varmus 80th Birthday | 09/01/19 | 08/31/20 | 3,000 | * |
| Generation Bio Co. | RNA and Oligonucleotide Therapeutics | 12/21/18 | 12/20/19 | 3,000 | |
| Intellia Therapeutics, Inc. | RNA and Oligonucleotide Therapeutics | 12/21/18 | 12/20/19 | 5,000 | |
| Ionis Pharmaceuticals, Inc. | RNA and Oligonucleotide Therapeutics | 12/21/18 | 12/20/19 | 3,000 | |
| Kymera Therapeutics Inc. | Ubiquitin | 03/15/19 | 03/14/20 | 2,500 | * |
| Levin Biosciences | RNA and Oligonucleotide Therapeutics | 12/21/18 | 12/20/19 | 1,000 | |
| Merck & Co. Inc. | System Immunology | 02/01/19 | 01/31/20 | 5,000 | * |
| | Microbiome | 05/15/19 | 12/31/19 | 5,000 | * |
| Merck Sharp & Dohme Corp. | Harold Varmus 80th Birthday | 09/01/19 | 08/31/20 | 10,000 | * |
| | Microbial Pathogenesis & Host Response | 06/25/19 | 12/31/19 | 5,000 | * |
| Moderna Therapeutics, Inc. | 84th Symposium RNA Control & Regulation | 05/01/19 | 04/30/20 | 2,500 | * |
| | RNA & Oligonucleotide Therapeutics | 12/21/18 | 12/20/19 | 5,000 | |
| Novartis Institutes for Biomedical Research, Inc. | Harold Varmus 80th Birthday | 09/01/19 | 08/31/20 | 10,000 | * |
| Pacific Biosciences of California, Inc. | Microbiome | 06/01/19 | 05/31/20 | 5,000 | * |
| Pfizer Inc. | Harold Varmus 80th Birthday | 09/01/19 | 08/31/20 | 10,000 | * |
| Sanofi | Harold Varmus 80th Birthday | 09/01/19 | 08/31/20 | 10,000 | * |
| Stoke Therapeutics, Inc. | 84th Symposium: RNA Control and Regulation | 05/01/19 | 04/30/20 | 5,000 | * |
| | RNA & Oligonucleotide Therapeutics | 12/21/18 | 12/20/19 | 5,000 | |
| Surrozen, Inc. | Harold Varmus 80th Birthday | 09/01/19 | 08/31/20 | 10,000 | * |
| System1 Biosciences, Inc. | Development and 3-D Modeling of the Human Brain | 11/01/19 | 10/31/20 | 10,000 | * |
| The Orphan Disease Center | Blood–Brain Barrier | 11/01/18 | 10/31/20 | 20,000 | |
| Vertex Pharmaceuticals, Inc. | RNA and Oligonucleotide Therapeutics | 12/21/18 | 12/20/19 | 12,000 | |
| ViiV Healthcare Company | Retrovirus | 02/26/19 | 02/25/20 | 20,000 | * |
| Zymergen Inc. | History of Biology and Biotechnology Research: Yeast Research, Insights, and Breakthroughs | 08/01/19 | 07/31/20 | 5,000 | * |
| *Library Support* | | | | | |
| The Ellen Brenner Memorial Fund | | 12/15/19 | 12/14/20 | 6,300 | * |
| Dr. and Mrs. Philip Goelet | | 12/21/16 | 12/20/19 | 10,000 | |
| The New York State Education Department | | 07/01/19 | 06/30/20 | 4,227 | * |
| *Preprint Server for Biology* | | | | | |
| Anonymous | Dr. Inglis | 06/01/18 | 05/31/20 | 856,440 | |
| Anonymous | Dr. Inglis | 05/01/17 | 04/30/22 | 1,291,958 | |

[1]Awarded, including direct and indirect costs
*New or competing renewals or supplements awarded in 2019

## DNA LEARNING CENTER GRANTS

| Grantor | Program | Duration of Grant | 2019 Funding† |
|---|---|---|---|
| **FEDERAL GRANTS** | | | |
| National Institutes of Health | *Barcode Long Island* | 7/14–3/21 | $135,178 |
| National Science Foundation | *Biotechnology in American High Schools: Continuing Research* | 9/18–8/19 | 28,048 |
| National Science Foundation | *Implementing DNA Barcoding for Course-Based Undergraduate Research Experiences* | 10/18–9/23 | 229,880 |
| National Science Foundation | *MaizeCODE: An Initial Analysis of Functional Elements in the Maize Genome* | 6/16–5/19 | 160,603 |
| National Science Foundation | *CyVerse: Cyberinfrastructure for the Life Sciences* | 8/18–7/23 | 187,183 |
| National Science Foundation | *The iPlant Collaborative: Cyberinfrastructure for the Life Sciences* | 9/13–8/19 | 171,068 |
| National Science Foundation | *RCN-UBE: Establishing a Genomics Education Alliance: Steps Towards Sustainability* | 9/18–8/20 | 45,107 |
| National Science Foundation | InnovATE*BIO* National Biotechnology Education Center | 10/19–9/24 | 0 |
| **NONFEDERAL GRANTS** | | | |
| Alfred P. Sloan Foundation | *DNA Center NYC* Start-up | 6/18–6/21 | 57,244 |
| Beijing No. 166 High School | Chinese Collaboration Agreement | 7/19–6/22 | 23,161 |
| Breakthrough Prize Foundation | Laboratory Design and Teacher Training for Breakthrough Junior Challenge Prize Winners | 12/15–12/19 | 7,625 |
| Health Park | Health Park Agreement | 12/15–12/20 | 7,491 |
| National Grid Foundation | Genetics Education Program | 7/19–6/20 | 0 |
| Pinkerton Foundation | *Urban Barcode Research Program* | 1/19–5/20 | 93,532 |
| Richard Lounsbery Foundation | Developing Independent Student Marine Biodiversity Research Using eDNA | 10/17–6/20 | 6,024 |
| The Simons Foundation | *Urban Barcode Research Program* | 12/17–8/20 | 104,410 |
| William Townsend Porter Foundation | *Harlem DNA Lab for Underprivileged Students* | 1/19–1/20 | 13,500 |
| Nature's Bounty Foundation | Vitamin Engineering Lab | 8/18–11/19 | 25,000 |
| New York Harbor Foundation, Inc. | Billion Oyster Project | 3/19–9/19 | 6,000 |

**The following schools and school districts participated in the *Curriculum Study* program:**

| | | | |
|---|---|---|---|
| Bellmore–Merrick Central High School District | $3,500 | Long Beach Union Free School District | $3,150 |
| East Meadow Union Free School District | $3,500 | Massapequa Union Free School District | $3,150 |
| Elwood Union Free School District | $2,100 | North Shore Central School District | $2,100 |
| Fordham Preparatory School | $2,100 | Oceanside Union Free School District | $2,100 |
| Half Hollow Schools Central School District | $2,100 | Oyster Bay–East Norwich Central School District | $2,100 |
| Harborfields Central School District | $2,100 | Plainview–Old Bethpage Central School District | $2,100 |
| Herricks Union Free School District | $2,100 | Portledge School | $3,150 |
| Island Trees Union Free School District | $2,100 | Port Washington Union Free School District | $2,100 |
| Jericho Union Free School District | $3,500 | Roslyn Union Free School District | $2,100 |
| Levittown Union Free School District | $2,100 | Syosset Central School District | $3,500 |
| Locust Valley Central School District | $2,100 | Yeshiva University High School for Girls | $2,100 |

**The following schools and school districts each contributed $1,000 or more for participation in the *Genetics as a Model for Whole Learning* program:**

| | | | |
|---|---|---|---|
| Bayshore Union Free School District | $2,695 | Great Neck Union Free School District | $12,700 |
| Berkeley Carroll School, Brooklyn | $4,280 | Green Vale School, Old Brookville | $1,662 |
| Cold Spring Harbor Central School District | $15,260 | Greenwich Country Day School, CT | $5,280 |
| Commack Union Free School District | $2,530 | Half Hollow Hills Central School District | $11,070 |
| East Williston Union Free School District | $1,347 | Hicksville Union Free School District | $1,540 |
| Elwood Union Free School District | $9,537 | Hofstra University Science and Technology Entry Program | $2,200 |
| Floral Park–Bellerose Union Free School District | $8,250 | Holy Child Academy, Old Westbury | $1,800 |
| Garden City Union Free School District | $12,600 | Huntington Union Free School District | $2,880 |

†Includes direct and indirect costs.

| | | | |
|---|---|---|---|
| Island Park Union Free School District | $2,160 | PS-IS 178, Queens | $1,600 |
| Kings Park Central School District | $3,300 | Rockville Centre Union Free School District | $11,880 |
| Locust Valley Central School District | $8,817 | Roslyn Union Free School District | $5,775 |
| Massapequa Union Free School District | $1,000 | Scarsdale Union Free School District | $6,924 |
| North Bellmore Union Free School District | $3,850 | School of the Holy Child, Rye | $1,555 |
| Northport–East Northport Union Free School District | $1,100 | Smithtown Union Free School District | $8,800 |
| Oceanside Union Free School District | $1,800 | South Huntington Union Free School District | $7,315 |
| Our Lady of the Hamptons Regional Catholic School, | | St. Patrick's School, Huntington | $2,160 |
| Southampton | $1,440 | Syosset Union Free School District | $46,255 |
| Oyster Bay–East Norwich Central School District | $1,320 | Three Village Central School District | $3,190 |
| Port Washington Union Free School District | $6,820 | Wantagh Union Free School District | $3,140 |

## BANBURY CENTER GRANTS

| Grantor | Program | *2019 Funding* |
|---|---|---|
| **NONFEDERAL SUPPORT** | | |
| 2Blades Foundation | The Plant Microbiota | $5,000 |
| AbbVie | Cancer Immunotherapy: Where to Go Next | 5,000 |
| Alfred P. Sloan Foundation | Microbiology of the Built Environment | 67,000 |
| Bill & Melinda Gates Foundation | Intermediate Indicators for Impact: The Art and Science of Effective Definition and Use of Prevention Indicators in the HIV Response | 51,177 |
| Boehringer Ingelheim Fonds | Communicating Science—Boehringer Ingelheim Fellows Retreat | 55,721 |
| Carney Institute for Brain Science at Brown University | Computational Psychiatry | 6,010 |
| Catalog DNA | DNA for Digital Storage II | 10,000 |
| Cold Spring Harbor Laboratory Corporate Sponsor Program | Emerging Issues of Privacy, Trust, and Societal Benefit from Consumer Genomics | 44,833 |
| Cold Spring Harbor Laboratory Corporate Sponsor Program | Reconceptualizing the Challenges of Direct-to-Consumer Health Products | 32,412 |
| Cold Spring Harbor Laboratory Corporate Sponsor Program | Bridging the Research-to-Practice Chasm in Digital Mental Health | 26,820 |
| Cold Spring Harbor Laboratory Corporate Sponsor Program | Integrated Control of Feeding and Energy Balance by Hypothalamic and Hindbrain Circuits | 25,417 |
| Cold Spring Harbor Laboratory Corporate Sponsor Program | CaMKII and Its Role as a Self-Tuning Structural Protein at the Synapse | 22,651 |
| Cold Spring Harbor Laboratory Corporate Sponsor Program | The Plant Microbiota | 21,115 |
| Columbia DSI | DNA for Digital Storage II | 2,000 |
| Conagen Inc. | DNA for Digital Storage II | 9,000 |
| Cygnal Therapeutics | The Nervous System in Cancer | 65,334 |
| Genentech, Inc. | Cancer Immunotherapy: Where to Go Next | 31,550 |
| Genentech, Inc. | Liquid Biopsies | 30,000 |
| Indigo Ag | The Plant Microbiota | 10,000 |
| Kallyope Inc. | Integrated Control of Feeding and Energy Balance by Hypothalamic and Hindbrain Circuits | 15,000 |
| KWS SAAT SE | The Plant Microbiota | 6,215 |
| Kyoto University | CaMKII and Its Role as a Self-Tuning Structural Protein at the Synapse | 19,430 |
| Lustgarten Foundation | Lustgarten Foundation Scientific Advisory Board Meeting | 24,760 |
| MedImmune | Integrated Control of Feeding and Energy Balance by Hypothalamic and Hindbrain Circuits | 10,000 |
| Microsoft Corporation | Bridging the Research-to-Practice Chasm in Digital Mental Health | 5,000 |
| Microsoft Corporation | DNA for Digital Storage II | 5,000 |
| MyHeritage | Emerging Issues of Privacy, Trust, and Societal Benefit from Consumer Genomics | 10,000 |
| Northwell Health–Cold Spring Harbor Laboratory Affiliation | Cancer Fibroblasts and Therapies | 61,014 |
| Northwell Health–Cold Spring Harbor Laboratory Affiliation | Glioblastoma: Why Is Impactful Science So Hard to Translate? | 60,471 |
| O'Hara & Co., Ltd. | CaMKII and Its Role as a Self-Tuning Structural Protein at the Synapse | 2,322 |
| Rhythm Pharmaceuticals | Integrated Control of Feeding and Energy Balance by Hypothalamic and Hindbrain Circuits | 10,000 |
| Rita Allen Foundation | Rita Allen Foundation Scholars Symposium 2019 | 65,165 |
| Society of Biological Psychiatry | Computational Psychiatry | 20,000 |
| Sutter Instruments | CaMKII and Its Role as a Self-Tuning Structural Protein at the Synapse | 500 |
| Syngenta | The Plant Microbiota | 6,215 |
| The Mark Foundation for Cancer Research | Liquid Biopsies | 40,000 |
| The William K. Warren Foundation (grant to Laureate Institute for Brain Research) | Computational Psychiatry | 20,000 |
| Twist Biosciences | DNA for Digital Storage II | 4,000 |

# CORPORATE SPONSOR PROGRAM FOR MEETINGS SUPPORT

Contributions from the following companies provide core support for the Cold Spring Harbor meetings program: **Corporate Benefactors:** Estée Lauder Companies; Regeneron; Thermo Fisher Scientific; **Corporate Sponsors:** Agilent Technologies; Bayer; Bristol-Myers Squibb Company; Calico Labs; Celgene; Genentech, Inc.; Merck; New England BioLabs; Pfizer; **Corporate Partners:** Alexandria Real Estate; Enzo Life Sciences; Gilead Sciences; Lundbeck; Novartis Institutes for Biomedical Research; Sanofi.

The Laboratory acknowledges the generosity of the following companies who loaned equipment and reagents to the various courses: AD Instruments; Addgene; AG Scientific, Inc.; Agilent Technologies; A-M Systems Inc.; Ametek; Andor Technology; Antibodies-Online; Aves Labs; Bangs Laboratories, Inc.; BD Life Sciences; Biolegend; Bio-Rad Laboratories; BioTek Instruments; Bitplane; Bruker Corporation; Campden Instruments, LTD; Charles River Laboratories; Coherent, Inc.; Conoptics; Cosmo Bio Co., Ltd.; CrystaLaser; CrystalGen Inc.; Electron Microscopy Sciences; Epicypher, Inc.; Eppendorf North America; GE Healthcare; GenScript Biotech; Hamamatsu Photonics; Holoeye; Intan Technologies, LLC; Intelligent Imaging Innovations; Labcyte Inc.; Leica Biosystems; Leica Microsystems Inc.; Macherey-Nagel Inc.; Molecular Devices; Morrell Instruments Co., Inc.; Narishige International USA; New Era Syringe Pump, Inc.; Nikon Instruments Inc.; Photometrics; Promega Life Sciences; ProteinSimple, Inc.; QSonica, LLC; RC Testing LLC; Scientifica; Singer Instruments; Sony Biotechnology; Sunrise Science Products; Sutter Instrument Company; Taconic Biosciences, Inc.; Takara Bio USA; the Jackson Laboratory; Thermo Fisher Scientific; Thorlabs, Inc.; Vector Laboratories; Vidrio Technologies, LLC; World Precision Instruments.

# DEVELOPMENT

The year 2019 was another stellar one for Cold Spring Harbor Laboratory. With much thanks to our supporters, $7.3 million was raised in annual support. These funds come through our three major events—the Golf Tournament at Piping Rock Club, the Women's Partnership for Science lunch, and the Double Helix Medals dinner—as well as our annual appeal and community group fund-raising. The importance of this funding cannot be overstated as it provides flexibility to support innovative research projects.

In October, Governor Andrew Cuomo helped President Bruce Stillman, Board Chair Marilyn Simons, and others cut the ribbon and mark the official re-opening of Demerec Laboratory. At the ceremony, Governor Cuomo said, "Cold Spring Harbor is at the forefront of this critical work, turning innovative research into new products and treatments. The new therapeutics center will provide the twenty-first century tools the lab's scientists need to keep saving lives and it will help grow the Long Island biomedical research corridor and the region's economy." Several exciting new scientists have been hired for this space, whose research is focused on cancer, metabolism, and the brain–body connection.

In addition to Demerec, Cold Spring Harbor completed construction of a new Organoid Facility. Dr. Dave Tuveson developed this breakthrough technology platform where three-dimensional cell culture systems reproduce a patient's tumor in a dish and recapitulate tumor sensitivities to chemotherapy and other drugs. CSHL's Organoid Facility will advance the development of high-throughput, affordable organoid-based clinical tests and help broadly disseminate these capabilities to clinical and research institutions across the country.

We cannot thank our donors enough for partnering with us as we continue to be at the forefront of scientific research and education. 2020 will surely bring excitement as we look forward to the opening of DNA Learning Center NYC, located in Brooklyn.

Thank you!

**Charles V. Prizzi**
*Vice President for Advancement and Special Advisor to the President*



Governor Andrew Cuomo, Nassau County Executive Laura Curran, and other local politicians join CSHL President and CEO Dr. Bruce Stillman, Chairman Dr. Marilyn Simons, and Honorary Trustee Dr. Jim Simons at the official reopening of Demerec laboratory.

# Cold Spring Harbor Laboratory Corporate Advisory Board

The Corporate Advisory Board (CAB) is comprised of prominent business leaders from the tristate community and is a vital source of funding and outreach for Cold Spring Harbor Laboratory. Board members are the driving force behind the Laboratory's annual golf outing at Piping Rock Club, which raises critical unrestricted funding for research and education programs. CAB president Eddie Chernoff chaired the 26th annual CSHL outing, which honored Dill Ayres. The CAB members also participate in other events and fundraisers for the Lab and are instrumental "ambassadors" to the community.



Golf Honoree Dill Ayres with his wife Sophie

## Corporate Advisory Board 2019
Edward A. Chernoff, *Chairman*

Michael Aboff, Aboff's Inc.
David Altman, Brown & Altman, LLP
Paul Amoruso, Oxford & Simpson Realty
Todd Andrews, Centerbrook Architects and Planners
Rocco S. Barrese, Dilworth & Barrese
Edward Blaskey, Sterling National Bank
Thomas J. Calabrese, Daniel Gale Sotheby's International Realty
Christopher Callaghan
John D. Catalano, Catalano Enterprises, LLC
Richard A. Catalano, KPMG, L.L.P.
Jonathan Connors, Wells Fargo
Marian Conway, Ph.D., New York Community Bank Foundation
Philip D'Avanzo, Cushman & Wakefield
Gregory DeRosa, Roanoke Holdings
Robert Dickstein, Ph.D.
David Einbinder, First Development Corporation
Jim Ford, Eppendorf North America
Brian Fox, McKinsey & Company
Amit Gandhi, M&R Hotel Group
Tom Giarraputo, Executive Cleaning Services, LLC
Lawrence Goodman, Curtis, Mallet-Prevost, Colt & Mosle L.L.P.
Thomas Gsell, R2DConsulting, LLC
Richard W. Humann, H2M architects + engineers
Robert Isaksen, Bank of America
Alan L. Jakimo, Sidley Austin LLP
Patricia Janco-Tupper, Capital Group

John C. Kean III, Kean Development Company
Michael Keenan, Wells Fargo Bank, N.A.
Norman Kelker, Ph.D., Enzo Life Sciences
Errol Kitt, GEI Consultants
Amy Koreen, M.D.
Andrew Kurita, Kettle Hill Capital Management, LLC
Laurie J. Landeau, V.M.D.
Brian D. Lee, Newmark Grubb Knight Frank
David Lessing, Lessing's
Kyle Markland, Bethpage Federal Credit Union
Jeffrey L. Martin, M.D., Sight MD
Mark McAteer, The Laurel Group
Victoria Sagona Meagher
Stephen F. Melore, Farrell Fritz, P.C.
Richard Nattis, M.D., SightMD
Robert Palatnick, DTTC
John G. Passarelli, M.D.
David Peikon
Patricia Petersen, Daniel Gale Sotheby's International Realty
Joseph Roberto, BankUnited, N.A.
Gareth Roberts, HSBC
Stephen Ross, Ph.D., Nikon
Don Saladino, Drive 495
Raju Sarwal, M.D.
Edward Strohm, Three Strohm Sisters Family Foundation
John Topolovec, TD Bank
Craig A. Weiss, T. Weiss Realty Corp.

# Cold Spring Harbor Laboratory Association

Under the leadership of CSHL Association President Michele Celestino, the CSHLA community raised $7.3 million through the annual fund campaign as well as fundraising and outreach events. The annual Dorcas Cummings Symposium dinner parties are a unique tradition where CSHLA Directors and other community members invite CSHL and visiting scientists into their homes for dinner. This year's speaker was Jennifer Doudna. The 26th Annual Golf Tournament was once again chaired by Director Eddie Chernoff and featured an added tennis tournament. The 18th annual Women's Partnership for Science Luncheon took place September 28 and featured CSHL Associate Professor Mikala Egeblad, who discussed her lab's work in breast cancer research. The Double Helix Medals Dinner was held November 6 at the American Museum of Natural History, where we honored Dr. Nancy Wexler for her work in Huntington's disease research and Boomer Esiason for his advocacy for cystic fibrosis. The CSHLA Directors continue to represent CSHL as community ambassadors and we are grateful for their service.

## Officers

Michele Munn Celestino, *President*
Kristin Olson Smith, *Vice President*
James A. Woods, *Treasurer*
Virginia Knott, *Secretary*

## Directors

Elizabeth Ainslie
Mary Auersperg
Hans E.R. Bosch
Vince Breitenbach
Lisa Broadbent
Barbara W. Callaghan
Eric Carlstrom
Edward A. Chernoff
Frank DellaFera

Nelson DeMille
Anna Saladino Gevinski
Gina Gherlone
David Goldring
Ronald A. Gottlieb
Mark Hamer
Nancy Lippman Israeli, M.D.
Carissa Jordan
Ashley Jostrom
Joel Kassimir, M.D.
Terri Keogh
Peter J. Klein
Madelyn Lombardi
Michael Maturo
Marcia Kramer Mayer
Michèle Bahnik Mercier
Mickie J. Nagel

Eileen Otto
Whitney F. Posillico
Alicia Zarou Scanlon
Lonnie Shoff
Hope Geier Smith
Debbie Parmet Sondock
Heather Spehr
David M. Stark
Mary Striano
Nancy M. Szigethy

## Honorary Directors

Mary D. Lindsay
Anne R. Meier
Cathy Cyphers Soref
Pearl F. Staller
Cynthia R. Stebbins



CSHLA Directors

# Honor Roll of Donors

2019 Golf Honoree and Former CSHL COO Dill Ayres (*center*) with Trustee Ed Travaglianti and Former Trustee Steve Lessing

*Deceased

CSHL President Bruce Stillman with 2019 Double Helix Medals Honoree Nancy Wexler and Trustee Joanne Berger-Sweeney

Rita and Arthur Levinson
LUI Che Woo Prize on behalf of Sir James D. Wolfensohn
Dr. and Mrs. Jeffrey L. Martin
Dr. Marcia Kramer Mayer
Breast Cancer Awareness Day in memory of Elizabeth McFarland
Dr. and Mrs. Richard Nattis
Estate of Thomas Parissidi
Michelle Paternoster Foundation
Regeneron Pharmaceuticals
Christina Renna Foundation
Robertson Family Fund
Charles and Marie Robertson Foundation
The Mary Ruchalski Foundation
Susan Wasserstein and George Sard
Dr. and Mrs. Raju Sarwal
Mr. and Mrs. Harry Slatkin
Kristen and Dustin Smith
Mr. and Mrs. Theodore Smith III, The Geier Foundation
Swim Across America Nassau–Suffolk
The Wasily Family Foundation
Samuel H. Wolcott

### Contributions of $10,000 +

Mary and Paul Auersperg
Austin's Purpose
Mr. and Mrs. W. Dillaway Ayres, Jr.
Breitenbach Family Foundation
Broad Hollow Bioscience Park
Lisa and Tim Broadbent
The Broder Family Foundation
Mr. and Mrs. Christopher C. Callaghan
Capital Group
Eric and Keri Carlstrom
Rita and Frank Castagna Family Foundation
Michele and Jim Celestino
Dr. and Mrs.* Bayard D. Clarkson

Maria and Geoff de Lesseps, Alpha Omega Charitable Foundation
Barbara J. Amonson and Vincent J. Della Pietra
Nelson DeMille
Dr. Lester Dubnick*
Liete and Mark Eichorn
Mr. and Mrs. Albert Gherlone
The GoGo Foundation
Oliver R. Grace, Jr., The Achelis and Bodman Foundation
Mr. and Mrs. Mark W. Hamer, Harvest Real Estate Services
Irving Hansen Foundation
Susan T. Harris
Nancy and Ron Israeli
James and Carissa Jordan
Ashley and Gabriel Jostrom
Dr. and Mrs. Joel J. Kassimir
Norman Kelker, Enzo Life Sciences, Inc.
Peter and Terri Keogh
Mr. and Mrs. Errol Kitt
Andrea B. and Peter D. Klein
Gordon Lamb, Jefferson Family Charitable Foundation
Brian Lee, Newmark Grubb Knight Frank
David Lessing, Lessing's Inc.
Madelyn and Carl Lombardi
Mr. and Mrs. John Macaskill
Edward E. Matthews
Paula and Michael Maturo, RXR Realty
The Meier and Linnartz Family Foundation
Thomas A. Moore
Monique and Doug Morris
Nina Monell Morton, The Ambrose Monell Foundation
Mickie and Jeff Nagel
National Grid Foundation

Northwell Health
University of Notre Dame
Susanne Olin, Olin Family Foundation
Wendy O'Neill
Eileen and Jonathan Otto
David P. Pearson
Pfizer
William Townsend Porter Foundation
Mrs. Richardson Pratt, Jr. and Family, The Spionkop Charitable Trust
The Thomas and Marina Purcell Family Foundation
Rauch Foundation
John R. Reese
Amy Falls and Hartley Rogers
Stephen Ross, Nikon Instruments, Inc.
Pamela and Richard Rubinstein Foundation
Alicia Zarou Scanlon and Richard T. Scanlon
Joan Smith and Jason Sheltzer
Michael and Lonnie Shoff
John Sobolewski
David Stark, Teva Pharmaceuticals
Dr. and Mrs. Bruce W. Stillman
The Lauri Strauss Leukemia Foundation
Mary and Vincent Striano, Striano Electric Co., Inc.
Nancy Szigethy
Anne Drackett Thomas
Joan P. Tilney
Mr. and Mrs. Edward Travaglianti
Mr. and Mrs. George Tsunis
Mr. and Mrs. Paul A. Vermylen, Jr.
Dr. and Mrs. Michael Wigler
Elisabeth R. Woods Foundation
Ann Eden Woodward Foundation
Susan and Bob Wright
The Bradley Zankel Foundation



CSHLA Directors Kristin Olson Smith and Barbara Callahan with Erin Talamas and Molly Henry at the Women's Partnership for Science lunch

*Deceased

Double Helix honoree Boomer Esiason (*right*) with CSHL President Bruce Stillman and Northwell Health CEO Michael Dowling

Sal Princiotta, Medline Industries, Inc.
John and Cheryl Pufahl
Joe Roberto, BankUnited, N.A.
Gareth Roberts, HSBC Bank USA, NA
Lari and Frank Russo
Michelle Lewis and Eric Salzman, Lewis Family Fund
Jeremy Shao, Infinity Interactive, Inc.
Mr. and Mrs. Douglas S. Soref
Heather and Richard Spehr
Michele Miroff and Carl Studer
Three Strohm Sisters Family Foundation
Trinity College
Trisomy 18 Foundation
John B. Vermylen
Craig Weiss, T. Weiss Realty Corp.
Deborah Norville and Karl Wellner
Georgene and Steven Winick

## Contributions of $5,000 +

Michael Aboff, Aboff's
David N. Altman, Brown & Altman, LLP
Paul Amoruso, Oxford & Simpson Realty Services, Inc.
Fred and Janet Baron
Ed Blaskey, Sterling National Bank
Robert Brennan, Ripco Real Estate LLC
Robert Butler, NetJets
Mr. and Mrs. Thomas J. Calabrese, Jr.
Mr. and Mrs. George W. Carmany III
John D. Catalano
Centerbrook Architects and Planners, LLC
Rita M. Cleary
Mr. and Mrs. Jonathan Connors
Ann Conway, Dau Family Foundation
Dr. Marian Conway, New York Community Bank Foundation
Mr. and Mrs. James C. Cook
Jane Duggan
Lee MacCormick Edwards Charitable Foundation
Bob and Diane Fagiola
Michael L. Focazio, Carissa Maringo Fund
Jim Ford, Eppendorf
Fortunato Sons Contracting, Inc.
Putney Cloos and Brian Fox
Mr. and Mrs. Amit Gandhi
Tom Giarraputo, Executive Cleaning Services
Glen Cove C.A.R.E.S.
Randall and Mary Hack Foundation
Robin Hadley
Hearts Who Care
Mary Jane Helenek
Jim and Robin Herrnstein
Robert Isaksen, Bank of America Merrill Lynch
Alan Jakimo
Michael Keenan, Wells Fargo Private Bank

Drs. Amy and Roger Koreen
Andrew Kurita
Mr. and Mrs. Richard Leeds
Robert T. Lindsay
Amanda and Tom Lister
Long Island 2 Day Walk to Fight Breast Cancer
Robert L. Marcus
Kyle Markland, Bethpage Federal Credit Union
Masthead Cove Yacht Club
Mark McAteer, The Laurel Group
Stephen F. Melore, Farrell Fritz, P.C.
Mitchells
Norwood Foundation, Inc.
Valerie Ohrstrom
Robert and Lauryl Palatnick
Patricia J. Petersen, Daniel Gale Sotheby's International Realty

## Contributions of $1,000 +

Debra and Scott Arenare
Hal and Tricia Avidano
Deanna Bahel
Dorothy T. Baldwin
Baranay Family Foundation
Stephen Barrese, Dilworth & Barrese, LLP
Mr. and Mrs. Dominic P. Bencivenga
Kathy and Gene Bernstein
Elizabeth Bertani
Elizabeth B. Borden, Katharine Bradford Foundation
Dr. Michael Botchan
Dr. William Braden, Porter Braden Fund at the Rhode Island Foundation
Laura Louise Breyer
Sergey Butkevich
Stephen and Carol Canter
Devon and Tommy Carroll
Louise M. Parent and John Casaly
Mary and Richard Catalano



Lel Gimbel, CSHL President Bruce Stillman, CSHLA Directors Elizabeth Ainslie and Alicia Zarou Scanlon at the Double Helix Medals

The Mary Ruchalski Foundation presented a check to Dr. Chris Vakoc and his lab members to support sarcoma research

The Elizabeth and Frank Chau Family
    Foundation
Susan and Glenn Cohen
Diana Conte
Cosel-Pieper Family Foundation
Lucy P. Cutting
Del Vecchio Family Foundation
Tracy Dellomo, UBS Financial Services
David Einbinder
Mr. and Mrs. Frederick W. Engel
Risa and Noah Finkel Charitable Fund
Floral Park Police Benevolent
    Association
The Frankfurth Family Foundation
Anneke and Howard Gaber
E. Maxwell Geddes, Jr.
Carolyn and Leonard Gero
Pierre and Paula Gonthier Family
    Foundation
Mr. and Mrs. William H. Grover
Dr. and Mrs. Thomas C. Gsell
Qing Gu
Lynn and Frank Gundersen
Reverend Linda Peyton Hancock
John Hanson
Hargraves Family
Mr. and Mrs. Stuart Henry
Nouria Hernandez and Winship Herr
William A. Herzog
Joan Morthland Hutchins
Walter B. James Fund No. 2
James D. Karam, Ph.D.
Kathy and Dr. Victor Klein
The Klinsky Family Charitable Fund
Mr. and Mrs. James Large
Edward N. Lee, KPMG LLP
Fred Leonardo, Electronix Systems
    Central Station Alarms, Inc.
James Jiangyan Li
Mr. and Mrs. Neil T. McGoldrick

Pamela M. Robb-Melius and Gary A. Melius
Walter and Sheila Mondschein
Mr. and Mrs. Stephen V. Murphy
Nature's Answer Foundation
James J. Norman
Patricia and Hugh O'Kane, Jr.
Mr. and Mrs. Frank T. O'Keefe
Kerrill O'Mahony
Philip M. Panarelli
Debbie Parmet
Rhoda Parmet
Mr. and Mrs. Nicholas B. Paumgarten
Mr. and Mrs. David Peikon
Dr. and Mrs. Anthony F. Piazza
Pilkington Family Fund
Carole Pittelman, The Litwin Foundation,
    Inc.

Nicole and Charlie Prizzi
Bonnie and Bennett Rechler
Robert Regan
Bernard and Anne Reynolds
Rich Roberts
Linda K. Rodgers
Arthur M. Rogers, Jr.
Dr. and Mrs. Gerald M. Rubin
Mr. and Mrs. Adam Savarese
Mr. and Mrs. Kristopher J. Selden
Mr. and Mrs. Enrique F. Senior
Drs. Mona and David Spector
James and Jane Spingarn
Frank Spokane
Mrs. David S. Taylor
Cora Michalis Thomas
Prof. Jeremy W. Thorner
Pamela M. Thye
Louise Giffuni Tiernan Foundation
Patricia W. Timpson
Leslie P. Farhangi and John P. Tuke
Universe Kogaku (America), Inc.
George Vande Woude
Sarah Vermylen Trust
Richard John Walk
Hweigene H. Wang
Douglas A. Warner III
Christine Wasserstein
Gerald I. White
Sandy and Jennifer Williams

**In-Kind Gifts**

Americana Manhasset
Sophie and Dill Ayres
Banfi Vintners
Besito
Bottles and Cases
Lisa Broadbent
Brooks Brothers



Cold Spring Harbor Laboratory commemorated the naming of the SightMD facility, supported by Heidi and John Passarelli, Roseann and Jeff Martin, Dayna and Rich Nattis, and Mona and Raju Sarmal.

Butera's Restaurant of Woodbury
Barbara Callaghan
John D. Catalano
Claytime
Cold Spring Harbor Wine Shoppe
Country Club Studios
Curtis, Mallet-Prevost, Colt & Mosle LLP
Daniel Gale Sotheby's International
    Realty Inc.
Nelson DeMille
Drive 495
Eileen Kathryn Boyd Interiors
Eleanor Stone NYC
John W. Engeman Theater
Estee Lauder
Fiorello Dolce
Esther Fortunoff
Frassanito Jewelers
Candido Fuentes-Felix, M.D.
GarageTek
Gina and Al Gherlone
Girl Obsessed
Walter and Dagnia Goldschmidts
Harbor Mist Restaurant
Harlan Estate
Huntington Country Club
Huntington Indoor Tennis
Ron Israeli, M.D.
Jewel
Jonathan's Ristorante
Kimberly Gorman Muto Photography
Living
MARS
J. McLaughlin
Mission Nutrition
Mitchells Huntington
Molton Brown
Nest
Nikon Instruments, Inc.
Oheka Castle Hotel and Estate
Heidi and John Passarelli
Beau R. Peelle, Ph.D.
David Peikon
Poll Restaurants
Prime Restaurant
Red Restaurant
Annabel Romero
Safavieh
Sandbar
Raju Sarwal, M.D.
Scarsella's
Tom Schaudel
SGS Original Art, Susan Graham Switala
The Shed
Sports Illustrated Golf Group
Priya Sridevi
Sterling National Bank
Grace and Bruce Stillman
TD Bank
Title Boxing Club of Huntington
Jan Tozzo, Glass.Pieces.Studio
Leah Trabich
Cathy Freudenberg Traykovski

Tres Jolie Salon
Marjorie van de Stouwe, M.D.
Walters Faith
W.B. Mason
Wempe
Windham House

## Contributions in honor of

Dill Ayres
Jamie Nicholls and Fran Biondi
Pien Bosch
Joseph Burke
Mr. and Mrs. William J. Candee III
Kristina Perkin Davison
Boomer Esiason
Jeffrey E. Kelter
Ginny & David Knott
Mary D. Lindsay
Lisa and Jonathan Lown
Laurie Landeau and Robert Maze
Victoria S. Meagher
Larry Norton
Kate O'Phelan
Sloane Dunlap Petrila
Renna Family
Susan Wasserstein and George Sard
Thomas A. Saunders III
Dr. Alfred and Barbara Sforza
Kristin Olson Smith
Freddie Staller
Bruce W. Stillman
Georgene and Steven Winick
Evelyn M. Witkin

## Contributions in memory of

Linda Aldhefery
Sandy Aranoff
Christopher Battista
Giuseppe Bertani
Patricia Boudreau
Ellen Brenner
Carol A. Buonaiuto
Robert Cafferkey
John Paul Cleary
Mary and Richard Cyphers
Marilyn DeLalio
Sandy DeMille
Maryliz Dickerson
Daniel C. DiNapoli
Lester Dubnick
Craig Dumain
Claire Martin Fairman
Thomas Farley
Christine Fazio
Brian J. Flaherty
Nancy Slaughter Gay
Dr. Christopher G. Goff
Craig Goldman
Eileen Goldwyn
Melvin and Barbara Gorelick
Lorraine and Oliver Grace

Bruce Guthart
Teresa Haire
Charles Harris
Willa N. Herzog
Barbara Ann Hickey
Barbara Hoover
Sumalee Huangthaisong
Andy Kaplan
Shirley Katzter
Ruth Kirsh
Vivian Kornet
Robert Lippman
Thomas Mantovi
Judith Marcus
Carissa Maringo
Olga Markarian
Edwin Marks
Elizabeth McFarland
Frank Morelli
Monica Mowery
Francis O'Connor
Kevin O'Connor
Gerald David Olin
Gertrude Ornstein
Clive Peacock
Dr. Cecile Pickart
Gail Poinelli
Kenneth Rudolf
Ethel Ryder
Diane Emdin Sachs
Micky Sayres
Ralph Schelin
June Schlanger
Mike Schramm
Andrew Seligson
Alan Siegel
John D. Simpson and Adelaide Shirley
    Donohue Simpson
David Smith
Jason Stewart
Rajani Lakshmi Velivela
Sister Catherine Jean Walkowski
John Wandrey
Werner (Chuck) Weiss
Edna Yurasko



Women's Partnership for Science honoree Pien Bosch (*right*) with Priscilla Press

# LABORATORY MANAGEMENT

**Bruce W. Stillman, Ph.D.**, President & CEO
**John P. Tuke**, Chief Operating Officer

## RESEARCH

**David L. Spector, Ph.D.**
*Director of Research*

**Walter L. Goldschmidts, Ph.D.**
*Vice President, Executive
Director of Sponsored Programs*

**Sydney Gary, Ph.D.**
*Director, Research Operations*

### Professors

Jeff Boyd, Ph.D.
Douglas Fearon, M.D.
Hiroyasu Furukawa, Ph.D.
Thomas Gingeras, Ph.D.
*Professor and Head Functional
Genomics*
Z. Josh Huang, Ph.D.
David Jackson, Ph.D.
Leemor Joshua-Tor*, Ph.D.
*Program Chair, Cancer and
Molecular Biology*
Adam Kepecs, Ph.D.
Alexei Koulakov, Ph.D.
Adrian Krainer, Ph.D.
*Deputy Director, CSHL
Cancer Center*
Bo Li, Ph.D.
Zachary Lippman*, Ph.D.
Robert Martienssen*, Ph.D.
*Program Chair, Genomics and
Plant Biology*
W. Richard McCombie, Ph.D.
Alea Mills, Ph.D.
Partha Mitra, Ph.D.
Adam Siepel, Ph.D.
*Program Chair, Simons Center
Quantitative Biology*
Nicholas Tonks, Ph.D
*Deputy Director, CSHL Cancer
Center*
Lloyd Trotman, Ph.D.
David Tuveson, M.D., Ph.D.
*Director, CSHL Cancer Center*
Christopher Vakoc, M.D., Ph.D.
Linda Van Aelst, Ph.D.
Michael Wigler, Ph.D.
Anthony Zador, M.D., Ph.D.

### Associate Professors

Dinu Albeanu, Ph.D.
Anne Churchland, Ph.D.
*Program Chair, Neuroscience*
Mikala Egeblad, Ph.D.
Jesse Gillis, Ph.D.
Christopher Hammell, Ph.D.
Molly Hammell, Ph.D.
Ivan Iossifov, Ph.D.
Alexander Krasnitz, Ph.D.
Dan Levy, Ph.D.
Saket Navlakha, Ph.D.
Pavel Osten, M.D., Ph.D.
Stephen Shea, Ph.D.
Raffaella Sordella, Ph.D.
Doreen Ware**, Ph.D.

### Assistant Professors

Alexander Dobin, Ph.D.
Camila dos Santos, Ph.D.
Tatiana Engel, Ph.D.
Tobias Janowitz, M.D., Ph.D.
Justin Kinney, Ph.D.
Peter Koo, Ph.D.
Je Hyuk Lee, M.D., Ph.D.
David McCandlish, Ph.D.
Vincent Pedmale, Ph.D.
Jessica Tollkuhn, Ph.D.

### Research Professor

Darryl Pappin, Ph.D.

### Research Associate Professor

Kenneth Chang, Ph.D.

### Research Assistant Professors

Scott Lyons, Ph.D.
Jonathan Preall, Ph.D.
*Research Assistant Professor
and Head of Genomics
Tech Development*
Michael Ronemus, Ph.D.
Andrea Schorn, Ph.D.
Johannes Yeh, DDS, Ph.D.
Seungtai Yoon, Ph.D.

### Senior CSHL Fellow

Lingbo Zhang, Ph.D.

### CSHL Fellows

Semir Beyaz, Ph.D.
Hannah Meyer, Ph.D.
Jason Sheltzer, Ph.D.

### Clinical Fellows

Rintaro Koike, M.D.
Koji Miyabayashi, M.D., Ph.D.
Kenneth Yu, M.D.

### Adjunct Professors

Richard Barakat, M.D., M.B.A.
Gregory Hannon, Ph.D.
Scott Lowe, Ph.D.
Larry Norton, M.D.

### Adjunct Associate Professors

Gurinder Atwal, Ph.D.
Michael Schatz, Ph.D.

### QB Visiting Scholar

Bhubaneswar Mishra, Ph.D.

### Senior Computer Scientists

Peter Andrews, Ph.D.
Boris Yamrom, Ph.D.

### Computer Scientists

Osama El Demerdash, Ph.D.
Vivek Kumar, Ph.D.

### Senior Research Investigator

Zihua Wang, Ph.D.

### Research Investigators

Joan Alexander, M.D.
Lindsey Baker, Ph.D.
Elad Elkayam*, Ph.D.
Clary Evans, MBChB
Qing Gao, M.D., Ph.D.
Carla Gauss, Ph.D.
Sara Goodwin, Ph.D.
Ana Henriques da Costa, Ph.D.
Gilbert Henry, Jr., Ph.D.
Manzar Hossain, Ph.D.
Jonathan Ipsaro, Ph.D.

Yoon-Ha Lee, Ph.D.
Youngkyu Park, Ph.D.
Michael Regulski, Ph.D.
Benjamin Roche, Ph.D.
Katharina Schmack, M.D.,
Ph.D.
Yi-Jun Sheu, Ph.D.
Damianos Skopelitis, Ph.D.
Asya Stepansky, Ph.D.
Kai Yu, Ph.D.
Huiqing Zhan, Ph.D.

### Manager, Clinical and Translational Research

Jessica Gotterer

### Visiting Scientists/ Collaborative Scientists

Taimour Baslan, Ph.D.
David Bonda, M.D.
David Chitty, D.O.
Ebru Demir, Ph.D.
Augusto Diniz, Ph.D.
Saikat Nandi
Aaron Nizam, M.D.
Lindsay Puckett
Kara Raphael, M.D.
Xiaonan Richard Sun, Ph.D.,
M.D.
Herve Tiriac, Ph.D.
Prashant Vempati, M.D.
Michael Wulf
Kenny Ye, Ph.D.

### Visiting Professors

Matthew Ridley, Ph.D.
Stuart Weisbrod, Ph.D.

### Director, Gene Targeting Shared Resources

Leyi Li, Ph.D.

### Senior Computational Fellow

Megan Crow, Ph.D.

### Translational Fellow

Fieke Froeling, M.D., Ph.D.

**Postdoctoral Fellows**

Oluwadamilare Adebambo, Ph.D.
Cina Aghamohammadi, Ph.D.
Xu An, Ph.D.
Oliver Artz, Ph.D.
Balasooriya Balasooriya, Ph.D.
Walter Bast, Ph.D.
Pascal Belleau, Ph.D.
Matthias Benoit*, Ph.D.
Kuhulika Bhalla, Ph.D.
Sonam Bhatia, Ph.D.
Francesco Boato, Ph.D.
Christopher Bonham, Ph.D.
Marion Bosc, Ph.D.
Christian Bravo-Rivera, Ph.D.
Robert Bronstein, Ph.D.
Jonathan Cahn*, Ph.D.
Hong Goo Chae, Ph.D.
Spyridon Champeris
    Tsaniras, Ph.D.
Chen Chen, Ph.D.
Wei Chia Chen, Ph.D.
Xiaoyin Chen, Ph.D.
Yushu Chen, Ph.D.
Quentin Chevy, Ph.D.
Tsung Han Chou, Ph.D.
Tae Mun Chung, Ph.D.
Samantha Cyrill, Ph.D.
Juliane Dassler Plenker, Ph.D.
Cristiane de Santis Alves
    Rosa, Ph.D.
Hanfei Deng, Ph.D.
Jakub Dolata*, Ph.D.
Noah Dukler, Ph.D.
Roman Dvorkin, Ph.D.
Ela Elyada, Ph.D.
Mary Feigman, Ph.D.
Miriam Fein, Ph.D.
Stephan Fischer, Ph.D.
Matthew Fisher, Ph.D.
Leonie Foerschner, M.D.
Talitha Forcier, Ph.D.
Daniel Fuerth, Ph.D.
Akihiro Funamizu, Ph.D.
Alessandro Furlan, Ph.D.
Yuan Gao, Ph.D.
Ankur Garg, Ph.D.
Mikhail Genkin, Ph.D.
Fabian Gerth, Ph.D.
Debarati Ghosh, Ph.D.
Nicholas Gladman, Ph.D.
Olivier Gschwend, Ph.D.
Wuqiang Guan, Ph.D.
Priyanka Gupta, Ph.D.
Rasmani Hazra, Ph.D.
Jia He*, Ph.D.
Xueyan He, Ph.D.

Anat Hendelman, Ph.D.
Diego Hernandez Trejo, Ph.D.
Longwen Huang, Ph.D.
Yuhan Huang, Ph.D.
Dhananjay Huilgol, Ph.D.
Stellar Hur, Ph.D.
Alessandra Inguscio, Ph.D.
Matt Jaremko, Ph.D.
Li-Feng Jiang-Xie, Ph.D.
Risa Karakida, Ph.D.
Gukhan Kim, Ph.D.
Brian Kinney, Ph.D.
Munenori Kitagawa, Ph.D.
Olaf Klingbeil, Ph.D.
Hideki Kondo, Ph.D.
Choon-Tak Kwon, Ph.D.
Christopher Langdon, Ph.D.
Jae Hyung Lee, Ph.D.
Seung Cho Lee*, Ph.D.
Bo Li, Ph.D.
Jiayun Li, Ph.D.
Shujing Li, Ph.D.
Siran Li, Ph.D.
Yi Li, Ph.D.
Kuan-Ting Lin, Ph.D.
Bodu Liu, Ph.D.
Lei Liu, Ph.D.
Bin Lu, Ph.D.
Vinay Mandati, Ph.D.
Grinu Mathew, Ph.D.
Katherine Matho, Ph.D.
Olivia Mendivil Ramos, Ph.D.
Kevin Michalski, Ph.D.
Sushanta Mishra, Ph.D.
Andrea Moffitt, Ph.D.
Hemanth Mohan, Ph.D.
Haiwei Mou, Ph.D.
Rodrigo Munoz Castaneda,
    Ph.D.
Adriana Munoz Jimenez, Ph.D.
Simon Musall, Ph.D.
Sukalp Muzumdar, Ph.D.
Arun Narasimhan, Ph.D.
Farhana Nasrin, Ph.D.
Gal Nechooshtan, Ph.D.
Kin On*, Ph.D.
Torben Ott, Ph.D.
Huiwu Ouyang, Ph.D.
Meng Ouyang, M.D., Ph.D.
Dennis Plenker, Ph.D.
Anna Posfai, Ph.D.
Yongjun Qian, Ph.D.
Mohammad Rahman, Ph.D.
Nissim Ranade, Ph.D.
Ashlan Reid, Ph.D.
James Roach, Ph.D.
Thiago Santos Gouvea, Ph.D.

Armin Scheben, Ph.D.
Florian Scheufele, Ph.D., M.D.
Klaske Schukken, Ph.D.
Donglai Shen, Ph.D.
Yang Shen, Ph.D.
Yanliang Shi, Ph.D.
Atsushi Shimada*, Ph.D.
Daniele Silva Pereira
    Rosado, Ph.D.
Timothy Somerville, Ph.D.
Sweta Srivas, Ph.D.
Sarah Starosta, Ph.D.
James Sturgill, Ph.D.
Lijuan Sun, Ph.D.
Xueqin Sun, Ph.D.
Hamsini Suresh, Ph.D.
Johanna Syrjanen, Ph.D.
Julian Taranda, Ph.D.
Hsiu-Chi Ting, Ph.D.
Claudia Tonelli, Ph.D.
Michael Tramantano, Ph.D.
Thu Tran, Ph.D.
Benno Traub, Ph.D., M.D.
Anne Urai, Ph.D.
Raditya Utama, Ph.D.
Anand Vasudevan, Ph.D.
Prabhadevi Venkataramani, Ph.D.
Ledong Wan, Ph.D.
Bo Wang, Ph.D.
Bor-Shuen Wang, Ph.D.
Minghui Wang, Ph.D.
Xingang Wang, Ph.D.
Xinyi Wang, Ph.D.
Zhikai Wang, Ph.D.
Yiliang Wei, Ph.D.
Sze Wong, Ph.D.
Xiong Xiao, Ph.D.
Yunyao Xie, Ph.D.
Xiaosa Xu, Ph.D.
Alexander Xue, Ph.D.
Tao Yang, Ph.D.
Zhaolin Yang, Ph.D.
Min Yao, Ph.D.
Toyoki Yoshimoto, Ph.D., M.D.
Li Yuan, Ph.D.
Dinar Yunusov, Ph.D.
Sophia Zebell*, Ph.D.
Xian Zhang, Ph.D.
Xiang Zhao, Ph.D.
Yixin Zhao, Ph.D.
Juannan Zhou, Ph.D.
Imanol Zubiete Franco, Ph.D.

**Graduate Students**

Dexter Adams
Disha Aggarwal
Lyndsey Aguirre

Seamus Balinth
Batuhan Baserdem
Benjamin Berube
Brianna Bibel
Anna Lien Bouhuis
Sara Boyle
Leah Braviner
Katherine Broekman
Yuxin Cen
Kung-Chi Chang
Derek Cheng
King Hei Cheng
Hsiang-Chen Chou
Charlie Chung
Danielle Ciren
Alberto Corona
Patrick Cunniff
Marie Dussauze
Ilgin Ergin
Miriam Ferrer Gonzalez
Connor Fitzpatrick
William Galbavy
Nicholas Gallo
Jordan Gegenhuber
Sanchari Ghosh, Ph.D.
Hamza Giaffar
Jacqueline Giovanniello, Ph.D.
Kristina Grigaiyte, Ph.D.
Amritha Varshin Hanasoge
    Somasundara
Benjamin Harris
Kelly Hills-Muckey
Yixin Hu
Yuzhao Hu
Elizabeth Hutton
Shruti Iyer
David Johnson
Young Jin Kim
Christopher Krasniak
Asad Aziz Lakhani
Matthew Lee
Cristian Lopez-Cleary
Shaina Lu
Devon Lukow
Dennis Maharjan
Diogo Maia e Silva, M.D.
Craig Marshall
Katarina Meze
Judith Mizrachi
Ziyi Mo
Philip Moresco
Manojit Mosur Swamynathan
Alexandra Nowlan
Kathryn O'Neill
Tobiloba Oni
Alexa Pagliaro
Michael Passalacqua

*Employee of the Howard Hughes Medical Institute

Tzvia Pinkhasov
Sofya Polyanskaya
Zhe Qian
Leonardo Jared Ramirez Sanchez
Ricardo Raudales
Suelynn Ren
Jenelys Ruiz
Deborah Rupert
Danilo Segovia
Luqun Shen
Margaret Shevik
Padmina Shrestha
Sergey Shuvaev
Nicole Sivetz
Dongyan Song
Olya Spassibojko
Martyna Sroka
Joshua Steinberg
Ziqi Tang
Jennifer Thalappillil
Shushan Toneyan
Ngoc Tran
Gaurang Trivedi
Julia Wang
Jonathan Werner
Xiaoli Wu
Cole Wunderlich
Wenbo Xu
Ran Yan
Chaoqun Yin
Zhezhen Yu
Chengxiang Yuan
Narges Zali
Anqi Zhang
Qian Zhang
Siwei Zhang, Ph.D.

**Computational Science**

Saiteja Abbu
Samik Banerjee
Daniel Blibaum
Kapeel Chougule
Astrid Deschenes
Corey Elowsky
Evan Ernst*
Nathan Fox
Yuan Hao, Ph.D.
Hussein Hijazi, Ph.D.
John Hover
Bingxing Huo, Ph.D.
Ying Jin, Ph.D.
Benjamin Kaminow
Jude Kendall
Melissa Kramer
Sunita Kumari, Ph.D.
John Lee, Ph.D.
Forrest Li

Xu Li
Meng Kuan Lin, Ph.D.
Zhenyuan Lu
Steven Marks
Ivar Meijs
Fahimeh Mirhaj
Andrew Olson
Xiaoli Qi, Ph.D.
Ritika Ramani
Oliver Tam, Ph.D.
Ammar Tareen, Ph.D.
Marcela Tello-Ruiz, Ph.D.
Marygrace Trousdell
Liya Wang, Ph.D.
Xiaofei Wang, Ph.D.
Xuehong Wei
Matthew Wroten, Ph.D.
Fengyuanshan Xu
Miu ki Yip

**Research Associates**

Habeeb Al Sudani
Carmelita Bautista
Yon Chang
Hilary Cox
Jorg Drenkow
Eric Gamache
Hyun Soo Kim, Ph.D.
Lisa Manche
Joseph Merrill
Ramesh Palaniswamy
Suzanne Russo
Caizhi Wu, M.D.
Melody Wu, Ph.D.
Lifang Zhang, Ph.D.

**Scientific Project Managers**

Katherine Brenner
Kathleen Kelly, Ph.D.
Alexander Vaughan, Ph.D.

**Laboratory Facility Managers**

Sabrina Boettcher
Amy Brady
Patricia Brady
Julie Cheong
Robert Eifert
Elena Ghiban
Stephanie Goldsmith*
Calvin Lee
Pamela Moody
Timothy Mulligan
Eleni Noutsos, Ph.D.
Joseph Simorowski*
Dennis Thomas, Ph.D.
Ming Wang

Tse-Luen Wee, Ph.D.

**Scientific IT Support**

John Pane
David Trimboli
Peter Van Buren

**Scientific Administrators**

Maria Anzaldi
Lorraine Baldwin
AnnMarie Ballew
Christine Bedell
Patricia Bird
Barbara Cascone
Melissa Daley
Becky Dong
Suzan Downey
Susan Fredericks
Irene Gill
Denise Hallinan
Lisa Kimoto
Caryn Koza
Idee Mallardi
Brenda Mejia
Lindsay Monaco
Sandra Moran
Kylie Parker
Catherine Scalise
Dorothy Tsang Hu
Madeline Wisnewski

**Research Technicians**

Ashley Adler
Valeria Aguillon Rodriguez
Andalus Ayaz
Anastasiia Bludova
Natasha Bobrowski-Khoury
Lily Brine
Erin Brosnan
Denise Cahn
Megan Chamberland
Lisa Christensen
Devon Cowan
Jenesis Curtis
Cassidy Danyko
Martin Davis
Zarmeena Dawood
Abhishek Doshi
Rhonda Drewes
Sintia Escobar Avelar
Onur Eskiocak
Audrey Fahey
Margaret Falkowski
Christy Felice
Eugene Fong
Jared Freedman

Zachary Galluzzo
Ya Gao
Libia Garcia
Nicole Gemmill
Steven Gluf
Antoine Gruet
Inessa Hakker
Kaarina Hanington
Brianna Harman
Joshua Hatfield
Cheng Wen Hsu
Fangle Hu
Qingting Hu, M.D.
Seraya Jones*
Clancy Kelahan
Anup Khanal
Jennifer Kim*
Robert Kwapich
Alana Lashley
Yan Li
Ying Hsiu Liu
Landon Lodato
Matthew Lucito
Krisann Marquart
Senem Mavruk Eskipehlivan
Colin McLaughlin
Joseph Milazzo
Kristin Milicich
Stephanie Muller
Payal Naik
David Ng
Joseph O'Rourke
Jie Ouyang
Katherine Papciak
Sarah Park
Hardik Patel
Ryan Peters
Vindhya Pillai
Craig Podszus
Umamaheswari Ramu
Claire Regan
Michael Riggs
Gina Robitaille
Julie Rosenbaum
Polona Safaric Tepes
Erin Sausville
Stephen Savoia
John Scheuring, Jr.*
Jennifer Shapp
Regina Shaw
Caroline Shipsey
Aubrey Siebels
Noriko Simorowski
Tara Skopelitis
Natalia Stec
Logan Stecher
Yu-Chi Sun

*Employee of the Howard Hughes Medical Institute

Abhishek Vijay Thakkar
Cassandra Viola
Wiktor Wadolowski
Jing Wang
Robert Wappel
Shih Ting Yang
Brian Yueh
Shupei Zhang
Weixin Zhong, Ph.D.

### Laboratory Facility Staff

Rachel Rubino, D.V.M.
*Director of Lab Animal
    Resources/Attending
    Veterinarian*
Doris Abrego
Alexander Aguilera
Carlos Ballon
Lisa Bianco
Amanda Bjertnes
Oral Bryan
Amy Burrows
Manuel Canas
Roberto Castro
Jodi Coblentz
Steven Collins
Kristin Cox
Charlene De Poto
Maria Demott
John Donnelly
Eileen Earl
Susan Guterman
Jill Habel
Thomas Hicks
Tristan Jarrett
Amanda Jensen
Michael Labarbera
Andrew Labriola
Dylan Marcus
Alyssa Molloy
Maria Mosquera
Gustavo Munoz
Angel Oliveira
Gregory Orkiszewski
Anup Patel
Jason Pellegrini
Bryan Rose
Laura Stonebridge
Walter Vigil Garcia

### Media Makers/ Farm Assistants

Nancy Bolanos
Terrance Chisum
Meredith Garrison
Adriana Hincapie
Kelly Ann Holfester
Andrew Krainer

Michele Murrell
Shouyun Qiao
Amy Qiu
Kyle Schlecht

## EDUCATION

### Banbury Center

Rebecca Leshan, Ph.D.
    *Director*
Michelle Corbeaux
Allison Eichler
Barbara Polakowski

### CSHL Press

John R. Inglis, Ph.D.
    *Executive Director*
Barbara Acosta
Janet Argentine
Sonali Bhattacharjee, Ph.D.
Kevin John Black
Cynthia Blaut
Carol Brown
Kathleen Bubbeo
Kathleen Cirone
Laureen Connell, Ph.D.
Marie Cotter
Jennifer DeLeon, Ph.D.
Ahmet Denli, Ph.D.
Corrisa Farmer
Bibiane Garite
Michael Henigman
Samantha Hindle, Ph.D.
Nancy Hodson
Deborah Jarski
Brandon Kelly
Catherine Kelly
Tara Kulesa
Dana Macciola
Wayne Manos
Mala Mazzullo
Joanne McFadden
Mary Mulligan
Stephanie Novara
Stephen Nussbaum
Dorothy Oddo
Robert Redmond
Theodore Roeder
Mary Sabala
Diane Schubach
Richard Sever, Ph.D.
Inez Sialiano
Marcie Siconolfi
Michael Siragusa
Maria Smit, Ph.D.
Hillary Sussman, Ph.D.
Linda Sussman
Denise Weiss

### DNA Learning Center

David A. Micklos
    *Executive Director*
Lina Bader
Louise Bodt
Justin Burke
Megan Capobianco
Elna Carrasco
Lauren Correri
Heather Cosel-Pieper
Cristina Fernandez Marco, Ph.D.
Cornel Ghiban
Jennifer Hackett, Ph.D.
Daniel Jacobs
Brittany Johnson
Mary Lamont
Susan Lauter
Melissa Lee
Allison Mayle, Ph.D.
Amanda McBrien
Colleen McKechnie
Valerie Meszaros
Eric Nash, Ph.D.
Lydia Paradiso
Michael Paul
Sharon Pepenella, Ph.D.
Jeffry Petracca
Colette Riccardi
Lina Ruiz Grajales
Jason Williams
Chun-hua Yang
Xiaoqun Zhang

### Library and Archives

Ludmila T. Pollock
    *Executive Director*
Paula Abisognio
Thomas Adams
Tara Bonet-Black
Clare Clark
Matthew Dunn, Ph.D.
Adrian Gomez, Ph.D.
Alec Micklos
Elizabeth Pessala, J.D.
Stephanie Satalino
Gail Sherman
Alistair Sponsel, Ph.D.

### Meetings & Courses Program

David J. Stewart, Ph.D.
    *Executive Director*
Alfred Baman
Edward Campodonico
Cathleen Carr
Joseph Carrieri
Meredith Cassuto
Betty Cavallaro
Reginald Cayo

Katiria Ceballos
Maoyen Chi, Ph.D.
Wendy Crowley
Matthew Denninger
William Dickerson
Alicia Franco
Bradley Frey
Constance Hallaran
Merrill Hughes
Patricia Klieger
Charla Lambert, Ph.D.
    *CSHL Diversity, Equity, and
    Inclusion Officer*
Rachel Lopez
Margaret Malone
Dennis McGuire
Andrew Mendelsohn
Maureen Morrow
Andrea Newell
Kenneth Orff
Valerie Pakaluk
Jonathan Parsons
Ira Russo
Shannon Silva
Mary Smith
Erica Stang
James Whitaker
Barbara Zane

### Office of the President

Terri I. Grodzicker, Ph.D.
    *Dean of Academic Affairs/
    Editor,* Genes & Development
Jan A. Witkowski, Ph.D.
    *WSBS Professor*

### Watson School of Biological Sciences

Alexander Gann, Ph.D.
    Lita Annenberg Hazen *Dean*
Kimberly Creteur
Kimberley Graham
Alyson Kass-Eisler, Ph.D.
Monn Monn Myat, Ph.D.

## ADMINISTRATIVE DEPARTMENTS

### Advancement

Charles V. Prizzi
*Vice President, Advancement and
    Special Advisor to President*
Jane Donaldson, Ph.D.
Diane Fagiola
Stephanie Garrett
Michael Gurtowski
Bridgette Johnson
Lital Lynn, Ph.D.

Karen Orzel
Tara Ruocco
Jill Stone

**Business Development and Technology Transfer**

Andrew Whiteley
*Vice President*
Vladimir Drozdoff, J.D., Ph.D.
Elizabeth Hand
Cynthia Miller
Radhakrishnan Narayanan, Ph.D.

**Central Administration**

Carol Gundersen
Carol Link
Elizabeth Panagot
Denise Roberts, Ph.D.

**Culinary Services**

James Hope
*Director*
Alan Aguirre
Jeannette Amato
Eliseo Amaya
Ingrid Amaya
Francis Bowdren
Maryam Brinkwart
Elmer Canales
William Carmona
Dilicia Castro Najera
Francis Cayamcela
James Charvis
Marvin Chavez
Rene Chavez Contreras
Melissa Davis
Christina DeDora
Erica Douglas
Suzanne Esposito
Saira Estrella
Bartola Garcia
Luis Garcia
Andres Gonzalez
Silvia Gonzalez
Jennifer Gordon
Bartholomew Isaac
Coleen Jackman
Kamal Johnson
Elizabeth Kirsch
Gabrielle Korwan
Deborah Koslosky
Pamela Lancellotti
Jennifer LoSordo
Juan Magana
Jose Maradiaga Gonzalez
Ann May
Thomas McIlvaine
Frederik Munter

Adalia Pardo
Joel Pena Corvera
Jose Reyes
Jesus Rivera
Susana Roman
Marlenis Romero
Carol Rupp
Robert Scherer
Karen Scola
Lisa Spero
Jose Suarez
Selma Turk
Jose Velasquez
Hector Vigil Garcia
Maria Villatoro
Maria Vilorio
Alexandros Vogiatzis
Tracy Woolfson
Lidia Zuleta

**Facilities**

Stephen Monez
*Vice President, Chief Facilities Officer*
Robert Albinson
Leslie Allen
Peter Anderson
Lucas Arevalo
Kyle Barnard
William Bishop
Caitlin Carroll
Dessie Carter
Daniel Chapman
Anthony Ciccarella
Christopher Clark
Robert Collins
Juan Colocho
Drew Comer
Timothy Cotone
Peter Dale
John Damico
Nicholas Dassau
Robert Dickerson
Glen DiMaria
Kevin Donohue
Sarah Drayzen
Wesley Dreusike
Paul Edwards
Elias Eifert
Rafael Estrella
Michael Forgione
Vito Gassi
Robert Gensel
Erick Greene
Leo Greene
Stephen Gregorovic
Philip Grella, Jr.
Wilman Gutierrez

Wayne Hamilton
Jameel Hodges
Gerard Holler
Jason Holzman
Adam Hopper
Joseph Houser
Matthew Hoyer
Christopher Hubert
Louis Hunter
Michael Hutchinson
Randal Jones
Teri Jurkschat
Daniel Jusino
Jeffrey Klaverweiden
Robert Kopetic, Jr.
Paulo Krizanovski
Oscar Lastra
Joseph Leonard
Walter Leute
Zaharia Likoka
Jian Lin
Charles Mandell
Gregory Marotta
Patricia McAdams
Tahid McCoy
Christopher McEvoy
Kevin McHugh
Daniel Miller
Susan Montalvo
Oscar Moreira
Guido Mosquera
Mary Muno
Michael Murphy
Kenneth Nastri
Herbert Neazer
Michael Nolasco
Wayne Pav
Christopher Pearson II
Jose Pena Corvera
Saul Pena Corvera
Sandro Peralta Zavala
John Pisciotta
Manuel Ramones
Wilson Ramones
Steven Randall
Elisha Rendal Sestoso
Jaqueline Reynolds
Alfredo Rivera, Jr.
Ronald Romani
Frank Russo
Richard Rutherford
Jeffrey Sanchez
Eric Sappell
Steven Sardone
John F. Schindler
John L. Schindler
Claudia Schmid
Lauren Schmidt
Bethany Scola

Cesar Sisalima
Daniel Stahl
Raymond Stewart
John Sundman
Steven Tang
Hock Yew Teo
Leigh Tyrrel
James Van Wie
Fredy Vasquez
Trinidad Velasquez
Wilfredo Velasquez
Benjamin Veneable III
Kyle Welton
Leslie Wenzel
Tracey Anne White
Graham Wildt
Daniel Zimmer

**Finance and Accounting**

Lari C. Russo
*Chief Financial Officer*
Sharon Bense
Elise Best
Karine Boyce
Gabrielle Conde
Flora De Leon
Damian Desiderio
Maryanne Detweiler
Karen Filasky
Mary Ellen Goldstein
Karen Heathwood
Evelyn Imbriani
Winston Li
Mary Mastropaolo
Daniel McClure
Alison McDermott
Denise Miller
Kristine Murphy
Patricia Penner
Nicole Povolny
Nicole Solomon
Patricia Urena

**Human Resources**

Katherine G. Raftery
*Vice President, Chief Human Resources Officer*
Andres Alarcon
Wendy Alexander
Lynn Carmen
Camille Cava-Belluscio
Patricia Crawford
Sandra Dunne
Jessica Heckel
Lesley Inglis
Laura Magri
Laura Moran
Elsy Padro

Victoria Panebianco
Victoria Roberts
Patricia Schoeffler
Samantha Walbridge

## Information Technology

Hans Erik Aronson
   *Vice President, Chief
   Information Officer*
Jean Bernard Agenor
Alex Barton
Walter Benka III
Paul Braun
Frank Falco
Todd Heywood, Ph.D.
Thomas Karow
Thomas Keller
Leila Kelly
Sean Kelly
Matthew Lindsey
Michael Malave
Louis Malfi
Leah Noreman
Barbara Peters
Vincent Reale
Steven Ruggiero
Salvatore Serafino
Michael Skuthan
David Jonathan Castillo Torres
Jeffrey Wilkens

## Legal

Debra Arenare, J.D.
   *Vice President, General Counsel*
Alexander Sosnovski

## Procurement

Michael Marchesiello
   *Vice President*
Deborah Aufiero
Kimberly Bronson
Susan De Angelo
Jeffrey DuPree
Elizabeth Janow
Anne Knoch
Christopher Oravitz

Barbara Purcell
Jorge Ramirez
Krystyna Rzonca
Gerard Theophilidis

## Public Affairs

Dagnia Zeidlickis
   *Vice President, Communications*
Margot Bennett
Christopher Fergo
Jessica Giordano
Philip Renna
Sara Roncero-Menendez
Susan Runkowski
Brian Stallard

## Research Operations

Diane Esposito, Ph.D.
   *Director of Research Compliance
   /Research Investigator*
Robert Gerdes
Jaclyn Jansen, Ph.D.
Laura Lynn
Joan O'Connor
Priya Sridevi, Ph.D.
Julie Sutherland

## Sponsored Programs

Walter L. Goldschmidts,
   Ph.D.
   *Vice President, Executive
   Director of Sponsored
   Programs*
Carol Amella
Irina Armstrong
Carol DuPree
David Garner
Donna Gavin
William Hamilton
Jill Hemish, Ph.D.
Joanne Janelli
Philip King
Ruta Ledins
Josephine Marrali, J.D.
Cynthia McCormack

Dave Neale
Catherine Perdikoylis
Jaclyn Tenaglia

## RESEARCH STAFF DEPARTURES DURING 2019

### Professor

Robert Maki, M.D., Ph.D.

### Associate Professor

Arne Stenlund, Ph.D.

### Research Investigator

Christopher Faehnle, Ph.D.

### Adjunct Professor

Scott Powers, Ph.D.

### Visiting Scientists/ Collaborative Scientists

Mostafa Abdelhamid, Ph.D.
Dannielle Engle, Ph.D.
Robert Henley, Ph.D.
Bingxing Huo, Ph.D.
Carolina Navas, Ph.D.
Jean-Paul Noel, Ph.D.
Anirban Paul, Ph.D.
Hyun Jae Pi, Ph.D.
Heather Read, Ph.D.
Stanley Soroka, Ph.D.
Goran Tomic, Ph.D.
Tajinder Ubhi, Ph.D.
Charles Vanderburg, Ph.D.
Shouhui Yang, Ph.D.
Toyoki Yoshimoto, Ph.D.
Honghe Zhang, Ph.D.
Yue Zhao, Ph.D.

### Visiting Clinical

Philip Gilbo, M.D.

### Postdoctoral Fellows

Jean Albrengues, Ph.D.
Sara Ballouz, Ph.D.

Leah Banks, Ph.D.
Prasamit Baruah, Ph.D.
Yael Berstein, Ph.D.
Giulia Biffi, Ph.D.
Amit Blumberg, Ph.D.
Irene Casanova Salas, Ph.D.
Hannes Claeys, Ph.D.
Chand Parvez Danka
   Mohammed, Ph.D.
Edgar Demesa Arevalo, Ph.D.
Ahmed Elewa, Ph.D.
Hardeep Gumber, Ph.D.
Jaynee Hart, Ph.D.
Jianping He, Ph.D.
Yifei Huang, Ph.D.
Anbalagan Jaganathan, Ph.D.
Yinping Jiao, Ph.D.
James Jung, Ph.D.
Andreas Leonhardt, M.D.
Qian Li, Ph.D.
Jayon Lihm, Ph.D.
Wai Kit Ma, Ph.D.
Farzaneh Najafi, Ph.D.
Ellen Louise Noren Lindback,
   Ph.D.
Sapan Patel, Ph.D.
Nikolay Rozhkov, Ph.D.
Subramanian Sankaranarayanan,
   Ph.D.
Christine Scaduto, Ph.D.
Harrison Seidner, Ph.D.
Sebastian Soyk, Ph.D.
Yilin Tai, Ph.D.
Yusuke Tarumoto, Ph.D.
Roger Tseng, Ph.D.
Qingyu Wu, Ph.D.
Fang Xu, Ph.D.

## Computational Science

Mihail Bota, Ph.D.
Daniel Ferrante, Ph.D.
Matthew Moss
Manthan Shah
Joshua Stein, Ph.D.
Kannan Umadevi Venkataraju