
CSH Cold Spring Harbor Laboratory
2021 ANNUAL REPORT


CSH
Cold Spring Harbor Laboratory
2021 ANNUAL REPORT

ANNUAL REPORT 2021

Cold Spring Harbor Laboratory
One Bungtown Road
Cold Spring Harbor, New York 11724

www.cshl.edu

Managing Editors Margot Bennett, Philip Renna, Dagnia Zeidlickis
Production Editor Kathleen Bubbeo
Copy Editors John Schneider, Dorothy Brown, Diane Schubach
Proofreader Joy Jones
Production Manager Denise Weiss
Nonscientific Photography Constance Brukin, Len Marks Photography
Cover Designer Margot Bennett

Front Cover Photography: © 2021 Peter Aaron/OTTO

Contents

Officers of the Corporation

Marilyn H. Simons, Ph.D., *Chair*
Charles I. Cogut, *Vice-Chair*
Robert D. Lindsay, *Vice-Chair*
Paul J. Taubman, *Vice-Chair*
Elizabeth McCaul, *Treasurer*
Robert W. Lourie, *Secretary*
Bruce W. Stillman, Ph.D., *President and CEO*
John P. Tuke, *Chief Operating Officer*

Board of Trustees

Life Trustee

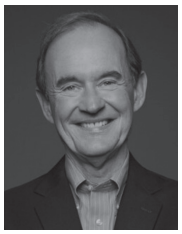
Jamie C. Nicholls
Chair Emeritus, New York, NY

Individual Trustees

Christine Anderson
Senior Managing Director, Global Public Affairs and Marketing, Blackstone

Lalit R. Bahl, Ph.D.
Senior Research Scientist, Renaissance Technologies Corp.

Joanne Berger-Sweeney, Ph.D.
President, Professor in Neuroscience, Trinity College



David Boies
Chairman, Boies, Schiller & Flexner LLP

Michael R. Botchan, Ph.D.
Richard and Rhoda Goldman Distinguished Professor of Biochemistry, University of California, Berkeley

Elizabeth Cogan Fascitelli
Goldman, Sachs & Co. (Retired Partner)

Charles I. Cogut
Retired Partner, Simpson Thacher & Bartlett LLP

Elaine Fuchs, Ph.D.
Rebecca C. Lancefield Professor/HHMI Investigator, Rockefeller University/Howard Hughes Medical Institute

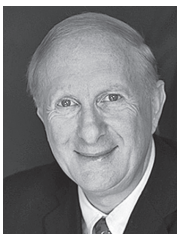
Jeffrey E. Kelter
Chief Executive Officer, KSH Capital; Chairman, Jack Creek Investment Corp.

Laurie J. Landeau, V.M.D.
General Manager, Listowel, Inc.

Robert D. Lindsay
Co-Managing Partner, Lindsay Goldberg, LLC

Robert W. Lourie, Ph.D.
Head of Future Research, Renaissance Technologies LLC

Elizabeth McCaul
ECB Representative, Supervisory Board, European Central Bank



Howard L. Morgan, Ph.D.
Chairman, B Capital Group, Managing Member, MFCIF, LLC

Jeanne Moutoussamy-Ashe
New York, NY

Lyon Polk
Founder, The Polk Wealth Management Group at Morgan Stanley

Bruce Ratner
New York, NY

Geoffrey Robertson
Director of Business Assistance, Vermont Sustainable Jobs Fund

Individual Trustees *(continued)*

Thomas A. Saunders III
CEO, Ivor & Co., LLC

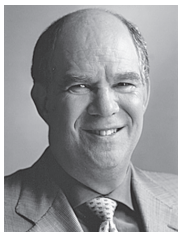
Douglas Schloss
President and CEO, Rexford Management

Marilyn H. Simons, Ph.D.
President, MJS Foundation; Co-chair, Simons Foundation

Laura Slatkin
Co-Founder and Board Chair, NEXT for AUTISM

Kristin Olson Smith
Managing Director, Alternative Investments, Goldman, Sachs & Co.

Bruce W. Stillman, Ph.D.
President and Chief Executive Officer, Cold Spring Harbor Laboratory



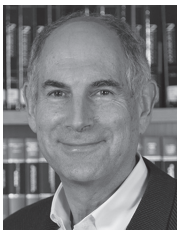
James M. Stone, Ph.D.
Chairman, The Plymouth Rock Company

Karel Svoboda, Ph.D.
Vice President and Executive Director, Allen Institute for Neural Dynamics

Paul J. Taubman
Chairman and CEO, PJT Partners Inc.

Diana L. Taylor
New York, NY

Edward Travaglianti
President, TD Bank Long Island



Stuart Weisbrod, Ph.D.
Weisbrod Family Office, LLC

George D. Yancopoulos, M.D., Ph.D.
Chief Scientific Officer & President, Regeneron Pharmaceuticals, Inc.

Roy J. Zuckerberg
Senior Director, Goldman Sachs & Co.

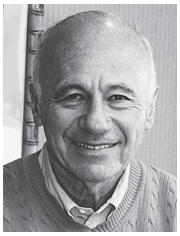
Honorary Trustees

Bayard Clarkson, M.D.
Chair Emeritus, Memorial Sloan Kettering Cancer Center

Charles F. Dolan
Oyster Bay, NY

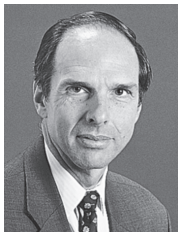
Helen Ann Dolan
Oyster Bay, NY

Lola N. Grace
Founder, Middle East Children's Institute



Leo A. Guthart
Founder and CEO, Topspin Partners

Nancy Abeles Marks
Nancy Marks Interiors and Carl Marks & Co., Inc.

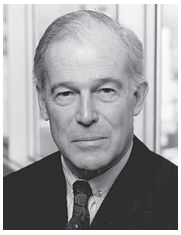


Eduardo G. Mestre
Chair Emeritus, Vice Chairman, Evercore Partners

William S. Robertson
Chairman, Robertson Foundation for Government

Alan Seligson
NAK International

James H. Simons, Ph.D.
Chairman, Simons Foundation



Henry Wendt III
Friday Harbor, WA

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 includes 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 members are world leaders in their respective fields and as such are able to provide advice on the different research areas of the Laboratory.

Governance

The Laboratory is governed by a Board of Trustees of up to 40 members that meets at least three times a year. Authority to act for the Board of Trustees between meetings is vested in the Executive Committee of the Board. 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 includes business and community leaders and scientists from major educational and research institutions.

The Laboratory is chartered as an educational corporation by the Board of Regents for and on behalf 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, School of Biological Sciences" and to confer the degrees of Doctor of Philosophy (Ph.D.), Master of Science (M.S.), and Doctor of Science (Sc.D.), Honorary.

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

Committees of the Board

Executive

Marilyn Simons, Chair
Casey Cogut
Robert Lourie
Elizabeth McCaul
Howard Morgan
Jamie Nicholls
Bruce Stillman
James Stone
Paul Taubman
Roy Zuckerberg

Advancement

Howard Morgan, Chair
Christine Anderson
David Boies
Casey Cogut
Jeffrey Kelter
Laurie Landeau
Robert Lourie
Charlie Prizzi
Laura Slatkin
Bruce Stillman
Paul Taubman
Diana Taylor
John Tuke
Roy Zuckerberg

Commercial Relations

James Stone, Chair
Casey Cogut
Robert Martienssen
Bruce Stillman
John Tuke
Christopher Vakoc
Stuart Weisbrod
Andrew Whiteley

DNA Learning Center

Laurie Landeau, Chair
Casey Cogut
MaryAnne Gilmartin
Yolanda Lyle
Howard Morgan
Jeanne Moutoussamy-Ashe
Douglas Schloss
Marilyn Simons

Audit and Risk

Douglas Schloss, Chair
Lalit Bahl
Elizabeth Cogan Fascitelli
Leo Guthart (Honorary)
Elizabeth McCaul

Investment

Jamie Nicholls, Chair
Alan Breed
Elizabeth Cogan Fascitelli
Lyon Polk
Thomas Purcell
Boaz Sidikaro
Kristin Olson Smith
John Tuke

Facilities

Jeffrey Kelter, Chair
Elizabeth Ainslie
Hans Bosch
Elizabeth Cogan Fascitelli
Elaine Fuchs
Steven Monez
Frank Sciame
Bruce Stillman
John Tuke

Finance

Elizabeth McCaul, Chair
Elizabeth Cogan Fascitelli
Jamie Nicholls
Lari Russo
Douglas Schloss
Marilyn Simons
John Tuke

Academic Affairs

James Stone, Chair
Lalit Bahl
Michael Botchan
Elaine Fuchs
Jacob Goldfield
Laurie Landeau
Robert Lourie
Bruce Stillman
Karel Svoboda

Business Development

Frank Brosens
Bryan (Hunt) Lawrence
William Spire
James Stone
Paul Taubman
Stuart Weisbrod
Andrew Whiteley

Nominating

Paul Taubman, Chair
Jeffrey Kelter
Robert Lindsay
Elizabeth McCaul
Bruce Ratner
Thomas Saunders
Douglas Schloss
Marilyn Simons
Laura Slatkin
Bruce Stillman
George Yancopoulos
Roy Zuckerberg

Robertson Research Fund (Board of Trustees)

Bruce Stillman (President)
Geoffrey Robertson (Vice President)
Walter Goldschmidts
William Gridley
Leo Guthart
Andrew McKibben
Walter Meier
Lyon Polk
Marilyn Simons

Robertson Research Fund, Inc. Investment Committee

William Gridley
Jamie Nicholls
Lyon Polk
Geoffrey Robertson
John Tuke

Compensation Review and Advisory

Marilyn Simons, Chair
Jamie Nicholls
Douglas Schloss
Diana Taylor

Alan Seligson (1930–2021)

Alan Seligson, who died on June 17, 2021, was a longtime friend and supporter of Cold Spring Harbor Laboratory. He served on the Board of Trustees from 2005 to 2013 and then was elected an Honorary Trustee. Alan and Edith, his wife, first became associated with the Laboratory in 1988 when their son, Andrew, was diagnosed with neuroblastoma, at age 17. The Seligsons' neighbor was Arthur Merrill, a long-standing member and director of the Long Island Biological Association. Merrill introduced the Seligsons to Jim Watson and to the Laboratory's research-intensive programs on cancer. Tragically, nothing could be done for Andrew, who died eight years after the diagnosis had been made.

Those who know the Seligsons will appreciate that their response to this devastating loss was a determination to do all they could to foster research into the causes of cancer that sometime in the future would lead to effective therapies. They decided to promote the work of the bright young scientists who come to CSHL early in their career, and they established the Andrew Seligson Memorial Fellowships to support postdoctoral fellows. This later developed into the CSHL Clinical Fellows Program, a new initiative at CSHL that attracted talented clinical oncologists to spend time working in a cancer research laboratory before embarking on a career in cancer research and treatment. In 2008, Alan and Edith made a most generous gift for a laboratory dedicated to cancer research, named for Andrew. Most recently, the Seligsons ensured that the Laboratory will be able to continue cancer research at the highest level by endowing the Edith and Alan Seligson Professor of Cancer Research. The inaugural recipient of the professorship is Chris Vakoc, M.D., Ph.D., whose work focuses on identifying key targets for cancer therapy and translating these discoveries into the clinic.

Alan and Edith have made many other contributions to CSHL, particularly the Double Helix Medal event, the Women's Partnership for Science Program, and, not surprisingly given Alan's passion for playing golf, to the annual DNA Learning Center Golf Tournament.

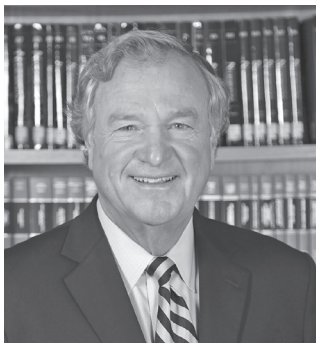
Alan was born on June 6, 1930, on the south shore of Long Island. He joined the Navy, serving on the USS Glynn in the Korean War, achieving the rank of Lieutenant Commander and retiring as a full commander. On leaving the Service, Alan met and married Edith Marks, and three children—Nancy, Andrew, and Kate—were born. Both Nancy and Kate also became associated with the Laboratory, and Kate served as director of the Cold Spring Harbor Laboratory Association for four years.

Alan founded his own company, NF&M International, which specialized in the highest-quality titanium products, primarily for the aerospace industry. The company flourished, becoming one of the leading manufacturers in that market. But it was not all work. Quite apart from his golf,

Alan was a lover of fine art, and he and Edith assembled an enviable collection of modern art and sculpture. He was a Trustee of the Nassau County Museum of Art and served as President of the Board.

Alan remained active on the Laboratory's Board and avidly followed the progress of cancer research at institutions worldwide as well as at CSHL until his last days. On a personal level, Alan was very engaging, kind, and thoughtful, with a wonderful sense of humor. The entire Seligson family are close friends with my wife, Grace, and me, and we very much enjoyed visiting their wonderful home in the Hamptons and their home in Locust Valley. Alan will be missed by all CSHL staff who knew him.

Bruce Stillman
Cold Spring Harbor Laboratory



Edward Travaglianti (1949–2021)

Edward (Ed) Travaglianti's career was spent in banking and finance, and it was his financial acumen and expertise, his networking, and his dedication to Cold Spring Harbor Laboratory that made him such a valuable member of our Board of Trustees. He died on August 24, 2021.

Ed was born in Bensonhurst, Brooklyn, in 1949. He attended St. Francis College, Brooklyn, graduating in 1970. He joined Franklin National Bank to begin a training program, but the bank was acquired by the European American Bank (EAB) in 1974. By 1991, Ed had risen to be Chairman and CEO of EAB. Ten years later, EAB in its turn was acquired by Citibank and Ed was appointed President of Citibank's division of Commercial Markets. Ed left Citibank in 2002 and moved to Commerce Bank in 2004. This was taken over by the Canadian-based Toronto-Dominion Bank, and Ed was appointed President of TD Bank Long Island, a post that he held until his death. Because of his senior position in Long Island banks, Ed had a remarkable network of business contacts, and he was a trusted advisor to philanthropic organizations, particularly those that were family-based.

As his son, Edward Travaglianti, Jr., has said of his father: "He always had a very philanthropic heart," a sense of community service clearly instilled by Ed's parents. He spoke of them as saying that "…everything we would do going forward, had to be anchored by ethical, moral, fair and equitable behavior. The axiom of treating others as you would like to be treated was a baseline." Indeed this philosophy defined who Ed was as a person.

Ed joined the CSHL Board of Trustees in 2004 and was immediately put to work serving on three key committees: the Executive, Finance, and Investment Committees. He was a member of these committees for 17 years. In 2011 he joined the Audit Committee and was still serving on all four committees in 2021. Ed was also much involved with the DNALC, serving on its Corporate Advisory Board as well as the Trustees' DNALC Committee. He and his wife were regular contributors to the Laboratory's Double Helix Medal event.

Through his extensive connections on Long Island, Ed and CSHL supporter Robert Vizza introduced CSHL to the DeMatteis Family Foundation, which supported the building of a cancer research building on the Hillside section of our campus. Ed was also instrumental in introducing us to the Don Monti Memorial Research Foundation, which was supporting cancer patient care at what has now become Northwell Health. With support from Joseph and Tita Monti and their daughter, Caroline Monti Saladino, the family supported a major cancer research program aimed at identifying new targets for cancer therapy. This program has already resulted in two drugs that are in clinical trials for leukemia or lung cancer. Thus, Ed's advocacy for CSHL as a Trustee has had a broad impact.

Ed's contributions to the broader Long Island community were far wider than just CSHL and he was a board member of NYU Winthrop Hospital (now NYU Langone Hospital—Long Island), the Tomorrow's Hope Foundation, and Long Island University. Ed was a past Chairman of the Long Island Association, the nonprofit and nonpartisan business organization in the region. He served as Executive Vice President of the Don Monti Memorial Research Foundation, which supports cancer research.

Ed was dedicated to supporting and promoting the Laboratory. It was always delightful to see the pleasure he took in participating in the Convocations of the School of Biological Sciences, mixing with the students about to receive their Ph.D. degree. We were pleased to name a seminar room in the recently renovated Demerec Laboratory after Ed.

Ed was very engaged in the life of CSHL, but importantly he was also a wonderful friend, along with his wife Pat, to Grace and me. He provided valuable advice and even on occasion addressed employees at certain times to convey the support of the Board of Trustees for important decisions and new programs. He is missed by all who knew him.

Bruce Stillman
Cold Spring Harbor Laboratory

PRESIDENT'S REPORT

This essay highlights the diverse programs of Cold Spring Harbor Laboratory (CSHL), revealing how seemingly separate initiatives strategically intertwine to make this institution central to the future of science. I believe that the interconnectivity of our research and education programs is CSHL's greatest strength. Combined, they enhance the impact of science conducted on our Long Island campus and, with altruistic purpose, advance life science throughout the world.

CSHL is a unique research and education institution in both form and function. It supports extraordinary in-house research, trains visiting scientists, publishes resources used worldwide, and serves as the top meeting place for biologists. The physical attributes of the picturesque campus attract the world's scientists here, but the real draw is deeper. The human interactions catalyzed by the research and education programs make CSHL the absolute center of contemporary life science.

From this core, we promote and advance science for a better world—and, of course, there is more to discover. We are planning new research and education facilities that are critical foundations for our future.

> *"The interconnectivity of our research and education programs is CSHL's greatest strength."*

Connectivity in Research

At CSHL, scientists mingle and share ideas on a campus deliberately designed to encourage interactions. In a village-like atmosphere, in-house scientists work in small-scale buildings—which means people have to leave their own workplace often. Shared eating spaces and beautiful grounds contribute to a culture that promotes thought and interaction. But we are not resting on our laurels. There is a growing need for new space to accommodate scientists in exciting new areas of research.

Our decentralized architecture and institutional structure reflect a defining organizational feature of CSHL. We have no departments, so scientists working in different fields readily learn from and collaborate with one another. It is easy to shift directions and explore new areas. That kind of flexibility is often challenging in traditionally structured academic institutions.

We also foster collaboration through our hiring strategy. We seek out promising early-career scientists in the fields of science we think are important for the future. At the same time, we make sure their interests overlap with those of existing faculty. This connectivity is particularly important when we are expanding into new research areas.

One example is computational neuroscience, which was a nascent field when I appointed our first computational neuroscientists in the late 1990s. To ease into this field, we recruited Tony Zador, who had a background in both computational and experimental neuroscience. Tony provided a bridge to the pure computational biologists who would follow later. His recruitment was facilitated by advice from a few prominent neuroscientists I knew because they had attended meetings and courses at CSHL. Today, CSHL computational neuroscientists Alex Koulakov and Tatiana Engel are leaders in a field that has become an essential component of any advanced experimental neuroscience program.

Inevitably, Tony's research program was shaped by unanticipated interactions with his new CSHL colleagues, including those working on cancer. Inspired by advanced genetic screening research within our Cancer Center, Tony developed an efficient and powerful DNA barcoding method for mapping the connections between the billions of neurons in the brain. His research on

neuronal connectivity is now a new bridge between neuroscience and artificial intelligence (AI). Indeed, we have recently recruited Ben Cowley, who employs AI methods to understand how the brain computes external information. This is the start of a new CSHL NeuroAI Program that will be facilitated by the campus expansion plan.

The Crossroads of Biology

I often hear from colleagues that the first touchpoints scientists have with CSHL are the books, journals, and technical manuals published by CSHL Press. These world-renowned resources explain the techniques and methods scientists use in research. More recently, CSHL founded the widely used preprint servers for biology and medicine, bioRxiv and medRxiv. The preprint servers are digital platforms for disseminating research before it is published in peer-reviewed journals.

After posting more than 21,500 COVID-19–related articles by the end of 2021, bioRxiv and medRxiv are recognized as a critical part of the world's response to the pandemic, informing scientists, journalists, and policy makers about the latest research.

The pandemic interrupted our ability to host the 9,000 scientists who usually visit the CSHL main campus and the Banbury Center for meetings and courses. Historically, scientists attend meetings here to listen and learn from each other and find collaborators. Spending anywhere from a few days to several weeks in residence, informal interactions among scientists can be just as impactful as the scheduled talks. Indeed, the Blackford Bar is famous among meeting attendees. Many CSHL faculty can readily share a story about an experiment that was inspired by a conversation had at the bar.

During the pandemic, our auditorium and poster pavilion were empty, as were our picnic tables, cafeterias, and bar. But although people could not come to campus, their avid attention to preprint servers and virtual attendance at online meetings revealed a large unmet need for ways to exchange scientific ideas. The virtual meetings attracted far more participants than we would have had the capacity to bring together in our current facilities. Now, post-pandemic, we are hosting hybrid meetings, but there is a clear need to add and improve on-campus housing for our visitors who are eager to return for the on-site experience. This is key to our campus expansion plan.

Exchange and Outreach

By training course attendees in emerging and advanced technologies, we are broadly accelerating scientific progress across the world. Visiting course participants apply what they have learned while at CSHL in their own laboratories. I am struck that 11 former course participants went on to receive Nobel Prizes in fields they studied here.

The exchange goes both ways. When visiting scientists teach or participate in courses, new techniques are brought directly to this campus. Our in-house scientists can rapidly deploy the latest techniques in their laboratories. Moreover, some meetings visitors return later to become integral parts of our campus community. I first visited CSHL as a second-year graduate student to speak at the 1978 CSHL *Symposium on DNA Replication and Recombination*, returning nine months later to begin my independent research career.

Scientists benefit from discussing current findings and exchanging lessons from the past. After more than 130 years of helping shape the direction of biomedical research and assembling an exceptionally rich collection of archival materials, CSHL educates scientists and the public about the history of the life sciences. Our Archives include primary records, personal items, and oral-history interviews, many of which have been digitized.

The Center for Humanities and the History of Modern Biology, started by Mila Pollock, organizes meetings on the origins of major discoveries in biology and medicine. In 2021, we

hosted a virtual event featuring the research of three prominent, Nobel Prize–winning women: Elizabeth Blackburn, Jennifer Doudna, and Dorothy Crowfoot Hodgkin, presented by her biographer Georgina Ferry. These three outstanding scientists have inspired the current generation of early-career scientists.

Since 1988, through outreach programs to the public, we have invested in building a pipeline of aspiring scientists from middle school on up. Serving middle, high school, and undergraduate teachers and students, the DNA Learning Center (DNALC) is a key part of the Laboratory's laudable cycle between research and education. The DNALC provides opportunities for our scientists to incorporate public outreach and education into their research programs. Importantly, the program educates students and their families about new advances in genetic research that may impact their health and well-being. Starting on Long Island, we have expanded our New York footprint to include Harlem and Westchester; just this year we opened DNALC NYC in Brooklyn. We hope to inspire more students to pursue a life of scientific discovery.

"CSHL is an inclusive place for all, both in-person and now online. We come together to listen to and learn from each other, find collaborators, and conduct research."

Hope Sparks New Fields

Through our convening power and programs that facilitate exchange, CSHL has also helped launch entirely new fields of inquiry. Complementing our larger scientific conferences, the Banbury Center hosts small scientific meetings with an outsized impact on society. A 1988 meeting on *DNA Technology and Forensic Science* led to the founding of the Innocence Project, an organization that works to exonerate incarcerated people who have been wrongly convicted of crimes. A 1994 meeting helped persuade the National Science Foundation to fund the first complete sequencing of a plant genome—an international effort that was led in part by CSHL scientists Rob Martienssen and W. Richard McCombie. More recently, a series of meetings on Lyme disease led to improved diagnostics and spurred efforts to develop an effective vaccine.

Ideas that emerge at the Banbury Center cycle back and spark new research directions at CSHL. A good example is the 2014 Banbury meeting on rhabdomyosarcoma, a rare cancer mostly afflicting children and young adults. The meeting was funded by three local foundations, including one established by CSHL's Philip Renna (who lost his teenage daughter to the disease in 2007).

At the time, only a few researchers around the world were studying this often-fatal disease. The Banbury meeting brought clinicians and researchers together to define the most pressing questions in sarcoma biology. They mapped out a discovery process that could lead to effective therapies. Soon after, we began the sarcoma initiative at CSHL, led by cancer biologist Chris Vakoc. Initially funded by the local foundations, the initiative has attracted additional funding, and Vakoc's team has already identified promising drug targets. The Business Development and Technology Transfer team has found commercial partners to pursue potential therapeutics that are now in preclinical development.

The Future of Science Starts Here

CSHL is a destination for scientists. They come at different points in their careers, as students, trainees, faculty, and leaders. CSHL is an inclusive place for all, both in-person and now online. We come together to listen to and learn from each other, find collaborators, and conduct research.

We've created this unique environment so scientists can advance research that will benefit society, both economically and through improved health and well-being. Cycles of research and education interact to build CSHL and the ever-advancing fields of the life and medical sciences. I am excited to work with our scientists, educators, and supporters to plan a campus expansion that builds strong foundations for the future. Most of all, I hope to see the picnic tables and the bar full of life again in 2022.

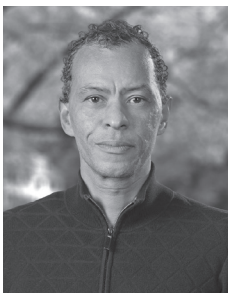
Bruce Stillman, Ph.D., F.R.S.
President and Chief Executive Officer

Highlights of the Year

Research

In 2021, hundreds of scientists working in nearly 60 Cold Spring Harbor Laboratory (CSHL) research groups published their findings in the world's major research journals. Their efforts reflect the full spectrum of CSHL's programs in Cancer, Neuroscience, Plant Biology, Quantitative Biology, Genomics, and a new chemistry lab. The following is a sampling of this year's important findings.

Pinpointing Cancer's Origins



L. Trotman

CSHL Adjunct Professor Pavel Osten and Professor Lloyd Trotman have pioneered a new method to track how prostate cancer progresses in mice, from its birth to its spread into other tissues. The pair combined their expertise in whole-organ imaging and prostate cancer to track how a single cancer cell can grow into a tumor and spread to other organs. For the first time, researchers can study a tumor as it grows in a setting that accurately mimics the disease in real life. The study was led by Julian Taranda, a former postdoc in the Osten laboratory, and was published in *Cell Reports*.

The researchers used a virus to transform as little as one normal mouse prostate cell into a cancerous cell. This lone cancer cell was located using a microscope technique called whole-organ serial two-photon tomography. The tomography machine is fully automated. It takes an image of all the cells on the top layer of an organ, slices off that piece, images what is in the next layer, and repeats the process until it has photographed the entire organ. Then, using artificial intelligence, a computer creates a 3D reconstruction of the organ at single-cell resolution. The scientists hope this versatile new method will help tackle unexplored questions about the early steps of cancer's growth and escape into other organs, wherever it starts.

Cushy Homes for Cancer

M. Egeblad

Cancer cells live in a complex neighborhood populated by immune cells, blood vessels, and other structures. Cells signal each other as to who they are and what they want. Cancer cells may hide their identity so they can grow and spread more easily. CSHL Professor Mikala Egeblad wants to know what creates an ideal microenvironment for a wandering cancer cell. She reasons that if she can decode the signals in the neighborhood, she may find ways to harness those signals to defeat cancers.

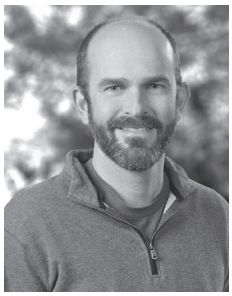
Breast and ovarian cancer cells can reprogram immune cells to make them into "tumor-associated macrophages" (TAMs), which act like a security team hired by the cancer cells to protect themselves. Egeblad's laboratory found that interferon-gamma plus an immune system activator re-reprogrammed breast cancer TAMs so they would attack the cancer. The two drugs were also extremely effective against ovarian cancer in mice. In both cancers, the treatment slowed metastases and made ovarian tumors more susceptible to chemotherapy.

Lessons learned about the cancer microenvironment can be applied to COVID-19. Neutrophils, a type of immune cell, can form sticky neutrophil extracellular traps, or NETs, which normally trap pathogens. However, too many NETs can be toxic—especially in the lungs, where they can damage tissue and cause respiratory distress. Thus neutrophils, like macrophages, could become therapeutic targets.

Cancer's Sweet Tooth

CSHL Professor Christopher Vakoc and his laboratory discovered that acute myeloid leukemia (AML) cells depend on a single transporter to get the essential sugar inositol into the cell. Cancers

streamline certain cell processes, "putting all their eggs in one basket" with a single pathway. Vakoc can then develop treatments to knock out that remaining pathway and kill the cancer cells.



C. Vakoc

Most cells either get inositol from the bloodstream (it is present in many foods) or make it themselves. Because there is plenty of sugar available outside the cell, some cancer cells decided to rely on the inositol transporter to capture it and stop making it inside the cell. If researchers can find a treatment that can turn off or block this transporter, the cancer cells would starve. This method would leave normal cells unharmed because they can make inositol on their own. Vakoc reported his findings in *Cancer Discovery*.

Vakoc says his work suggests a few roads to developing a therapeutic: "You could make an antibody that just sticks to this transporter. It doesn't need to get into the cell, and it could shut off the transport function. The other possibility, from a drug development point of view, is inositol. You could build a molecular medicine that sort of looks like inositol, but has a few chemical differences that can clog the transport function."

Predicting Cancer's Path

Assistant Professor David McCandlish and collaborators used the statistical method of density estimation in a new way: to predict how combinations of genetic mutations cause different types of tumors. McCandlish says, "This is what's fascinating about mathematical research. Sometimes you see connections between topics that seem so different, but at a mathematical level, they use the same ideas."

D. McCandlish

McCandlish mapped the combinations of mutations most likely to occur in a particular protein and in the same cancer cell. It is straightforward to predict the co-occurrence of a couple of events, like how often you might find two people of the same height in a group. But for complex biological sequences, such as the hundreds of amino acids that make up a protein, predicting the probability of each potential sequence becomes astonishingly complex. "Sometimes, with one mutation in a protein sequence, the protein works fine," explains McCandlish. "And with a second mutation, it still works fine, but then if you put the two of them together, you get a broken protein. We've been trying to come up with methods to model interactions between any number of mutations." Their new method can predict how hundreds of thousands of different mutation combinations impact the function of a protein.

The team published the study in the *Proceedings of the National Academy of Sciences* and made their density estimation software available publicly.

Estrogen Gives Mice the Moves

Female animals are most active when estrogen levels are high, increasing their chances of encountering a mate when pregnancy is most likely. Mice with low levels of estrogen are more sedentary than ones with high levels. Women also become more sedentary as estrogen levels decrease during menopause. CSHL Assistant Professor Jessica Tollkuhn and her collaborators at the University of California, San Francisco, have now traced this hormone-driven activity to a cluster of estrogen-sensitive cells in the brain.

J. Tollkuhn

Tollkuhn and her team study the profound impact estrogen has on the brain, where it not only influences activity levels, but also modulates mood, alters sleep patterns, and helps control body temperature. In the brain, the hormone latches onto an estrogen receptor (ERα), changing the activities of specific genes. Tollkuhn and colleagues have found nearly 2,000 sites within the genome that interact with ERα, suggesting the hormone regulates hundreds of different genes in the brain.

One of those genes is *Mc4r*. The team's experiments in mice revealed how the hormone provokes signaling changes inside estrogen-sensitive neurons. Importantly, they found they can mimic these effects without increasing estrogen exposure, simply by activating *Mc4r* in the relevant neurons.

The team's findings suggest it may be possible to develop targeted therapies that restore specific benefits of estrogen signaling—without the side effects of hormone replacement.

Every Brain Cell Counts

P. Osten

CSHL Adjunct Professor Pavel Osten and his laboratory mapped cells and connections within the mouse primary motor cortex. They categorized different cell types throughout the brain in a quantitative brain-wide (qBrain) catalog. With this method, the researchers can standardize a 3D map of the 100 million neurons in the mouse brain and, in the near future, even the 100 billion neurons in the human brain.

Osten's technique starts with labeling brain cells of interest to identify classes of cells or particular pathways. The brain is then preserved and imaged automatically at high resolution. Each brain is analyzed by a computer that can count the cells or trace the pathways, comparing it to previously mapped brains. The entire process, automated after the brain is preserved, takes 12 to 32 hours per mouse. The researchers then compare new brains to their standardized 3D maps to figure out gender differences within a species, development stages, and diseases. They already discovered anatomical differences between male and female mouse brains associated with behavioral differences.

Osten and his laboratory were a part of the founding group of scientists for the NIH-funded Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative–Cell Census Network (BICCN). Similar in scope to the Human Genome Project, BRAIN is providing a foundation for studying the mammalian brain with new methods.

Helpless No More

L. Van Aelst

Everyone faces stress occasionally, whether in school, at work, or during a global pandemic. CSHL Professor Linda Van Aelst studies how individuals respond to stress. Her laboratory studied the mouse gene *Ophn1*, which plays a critical role in developing brain cell connections, memories, and stress tolerance. When *Ophn1* is removed in a specific part of the brain, mice express depression-like helpless behaviors. The researchers found three agents, including a drug called fasudil, that could reverse this effect.

To test for stress responses, the researchers put mice into a two-room cage with a door in between. The floor in one room provides a light shock to their feet, and normal mice leave that room. But animals lacking *Ophn1* sit helplessly in that room without trying to escape.

Van Aelst's laboratory deleted the *Ophn1* gene in different brain regions. Removing *Ophn1* from a circuit in the prefrontal cortex known to influence behavioral responses and emotion induced the helpless phenotype. Overactivity in a particular part of the circuit was the key.

In humans, mutations in the *Ophn1* gene cause a rare X-linked disease that includes poor stress tolerance. Van Aelst hopes that understanding the complex circuit behind *Ophn1*-related stress responses will lead to better treatments for humans. The team published their results in *Neuron*.

Robots Explain Their Thoughts

Brain-like artificial networks are often referred to as a "black box" because researchers do not know how they learn and make predictions. CSHL Assistant Professor Peter Koo and his team reported new ways to peek inside the box and identify key features on which the computer program relies, particularly when trying to analyze complex RNA and DNA sequences.

P. Koo

Residual Bind is a type of AI program called a deep neural network. It predicts the ability of RNA sequences to bind to proteins. Koo and his team developed a new method, called global importance analysis, that "quizzes" this AI program to figure out what rules it learned on its own and whether they are the right ones. They discovered that the network considered more than just the spelling of a short stretch of RNA. It factored in how the RNA strand might fold over and bind to itself, how close one RNA pattern is to another, and other features.

Koo's team also reported a new way to train a type of AI network, called a convolutional neural network, to predict the function of DNA sequences. This new method allows machine learning researchers to identify some key features that lead to the computer's decision-making process.

Koo and his colleagues published their findings in *Nature Machine Intelligence* and *PLOS Computational Biology*.

Shady Communications

Shade avoidance is a vital survival strategy for plants, but it's a problem for farmers, says CSHL Assistant Professor Ullas Pedmale. When a plant finds itself in the shade, it directs its resources to reach for the light. This shade response limits the density of crop plantings and thus limits yields. Pedmale and his laboratory discovered a group of proteins called WRKYs that are responsible for stunting root growth in the shade.

U. Pedmale

Pedmale's team compared the roots of tomato and *Arabidopsis thaliana* seedlings grown in light to the shorter, less developed roots of plants grown in shade. Postdoc Daniele Rosado and colleagues found hundreds of genes that plants use to respond to stress were switched on in the shade-grown plants—including dozens that encode WRKYs.

To confirm that WRKYs limit root growth, Pedmale's team engineered plants in which specific WRKY genes were highly active in light and shade. They found that plants with high levels of certain WRKY proteins grew the same stunted roots seen in shade-grown plants, even when provided with plenty of light. The plant's stems, in contrast, grew at a normal rate.

Pedmale hopes this work will help researchers develop plants that can thrive under more crowded conditions, withstand extreme weather, and pull carbon dioxide out of the air into extensive root systems. The research was published in *Plant Physiology*.

Corn Evolution

Doreen Ware, a CSHL adjunct professor and research scientist at the U.S. Department of Agriculture, and her colleagues published the genome sequences of 26 different strains of corn in *Science*. They describe a large portion of the genetic diversity found in modern corn plants and reveal new genetic insights valuable for optimizing the crop for changing climates.

D. Ware

Like a continental landscape, genomic maps have areas that are full of features (like well-mapped cities), whereas others are more like deserts (vast and uncharted). With recent techniques, the team of scientists charted difficult stretches of the genome, even the deserts. These complete genomes allow researchers to locate and study both important crop genes and the nearby regions that regulate their use. Ware notes, "we had little access to the regulatory architecture of corn before."

The new collection reveals how the corn genome was shuffled over time. Ware says, "Different strains have experienced different environments. For example, some came from tropical environments, others experienced particular diseases, and all those selective pressures leave a footprint of that history."

Equipped with more detailed maps of the corn genome, scientists have a head start in developing crops for a rapidly changing climate. Ware explains, "The genomes provide broader insights into corn genetics, and this, in turn, can be used to start optimizing corn to grow in future environments."

Instant Polymers


J.E. Moses

A multi-institutional team of chemists, including CSHL Professor John E. Moses, Nobel laureate K. Barry Sharpless from Scripps Research, and Han Zuilhof of Wageningen University found a way to modify and use a dangerous gas called SOF_4 as building blocks for new products. In a paper in *Nature Chemistry*, they describe a new set of modifiable polymers made from SOF_4.

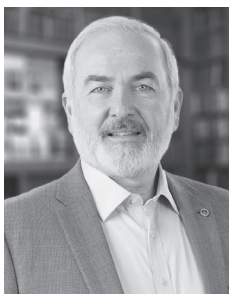
The team used a type of rapid and reliable chemistry known as click chemistry to "click" molecules together without producing toxic byproducts. The SOF_4 molecule acts as a hub to link together diverse components into a modular family of new—and potentially valuable—drugs and polymers.

The reactions are fast and produce very little waste or dirty by-products. Moses says, "That's why click chemistry is great. These polymers could be made in one day. As long as we have the gas, we could do all that chemistry in one day, make a polymer, and post-modify it in one day. That's incredibly fast."

This new chemistry will allow scientists to generate a vast new library of polymers, each with its own distinct properties and applications in drug discovery and material science. Moses says, "The opportunity for these polymers, I think, is infinite. There are so many things we can do with it. We're limited by our imagination."

Research Faculty

Awards


A. Krainer

Professor Adrian Krainer was awarded the 2021 Wolf Prize in Medicine for his contributions in biochemistry and molecular genetics and his fundamental mechanistic discoveries regarding RNA splicing. Awarded by the Wolf Foundation and the President of the State of Israel, the prize honors people for advancing science and art for humanity, for friendship between peoples, and for contributing to the creation of a better world.

Krainer is best known for his work on RNA splicing and the development of Spinraza®, the first FDA-approved treatment for spinal muscular atrophy (SMA). SMA is a neurodegenerative disease that is the leading genetic cause of infant death. For this research, Adrian was awarded the Jacob and Louise Gabbay Award in Biotechnology and Medicine, which recognizes scientists in academia, medicine, or industry whose work had outstanding scientific content and significant practical consequences in the biomedical sciences.


D. Tuveson

The Pew Charitable Trusts granted Adrian and his collaborator Paola Haeger Soto, an associate professor at the Universidad Católica del Norte in Chile, a $200,000 Innovation Fund grant. The researchers will investigate the biology behind fetal alcohol syndrome. The Pew Innovation Fund encourages cross-disciplinary collaborations to tackle biological questions in new ways.

Director of the CSHL Cancer Center David Tuveson was inaugurated president of the American Association for Cancer Research (AACR). Tuveson has dedicated his career to identifying new ways to diagnose and treat pancreatic cancer, a highly lethal disease. As AACR president, Tuveson will bring his distinguished scientific and clinical expertise in cancer to one of the oldest and largest cancer research organizations in the world.

Z. Lippman

J.E. Moses

H. Furukawa

T. Janowitz

This year, the National Cancer Institute (NCI) renewed the CSHL Cancer Center grant with a $4.5 million annual award. Led by David, the center explores the fundamental biology of human cancer. CSHL first received funding and NCI designation in 1987. Today, CSHL researches many cancer types, such as breast, prostate, leukemia, glioma, pancreatic, sarcoma, lung, and melanoma.

The National Academy of Sciences (NAS) elected Professor & HHMI Investigator Zach Lippman as a member. Members are chosen for their original and significant contributions to science and the world. Lippman has made major discoveries in plant genetics, studying the genes that control the size and yield of important crops.

The Royal Society of Chemistry awarded an international team of scientists the first-ever Organic Division Horizon Prize: the Robert Robinson Award in Synthetic Organic Chemistry. Professor and Fellow of the Royal Society of Chemistry John E. Moses received this award as a part of a collaborative effort with prestigious institutions around the world, including Nobel laureate K.B. Sharpless of Scripps Research Institute. The prize recognizes the development of multidimensional click chemistry, a groundbreaking and innovative technology that creates new kinds of molecules.

Professor Hiroyasu (Hiro) Furukawa received the 2020 Nakaakira Tsukahara Memorial Award from the Japan Neuroscience Society for his research on the NMDA receptor, a key molecule in the brain involved in various types of memory and is implicated in several. The award recognizes scientists conducting innovative research in the life sciences and was created in memory of Osaka University Professor Nakaakira Tsukahara, who studied flexibility, plasticity, and mechanisms of learning and memory in the brain.

Assistant Professor Tobias Janowitz and CSHL Fellow Semir Beyaz were two of the first-ever Endeavor Awards recipients from the Mark Foundation for Cancer Research (MFCR). The Endeavor Awards were created to unite scientists across diverse areas of expertise with the goal of addressing urgent questions in cancer research. Tobias and Semir are looking for systemic biological changes in the entire body after cancer develops.

Semir is also leading a multi-institutional team to generate a comprehensive atlas of uterine cells from patients of diverse ancestry. The $1 million project is part of the $28 million Chan Zuckerberg Initiative (CZI) Ancestry Network for the Human Cell Atlas initiative to identify and characterize uterine cells, in particular all the endometrial cells that line the uterus in people of African, Hispanic, Asian, and Native American descent.

The CSHL-led project, "Multi-omics Maps of Human Endometrium in Diverse Ancestries," is a partnership between the New York Genome Center (NYGC), Northwell Health, and Weill Cornell Medical College. The CZI project builds on the partnership already in place working on the NYGC-led Polyethnic-1000 initiative. That initiative was launched in 2018 to help address cancer care inequities in underserved populations. It is studying at least 1,000 ethnically diverse patients across the New York City population.

Assistant Professor Lucas Cheadle was selected as a McKnight Scholar for his work on microglia, a specialized class of immune cells in the brain. Cheadle and his team identified an important role for microglia in sculpting neural circuits in response to sensory experience. The award encourages neuroscientists in the early stages of their careers to focus on disorders of learning and memory, and is awarded for work that would have immediate and significant impact on clinically relevant issues.

Lucas was awarded a Klingenstein-Simons Neuroscience Fellowship for his research on the developing brain. This fellowship supports innovative research by early-career investigators and


S. Beyaz


L. Cheadle

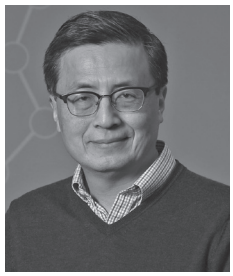

H. Meyer


J. Sheltzer

funds research into the mechanisms underlying a wide range of neurological and behavioral disorders.

Lucas also received the Rita Allen Scholar Award from the Rita Allen Foundation, which funds big ideas that aim to solve complex problems in science and civil society. Cheadle was selected for his neuroscience research.

UK Biobank awarded CSHL Fellow Hannah Meyer the Early Career Researcher of the Year Award for 2021. The award honors early-career researchers who have made significant scientific discoveries using UK Biobank's biomedical database. The award recognizes Meyer's study that examined the structure of the hearts of more than 18,000 individuals in the database.

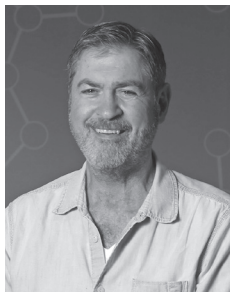

Z.J. Huang

CSHL Fellow Jason Sheltzer was awarded a TheoryLab Collaborative Grant from the American Cancer Society (ACS). TheoryLab™ is a new social media platform that allows cancer researchers to connect and launch collaborative projects. Sheltzer partnered with fellow TheoryLab™ user Rajan Kulkarni, an associate professor at Oregon Health & Science University, for this grant. Together, they are investigating whether having too many chromosomes affects how well cancer therapies work in melanoma patients.

Adjunct Professor Z. Josh Huang was awarded the National Institutes of Health (NIH) Director's Pioneer Award. The award supports scientists with highly innovative, "high-risk, high-reward" approaches to major challenges in biomedical, social science, and behavioral research. Huang was recognized for developing a new generation of precise cell engineering technologies to study the function of diverse cell types across animal organs and species. Josh credits the research program he started at CSHL two decades ago for making this award possible.


A. Moffitt

Senior Computational Fellow Andrea Moffitt was selected as a 2021 Leading Edge Fellow for her research on personalized genomic tools that help guide cancer therapies. Leading Edge is an initiative to improve gender diversity in the life sciences by providing women and nonbinary postdocs with a platform to present their work. She is the professional development chair of the Women in Science and Engineering (WiSE) group at CSHL.


D.A. Micklos

David A. Micklos, the founder and executive director of the DNA Learning Center, received the 2021 Bruce Alberts Award for Excellence in Science Education from the American Society for Cell Biology (ASCB). The award honors David for his innovation and leadership in science education at the DNALC, the world's first science center devoted to public genetics education. Each year, the ASCB recognizes individuals who have made significant contributions to the discipline and to the biology community over their careers.

I was awarded the 2021 Advance Global Impact Award from the Australian-based organization Advance.org. The annual prize honors an Australian who has had an extraordinary impact worldwide through their work; I was recognized for my cancer research using molecular biology and genetics, leadership, and influence in numerous fields and disciplines. The award was presented by


B. Stillman


A. Schorn


C. dos Santos


M. Egeblad


N. Bhattasali


D. Torre


S. Edwards


C. Prizzi

Professor Barry Marshall, the 2005 Nobel laureate in Physiology or Medicine and 2018 recipient of the Global Impact Award.

New Hires/Promotions

Andrea Schorn was promoted to Assistant Professor. Camila dos Santos was promoted to Associate Professor. Mikala Egeblad was promoted to Professor.

CSHL welcomed its first NeuroAI Scholar, Nikhil Bhattasali.

Douglas Torre joined CSHL as Vice President and Chief Information Officer. Shanique Edwards was recruited as Assistant Director, Research Operations.

Charles Prizzi was promoted to Senior Vice President & Special Advisor to the President.

Education Highlights

Meetings & Courses Program

CSHL hosted 28 virtual scientific meetings, six virtual advanced technology training courses, and two in-person courses this year. The Meetings & Courses Program has earned significant kudos within the scientific community by keeping alive the flame of scientific exchange during the COVID-19 pandemic. Attendance at the Laboratory's virtual meetings and courses was higher and more diverse than pre-pandemic levels because the virtual format eliminated space limitations and lowered the cost of participation. CSHL partnered with International Brain Research Organization, Chan Zuckerberg Initiative, and Regeneron to encourage broader participation from minority-serving institutions, low- and middle-income countries, and early-career scientists in need of financial aid.

Throughout the year, the scientific community expressed a strong interest in returning to the CSHL campus for in-person conferences and scientific training. The infrastructure to support this on-campus activity in the future is critical. Renovations to visitor housing begun during the pandemic are nearly complete, including conversion of cabins into family units and the addition of bathroom facilities. The Bush Lecture Hall has also undergone major renovation addressing some key design

The historic Bush auditorium, built originally in 1953, was newly renovated.

The new Bush fireplace room.

problems including noise abatement, lighting improvements, and new door and window hardware. At the Banbury Center, Sammis Hall was renovated to equip all bedrooms with private baths.

Banbury Center

The Banbury Center was closed to in-person meetings this year. Virtual options were unrealistic for most groups who value the Center's isolated and intimate setting for small meetings requiring deep discussion and productive workshops. The Center convened three virtual sessions to prepare for upcoming on-site meetings:

- *Environmental Consequences of Deep-Sea Mining* expert group met to (1) discuss the *scope* of comparisons between land-based and deep-sea mining; (2) review possible sources of *data*; and (3) propose *methodologies* for making comparisons. The expert group, excluding those still facing travel restrictions, will meet at Banbury in April 2022.

Sammis Hall blossoms

- *CSHL Technology and Education Council*, an advisory group of senior scientists who are members of the Laboratory's Corporate Sponsor Program, discussed diversity in clinical trials, indirect effects of COVID-19, and challenges of data. They helped develop themes for future Banbury Center meetings.
- *Making Career-Spanning Learning in the Life Sciences Inclusive and Effective for All* held a kickoff meeting to consider meeting objectives and start case study work on challenges faced by short-format trainers in the life sciences. An in-person meeting will take place in May 2022.

The Center celebrated three publications that resulted from prior Banbury meetings. Participants in the 2019 in-person and 2020 virtual meetings on *Bridging the Research-to-Practice Chasm in Digital Mental Health* published the "Banbury Forum Consensus Statement on the Path Forward for Digital Mental Health Treatment" in *Psychiatric Services*. The 2020 *MAVEN Project* leadership team produced "Introducing the MAVEN Leadership Training Initiative to diversify the scientific workforce" in *eLife*. The *Copper Cancer Consortium* published "Connecting copper and cancer: from transition metal signaling to metalloplasia" in *Nature Reviews Cancer*.

School of Biological Sciences

There were two U.S. and six international students in the twenty-third incoming class. By 2021, 127 students received their Ph.D. degrees from CSHL. One-third of our graduates are in faculty positions, with most of them in tenure-track positions at major U.S. or foreign research institutions. One-third work in industry. Other graduates are postdoctoral fellows, doing research in an academic setting, or are pursuing other scientific careers, such as university administrators, CEOs of biotech companies, and consultants. Applications for 2022 admissions were at a historic high.

The School secured a Postbaccalaureate Research Education Program (PREP) grant from the National Institutes of Health to start a new program for students from minority groups underrepresented in the sciences. The program is for recent graduates. Starting in 2023, four CSHL PREP postbacs will spend a year conducting research with a faculty member, attending courses, professional and skills development workshops, and preparing to matriculate into top-level graduate programs. With this program, CSHL now has undergraduate, postbac, and Ph.D. programs.

School of Biological Sciences entering class of 2021

All CSHL postdoctoral fellows and graduate students are now enrolled in a professional development initiative of the New York Academy of Sciences, the Science Alliance. The initiative is a consortium of universities, teaching hospitals, and independent research facilities in the New York City metro area. This year, CSHL supported a biannual conference, "What Can You Be with a Ph.D."

Because of the pandemic, the Undergraduate Research Program was conducted virtually this year. The Partners for the Future Program for high school students continues to grow with increased numbers of students, particularly from underrepresented communities.

DNA Learning Center

CSHL is the largest provider of biotechnology instruction at the precollege level in the United States. The DNA Learning Center (DNALC) engages more than eight million students across the globe in hands-on experiments with DNA. Together with the City University of New York (CUNY), CSHL opened the DNA Learning Center NYC (DNALC NYC) at the New York City College of Technology in Brooklyn on September 24, 2021. The 18,000-square-foot facility will serve 30,000 New York City middle- and high-school students annually, with a focus on engaging underrepresented communities. Two of the DNALC NYC lab classrooms are dedicated to college-level research courses specially designed for City Tech students. It is the newest and largest of thirteen CSHL DNALC teaching facilities in the United States.

DNA Learning Center (DNALC) Assistant Director Amanda McBrien was named in Crain's New York Notable Nonprofits and Philanthropy list. She was chosen for her work adapting the DNALC's programs to keep students learning hands-on science during the COVID-19 pandemic. The list recognizes people who have delivered innovative programs in response to safety concerns or other challenges in the wake of the pandemic, and Amanda pivoted the DNALC's educational programs to create a unique blend of online classes and at-home science kits.

A. McBrien

After a year of remote learning, DNALC rebooted in-person summer camps at 50% occupancy, including camps at DNALC NYC. With strong virtual camp attendance, total summer camp attendance was up by 14% over the previous year. School year field trips were severely limited by the pandemic. With the gradual relaxing and local

Newly opened CSHL DNA Learning Center NYC provides hands-on learning to students.

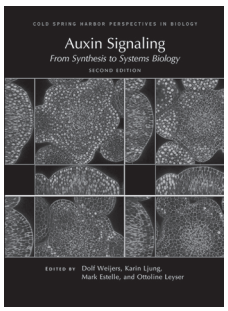
New York City's 110th Mayor Eric Adams cut the ribbon at the opening of the CSHL DNA Learning Center NYC.

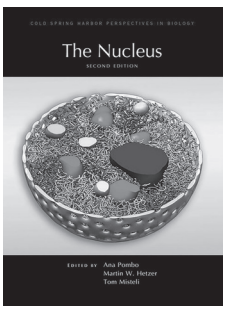
interpretation of CDC and local government guidelines, full classes of students from some Long Island school districts began to come back to the DNALC facility in Cold Spring Harbor.

Cold Spring Harbor Laboratory Press

CSHL Press published five new titles in 2021, four of them derivatives of the *Perspectives in Biology* and *Perspectives in Medicine* journals: *Influenza: The Cutting Edge*; *T-Cell Memory*; *Ahead of the Curve: Women Scientists at the MRC Laboratory of Molecular Biology*; *Auxin Signaling: From Synthesis to Systems Biology*; and *The Nucleus*.

The strong performance of our book program was driven by highly effective direct-to-customer marketing and e-commerce efforts. Offering electronic editions has helped drive this success: in 2021, more than 40% of Press website sales included an e-book, either as a companion to a print edition or a stand-alone purchase.





Editorially, CSHL's scientific journals continue to achieve recognition for high standards of quality and relevance. Judged by impact factor, *Genes & Development* and *Genome Research* remain highly positioned in their disciplines among primary research journals. Both journals continue to rank in the top 1% of the 9000+ journals in the Science Citation Index. *Perspectives in Biology* is in the top quartile of Cell Biology and *Perspectives in Medicine* in the top quartile of Research & Experimental Medicine.

Life Science Alliance is an open-access journal owned and published jointly by CSHL, the European Molecular Biology Organization, and Rockefeller University. Launched in April 2018, the journal published 198 new articles in 2021. Usage of the journal's content across all platforms grew more than 18% over the previous year.

Preprints in Biology and Medicine

A preprint is a research manuscript made freely available by its authors on a public website—a preprint server. The Laboratory's preprint server, bioRxiv, is the world's largest source of preprints in biology, with more than 148,000 unique manuscripts, increasing at a rate of more than 3,000 each month. In 2021, nearly 37,000 new manuscripts were posted.

bioRxiv contains the work of 580,000 authors with 45,000 institutional addresses in 151 countries and dependent territories. The largest subject categories are neuroscience (18%), bioinformatics (9%), and microbiology (9%). At least 70% of manuscripts posted to bioRxiv are published in a journal within two years: nearly 81,000 papers first posted to bioRxiv have been published in 4,000 journals.

medRxiv was launched in 2019 as a complement to bioRxiv. It is owned and operated by CSHL and managed in partnership with Yale University and BMJ, the global health information provider. Generous funding is provided by the Chan Zuckerberg Initiative.

As of February 15, 2022, 29,500 unique manuscripts were posted to medRxiv from 157 countries. The largest subject categories on medRxiv are Infectious Diseases (25%), Epidemiology (19%), and Public and Global Health (9%). Usage of medRxiv increased dramatically in 2020, rising to 10 million page views per month at its peak; it has maintained levels of 6–7 million views since then. More than 10,500 papers first posted to medRxiv have been published in more than 2,300 journals.

The COVID-19 pandemic proved that preprint servers are made for crises in public health when there is a need to share the newest information with utmost speed. There was a surge of pandemic-related submissions to medRxiv and bioRxiv that began before the pandemic was widely recognized. Some 22,000 such articles are now available.

Board of Trustees

In 1890, the Brooklyn Institute of Arts and Sciences established a program for biology teachers on Long Island Sound that became the modern-day CSHL. Eric Adams, New York City's 110th Mayor, welcomed CSHL back to Brooklyn in 2021. "This state-of-the-art facility will be a hub for hands-on STEM laboratory education for all New York City students, and a destination for families to learn about genetics, genealogy, and more," he said. Thirty million dollars was raised for the DNA Learning Center facility at City Tech of the City University of New York. Donors included founding contributor Laurie Landeau Foundation, Achelis and Bodman Foundation, Booth Ferris Foundation, Office of the Brooklyn Borough President, Carson Family Charitable Trust,

From *left* to *right*: CSHL President and CEO Bruce Stillman, Double Helix Medal recipients Leonard Schleifer and Reggie Jackson, Chair of the CSHL Board of Trustees Marilyn Simons, and Double Helix Medal recipient George D. Yancopoulos.

J. Moutoussamy-Ashe

E. Cogan Fascitelli

Ellen and Casey Cogut, William Randolph Hearst Foundation, Jerome Levy Foundation, Terry and Bob Lindsay, the Perkin Fund, Pfizer Foundation, Alison Holtzschue and Doug Schloss, Simons Foundation International, Alfred P. Sloan Foundation, Danielle and Paul Taubman, Thompson Family Foundation, and Anne Wojcicki Foundation.

The Double Helix Medals Dinner raised a record $5 million honoring baseball legend Reggie Jackson and Regeneron Pharmaceuticals, Inc. cofounders Leonard Schleifer and George D. Yancopoulos. The event was chaired by trustees Ms. Jamie Nicholls and Mr. O. Francis Biondi, Jr., Mr. and Mrs. Jeffrey Kelter, Mr. and Mrs. Robert Lindsay, Drs. Marilyn and James Simons, and Mr. and Mrs. Paul Taubman. The 20th annual Women's Partnership for Science luncheon, attended by 300 people, raised more than $250,000 to support women scientists' research. A $3 million gift established Chris Vakoc as the first Alan and Edith Seligson Professor of Cancer Research.

Philanthropic contributions are crucial to the Laboratory's success. We mourned the loss of three champions of CSHL, trustees David Knott, Alan Seligson, and Ed Travaglianti. Contributing fresh perspectives are new trustees: photographer and public health advocate Jeanne Moutoussamy-Ashe, and businesswoman and philanthropist Elizabeth Cogan Fascitelli.

Library and Archives

2009 Nobelist and former CSHL Fellow Carol Greider donated her 1975–1992 lab notebooks and data to the Archives, which has a mission to share the historic impact of life science on society. Greider won the 2009 Nobel Prize for work on telomerase, a protein linked to aging and cancer. She made her pioneering discoveries as a graduate student in Elizabeth Blackburn's laboratory at the University of California Berkeley in the 1980s and continued to study telomerase as one of the first CSHL Fellows.

More than 800 virtual attendees joined a virtual event on October 18 about the life in science of three prominent Nobel Prize laureates, Dorothy Crowfoot Hodgkin, Jennifer Doudna, and Elizabeth Blackburn. Presented by the Center for Humanities and the History of Modern Biology, the event was organized by Nancy Hopkins, Amgen Inc. Professor of Biology Emerita, Massachusetts Institute of Technology, Mila Pollock, Executive Director, CSHL Library &

Nobel laureate and CSHL alumna Carol Greider, circa 1985.

Archives, the Center for Humanities and the History of Modern Biology, and Jan Witkowski**,** Professor, CSHL School of Biological Sciences.

Susan Hockfield, president emerita of MIT, moderated the discussion with Georgina Ferry, biographer, author of *Dorothy Crowfoot Hodgkin: Patterns, Proteins and Peace, a Life in Science*; Jennifer Doudna, subject of *The Code Breaker: Jennifer Doudna, Gene Editing, and the Future of the Human Race* by Walter Isaacson; and Elizabeth Blackburn, co-author of *The Telomere Effect: A Revolutionary Approach to Living Younger, Healthier, Longer.*

Business Development and Technology Transfer

The first three quarters of 2021 proved to be a banner period for the biotech industry in raising capital and starting new companies. During this time, the Business Development and Technology Transfer team was able to close its largest-ever deal with an early-stage company (operating in stealth mode at press) to start up equity, provide five years of funding to access multiple drug targets, and, upon success, bring milestones and royalties to CSHL. This was very timely as biotech company valuations started to drop toward the end of the year.

Licensing and equity revenue received totaled $2.8 million and we received $0.55 million of patent reimbursement. Sponsored research booked in 2021 grew to $2.5 million on the back of the new deal mentioned above.

CSHL's science credentials in drug development were also recognized in the signing in 2021 of an institutional agreement with Autobahn Labs, a novel venture capital–backed early-stage accelerator company. The promise of this deal brings early-stage capital to fund drug development programs and advance new clinical assets to become investigational new drugs (IND). The first company to receive funding with this model is Lingbo Zhang's work on M4 inhibitors.

Fierce Biotech named Mestag Therapeutics™, a start-up inflammatory disease and immuno-oncology company, as one of its 2021 "Fierce 15" biotechnology companies. Mestag was founded in 2020 by an international team of researchers including Professor and Cancer Center Director David Tuveson. Mestag is developing new therapeutics for people affected by inflammatory disease and cancer. Tuveson and his colleagues target fibroblasts, the most common type of cell found in connective tissue. Mestag received $11 million in seed financing, jumpstarting its efforts to develop a first-in-class fibroblast therapeutic strategy in collaboration with CSHL.

CSHL spin-out company Envisagenics and Biogen Inc. announced a new collaboration to advance RNA research in central nervous system diseases by leveraging the Envisagenics artificial intelligence–based platform called SpliceCore®. Led by CEO Maria Luisa Pineda, who is a CSHL School of Biological Sciences alum, Envisagenics uses its proprietary algorithms to identify potential RNA splicing errors and design potential therapeutics to fix them. Biogen is a global biotechnology company that pioneers treatments for neurological diseases such as multiple sclerosis, spinal muscular atrophy, and Alzheimer's.

The SpliceCore® platform was developed by Martin Akerman when he was a postdoc in CSHL Professor Adrian Krainer's laboratory. Biogen and Ionis Pharmaceuticals collaborated with Krainer's team to release Spinraza®, the first FDA-approved treatment for spinal muscular atrophy, in 2016. Akerman, now co-founder and chief technology officer of Envisagenics, is using SpliceCore® to discover splicing errors and design treatments for other RNA splicing-related diseases.

CSHL is collaborating with Autobahn Labs, a new life sciences incubator, to catalyze the Laboratory's early-stage discovery programs into spin-out companies that commercialize transformational new therapies. CSHL is a world leader in basic biological research and in target discovery for human genetic diseases like cancer. Autobahn will provide up to $5 million to early-stage drug discovery projects that are beyond the academic funding model through its association with Samsara BioCapital, a leading life sciences investment firm, and Evotec SE, a global drug discovery

and development company. This collaboration will enable us to advance our basic biology expertise more quickly from the lab to the clinic.

Infrastructure

CSHL's Facilities Department worked on more than 70 capital projects during 2021.

Construction was completed on the new DNA Learning Center at City University of New York's City Tech facility in Brooklyn, New York. The grand opening was held at the 18,000-square-foot facility in September to kick off programming.

A full renovation and modernization of the 3,775-square-foot Bush Lecture Hall was completed in 2021. The stone patio space outside of Bush Lecture Hall and the Blackford Dining Hall was also replaced.

The revitalization of the 11 cabins for Meetings & Courses Program visitors approached completion. The newly renovated cabins will provide more than 12,000 square feet of renovated housing throughout the year.

The much anticipated rebuild of the historic Seawall on Cold Spring Harbor commenced in June. This project refurbishes approximately 1,300 linear feet of seawall. CSHL worked closely with regulatory agencies including the U.S. Army Corps of Engineers, New York State Department of Environmental Conservation, New York State Historic Preservation Office, and the Village of Laurel Hollow to design the new pile-supported concrete seawall with stone facia. Stones from the existing seawall were reused to maintain the historic integrity of the wall. The height of the wall is increased by 24 inches and is extended at both the north and south ends. Construction will be completed in 2022.

Vital housing renovations continued at the Easthouse Residence, Hershey House, and DeForest North and South locations on main campus and at Sammis Hall on the grounds of the Banbury Conference Center.

At Uplands Farms, construction of new state-of-the-art growth chambers continued, providing 288 square feet of growth area. The installation will be completed in the summer of 2022.

CSHL continued planning for a capital project to expand the facilities on the hillside of the main campus. Initial plans call for new laboratory buildings, housing for Meetings & Courses programs and other scientific visitors, and additional campus parking.

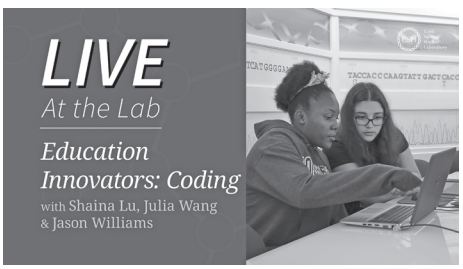
Construction underway on the historic CSHL seawall, summer 2021.

The Facilities Department completed an ASHRAE Level II Energy Audit in 2021 as part of the development of CSHL's energy and sustainability master plans. Programs to modernize and improve the heating, ventilation, air conditioning, electrical, and plumbing systems throughout facilities continue.

Community Outreach

COVID-19 health and safety restrictions limited in-person public programming at CSHL facilities. Virtual programming, the CSHL website CSHL.EDU, and social media channels continued to grow the institution's engagement with broader scientific and nonscientific audiences.

CSHL Public Presentations



January 5: Virtual Event: *Live @ the Lab Defeating Cancer: Working with Communities to Conduct Better Cancer Research* with Kevin Cassel, Dr.PH., Assistant Professor Population Sciences in the Pacific Program, University of Hawaii Cancer Center. Presented as part of the ongoing Roy J. Zuckerberg community engagement series from CSHL's NCI-designated Cancer Center.

March 24: Virtual Event: *Live @ the Lab: Adenovirus—A Look Back at Early CSHL Research* with Terri Grodzicker, Ph.D., CSHL Dean of Academic Affairs, and Bruce Stillman, Ph.D., CSHL President & Chief Executive Officer; moderated by Scientific American journalist and podcast host Steve Mirsky.

April 14: Virtual Event: *Live @ the Lab: Education Innovators*, with CRAIN'S award winner Amanda McBrien, Assistant Director, DNA Learning Center; moderated by CSHL Communications Department Creative Director Eliene Augenbraun, Ph.D.

April 28: Virtual Event: *Live @ the Lab: Education Innovators, CODING* with Shaina Lu, Ph.D. candidate, CSHL Women in Science (WiSE) Coding Camp Instructor, Julia Wang, Ph.D. candidate, CSHL WiSE Coding Camp Instructor, and Jason Williams, Assistant Director, External Collaborations, CSHL DNA Learning Center; moderated by Charla Lambert, Ph.D., Director of CSHL's Office of Diversity, Equity and Inclusion.

July 1: Virtual Event: *Live @ the Lab Panel Discussion about CSHL Programs and Experiences that Introduce Students to Biological Research*. Panelists included Diana Benedicto-Jimenez, Student, Paul D. Schreiber Senior High School, incoming class of CSHL Partners for the Future, Connor Fitzpatrick, Ph.D. candidate; Monn Monn Myat, Ph.D., Associate Dean, CSHL School of Biological Sciences; and Jason Williams, Assistant Director, External Collaborations, CSHL DNA Learning Center; moderated by Charla Lambert, Ph.D., Director of CSHL's Office of Diversity, Equity and Inclusion.

Social Media Outreach

In 2021, CSHL continued to grow its presence on social media by utilizing Facebook, Twitter, Instagram, and LinkedIn to engage the broader scientific and nonscientific audiences in research and education program developments. Through the daily promotion of stories about science, scientists, educators, and the campus community, CSHL's channels successfully grew from 2020:

Facebook reached 24,000 followers, Twitter followers grew by 20% to 33,000, Instagram followers totaled 4,000, and LinkedIn reached 11,000 followers. The CSHL monthly email newsletter achieved 10,000 subscribers.

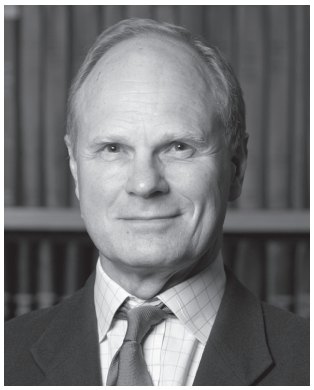
Looking Forward

In 2021 CSHL achieved historic milestones in its research and education programs, highlighting the importance of strategic investments in strong physical and intellectual foundations that facilitate future growth and evolution.

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

CHIEF OPERATING OFFICER'S REPORT

The year 2021 saw fewer disruptive changes to our planning and budgeting process than the year before.



J. Tuke

Financial Resources

The 2021 Budget maintained most of the resource-preserving constraints instituted in June 2020. Nonscientific hiring was suspended, with replacements closely scrutinized. Capital investment increased over that of 2020 but remained below what had been planned before the pandemic. Discretionary spending remained absent from budgets.

Within the Educational divisions, the Meetings & Courses Program (M&C) and the DNA Learning Center (DNALC) continued providing virtual programs for most of the year, with some in-person activity occurring in the fall when conditions allowed. Our graduate school, the School of Biological Sciences, was able to remain on campus because of small class sizes. The Banbury Center, given the nature of its conferences, was able to hold only a few virtual programs. The Press provided stable revenues as the downturn in subscriptions was more than offset by robust preprint activities.

Research revenues remained depressed from pre-COVID levels and were slightly lower than in 2020, primarily because of the departure of several Investigators and their grants.

Last year generated the highest level of philanthropic support ever to the Laboratory. The Double Helix Medals dinner in November at the American Museum of Natural History in New York City was by far the largest fund-raising event of the year and happened in person. Attendees were inspired by speeches from the medal recipients and the New York City mayor-elect.

The most stable source of revenue in 2021 was the endowment. Its value grew from $737 million to $796 million net of all gifts, the operating draw, and an investment return of 9.7%. The 10-year return of 8.2% exceeds inflation (CPI) and the 4.5% draw by 1.5%.

The operating result for 2021 was a balanced budget, thanks to excellent fund-raising performance and lower-than-budgeted operating expenses. Some projects and hiring were deferred to 2022.

Operating and Capital Matters

Learning to live and work with COVID-19 became easier and safer for the community as protocols put in place in 2020 became normalized. By autumn all staff had returned to the campus and 97% of employees had been vaccinated. On-site testing (mandated weekly for the unvaccinated) continued throughout the year, as did free vaccines and boosters at the Wellness Center.

The Laboratory took advantage of 10 months without visitors to make improvements to several buildings: the Bush Lecture Hall, Jones Laboratory, and the 11 cabins used for M&C visitors. Vacancies in several faculty residences allowed Facilities to make major upgrades to those buildings. Thanks to generous donors, the storm-ravaged residence next to Ballybung was rebuilt.

Rebuilding our 900-foot, 170-year-old historical seawall was the most significant capital project commenced in 2021. To comply with requests from the Army Corps of Engineers and the NYS Historic Preservation Office, the project's scope became more extensive than originally planned. But the result will be a feature of our landscape that retains its aesthetic charm while providing improved protection against storms and rising sea levels.

Human Resources

In 2021 the Laboratory did not lay off or furlough staff because of COVID. Instead, some staff switched into other roles. For example, the Facilities department benefited from help in providing more frequent cleaning related to COVID-19.

Like most of the country, after the Thanksgiving holiday, we saw a significant rise in COVID-19 infections. We instituted a temporary Remote Work Program for some staff to reduce on-site density. The Administration recognizes changes in the work environment and the desire by many staff for more flexible work arrangements. The overall benefit of that program, including its impact on productivity and equity, will be assessed in June 2022.

A mandatory vaccine policy was put in place early in 2022, allowing for medical and religious exemptions.

Maintaining Community While Embracing Change

Will 2022 bring an end to the disruptions and the uncertainty brought on by COVID-19, or will the pandemic come to an end? We cannot predict or control the future, but we must remain resilient and adhere to the values of our culture. As the challenges we face grow in complexity, collaboration and creativity are never in more demand. In 2022 we will welcome key new leaders—a chief information officer who began in late 2021, and a new CFO who starts in early 2022—both replacing long-standing administrators. Our success at the Laboratory will be defined by our ability to support each other, welcome new leaders and staff, and remain open to new ideas.

John P. Tuke
Chief Operating Officer

Long-Term Service

The following employees celebrated milestone anniversaries in 2021:

40 years	Rodney Chisum, *C. Hammell Laboratory*; Philip Renna, *Communications*
35 years	Chris Hubert, *Facilities*; Adrian Krainer, *Krainer Laboratory*; Susan Lauter, *DNA Learning Center*; Tim Mulligan, *Uplands Farm*
30 years	Kathy Cirone, *CSHL Press*; Patricia McAdams, *Building Services*; Chris Oravitz, *Procurement*; Michael Regulski, *Ware Laboratory*; Frank Russo, *Facilities*; Linda Van Aelst, *Van Aelst Laboratory*; Barbara Zane, *Meetings & Courses Program*
25 years	William Carmona, *Culinary Services*; Wendy Crowley, *Meetings & Courses Program*; Constance Hallaran, *Meetings & Courses Program*; Melissa Kramer, *McCombie Laboratory*; Marcie Siconolfi, *CSHL Press*
20 years	Andres Alarcon, *Human Resources*; Joe Carrieri, *Facilities*; Maoyen Chi, *Meetings & Courses Program*; Damian Desiderio, *Finance & Accounting*; Diane Fagiola, *Advancement*; Karen Filasky, *Finance & Accounting*; Bibi Garite, *CSHL Press*; Steve Gregorovic, *Security*; Adriana Hincapie, *Joshua-Tor Laboratory*; Louis Malfi, *Information Technology*; Alea Mills, *Mills Laboratory*; Stephanie Muller, *McCombie Laboratory*; Betsy Panagot, *Business Development & Tech Transfer*; Amy Qiu, *Lukey Laboratory*; Uma Ramu, *Martienssen Laboratory*; Cesar Sisalima, *Building Services*; Doreen Ware, *Ware Laboratory*
15 years	Peter Anderson, *Shipping & Receiving*; Carol Brown, *CSHL Press*; Becky Dong, *Research Operations*; Hiro Furukawa, *Furukawa Laboratory*; Walter Goldschmidts, *Office of Sponsored Programs*; Todd Heywood, *Information Technology*; Ying Hsiu Liu, *Krainer Laboratory*; Juan Magana, *Culinary Services*; Mala Mazzullo, *CSHL Press*; Krystyna Rzonca, *Procurement*; James Van Wie, *Security*; Maria Villatoro, *Shipping & Receiving*; Chris Yoon, *Yoon Laboratory*

(*Back row, left to right*) C. Prizzi, B. Panagot, A. Krainer, A. Alarcon, S. Lauter, J. Stone, T. Mulligan, M. Hutchinson, H.Y. Teo, B. Stillman; (*center row, left to right*) D. Ware, A. Mills, M. Siconolfi, J. Carrieri, B. Veneable III, L. Malfi, A. Qiu; (*front row, left to right*) M. Regulski, D. Spector, J. Schinndler, F. Russo, B. Garite, M. Wisnewski, L. Manche, C. Sisalima, P. Abisognio, A. Hincapie, C. Hubert, A. Kass-Eisler, P. Renna.

RESEARCH

CANCER: GENE REGULATION AND INHERITANCE

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. Her group has recently found that mammary glands react differently to a second pregnancy than they do to the first one, 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 utilizes 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 Camila's group is to improve the notion of the mammary epigenome during normal development and use this information to gain insight 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 insight 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 year, members of the Joshua-Tor laboratory explored the function of a very similar protein, called Argonaute 1, that 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 cancer and genetic diseases, and the means by which faulty splicing can be corrected. For example, they study 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 defect using a powerful therapeutic approach. It is 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 antisense technology to develop therapies for other diseases caused by splicing defects, including familial dysautonomia, and to target a cancer-specific alternative-splicing event that controls the Warburg effect. In addition, they are applying antisense technology to stabilize mRNAs that are destroyed by a process called nonsense-mediated mRNA decay (NMD), both to learn about the underlying mechanisms and to develop new therapies (e.g., for a nonsense allele in cystic fibrosis). The Krainer laboratory has also worked to shed light on how splicing factors and alternative splicing promote cancer progression in the context of breast, liver, brain, pancreatic, and blood malignancies. Finally, the laboratory continues to study fundamental mechanisms of splicing and its regulation, focusing on the precise recognition of highly diverse intronic and exonic pre-mRNA features by various spliceosome components.

John Moses' laboratory specializes in click chemistry, a powerful discovery method that relies upon the most robust chemical reactions to synthesize functional molecules. Small molecules are important because Nature's machinery, including proteins, enzymes, and receptors, evolved to interact selectively with molecular ligands, similar to how a key fits a lock.

Using click chemistry, they create molecular probes for studying biological systems that may lead to new treatments for deadly diseases, including cancer. For example, they developed a new class of therapeutic DNA binding ligand that interacts selectively with telomeric regions of the genome involved in cellular maintenance. Several of these telomere binding ligands show remarkable selectivity and potency against cancer cells and tumors.

Further developing click chemistry, they recently described a discovery method called diversity oriented clicking, which exploits a focused group of reliable click chemistry reactions to achieve structural diversity. Their diversity-clicking approach led to the discovery of a new group of antibiotics with activity against multidrug-resistant bacteria, including MRSA.

Through the application of click chemistry, the Moses group is committed to developing Chemistry *for* Biology at Cold Spring Harbor Laboratory.

Transposable elements make up half of our DNA. They control gene expression and have been a major evolutionary force in all organisms. The **Andrea Schorn** laboratory investigates how small RNAs identify and silence transposable elements when they become active during development and cancer.

Transposable elements or "mobile genes," which are closely related to viruses, promote active gene expression in a selfish manner. These elements are usually buried in inactive, condensed DNA by their host to prevent mutagenic damage. However, both stem cells in the embryo and cancer cells undergo genome-wide reprogramming that reactivates silent transposable elements. The Schorn laboratory is exploring how the host recognizes transposons among thousands of genes and noncoding DNA and specifically restricts transposon mobility.

They found that a highly conserved 18-nucleotide sequence motif is the Achilles' heel of a widespread class of transposable elements that are closely related to retroviruses such as HIV. These retroelements initiate replication at the 18-nucleotide binding site using transfer RNA (tRNA), an essential RNA component of the cell. In turn, cells produce short fragments of tRNAs that they discovered inhibit this class of retroelements. These tRNA fragments are processed from mature tRNAs under yet unknown conditions and potentially protect many cell types in eukaryotes. They are investigating under which conditions cells produce this class of small RNAs and assessing their impact on development and pluripotency. tRNA fragments are an ancient link between the "RNA interference" silencing machinery, transposons and genome stability, with potential roles in trans-generational inheritance and cancer.

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, and they are 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 upon the differentiation of mouse embryonic stem cells to neural progenitor cells. These data support a model where 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 when 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 a helicase, 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 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.

The research of the **Lingbo Zhang** laboratory focuses on decoding the role of metabolites, including micronutrients and neurotransmitters, in the tumor microenvironment of hematologic malignancies. They utilize a combination of CRISPR-Cas functional genomics and metabolomics approaches to systematically uncover key metabolites and their genetic effectors, including receptors and metabolic enzymes, to reveal mechanistic insights into and identify key drug targets for hematologic malignancies.

Together, the Zhang laboratory recently uncovered a series of critical metabolites in the tumor microenvironment, including acetylcholine and pyridoxal, and their genetic effectors as novel regulators of hematologic malignancies. They identified cholinergic receptor muscarinic 4 (CHRM4) as a novel regulator of early erythroid progenitor self-renewal and a therapeutic target for myelodysplastic syndrome (MDS). Their research uncovered the HematopoArc as a novel nervous system activity–based regulatory mechanism of hematopoietic stem and progenitor cell self-renewal. They also identified the vitamin B6 pathway as a nutritional and metabolic dependency in acute myeloid leukemia (AML) that coordinates nucleotides and putrescine metabolism specifically required for leukemia maintenance. Their research uncovered this vitamin B6 pathway as a pharmacologically actionable target to treat leukemia with minimal bone marrow suppression effect. Through collaborations with medicinal chemists, they are building pharmacologic approaches to target these novel regulators, and their findings will help to treat devastating hematologic malignancies, including refractory anemia, myelodysplasia, and leukemia.

UNDERSTANDING THE EPIGENETIC REGULATION OF NORMAL AND MALIGNANT MAMMARY GLAND DEVELOPMENT

C. dos Santos M. Callaway N. Dong S. Lewis
C. Chen A.V. Hanasoge Somasundara J. Rail
M. Ciccone S. Henry J. Ruiz-Ortiz
S. Cyrill A. Kaleem M. Trousdell

Defining a Role for BPTF Inhibition in Blocking Breast Cancer Progression

C. Chen, M. Ciccone

The goal of this project is to define the effects of bromodomain protein transcription factor (BPTF) inhibition on controlling the development and progression of mammary tumors. To date, we have established distinct roles for the epigenetic factor utilizing BPTF in a series of murine models of mammary tumorigenesis.

We have monitored mammary tumor development in a cohort of Brca1 knockout (KO) female mice (basal/triple-negative breast cancer [TNBC] model) and observed mammary tumors developed within 2–9 months after tumor-assisted macrophage (TAM) administration in all wild-type (wt) mice ($n = 12$). In marked contrast, Bptf KO mice did not develop mammary tumors during the 1-year observation period, suggesting that Bptf depletion significantly blocked mammary tumorigenesis in Brca1-deficient female mice.

Alternatively, tumor monitoring experiments suggest that *Bptf* deletion delayed the onset and progression of mouse mammary tumor virus–polyoma middle tumor antigen (MMTV-PyMT) mammary tumors (luminal-like breast cancer model). Histological and immunofluorescence analysis indicated smaller and rare lung metastasis in Bptf KO mice. Moreover, the effects of Bptf loss on metastatic potential of MMTV-PyMT mammary tumors were further validated utilizing organoid-based migration assays (trans-well) and in vivo metastasis assays. Utilizing single-cell RNA sequencing (scRNA-seq), we determined that Bptf KO MMTV-PyMT mammary tumors lack a specific population of epithelial cells that bear a transcriptional profile of metastatic cancer cells, thus representing a possible basis for decreased lung metastasis. We also found that populations of basal-like lineages gained the expression of estrogen receptor alpha (ERα) after Bptf deletion. Treatment of female mice with tamoxifen, an ERα antagonist, exclusively inhibited the growth of orthotopically transplanted Bptf KO mammary tumors. These results suggest that Bptf inhibition could represent a novel strategy to convert hormone-negative breast cancer subtypes into those that express ERα and therefore increase their response to hormone-targeting therapies. This hypothesis is currently being addressed in classical models of human TNBC, including cell lines and 3D organoid cultures.

Characterization of Pharmacological BPTF Inhibition in Human Cancer Cell Lines

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

The goal of this project is to characterize the effects of new small-molecule inhibitors of BPTF function on survival and proliferation of human breast cancer cells. To date, we employed assay for transposase-accessible chromatin using sequencing (ATAC-seq) data analysis and found that MCF7 cells bearing BPTF single-guide RNAs (sgRNAs) (BPTF KO) or those treated with S-AU-1, the first-in-class BPTF inhibitor, are enriched for open chromatin regions associated with ERα DNA binding domains and genes downstream of the ERα transcriptional program. These results support inhibition of BPTF function related to transcription alterations associated with ERα. More recently, a new generation of BPTF inhibitors was tested on the normal mouse cell line EPH4 to characterize its potency and specificity for BPTF in comparison to S-AU-1. We checked gene expression of known downstream targets of BPTF as a measure of how well the drugs were inhibiting BPTF. We also performed a

caspase activity assay to check if the compounds were inherently toxic to the cells. We found that these new compounds can inhibit BPTF while keeping caspase activity lower than S-AU-1.

Investigation of the Epigenetic Modifications Brought About by Pregnancy in Mammary-Derived Organoid Cultures

M. Ciccone, M. Trousdell, J. Rail

We have 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 have determined that cultures derived from post-pregnancy mammary extracellular matrices (MECs) selectively depend on the activity of the Polycomb repressive complex 2 component Ezh2 to regulate gene expression and branching morphogenesis in response to pregnancy hormones. More recently, we found that EZH2 inhibition results in the selective up-regulation of cell senescence markers in post-pregnancy mammary organoids. These results suggest that EZH2 may be suppressing the pre-senescence state of post-pregnancy MECs in response to pregnancy, thus allowing for their rapid reactivation of pregnancy-induced programs when re-exposed to pregnancy hormones.

Exploring the Plasticity of the Mammary Gland Using Organoids

M. Ciccone, J. Rail, S. Henry, S. Lewis, J. Ruiz-Ortiz

Building on our work using organoids to study pregnancy-related epigenetic modifications to the mammary gland, we have optimized organoid culture of male and female mammary glands. This has allowed us, with imaging and single-cell RNA sequencing analysis, to study the plasticity of mammary tissue in males and females in response to reproductive hormones. We are currently studying the extent to which organoid culture recapitulates the cellular diversity of the mammary gland. This work is relevant for a deeper understanding of basic tissue homeostasis and plasticity. From a clinical perspective, it may offer a methodology for modeling and better understanding breast cancer risk in male-to-female transitioning patients who are exposed to hormone replacement therapy for extended periods of time.

Generation of a Single-Cell RNA Sequencing Signature to Define Cellular Heterogeneity in the Developing Mammary Gland

S. Henry, M. Trousdell, S. Cyrill [in collaboration with Y. Zhao and A. Siepel, CSHL]

One of the most difficult processes of single-cell RNA sequencing analysis is defining the cellular identity of clusters. This is exacerbated by an inconsistency and a limited marker profile to define cellular heterogeneity across the mammary gland. We have utilized transcriptional profiles from FACS-isolated mammary cells to expand our molecular signature, enabling us to define mammary resident cell identities at the single-cell level. This work has resulted in a recent publication (Henry et al. 2021) in which we have successfully generated a list of markers to be utilized to define epithelial and nonepithelial cells in both mouse and human single-cell data. Additionally, we have integrated single-cell data from mouse and human epithelial tissue, giving us insight into evolutionary conserved cell identities.

Developing Models to Study Cellular Composition Changes across Pregnancy Cycles in the Human Mammary Gland

J. Ruiz-Ortiz, A. Kaleem

This project is focused on characterizing human breast tissue development during pregnancy at a single-cell level in a dynamic manner. To achieve this, we are employing scRNA-seq on samples donated by healthy women who have never been pregnant (nulliparous) or have been pregnant at least once (parous), with the goal to define cellular identities and differentially expressed genes that are influenced by parity in breast tissue. Additionally, we are carrying out scRNA-seq and scATAC-seq on 3D organoid cultures derived from human mammary gland epithelial cells in order to assess dynamic molecular changes to the mammary gland in exposure to pregnancy hormones. Ultimately,

pairing scRNA-seq data with scATAC-seq data will also allow us to identify epigenomic mechanisms contributing to mammary gland development during pregnancy and persistent parity-associated changes in the parous mammary gland. Furthermore, comparing data derived from human and murine mammary glands will help us identify conserved parity-associated molecular signatures throughout evolutionary history.

Defining the Molecular Basis of Pregnancy-Induced Breast Cancer Protection

A.V. Hanasoge Somasundara, M. Trousdell, S. Cyrill

Women harboring BRCA1 mutations with a full-term pregnancy before the age of 25 benefit from pregnancy-induced breast cancer protection. Therefore, we developed an inducible mouse model of Brca1 loss of function, for the purpose of investigating how pregnancy-induced changes influence Brca1 null mammary tumor development. Using age-matched, TAM-treated, Brca1 KO nulliparous and parous female mice, we have shown that a full pregnancy cycle decreases the frequency of Brca1 KO mammary tumors by 80%.

We next utilized scRNA-seq and flow cytometry analysis to define parity-induced changes that inhibit tumor growth. We determined that pregnancy induces the expansion of specific subtypes of natural killer T (NKT) cells during the late stages of involution, which preferentially populates the fully involuted mammary tissue. NKT cell expansion was linked to increased expression of CD1d on the surface of post-pregnancy MECs, which was associated with the stable gain of active transcription markers at the CD1d loci and increased mRNA levels. We further demonstrated that gain of CD1d expression on post-pregnancy MECs and expansion of NKT cells were observed in mammary tissues that failed to develop premalignant lesions and tumors in response to oncogenic stressors such as either cMYC overexpression (CAGMYC) or loss of Brca1 (Brca1 KO), thus connecting pregnancy-induced molecular changes to alteration of immune microenvironment and lack of mammary oncogenesis. We are currently working on single-cell T-cell receptor (TCR) profiling and TCR overexpression methods to reprogram pre-pregnancy NKT cells, an approach that could allow for the development of adoptive transfer strategies to block cancer development.

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

S. Cyrill, S. Henry, M. Callaway, S. Lewis, A. Kaleem

Over the last year, we have worked toward characterizing the influence of urinary tract infection (UTI)-elicited signals on mammary gland homeostasis, involution, and oncogenesis. Using an induced model of uropathogenic *Escherichia coli* infection to study the influence of UTI-elicited signals on the mammary gland in nulliparous mice, we noted alterations to both the immune and extracellular matrix (ECM) compartments of the mammary gland. Using a combination of imaging techniques and flow cytometry, we characterized these changes and are now extending this model of induced UTIs to mice undergoing a natural pregnancy. We have also identified systemic factors that may mediate the effects of UTIs on the mammary gland and are investigating them further as prophylactic targets using mammary-derived organoid models. Finally, we induced UTIs in a murine model of Brca1 deletion to assess the impact of UTI-induced signals on tumorigenesis. With systemic Brca1 deletion, UTI-bearing mice developed malignancy faster than their UTI-free counterparts. We are now extending this analysis to a murine model with mammary-specific Brca1 deletion using mammary transplantation.

Elucidating Key Physical and Molecular Drivers of Gestational Breast Cancer

M. Callaway

This project aims to understand key drivers of gestational breast cancer, or cancer that arises during or shortly after pregnancy. Because of the high recurrence and poor prognosis of gestational breast cancer, we seek to characterize distinct cellular and tissue alterations occurring during pregnancy that may prime breast tissue to support disease progression and recurrence post-pregnancy. Preliminary evidence suggests gestational hormones promote normal mammary epithelial cells to assume a mesenchymal-like state with enhanced migratory capacity. We hypothesize that systemic pregnancy hormones may alter ECM composition and architecture in the breast, generating

structures conducive to epithelial cell dissemination. Further, we suspect that hormone signaling during pregnancy promotes gestational cancer and recurrence by promoting epithelial cell dissemination into and along extracellular matrix architectures. This project will employ scRNA-seq, 3D organoid systems, high-resolution live imaging, mouse models of pregnancy and mammary cancers, and live human samples for organoid culture. The overall goal is to better understand drivers of gestational breast cancer and identify new targets for therapeutics to improve survival and prognosis for patients with gestational breast cancer.

Defining the Impact of Aging, Pregnancy, and Menopause on Normal and Malignant Mammary Gland Development

S. Henry

The goal of this project is to elucidate the precise modification associated with late age of first pregnancy and onset of menopause in mice and to examine how these modifications impact mammary gland development and oncogenesis. To gain insight into the effects of late age of first pregnancy, we plan to utilize scRNA-seq and organoid derived from nulliparous and parous aged mice (>6 mo). Preliminary analysis of organoid cultures actually indicates an increase in size of aged mammary organoids, but minimal formation of branching structures, which suggests that there is some deficiency in the developmental process of the mammary gland in a later age of first pregnancy. To examine how late age of first pregnancy impacts oncogenesis, we plan to utilize our CAGMYC mouse model. From this model we will derive organoids for aged nulliparous and parous mice and examine how DOX treatment impacts mammary gland organoid structure. Last, we will utilize single-cell RNA sequencing and single-cell ATAC sequencing to examine how menopause impacts previously described pregnancy-induced alterations, ultimately defining how menopause influences pregnancy-induced unprotection.

PUBLICATIONS

Hanasoge Somasundara AV, Moss MA, Feigman MJ, Chen C, Cyrill SL, Ciccone MF, Trousdell MC, Vollbrecht M, Li S, Kendall J, et al. 2021. Parity-induced changes to mammary immune microenvironment control NKT cell expansion and mammary oncogenesis. *Cell Rep* **37:** 110099.

Henry S, Trousdell MC, Cyrill SL, Zhao Y, Feigman MJ, Bouhuis JM, Aylard DA, Siepel A, dos Santos CO. 2021. Characterization of gene expression signatures for the identification of cellular heterogeneity in the developing mammary gland. *J Mammary Gland Biol Neoplasia* **26:** 43–66.

Truong TH, Benner EA, Hagen KM, Temiz NA, Kerkvliet CP, Wang Y, Cortes-Sanchez E, Yang C, Trousdell MC, Pengo T, et al. 2021. PELP1/SRC-3-dependent regulation of metabolic PFKFB kinases drives therapy resistant ER^+ breast cancer. *Oncogene* **40:** 4384–4397.

Zahid H, Buchholz CR, Singh M, Chan A, Nithianantham S, Ciccone M, Shi K, Aihara H, Fischer M, Schönbrunn E, et al. 2021. New design rules for developing potent, cell active inhibitors of the Nucleosome Remodeling Factor (NURF) via BPTF bromodomain inhibition. *J Med Chem* **64:** 13902–13917.

NUCLEIC ACID REGULATORY PROCESSES

L. Joshua-Tor D. Adams, A. Axhemi, B. Bibel, L. Braviner, A. Garg, S. Goldsmith, S. Gonuguntla, J. Ipsaro, M. Jaremko, S. Jones, M. Licht, K. Meze, K. On, J. Scheuring

We study the molecular basis of nucleic acid regulatory processes—RNA interference (RNAi) and DNA replication in particular. We apply the tools of structural biology, biochemistry, and biophysics to protein complexes associated with these processes to elucidate how they work. Cryo-electron microscopy (cryo-EM), X-ray crystallography, 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.

The let-7 Regulatory Network

K. Meze, A. Axhemi

Although many studies focus on the identification of microRNA (miRNA) targets and the various downstream mechanisms of gene silencing, we are interested in the regulation of a particular developmental miRNA, let-7. The pluripotency factor Lin28 inhibits the biogenesis of the let-7 family of mammalian miRNAs. Let-7 regulates proliferation genes such as *MYC*, *HMGA2*, and *RAS* and thus plays an important role in stem-cell differentiation and acts as a tumor suppressor. In stem cells, the pluripotent factor, Lin28, binds to precursor let-7 (pre-let-7) hairpins, triggering the 3′ oligouridylation activity of TUT4/7. The oligoU tail added to pre-let-7 serves as a decay signal, as it is rapidly degraded by the exonuclease Dis3L2. Genetic disruption of *DIS3L2* is the primary cause of Perlman syndrome, a congenital disorder leading to fetal overgrowth and an increased susceptibility to Wilms' tumor development. In somatic cells, in the absence of Lin28, TUT4/7 promotes let-7 biogenesis by catalyzing single uridine addition to a subset of pre-let-7 pre-miRNA. We have been studying the molecular basis and mechanism of Lin28-mediated recruitment of TUT4/7 to pre-let-7 and its effect on the uridylation activity of TUT4/7, switching it from a monouridylation activity to an oligouridylase, and the subsequent degradation of pre-let-7 by Dis3L2.

Dis3L2 is a 3′-5′ exoribonuclease that can independently degrade structured RNA substrates and apart from pre-let-7, its substrates include many coding and noncoding 3′ uridylated RNAs. Earlier, we determined the basis for Dis3L2's substrate recognition, but the mechanism of structured RNA degradation by this family of enzymes was unknown. We have now characterized the discrete steps of the degradation cycle by determining cryo-EM structures representing snapshots along the RNA turnover pathway and measuring kinetic parameters for single-stranded RNA (ssRNA) and double-stranded RNA (dsRNA) processing. We discovered a dramatic conformational change that is triggered by the dsRNA, involving repositioning of two cold-shock domains by 70Å (Fig. 1). This movement exposes a trihelix-linker region, which acts as a wedge to separate the two RNA strands. Furthermore, we show that the trihelix linker is critical for dsRNA, but not ssRNA degradation. Our findings reveal the conformational plasticity of this enzyme, and detail a novel mechanism of structured RNA degradation.

Molecular Mechanisms of piRNA Silencing

J. Ipsaro [in collaboration with A. Palmer III, Columbia University Medical Center]

Maintaining genomic integrity across generations is critical for the survival of a species. To help ensure this, higher eukaryotes have evolved a germline-specific RNAi system known as the Piwi-interacting RNA (piRNA) pathway. piRNA silencing specifically protects against the activity of transposons—mobile genetic elements that can integrate into distant genomic sites and potentially cause gene dysregulation. Unlike the well-characterized small

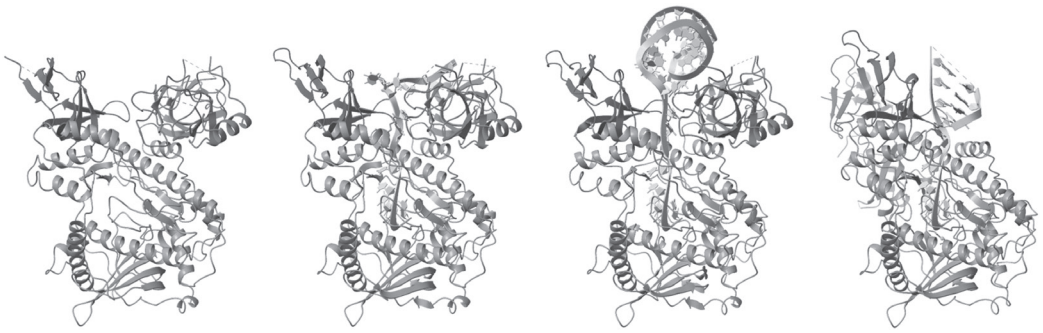


Figure 1. Snapshots along the degradation pathway of structured RNAs. (*Left to right*) RNA-free human Dis3L2 is pre-organized into a vase conformation to bind RNA substrates (yellow), with a 7-nt deep tunnel leading to the nuclease active site. When a substrate with a long single-stranded U-tail binds, it enters the tunnel containing a large number of U-specific binding sites. When the RNA overhang is shortened to ~12 nt, additional contacts are made with the structured, double-stranded RNA region of the substrate. Further shortening of the overhang leads the structured portion of the RNA to run into the enzyme and trigger a large rearrangement of the two cold shock domains (CSDs) (pink, orange) to the prong conformation shown in the *right* panel. This allows the base of the double-stranded RNA to access a module in the RNB domain (blue) that acts as a wedge to separate the two RNA strands and allow entry of one of the strands into the narrow, shortened tunnel leading to the active site. In this way, the enzyme ensures continued RNA degradation during RNA-duplex unwinding.

interfering RNA (siRNA) and miRNA pathways, piRNA biology can function both in the nucleus and cytoplasm. Dozens of proteins have been implicated in piRNA silencing, and the activity of most of these remains mysterious.

We investigated the mechanism of one such piRNA factor, Asterix/Gtsf1, using a combination of biochemical, biophysical, structural, and sequencing-based techniques. Recombinant expression and characterization of Gtsf1 implicated it as an RNA-binding protein, with the protein's first CHHC zinc finger domain being necessary and sufficient for RNA interaction. Structure determination of the protein by nuclear magnetic resonance (NMR) revealed the atomic details of the RNA-binding interface. Enhanced cross-linking and immunoprecipitation (eCLIP) experiments in cell culture demonstrated that Asterix/Gtsf1 predominantly binds transfer RNAs (tRNAs), which was corroborated by cryo-EM structures of Gtsf1 in complex with co-purifying RNAs.

In probing the relationship of tRNAs to a protein known to be involved in piRNA silencing of retrotransposons, we investigated the effects of Asterix silencing on specific transposon classes. We realized that Asterix preferentially silences long terminal repeat (LTR) retrotransposons, which depend on and share an evolutionary history with tRNAs. Taken together, these data intriguingly suggest that Asterix/Gtsf1 exploits tRNA dependence of LTR transposons as a means to identify them for silencing.

PUBLICATIONS

Ipsaro JJ, O'Brien PA, Bhattacharya S, Palmer AG III, Joshua-Tor L. 2021. Asterix/Gtsf1 links tRNAs and piRNA silencing of retrotransposons. *Cell Rep* **34:** 108914. doi:10.1016/j.celrep.2021.108914

Meze K, Axhemi A, Thomas DR, Doymaz A, Joshua-Tor L. 2021. A shape-shifting nuclease unravels structured RNA. bioRxiv doi:10.1101/2021.11.30.470623

In Press

Ipsaro JJ, Joshua-Tor L. 2022. Developmental roles and molecular mechanisms of Asterix/GTSF1. *WIREs RNA* doi:10.1002/wrna.1716

RNA SPLICING

A.R. Krainer M. Daley, Y. Ishigami, Y.J. Kim, A. Kral, K-T. Lin, Y.H. Liu, L. Manche, E.E. Nakagaki-Silva, M. Rahman, D. Segovia, D. Voss, L. Wan, M.S. Wong, L. Yang, Q. Zhang

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. 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 mRNA–protein complex (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 various disease contexts. A summary of our recently published studies is provided below.

Experimental and Computational Analyses of Splicing

The *SMN2* intronic silencer element ISS-N1 is the target of the antisense oligonucleotide (ASO) drug nusinersen (Spinraza®), which we previously developed in collaboration with Ionis Pharmaceuticals and which became the first approved treatment for spinal muscular atrophy (SMA) in 2016. We previously determined that nusinersen's mechanism of action involves steric blocking of binding by the splicing repressors hnRNPA1/A2 to ISS-N1. However, this mechanism did not rule out the potential involvement of other *trans*-acting factors or RNA secondary structure. We collaborated with Brage Andresen (University of Southern Denmark, Odense) to search for additional regulatory factors. Using RNA-affinity purification and iTRAQ tandem mass spectrometry (MS/MS) analysis, as well as surface plasmon resonance imaging, the SR protein splicing factor SRSF10 was shown to also bind ISS-N1. Two different SRSF10 isoforms have carboxy-terminal RS domains of different lengths, and they repress exon 7 inclusion to different extents. Splice-switching ASOs were used to modulate alternative splicing of *SRSF10* pre-mRNA and alter the

distribution of the long and short isoforms. The results suggested that the relative amounts of the SRSF10 isoforms are important for splicing regulation.

Intronic splicing enhancers and silencers (ISEs and ISSs) are two groups of splicing-regulatory elements (SREs) that typically consist of short RNA-sequence motifs; their mutation or dysregulation of their cognate RNA-binding proteins frequently causes aberrant splicing and results in disease. To comprehensively characterize short SREs, in collaboration with former laboratory member Yimin Hua (Nanjing Normal University, China), we used an *SMN2* minigene and generated a complete pentamer-sequence library that comprises all possible combinations of 5 nucleotides in intron 7, at a fixed site downstream of the 5′ splice site (5′ ss). We systematically analyzed the effects of all 1,023 mutant pentamers on exon 7 splicing, in comparison to the wild-type minigene, in HEK293 cells. We found that the majority of pentamers significantly affect exon 7 splicing: 584 of them are stimulatory and 230 are inhibitory. To identify actual SREs, we utilized motif set enrichment analysis (MSEA) and found groups of stimulatory and inhibitory SRE motifs. We experimentally validated several strong SREs in *SMN1/2* and *MAPT* minigene settings. These results provide a resource for understanding how short RNA sequences regulate splicing. Novel SREs can now be explored further to elucidate their mechanisms of action.

In collaboration with Justin Kinney (CSHL), we previously generated and analyzed complete libraries of all possible 5′-ss sequences in three different minigene contexts, using barcoding and massively parallel RNA sequencing after transient transfection. Among other findings, these data revealed pairwise epistatic interactions between certain nucleotides of the 5′ splice site, such as a positive epistatic interaction between G at the –1 position and G at the +5 position. More recently, we collaborated with David McCandlish (CSHL) to analyze the resulting data and explore higher-order genetic interactions that affect splicing efficiency. Specific predictions made on the basis of their empirical variance component regression analysis were then validated experimentally. These models can predict which nucleotides and positions contribute more to initial 5′-ss recognition by U1 snRNP, and hence are more likely to result in aberrant splicing when mutated in the context of human diseases.

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.

The gene encoding KRAS GTPase is recurrently mutated in pancreatic ductal adenocarcinoma (PDAC), triggering the formation of acinar-to-ductal metaplasia (ADM) and pancreatic intraepithelial neoplasia (PanIN). However, the majority of pancreatic cells from KC (*LSL-Kras$^{G12D/+}$; Pdx-1-Cre*) mice expressing KRAS^{G12D} remain morphologically normal for a long time, suggesting that compensatory feedback mechanisms buffer aberrant KRAS^{G12D} signaling. We discovered a feedback mechanism by which the ubiquitously expressed, conserved oncogenic splicing factor SRSF1 is down-regulated in normal KC-mouse pancreas cells with KRAS^{G12D} and in pancreas organoids acutely expressing KRAS^{G12D}. Conversely, we found that SRSF1 is up-regulated in PDAC precursor lesions

and tumors. Increasing SRSF1 in the pancreas disrupted cell homeostasis by activating MAPK signaling—in part by up-regulating interleukin 1 receptor type 1 (IL1R1) through alternative splicing–regulated mRNA stability—which in turn accelerated PDAC. These results demonstrate the involvement of SRSF1 in the pancreatic cell's homeostatic response against $KRAS^{G12D}$, dysregulation of which facilitates PDAC tumorigenesis.

Antisense Technology Development

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. As mentioned above, Spinraza targets a splicing silencer located in intron 7 of the *SMN2* pre-mRNA, and by blocking the binding of splicing repressors, it causes greater 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.

We previously developed another ASO to correct the splicing defect resulting from a T to C mutation at position +6 of intron 20 in *IKBKAP/ELP1*. Familial dysautonomia patients are homozygous for this mutation. We are currently collaborating with the n-Lorem Foundation to take this approach to the clinic. n-Lorem recently completed medicinal-chemistry and toxicology assessments that validated the original lead ASO and a derivative of it as clinical candidate and backup ASOs. We anticipate that a first-in-human study will take place at NYU Langone's Dysautonomia Center, led by Horacio Kaufmann, M.D. We also explored an alternative therapeutic strategy, in collaboration with Dr. Masatoshi Hagiwara (Kyoto University), whose laboratory developed a small-molecule splice modulator dubbed RECTAS. Knockdown experiments showed that efficient *ELP1* exon 20 inclusion requires SRSF6 binding to an intronic splicing enhancer in intron 20. Moreover, RECTAS interacts directly with CLK kinases and enhances SRSF6 phosphorylation. Consistent with this mechanism, RECTAS promotes exon 20 inclusion in mutant *ELP1*, whereas CLK inhibitors promote exon skipping. We then showed target engagement in vivo after oral administration of RECTAS in mutant-*ELP1* transgenic mice. Thus, small molecules that modulate the phosphorylation state of SR proteins could be developed as therapeutics for diseases involving defective splicing.

We are also pursuing approaches to inhibit NMD of specific genes. Low *CFTR* mRNA expression due to NMD is a major hurdle in developing a therapy for cystic fibrosis (CF) caused by the *W1282X* mutation in the *CFTR* gene. The truncated CFTR-W1282X protein retains partial function, so increasing its levels by inhibiting NMD of its mRNA is expected to be beneficial. Because NMD regulates the normal expression of many genes, gene-specific stabilization of *CFTR-W1282X* mRNA expression is more desirable than general NMD inhibition. We previously showed that synthetic ASOs designed to prevent binding of exon-junction complexes (EJCs) downstream of premature termination codons (PTCs) attenuate NMD in a gene-specific manner. Therefore, to block the binding of the EJCs to the three expected sites downstream of the PTC, we screened for and optimized a cocktail of three ASOs that specifically increases the expression of *CFTR* W1282X mRNA and CFTR protein in transfected human bronchial epithelial cells. We further showed, via Ussing chamber assays, that this treatment increased the CFTR-mediated chloride current.

As an alternative approach, we developed an exon-skipping ASO to achieve gene-specific NMD evasion. A cocktail of two splice site–targeting ASOs induced the expression of *CFTR* mRNA without the premature termination codon–containing exon 23 (CFTR-Δex23), which is an in-frame exon. Treatment of human bronchial epithelial cells with this cocktail of two ASOs that target the splice sites flanking exon 23 resulted in efficient skipping of exon 23 and expression of CFTR-Δex23 protein. The splice-switching ASO cocktail increased the CFTR-mediated chloride current in human bronchial epithelial cells, demonstrating that the CFTR-Δex23 protein has residual activity. Both of these preclinical studies—the above EJC-blocking approach and the exon-skipping approach—set the stage for clinical development of an allele-specific ASO-cocktail therapy for CF caused by the W1282X mutation.

PUBLICATIONS

Abou-el-Enein M, Angelis A, Appelbaum AR, Andrew NC, Bates SE, Bierman AS, Brenner MK, Cavazzana M, Caligiuri MA, Clevers H, et al. 2021. Evidence generation and reproducibility in cell and gene therapy research: a call to action. *Mol Ther Meth Clin Dev* **22:** 11–14.

Ajiro M, Awaya T, Kim YJ, Iida K, Denawa M, Tanaka N, Shibata S, Sakamoto T, Studer R, Krainer AR, et al. 2021. Therapeutic manipulation of *IKBKAP* mis-splicing with a small molecule to cure familial dysautonomia. *Nat Comm* **12:** 4507.

Frederiksen SB, Holm LL, Larsen MR, Doktor TK, Andersen HS, Hastings ML, Hua Y, Krainer AR, Andresen BS. 2021. Identification of SRSF10 as a regulator of *SMN2* ISS-N1. *Hum Mutat* **42:** 246–260.

In Press

Gao Y, Lin K-T, Jiang T, Yang Y, Rahman MA, Gong S, Bai J, Wang L, Sun J, Sheng L, et al. 2022. Systematic characterization of short intronic splicing-regulatory elements. *Nucl Acids Res* **50:** 731–749.

Kim YJ, Nomakuchi T, Papaleonidopoulou F, Yang L, Zhang Q, Krainer AR. 2022. Gene-specific nonsense-mediated mRNA decay targeting for cystic fibrosis therapy. *Nat Comm* **13:** 2978.

Kim YJ, Sivetz N, Layne J, Voss D, Yang L, Zhang Q, Krainer AR. 2022. Exon-skipping antisense oligonucleotides for cystic fibrosis therapy. *Proc Natl Acad Sci* **119:** e2114858118.

Ma WK, Voss DM, Scharner J, Costa ASH, Lin K-T, Jeon HY, Wilkinson JE, Jackson M, Rigo F, Bennett CF, et al. 2022. ASO-based PKM splice-switching therapy inhibits hepatocellular carcinoma cell growth. *Cancer Res* **82:** 900–915.

Marasco LE, Dujardin G, Sousa-Luís R, Liu YH, Stigliano J, Nomakuchi T, Proudfoot NJ, Krainer AR, Kornblihtt A. 2022. Counteracting chromatin effects of a splicing-correcting antisense oligonucleotide improves its therapeutic efficacy in spinal muscular atrophy. *Cell* **185:** 2057–2070.e15.

Wan L, Lin K-T, Rahman MA, Wang Z, Jensen MA, Park Y, Tuveson DA, Krainer AR. 2022. Loss of $KRAS^{G12D}$ feedback regulation involving splicing factor SRSF1 accelerates pancreatic cancer. bioRxiv doi:10.1101/2021.10.13.464210

Zhoua J, Wong MS, Chen W-C, Krainer AR, Kinney JB, McCandlish DM. 2022. Higher-order epistasis and phenotypic prediction. *Proc Natl Acad Sci* **119:** e2204233119. doi:10.1073/pnas.220433119

CLICK CHEMISTRY, SYNTHESIS, AND CHEMICAL BIOLOGY

J. Moses D. Bhunia R. Koelln S. Sun
J. Homer W. Lam N. Upadhyay
R. Johnson N. Pawar D. Vishwakarma

Taking inspiration from nature, click chemistry is a synthesis philosophy that uses robust and versatile near-perfect reactions to rapidly build up molecular complexity, forming linkages between simple building blocks. Click reactions meet stringent criteria: they are scalable, high-yielding, selective, and wide in scope, have functional group tolerance, operate under mild reaction conditions, and generate inoffensive by-products.

The Moses group is focused on applying state-of-the-art click methodologies in two main ways. First, the fidelity and selectivity of click reactions make them perfectly suited for the late-stage derivatization of biologically active scaffolds, allowing the synthesis of bioconjugate probes that explore therapeutic targets and mechanisms of action. Additionally, click chemistry is a powerful discovery tool that enables the high-throughput synthesis of diverse chemical libraries for screening in biological assays. Our principal areas of research include methodology development and application of click technology in the chemistry of cancer and infectious disease.

SuFExable Polymers with Helical Structures

This work was done in collaboration with K.B. Sharpless and P. Wu (Scripps) and H. Zuilhof (Wageningen University).

We have employed SuFEx in the synthesis of an unprecedented class of polymer with potential application as materials. The synthesis of these structurally diverse SOF_4-derived copolymers was achieved through the SuFEx coupling of a selection of bis(iminosulfur oxydifluorides) and bis(aryl silyl ethers) (Fig. 1A). The sulfur-linked polymer presents two key characteristics: (i) the [–N = S(=O)F–O–] polymer backbone linkages are themselves SuFExable and can undergo precise post-polymerization modification (PPM) with phenols or amines to yield branched functional polymers, and (ii) studies of individual polymer chains reveal unprecedented helical structures (Fig. 1B). The PPM of polymer backbones has historically proven to be incredibly challenging. In this work, we have successfully decorated the backbone of our polymers with a range of antibiotics and antiviral drugs, creating novel functional polymers with a wide range of potential applications.

Diversity-Oriented Clicking

This work was done in collaboration with K.B. Sharpless and D.W. Wolan (Scripps).

Our laboratory pioneered diversity-oriented clicking (DOC), a unified click approach for the modular synthesis of lead-like structures through the application of a wide range of click transformations. DOC evolved from the concept of achieving "diversity with ease" by combining classic C–C π-bond click chemistry with recent developments in connective SuFEx technologies. We showcased 2-substituted-alkynyl-1-sulfonyl fluorides (SASFs) as a new class of connective hub in concert with a diverse selection of click-cycloaddition processes. Through the selective DOC of SASFs with a range of dipoles and cyclic dienes, we have accessed a diverse click library of 173 unique functional molecules in minimal synthetic steps (Fig. 2). The SuFExable library comprised 10 discrete heterocyclic core structures derived from 1,3- and 1,5-dipoles; whereas reaction with cyclic dienes yields several three-dimensional bicyclic Diels–Alder adducts. Growing the library to 278 discrete compounds through late-stage modification was made possible through SuFEx click derivatization of the pendant sulfonyl fluoride group in 96-well plates—demonstrating the versatility of the DOC approach for the rapid synthesis of diverse functional structures. A portion of our DOC library was screened against methicillin-resistant *Staphylococcus aureus* (MRSA) (strain USA300). A preliminary screen at a concentration of 200 μM revealed 16 hit compounds (11% hit rate), five of which had significantly increased antibiotic activity compared to methicillin. Hence, using the DOC approach, we identified a new class of antibiotics effective against drug-resistant *S. aureus*.

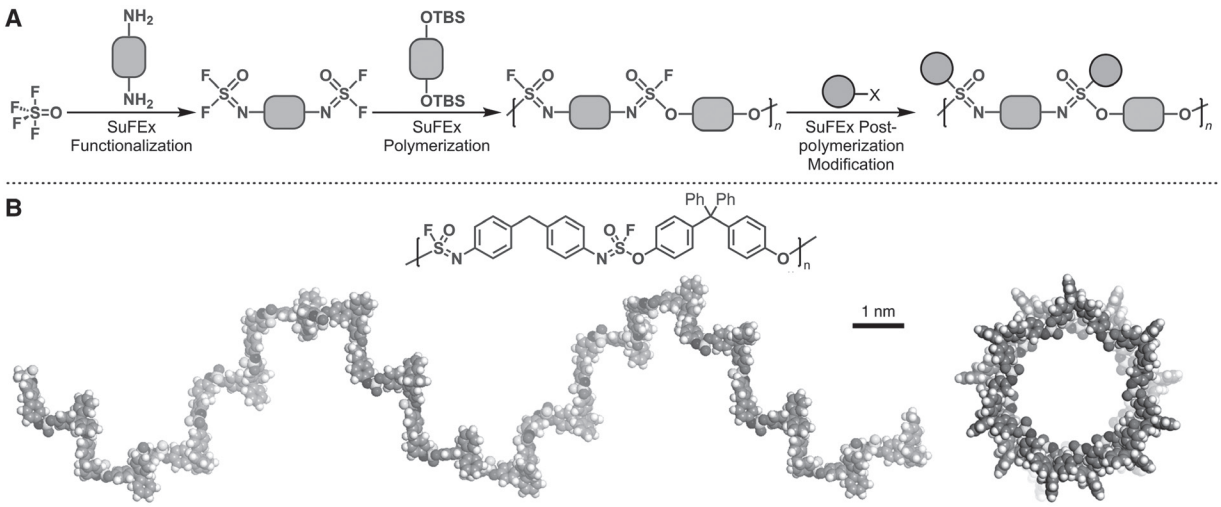


Figure 1. (*A*) Synthesis of polymers and postpolymerization modification using SuFEx click chemistry. (*B*) 2D chemical and 3D PCFF-optimized structure of SuFEx polymer. The polymer consistent force field (PCFF)-optimized structure demonstrates the helical nature of the polymer wires.

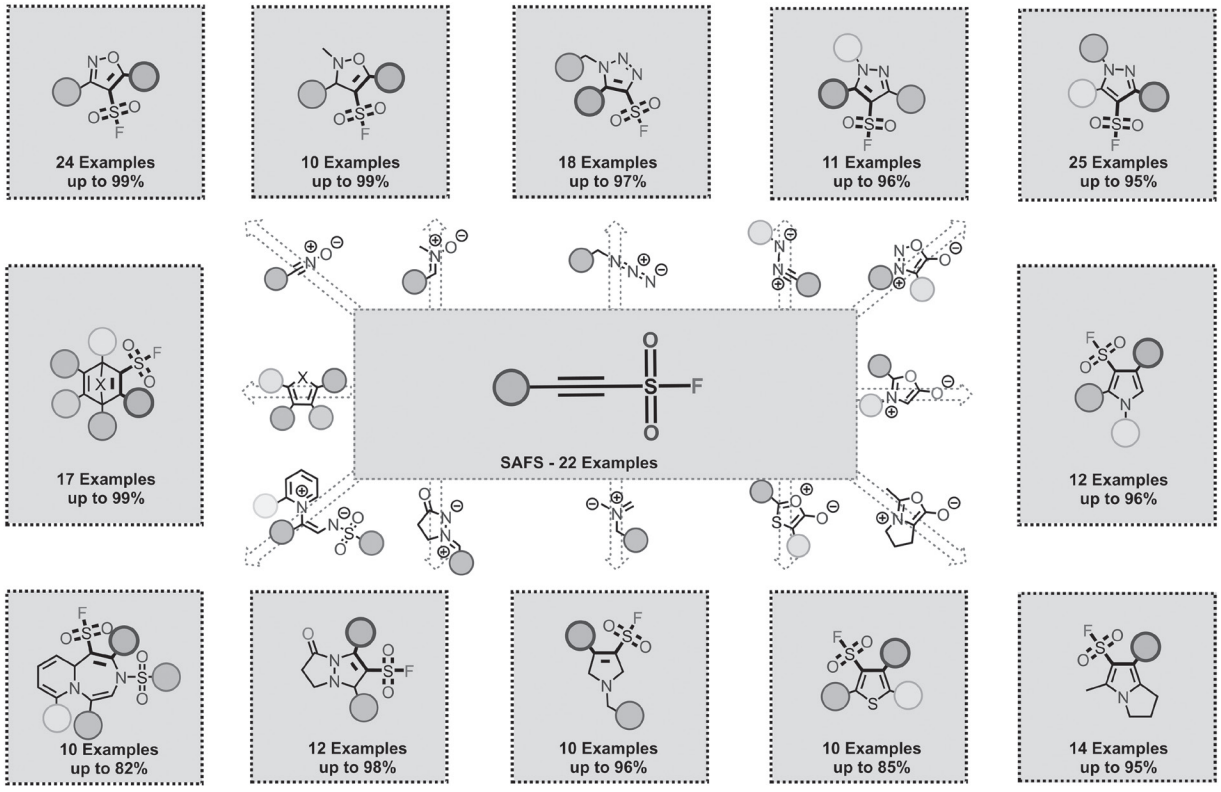


Figure 2. The reaction of 2-substituted-alkynyl-1-sulfonyl fluoride (SASF) hubs with a selection of 1,3-dipoles, 1,5-dipoles, and dienes to assemble a SuFExable 173-compound diversity-oriented clicking (DOC) library**.** The library can be further derivatized via SuFEx click reactions at the pendant sulfonyl fluoride moiety.

Accelerated SuFEx Click Chemistry

Classical SuFEx reactions form stable S–O linkages upon exchange of S–F bonds with aryl silyl ether substrates and, although near-perfect in their outcome, have disadvantages that leave room for improvement. For example, high catalyst loading is sometimes necessary, particularly when reaction conditions lead to catalyst degradation. Another consideration is the steric bulk of the silyl ether group. More steric bulk around the silicon center tends to lead to an increase in reaction time from minutes to several hours. Additionally, SuFEx reactions using aliphatic substrates rather than aryl substrates are rare and difficult. We have developed accelerated SuFEx click chemistry (ASCC), an improved SuFEx method for the efficient and catalytic coupling of aryl and alkyl alcohols with a range of SuFExable hubs. We have found that hindered guanidine base 2-*tert*-butyl-1,1,3,3-tetramethylguanidine (BTMG; Barton's base) is a superb SuFEx catalyst that, when used in synergy with silicon additive hexamethyldisilazane (HMDS), yields stable S–O bond linkages in a single step—often within minutes. The powerful combination of BTMG and HMDS reagents allows for catalyst loadings as low as 1.0 mol% and, in congruence with click principles, provides a scalable method that is safe, efficient, and practical for modular synthesis. Using this improved method, we are able to synthesize compound libraries in excellent yield without the need for purification steps (Fig. 3). ASCC expands the number of accessible SuFEx products and has prompted numerous collaborations for our group, both across Cold Spring Harbor Laboratory and internationally.

Aminium Cation-Radical Catalyzed Selective Hydration of (*E*)-aryl Enynes

The hydration of carbon–carbon triple bonds is an important and atom-economic synthetic transformation. The hydration of terminal alkynes can either give the ketone (Markovnikov pathway) or the aldehyde (anti-Markovnikov pathway) product. In contrast, the hydration of nonsymmetrical internal alkynes can lead to a mixture of either possible ketone regioisomer. Our group has recently reported a mild, selective catalytic method for hydration of (*E*)-aryl enynes to their corresponding Markovnikov enones (Fig. 4A). This transformation has been significantly challenging because of the potential for competing alkene hydration. Previous examples have used toxic mercury(II) salts or strong Brønsted acids to access the desired enone product. We were able to use the bench-stable aminium salt, tris(4-bromophenyl)ammoniumyl hexachloroantimonate (TBPA), to achieve the selective hydration of (*E*)-enynes to (*E*)-enones in excellent yield under neutral, metal-free conditions. Our reaction is both chemoselective and diastereoselective, delivering products in excellent yields from both terminal and internal alkynes. The synthesis of biologically important (*E*)-3-styrylisocoumarins (Fig. 4B), including a formal synthesis of the natural product achlisocoumarin III, demonstrates the utility of this novel transformation.

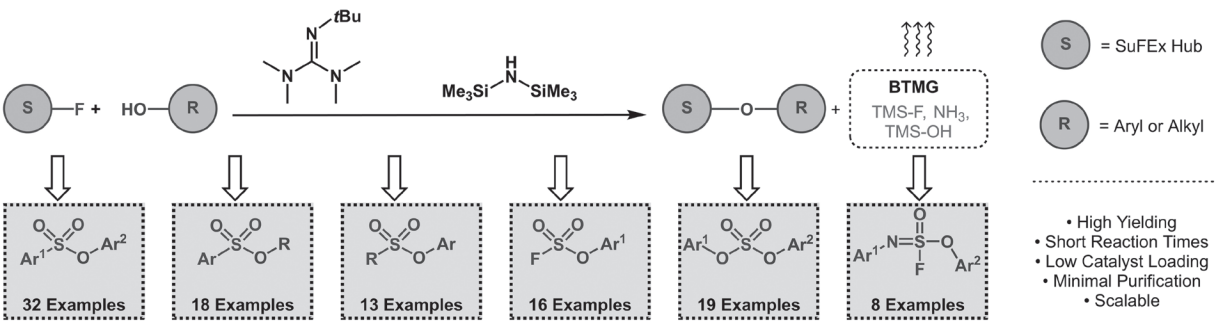


Figure 3. A variety of SuFEx hubs, including aryl and alkyl sulfonyl fluorides, sulfuryl fluoride, aryl fluorosulfates, and iminosulfur oxydifluorides, were reacted with both aryl and alkyl alcohols to form more than 100 compounds using our accelerated SuFEx click chemistry (ASCC) conditions.

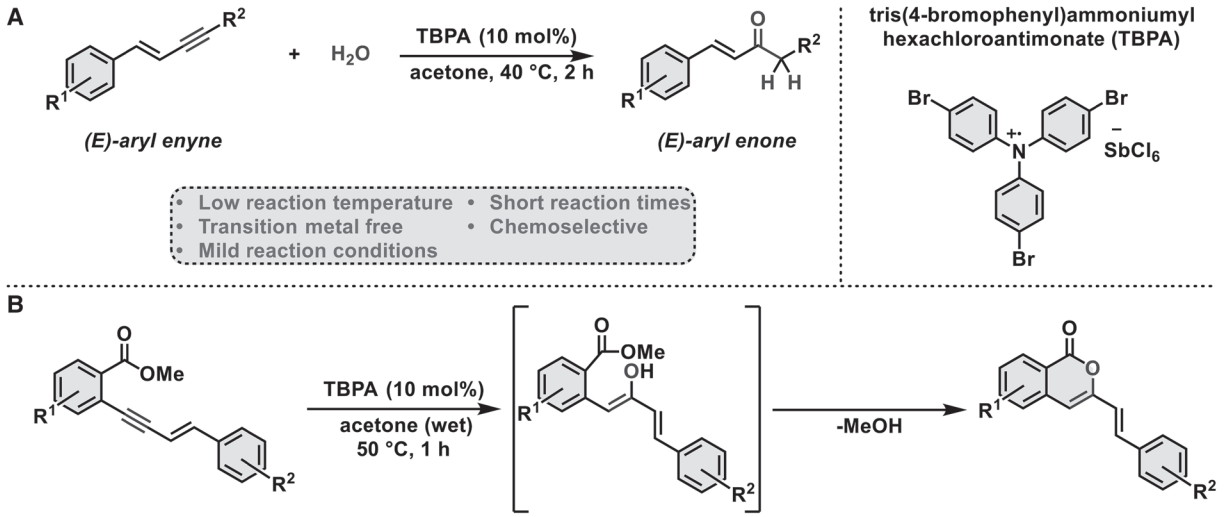


Figure 4. (*A*) General reaction scheme for new enyne to enone transformation. (*B*) Application of this new methodology to synthesize molecules with 3-styrylisocoumarin cores, a motif prevalent in several natural products.

Jerantinine A—Sustainable Synthesis and Analog Generation

In 2018, the Moses group developed a concise and sustainable synthesis for the natural product (-)-jerantinine A from (-)-tabersonine. Jerantinine A is a natural metabolite derived from *Tabernaemontana corymbosa* and exhibits potent antineoplastic activity via multiple mechanisms. Although both inhibition of microtubule formation and modulation of splicing factor 3b subunit 1 (SF3B1) have been implicated, other modes of action of jerantinine A remain obscure. We hypothesize that there are additional targets that contribute to the cytotoxic effects of jerantinine A and, to identify these, we have utilized our SuFEx protocols to access a number of bioconjugates of this natural product. We have synthesized analogs containing biotin for streptavidin pull-down assays as well as novel diazirine photoaffinity probes to further our understanding of jerantinine A's pharmacodynamics and antineoplastic activity. These results, combined with our scalable and sustainable route to the precursor (-)-tabersonine and high-throughput SuFEx derivatization, position us well for future lead discovery and optimization.

PUBLICATIONS

Giel MC, Barrow AS, Smedley CJ, Lewis W, Moses JE. 2021. Aminium cation-radical catalysed selective hydration of (*E*)-aryl enynes. *Chem Commun (Camb)* **57:** 6991–6994. doi:10.1039/d1cc02257a

Li S, Li G, Gao B, Pujari SP, Chen X, Kim H, Zhou F, Klivansky LM, Liu Y, Driss H, et al. 2021. SuFExable polymers with helical structures derived from thionyl tetrafluoride. *Nat Chem* **13:** 858–867. doi:10.1038/s41557-021-00726-x

In Press

Smedley CJ, Homer JA, Gialelis TL, Barrow AS, Koelln RA, Moses JE. 2022. Accelerated SuFEx click chemistry for modular synthesis. *Angew Chem Int Ed Engl* doi:10.1002/anie.202112375

ENDOGENOUS RETROVIRUSES IN MOUSE AND HUMAN AND THEIR REGULATION BY SMALL RNAs

A.J. Schorn S. Constantinou M. Peacey
H. Cullen J.I. Steinberg

Sequence Variation and Evolution of the tRNA Primer Binding Site Essential for Retrovirus Replication

H. Cullen, A.J. Schorn

Transposable elements and retroviruses are a threat to genome stability. They are usually embedded in inactive chromatin, but epigenetic reprogramming during development and disease erases repressive chromatin marks and small RNA–mediated silencing becomes crucial. We found that small RNAs derived from the 3′ end of mature tRNAs (3′-tRFs) are highly expressed in cells undergoing epigenetic reprogramming and target long terminal repeat (LTR) retrotransposons at their conserved tRNA primer binding site (PBS). The use of host tRNAs as a primer for reverse transcription and replication is a hallmark of all LTR retroelements, which include endogenous retroviruses (ERVs) in mammals but also closely related infectious retroviruses such as HIV. Their dependency on tRNA makes this large class of retroelements vulnerable to inhibition by 3′-tRFs. We analyzed PBS sequences in the mouse and human genome and found that their sequence variation is shaped by the conflict of successful tRNA priming for replication *versus* evasion of silencing by 3′-tRFs (Cullen and Schorn, *Viruses* **12:** 192 [2020]). Strikingly, the most active ERV retroelements carry distinct mutations in their PBS that are permissible for priming but decrease binding to tRFs. Our phylogenetic analysis of several ERV families indicates repeated, independent mutation events at the PBS of young and actively replicating ERVs, suggesting a selective advantage of these mutations. The phylogenetic analysis also revealed that early transposon (ETn) and *Mus musculus* particle D (MusD) replicate like other Retroviridae: In contrast to Gypsy and Copia LTR retrotransposons, one strand is copied from their lysine tRNA primer (perfect match with tRNA sequence) and one strand is a copy of their genomic RNA (perfect match with parental transposon). Actively transposing elements of the ETn, MusD, and intracisternal A particle (IAP) families tolerate up to three insertions-deletions in the 18 nt long PBS motif. These mutations accumulated at the same position within the PBS that have been found permissive for tRNA priming and replication of the HIV-1 retrovirus, suggesting that tRNA priming as well as evasion from targeting by tRFs shape PBS sequences in ERVs. Interestingly, we found perfect conservation of the PBS in noncoding ERVs that are no longer actively replicating but may benefit from regulation by 3′-tRFs as their RNA was co-opted by the mammalian host.

Biogenesis of tRNA Fragments (tRFs)

A.J. Schorn, J.I. Steinberg, S. Constantinou [in collaboration with R. Bonasio, University of Pennsylvania, Philadelphia]

Genome-wide reprogramming resets epigenetic marks to enable pluripotency in the developing embryo, and tRFs are able to restrict retrotransposition at these critical time points. The ten-eleven translocation 2 (TET2) oxidase previously believed to drive the decline of DNA methylation during preimplantation reprogramming instead preferentially binds RNA, specifically tRNAs in mouse embryonic stem cells (mESCs). In collaboration with Roberto Bonasio (University of Pennsylvania), we found that TET2 regulates tRF production by cytosine 5-hydroxymethylation of tRNAs (He et al. 2021). Using stringent affinity tags introduced at the *Tet2* locus, the Bonasio laboratory purified and sequenced TET2 cross-linked RNAs from mESCs and found a high enrichment for tRNAs. RNA immunoprecipitation with an antibody against 5-hydroxymethylcytosine (hm5C) recovered tRNAs that overlapped with those bound to TET2 in

cells. Mass spectrometry analyses revealed that TET2 is necessary and sufficient for the deposition of the hm5C modification on tRNA. We found that *Tet2* knockout in mESCs affected the levels of several small noncoding RNAs originating from TET2-bound tRNAs that were enriched by hm5C immunoprecipitation. These results suggest a new function of TET2 in promoting the conversion of 5-methylcytosine to hm5C on tRNA and regulating the processing or stability of different classes of tRNA fragments. TET2 is highly expressed during epigenetic reprogramming in mouse embryos. Our findings indicate that RNA modifications not only determine tRNA maturation but also processing into tRFs during reprogramming.

3′-tRFs are processed from full-length tRNAs under yet-unknown conditions and potentially protect many cell types in eukaryotes. We quantified tRFs and full-length, "parental" tRNA molecules by hydrolysis-based sequencing and revealed several notable patterns across cell types. Specific tRNAs (four of approximately 55 isotypes) are cleaved more frequently than expected by their expression level in a pattern conserved between mouse and human. Moreover, specific tRNA isoacceptors that differ by only a few nucleotides are processed into tRFs preferentially. For example, lysine^{3}- and lysine^{1,2}-tRNA are expressed at similar levels, but lysine^{3}-tRFs are strongly enriched in several cell lines across multiple replicates whereas lysine^{1,2}-tRFs are not. We also profiled 3′-tRF level during different phases of the cell cycle using fluorescent ubiquitination-based cell cycle indicator (FUCCI) cell lines and observed an increase for a specific class of 3′-tRFs (tRF3a) during the G_2/M phase of the cell cycle, whereas microRNAs (miRNAs) and the miRNA-like 3′-tRFs (tRF3b) level did not change during the cell cycle. tRF3a and tRF3b fragments inhibit retrotransposition through different mechanisms. These findings confirmed that these two classes of 3′-tRFs are also processed independently from one another and suggest 3′-tRF production is not merely determined by tRNA amounts but conserved between species and cell types from select tRNAs (unpublished).

Regulation of ERV-Derived Long Noncoding RNAs and Genes

M. Peacey, A.J. Schorn

Because of the perfect sequence complementarity of 3′-tRFs to the PBS of endogenous retroviral sequences, they have tens of thousands of targets in mammalian genomes. Several imprinted genes are derived from ERVs and an estimated 6%–30% of transcripts in mouse and human are driven by retrotransposon sequences that define tissue-specificity and gene-regulatory networks during development. More than 800 LTRs act as alternative promoters and first exons to drive stage-specific gene expression in mammalian oocytes and the developing zygote. Expression of the noncoding murine MERV-L and human HERV-H elements drive totipotency and pluripotency, respectively, in embryonic stem cells. Both retroelements are targets of 3′-tRFs (our unpublished results). As per our genome-wide analysis, both elements exhibit remarkable sequence conservation at their PBS sites, although they no longer replicate nor require tRNA binding. We are examining whether tRFs regulate expression of ERV-derived long noncoding RNAs (lncRNAs) or genes and whether they play an active role in shaping stem cell potency. In addition to their role in early development, ERVs are frequently epigenetically reactivated as cryptic promoters in cancer that drive oncogene expression, and tRFs are highly expressed in many cancer cell lines. We have used miRNA-target site prediction algorithms to identify putative targets of 3′-tRFs in the human transcriptome. We found 4,480 candidate transcripts representing 1,587 genes with target sites for just one isotype of 3′-tRFs (histidine), suggesting numerous targets to explore experimentally and great regulatory potential for tRFs.

PUBLICATION

He C, Bozler J, Janssen KA, Wilusz JE, Garcia BA, Schorn AJ, Bonasio R. 2021. TET2 chemically modifies tRNAs and regulates tRNA fragment levels. *Nat Struct Mol Biol* **28:** 62–70.

REGULATION OF GENE EXPRESSION

D.L. Spector D. Aggarwal S. Bhatia M. Gandhi 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 to produce an RNA molecule, some of which code for 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 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, B. Liu, W. Xu, S. Russo, L. Brine
[in collaboration with F. Rigo and R. MacLeod, 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 >200 nucleotides in length and are referred to as long noncoding RNAs (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 to evaluate their mechanism of action and potential as therapeutic targets.

Metastasis-associated lung adenocarcinoma transcript 1 (*MALAT1*) is a nuclear-localized, ubiquitously expressed lncRNA. It has been shown to be overexpressed in more than 20 different cancer types, including breast cancer. Our laboratory has previously shown that knockout or knockdown of *Malat1* in the MMTV-PyMT mouse model of luminal B breast cancer led to reduction in overall tumor burden, differentiation of the primary tumors, and a significant reduction in metastasis to the lungs. We are currently evaluating the mechanism of action of *MALAT1* in patient-derived breast tumor organoid models. These three-dimensionally cultured models provide a more accurate representation of the patient tumor and subtype diversity than existing in vitro models, and they also recapitulate cell–extracellular matrix interactions unlike cell lines grown in 2D culture.

We have identified 10 triple-negative breast tumor patient-derived organoid (PDO) lines that display >75% *MALAT1* knockdown (KD) using two different human-specific *MALAT1*-targeting antisense oligonucleotides (ASOs). Using two of these PDO lines, we established patient-derived breast tumor organoid xenografts (PDO-Xs) in NOD-SCID gamma mice by injecting 50,000 organoids bilaterally into each of the #4 mammary fat pads. The NH85TSc PDOs ($n = 3$) developed into palpable tumors within 2 wk postinjection. The mice were then treated subcutaneously with 50 mg/kg of either a *MALAT1*-targeting ASO or scrambled nontargeting ASO 2 ×/week for 8 wk. At 2 cm end point, the tumors and organs such as lungs, liver, lymph nodes, brain and spleen were collected and sectioned for hematoxylin and eosin (H&E) staining. Using an antibody against human mitochondria, we identified PDO-derived lung micrometastases via immunohistochemistry (IHC) on multiple lung sections. We observed ~50% reduction in metastasis to the lungs upon *MALAT1* KD in vivo. Half of the primary tumors were collected for RNA extraction. We observed an average of 70% *MALAT1* KD in the primary tumors via quantitative reverse transcription polymerase chain reaction (qRT-PCR), which was also confirmed by single-molecule RNA fluorescence in situ hybridization (FISH). PE150 mRNA sequencing revealed 1,599 differentially expressed genes and 59 alternatively spliced genes upon *MALAT1* KD in the primary tumors. Gene Set Enrichment Analysis (GSEA) revealed that *MALAT1* KD resulted in a

down-regulation of epithelial to mesenchymal transition, hypoxia, and inflammatory response pathways. We are now performing RNA sequencing (RNA-seq) on in vitro NH85TSc PDOs upon *MALAT1* KD. The second PDO line, HCM-CSHL-0366-C50, formed palpable tumors 7 wk following injection of organoids. The mice were then treated with ASOs for 10.5 wk and displayed an average of 86% *MALAT1* KD in the primary tumors via qRT-PCR. However, the mice did not develop any lung metastases, and RNA-seq of the primary tumors revealed *MALAT1* to be the only differentially expressed gene upon *MALAT1* KD. The in vivo data for PDO line HCM-CSHL-0366-C50 matched the initial in vitro RNA-seq results. This indicates that there are potential differences between patient-derived organoid lines that govern whether they are *MALAT1*-sensitive. We are working on identifying these potential differences in signaling pathways or mutation profiles between these PDO lines. This can provide us with some insight into which patient populations might be more suited for *MALAT1* targeted therapy.

Mammary tumor–associated RNA 42 (*MaTAR42*) was found to be up-regulated in MMTV-PyMT and MMTV-Neu-NDL mammary tumors compared to normal mammary epithelial cells. The *MaTAR42* gene is conserved between mouse and human. ASO-mediated knockdown of *MaTAR42* in MMTV-PyMT tumor-derived mammary tumor organoids reduced cell proliferation and the branching phenotype. In addition to the previously identified phenotypes in which *MaTAR42* promoted cellular migration/invasion and cell-extracellular adhesion, we investigated the morphological dynamics of mammospheres derived from *MaTAR42*-overexpressing cells by time-lapse microscopy. We have observed that *MaTAR42*-overexpressing 4T1 cell–derived mammospheres cultured in Matrigel acquired a collective migration/invasion phenotype, which was not observable in control cell–derived mammospheres. To observe subcellular morphological changes resulting from *MaTAR42* overexpression, sections of mammospheres derived from *MaTAR42*-overexpressing or control 4T1 cells cultured in Matrigel were imaged by transmission electron microscopy (TEM). The results indicated that *MaTAR42*-overexpressing mammospheres were hollow structures and exhibited reduced cell–cell adhesion compared to control cells, which is correlated with enhanced cell migration and invasion.

In vivo, *MaTAR42*-overexpressing 4T1 cell–derived xenograft tumors in BALB/c mice exhibited significantly more rapid tumor growth and higher lung metastases. Mechanistically, we previously revealed that *MaTAR42* sensitized cells to epithelial–mesenchymal transition (EMT) induced by TGF-β. In addition, we performed bulk RNA-seq and single-cell RNA-seq from *MaTAR42*-overexpressing 4T1 cell–derived mammospheres. Both single-cell RNA-seq and bulk RNA-seq analysis revealed that ectopic overexpression of *MaTAR42* up-regulated a number of genes that are associated with EMT and hypoxia response. Several hypoxia-induced genes, including *Il33, Car9, Mmp9, Mgp, Il11, Hhpi* and *Slpi*, were found to be up-regulated in *MaTAR42*-overexpressing 4T1 cell–derived mammospheres, and expression of these genes was positively correlated with *MaTAR42* expression at single-cell resolution. Moreover, a dual luciferase assay examining hypoxia-response element–driven promoter activities showed that although in normoxia, ectopic overexpressed *MaTAR42* did not activate a hypoxia-responding transcriptional program by itself, overexpression of *MaTAR42* significantly enhanced hypoxia-driven promoter activation by 3.5-fold. These results indicate a potential role of *MaTAR42* in regulating a hypoxia-induced transcriptomic shift, including EMT. Importantly, expression of the *MaTAR42* human ortholog was positively correlated with that of several hypoxia-induced genes, including *MGP, IL11, HHPI,* and *SLPI*, in various human breast cancer organoid and breast cancer cell lines. To study the molecular mechanisms by which *MaTAR42* regulates the transcriptomic shift, we performed RNA antisense purification coupled with mass spectrometry (RAP-MS) analysis. The results identified Casein Kinase 2 (CK2α) to specifically associate with *MaTAR42*, which was verified by immunoblot. In addition, cellular fractionation and FISH assays indicated a nuclear localization of *MaTAR42*. Because CK2 was previously reported to regulate activities of several transcription factors by phosphorylation, these results suggest that *MaTAR42* may act as a scaffold/chaperone to direct the kinase activity of CK2α in the nucleus for the transcription of particular genes.

Over the past few years, we have developed a breast tumor PDO biobank (see below) with the ultimate goal of identifying potential lncRNA therapeutic targets. From a differential RNA-seq screen, we identified lncRNAs that are overexpressed in these organoid

samples. *Linc01235* was identified as an interesting candidate based on expression level and fold change as compared to normal organoids. When further examined by single-cell RNA-seq analysis, this lincRNA exhibited an interesting expression profile. Analysis of single-cell RNA-seq data demonstrated the PDOs comprise several different cell types, including luminal cells, basal cells, and luminal progenitor cells. *linc01235* was shown to be highly expressed in the luminal progenitor cells in multiple breast tumor organoid lines. To further examine the role of this lncRNA, we selected three PDO lines from our organoid biobank that showed a high expression level of *linc01235*. We found that tumor progenitor cells account for >90% of the cells in these breast tumor organoid lines based on flow cytometry experiments. To begin to assess the function of *linc01235*, we developed ASOs to knock down this lncRNA in these three tumor organoid lines. Preliminary data demonstrated that ASO knockdown resulted in an ~60% reduction in cell proliferation (MTT assay).

To further evaluate the role of *linc01235* in breast cancer progression, the CRISPR-Cas9 system was used to knock out this lncRNA in MDA-MB-468 triple-negative breast cancer (TNBC) cells. To knock out this lncRNA, two guide RNAs were designed to target around the promoter region of *linc01235* to delete the transcription start site. As a result, the expression of *linc01235* was successfully reduced by ~90% compared with the control cell line. Multiple functional assays relevant to cell growth, including cell proliferation, and cell viability assays, were conducted. The knockdown cells showed ~50% reduction in cell proliferation, consistent with ASO-mediated knockdown studies.

By examining co-expression analysis in the TCGA database and in multiple TNBC organoid lines established in our laboratory, we identified a neighboring gene of *linc01235*, *NFIB*, to have high correlative expression ($p < 0.05$). NFIB is a transcription factor and has been reported to be a potential oncogene in TNBC (Liu et al., *J Pathol* **247:** 186 [2019]). Loss of NFIB shows phenotypes similar to those with the loss of *linc01235* in the MDA-MB-468 cell line (Liu et al., *J Pathol* **247:** 186 [2019]; Moon et al., *Mol Oncol* **5:** 538 [2011]). To examine whether NFIB is regulated by *linc01235*, qRT-PCR analysis was performed to determine the expression level of *NFIB* upon ASO knockdown of *linc01235* or CRISPR-Cas9 knockout. We found that a ~60% reduction of *linc01235* led to an ~35%–40% lower expression level of *NFIB*. Importantly, transfection of full-length *linc01235* into the knockout cells rescued the expression of NFIB, demonstrating that *NFIB* is a direct downstream target of *linc01235* and *linc01235* functions in *trans*.

Human Breast Tumor Organoid Project

S. Bhatia, M. Gandhi, P. Naik, S. Russo [in collaboration with K. Kostroff and Arvind Rishi, 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. Thus, they represent an excellent system for identifying new therapeutic targets and for drug development and screening in a patient-specific manner. We performed a comprehensive genomic, transcriptomic, and cellular analysis of 87 breast cancer PDO lines developed in our laboratory, out of which 50 have been fully validated as being tumor-derived and retain key genomic properties of breast cancers, including single-nucleotide variants (SNVs) and copy number alterations (CNAs). We also performed a comprehensive transcriptomic analysis in which we identified signatures associated with different subgroups of breast cancers and identified predominant enrichment of basal-like, proliferative, and luminal progenitor-like signatures associated with TNBC PDOs. Single-cell RNA-seq identified multiple cell types that can be used to study various tumor-specific properties in future studies. This analysis will form the basis for downstream utilization of these PDOs and will serve as a comprehensive resource for studying TNBC.

We have used a panel of 26 cancer therapeutics to screen a cohort of 22 validated PDOs, predominantly composed of TNBC PDOs, and several normal-derived organoids. This panel includes current standard-of-care (SOC) therapies, such as doxorubicin, paclitaxel, 5FU, and cisplatin, and novel targeted agents that were short-listed based on the molecular profile of these tumors, such as afatinib (for *EGFR*-high PDOs), capivasertib (for *PIK3CA*- or *PTEN*-mutated PDOs), and nirogacestat (for *NOTCH1*-active PDOs). The purpose of this study is to assess the extent to which TNBC PDOs can be used for drug testing ex vivo to guide patient treatment. We have found that these PDOs showed a differential response to various

agents, especially to current standard-of-care cytotoxic chemotherapeutics. Using our molecular data, we have identified signatures associated with PDOs that do not respond well to chemotherapy. Specifically, we find signatures of luminal breast cancers and PTEN inactivation to be associated with poor response to chemotherapies, whereas T-cell activation signatures in PDOs seem to be associated with PDOs that respond well. We are currently in the process of validating these in-patient tumor sections for specific markers based on the observed signatures (e.g., CD8 immunolabeling for T-cell signature) and in larger patient population expression databases.

In addition, using our RNA-seq expression data from TNBC and normal PDOs, we performed differential gene expression analysis and we identified approximately 200 lncRNAs that are up-regulated in TNBC PDOs versus normal breast organoids. We cross-referenced this list with lncRNAs found in the basal-like cohort of TCGA breast cancer cases to identify a robust list of intergenic TNBC/basal-specific lncRNA candidates. Over the next year, we will be expanding our in silico analysis to identify gene networks that these TNBC-specific lncRNAs are associated with to gain more insight into their function. Based on this analysis we will further prioritize the candidate list to perform comprehensive functional and molecular analyses on these TNBC-specific lncRNA candidates to evaluate their potential as therapeutic targets to treat TNBCs in a more targeted manner.

Invasive lobular carcinoma (ILC) is the second most common type of breast cancer, representing 10%–15% of diagnosed invasive breast cancers, and it is hormonally driven by estrogen and progesterone receptors (ERs and PRs). We recently initiated a project to assess the role of lncRNAs in ILC using PDOs as a model system. We have received 20 ILC tumor samples for the preparation of PDOs and they are at various stages of expansion. Organoid lines are being taken through a multistep validation pipeline involving phenotypic evaluation, IHC (for hormone receptors ER and PR), genomic profiling including CNV and SNP identification, and transcriptome profiling. Our goal is to identify novel estrogen-responsive lncRNAs and to further investigate their therapeutic potential in ILC. We designed a drug panel consisting of ER signaling modulators based on routinely used drugs for treating ILC patients in the clinical setting. These drugs belonged to three major classes: selective estrogen receptor modulators (SERMs), selective estrogen receptor degraders (SERDs), and aromatase inhibitors (AIs). As a first step, we tested different drug concentrations at multiple time points, validated the down-regulation of ESR1 using qRT-PCR, and chose the lowest possible concentration of each respective drug for treating ILC organoids. Next, as proof of concept, we treated a validated ILC organoid line, LNS055, with the ER signaling modulators anastrozole, exemestane, fulvestrant, and tamoxifen and isolated total RNA for RNA-seq analysis to identify lncRNAs responsive to modulation of estrogen signaling. Using the Galaxy pipeline, we identified 22 lncRNAs that were dysregulated in LNS055 upon ER signaling modulation. Next steps are under way to evaluate additional ILC organoid lines and to consolidate the estrogen-responsive lncRNA signature to be further subjected to a CRISPR screen for functional studies.

Role of the Long Noncoding RNA *Platr4* in Lineage Commitment

R. Hazra, L. Brine [in collaboration with S. Lyons, CSHL; J.E. Wilkinson, University of Michigan Medical School]

We have been investigating the function of a mouse embryonic stem cell (mESC)-specific lncRNA, *Platr4* (pluripotency-associated transcript 4). *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 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. We found that deletion of *Platr4* in ESCs disrupted mesoderm/endoderm lineage specification, specifically the cardiac lineage, while preserving self-renewal and pluripotency. We further showed that *Platr4* is necessary for the activation of the core cardiac gene expression network that induces cardiac transcription factors (Mesp1, Gata4, Tbx5, and Nkx2.5) and initiates EMT by inducing N-cadherin and reducing E-cadherin expression.

To determine the expression of *Platr4* in vivo, we generated a *Platr4*-KO mouse in the C57BL/6J strain

using two single-guide RNAs (sgRNAs) targeting the transcription start site and first exon of *Platr4* as described earlier, and three KO founders were genotyped and used to establish three independent lines. The *Platr4*-KO alleles were backcrossed to C57/BL6 background for t10 generations to yield a pure C57/BL6 genetic background. Heterozygotes intercrossed with wild-type (WT) mice generated WT and *Platr4*-KO littermates. All homozygous KO mice used for breeding were fertile, and they did not show any gross physical abnormalities compared to WT mice. Interestingly, we observed sudden death of young adult and adult mice (40%) in the KO cohort but not in the WT cohort. When we examined the hearts of the young adult dead mice ($N = 3$) and live mice ($N = 10$ for both WT and KO cohorts) by H&E staining of tissue sections, we found 60% of the knockout mice exhibited valve defects with fibrocartilaginous metaplasia, fibro-osseous metaplasia, and mucinous degeneration compared to WT ($N = 10$). In addition, we have found perivascular and myocardial mineralization (60% KO mice, $N = 10$/group) and myocardial atrophy and fibrosis. To further assess heart defects, we performed echocardiography analysis using WT versus KO mice. Echocardiography is used to visualize the cardiovascular structures and measure cardiac function in mice due to its advanced spatial–temporal resolution. Our findings showed a significant 30% decrease in the percentage of fractional shortening (%FS) without altering the percentage ejection fraction (%EF) in KO compared to WT mice. This observation supports our in vitro data of cardiomyocyte dysfunction because %FS indicates changes in left ventricle (LV) chamber size and myocyte contractility. We further demonstrated increased ventricular wall thickness and ventricular mass in KO versus WT mice, indicating ventricular hypertrophy. Interestingly, significant reduction of cardiac output (CO) and heart rate (HR) in KO compared to WT mice may explain our observation of heart failure and sudden death because CO is an important measuring parameter of cardiac dysfunction. Thus, genetic loss of *Platr4* impacts cardiac development both in vitro as well as in vivo.

To explore the transcriptional control of *Platr4* in the ESC state in order to understand its function during differentiation, we analyzed global gene expression profiles using poly(A^+) RNA-seq from control versus *Platr4*-KO ESCs. To this end, a large number of differentially expressed genes (DEGs) were identified between *Platr4*-KO and control ESCs. The top 10 significantly down-regulated genes were validated by qRT-PCR. DAVID GO term pathway analysis in control versus *Platr4*-KO ESCs showed that DEGs were significantly enriched into eight significant pathways, including extracellular matrix (ECM)–receptor interaction, focal adhesion, and PI3K-Akt signaling pathway, each of which plays a vital role in mammalian development. We further performed an RNA-seq experiment using control versus *Platr4*-KO embryoid bodies (EBs), and DEGs were analyzed at different days (D0-D12). DAVID GO term analysis showed DEGs were significantly enriched in eight pathways, including ECM–receptor interaction, focal adhesion, PI3K-Akt signaling pathway, pathways in cancer, proteoglycans in cancer, endocytosis, coagulation cascade, and amoebiasis. Next, customized GSEA analysis showed that DEGs were significantly down-regulated in the mesoderm and endoderm lineages without altering the neuroectoderm lineage, consistent with our previous observation that *Platr4* is critical to mesoderm and endoderm lineage specification. These results indicate that *Platr4* functions in *trans* and interacts with transcription factors (TFs) to regulate the expression level of target genes in ESC state. Therefore, we performed iRegulon (Cytoscape) analysis to predict the potential TFs from the co-expressed gene set of *Platr4* lncRNA. We used a normalized enrichment score (NES) of >3 for the significant enrichment of target TFs. iRegulon interactome maps with TFs using up-regulated and down-regulated DEGs were found.

iRegulon analysis revealed that *Platr4* might exert its function via the regulation of transcription factors of the Tead family and its known downstream target gene, *Ctgf*. First, we found that the expression levels of Tead1, 2, 3, and 4 RNA and protein were equivalent in control and *Platr4*-KO ESCs. Next, we performed small interfering RNA (siRNA) knockdown analysis of each Tead family member and found that they did not affect *Platr4* expression level, although the immediate downstream targets of Tead4 (*Ctgf*, *Gli2*, and *Vgll3*) were significantly down-regulated. Interestingly, we found that *Ctgf* is one of the top five significantly down-regulated genes in RNA-seq data. Furthermore, we found that both the RNA and protein levels of Ctgf were significantly reduced in *Platr4*-KO versus control ESCs. Moreover, ectopic expression of *Platr4* in *Platr4*-KO ESC resulted in a corresponding increase in the mRNA level of *Ctgf*, but not the

level of *Gli2* or *Vgll3*. These results suggest that *Ctgf* is a direct downstream target of *Platr4*, and it functions in *trans*. We next performed CRISPR-Cas9 KD using sgRNAs targeting *Ctgf* in ESC puromycin-resistant cells for downstream functional assays. Both the RNA and protein expression levels of Ctgf in *Ctgf*-KD ESC were verified. Control and *Ctgf*-depleted ESCs were induced to differentiate via withdrawal of leukemia inhibitory factor (LIF) by allowing cells to aggregate into EBs by the hanging drop method. We found that the reduction of *Ctgf* in EBs at day 12 phenocopied the *Platr4*-KO EBs and exhibited a significant reduction in spontaneous contracting EBs: 24% of EBs showed an observable beat at day 12 in control cells compared to 14% in *Ctgf*-KD cells. Notably, ectopic expression of the full-length *Ctgf* gene in *Platr4*-KO ESC increases the percentage of contracting EBs in *Platr4*-depleted cells. Together, these data suggest that *Ctgf* is a potential regulator of cardiomyocyte differentiation and a critical downstream target of *Platr4*.

Ctgf is a direct downstream target of the Tead transcription factor. Because we found significant enrichment of the Tead family in down-regulated *Platr4*-DEGs, we performed RNA immunoprecipitation (RIP) using antibodies against Tead1, 2, and 4 to assess specific interactions with *Platr4*. Interestingly, we found a specific interaction of *Platr4* with Tead4 but none with Tead1, Tead2, and Tead3. Next, using in vitro biotin-RNA pull-down, we further confirmed this specific interaction of Tead4-*Platr4* in ESCs, but not with Tead1 and Tead2. In addition, to examine the regulatory role of Tead4 on Ctgf in ESCs, we performed either ectopic overexpression or siRNA-mediated down-regulation of Tead4 to demonstrate the up-regulation or down-regulation of *Ctgf*, respectively. These results verify that modulation of Tead4 expression level influences the expression level of Ctgf in ESCs, suggesting that the *Platr4*/Tead4 ribonucleoprotein particle (RNP) complex is essential for regulating *Ctgf*. We further performed chromatin immunoprecipitation (ChIP) using a Tead4 antibody, and multiple qPCR primer pairs showed that Tead4 has a high occupancy over the *Ctgf* promoter, as also suggested in other studies. Notably, the occupancy was impaired in *Platr4*-KO ESCs and was able to be restored upon ectopic expression of *Platr4* in *Platr4*-KO ESCs. Together, these results provide compelling evidence that the interaction of Tead4 protein with *Platr4* lncRNA is required for Tead4 to bind to regulatory motifs in the *Ctgf* gene in ESCs. This is the first demonstration that the interaction between a lncRNA and Tead4 is necessary to modulate a downstream target gene and regulate cardiac lineage differentiation.

Random Autosomal Monoallelic Gene Expression and Differentiation

B. Balasooriya

Random autosomal monoallelic (RAM) gene expression may partly explain the variable penetrance of disease-associated genes, 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. Using a previously developed live cell imaging system and nascent RNA-FISH assays, we observed allelic pairing and localization at the nuclear periphery when RAM genes are biallelically expressed in mESCs, but upon differentiation to neural progenitor cells (NPCs) they are monoallelically expressed and only the active allele localizes to the nuclear periphery. Over the past year, using nascent RNA-FISH and ChIP-seq, we found that autosomal monoallelic genes are preferentially expressed at the nuclear periphery. We further identified this nuclear territory as a Sun-associated compartment. By performing ChIP-seq for Sun1, we discovered the presence of a $(TAAA)_n$ DNA motif ($n = 5$ to 8) in most of the monoallelic genes. In our preliminary immunofluorescent assays, we observed that Sun1 colocalized with the $(TAAA)_n$ DNA motif in ESCs and NPCs. To evaluate whether Sun1 acts as a "molecular hook" with respect to the $(TAAA)_n$ motifs, we performed an electromobility shift assay (EMSA) using recombinant SUN1 protein. Forty-two amino acids at the amino terminus of Sun1 are conserved between human SUN1 and mouse Sun1, and this region's molecular function is thus far not identified. Using biotinylated double- and single-stranded $(TAAA)_8$ DNA motif oligos, we found that SUN1 protein does not physically bind to the $(TAAA)_8$ DNA motif. Further, to confirm the above observation, we performed an EMSA assay using nuclear proteins extracted from mouse ESCs and NPCs. In this assay, we observed that a double-stranded $(TAAA)_8$ DNA oligo does not interact with any nuclear protein in vitro, but the sense as well as the antisense DNA motif oligos independently interact with multiple protein

complexes, resulting in multiple biotin bands. In our preliminary immunofluorescent assays performed using cells transfected with a biotinylated $(TAAA)_8$ DNA oligo, we noticed that the double-stranded $(TAAA)_8$ DNA motif also colocalized with Sun1. Therefore, based on these in vitro and in vivo data, we speculate that the allele localization mechanism for monoallelic genes could involve a "DNA unzipping" machinery to facilitate the DNA strand separation, which allows the "molecular hooks" to bind to the DNA motif. Possibly becasue the potential "DNA unzipping" machinery may not be present in the in vitro assay setting, we did not observe any double-stranded motif/protein binding events in vitro. We are continuing our investigation to identify the "molecular hook" that anchors the actively transcribing alleles of monoallelic genes at the Sun1 protein territories.

Using single-cell RNA-seq data from ESCs and cardiac lineage cells, we have constructed maternal and paternal allele-specific histone signature models for deterministic (DeMA) and random (RaMA) autosomal monoallelic genes by in silico data mining approaches. To experimentally validate these allele-specific histone signature models, as a proof of principle, we assessed the four maternal allele–associated histone signatures (H3K27ac, H3K4me3, H3K79me2, and H3K79me3) and two paternal allele–associated histone signatures (H3Kme1 and H3K36me3) in ESCs using a CUT&Tag assay. We constructed sequencing libraries from approximately 200,000 nuclei for each histone signature and sequenced them using the Illumina MiSeq sequencing platform to obtain 300 paired-end reads. We obtained, on average, five million reads per library, which exceeds the required read depth (approximately two million reads) per library constructed using the CUT&Tag approach for chromatin profiling. We analyzed each histone signature–associated chromatin profile using our allele-specific DNA analysis pipeline to obtain maternal and paternal alleles associated with each chromatin signature. By incorporating this histone-specific allele data with DeMA and RaMA genes from ESCs, we demonstrated that 78.4% of the maternal DeMA genes overlapped with the genes enriched for maternal-specific histone signatures and 21% of the paternal DeMA genes overlapped with the genes enriched for paternal-specific histone signatures. In addition, we found that 85.6% of the RaMA genes in ESCs also overlapped with the genes–enriched for RaMA gene–enriched histone signatures.

PUBLICATIONS

Chang K-C, Spector DL. 2021. *MaTAR25*: a long non-coding RNA involved in breast cancer progression. *Mol Cell Oncol* **8:** 2. doi:10.1080/23723556.2021.1882286

Yu AT, Berasain C, Bhatia S, Rivera K, Liu B, Rigo F, Pappin DJ, Spector DL. 2021. *PHAROH* lncRNA regulates Myc translation in hepatocellular carcinoma via sequestering TIAR. *eLife* **10:** e68263. doi:10.7554/eLife.68263

DNA REPLICATION AND CHROMATIN INHERITANCE

B. Stillman K. Bhalla, H-C. Chou, A. Cocozzelli, M. Hossain, Y. Hu, G. McDermott, F. Nasrin, M. Quyang, J. Shapp, Y-J. Sheu, M. Tramantano, N. Zali

Biochemistry of Pre-Replicative Complex Assembly

The focus of the current year of research has remained the investigation of the mechanism and regulation of the initiation of DNA replication in eukaryotes, including using yeast as a model system. In 2004, we started to incorporate cryo-electron microscopy (EM) into the biochemical analysis of pre-replicative complex (pre-RC) assembly; however, at that time cryo-EM produced structures with ~20–30 Å resolution and required creating fusion proteins with suitably large protein tags to map the location of subunits of multiprotein complexes such as the origin recognition complex (ORC). More recently, thanks to advances in cryo-EM technology and software, it is possible to obtain high-resolution structures of protein complexes bound to origins of DNA replication. In collaboration with Huilin Li (formerly at Brookhaven National Laboratory and now at the Van Andel Institute in Michigan) and Christian Speck (a former postdoctoral fellow and now at Imperial College, London), we contributed to determining the structures of many intermediate complexes in the pathway to assemble the pre-RC from purified proteins, including ORC, Cdc6, Mcm2-7, and Cdt1 (see Fig. 1). In the last year we determined the structure of the ORC-Cdc6 complex bound to double-stranded origin DNA that shows that the DNA is bent, poised to be inserted into a channel in the incoming Mcm2-7 hexamer that is chaperoned by Cdt1 after it binds to ORC-Cdc6 (see structure C in Fig. 1). We previously demonstrated that the Mcm2-7 hexamer winged helix domains of Mcm3 and Mcm7 first engage with the winged helix domains of Orc2 and Cdc6 in this structure, and then other winged helix domains in ORC and Mcm2-7 engage each other to create a highly structured protein complex before the DNA is inserted into a channel between the Mcm2 and Mcm5 subunits of the Mcm2-7 hexamer to create the ORC-Cdc6-Cdt1-Mcm2-7 (OCCM) structure. Thus, these studies reported over a number of years have contributed to a detailed understanding of the assembly of pre-RCs at origins of DNA replication in the yeast *Saccharomyces cerevisiae*. A number of laboratories, including the collaboration mentioned above and the laboratories of Alessandro Costa and John Diffley (Francis Crick Institute) and the laboratories of Bik-Kwoon Tye and Ning Gao (University of Hong Kong and Peking University in Beijing) have also contributed structures, including the most recent M-O intermediate (structure G) in Figure 1.

As described in previous reports, these structures, including our structure of the OCCM on origin DNA published in 2017 (complex E in figure) revealed how ORC from *S. cerevisiae* recognizes the replication origin in a DNA sequence–specific manner. We demonstrated that an α-helix in the Orc4 subunit of ORC was responsible in part for recognizing specific base-pairs in a major groove of the origin DNA. Only a small clade of budding yeasts that include *S. cerevisiae* and *Kluyveromyces lactis* have this α-helix insert in Orc4 and as shown by collaborative studies with Leemor Joshua-Tor and her colleagues, the structure of ORC from human cells only has a short protein loop instead of the α-helix. Amino-acid sequence comparisons of Orc4 from many other budding yeasts, other fungi, all animals, and all plants showed that Orc4 from these species also lacks the α-helix, suggesting that origins of DNA replication are not DNA sequence–specific. Indeed, it is known that ORC for human and *Drosophila* bind to DNA in a DNA sequence–independent manner. Therefore, how they determine the location of origins of DNA replication, or more specifically how ORC interacts with the chromosome and localizes starts sites for DNA replication, is not

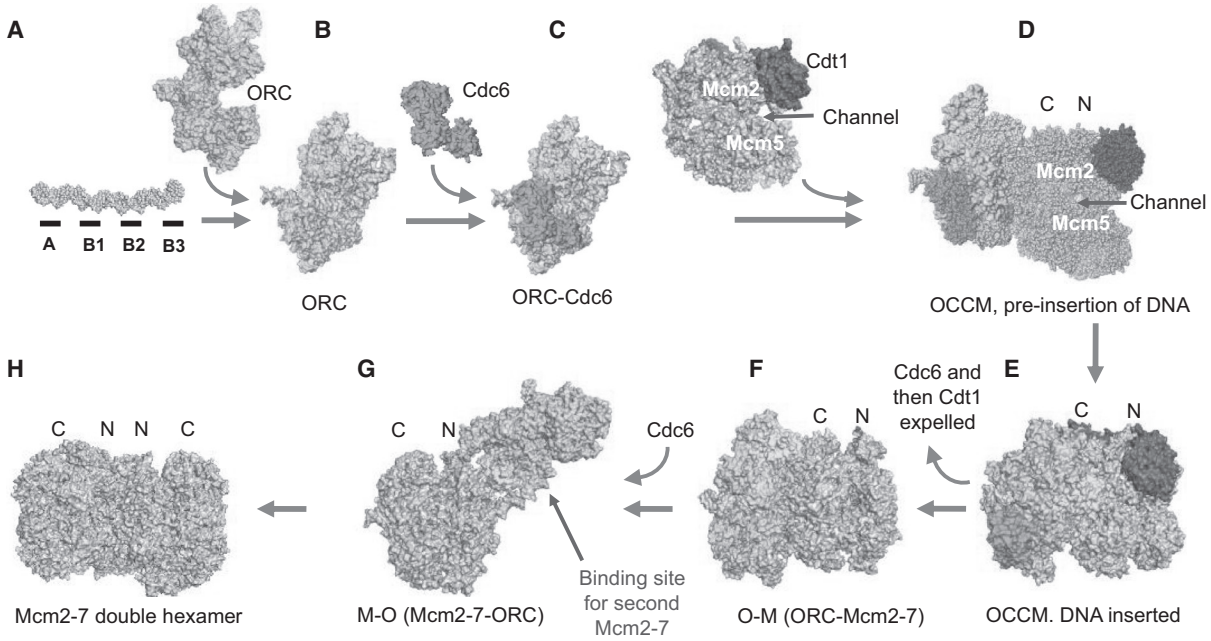


Figure 1. Pre-replicative complex assembly requires a seven-step process. (*A*) The genetic sequence that acts as the starting point for replication in *Saccharomyces cerevisiae* DNA consists of four elements (black segments), with the A and B2 elements binding ORC in opposite orientations. (*B*) ORC (lime green) first binds to the A and B1 elements. (*C*) ORC recruits Cdc6 (green). (*D*) ORC-Cdc6 recruits the Cdt1-Mcm2-7 complex (Cdt1 [blue] and Mcm2-7 [light blue]; Mcm2 and Mcm5 are parts of the six proteins in the complex) so that the DNA is aligned to a channel in the Mcm2-7 hexamer. The Mcm2-7 complex is oriented based on the structure of the protein components; with a carboxy-terminus (or extremity; Mcm-C or "C") at one end, and an amino-terminus (Mcm-N or "N") at the other. At this stage, Mcm-C binds ORC-Cdc6, (*E*) creating the MO complex. As a result, the ORC is now orientated toward the other end of the Mcm2-7-N. (*F*) ORC can now recruit a second Cdc6, creating a binding site for a second Cdt1-Mcm2-7 complex that is loaded in an orientation opposite to the first Mcm2-7. (From Stillman 2022.)

understood. We are continuing to study ORC and CDC6 from human cells and their interaction with DNA and with chromatin modifying proteins.

The *S. cerevisiae* and *K. lactis* clade of budding yeasts have a dense genome with ~70% of the genome encoding protein, with few introns and little repetitive DNA. To avoid conflicts between replication and transcription, it makes sense to place origins of DNA replication in intergenic regions, and this may well be the reason why the acquisition of DNA sequence–specific origins of DNA replication in these yeasts evolved. Locating pre-RCs to intergenic regions would allow the initiation of DNA replication to occur at sites that do not compete with transcription of genes, thereby limiting replication-transcription conflicts that can cause genome instability. However, many budding yeasts lack the α-helix insert in Orc4 and one example is the budding yeast *Yarrowia lipolytica*. This yeast has been used by industry for large-scale production of protein, small-molecule organic compounds, and fine chemicals, as well as for the degradation of crude oil and hydrocarbons. Not much is known about how the six chromosomes that make up the *Y. lipolytica* genome are replicated. A few autonomously replicating sequences (ARSs) that allow replication of a small circular plasmid have been isolated and these genetic elements consist of physically separatable centromere (*CEN*) and replication origin (*ori*) sequences. Both the *CEN* and *ori* sequences are required to maintain replicating plasmids, but only a limited number of *ori* sequences that are associated with *CEN* sequences have been characterized to date and they do not appear to be DNA sequence–specific. Thus, we have embarked on a project to compare and contrast the mechanism and control of the initiation of DNA replication in *S. cerevisiae* that has sequence-specific origins with origins of replication in *Y. lipolytica* and human cells.

Toward this end, we have, in collaboration with Doreen Ware's laboratory, sequenced the genomes of both the Mat A and Mat B strains that are derived from original isolates from France using both long-read nanopore sequencing and Illumina short-read sequencing to obtain a near-complete *Yarrowia* genome. The *Y. lipolytica* genome is 1.6 times larger than the *S. cerevisiae* genome but has roughly the same number of protein-coding genes. Thus, the predicted increase in intergenic DNA sequences may provide a greater probability for DNA sequence–independent initiation to be located away from protein-coding regions in the genome. We are currently investigating the mechanism of pre-RC assembly in this yeast with the view that it may be a model for higher eukaryotes such as human cells.

DNA Replication in Human Cells

We continue to study the mechanism of DNA replication and its control in human cells, and last year reported that ORC and CDC6 from human cells have intrinsically disordered regions (IDRs) that bind a number of regulatory proteins—including protein phosphatases and protein kinases. We have also compared and contrasted the replication of DNA in normal and cancer cells and in the past have identified the DEAD-box RNA helicase enzyme DDX5 as a regulator of DNA replication in some cancer cells where is it essential, but interestingly it is not essential in other cancer cells and in normal, diploid human fibroblast cells. Previous reports from this laboratory have shown that DDX5 is also essential for the proliferation of acute myeloid leukemia cells in the mouse, but DDX5 inhibition in normal tissues in the entire animal is well tolerated and thus DDX5 is a potential target for cancer or leukemia therapy.

More recently, we investigated why DDX5 is essential in some cells and yet is dispensable in others, including normal human cells. We have shown previously that DDX5 occupies the E2F1-regulated promoter of genes required for DNA replication where RNA polymerase II loading and E2F1 is controlled by the transcriptional repressor retinoblastoma protein (RB). However, a broad range of cell cycle–regulated genes, particularly those that are required for mitosis, are controlled by a separate multiprotein complex known as the DREAM complex (the dimerization partner [DP], retinoblastoma [RB]-like, E2F, and MuvB). The DREAM complex was originally identified in *Drosophila melanogaster* with a role in transcriptional repression. Its composition is regulated differently at distinct phases of the cell cycle. The core DREAM complex is composed of Lin9, Lin37, Lin54, Lin52, and RbAp48 (the human homologs of *Drosophila* Mip130, Mip40, Mip120, dLin52, and Caf1p55). The DREAM complex is involved in regulation of genes that bind the E2F4 transcription factor during the G_0/G_1 phase of the cell cycle and the B-myb/FOXM1 transcription factors during in S/G_2. However, these core DREAM complex proteins can associate with the retinoblastoma (RB)-related proteins p130 and p107 to form a repressive transcription complex. Many cell cycle–regulated genes are repressed by both pRB-E2F1 and p130 containing DREAM core complex.

Our analysis of the effects on gene expression by RNA sequencing following depletion of DDX5 from cells in which it is either essential or dispensable has uncovered a new role for DDX5 in cancer cells in regulation of both G1/S- and G_2/M-specific cell cycle genes for cell proliferation in the subset of breast and colon cancer. Genes that were down-regulated upon DDX5 depletion in cells that require DDX5 for cell proliferation were predominantly involved in G_1-S phase transition or in the transition from G_2 to mitosis. We then showed that DDX5 interacts with the p130 transcriptional repressor protein and disrupts the interaction between p130 and the core DREAM complex proteins. These results suggested that DDX5 binds to p130, thereby relieving repression of genes that are regulated by the DREAM complex. These results show certain cancer cells that have acquired a dependence on DDX5 do so to counteract the repressive p130-DREAM complex. Interestingly, we demonstrated that in cells that can proliferate in the absence of DDX5, the levels of p130 are low, suggesting that they do not need to inhibit p130 and that the transcriptionally active DREAM complexed are constitutively active.

PUBLICATIONS

Chou H-C, Bhalla K, Demerdesh OE, Klingbeil O, Hanington K, Aganezov S, Andrews P, Alsudani H, Chang K, Vakoc CR, et al. 2021. The human Origin Recognition Complex is essential for pre-RC assembly, mitosis and maintenance of nuclear structure. *eLife* **10:** e61797. doi:10.7554/eLife.61797

Feng X, Noguchi Y, Barbon M, Stillman B, Speck C, Li H. 2021. The structure of ORC–Cdc6 on an origin DNA reveals the mechanism of ORC activation by the replication initiator. *Nat Commun* **12:** 3883. doi:10.1038/s41467-021-24199-1

Hossain M, Bhalla K, Stillman B. 2021. Multiple, short protein binding motifs in ORC1 and CDC6 control the initiation of DNA replication. *Mol Cell* **81:** 1951–1969. doi:10.1016/j.molcel.2021.03.003

Sheu Y-J, Karakida R, Gillis J, Stillman B. 2021. Prevalent and dynamic binding of the cell cycle checkpoint kinase Rad53 to gene promoters. bioRxiv doi:10.1101/2021.05.23.445333

Stillman B, Matos Rodríguez FV. 2021. Putting innovation in our educational DNA. Op ed. *New York Daily News*. September 24, 2021. https://www.nydailynews.com/opinion/ny-oped-putting-innovation-in-our-educational-dna-20210924-t6hrhx27b5etjitpxlzhwxch2q-story.html

In Press

Stillman B. 2022. Replication: the remarkable gymnastics of ORC. *eLife* **11:** e76475. doi:10.7554/eLife.76475

CHARACTERIZING AND TARGETING ONCOGENIC TRANSCRIPTION FACTOR COMPLEXES

C.R. Vakoc A. Alpsoy, P. Cunniff, S. Espinosa, C. Fitzpatrick, Y. Gao, Y. Huang, V. Kechejian, O. Klingbeil, P. Kumar, J. Liverpool, C. Lopez-Cleary, D. Maia-Silva, J. Milazzo, B. Nalbant, S. Polyanskaya, D. Skopelitis, M. Sroka, A. Vasudevan, Y. Wei, X. Wu, Z. Yang, T. Yoshimoto

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/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 transcription factors 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.

Transcription Factor Dependencies in Lung Cancer

A. Alpsoy, S. Espinosa, C. Fitzpatrick, P. Kumar, Y. Huang, D. Skopelitis, 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 transcription factor 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, Rb2* in the normal tuft cell lineage of the mouse lung using a Cre knock-in 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 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*, which functions as a POU2F3 cofactor in tuft cell lung cancer. Ongoing mechanistic

studies aim to define the detailed mechanism that connects *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.

We recently initiated a project that seeks to investigate the function of NRF2 as a transcription factor dependency in non-small-cell lung cancer.

Novel Dependencies in Acute Myeloid Leukemia

S. Polyanskaya, Y. Wei, Z. Yang

An enhanced requirement for extracellular nutrients is a hallmark property of cancer cells. We optimized an in vivo genetic screening strategy for evaluating dependencies in acute myeloid leukemia (AML), which led to the identification of the myo-inositol transporter SLC5A3 as a unique vulnerability in this disease. In accord with this transport function, we demonstrate that the SLC5A3 dependency reflects a myo-inositol auxotrophy in AML. Importantly, the commonality among SLC5A3-dependent AML lines is the transcriptional silencing of *ISYNA1*, which encodes the rate-limiting enzyme for myo-inositol biosynthesis, inositol-3-phosphate synthase 1. We used gain- and loss-of-function experiments to demonstrate a synthetic lethal genetic interaction between *ISYNA1* and *SLC5A3* in AML, which function redundantly to sustain intracellular myo-inositol. Transcriptional silencing and DNA hypermethylation of *ISYNA1* occur in a recurrent manner in human AML patient samples, in association with the presence of *IDH1*/*IDH2* and *CEBPA* mutations. Collectively, our findings reveal myo-inositol auxotrophy as a novel form of metabolic dysregulation in AML, which is caused by the aberrant silencing of a biosynthetic enzyme.

Hundreds of genes become aberrantly silenced in AML, with most of these epigenetic changes being of unknown functional consequence. Here, we demonstrate how gene silencing can lead to an acquired dependency on the DNA repair machinery in AML. We make this observation by profiling the essentiality of the ubiquitination machinery in cancer cell lines using domain-focused CRISPR screening, which revealed Fanconi anemia (FA) proteins UBE2T and FANCL as unique dependencies in AML. We demonstrate that these dependencies are due to a synthetic lethal interaction between FA proteins and aldehyde dehydrogenase 2 (ALDH2), which function in parallel pathways to counteract the genotoxicity of endogenous aldehydes. We show that DNA hypermethylation and silencing of *ALDH2* occur in a recurrent manner in human AML, which is sufficient to confer FA pathway dependency. Our study suggests that targeting of the ubiquitination reaction catalyzed by FA proteins can eliminate ALDH2-deficient AML.

AML cells rely on phospho-signaling pathways to gain unlimited proliferation potential. Here, we use domain-focused CRISPR screening and identify the nuclear phosphatase SCP4 as a dependency in AML, whereas this enzyme is dispensable in normal hematopoietic progenitor cells. Using CRISPR exon scanning and gene complementation assays, we found that the catalytic function of SCP4 is essential in AML. Through mass spectrometry analysis of affinity-purified complexes, we identify the kinase paralogs STK35 and PDIK1L as binding partners and substrates of the SCP4 phosphatase domain. We show that STK35 and PDIK1L function catalytically and redundantly in the same pathway as SCP4 to maintain AML proliferation and to support amino acid biosynthesis and transport. We obtained evidence that SCP4 regulates STK35/PDIK1L through two distinct mechanisms: catalytic removal of inhibitory phosphorylation and by promoting kinase stability. Our findings reveal a phosphatase-kinase signaling complex that supports the pathogenesis of AML.

Pediatric Sarcoma

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

Alveolar rhabdomyosarcoma (aRMS) is a rare muscle cancer that affects primarily children and adolescents. Although the disease bears a low overall mutational burden, >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 to the transactivation domain of FOXO1. Fusion-positive tumors are the most aggressive, with 4-yr 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 challenges associated with designing drugs that target transcription factors. 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, fluorescence-activated cell sorting (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.

The EWS-FLI1 fusion in Ewing sarcoma is also of interest to our laboratory. 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 has a powerful lineage dependency in rhabdomyosarcoma cells. The EWS-FLI1 fusion oncoprotein deregulates transcription to initiate the pediatric cancer Ewing sarcoma. Here, we used a domain-focused CRISPR screen to implicate the transcriptional repressor ETV6 as a unique dependency in this tumor. Using biochemical assays and epigenomics, we show that ETV6 competes with EWS-FLI1 for binding to select DNA elements enriched for short GGAA repeat sequences. Upon inactivating ETV6, EWS-FLI1 overtakes and hyperactivates these *cis*-elements to promote mesenchymal differentiation, with *SOX11* and *NRTK1* being key downstream targets. We show that squelching of ETV6 with a dominant-interfering peptide phenocopies these effects and suppresses Ewing sarcoma growth in vivo. These findings reveal a strategy for neutralizing the EWS-FLI1 oncoprotein by reprogramming its genomic occupancy.

Pancreatic Cancer

P. Cunniff, O. Klingbeil, D. Maia-Silva

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 single guide RNAs (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. Two other targets we are investigating are p63 and KLF5 using marker-based genetic screens to reveal critical interaction partners.

PUBLICATIONS

Chou HC, Bhalla K, Demerdesh OE, Klingbeil O, Hanington K, Aganezov S, Andrews P, Alsudani H, Chang K, Vakoc CR, et al. 2021. The human origin recognition complex is essential for pre-RC assembly, mitosis, and maintenance of nuclear structure. *Elife* **10:** e61797. doi:10.7554/eLife.61797

Fiskus W, Mill CP, Nabet B, Perera D, Birdwell C, Manshouri T, Lara B, Kadia TM, DiNardo C, Takahashi K, et al. 2021. Superior efficacy of co-targeting GFI1/KDM1A and BRD4 against AML and post-MPN secondary AML cells. *Blood Cancer J* **11:** 98. doi:10.1038/s41408-021-00487-3

Hutton ER, Vakoc CR, Siepel A. 2021. ACE: a probabilistic model for characterizing gene-level essentiality in CRISPR screens. *Genome Biol* **22:** 278. doi:10.1186/s13059-021-02491-z

Laubscher D, Gryder BE, Sunkel BD, Andresson T, Wachtel M, Das S, Roschitzki B, Wolski W, Wu XS, Chou HC, et al. 2021. BAF complexes drive proliferation and block myogenic differentiation in fusion-positive rhabdomyosarcoma. *Nat Commun* **12:** 6924. doi:10.1038/s41467-021-27176-w

Sroka MW, Vakoc CR. 2021. An epigenetic tipping point in cancer comes under the microscope. *Nature* **590:** 399–400. doi:10.1038/d41586-021-00002-5

Takao S, Forbes L, Uni M, Cheng S, Pineda JMB, Tarumoto Y, Cifani P, Minuesa G, Chen C, Kharas MG, et al. 2021. Convergent organization of aberrant MYB complex controls oncogenic gene expression in acute myeloid leukemia. *eLife* **10:** e65905. doi:10.7554/eLife.65905

Yang Z, Wu XS, Wei Y, Polyanskaya SA, Iyer SV, Jung M, Lach FP, Adelman ER, Klingbeil O, Milazzo JP, et al. 2021. Transcriptional silencing of *ALDH2* confers a dependency on Fanconi anemia proteins in acute myeloid leukemia. *Cancer Discov* **11:** 2300–2315. doi:10.1158/2159-8290.CD-20-1542

CANCER: GENETICS

Jeff Boyd's research interests have long been in somatic molecular genetics, inherited genetics, and genomics of gynecologic (endometrial and ovarian) and breast cancers. Current areas of focus include elucidating the early natural history of epithelial ovarian carcinoma using combined histopathologic and molecular genetic techniques. The histogenesis of ovarian carcinoma remains very poorly understood, which is remarkable for a relatively common solid tumor type. With the recent advent of technology allowing for the interrogation of cancer "omics," a related research goal is to perform a comprehensive genomic characterization of putative ovarian cancer stem/progenitor cells. With respect to breast cancer, the Boyd laboratory focused on two novel hypotheses. The first is that primary breast cancers do not metastasize (linear model), but rather that primary and metastatic breast cancer lesions arise concomitantly from a small population of breast cancer stem cells (parallel model). This hypothesis is based on epidemiologic and clinical data, but awaits rigorous testing in the laboratory. Second is the hypothesis that a majority of breast cancers occur in a genetically susceptible minority of the population. This hypothesis is also based on epidemiologic and clinical data, and requires a polygenic basis for "genetic susceptibility" in this context. Finding the right laboratory and computational approach(es) to this hypothesis is a great challenge, but a tractable one. Finally, the Boyd group is exploiting the advent of next-generation sequencing technology and Dr. Boyd's career-long desire to improve cancer patient care by creating a clinical molecular diagnostics program (Center for Genomic Medicine) at our clinical affiliate, the Northwell Health Cancer Institute.

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 showed subtle differences in

cell–cell/extracellular matrix (ECM) signaling and gene expression genome-wide in situ. By clustering transcripts into functionally or morphologically discrete regions, the laboratory found 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. The laboratory uses in situ sequencing, cell lineage tracing, and single-cell profiling to understand how noncoding RNA affects tumor cell evolution in their native context. The long-term goal is to develop better tumor classification tools and anticancer therapeutics using an 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. Mills 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 Mills 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 which 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 exhibit 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. The **Jason Sheltzer** laboratory's work has demonstrated the existence of several surprising phenotypes that are shared among cells with different chromosomal imbalances. They demonstrated 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. They 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 demonstrated that aneuploidy can paradoxically function as a barrier to tumor growth. They are currently continuing their investigation of the role of aneuploidy in cancer. They are also applying CRISPR-Cas9-mediated genome engineering to develop novel mouse models in order to explore 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, the Sheltzer laboratory hopes 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 exhibiting 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 a 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 CELL FATE SPECIFICATION DURING ANIMAL DEVELOPMENT

C.M. Hammell K. Hills-Muckey N. Stec
B. Kinney J. Wang
H. Ouyang

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 key regulatory proteins are established. One half of the laboratory studies how oscillatory gene expression during the development of *Caenorhabditis elegans* larvae contributes to the establishment and stability of postembryonic cell fate specification. The second half of the laboratory studies how protein translation is regulated by controlling mTORC1 signaling.

Dynamic Patterns of Transcription Are Paused during Starvation and BLMP-1 and ELT-3 Are Required for the Resumption of Normal Development When Food Is Available

N. Stec, K. Hills-Muckey

Development of *C. elegans* larvae in laboratory conditions is both rapid and continuous because of an abundant, controlled food supply and standardized growth conditions. In contrast, the natural environment in which *C. elegans* larvae normally live is far more diverse and dynamic, leading to a more punctuated developmental trajectory. Acute food removal induces a defined developmental diapause that can occur immediately after each larval molt. The diapause is part of a developmental checkpoint that is distinct from dauer development and controlled by insulin signaling. To determine whether global aspects of cyclical transcription are altered during acute starvation, we monitored the expression of four transcriptional reporters that exhibit distinct phases of expression. These reporters include targets of BLMP-1 and ELT-3 (*(minCE)gst-5::GFP-pest*, *ZK180.5::GFP-pest,* and *mlt-10::GFP-pest*) as well as a transcriptional reporter of *blmp-1,* because chromatin immunoprecipitation sequencing (ChIP-seq) data indicates that it is expressed in a cyclical expression pattern, and BLMP-1 activity regulates its own expression through direct positive feedback. We grew populations of L1-staged animals under nutrient-rich conditions for 22 h and then one-half of each sample was allowed to continue development in nutrient-replete conditions while the other half was acutely starved (Fig. 1A). Acute food removal results in >90% of transgenic animals arresting development with undivided vulval precursor cells after 24 h, indicating that larval development has been paused. During the starvation period, we monitored green fluorescent protein (GFP)-pest expression in fed and starved populations. In contrast to the cyclical expression of each reporter under nutrient-replete conditions, GFP-pest expression for each of the reporters was extinguished during starvation conditions with similar kinetics (Fig. 1B). This transcriptionally inactive state was maintained throughout the starvation period. Re-fed animals resumed normal, cyclical expression patterns within a few hours after food exposure (Fig. 1B). These experiments demonstrate that oscillatory transcription is inhibited by the starvation-induced developmental diapause and that wild-type animals can re-initiate these transcriptional patterns in a nutrient-dependent manner.

To examine whether BLMP-1-dependent chromatin remodeling is also modulated by nutrient-dependent signaling, we monitored compaction of *lin-4* locus in animals that have entered the acute starvation–dependent developmental diapause. At the mid-L3 stage (first vulval precursor cell [VPC] division), a developmental time point that precedes *lin-4* transcription during the L3 stage, the *lin-4* locus was decompacted (Fig. 1C). We then rapidly removed the bacteria food source to induce the L4-specific developmental diapause. After 22 h of

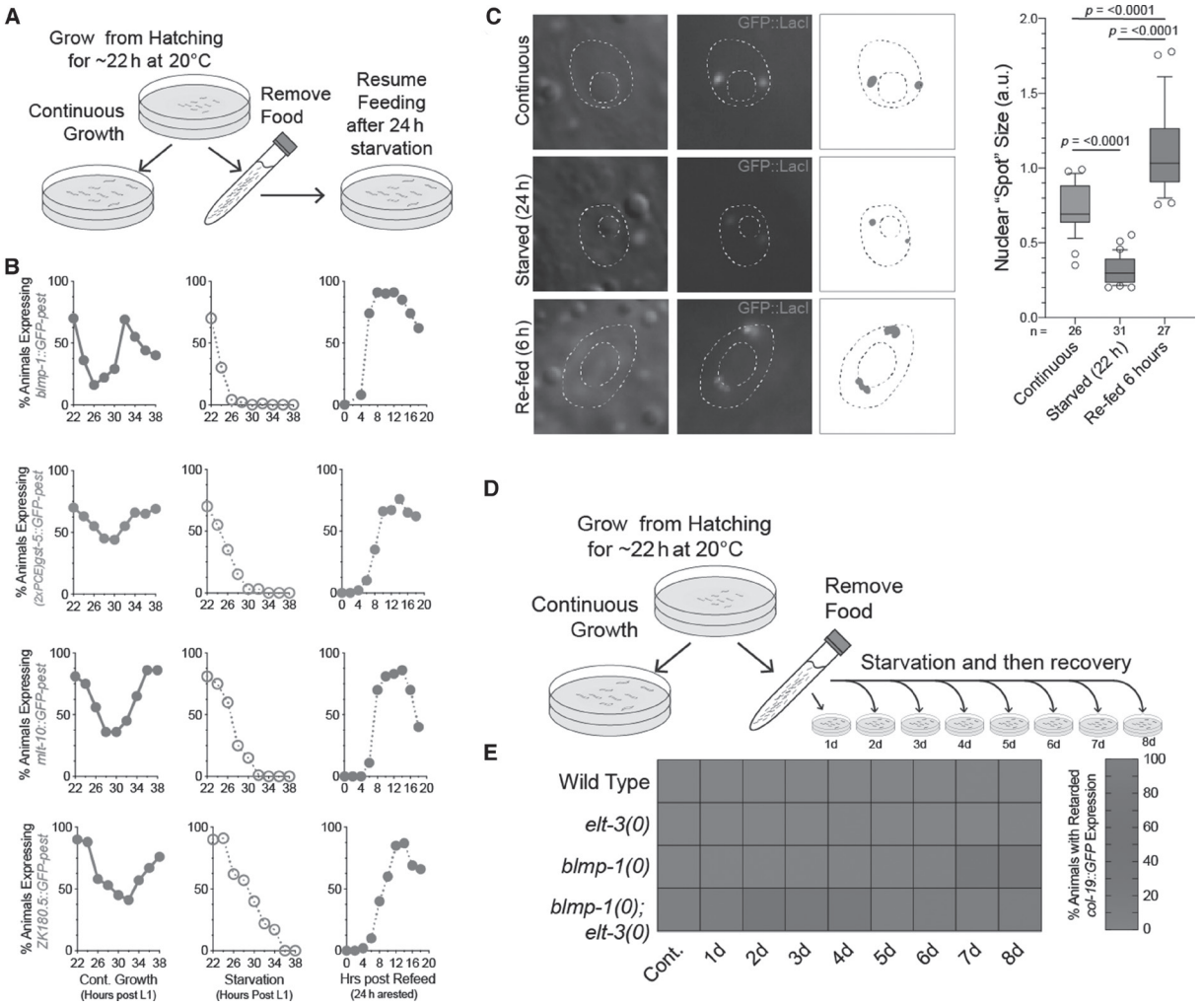


Figure 1. Dynamic gene expression is halted during nutrition-mediated developmental arrest, and BLMP-1 and ELT-3 are essential for the recovery of normal temporal patterning after starvation. (*A*) A schematic of the starvation experiment used to measure nutrient-dependent changes in the expression patterns of cyclically expressed genes. (*B*) Continuous cycling of gene expression is arrested during starvation conditions and is reinitiated in a coordinated manner when animals resume development. For each time point, 70–200 animals were scored. (*C*) The *lin-4* locus is compacted during starvation conditions in lateral seam cells. (*D*) A schematic of the starvation-mediated arrest and refeeding experiment used to measure the ability of animals to resume normal temporal patterning after starvation. (*E*) The resumption of normal temporal patterning after starvation. The resumption of normal temporal patterning after starvation requires *blmp-1* and *elt-3* activity. Defects in temporal development were monitored by measuring the *col-19::GFP* expression phenotypes in refed animals after they resumed growth for 24 h after the indicated period of starvation. ($n \geq 100$ per condition.)

starvation, >90% of animals arrested at the invagination stage of vulval development (L4.1-L4.2), which precedes fusion between the anchor cell and VPCs on the ventral surface of the L4 larva. Analysis of lateral seam cell GFP-LacI puffs sizes in starved animals demonstrated that the *lin-4* loci in seam cells were dramatically compacted when compared to the same loci in continuously developing animals (Fig. 1C). Replating of these starved animals led to the rapid decompaction of the *lin-4* locus at a time point when BLMP-1 expression had resumed (Fig. 1C). Interestingly, the level of decompaction is greater in recovering animals than those found in nonstarved, continuously developing animals (Fig. 1C). We also monitored *lin-4::mCherry*/LacO loci in the seam cells of starved, L1-stage animals and also found that loci compacted in these starvation

conditions. In sum, these experiments indicate that dynamic chromatin rearrangements are modulated by nutrient sensing.

Changes in BLMP-1 expression and the dramatic compaction and decompaction of BLMP-1 target chromatin during starvation and recovery, respectively, suggests that BLMP-1 could play an essential role in modulating transcriptional output of critical developmental genes in dynamic environments. To determine whether animals that lack *blmp-1* and/or *elt-3* can robustly adapt to starvation, we repeated the food experiments as described in Figure 1D but maintained the animals in starvation conditions for increasing periods of time. At the end of the starvation period, animals were replated on normal food and monitored for developmental phenotypes by scoring changes in *col-19::GFP* expression when the animals had reached adulthood (Fig. 1D). In these experiments, wild-type and *elt-3(0)* larvae re-initiate normal temporal patterning after prolonged bouts of starvation (Fig. 1D). In contrast, a minor fraction of *blmp-1(0)* mutants failed to express *col-19::GFP* in hyp7 cells, a phenotype associated with weak reiterative heterochronic mutants. The relatively mild *blmp-1(0)* phenotypes were dramatically enhanced in *blmp-1(0); elt-3(0)* double mutants. These results indicate that BLMP-1 and ELT-3 are essential for resuming normal temporal patterning after continuous development has been interrupted by starvation.

The *lin-4* MicroRNA Regulates the Expression of Multiple Targets at Sequential Stages of Postembryonic Development

K. Hills-Muckey

A majority of our work on the developmental clock has focused on studying phenotypes that are associated with the trans-acting DNA-binding proteins that control transcriptional processes. In the last year, we shifted focus to making various mutations in the *cis*-regulatory architecture of genes that are directly regulated by clock components. We initially focused our attention on conserved regulatory elements that are important for driving cyclical transcription of the *lin-4* microRNA gene. This gene was chosen because mutations result in clear, quantifiable developmental phenotypes, and there were highly conserved regions upstream of the *lin-4* gene that are likely important for transcriptional control. To perform a structure/function analysis of the lin-4 regulatory regions, we targeted a number of engineered versions of the *lin-4* locus to chromosome I using CRISPR in the context of animals that lack the *lin-4* gene. This analysis enabled us to determine that the *lin-4* gene is regulated by two partially redundant enhancer elements (Fig. 2A–C). We used a variety of genetic approaches to determine that the distal enhancer element was the dominant enhancer that drives optimal levels of the lin-4 transcript. RNAi depletion of *ain-1*, encoding a nonessential component of the micro RNA–induced silencing complex (miRISC), has little effect on the temporal development of wild-type worms but dramatically enhances retarded developmental phenotypes in animals lacking the BLMP-1-bound PCE (pioneer control enhancer). Specifically, ΔPCE animals exposed to *ain-1*(RNAi) fail to express the adult-specific *col-19::GFP* reporter after the L4 stage (Fig. 2D,E). Surprisingly, ain-1(RNAi) causes ΔPCE mutant animals to reiterate the L2 stage as measured by quantification of cell lineage programs of the lateral hypodermal blast cells (Fig. 2F,G). This indicates that *lin-4* likely regulates the expression of additional downstream temporal targets in the L2 in addition to regulating LIN-14 expression in the L1 stage. We are currently exploring the idea that *lin-4* regulates the expression of hbl-1 via complementary sites in its 3′ UTR.

Development of a Controllable Degradation System to Study the Spatial and Temporal Functions of Target Genes

K. Hills-Muckey

A large portion of our molecular and genetic approaches utilizes mutations that reduce the activity of particular genes of interest. This approach is straightforward for nonessential genes that disrupt development but do not lead to animal lethality. In a number of systems, the auxin-degradation system has been developed to specifically target the destruction of virtually any protein of interest. This system relies on the heterologous expression of a single *Arabidopsis thaliana* F-box protein, transport inhibitor response 1 ($_{At}$TIR1). In this system, exogenous auxin (indole-3-acetic acid; IAA) enhances the ability of $_{At}$TIR1 to function as a substrate recognition component that

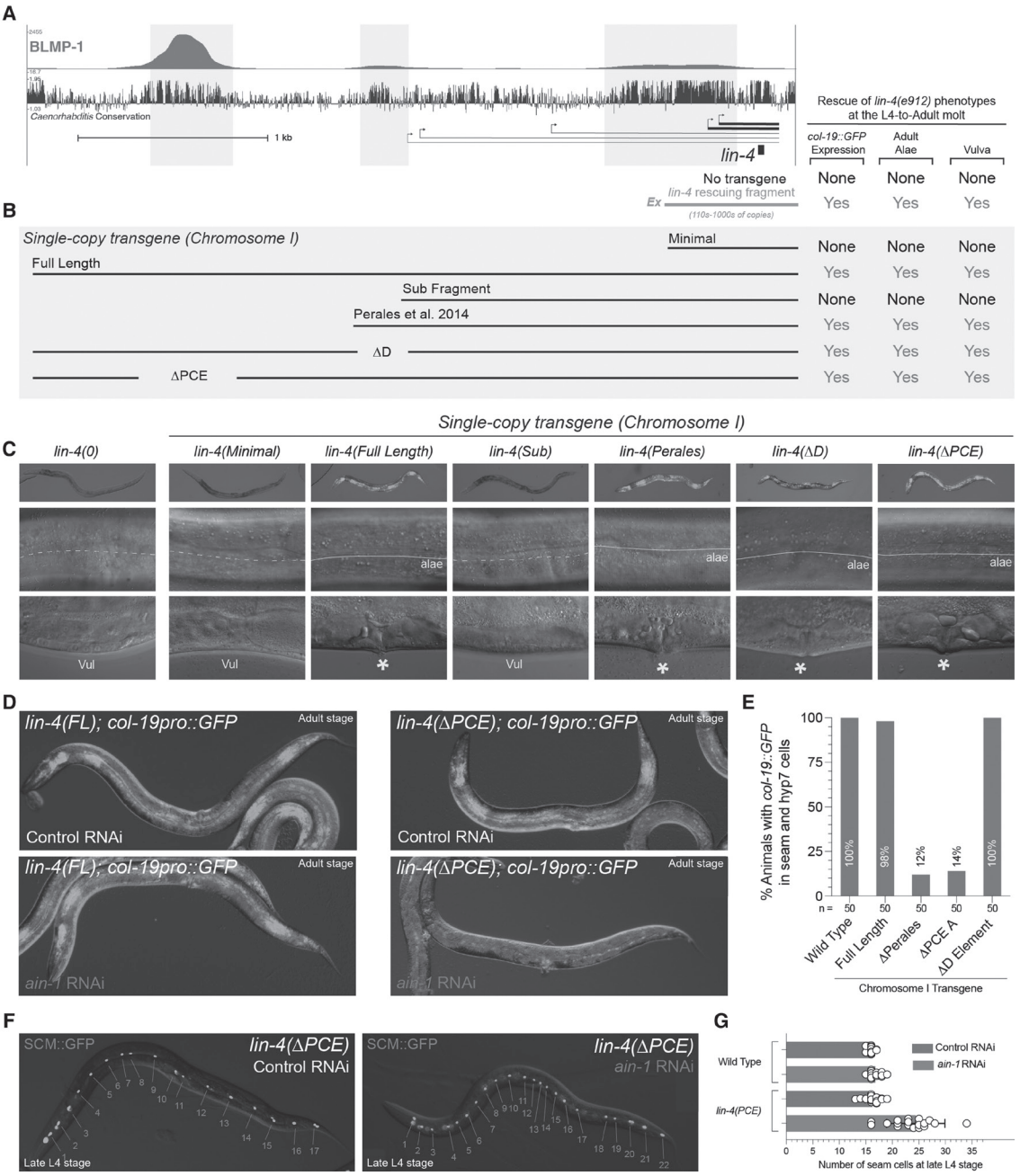


Figure 2. lin-4 expression is regulated by two partially redundant enhancer elements and the PCE element functions to regulate L2-specific cell division programs. (*A*) A schematic diagram of the lin-4 locus illustrating the major BLMP-1 binding sites (light blue), the regions conserved with other nematode species, and the five major lin-4 primary transcripts. (*B*) Diagram of the various rescuing transgenes as well as the ability of these rescuing transgenes to correct the *col-19::GFP*, adult alae, and vulval phenotypes of *lin-4(0), lin-4(e912),* developmental phenotypes. (*C*) Pictographs depicting the col-19:GFP, adult alae, and vulval phenoytpes of individual animals expressing the indicated rescuing transgene. (*D*) *col-19::GFP* expression is defective in animals that lack the PCE element of the lin-4 promoter when miRNA activity is compromised via RNAi of *ain-1*. (*E*) Quantification of the col-19::GFP mis-expression phenotypes of various lin-4 rescuing constructs. (*F*) *lin-4(PCE); ain-1(RNAi)* animals harbor additional skin stem cells (seam cells) compared to wild type. (*G*) Quantification of the seam cell numbers of wild type or *lin-4(PCE)* exposed to control RNAi or *ain-1(RNAi).* The excessive number of seam cells in *lin-4(PCE); ain-1(RNAi)* animals is indicative of a reiteration of L2-specific seam cell division programs. (Perales et al., *PLoS Genet* **10:** e1004486 [2014].)

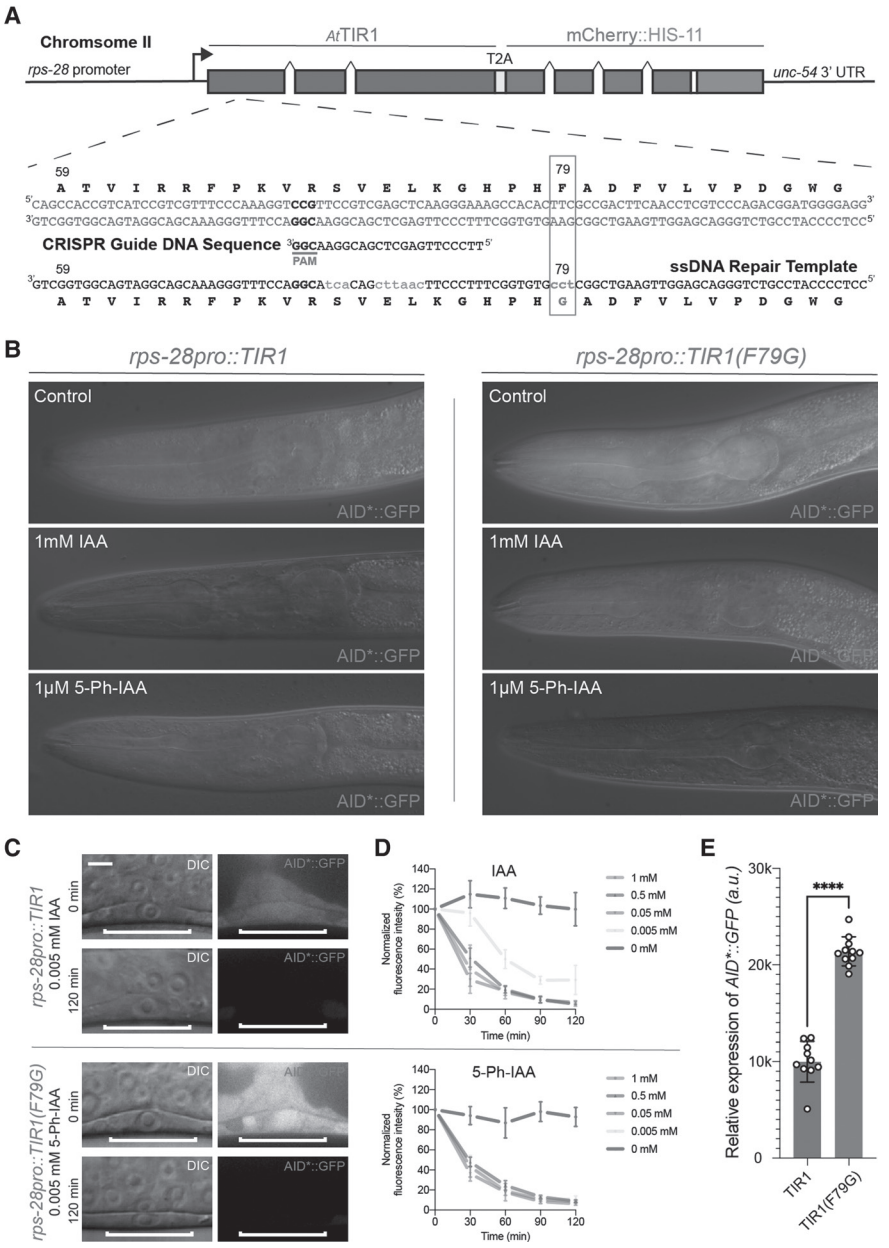


Figure 3. Mutation of phenylalanine 79 to glycine in the TIR1 protein switches the specificity of the auxin degradation system to one that is now responsive to 5-Ph-IAA. (*A*) Structure of *cshIs128* that encodes both a $_{At}$TIR1(WT) protein and an autocatalytically cleaved nuclear localized mCherry::HIS-11 reporter driven by a ubiquitously expressed ribosomal protein (*rps-28*) promoter. Below the gene structure is the coding sequence of the region of $_{At}$TIR1(WT) that was mutagenized via CRISPR and homology-directed repair (HDR) using a co-injected single-stranded DNA (ssDNA) oligo. (*B*) Representative mid-larval staged animals expressing AID*::GFP and one of two indicated $_{At}$TIR1 variants. Animals were grown continuously on untreated nematode growth media (NGM) plates or NGM plates including the indicated auxin analog at the listed concentration. (*C*) Micrographs of early L3 staged animals expressing AID*::GFP and one of the two $_{At}$TIR1 variants before and 120 minutes after the addition of the auxin analog. (*D*) Rates of AID*::GFP degradation were determined by quantifying AID*::GFP in early L3 staged P6.p cells in animals coexpressing the indicated $_{At}$TIR1 variant following auxin analog treatment. Data presented as the mean and SD ($n \geq 10$ animals examined for each time point). (*E*) Quantification of the relative expression levels of the AID*::GFP reporter in P6.p cells of early L3 staged animals that were grown on control plates. Data presented as the median with SD ($n \geq 10$ animals examined for each TIR1 transgene and **** = $p < 0.0001$ by a Mann–Whitney U-test).

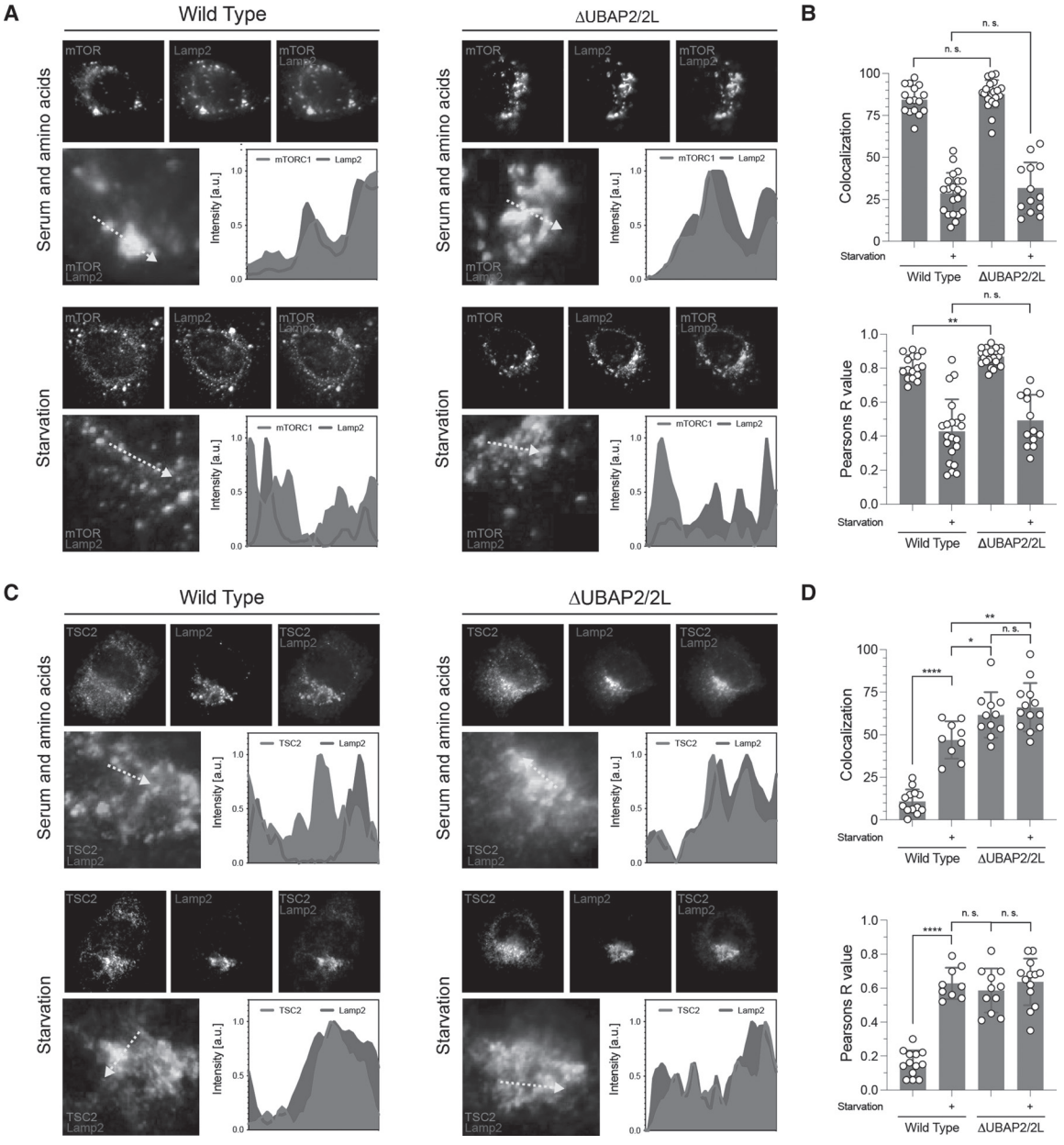


Figure 4. Deletion of UBAP2 and UBAP2L results in constitutive localization of TSC2 to the lysosomal surface, where it inhibits mTORC1 kinase activity. (*A*) mTORC1 localization from the cytoplasm to the lysosomal surface is dynamically regulated. In serum- and amino acid–replete conditions, mTORC1 is recruited to the lysosomal surface, where its kinase activity is stimulated by GTP-bound Rheb. When amino acids and serum are removed from the growth medium, mTORC1 is relocalized to the cytoplasm. Removal of UBAP2/2L expression does not alter these dynamics. (*B*) Quantification of the spatial overlap of mTOR with the resident lysosomal surface marker LAMP1. (*C*) The localization of the tumor suppressor tuberous sclerosis complex is also dynamically altered by changes in environmental conditions. In normal growth media, TSC2 is localized throughout the cytoplasm. When serum is removed from the growth medium, TSC2 translocates to the lysosomal surface, where it can dampen the stimulatory activity of Rheb on mTORC1 kinase activity. In contrast to wild-type cells, ΔUBAP2/2L cells constitutively localize TSC2 to the lysosomal surface. (*D*) Quantification of the overlap of TSC2 and LAMP2 localization in wild-type and ΔUBAP2/2L cells.

adapts engineered degron-tagged proteins to the endogenous *C. elegans* E3 ubiquitin ligases complex (SKR-1/2-CUL-1-F-box [SCF]), targeting them for degradation by the proteosome. While this system has been employed to dissect the developmental functions of many *C. elegans* proteins, we have found that several auxin-inducible degron (AID)-tagged proteins are constitutively degraded by $_{At}$TIR1 in the absence of auxin, leading to undesired loss-of-function phenotypes. In the last year, we generated an orthogonal auxin-derivative/mutant $_{At}$TIR1 pair (*C. elegans* AID version 2 [*C.e.*AIDv2]) that transforms the specificity of allosteric regulation of TIR1 from IAA to one that is dependent on an auxin derivative harboring a bulky aryl group (5-Ph-IAA). We find that a mutant $_{At}$TIR1(F79G) allele that alters the ligand-binding interface of TIR1 dramatically reduces ligand-independent degradation of multiple AID*-tagged proteins. In addition to solving the ectopic degradation problem for some AID targets, addition of 5-Ph-IAA to culture media of animals expressing $_{At}$TIR1(F79G) leads to more penetrant loss-of-function phenotypes for AID*-tagged proteins than those elicited by the $_{At}$TIR1-IAA pairing at similar auxin analog concentrations (Fig. 3). The improved specificity and efficacy afforded by the mutant $_{At}$TIR1(F79G) allele expands the utility of the AID system and broadens the number of proteins that can be effectively targeted with it.

UBAP2 and UBAP2L Are Required for Normal mTORC1 Signaling and the Tumor Suppressor Tuberous Sclerosis Complex Is Constitutively Localized to the Lysosomal Surface in ΔUBAP2/2L Cells

H. Ouyang, J. Wang

Previous findings from our laboratory indicate that UBAP2 and UBAP2L function redundantly to promote mTORC1 signaling. This regulatory role for UBAP2 and UBAP2L ensures that mTORC1-dependent phosphorylation of protein translation machinery components (including RSP6 kinase and RSP6) that stimulate translation and other proteins that inhibit translation (e.g., 4EBP proteins) are properly regulated. In the last year we explored several candidate mechanisms for UBAP2/2L regulation of mTORC1. mTORC1 is regulated at a number of levels that include the controlled localization of both mTORC1 and its negative regulator tuberous sclerosis complex (TSC) to the lysosomal surface. Under nutrient-replete conditions and ample growth factors,

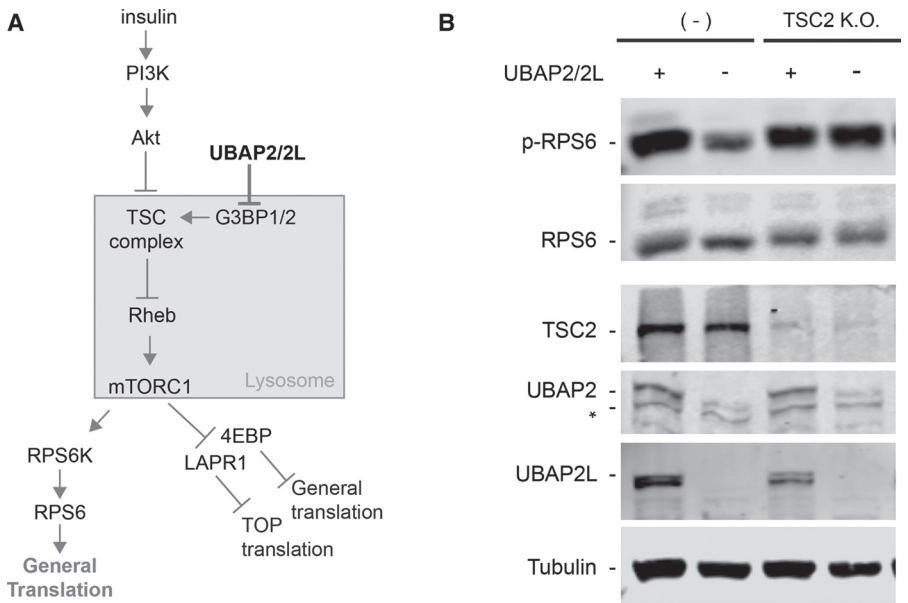


Figure 5. The phenotypes associated with deletion of UBAP2 and UBAP2L are suppressed by removing TSC2. (*A*) A schematic diagram of mTORC1 signaling indicating the proposed regulatory interactions among UBAP2/2L, TSC2, and mTORC1. (*B*) Deletion of TSC2 in UBAP2/2L cells restores normal mTORC1 signaling in ΔBAP2/2L cells.

mTORC1 is recruited to the lysosomal surface, where its kinase activity is stimulated by the lysosomal resident protein Rheb. When amino acids and serum are removed, the tumor suppressor TSC is recruited to the lysosomal surface, where it activates Rheb GTPase activity, removing its ability to stimulate mTORC1. This regulatory step, coupled to the activation of other GTPases that tether mTORC1 to the lysosomal surface, leads to a dampening of mTORC1 kinase activity. We localized both mTORC1 and TSC2 in wild-type and UBAP2/2L-deleted strains. We found that mTORC1 localization was not altered in UBAP2/2L cells (Fig. 4). In contrast, we found that TSC2 was constitutively localized to the lysosomal surface, where it inactivates Rheb in ΔUBAP2/2L cells. These results led to the model that mTORC1 activity is continuously repressed in UBAP2/2L cells and this likely is the cause of repressed translational activation in these cells. Consistent with this model, deletion of TSC2 in ΔUBAP2/2L cells suppresses mTORC1 signaling defects (Fig. 5B). We hypothesize that UBAP2/2L controls the activity of one or more proteins that functionally tether TSC2 to the lysosomal surface.

PUBLICATIONS

Abbatemarco S, Bondaz A, Schwager F, Wang J, Hammell CM, Gotta M. 2021. PQN-59 and GTBP-1 contribute to stress granule formation but are not essential for their assembly in *C. elegans* embryos. *J Cell Sci* **134:** jcs258834. doi:10.1242/jcs.258834

Carlston C, Weinmann R, Stec N, Abbatemarco S, Schwager F, Wang J, Ouyang H, Ewald CY, Gotta M, Hammell CM. 2021. PQN-59 antagonizes microRNA-mediated repression during post-embryonic temporal patterning and modulates translation and stress granule formation in *C. elegans*. *PLoS Genet* **17:** e1009599. doi:10.1371/journal.pgen.1009599

Hills-Muckey K, Martinez MAQ, Stec N, Hebbar S, Saldanha J, Medwig-Kinney TN, Moore FEQ, Ivanova M, Morao A, Ward JD, et al. 2021. An engineered, orthogonal auxin analog/$_{At}$TIR1(F79G) pairing improves both specificity and efficacy of the auxin degradation system in *Caenorhabditis elegans*. *Genetics* **220:** iyab174. doi:10.1093/genetics/iyab174

Stec N, Doerfel K, Hills-Muckey K, Ettorre VM, Ercan S, Keil W, Hammell CM. 2021. An epigenetic priming mechanism mediated by nutrient sensing regulates transcriptional output during *C. elegans* development. *Curr Biol* **31:** 809–826.e6. doi:10.1016/j.cub.2020.11.060

SPATIAL PROFILING OF GENETIC VARIANTS IN SITU

J.H. Lee D. Chitty D. Furth D. Ghosh C. McLaughlin R. Peters X. Wang C. Yuan

Next-Generation Sequencing Chemistry-Enabled Cytology in Cancer Detection

D. Ghosh, C. McLaughlin

Timely cancer detection is important for the prevention of tumor progression or relapse. If premalignant, early, or residual tumor cells could be seen with certainty based on the mutational signature in addition to protein biomarkers, it could lead to a more definitive diagnosis and early treatment. Therefore, we developed a scalable way to interrogate the mutational landscape in situ using RNA-templated sequencing-by-ligation (RSBL). We also developed a combinatorial encoding scheme called anticodon oligonucleotides (ACOs) so harmful codons could be identified de novo without sequencing. These advances are enabling flow cytometry–based applications capable of detecting rare tumor cells based on coexpressed mutations. Beyond cytogenetics, our platform (heuristic in situ targeted oligonucleotide sequencing, or HISTO-seq) opens the door to tumor pathology integrating the genetic landscape in disseminated tumor cells, as well as neoantigens and their receptors in immunotherapy.

Many available cancer detection methods rely on antibodies against cell type– or lineage-specific proteins. As a result, they do not capture early or disseminated tumor cells with a variable expression of indirect biomarkers (Pommier et al., *Science* **360:** 4908 [2018]). Although next-generation sequencing (NGS) is more sensitive, it may fail to provide a definitive proof of malignancy because of background mutations. Also, it needs additional radiologic or pathologic findings to justify therapeutic intervention, which is difficult in the setting of early cancer. If one could unambiguously demonstrate the existence of malignant tumor cells using clonal mutations, it could potentially lead to a more confident diagnosis and earlier treatment.

Currently, there are several different approaches to assessing single-nucleotide variants (SNVs) in situ. One approach uses allele-specific RNA in situ hybridization (ISH), which requires customizing and validating individual probes (Baker and Graham, *Methods Mol Biol* **2148:** 349 [2020]). Another approach is to genotype the DNA after reverse transcribing the mRNA, but this involves the use of nonstandard primers that do not scale (Larsson et al., *Nat Methods* **7:** 395 [2010]). In contrast, NGS chemistries can genotype millions of DNA molecules effortlessly without having to optimize each and every probe (Shendure et al., *Science* **309:** 1728 [2005]). This motivated us to ask whether we could develop NGS chemistry-enabled ISH to "paint" the mutational landscape in situ.

First, we tested the feasibility of a direct RSBL chemistry. Here, a DNA primer was hybridized to an RNA template immobilized on beads. Then, the primer was extended using partially degenerate oligonucleotides and a DNA end-joining ligase (Lohman et al., *Nucl Acids Res* **42:** 1831 [2014]) (SplintR, NEB). The analysis of the RSBL product revealed that the specificity approached that of SBL in NGS. The gel fragment analysis showed that ~95% of the DNA primer participated in RSBL and that 99% of the ligation product contained a correct base at the ligation junction (Fig. 1A).

To implement RSBL in situ, we used rolling circle amplification (RCA) based on STARMAP (Wang et al., *Science* **361:** 5691 [2018]). Briefly, a DNA primer was hybridized to the mRNA in fixed cells, and then partially degenerate RSBL oligonucleotides were ligated to the primer. Subsequently, the ligation product was splinted by another RNA ISH primer for intramolecular ligation, RCA, and fluorescent probe hybridization. For mouse *ACTB* mRNA, 60% of the DNA primers participated in RSBL, whereas the efficiency for human *MALAT1* lncRNA was 87%.

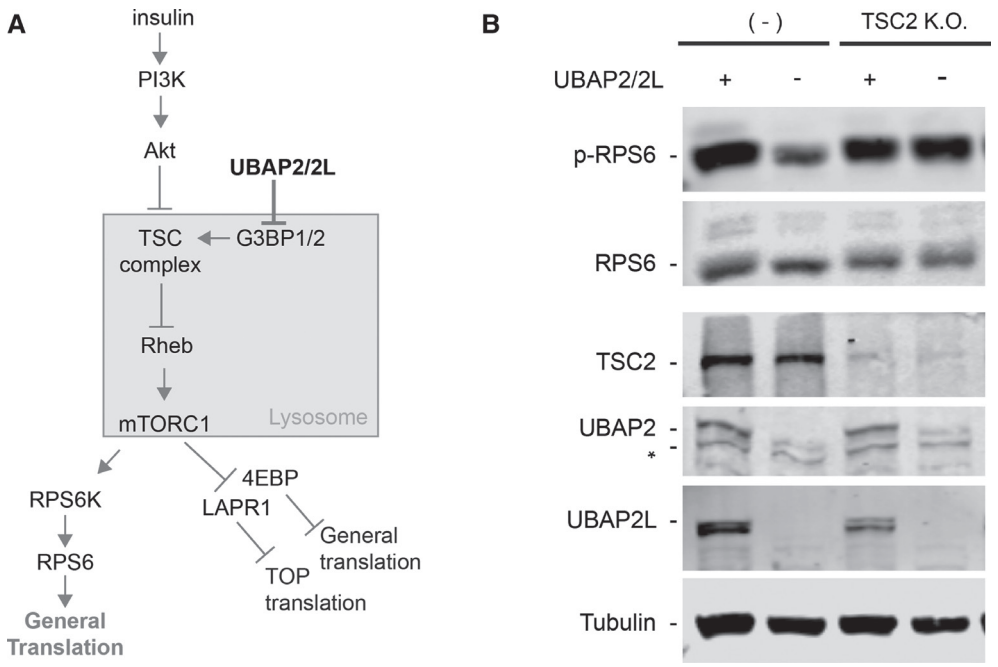


Figure 1. Sensitive direct RNA sequencing/genotyping method in single cells for detecting rare cancer clones.

Interestingly, RSBL probes against the 5′ *MALAT1* region resulted in a different localization pattern compared to the 3′ region, consistent with the cytoplasmic export of the latter (Wilusz et al., *Cell* **135:** 919 [2008]). We then tested 11 more genes using RSBL probes, revealing that their detection efficiency approached that of simple ISH probes for the majority of the probes (mean 60 ± 38%) (Fig. 1B,C).

We then compared RSBL to single-molecule FISH (smFISH) using Molecular Cartography from Resolve. The average RCA signal count was reproducible among the 12 RSBL targets ($R^2 = 0.8$), but the expression values between RSBL and smFISH were only moderately correlated ($R^2 = 0.4$) (Fig. 1D). Interestingly, we observed that the RCA efficiency was significantly influenced by the probe sequence even on the same transcript. This contributed to the variable RSBL sensitivity that ranged from 1% to 89% (mean 16 ± 24%) across different loci, in addition to the target accessibility.

To examine the base specificity of RSBL, we created a simple one-color readout in which only the desired variant-specific RSBL probes contained a signal amplification sequence. For mouse and human

ACTB mRNA, the correct allele was observed with >98% specificity with a false-positive rate of 0.4 ± 0.2 spots/cell. We then tested 20 additional mRNA loci known to be mutated in MCF7 cells. To calculate the base specificity, both wild-type and mutant alleles were represented using a mixed base for detection, whereas the other two alleles provided a measure of false positives. Here, the base specificity was as high as 95 ± 5.8% with a false positive rate of 0.28 ± 0.13 spots/cell (Fig. 1E).

Our eventual goal was to identify rare tumor cells based on genotype. However, point mutations are common also in normal tissues, especially with aging (Li et al., *Nature* **597:** 398 [2021]). In NGS-based early cancer detection, therefore, multiple co-occurring mutations are used to reduce false positives (Abbosh et al., *Nature* **545:** 446 [2017]), implying that doing so in situ might also be important. To identify a large number of coding variants de novo without having to sequence, we developed a color-coding scheme called anticodon oligonucleotides (ACOs) for RSBL. Here, the first three bases were encoded based on a codon table, and IUPAC ambiguity codes were used to represent degenerate mutations or wobble codon bases. As a result, fewer than three RSBL oligonucleotides were needed to represent all possible missense mutations for each codon. This also made it possible for us to generate multiple site-specific RSBL probes even in the absence of sequencing data, such as the 20 previously described mutations in MCF7 (Fig. 1F,G).

We then asked whether pooled variant screening could be used as a simple clinical assay, especially for assessing minimal residual disease (MRD). After screening for sufficiently sensitive and specific ACOs in MCF7 cells, six probes were chosen for pooled RSBL in adherent and suspended MCF7 cells, demonstrating up to approximately 50 fluorescent spots/cell. We then optimized the protocol so the signal from preligated ACO control probes could fill the entire volume of the cell (Fig. 1H). We are now optimizing the ligation step so the labeled cells in the blood sample could be used for fluorescence-activated cell sorting (FACS) analysis. Based on the observed false-positive rate, we expect the sensitivity to better that of existing MRD assays for acute myeloid leukemia (AML) and acute lymphocytic leukemia (ALL) (10^{-4} to 10^{-6}).

We report a simple in situ cell labeling technique incorporating an NGS chemistry. We also introduce the ACO concept that represents the intersection of protein detection and DNA sequencing. By tweaking a well-known NGS dinucleotide encoding scheme in SOLiD (Valouev et al., *Genome Res* **18:** 1051 [2008]), ACO was made to recognize amino acid rather than nucleotide identities. Because of its simplicity and scalability, it may eventually be possible to see rare cell clones containing any number of sequence tags using more accessible laboratory techniques.

In terms of future applications, we hope to adapt our assay for MRD assessment in blood cancers. The genotype-based approaches might be particularly useful in the context of regenerating bone marrow or clonal hematopoiesis approaches that delay early or residual leukemia cell detection. For solid tumors, this could enable the identification of disseminated tumor cells, as well as neoepitope and immune cell signatures in the tumor microenvironment. Beyond clinical medicine, we believe that the ability to map and isolate rare single cells based on cell lineage information could reveal important insights about cell fate determination in early development.

The concepts and tools described in our study owe much to the insights that led to the development of NGS, single-molecule imaging, and spatial genomics over the years. We simplified some of these ideas for clinical applications; however, our method is closely related to highly multiplexed spatial methods and can easily be multiplexed. This could eventually permit comprehensive mapping of RNA polymorphisms, mutations, isoforms, and modifications in multiple tissue types. If used in conjunction with expansion microscopy techniques (Alon et al., *Science* **371:** 2656 [2021]), it might even be possible to visualize the RNA regulatory events with a nano-scale resolution inside the cell. Over time, our hope is that such methods will provide more insights into the spatially segregated genetic mosaicism in development, aging, and cancer.

GENETIC AND EPIGENETIC BASIS OF CANCER AND NEURODEVELOPMENTAL SYNDROMES

A. Mills S. Balinth, C. Ballon, J. Boockvar, Y. Chang, M. Fisher, D. Johnson, H. Liang, J. Park, P. Shrestha, S. Sun, C. Wu

Our team elucidates genetic and epigenetic mechanisms that impact human disease so more effective treatments can be developed in the future. We focus on cancer and neurodevelopmental syndromes by discovering genes that impact these conditions, figuring out how their encoded proteins normally work, and determining how their perturbation contributes to human disease. Our findings have impacted how clinicians analyze and treat patients with these syndromes.

Our major discoveries have been:

- Revealing the role of P63 in development, cancer, and aging
- Defining the genetic basis of autism
- Discovering *CHD5* as a gene that prevents cancer
- Identifying new epigenetic vulnerabilities of human glioblastoma

p63 in Development, Aging, and Cancer

We have had a long-standing interest in P63, a P53-related protein Dr. Mills discovered that is a master regulator of development. We have focused on how P63 normally works, as well as how its deregulation leads to disease processes. Although human P63 looked very similar to the P53 tumor suppressor, which is mutated in the majority of cancers, its function was completely unknown when Dr. Mills first found it. The Mills laboratory discovered that loss of p63 (the mouse version of human P63) leads to premature aging, and that it is required for replenishing stem cells of stratified epithelia such as the skin. Indeed, loss of p63 in mice causes features of aging such as curvature of the spine, hair loss, and severe skin lesions. Levels and types of P63 are crucial, as we discovered that excessive expression of one version of P63 (ΔNP63α) causes carcinoma—the most prevalent type of human cancer—whereas a different version of P63 (TAp63) prevents mesenchymal cancers such as sarcoma. Our finding that TAP63 effectively inhibits tumor growth in cells lacking P53 was paradigm-shifting, as it indicates that TAP63 prevents cancer completely independently of P53.

We had initially discovered that p63 was essential for mouse development, as its loss causes depletion of stem cells and birth defects of the limbs, skin, and palate. Our findings led others to interrogate P63 in humans and to reveal that its mutation causes seven different syndromes characterized by birth defects similar to those we found in the mice we engineered to lack p63. By generating mouse models for one of these human syndromes, ectrodactyly, ectodermal dysplasia, clefting (EEC) syndrome, we found out 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 advanced our P63 projects on several fronts. First, we published a collaborative study with the Blazer laboratory, in which we elucidate P63's role in thymic development (Stefanski et al. 2022). Second, our team identified a signaling cascade driven by the bromodomain protein BRD4 that up-regulates ΔNP63α to promote cancer stem-like properties (Fisher et al. 2021). This work has important therapeutic implications for the treatment of squamous cell carcinomas. In collaboration with clinical oncologists at Northwell Health, we are studying how P63 impacts human carcinomas of the head and neck, cervix, and salivary gland—tissues in which we showed p63 to be essential in mice.

Genetics of Autism Spectrum Disorder

Our laboratory revealed that one of the most common genetic alterations in autism—deletion of a 27-gene cluster on human chromosome 16—causes autism-like features. We used chromosome engineering—a technology that allows us to generate precise chromosome rearrangements in the mouse—to create mouse models with chromosome deletions corresponding to those at human chromosome 16, providing the first functional evidence that inheriting fewer copies of genes in this region leads to autism-associated characteristics. Our mouse models have autism-like behaviors such as hyperactivity, 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 have changes in brain architecture that we detected using magnetic resonance imaging (MRI). Our work provides functional evidence for the genetic basis of autism that was not previously appreciated. Our collaborative work reveals that genes deleted in autism affect sleeping patterns by impacting rapid eye movement and neural oscillation. We have been putting significant effort into homing in on the critical genes by generating a series of mouse models with subdeletions of the region corresponding to the chromosome 16 region we identified in humans. These subdeletion models will be invaluable for the community to pinpoint the genes responsible for autism as well as for designing effective treatments.

CHD5, a New Tumor Suppressor

We discovered *CHD5* as a tumor suppressor mapping to 1p36—a region of our genomes frequently deleted in a variety of human cancers. Although ample evidence that a tumor suppressor was located in this region had existed for more than three decades, this tumor suppressor had not been identified. Using chromosome engineering to generate mice with deletions and duplications of the genomic region corresponding to human 1p36, we discovered *CHD5* as the tumor suppressor gene and found that its product turns on a network of tumor suppressors. In addition, we showed that *CHD5* is frequently deleted in human glioma.

We advanced this fundamental discovery by 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 Chd5 (the mouse version of human CHD5) uses its plant homeodomains to bind histone H3, and that this interaction is essential for tumor suppression. This 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. In fact, recent reports show that patients with high levels of *CHD5* have much better overall survival than those with low levels. We found that CHD5 is essential for packaging DNA, and that loss of Chd5 in engineered mice causes improperly packaged DNA that is prone to DNA damage. 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 collaborated with clinicians at Northwell Health 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. We also found Chd5 to be 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. Over the past year we have been 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. These studies have led to collaborations with neuro-oncological surgeons at Northwell Health in which we are able to grow organoids from patients with glioblastoma—a deadly cancer for which there are few successful treatment options. We have extended this research program significantly, which over the past year allowed us to identify new epigenetic vulnerabilities of human glioblastoma. We believe that these targets provide inroads toward more successful treatments for patients with this devastating malignancy of the brain.

PUBLICATIONS

Fisher ML, Balinth S, Hwangbo Y, Wu C, Ballon C, Wilkinson JE, Goldberg GL, Mills AA. 2021. BRD4 regulates ΔNp63α to drive a cancer stem cell phenotype in squamous cell carcinoma. *Cancer Res* **81:** 6246–6258. doi:10.1158/0008-5472.CAN-21-0707

In Press

Stefanski H, Xing Y, Nicholls J, Jonart L, Goren E, Taylor P, Mills A, Riddle M, McGrath J, Tolar J, et al. 2022. P63 targeted deletion under the FOXN1 promoter disrupts pre- and post-natal thymus development, function and maintenance as well as induces severe hair loss. *PLoS ONE* **17:** e0261770. doi:10.1371/journal.pone.0261770

CANCER, CORONAVIRUS, AND HUMAN GENETICS

M. Wigler J. Alexander, P. Andrews, B. Debmalya, A. Gruet, I. Hakker, J. Kendall, S. Li, E. McKenna, A. Moffitt, H. Orzuela, A. Peyser, C. Podszus, M. Riggs, M. Ronemus, E. Rose, J. Rosenbaum, L. Shanley, D. Stauder, A. Stepansky, A. Tandon, D. Trimboli, Z. Wang, M. Wroten, Z. Yu

Because of the coronavirus epidemic, our laboratory retooled in 2020 to work on the virus. Our efforts combine methodology development, computational analysis, and clinical and population discovery. In addition to the coronavirus research, 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, A. Gruet, I. Hakker, J. Kendall, S. Li, A. Moffitt, C. Podszus, M. Riggs, M. Ronemus, J. Rosenbaum, D. Stauder, 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; T. Baslan and R. Levine, Memorial Sloan-Kettering Cancer Center; A. Zetterberg, Karolinska Institutet; L. Muthuswamy, Beth Israel Deaconess Medical Center; M. Kemeny, Icahn School of Medicine at Mount Sinai; S. Spivack, Albert Einstein College of Medicine; E. Li, Stony Brook University Medical Center; D. Levy and A. Krasnitz, CSHL]

We continue our efforts to integrate our genomic methods within the healthcare arena. We have initiated new collaborations with several academic medical centers in the metropolitan area. Our previous work dealt mainly with breast, endometrial, and ovarian cancers as well as acute myeloid leukemia, and we have now extended these analyses to lung and colorectal cancers. Using the MASQ (multiplex accurate sensitive quantitation) method developed in our laboratory, we assess tumor load in two components derived from blood (circulating epithelial cells and cell-free DNA from plasma) and cell-free DNA from urine. With the establishment of a robust logistical pipeline for receiving samples from our clinical collaborators, we have been able to handle an ever-increasing number of specimens covering a wide range of sample types, including various tissues and biopsy washings. The results, generated from our application of MASQ to these specimens, demonstrate the clinical utility of MASQ in assessing the efficacy of treatment and detecting minimal residual disease—and potentially for early detection.

The MASQ technique incorporates genetic mutations unique to a given tumor. Because it requires this individual customization, we have begun to explore an alternative method for tracking cancer genomes: variation in length of microsatellites, tracts of repetitive DNA in which certain motifs (ranging in length from 1 to 6 base pairs) are repeated. Instability within these regions is a hallmark of cancer. Until now, measuring microsatellite lengths has been very difficult, as the same instability that makes microsatellites excellent markers for cancer also makes them highly unstable for DNA replication and sequencing. But by using the muSeq technique we have developed (see Genomics section below), we can accurately assess instability of these regions using a single hybrid capture panel to identify tumor-specific variants and to track length variation over time. The analysis can be customized to the patient's tumor but does not require patient-specific reagents, allowing for vast reductions in time and cost. In our initial experiments, the custom capture panels have provided 2,000-fold enrichment of 1,260 microsatellite loci with ~40% of reads on target. Capture is nearly uniform across all targets, and our error estimates from synthetic template molecules have been confirmed in actual human DNA samples. Further development of this promising approach will be a major focus of our continuing efforts.

Autism Genetics

J. Kendall, M. Ronemus, M. Wroten [in collaboration with I. Iossifov, CSHL and the New York Genome Center; D. Levy, A. Krasnitz, H. Meyer, and T. Janowitz, CSHL; K. Baldwin, Columbia University; K. Ye, Albert Einstein College of Medicine; A. Buja and A. Krieger, Wharton School of the University of Pennsylvania]

In collaboration with Kenny Ye (Albert Einstein College of Medicine) and Andreas Buja and Abba Krieger (Wharton School of the University of Pennsylvania), we developed a method to measure the extent to which siblings, concordant and discordant for autism, share their parental genomes using genotypes derived from whole-genome data or chips. We applied the method to approximately 1,300 pairs of concordant and approximately 4,500 pairs of discordant siblings from the Simons Simplex Collection (SSC), Autism Genetic Research Exchange (AGRE), and Simons Foundation Powering Autism Research for Knowledge (SPARK) collections of autistic families. Surprisingly, we observed that the fathers have an increased sharing consistent with them carrying causal determinants in at least half of the multiplex families, more than the families in which the mother has a causal variant. With lesser significance, the discordant siblings from the simplex families corroborate the observation: the antisharing tended to be stronger for the paternal genome. The more extensive sharing of paternal than maternal genomes contradicted our expectations that mothers will be the primary source of damaging variants in the high-risk multiplex families and has forced us to rethink our unified hypothesis of the genetic contribution to autism.

We are investigating the effects of autism-linked mutations on gene expression to explore the molecular mechanisms that may underlie the disorder. In collaboration with Kristin Baldwin (Columbia University) and the New York Genome Center, we have previously conducted pilot studies on the effects of de novo mutation on 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. Building on the success of these efforts, the Simons Foundation Autism Research Initiative (SFARI) funded a large-scale project to generate 4,000 RNA-sequencing (RNA-seq) profiles, which correspond to all children in the SSC. We have generated approximately 3,500 of these profiles so far, and we expect to complete the data generation by February 2022. We expect that the project will yield 200–300 de novo noncoding variants associated with the perturbed expression or splicing of nearby genes. Following this, we hope to use CRISPR and cell transformation to determine whether any of the candidate de novo noncoding variants are responsible for the observed changes in gene expression.

The paternal bias in transmission that we observe has led us to test the hypothesis that this effect arises from maternal–fetal antigenic incompatibility. Following two separate paths, in an effort led by Tobias Janowitz, we are developing mouse models of maternal–fetal incompatibility, and in collaboration with Alexander Krasnitz and Hannah Meyer, we are applying computational approaches to identify traces of maternal–fetal incompatibility. Initial data from the mouse models indicates that in mothers undergoing stimulation of the immune system, placental weight is lower and mouse pups are deficient in normal social interactions. We are currently investigating these preliminary findings in greater detail.

Genomics

P. Andrews, I. Hakker, S. Li, E. McKenna, A. Moffitt, H. Orzuela, A. Peyser, E. Rose, L. Shanley, Z. Wang [in collaboration with D. Levy, CSHL]

BAG-seq is a refinement of the BAG (balls of acrylamide gel) method we use to embed individual cells and crosslink the nucleic acids to a polyacrylamide matrix. We are using BAGs in a number of approaches: to define host cells within tumors, to characterize tumor heterogeneity, and to classify the expression patterns of neurons. When studying the relationship between genomes and transcriptomes in cancer, full understanding of the cooperation and competition between cancer cells and stroma, as well as the emergence of malignant from premalignant cells, is essential. Therefore, we found it valuable to combine both DNA and RNA from individual cells to distinguish the cells of the host from those of the tumor and then determine the respective transcriptional states. Our approach was to develop a new high-throughput method to simultaneously capture genomic DNA and RNA transcripts of single nuclei isolated from frozen biopsies into BAGs, and then to separate the two types of nucleic acid by bioinformatics ("BAG-seq"). We applied BAG-seq to four types of endometrial cancer and studied more than 40,000 nuclei. Using this hybrid method, the DNA component can readily distinguish stroma and tumor subpopulations, whereas the RNA component can robustly distinguish types of

stroma and distinct transcriptional states of tumor cells. By mapping the connections between DNA clones and RNA clusters of the same tumor using this hybrid platform, we found correlations in genomic transcription, most notably that different patients show common expression of stromal components. We also have found that some cells with tumor genomes match the transcriptional states of normal stroma.

De novo muSeq is a computational method that builds on our earlier work and uses short (300-bp) reads to generate longer contigs (~10 kb) by first mutating the template molecules using incomplete sodium bisulfite conversion of unmethylated cytosine to uracil. This random mutation pattern distinctly marks each initial template molecule, and we use this uniqueness to reconstruct each of the initial template molecules. We aggregate templates from the same haplotype to obtain specific assemblies that are free of mutational signatures and have extremely low error rates. To test the approach, we have applied this method to a quad family to generate fully phased haplotypes: both regions of low variation, in which 1-kb to 2-kb stretches of DNA typically separate heterozygous variants, and regions of extremely high variation, such as the entire 4.6-kb stretch that encompasses *HLA-B*, were successfully assembled. These phased assemblies did not use a reference genome, and they represent how mutational sequencing can effectively convert short-read methods into long-read sequencing.

Coronavirus

P. Andrews, B. Debmalya, A. Gruet, S. Li, A. Moffitt, M. Ronemus, J. Rosenbaum, A. Tandon, Z. Wang
[in collaboration with D. Levy and T. Gingeras, CSHL; S. Perlman, University of Iowa; J. Rothman, Columbia University; D. Donoho, Stanford University]

Because of the ongoing COVID-19 pandemic, we have refocused some of our efforts. Very briefly, (1) we have designed a prototype system to detect and quantitate viral RNA in patients, based on the principles of MASQ. We envision a multisample, multisite assay that could be used for any respiratory illness. The test is inexpensive, and it can be performed on a very large scale using next-generation sequencing technology. Given the current widespread availability of testing, we have turned our focus elsewhere. (2) We have continued to test ideas for blocking viral replication by inhibiting the translation of its proteins through binding of specific complementary synthetic nucleotide derivatives. If our methods are successful, they will also have potential application in the treatment of diseases of genetic origin by allowing us to modulate the expression of target genes. Using a reporter construct in cell lines, we have successfully blocked translation by targeting pairs of antisense oligos to the 5′-untranslated region, and we are now expanding this analysis. (3) In collaboration with David Donoho's group at Stanford, we have created a computer simulation of coronavirus spread utilizing mathematical models that incorporate factors such as viral multiplicity of infection (MOI), social distancing and the wearing of masks. This will allow us to make better predictions to guide social policy on testing and vaccination. The simulation is currently in beta form, and we expect to release it publicly in the near future.

PUBLICATION

Yoon S, Munoz A, Yamrom B, Lee Y-H, Andrews P, Marks S, Wang Z, Reeves C, Winterkorn L, Krieger AM, et al. 2021. Rates of contributory de novo mutation in high and low-risk autism families. *Commun Biol* **4:** 1026. doi:10.1038/s42003-021-02533-z

Semir Beyaz's research interrogates the functional consequences of diets for immune recognition and response pathways that play critical roles in cancer immunity. 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 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. 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.

Why do patients with cancer (irrespective of cancer type) frequently experience systemic symptoms like pain, cognitive impairment, deficits in appetite, and disrupted sleep/wake cycles? What is the underlying biology governing these phenomena, and how can this biology be leveraged to improve peoples' lives? To answer questions such as these, the **Jeremy C. Borniger** laboratory investigates bidirectional communication between the brain and periphery in the context of cancer. The laboratory aims to determine how tumors disrupt neural circuit function, how aberrant cellular activity promotes cancer-associated systemic dysfunction, and how reciprocal outputs from the brain regulate cancer growth and metastasis. Specifically, the Borniger laboratory uses techniques from systems neuroscience (e.g., optogenetics, calcium imaging, circuit mapping, electrophysiology, and behavioral assays) to dissect how factors in the tumor microenvironment alter host physiology and behavior. Recent work has focused on how central neuromodulator populations participate in cancer-associated sleep and metabolic disruption. The laboratory discovered that nonmetastatic mammary tumors distally alter immune and endocrine signaling to aberrantly activate lateral hypothalamic hypocretin/orexin (HO) neurons. This resulted in disrupted sleep and hepatic glucose metabolism, the latter being driven by the sympathetic nervous system (Borniger et al., *Cell Metabolism* **28:** 118 [2018]). This research, in combination with clinical work, will facilitate the development of novel treatments to improve outcomes for patients with 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. Egeblad 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 a 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.

The **Douglas Fearon** laboratory studies the interaction between cancer and the immune system. Their 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 uncovering 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. Some of the Fearon laboratory's 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.

How do tumors interact with the biology of the host system? What can be learned 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 **Tobias Janowitz** 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. The Janowitz laboratory's work, therefore, confirms that cancer cannot be understood, and probably not 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.

The **Michael James Lukey** laboratory aims to understand the nature and regulation of metabolic adaptations in the different stages of cancer, and then to develop therapeutic strategies that target resulting vulnerabilities. Proliferative signals in mammalian cells drive biosynthetic programs that support cell growth and replication. In healthy cells this process is tightly regulated by growth factors, but in cancer cells oncogenic lesions can result in continuous signaling to the metabolic machinery. Oncogene-driven metabolic reprogramming supports tumorigenesis but renders cells sensitive to specific metabolic stresses, a phenomenon that is exploited for cancer therapy. Because the distribution of nutrients varies markedly between organs, cancer cells growing at different sites in the body—and in different regions of the tumor microenvironment—must employ a range

of metabolic strategies to fuel their growth. The Lukey laboratory is especially interested in the biochemical processes underlying nutrient sensing and metabolic/redox homeostasis, including regulation of the protein posttranslational modification landscape by reactive metabolites. They are also exploring the reciprocal connections between tumor metabolism and host physiology, recognizing that metabolic therapies must be designed to synergize with, and not to antagonize, the anti-tumor immune response.

Darryl Pappin's laboratory develops chemical and computational methods for analysis of proteins and peptides. These are fundamental tools for proteomics, and they are vital in many fields of biological investigation. Proteins and peptides are typically analyzed via mass spectrometry, a method that involves fragmenting samples by colliding them with gas atoms in a vacuum. Masses of the resulting fragments are measured, and computer algorithms match the results with known or predicted molecules whose amino acid sequences are either known or inferred. Pappin has developed search engines for mass spectrometry data that enable investigators to sift hundreds of thousands of experimental spectra at a time for database matches. Along with chemical methods for quantifying protein changes, the laboratory has applied these approaches to the characterization of redox and glycosylation changes in pancreatic cancer, identification of cell-surface antigens for cancer diagnostics, and both protein and metabolic changes in prostate cancer and leukemia. The Pappin laboratory is also exploring methods to reduce sample complexity via called chemical sorting. This includes the use of chelation to enrich phosphopeptides from the total peptide pool and the use of biotin-tagged small molecules to segregate free thiol groups for redox proteomics.

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 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 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 long-term depression, 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.

LOCAL AND SYSTEMIC INTERACTIONS BETWEEN CANCER AND THE NERVOUS SYSTEM

J.C. Borniger A. Berisha A. Gomez
N. Francis Y. Wu

Role of Sleep Disruption in Antitumor Immunity

N. Francis, J. Borniger

Postdoc Nikita Francis is tackling an important problem in the field aiming to understand how sleep disruption influences downstream antitumor immune responses in mouse models of breast cancer. Sleep disruption is one of the most prevalent disturbances experienced by breast cancer patients. Up to ~90% of breast cancer patients report sleep disruption and associated fatigue throughout diagnosis, surgery, treatment, remission, and relapse. Accumulating evidence demonstrates that disruptions in the amount or structure of sleep can impair peripheral immune responses. Whether this functionally impairs antitumor immune responses, and the mechanisms underlying this phenomenon, are currently undefined.

Using an automated sleep deprivation protocol (8 hours/day for 3 weeks), Nikita examined how disruptions in sleep/wake states influenced tumor development, metastasis, and associated immune cell dynamics in a mouse model of breast cancer. Preliminary data suggest that mice with tumors that underwent the 3-week sleep deprivation protocol had a higher rate of metastatic spread to the lungs, a common site of breast cancer metastasis. Additionally, these mice had elevated circulating concentrations of the chemokines CCL22 and CXCL1, which are involved in immune cell trafficking and functional polarization. CXCL1 promotes breast cancer migration and invasion in both human and mouse breast cancer models. It is further overexpressed in metastatic lung lesions secondary to primary breast cancer. CCL22, a CCR4 ligand expressed by dendritic cells and macrophages, is highly abundant in breast cancer and is associated with lung metastatic burden. Likewise, IGFBP-1, a protein associated with tumor proliferation and migration, is also altered in the plasma of tumor-bearing mice that underwent sleep deprivation. As many growth factors, chemokines, and cytokines that we observed to be changed in our model are part of the senescence-associated secretory phenotype (SASP), we are testing the working hypothesis that sleep disruption promotes immune cell senescence to enhance cancer progression. This work is forthcoming.

Lateral Proteome Transfer between Sensory Neurons and Breast Cancer Cells

Y. Wu, J. Borniger

A primary question we are interested in is how do peripheral neurons and breast cancer cells interact at the cellular/local level? Additionally, do these local interactions benefit cancer progression? Postdoc May Wu co-cultured dorsal root ganglion cells (including neurons, Schwann cells, fibroblasts) with mouse breast cancer cells and assessed cancer cell proliferation using MTT assays. Notably, cancer cell proliferation was effectively doubled only when cancer cells were co-cultured with mitotically active dorsal root ganglion (DRG) cells (i.e., nonneuronal cells), an effect that was eliminated by pretreating DRG cells with mitotic inhibitors (i.e., leaving only postmitotic neurons). This suggests that DRG stromal components (e.g., fibroblasts, Schwann cells) are essential for the proliferative effect.

Using live-cell imaging, May demonstrated that calcein-labeled breast cancer cells move more and at a faster speed when co-cultured with DRG cells without, but not with, mitotic inhibitor pretreatment. Indeed, co-culturing breast cancer cells with DRG neurons alone or using DRG-conditioned media had no effect on cell migration distance or speed. Together, these data suggest that nonneuronal DRG cells are critical for the pro-cancer effects, and these effects are dependent on direct contact, rather than paracrine/long-distance signaling.

May aimed to get a better look at what was really going on between neurons and cancer cells directly. To accomplish this, she developed a scanning electron microscopy (SEM) protocol for imaging co-cultured neurons and cancer cells at high magnification (up to 2500x) and resolution. Under SEM, EO771 murine breast cancer cells took on an elongated morphology and seemed to squeeze themselves into the neuronal network during migration. Spindle-shaped Schwann cells on axons and flat, adhered fibroblasts were also observed. In general, May observed that fibroblasts tightly adhere to the bottom of the culture plate, whereas cancer cells migrate in a middle layer with nonadherent axons covering the heterogeneous mix of cells like taut strings off the bottom (see Fig. 1D).

We hypothesized that direct contact between neurons and breast cancer cells mediates the transfer of cellular components (e.g., proteins, organelles, mRNAs) between each cell type, which may facilitate cancer cell survival and proliferation. Based on these previous experiments, May established a powerful proteomic approach in the lab, which uses azide-tagged noncanonical amino acids (ncAAs) to specifically label newly translated proteins in vitro and in vivo in a cre-dependent manner (see Fig. 1) (Alvarez-Castelao et al., *Nat Biotechnol* **35:** 1196 [2017]). Metabolic nascent proteome labeling with ncAAs is an unbiased method to explore protein exchange between cancer cells in the tumor microenvironment (TME). ncAAs can cross cell membranes and be charged onto me-

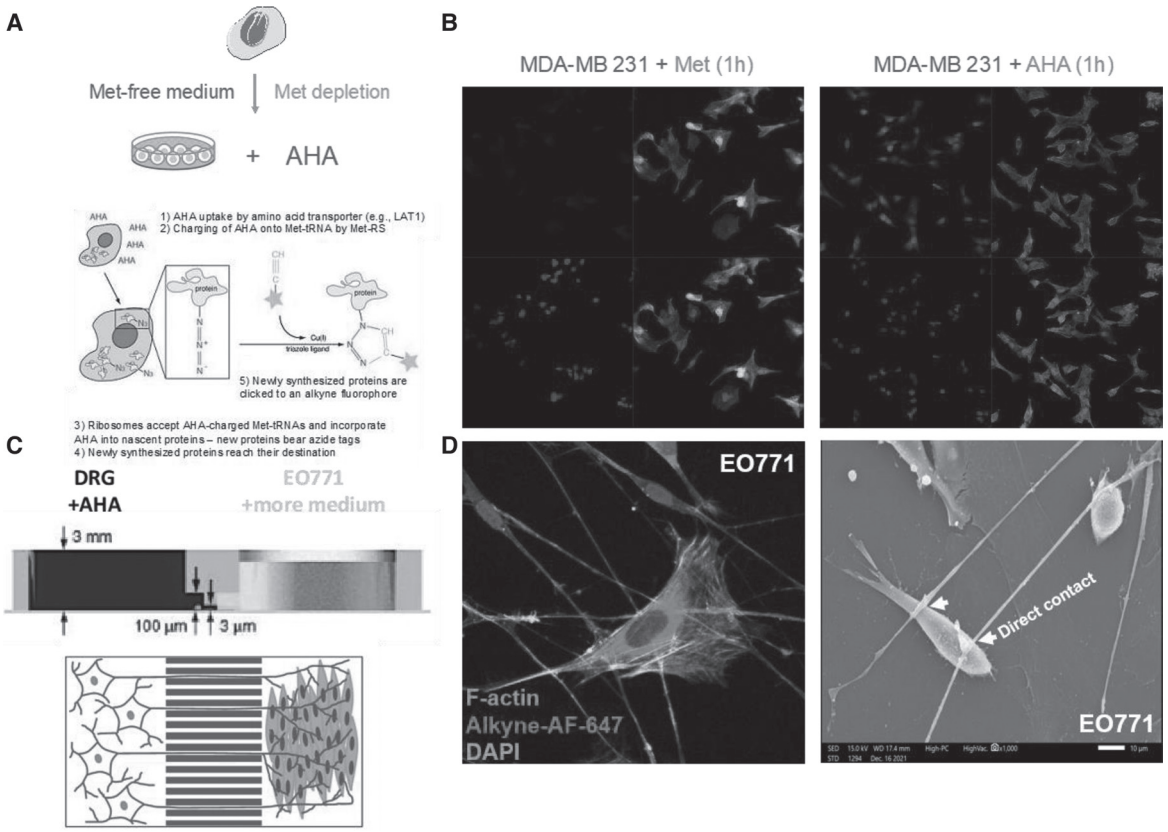


Figure 1. Lateral proteome transfer between dorsal root ganglion (DRG) cells and breast cancer. (*A*) Mechanism of nascent proteome labeling using BONCAT/FUNCAT click chemistry. (*B*) (*Left*) Control experiment demonstrating no labeling (red) when cells (human breast cancer) are cultured in media containing methionine; (*right*) nascent proteome labeling when cells were cultured in media containing the noncanonical amino acid (ncAA) L-azidohomoalanine (AHA) (red). (*C*) Experimental design using a dual-chamber microfluidic device to separate neuronal and cancer cell compartments. Only axons can cross to the cancer cell chamber from the neuronal side through microgrooves. (*D*) (*Left*) Neuronal proteome (red) is visible in cancer cells (EO771) on the other side of the microfluidic device, indicating lateral transfer through axons traversing the microgrooves. (*Right*) Scanning electron microscopy image of EO771/DRG co-cultures demonstrating direct axonal contact between neurons and breast cancer cells.

thionine transfer RNAs (tRNAs) by the endogenous methionyl-tRNA synthetase (MetRS) or by a mutant MetRS*. During protein synthesis, ncAAs will be introduced in the place of methionine (Met), resulting in the introduction of azide groups into the newly synthesized proteins. These azide groups on the ncAAs can be used to tag proteins with either a biotin-alkyne tag (BONCAT) or a fluorescent-alkyne tag (FUNCAT) through a click cycloaddition reaction. Biotin-tagged nascent proteins can be quantified by immunoblot analysis or enriched from the overall proteome pool by affinity purification and identification using liquid chromatography-mass spectrometry (LC-MS).

May tested the technique first in cancer cell or neuronal monocultures. After letting the cells adhere overnight and then replacing complete media with methionine-deficient media, she used the ncAA L-azidohomoalanine (AHA) to label all newly translated proteins in the span of 1 hour. After labeling, cells were fixed and clicked to an alkyne-fluorophore. This reliably labeled all nascent proteins that were translated in the 1-hour time window. Red fluorescence was observable in cell bodies, and in the neuronal cultures, signal could be detected in distal structures like axons and dendrites. May then repeated the experiment except after metabolically labeling neuronal cultures, she stopped the process, washed, and supplemented the cells with methionine-replete media. Then she added breast cancer cells and allowed the cells to interact overnight. To our surprise, May was able to detect nascent neuronal proteins within co-cultured cancer cells, demonstrating lateral proteome transfer between the two cell types. However, these experiments could not determine whether this transfer was an active process or simply the result of leakage or residual AHA labeling after washing. To control for this possibility, May repeated the experiment in a microfluidic chamber that allows for fluidic isolation of neurons and cancer cells separately, allowing contact only via axonal projections through microgroove channels (Fig. 1). AHA was supplemented into the side of the chamber with neuronal soma to label nascent proteins, and then cancer cells were seeded into the right chamber. A larger volume of media was added into the right side to create a hydrostatic gradient preventing any AHA from flowing to the cancer cell side. Even with the hydrostatic pressure preventing AHA diffusion into the cancer cell compartment, we were able to observe neuronal proteins present in cancer cell soma, demonstrating lateral proteome transfer through a direct contact mechanism. We are currently exploring this phenomenon in more detail.

Neuronal Circuits underlying HPA-Axis Circadian Dysfunction in Breast Cancer

A. Gomez, J. Borniger

Circulating glucocorticoids (GCs) (e.g., cortisol, corticosterone) typically display a circadian rhythm in their production and secretion. The circadian release of endogenous glucocorticoids is essential in preparing and synchronizing the body's daily physiological requirements. Disruption in the rhythmic activity of glucocorticoids has been observed in individuals with a variety of cancer types (breast, ovarian, lung, bone), and subsequent blunting of this rhythm predicts cancer mortality and quality of life. This suggests that a disrupted glucocorticoid rhythm is potentially a shared phenomenon across cancers. However, the extent of how this disrupted rhythm can be recapitulated in rodent models, and the causal mechanisms that link glucocorticoid rhythm dysfunction and cancer outcomes, remain preliminary at best.

The neural circuitry regulating daily glucocorticoid activity has been well characterized and is maintained, in part, by the coordinated response of the hypothalamic-pituitary-adrenal (HPA) axis, consisting of the suprachiasmatic nucleus (SCN; the central circadian clock), corticotropin-releasing hormones of the paraventricular nucleus of the hypothalamus (PVN^{CRH}), the anterior pituitary gland, and the adrenal cortex. Consequently, we set out to examine whether we could recapitulate clinical glucocorticoid dysfunction in spontaneous and transplantable syngeneic mouse models of breast cancer progression, and if this is due to cancer-induced disruption within brain hypothalamic nuclei—namely, the SCN and PVN. To date, we have observed a blunting of glucocorticoid rhythms across the day, as measured by fecal corticosterone collection (every 6 hours), across multiple time points during tumor progression. Not only do glucocorticoid levels throughout the day appear lower in response to tumor progression, as compared to their tumor-free baseline, but the slope of the rhythm appears diminished as well, in line with clinical findings.

As central regulators highly influence the outcome of glucocorticoid rhythms, we further examined how peripheral tumors shape hypothalamic activity within the brain. Preliminary fiber photometry data show that PVN^{CRH} neurons exhibit an enhanced response during tumor progression, as compared to baseline (no tumor) activity. In conjunction with the increased PVN^{CRH} activity, we have observed an increase in size of the adrenal cortex during tumor progression. Taken together, this suggests that there may be an overactive HPA response during tumor progression, which, in turn, may result in subsequent negative feedback on glucocorticoid rhythms. Current and future studies aim to (1) examine whether tumor progression disrupts SCN activity, as synchronization of the PVN^{CRH} neurons depends on SCN input; (2) identify the specific neural populations that map onto the tumor by utilizing polysynaptic viral approaches; and (3) determine whether manipulation, by use of optogenetics and CRISPR-Cas9, of the PVN and SCN can restore normal glucocorticoid rhythms in tumor-bearing mice, and whether this can subsequently impact tumor progression. Through these series of experiments, we aim to provide evidence for whether a causal role exists for glucocorticoid rhythms in breast cancer progression.

Neuromodulation of Brainstem Noradrenergic Neurons in Breast Cancer

A. Berisha, J. Borniger

Technician Adrian Berisha is tackling a major question regarding the role of central autonomic output nuclei in noradrenergic regulation of cancer growth and progression. For several years now, evidence demonstrating the role of norepinephrine (noradrenaline) in multiple cancer-related processes such as tumor blood vessel development (Zahalka et al., *Science* **358:** 321 [2017]), immune responses (Devi et al., *Immunity* **54:** 1219.e7 [2021]), and metabolic reprogramming (Thaker et al., *Nat Med* **12:** 939 [2006]; Sloan et al., *Cancer Res* **70:** 7042 [2010]) has accumulated. All of this work, however, used local norepinephrine administration, direct sympathetic nerve stimulation, or psychological stress to boost autonomic tone. This leaves a major hole in our understanding of adrenergic regulation of cancers. That is, how does the brainstem locus coeruleus (LC), the major source of norepinephrine in the brain (LC^{NE}), contribute to downstream cancer outcomes?

Using optogenetics, pseudorabies virus tract tracing, and fiber photometry, Adrian is examining how manipulation of LC^{NE} neurons alters breast cancer progression, how cancer itself alters the activity of these neurons, and how these cells are polysynaptically connected to the mammary gland (site of tumorigenesis). Preliminary studies demonstrate that daily stimulation (5 Hz optogenetic stimulation for 20 sec every minute for 1 hour) of LC^{NE} neurons promotes breast cancer growth in a syngeneic model. Additionally, Adrian observed pseudorabies-labeled neurons in the LC after injection of the retrograde tracer into the mammary gland, suggesting that there is a direct (synaptic) connection between the LC and mammary gland that may be functionally relevant to its influence over cancer outcomes. Adrian is currently increasing the number of biological replicates in these experiments and is setting up flow cytometry protocols to examine how manipulation of the LC^{NE} circuit influences intratumor immunity.

Internal Collaborations

N. Francis, Y. Wu, A. Berisha, A. Gomez, J. Borniger

Postdoc Nikita Francis is working with Mary Doherty in CSHL Professor Lloyd Trotman's laboratory to map, monitor, and manipulate nerves within the prostate and at prostate cancer metastatic sites (primarily liver).

Technician Adrian Berisha is collaborating with Julia Wang in CSHL Assistant Professor Tatiana Engel's laboratory to develop automated electroencephalography/electromyography (EEG/EMG) classification protocols to identify typical and atypical brain oscillatory states in healthy mice and those with cancer cachexia. The cachexia model implemented was supplied and coordinated by CSHL Assistant Professor Tobias Janowitz. Adrian is further collaborating with Professor Linda Van Aelst's group to examine EEG data from mice lacking oligophrenin-1 (OPHN1) globally or in specific brain cells.

Several members of the laboratory have participated in a large collaboration with CSHL Associate Professor Pavel Osten's laboratory to map the brain-wide neural activity response to breast cancer development in an unbiased manner using serial two-photon tomography (STPT). Our group is additionally characterizing sleep changes in mice lacking microglia in collaboration with CSHL Assistant Professor Lucas Cheadle's laboratory.

External Collaborations

A. Berisha, A. Gomez, J. Borniger

Adrian Gomez is leading a collaboration with Northwell Health Professor Kevin Tracey to understand how the nervous system conveys inflammatory signals from the periphery to brain centers controlling body temperature and sleep. Specifically, he is focused on vagal afferent signals and how they are relayed to the median preoptic nucleus (MnPO) in response to systemic IL-1 or LPS.

Adrian Berisha is collaborating with Stanford Assistant Professor Erin Gibson's laboratory to examine how cytotoxic chemotherapy (i.e., methotrexate) influences sleep/wake states and circadian rhythms. The goal is to find a cellular mechanism contributing to sleep and circadian disruption in response to this toxic treatment.

PUBLICATIONS

Borniger JC. 2021. Cancer as a tool for preclinical psychoneuroimmunology. *Brain Behav Immun Health* 1**8:** 100351.

Borniger JC, de Lecea L. 2021. Peripheral lipopolysaccharide rapidly silences REM-active LH^{GABA} neurons. *Front Behav Neurosci* **15:** 26.

Francis N, Borniger JC. 2021. Cancer as a homeostatic challenge: the role of the hypothalamus. *Trends Neurosci* **44:** 903–914.

Silva T, Borniger JC, Alves MJ, Correa DA, Zhao J, Fadda P, Toland AE, Takakura AC, Moreira TS, Czeisler CM, Otero JJ. 2021. Machine learning approaches reveal subtle differences in breathing and sleep fragmentation in *Phox2b*-derived astrocyte ablated mice. *J Neurophysiol* **125:** 1164–1179.

THE MICROENVIRONMENT: HOST CONTRIBUTIONS TO METASTASIS AND COVID-19

M. Egeblad J. Adrover, J. Curtis, J. Daßler Plenker, L. Foerschner, T. Fujii, X. Han, X. He, D. Ng, H. Seidner, M. Shevik, N. Sivetz, L. Sun

Solid tumors are aberrant tissues 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, but become dysregulated during disease, including cancer and COVID-19.

In solid tumors, the stroma is also known as the tumor microenvironment. The Egeblad laboratory studies 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, a technique known as intravital imaging. This allows us to study how cancer cell proliferation, migration, and survival are influenced by stromal components in real time. Since 2020, we have also applied our knowledge of innate immune cell functions to understand the pathophysiology of COVID-19.

Activating a Collaborative Innate-Adaptive Immune Response to Control Metastasis

L. Sun, X. He, D. Ng [in collaboration with S. Adams, NYU; I. McNeish, Imperial College London]

It was discovered more than a century ago that intratumoral injection of dead bacteria, named "Coley's toxin" 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 (IFN-γ). 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 IFN-γ. First, we determined that MPLA with IFN-γ 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 + IFN-γ treatment (but not with either MPLA or IFN-γ given alone) killed ~70%–90% of cancer cells in 48 h. Next, we determined that intratumoral injection of MPLA with IFN-γ 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 IFN-γ 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 d with control treatment, to 106 d when mice were treated with MPLA and IFN-γ, and extended beyond our preset study end point of 5 mo when MPLA and IFN-γ treatment was combined with chemotherapy.

In vitro and in vivo analysis revealed the cellular mechanisms responsible for the effects of the combined MPLA + IFN-γ 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, we could document that $CD8^+$ T cells were stimulated through macrophage-secreted interleukin 12 (IL-12) and tumor necrosis factor α (TNF-α) when the macrophages had been treated with MPLA + IFN-γ. This suggests that MPLA + IFN-γ may also indirectly 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 + IFN-γ in vivo. MPLA and IFN-γ 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. More details can be found in Sun et al. (2021).

Longitudinal Intravital Imaging through Flexible, Clear, Silicone Windows

L. Maiorino, M. Shevik, X. Han, J. Adrover, H. Seidner, L. Foerschner [in collaboration with D. Tuveson, CSHL]

Intravital microscopy (IVM), the imaging of tissues in anesthetized animals, offers insights into the dynamics of physiological and pathological events at cellular resolution in intact tissues. IVM enables visualization of cell movement, division, and death at single-cell resolution. IVM has therefore been instrumental in the cancer biology field to help elucidate how cancer cells invade tissues and metastasize, interact with the surrounding microenvironment, and respond to drugs. In addition, IVM has been key to advancing our understanding of the complex mechanisms governing immune responses by providing insights complementary to ex vivo profiling approaches (e.g., flow cytometry). For instance, intravital imaging experiments have revealed details about immune functions as they relate to cell migration and cell–cell contacts and have offered a platform to quantitate spatiotemporal dynamics in response to injury or infection. Many of these processes can also be studied through histological staining, but only IVM allows the tracking of dynamic changes. Whereas a histological section offers a snapshot of the tissue at a given time, intravital imaging can track intercellular and subcellular events within the same tissue over time. Most IVM techniques are based on fluorescence microscopy, which because of light scattering makes imaging deeper tissues challenging. The tissue of interest, therefore, often needs to be surgically exposed with an often invasive and terminal procedure. Thus, depending on the organ site, the tissue can be imaged continuously for a period varying from a few to 40 h. Alternatively, the surgical insertion of a permanent imaging window permits the imaging of the same tissue sequentially over a period of days to weeks. Typical imaging windows comprise a glass coverslip in a biocompatible metal frame sutured to the mouse's skin. The prototypical intravital imaging window has so far been a metal ring containing a glass coverslip that is secured to the skin with sutures. Interference with free movement, the accumulation of exudate, and damage to the glass coverslip are common problems with using the prototypical metal-frame IVM windows, problems that may necessitate euthanasia. Moreover, the metal-frame window requires specialized production, and the surgical procedure can require extensive training. To address these issues, windows for long-term abdominal organ and mammary gland imaging were developed from a thin film of polydimethylsiloxane (PDMS), an optically clear silicone polymer previously used for cranial imaging windows. These windows can be glued directly to the tissues, reducing the time needed for insertion and making the windows highly versatile in terms of what organs can be imaged. PDMS is flexible, contributing to its improved durability in mice over time—at least a month. The Egeblad laboratory developed a method to cast the PDMS window around a stainless-steel grid to provide landmarks for repeated imaging of the same tissue regions. Longitudinal imaging is imaging of the same tissue region during separate sessions, and the embedded stainless-steel grid allowed localization of the same region for imaging days apart. Processes as diverse as mammary gland involution and the development of metastases from single disseminated cancer cells in the liver were visualized using this new window.

This new approach overcomes some of the most common issues with currently used metal-framed IVM windows. In the coming years, the technology will be employed to understand how a metastasis-supporting tumor microenvironment is established.

Stress-Induced Metastasis

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

Increasing evidence suggests that both intrinsic (genetic and epigenetic) changes in the cancer cells and extrinsic changes affecting the organism or the microenvironment can drive disseminated 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 adrenocorticotrophic 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 metastases, 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 to 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. Since 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. More details can be found in Almeida et al. (2021).

Neutrophil Extracellular Traps and COVID-19

J. Adrover, M. Shevik, J. Daßler Plenker [in collaboration with R. Schwartz, Weill Cornell Medicine; B. tenOever, Mount Sinai and NYU]

With the onset of the COVID-19 pandemic, we realized that the disease pathogenesis of COVID-19 had critical similarities to that of pathologies that are well documented to involve excess NET formation, including lung inflammation, acute respiratory distress syndrome (ARDS), and immunothrombosis. We were the first to demonstrate that NET levels are elevated in blood from patients with severe COVID-19, and that high NET levels correlate with poor prognosis. In addition, we were the first group to demonstrate NETs in lung tissue from autopsies of patients with COVID-19 and to show that NETs are components of atypical microthrombi found in patients with COVID-19.

Severe COVID-19 involves acute lung injury (ALI) or ARDS, which has few treatment options and a high in-hospital mortality rate. Excessive formation of NETs is recognized as a key contributor to ARDS. Upon injury, neutrophils infiltrate the lungs and form NETs. Because NETs are coated with granule-derived proteins, including proteases and histones that are highly cytotoxic, they can inflict severe damage on lung tissue, directly damaging the lung microvasculature, and promote thrombosis, leading to multiorgan damage and cardiovascular complications. Unfortunately, despite the well-recognized role of NETs in ARDS, the only FDA-approved NET-targeting drug is the inhaled drug dornase alfa (which is recombinant DNase I). DNase I can digest NETs present in the airways once they have formed. However, DNase I does not block NET formation or release, and in its inhaled form, it likely has minimal ability to digest NETs beyond the airways. We have now discovered that disulfiram—a drug that has been FDA approved since 1951 for alcohol use disorder—dramatically reduced NETs, increased survival, improved blood oxygenation, and reduced lung edema in a transfusion-related acute lung injury (TRALI) mouse model of ARDS. Disulfiram was recently shown to inhibit gasdermin D in macrophages and to increase survival after experimental sepsis in mice. Gasdermin D has been proposed to be critical for NET formation by forming pores in the nuclear and plasma membranes. Consistently, we found that disulfiram blocked gasdermin D–dependent NET formation efficiently ex vivo. Additionally, disulfiram conferred protection in SARS-CoV-2–infected golden hamsters: disulfiram treatment reduced the levels of NETs and perivascular fibrosis in the lungs, and the treatment also down-regulated innate immune and complement/coagulation pathways.

Our results support the notion that NETs are drivers of severe COVID-19 pathology. Additionally, we identify an existing FDA-approved drug as a NET formation inhibitor that can improve disease course in two rodent models of lung injury for which treatment options are limited—TRALI and several COVID-19s. Disulfiram has been used since 1951 and has a well-understood and generally manageable side effect profile. However, disulfiram treatment is not compatible with alcohol consumption because of its ability to inhibit aldehyde dehydrogenase. Nevertheless, its strong inhibitory effect on NET formation and its improvement of disease outcomes in multiple rodent models highlight the potential for the future development of safe and effective inhibitors of NET formation. More details can be found at Ackermann et al. (2021).

Drivers of Brain Metastasis

T. Fujii

Liver, lung, bone, and brain are the most common sites of breast cancer metastasis. Approximately 15%–30% of patients with metastatic breast cancer develop brain metastasis. The median survival time after patients develop brain metastasis is just 6–12 mo and there is no drug approved specifically for the treatment of brain metastasis; current treatments are limited to radiation and surgery. Therefore, identifying factors that may drive brain metastasis, so that these can be targeted, is a critical clinical need.

Manufactured forms of granulocyte-colony stimulating factor (G-CSF) are given to support neutrophil recovery after chemotherapy that has a high likelihood of causing treatment-related neutropenia, such as anthracycline-based regimens. G-CSF stimulates the bone marrow to produce mature granulocytes as well as the proliferation and differentiation of neutrophil precursors. G-CSF clearly has clinical benefits in the short term, significantly decreasing the risk of febrile neutropenia and neutropenia-related hospitalization. However, the long-term consequences of G-CSF

treatment are less clear. A clinical study showed that high expression levels of endogenous G-CSF in triple-negative breast cancer are associated with poor overall survival. Furthermore, G-CSF blockage significantly reduced lung metastasis in the 4T1 breast cancer mouse model. Several studies have also documented a direct link between G-CSF and formation of NETs. With regard to brain metastasis in breast cancer, a recent study demonstrated that blocking G-CSF resulted in fewer brain metastases in multiple mouse models.

Neutrophils can promote metastasis, including through the formation of NETs. In animal models, G-CSFs can specifically induce NETs to promote liver and lung metastasis. Given the mounting evidence that G-CSF promotes breast cancer metastasis in mouse models by acting on neutrophils, we hypothesized that patients who received G-CSF along with chemotherapy(ies) have higher incidence of brain metastasis than those who did not receive G-CSF. In the nonmetastatic setting, the majority of the patients receive G-CSF as a part of a standard care chemotherapy regimen. Thus, it is difficult to compare long-term outcome for patients who receive chemotherapy(ies) with or without G-CSF. In contrast, G-CSF is not routinely used in the metastatic setting. Therefore, to test our hypothesis, we identified patients at Northwell Health with de novo Stage IV breast cancer, without known brain metastasis at the time of initial diagnosis, from electronic medical records covering the period from 1/1/2013 to 12/31/2020. Univariate and multivariate logistic regression models were used to test the association between variables of interest, including G-CSF use, and brain metastasis. Among the 78 patients analyzed, 24 patients (31%) had received G-CSF along with chemotherapy at least once. In logistic regression models, G-CSF use was not a significant factor to predict brain metastasis ($P = 0.23$), with our study powered to identify a difference in brain metastasis incidence of 25% or above between G-CSF treated and untreated. Unexpectedly, occurrence of pulmonary embolism (PE) or deep venous thrombosis (DVT) was a significant predictive factor of brain metastasis ($P = 0.004$). PE and DVT are seen in patients with elevated NETs, and these results have therefore led us to undertake animal studies to determine whether DVT, PE, or other types of thrombosis, directly or via NETs, can promote the formation of brain metastasis. More details can be found at Fujii et al. (2021).

PUBLICATIONS

Ackermann M, Anders H-J, Bilyy R, Bowlin GL, Daniel C, De Lorenzo R, Egeblad M, Henneck T, Hidalgo A, Hoffmann M, et al. 2021. Patients with COVID-19: in the dark-NETs of neutrophils. *Cell Death Differ* **28:** 3125–3139. doi:10.1038/s41418-021-00805-z

Almeida AS, Fein MR, Egeblad M. 2021. Multi-colour flow cytometry for a comprehensive analysis of the tumour immune infiltrate in a murine model of breast cancer. *Bio-Protocol* **11:** e4012. doi:10.21769/BioProtoc.4012

Fujii T, Rehman H, Chung SY, Shen J, Newman J, Wu V, Hines A, Azimi-Nekoo E, Fayyaz F, Lee M, et al. 2021. Treatment with granulocyte-colony stimulating factor (G-CSF) is not associated with increased risk of brain metastasis in patients with de novo Stage IV breast cancer. *J Cancer* **12:** 5687–5692. doi:10.7150/jca.63159. eCollection 2021

Sun L, Kees T, Almeida AS, Liu B, He XY, Ng D, Han X, Spector DL, McNeish IA, Gimotty P, et al. 2021. Activating a collaborative innate-adaptive immune response to control breast and ovarian cancer metastasis. *Cancer Cell* **39:** 1361–1374. doi:10.1016/j.ccell.2021.08.005

THE IMMUNOLOGY OF MOUSE AND HUMAN PANCREATIC CANCER

D.T. Fearon J. Li R. Yan
P. Moresco J. Yang
J. Pearson M. Yao
Z. Wang

The failure of immunotherapy in human carcinomas correlates with the absence of T cells within nests of cancer cells. We have discovered that this circumstance may be caused by the active exclusion of T cells by stimulation of the chemokine receptor CXCR4 on T cells with the chemokine CXCL12. We extended our earlier finding that inhibiting CXCR4 in mice with pancreatic ductal adenocarcinoma (PDA) with the drug AMD3100/plerixafor caused intratumoral T-cell accumulation by conducting a phase 1 study in patients with either PDA or colorectal cancer (CRC).

Inhibition of the chemokine receptor CXCR4 in combination with blockade of the PD-1/PD-L1 T-cell checkpoint induces T-cell infiltration and anticancer responses in murine and human pancreatic cancer. Inhibiting CXCR4 in this experimental cancer medicine study by 1-wk continuous infusion of AMD3100 induces an integrated anticancer immune response that is detected by transcriptional analysis of paired biopsies of metastases (Biasci et al., *Proc Natl Acad Sci* **117:** 28960 [2020]). This integrated immune response occurs in three other examples of immune-mediated damage to noninfected tissues: rejecting renal allografts, melanomas clinically responding to anti-PD1 antibody therapy, and microsatellite instable colorectal cancers. Thus, signaling by CXCR4 causes immune suppression in human pancreatic ductal adenocarcinoma and colorectal cancer. This result was discussed in an invited clinical oncology journal (Fearon and Janowitz 2021).

We discovered that cancer cells have adapted to this immune suppressive function of CXCR4 by showing that cancer cells of human PDA, CRC, and breast cancer coat themselves with covalent CXCL12-keratin-19 (KRT19) heterodimers. Mouse PDA tumors containing control PDA cells exhibited the CXCL12-KRT19 coating, excluded T cells, and did not respond to treatment with anti-PD-1 antibody. Tumors containing PDA cells not expressing either KRT19 or TGM2 lacked the CXCL12-KRT19 coating, were infiltrated with activated $CD8^+$ T cells, and growth was suppressed with anti-PD-1 antibody treatment. Thus, carcinomas assemble a CXCL12-KRT19 coating to evade cancer immune attack.

PUBLICATIONS

Fearon DT, Janowitz T. 2021. AMD3100/Plerixafor overcomes immune inhibition by the CXCL12-KRT19 coating on pancreatic and colorectal cancer cells. *Br J Cancer* **125:** 149–151. doi:10.1038/s41416-021-01315-y

In Press

Wang Z, Moresco P, Yan R, Li J, Gao Y, Biasci D, Yao M, Pearson J, Hechtman JF, Janowitz T, et al. 2022. Carcinomas assemble a filamentous CXCL12-keratin-19 coating that suppresses T cell–mediated immune attack. *Proc Natl Acad Sci* **119:** e2119463119. doi:10.1073/pnas.2119463119

THE EFFECTS OF DISEASE ON THE WHOLE BODY

T. Janowitz B. Demestichas S. Kleeman N. Mourikis X. Zhao
M. Ferrer Gonzalez C. Koza E. von Gablenz

Our work focuses on the body's dynamic response to disease—in particular, to cancer. Utilizing laboratory and clinical research paired with statistical modelling and bioinformatics, we decipher the connections between metabolism, behavior, endocrinology, and immunology. This approach reflects the idea that cancer cannot be understood or treated by investigating tumors in isolation.

Cancer—A Disease of the Whole Body

We are conducting a comprehensive body of work examining the interactions among different organ systems during cancer progression, with a focus on neuroendocrine and immunological consequences, in murine models of cancer and in patients with cancer. In a preclinical study, we have discovered that interleukin-6-associated cancer causes systemic neutrophilia and reprograms neutrophil metabolism to induce oxidative glycolysis. This metabolic state is also present in cancer cells. Targeting the pathway uncouples tumor growth and overall survival: tumors grow less fast, but overall survival is reduced. This illustrates that host factors have to be considered during the development of treatment strategies for cancer.

We continue to study the molecular mechanisms by which cancer progression induces behavioral abnormalities, such as reduced appetite, lethargy, and anhedonia.

In recognition of the importance of the whole-body response to cancer, Dr. Janowitz coleads an international group of collaborating scientists to respond to the 2021 Cancer Grand Challenge on cancer-associated cachexia posed by Cancer Research UK and the U.S. National Cancer Institute. Cachexia is a highly debilitating wasting syndrome of the whole body that is experienced by patients with cancers such as colorectal, lung, and pancreatic cancer. The causes are

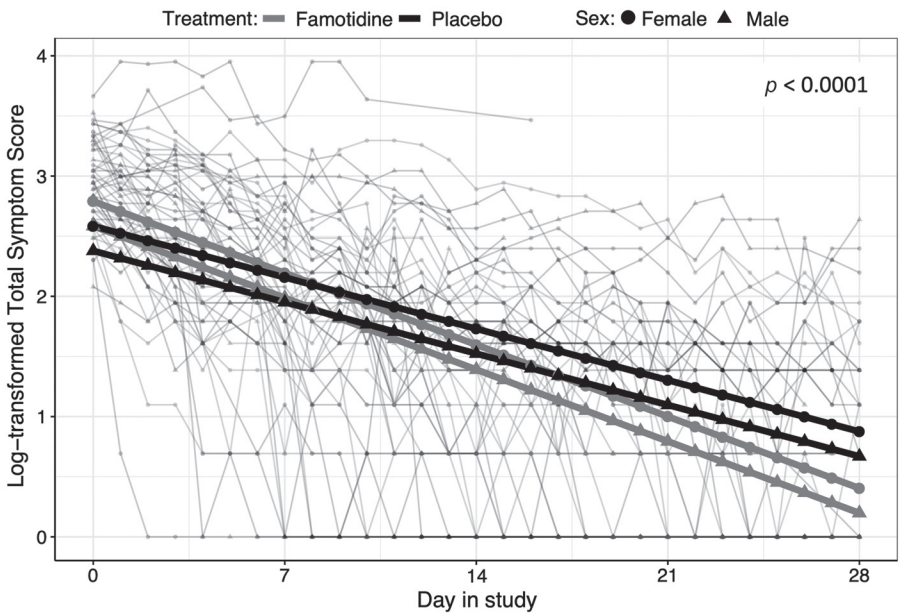


Figure 1. Symptom resolution analysis. The logarithmically transformed patient-level total symptom score (thin lines) and their estimated means based on linear mixed-effect model are shown for each study arm. The *p*-value for the interaction term of group and day in study is displayed.

largely unknown and treatment options are limited. The decision on the proposal is expected in June 2022.

COVID-19 Research

Substantial similarities between the pathways and organismal response pattern in COVID-19 and cancer exist. Therefore, we were able to pivot our work to contribute to treatment research for COVID-19. Sustained increased inflammation and immune activation (cytokine release syndrome) drive poor outcomes for patients with COVID-19. Modulating the host response during COVID-19 infection may offer benefits to the patients. In the last year, we completed a randomized,

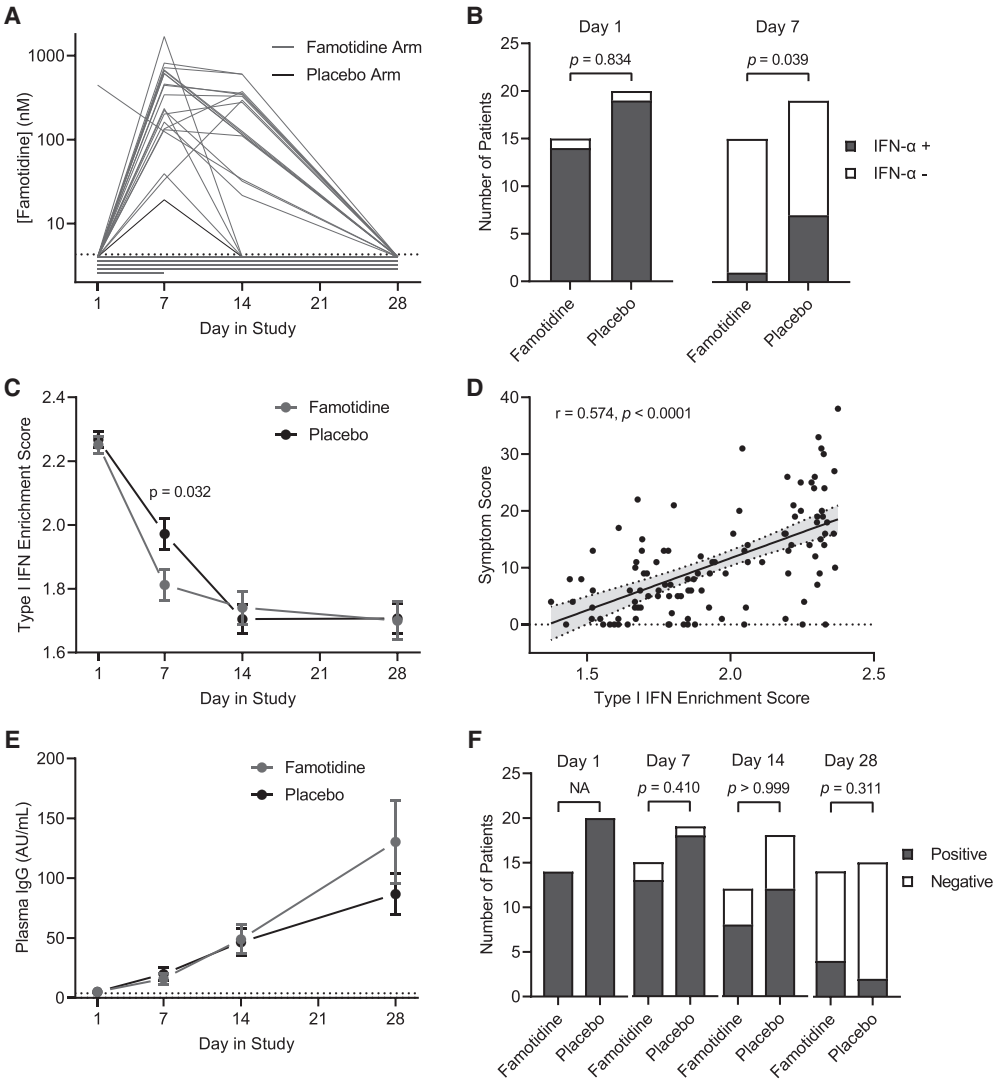


Figure 2. Effect of famotidine on inflammation, symptoms, and immunity. (*A*) The longitudinal plasma famotidine levels are displayed for patients enrolled in the famotidine arm and for the patient from the placebo arm with detectable plasma famotidine. (*B*) Participant numbers with detectable plasma interferon α levels in each arm at day 1 and day 7 of the trial are shown. Statistical comparison by chi-square test. (*C*) Enrichment scores for type I interferon response genes expressed in peripheral blood mononuclear cells (PBMCs) at days 1, 7, 14, and 28 are shown. (*D*) The correlation of symptom scores and enrichment scores for type I interferon is assessed. Statistical comparison by Spearman rank analysis. (*E*) The levels (mean and standard error of the mean) of class g immunoglobulins reactive to the SARS-CoV-2 core protein are plotted for days 1, 7, 14, and 28 for each study arm. (*F*) The number of study participants with reverse transcriptase polymerase chain reaction (RT-PCR)-detectable viral RNA extracted from nasal swabs on days 1, 7, 14, and 28 are shown for each trial arm. (IFN) interferon, (IgG) immunoglobulin type G.

placebo-controlled, double-blind, phase 2 clinical trial studying the effect of famotidine on symptomatic recovery and resolution of inflammation in patients with mild to moderate COVID-19 (NCT04724720). This trial was conducted in close collaboration with Northwell Health and the Feinstein Institutes for Medical Research.

Famotidine is a low-cost and safe histamine 2 receptor antagonist that is globally available. In laboratory studies, famotidine reduced type I interferon release from virally infected epithelial cells. In our trial, oral famotidine was safe and well tolerated. The primary end point was not reached. Based on the secondary end points, patients on famotidine experienced more rapid symptom resolution (Fig. 1) and earlier resolution of type I interferon levels and inflammation (Fig. 2A–D). Importantly, the emergence of an immune reaction against SARS-CoV-2 was not diminished (Fig. 2E–F).

We collected high-density data from patients. Therefore, our study may serve as a template for studies with fewer patients. This may improve access to clinical trial participation from more institutions globally and require less resource.

The study was a phase 2 trial and, therefore, additional clinical trials are needed and may leverage the knowledge generated by this study.

PUBLICATIONS

Fearon DT, Janowitz T. 2021. AMD3100/Plerixafor overcomes immune inhibition by the CXCL12-KRT19 coating on pancreatic and colorectal cancer cells. *Br J Cancer* **125:** 149–151. doi:10.1038/s41416-021-01315-y

Kolamunnage-Dona R, Berhane S, Potts H, Williams EH, Tanner J, Janowitz T, Hoare M, Johnson P. 2021. Sorafenib is associated with a reduced rate of tumour growth and liver function deterioration in HCV-induced hepatocellular carcinoma. *J Hepatol* **75:** 879–887. doi:10.1016/j.jhep.2021.05.01

Munir H, Jones JO, Janowitz T, Hoffmann M, Euler M, Martins CP, Welsh SJ, Shields JD. 2021. Stromal-driven and amyloid β–dependent induction of neutrophil extracellular traps modulates tumor growth. *Nat Commun* **29:** 683. doi:10.1038/s41467-021-20982-2

Wang Z, Moresco P, Yan R, Li J, Gao Y, Biasci D, Yao M, Pearson J, Hechtman JF, Janowitz T, Zaidi RM, Weiss MJ, Fearon DT. 2021. Carcinomas assemble a filamentous CXCL12-keratin-19 coating that suppresses T cell–mediated immune attack. *Proc Natl Acad Sci* **119:** e2119463119. doi:10.1073/pnas.2119463119

Williams EH, Flint TR, Connell CM, Giglio D, Lee H, Ha T, Gablenz E, Bird NJ, Weaver JMJ, Potts H, et al. 2021. CamGFR v2: a new model for estimating the glomerular filtration rate from standardized or non-standardized creatinine in patients with cancer. *Clin Cancer Res* **27:** 1381–1390. doi:10.1158/1078-0432.CCR-20-3201

Yang WH, Qiu Y, Stamatatos O, Janowitz T, Lukey MJ. 2021. Enhancing the efficacy of glutamine metabolism inhibitors in cancer therapy. *Trends Cancer* **7:** 790–804. doi:10.1016/j.trecan.2021.04.003

In Press

Brennan CM, Nadella S, Zhao X, Dima RJ, Jordan-Martin N, Demestichas BR, Kleeman SO, Ferrer M, von Gablenz EC, Mourikis N, et al. 2022. Oral famotidine versus placebo in non-hospitalised patients with COVID-19: a randomised, double-blind, data-intense, phase 2 clinical trial. *Gut* Feb 10. doi:10.1136/gutjnl-2022-326952

METABOLIC REPROGRAMMING DURING TUMORIGENESIS AND METASTASIS

M. Lukey J. de Ruiter-Swain O. Stamatatos
Q. Hu W-H. Yang
Y. Qiu

In the 1920s, Otto Warburg made the seminal discovery that glucose metabolism in tumors differs markedly from that in adjacent healthy tissue. A century of research has now established that metabolic reprogramming is a hallmark of tumorigenesis, essential for fulfilling the biosynthetic, bioenergetic, and redox requirements of malignant growth. Within the tumor microenvironment (TME) and during metastasis, cancer cells encounter fluctuating and restrictive nutrient environments and must therefore possess metabolic flexibility to make use of changing fuel sources. Our laboratory aims to understand the biochemical underpinnings of metabolic reprogramming during tumorigenesis and metastasis, with the goal of identifying context-specific metabolic essentialities that can be selectively targeted for cancer therapy sparing the normal metabolic physiology of the host.

Tumor Adaptation to Glutamine Blockade

The diversity and flexibility of tumor metabolism means that there is no universal metabolic phenotype of cancer. However, a small number of metabolic characteristics, such as increased consumption of glucose and glutamine and a shift toward fermentation alongside respiration, are broadly conserved across most cancer types. Glutamine is the most abundant amino acid in blood plasma, and its critical importance for cancer cell proliferation has been known since the establishment of the first cancer cell lines in the 1950s. Over the last decade there have been intense research efforts to target tumor glutamine metabolism for cancer therapy, but to date the results from clinical trials have been largely disappointing. We have been working to define the intrinsic determinants of "glutamine addiction" across the different molecular subtypes of breast cancer, and we are also exploring how highly glutamine-avid basal subtype breast cancers adapt to glutamine restriction within the TME or to pharmacological blockade of glutamine catabolism. Remarkably, in both cases, cancer cell survival becomes strictly dependent on a cytosolic pathway for serine biosynthesis that is dispensable under normal conditions. We have found that in this biological context, serine production itself is not the key function of the pathway, but instead a side product is able to compensate for the loss of glutamine-derived carbon. Since serine biosynthesis is a pharmacologically tractable process, we have been able to use our findings to develop a synergistic combination therapy approach that prevents breast cancer cell adaptation to glutamine blockade. In addition to glutamine's role as a major carbon source for cancer cells, the amide and amine groups are obligate nitrogen donors in several biosynthetic pathways. In a parallel study, we have traced a previously unreported nutrient scavenging mechanism that allows breast and pancreatic cancer cells to compensate for this loss of nitrogen supply during glutamine depletion in the TME.

Metabolic Adaptations and Interactions in the Brain Metastasis TME

The development of brain metastases is typically an end-stage event in cancer progression, with median survival after diagnosis measured in months regardless of the primary cancer type. However, brain metastasis is a biologically inefficient process, with cancer cells that traverse the blood–brain barrier showing a particularly high rate of attrition. This is likely due in part to the hostile nutrient environment encountered, with many nutrients, including glucose and glutamine, present at much lower levels in the brain interstitial fluid (ISF) than in blood plasma or the ISF of other tissues. Consequently, brain-metastatic cancer cells appear to undergo convergent evolution such that brain metastases arising from diverse primary tumors exhibit a conserved metabolic phenotype.

We are investigating the cancer cell–intrinsic metabolic adaptations required for growth in the brain and also the paracrine metabolic interactions that occur between metastatic cancer cells and brain-resident cells including both glial and neuron populations.

To facilitate these studies, we have developed a culture medium that models the nutrient composition of the brain ISF. Consistent with selection for cell-intrinsic metabolic reprogramming during metastasis, we observed that brain-metastatic triple-negative breast cancer (TNBC) cells readily grow in this physiological medium, whereas matched parental TNBC cells fail to proliferate. This has allowed us to conduct a whole-genome CRISPR-Cas9 screen to identify genes that are selectively essential for growth in physiological nutrient conditions, relative to the supraphysiological nutrient concentrations provided by standard culture media. In addition to cell-intrinsic adaptation, we have found that other cell types present within the brain metastasis TME provide metabolic support to cancer cells. To understand this phenomenon, we have applied a combination of untargeted metabolomics and stable isotope tracing approaches, which have revealed nutrient cycles between tumor and stromal cells that result in net transfer of nitrogen into the cancer cells.

Intersections of Cellular Signal Transduction and Metabolism

Most oncogenic signaling pathways drive changes in cellular metabolism, with the tumor oncogenotype being a key determinant of its ultimate metabolic phenotype. Reciprocally, numerous metabolites are able to influence cellular signal transduction pathways—for example, by providing precursors for protein prenylation, altering gene expression downstream of epigenetic changes, or by direct allosteric modulation of signaling proteins. Phospholipase enzymes lie at the intersection of signal transduction and metabolism, hydrolyzing phospholipids into bioactive signaling lipids that regulate diverse physiological processes. We have identified a new role for phospholipase signaling in the inner mitochondrial membrane (IMM), whereby the lipid product of the phospholipase reaction directly binds and activates IMM-localized metabolic enzymes. These associations occur at nanomolar concentrations and show exquisite specificity regarding the lipid tail length and presence/position of unsaturated bonds. In the case of mitochondrial glutaminase (GLS), we have identified lipids that can compete with and displace clinical GLS inhibitors, potentially constituting a resistance mechanism to this class of drug. We are therefore evaluating combination therapies that target GLS and phospholipase signaling, thereby shutting down a key metabolic pathway for cellular proliferation and simultaneously blocking a potential mechanism of resistance to this approach.

PUBLICATIONS

Yang W-H, Qiu Y, Stamatatos O, Janowitz T, Lukey MJ. 2021. Enhancing the efficacy of glutamine metabolism inhibitors in cancer therapy. *Trends Cancer* **7:** 790–804.

In Press

Petruzzelli M, Ferrer M, Schuijs M, Kleeman SO, Hall Z, Perera D, Raghunathan S, Vacca M, Gaude E, Lukey MJ, et al. 2022. Early neutrophilia marked by aerobic glycolysis sustains host metabolism and delays cancer cachexia. *Cancers* **14:** 963. doi: 10.3390/cancers14040963

MASS SPECTROMETRY LABORATORY

D.J. Pappin P. Cifani S. Costa E. Kilic F. Minicozzi

PHAROH lncRNA Regulates Myc Translation in Hepatocellular Carcinoma by Sequestering TIAR

D.J. Pappin [in collaboration with D. Spector, A. Yu, C. Berasain, S. Bhatia, K. Rivera, B. Liu, and F. Rigo, CSHL]

Hepatocellular carcinoma, the most common type of liver malignancy, is one of the most lethal forms of cancer. The Spector laboratory had previously identified a long noncoding RNA (lncRNA), Gm19705, that is overexpressed in hepatocellular carcinoma and mouse embryonic stem cells. This lncRNA was named RNA pluripotency and hepatocyte associated RNA overexpressed in hepatocellular carcinoma (HCC), or PHAROH.

Depletion of PHAROH impacts cell proliferation and migration, which can be rescued by ectopic expression of PHAROH. RNA-sequencing (RNA-seq) analysis of PHAROH knockouts revealed that a large number of genes with decreased expression contained a Myc motif in their promoter. Myc is decreased in knockout cells at the protein level, but not the mRNA level. It is also known that lncRNAs can act as structural scaffolds to promote interaction between protein complexes or to sequester a specific protein. Because modulation of PHAROH levels changes Myc protein levels but not mRNA levels to a significant degree, it was hypothesized that PHAROH might be regulating the translation of MYC through a protein mediator. To search for PHAROH-interacting proteins, the Spector laboratory used a pull-down method adapted from a previously published RNA antisense purification-mass spectrometry technique (RAP-MS). In lieu of pooling all available antisense capture biotinylated oligonucleotides (oligos), it was reasoned that individual oligos might be equally effective and could be used as powerful biological replicates. In addition, oligo-specific off-target effects might be minimized by verifying results with multiple oligos. The Spector laboratory screened through five 20-mer 3′-biotinylated DNA oligos that tiled the length of PHAROH and found that four out of the five oligos pulled down >80% of endogenous PHAROH, whereas the pull-down of a control RNA (PPIB) remained low. MS analysis by the Pappin laboratory identified this RNA-antisense pull-down as nucleolysin TIAR, a translational repressor, which binds to a 71-nt hairpin within PHAROH. The data suggest that PHAROH regulates MYC translation by sequestering TIAR, and as such represents a potentially exciting diagnostic or therapeutic target in hepatocellular carcinoma.

Oncogenic KRAS Engages an RSK1/NF1 Pathway to Inhibit Wild-Type RAS Signaling in Pancreatic Cancer

D.J. Pappin [in collaboration with D. Cheng, T. Oni, J. Thalappillil, Y. Park, H-C. Ting, B. Alagesan, N. Prasad, K. Addison, K. Rivera, L. Van Aelst, and D. Tuveson, CSHL]

Pancreatic ductal adenocarcinoma (PDAC) is a lethal malignancy with limited treatment options. Although activating mutations of the KRAS GTPase are the predominant dependency present in >90% of PDAC patients, targeting KRAS mutants directly has been challenging. Similarly, strategies targeting known KRAS downstream effectors have had limited clinical success because of feedback mechanisms, alternative pathways and dose-limiting toxicities in normal tissues. Therefore, identifying additional functionally relevant KRAS interactions in PDAC might allow for a better understanding of feedback mechanisms and unveil new therapeutic targets. The Tuveson and Pappin laboratories used proximity labeling to identify protein interactors of active KRAS in PDAC cells. They expressed fusions of wild-type (WT) (BirA-KRAS4B), mutant (BirA-KRAS4B^{G12D}), and nontransforming cytosolic double mutant (BirA-KRAS4B$^{G12D/C185S}$) KRAS with the BirA biotin ligase in murine PDAC cells. Mass spectrometry analysis revealed that RSK1 selectively interacted with membrane-bound KRAS^{G12D}, and it was shown that this interaction required NF1 and

SPRED2. It was also determined that membrane RSK1 mediated negative feedback on WT RAS signaling and impeded the proliferation of pancreatic cancer cells upon the ablation of mutant KRAS. The findings link NF1 to the membrane-localized functions of RSK1 and highlight a role for wild-type RAS signaling in promoting adaptive resistance to mutant KRAS-specific inhibitors in PDAC.

Prostate Tumor–Induced Stromal Reprogramming Generates Tenascin C That Promotes Prostate Cancer Metastasis

D.J. Pappin [in collaboration with K. Rivera, CSHL; S-H. Lin, Y-C. Lee, S-C. Lin, G. Yu, M. Zhu, J.H. Song, C. Logothetis, T. Panaretakis, G. Wang, and L-Y. Yu-Lee, MD Anderson Cancer Center]

Metastatic prostate cancer (PCa) in bone induces bone-forming lesions that enhance the progression of prostate cancer. The Lin laboratory had previously shown that PCa-induced bone originates from endothelial cells (ECs) that have undergone endothelial-to-osteoblast (EC-to-OSB) transition induced by tumor-secreted bone morphogenetic protein 4 (BMP4). In a collaboration with the CSHL Pappin laboratory, it was shown that EC-to-OSB transition led to changes in the tumor microenvironment that increased the metastatic potential of PCa cells. Conditioned medium (CM) from EC-OSB hybrid cells increased the migration, invasion, and survival of PC3-mm2 and C4-2B4 PCa cells. Quantitative iTRAQ mass spectrometry by the Pappin laboratory identified tenascin C (TNC) as one of the major proteins secreted from EC-OSB hybrid cells. TNC expression in tumor-induced OSBs was confirmed by immunohistochemistry of MDA PCa-118b xenograft and human bone metastasis specimens. Mechanistically, BMP4 increases TNC expression in EC-OSB cells through the Smad1-Notch/Hey1 pathway. In vitro studies showed that a TNC-neutralizing antibody inhibits EC-OSB-CM-mediated PCa cell migration and survival. TNC knockdown decreased, whereas the addition of recombinant TNC or TNC overexpression increased migration and anchorage-independent growth of PC3 or C4-2b cells. When injected orthotopically, PC3-mm2-shTNC clones decreased metastasis to bone, whereas C4-2b-TNC-overexpressing cells increased metastasis to lymph nodes. Mechanistically, it was found that TNC enhances PCa cell migration through $\alpha5\beta1$ integrin-mediated YAP/TAZ inhibition. These studies show that tumor-induced stromal reprogramming generates TNC that enhances PCa metastasis and suggest that TNC may be a target for PCa therapy.

PUBLICATIONS

Cheng DK, Oni TE, Thalappillil JS, Park Y, Ting HC, Alagesan B, Prasad NV, Addison K, Rivera KD, Pappin DJ, et al. 2021. Oncogenic KRAS engages an RSK1/NF1 pathway to inhibit wild-type RAS signaling in pancreatic cancer. *Proc Natl Acad Sci* **118:** e2016904118. doi:10.1073/pnas.2016904118

Lee YC, Lin SC, Yu G, Zhu M, Song JH, Rivera K, Pappin DJ, Logothetis CJ, Panaretakis T, Wang G, et al. 2021. Prostate tumor–induced stromal reprogramming generates tenascin C that promotes prostate cancer metastasis through YAP/TAZ inhibition. *Oncogene* **41:** 757–769. doi:10.1038/s41388-021-02131-7

Yu AT, Berasain C, Bhatia S, Rivera K, Liu B, Rigo F, Pappin DJ, Spector DL. *PHAROH* lncRNA regulates Myc translation in hepatocellular carcinoma via sequestering TIAR. *eLife* **10:** e68263. doi:10.7554/eLife.68263

PROTEIN TYROSINE PHOSPHATASES AND THE CONTROL OF SIGNAL TRANSDUCTION

N.K. Tonks S. Blockhuys L. Christensen Q. Hu Z. Qian P. Venkataramani
Y. Cen C. Felice V. Mandati 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 in the 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 the family of protein tyrosine phosphatases (PTPs), which Tonks discovered more than 30 years ago. 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 protein tyrosine phosphatase family of enzymes. Through this basic research, 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, with the restrictions resulting from the COVID pandemic, Stephanie Blockhuys decided to return to Sweden.

A New Dimension to Our Substrate-Trapping Technology

We have continued our studies on the impact of PTP1B inhibitors on models of HER2-positive breast cancer. The ability to define functional interactions between enzymes and their substrates is crucial for understanding biological control mechanisms; however, such methods face challenges in the transient nature and low stoichiometry of enzyme–substrate interactions. Now, we have developed a strategy that couples substrate-trapping mutagenesis to proximity-labeling mass spectrometry for quantitative analysis of protein complexes involving the protein tyrosine phosphatase PTP1B. Advantages of this new approach have been illustrated by application to PTP1B interaction networks in models of HER2-positive breast cancer and Herceptin resistance. Employing this methodology, we have identified multiple protein targets of PTP1B with established links to HER2-induced signaling and Herceptin resistance mechanisms. We have further demonstrated that inhibitors of PTP1B significantly reduced proliferation and viability in cell-based models of acquired and de novo Herceptin resistance in HER2-positive breast cancer. Overall, this opens up new possibilities for therapeutic application of PTP1B inhibitors in Herceptin-resistant HER2-positive cancers. Furthermore, this versatile approach can be readily integrated with evolving proximity-labeling platforms (TurboID, BioID2, etc.) and is broadly applicable across all PTP family members for the identification of conditional substrate specificities and signaling nodes in models of human disease. This work is currently under review.

CSH9885: A Small Molecule That Stabilizes the Reversibly Oxidized Form of PTP1B

Previously, we developed conformation-sensor antibodies that recognize the reversibly oxidized form of PTP1B (PTP1B-OX) selectively and stabilize this inactive state, inhibiting its reactivation by reducing agent and thereby inhibiting phosphatase activity. We demonstrated that expression of these recombinant antibodies in cells enhanced insulin-induced signal

transduction. Furthermore, we identified small molecules that could mimic the effects of the PTP1B-OX-directed conformation-sensor antibodies. In particular, we identified chelerythrine, which led to enhanced insulin and leptin signaling in cell models in a manner that was abrogated by expression of catalase, or NOX inhibitors, which suppress H_2O_2 levels. Furthermore, treatment of animal models with chelerythrine improved glucose tolerance and insulin sensitivity, coincident with enhanced phosphorylation of the insulin receptor β-subunit and AKT in liver. We are now preparing to submit a paper describing the development of an in vitro assay to measure stabilization of the reversibly oxidized form of PTP1B, which was used to screen an industrial-scale library of small molecules. The manuscript describes the primary screening strategy and the secondary assays that were integrated as quality control steps, to demonstrate target engagement and eliminate "bad actors" that cause aggregation, covalent modification etc. Four of the top candidates are highlighted, with a particular focus on CSH9885. We show that this compound engaged PTP1B-OX with improved specificity relative to chelerythrine. Furthermore, CSH9885 promoted insulin signaling in cell models and enhanced glucose homeostasis in the High-Fat Diet–fed mouse model of diet-induced obesity (DIO), through a mechanism based on stabilization of PTP1B-OX. This work illustrates a novel paradigm for inhibiting the signaling function of PTP1B. A unique feature of the approach is that it focuses on the critical pool of PTP1B that is responsible for regulation of insulin and leptin signaling—the pool of the enzyme that is acutely regulated by reversible oxidation following insulin and leptin stimulation. By focusing on this pool of PTP1B, the approach may minimize the potential for complications arising from inhibition of the native enzyme as a whole.

Targeting PTP1B Therapeutically

A current focus of the laboratory remains an examination of the effect of our allosteric inhibitors of PTP1B in various disease contexts. We are continuing to develop DPM-1001, a potent and specific inhibitor of PTP1B that is orally bioavailable and crosses the blood–brain barrier. This compound also exhibits a unique specificity and high affinity for copper that enhances its potency as an inhibitor of PTP1B. We are testing the ability of DPM-1001 to function as a copper chelator, particularly in a cancer context where elevated copper underlies a newly discovered form of "oncogene addiction"—*cuproplasia*. SAR syntheses around the structure of DPM-1001 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. Future studies will focus on further optimizing and characterizing such molecules in various cancer models.

Approaches that alleviate the inhibitory constraints imposed on T cells by cell surface checkpoints, such as PD-1, have revolutionized cancer management, but resistance is common. Tony Tiganis' laboratory (Monash University) has reported that PTP1B antagonizes T cell function and that its abundance is increased in intra-tumoral $CD8^+$ T cells to limit T cell anti-tumor immunity. They showed that deletion of PTP1B in endogenous or adoptively transferred T cells including chimeric antigen receptor (CAR) T cells, repressed tumor growth and enhanced the response to anti-PD-1 therapy. In collaboration with the Tonks lab, they showed similar results following treatment with our allosteric inhibitors of PTP1B. These findings identify PTP1B as an intracellular checkpoint that limits the cytotoxic activity of tumor-infiltrating T cells. Its deletion or inhibition can enhance the anti-tumor immunity of T cells and CAR T cells, and the response to anti-PD1 therapy. These studies provide further validation of PTP1B as a therapeutic target for cancer immunotherapy and we are continuing to work with Tony Tiganis to test our allosteric inhibitors of PTP1B in various cancer immunotherapy models. We propose that our studies highlight the potential for our inhibitors of PTP1B to exert simultaneous effects that are both tumor-intrinsic, via chelation of copper and inhibition of critical effects of PTP1B within the tumor, and tumor-extrinsic, via relieving the immune checkpoint function of PTP1B. This dual mechanism of action—a "combination inhibitor in one molecule"—may improve the therapeutic potential of PTP1B inhibitors.

In addition, we continue to assess the role of PTP1B, and the impact of our PTP1B inhibitors, in Rett syndrome, an X-linked neurological disorder

presenting with autistic features, which is caused primarily by mutations in a transcriptional regulator, methyl CpG binding protein 2 (MECP2). Previously, we demonstrated 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, which, in turn, disrupted cellular signaling. Pharmacological inhibition of PTP1B, with multiple structurally and mechanistically distinct small-molecule inhibitors, ameliorated the effects of MECP2 disruption on various phenotypes in mouse models of RTT. Currently, we have established a standardized protocol for administration of PTP1B inhibitors to our animal model and for examining their impact on movement, respiration, EKG, and other behavioral tests. Furthermore, we are integrating these studies with a proteomic analysis of effects on cell signaling. Taken together, our 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. We are working with DepYmed Inc. to take our PTP1B inhibitors into clinical trials for treatment of Rett syndrome patients. Furthermore, we are now extending these investigations to exploit PTP1B in a new approach to understanding and targeting aspects of neurodegenerative disease and inflammation.

Acute lung injury (ALI) can cause acute respiratory distress syndrome (ARDS), a lethal condition with limited treatment options that was highlighted as a common global cause of death due to COVID-19. When the pandemic took hold, we were wondering whether there was anything we could do to help, to provide an understanding of this aspect of the COVID disease and to suggest ways it could be treated. ARDS secondary to transfusion-related acute lung injury (TRALI) has been recapitulated preclinically by anti-MHC-I antibody administration to LPS-primed mice. In this model, we demonstrated that inhibitors of PTP1B prevented lung injury and increased survival. Treatment with PTP1B inhibitors attenuated the aberrant neutrophil function that drives ALI, and was associated with release of myeloperoxidase, suppression of neutrophil extracellular trap (NET) formation, and inhibition of neutrophil migration. Mechanistically, reduced signaling through the CXCR4 chemokine receptor, particularly to the activation of PI3Kγ/AKT/mTOR, was essential for these effects, linking PTP1B inhibition to promoting an aged-neutrophil phenotype (see Fig. 1). Considering dysregulated activation of neutrophils has been implicated in sepsis and causes collateral tissue damage, we demonstrated also that PTP1B inhibitors improved survival and ameliorated lung injury in an LPS-induced sepsis model and improved survival in the cecal ligation and puncture (CLP)-induced sepsis model. These data highlight the potential for PTP1B inhibition to prevent ALI and ARDS from multiple etiologies, and this effect of PTP1B inhibitors on neutrophil function has now opened up new opportunities in cancer.

Defining the Function of Receptor PTPδ (PTPRD) as a Potential Drug Target in Cancer

Current models suggest that receptor PTPs may exist either as active monomers or inactive homodimers, in which activity is suppressed due to reciprocal inhibition—the catalytic center of one PTP domain is occluded by a wedge motif from the apposing domain in the dimer. Essentially, this is the reciprocal of ligand-induced dimerization and activation of receptor PTKs. A major unresolved issue in the PTP field is the extent to which RPTP function is regulated by dimerization (potentially in response to ligand binding) in vivo and the extent to which this may be exploited to manipulate RPTP activity for therapeutic benefit.

Our studies have focused on a receptor PTP, PTPRD. Depending on context, PTPRD has been shown to elicit tumor-suppressive or tumor-promoting effects, the latter associated with activation of the proto-oncogene SRC. PTPRD dephosphorylates the inhibitory carboxy-terminal phosphorylation site of SRC, relieving auto-inhibition and activating downstream signals. PTPRD induced transformation of breast epithelial cells, including malignant morphological change and invasive growth. Furthermore, expression analyses have shown that PTPRD is upregulated in patients with metastatic breast cancer and is associated with poor survival rate. We have shown that following chemically induced dimerization of DmrB-tagged PTPRD in 293T cells, activation of SRC was suppressed and the PTPRD protein was degraded. In a complementary approach, we worked in collaboration with Johannes Yeh and the Antibody Shared Resource to generate antibodies that target the extracellular segment of PTPRD, with the goal of manipulating the dimerization status and activity of the

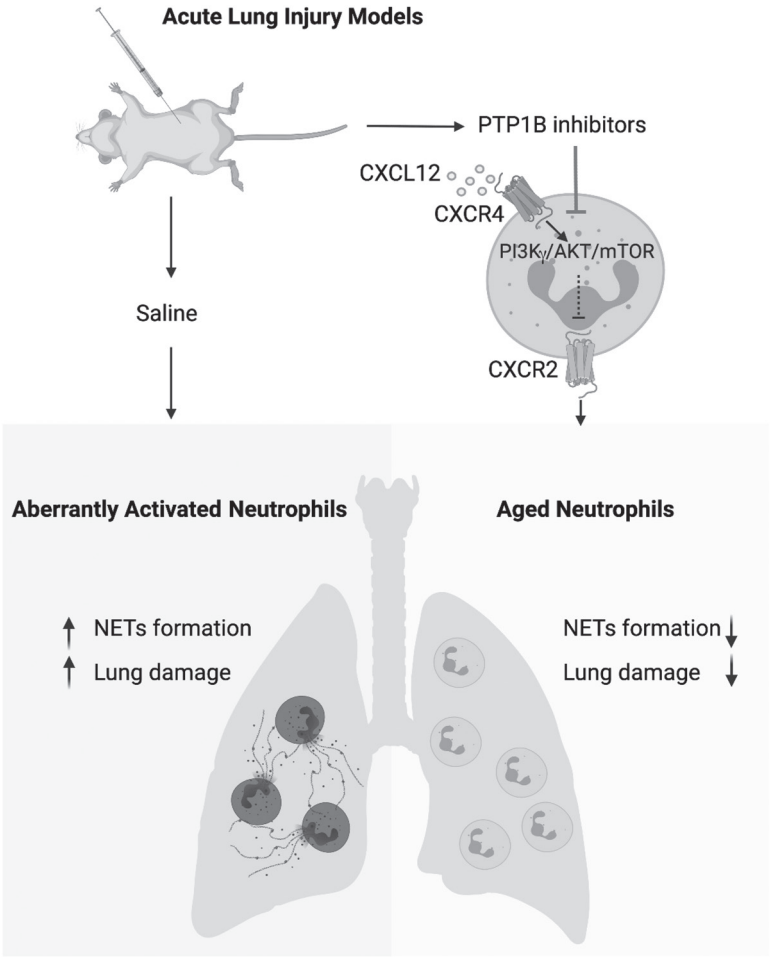


Figure 1. Pretreatment of PTP1B inhibitors improved survival and prevented lung damage in acute lung injury models. Three neutrophil-mediated acute lung injury (ALI) models were induced in this study: transfusion-related ALI (TRALI), LPS-induced, and cecal ligation and puncture (CLP)-induced sepsis models. Compared to saline treatment, administration of PTP1B inhibitors two hours prior to ALI induction enhanced survival, attenuated neutrophil extracellular trap (NET) formation, and ameliorated lung damage in ALI models. PTP1B inhibitors suppressed the PI3Kγ/AKT/mTOR-dependent CXCR4 pathway, leading to activation of CXCR2 signaling, which promoted an aged-neutrophil phenotype. Neutrophil aging was associated with decreased capacity to cause tissue damage, suggesting that the beneficial effects of PTP1B inhibitors in murine ALI models may be exerted through suppression of the inhibitory circuit of neutrophil aging.

phosphatase from outside the cell, thereby regulating intracellular signaling in cancer cells. Using a eukaryotic cell secretion system, we purified the ectodomain of PTPRD. We used it as an antigen to generate a pool of monoclonal antibodies, which included several with subnanomolar affinity that recognize the phosphatase expressed ectopically in 293T cells, where we demonstrated PTPRD was displayed on the cell surface as two non–covalently associated subunits. In addition, we validated antibody binding to endogenously expressed PTPRD in a metastatic breast cancer cell line, CAL51, in which PTPRD served a tumor-promoting function. We demonstrated that antibody treatment rapidly inhibited phosphatase activity and induced the degradation of PTPRD protein. As a result, treatment of CAL51 cells with this antibody led to inhibition of the SRC signal and suppressed PTPRD-dependent cell invasion. These data provide further validation that manipulating the function of receptor PTPs with antibodies to the extracellular segment may offer advantages over the active site–directed inhibitors and illustrates their potential for therapeutic application.

Developing Inhibitors of the Protein Kinases PIM and DYRK

This project represents a long-standing collaboration with the laboratories of Darryl Pappin and Leemor Joshua-Tor, and more recently with Dr. Yousef Al-Abed and his team at the Feinstein Institute. The foundation for this project was our purification and characterization of a small-molecule natural product that we demonstrated to have 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; it was our expectation that inhibitors of these kinases may have broad therapeutic utility. We completed an SAR optimization program to yield low-nanomolar, drug-like lead compounds that were ATP-competitive and displayed both improved potency and selectivity relative to the parent molecule. We identified compounds that inhibited DYRK preferentially, inhibited PIM preferentially, or inhibited both. Specificity was confirmed by profiling against a panel of 140 kinases. Predictive ADME analysis confirmed drug-like properties. We focused our attention on the DYRKs because of their interesting biology. DYRK1A is considered a potential therapeutic target for several cancers including glioblastomas, which are aggressive brain tumors that are unresponsive to chemo- or radiotherapy. We determined the crystal structure of DYRK1A with one of the inhibitors, which gave us deeper insight into the possible basis for specificity of the inhibitors for DYRKs over other kinases. 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. The sites were mapped to the hinge region of DYRK1A located near the ATP-binding site. By expressing these mutants in cancer cells, we established that the effects of our small-molecule drug candidates were due to "on-target" inhibition of DYRK, rather than "off-target" effects. Inhibition or knockout of DYRK1A reduced neurosphere proliferation and cell invasion using two glioblastoma cell lines, U87MG and LN229. Previous studies have reported that DYRK1A stabilizes EGFR by preventing its ubiquitination in EGFR-dependent glioblastoma; this is mediated by the sequestration of the E3 ubiquitin ligase (c-CBL) by Sprouty2 once it has been phosphorylated by DYRK1A at Thr75. Consistent with these findings, we show that pharmacological inhibition, or CRISPR-Cas9 knockout, of DYRK1A down-regulates EGFR levels and modulates the downstream MAPK signaling pathway. Treatment of DYRK1A knockout clones with the proteasome inhibitor MG132 also attenuated the down-regulation of EGFR. Finally, we demonstrated that DYRK inhibitor treatment, as well as *DYRK1A* ablation, abrogates tumor formation in U87MG subcutaneous xenograft models. Additional glioblastoma animal models are currently under investigation. Collectively, our data illustrate that targeting DYRK1A with these small-molecule drug candidates, alone or in combination with EGFR inhibitors, could be extremely beneficial in treating EGFR-dependent glioblastomas and other cancers, particularly non-small-cell lung cancer (NSCLC).

In Press

Ge EJ, Bush AI, Casini A, Cobine PA, Cross JR, DeNicola GM, Dou QP, Franz KJ, Gohil VM, Gupta S, et al. 2022. Connecting copper and cancer: from transition metal signalling to metalloplasia. *Nat Rev Cancer* **22:** 102–113. doi:10.1038/s41568-021-00417-2

Wiede F, Lu K-H, Du X, Zeissig MM, Xu R, Goh PK, Xirouchaki CE, Hogarth SJ, Greatorex S, Sek K, et al. 2022. PTP1B is an intracellular checkpoint that limits T cell and CAR T cell antitumor immunity. *Cancer Discov* **12:** 752–773. doi:10.1158/2159-8290.CD-21-0694

MECHANISMS AND TREATMENT OF METASTASIS

L. Trotman A. Aziz, W. Borges, H. Cox, M. Doherty, S. Escobar-Avelar, S. Kim, S. Kuang, M. Lee, G. Mathew, E. Meyer, J. Rey, M. Swamynathan

Resistance to therapy of metastatic prostate cancer (CRPC) is responsible for the deaths of some 30,000 U.S. men annually. Although there has been considerable progress in improving antihormone therapy against metastatic disease, this standard-of-care approach will invariably fail patients at some point.

Our focus is twofold: to understand the mechanisms driving primary prostate cancer to switch into lethal metastatic disease, and to develop novel therapeutic approaches. We study the human genetics behind the transition from indolent to lethal metastatic prostate cancer, in part through advanced functional modeling in mouse. By developing somatic transgene delivery into mouse prostate, we have succeeded in generating a unique, fast, and faithful mouse model for metastatic prostate cancer, termed RapidCaP. 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.

Independently, we aim to better understand how the PTEN tumor suppressor actually works. This has given us unique insights into how the process of endocytosis is intimately associated with tumor suppression by PTEN, allowing us to rethink the cell's anticancer system.

The Cancer's Needle in the Haystack: Whole Organ Imaging at Single-Cell Resolution

The Trotman laboratory has established a novel platform for revealing cancer by 3D reconstruction of entire organs at single-cell resolution in collaboration with Dr. Pavel Osten (Taranda et al. 2021). The approach uses serial two-photon tomography (STPT; Ragan et al., *Nat Methods* **9:** 255 [2012]), which works on fixed and agar-embedded mouse tissue that is placed under the objective of a two-photon microscope. A motor moves the tissue under the objective so that an entire several-centimeter-wide specimen can be imaged as a mosaic of individual tiled squares. After imaging of one section, a built-in vibrating blade microtome mechanically cuts off a tissue slice from the top, and then the steps of imaging and sectioning are repeated until the entire information of the tissue is collected. The microscope can record two fluorescent reporter channels for imaging. The Trotman laboratory has used this approach (originally designed for brain connectivity mapping) to solve the needle in a haystack problem of finding the origins of RapidCaP tumor initiation and to map the earliest steps of prostate cancer evolution (Taranda et al. 2021). This method allows us to visualize single metastatic cells in an entire liver lobe in the context of the whole organ. This revealed how pioneer metastatic cells are organized by the liver vasculature and how rare metastatic cells are found within the brain vasculature (Taranda et al. 2021). Regarding our focus on primary tumor progression to metastasis, our new technology yielded several surprising results. We found that metastatic escape can happen as early as 20 days after generation of the first tumor cell in the tissue. Furthermore, we observed that the physical properties of tumor cells and escaping cells are fundamentally different. Although the former show a strong increase in cell volume compared to normal prostate cells, reaching between 1000 and 3000 μm^3, the metastatic escapers are thinned down in volume to ~100 μm^3. They thus fit well into the stromal cell layers that surround the tumor. Now the team is studying what guides these transitions. To this end we use cell lines derived from the RapidCaP mouse at different time points. These can be divided into nonmigratory large cells and migratory thinner cells that serve as proxies for the tumor and metastatic escapers, respectively.

Endocytosis and Cancer

M. Lee, M. Doherty, G. Mathew, H. Cox, M. Swamynathan, W. Borges

The transduction of signals in the PTEN/PI3-kinase (PI3K) pathway is built around a phosphoinositide (PIP) lipid messenger, phosphatidylinositol trisphosphate, PI(3,4,5)P_3 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 PI(3,4,5)P_3 eraser domain, has prompted us to explore the relationship between PI3K signaling and endocytosis. We have now 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 work into the therapeutic potential of targeting endocytosis in cancer. We have elucidated how PTEN is recruited to endosomes and defined a set of protein players that are involved in this process. Now the team is asking whether the genes that encode endosomal recruitment factors of PTEN are involved in cancer.

Genomics of Lethal Human Prostate Cancer

S. Kuang, G. Mathew, H. Cox [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 prostate cancer (PC) are a crucial imprint of this variability. Therefore, one of our aims in the laboratory 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 after death from prostate cancer. Samples were obtained from the University of Washington, Seattle, and Northwell Health, New York. 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 CNA involving cancer genes has 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 2,914 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 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 following 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 RapidCaP 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 that had remained undiscovered based on bulk sequencing analysis.

PUBLICATIONS

Lyons SK, Plenker D, Trotman LC. 2021. Advances in preclinical evaluation of experimental antibody-drug conjugates. *Cancer Drug Resist* **4:** 745–754. doi:10.20517/cdr.2021.37

Taranda J, Mathew G, Watrud K, El-Amine N, Lee MF, Elowsky C, Bludova A, Escobar Avelar S, Nowak DG, Wee TL, et al. 2021. Combined whole-organ imaging at single-cell resolution and immunohistochemical analysis of prostate cancer and its liver and brain metastases. *Cell Rep* **37:** 110027. doi:10.1016/j.celrep.2021.110027

NOVEL STRATEGIES TO DIAGNOSE AND TREAT PANCREATIC CANCER

D. Tuveson K. Addison, L. Baker, E. Brosnan, G. Caligiuri, L. Demyan, A. Deschênes, A. Doshi, J. Filippazzo, V. Gaeth, A. Habowski, J. Kastan, S. Kim, D. King, R. Koike, W. Lan, S. Nadella, J. Nigri, Y. Park, H. Patel, J. Peluso, D. Plenker, M. Rubin, M. Shakiba, O. Standring, L. Surin, J. Thalappillil, H-C.(S.) Ting, C. Tonelli, C. Tsang, 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 using mouse models and our recently developed three-dimensional organoid cultures as ex vivo models of PDAC biology, which could inform novel strategies to detect and treat this currently incurable cancer. These tools have enabled new insights into the factors driving PDAC development that have the potential to improve patient care. Through our studies, we have identified NF1/RSK1 as a protein complex that exerts negative feedback on wild-type RAS via RSK1, giving insight into the development of combination strategies to target both mutant and wild-type KRAS in PDAC. By employing organoid and mouse models of PDAC, we have also begun to disentangle the cell types and signals responsible for patterning the PDAC microenvironment. For instance, we have uncovered that targeting Hedgehog signaling reprograms the tumor microenvironment, highlighting new strategies to specifically target tumor-promoting cancer-associated fibroblasts in PDAC.

Identification of Novel Kras Interactomes in Pancreatic Cancer

This work was done in collaboration with D.J. Pappin and L. Van Aelst (CSHL).

Oncogenic mutant KRAS drives almost all cases of PDAC. *KRAS* is mutated at the 12th amino acid residue located in the G-domain from glycine to either a valine (G12V) or, more commonly, aspartate (G12D) in most cases of PDAC. These mutations cause KRAS to constitutively trigger RAF/MEK and PI3K/AKT pathways, leading to increased cell proliferation and other pro-oncogenic behaviors. Although new therapies that target G12C, a different mutant form of KRAS more commonly found in other cancer types, have shown some clinical success, approaches to treat pancreatic cancer by blocking the G12V and G12D mutations in KRAS are not currently available. In addition, targeting the RAF/MEK and PI3K/AKT effector pathways has not been successful in PDAC because of dose-limiting toxicity. Therefore, we have focused on identifying the proteins that interact with KRAS as another source of potential therapeutic targets.

To determine whether there were additional oncogenic KRAS effectors, we set out to determine which proteins KRAS interacts with, and these interactors were specific to the wild-type or mutant forms of the protein. To accomplish this, we employed a proximity labeling strategy using the BirA biotin ligase. When fused to the amino-terminal end of KRAS, BirA covalently attaches biotin moieties to the lysines of nearby proteins. The biotinylated lysines are then recovered by streptavidin affinity pull-down and identified by mass spectrometry (MS) to reveal KRAS interaction partners. This proximity labeling strategy, referred to as BioID, allows us to discover proteins that only transiently interact with KRAS and therefore may have been missed by prior immunoprecipitation studies.

To label proteins proximal to KRAS in PDAC cells, we generated a series of *BirA-KRAS* constructs including *BirA-KRAS*^{G12D} (B-G12D) and *BirA-KRAS*^{WT} (B-WT) (Fig. 1A). To ensure we did not simply identify membrane proteins, we fused BirA to the membrane localization sequence of KRAS (B-CAAX) as a control. As an additional comparison, we generated a mutant KRAS-BirA fusion that was unable to localize to the plasma membrane due to loss of the prenylation site (B-G12D/C185S), which allowed us to identify proteins that only interacted with KRAS at the plasma membrane. Using cellular fractionation and immunofluorescence, we first validated that each BirA construct was localized in the expected

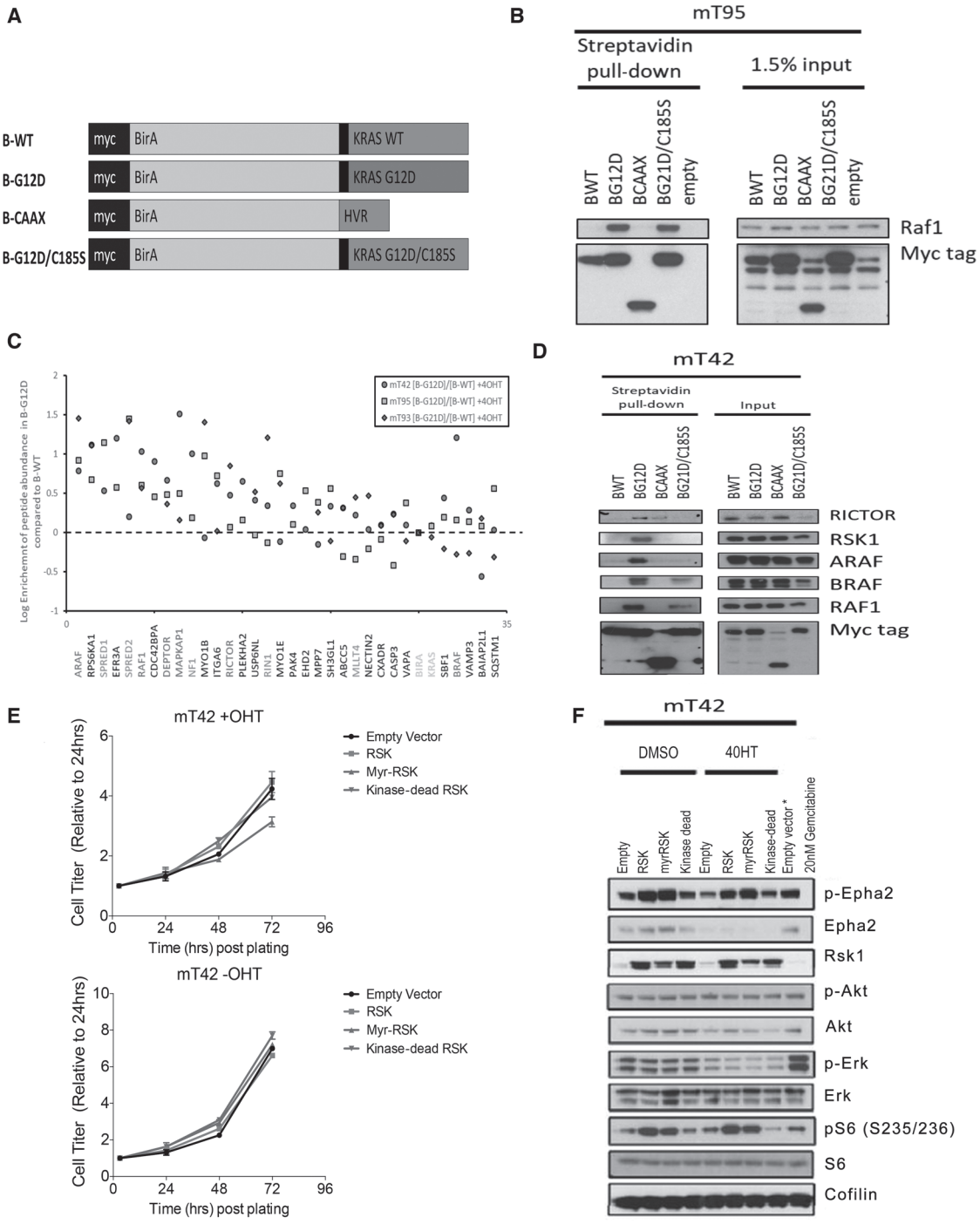


Figure 1. Identification of novel Kras interactomes in pancreatic cancer. (*A*) Design of BirA-KRAS fusion constructs. Myc-tagged BirA was fused to the amino terminus of murine wild-type (WT) KRAS4B or mutant KRAS4B-G12D with a short glycine linker (GGSG). Membrane-localized BirA was created by fusing the last 20 amino acids of murine KRAS to myc-tagged BirA. KRAS4BG12D/C185S was created by mutating the cysteine residue in the CAAX prenylation motif to mislocalize the protein. (*B*) mT95-2D FPC cells were infected to express B-WT, B-G12D, B-CAAX, B-G12D/C185S, or empty control. Cells were incubated with 50 μM biotin for 24 hours prior to lysis and streptavidin pull-down. Elutions and 1.5% input were probed for RAF1 and myc-tag BirA by western blot. (*C*) Enrichment of peptide abundances for the 32 candidate interactors as well as KRAS and BirA control peptides. (*Legend continues on following page.*)

cellular compartment (data not shown). These constructs were introduced into cell lines derived from FPC (*FRT-LSL-Kras^{G12V}-FRT; LSL-Trp53^{R172H}; Pdx1-Cre; R26-FlpOERT2*) mouse PDAC tumors, in which endogenous mutant KRAS^{G12V} can be excised upon application of 4-hydroxy-tamoxifen (4-OHT). When expressed in FPC PDAC cells, B-G12D preferentially biotinylated more RAF1 (a known KRAS interactor) than the other constructs (Fig. 1B).

Having established our approach to label proteins proximal to KRAS in PDAC cells, we next applied MS to identify KRAS-proximal proteins. We performed biotin labeling, streptavidin pull-down, and MS analysis on three biological replicate PDAC cell lines. To identify novel KRAS^{G12D} effectors, we compared the proteins enriched by B-G12D and B-WT, and also queried for membrane-specific candidates identified by comparing B-G12D to B-G12D/C185S. Across our biological replicates, we identified 32 proteins as KRAS^{G12D}-specific membrane interactors (Fig. 1C). Eleven of these proteins were known KRAS interactors or effectors, including RAF family members (ARAF, BRAF, RAF1), SPRED1, SPRED2, NF1, MLLT4, RIN1, and mTORC2 complex members (MAPKAP1, RICTOR, DEPTOR). In addition, new candidates such as PLEKHA2, MYO1E, and RPS6KA1 (RSK1) were identified in proximity to mutant KRAS^{G12D}. Among these, ARAF followed by RSK1 were the most highly ranked as consistently B-G12D enriched over B-WT. We validated RICTOR, RSK1, and ARAF as mutant KRAS interactors by western blot analysis of the BioID samples (Fig. 1D). Notably, an association between KRAS and RSK1 could not be detected using standard immunoprecipitation approaches (data not shown), demonstrating the importance of our BioID proximity labeling strategy to finding novel and transient interaction partners.

Next, we sought to better understand the interaction of KRAS with RSK1. To determine how RSK1 associates with mutant KRAS, we selectively depleted other KRAS interaction partners to determine whether this abrogated the ability of KRAS to interact with RSK1. Interestingly, RSK1 was no longer labeled by the B-G12D construct following depletion of the KRAS interactor NF1, suggesting that NF1 selectively mediates RSK1 interaction with oncogenic KRAS (Fig. 1E). As NF1 is known to require the proteins SPRED1 and SPRED2 for membrane recruitment, we next assessed whether these proteins are required for the RSK1-mutant KRAS interaction. Indeed, depletion of SPRED2 but not SPRED1 abolished RSK1 biotinylation (Fig. 1F), suggesting that a SPRED2/NF1 membrane complex bridges the interaction between RSK1 and oncogenic KRAS.

Previous studies had led to conflicting notions of RSK1 function: Some studies had suggested RSK1 promoted KRAS signaling, whereas others had found that when localized to the membrane, RSK1 inhibits RAS signaling. To address this, we generated a version of RSK1 with a myristylation (myr) tag, allowing the protein to be constitutively targeted to the membrane. Compared to empty vector, wild-type RSK1, or kinase-dead RSK1 expression, Myr-RSK1 expression in FPC cells markedly attenuated proliferation and p-ERK signaling following 4-OHT-mediated excision of oncogenic *Kras^{G12V}* (Fig. 1E, top panel, Fig. 1F), but had no effect on cellular fitness when oncogenic *Kras^{G12V}* was present (Fig. 1E, *bottom panel*, Fig. 1F). Based on these results, we hypothesized that RSK-mediated suppression of KRAS/ERK signaling was mediated through suppression of wild-type RAS signaling. In support of this, blocking oncogenic KRAS signaling in human MIA PaCa-2 PDAC cells that express KRAS^{G12C} with the G12C-specific

Figure 1. (*Continued from preceding page.*) Known KRAS interactors include RAF isoforms (purple), NF1/SPRED proteins (blue), mTORC2 (magenta), and RBD-containing proteins (orange) with the internal controls BirA and KRAS (yellow). Enrichment is plotted on a scale of log base 10. (*D*) BioID was performed on mT42 expressing B-WT, B-G12D, B-CAAX, or B-G12D/C185S followed by western blot analysis to confirm RICTOR, RSK1, ARAF, BRAF, and RAF1 as B-G12D substrates. (*E*) FPC cell line mT42 cells ectopically expressing RSK1, myristylation-tagged RSK (myr-RSK1), or empty vector control were plated for cell proliferation and treated with DMSO (as a control) or 4-OHT (to trigger Kras^{G12V} excision), and then measured by CellTiter-Glo luminescence assay every day for 4 days. (*F*) mT42 FPC cells ectopically expressing RSK1, myr-RSK1, or empty vector control were treated with DMSO or 4-OHT and analyzed by western blot. Gemcitabine was used as a proliferation control. Immunoblotting was performed for RSK1 with RAS pathways represented by ERK and AKT phosphorylation (p-Erk and p-Akt, respectively), and the RSK1 substrates, phosphorylated EPHA2 (p–Epha2), and phosphorylated S6 (p-S6).

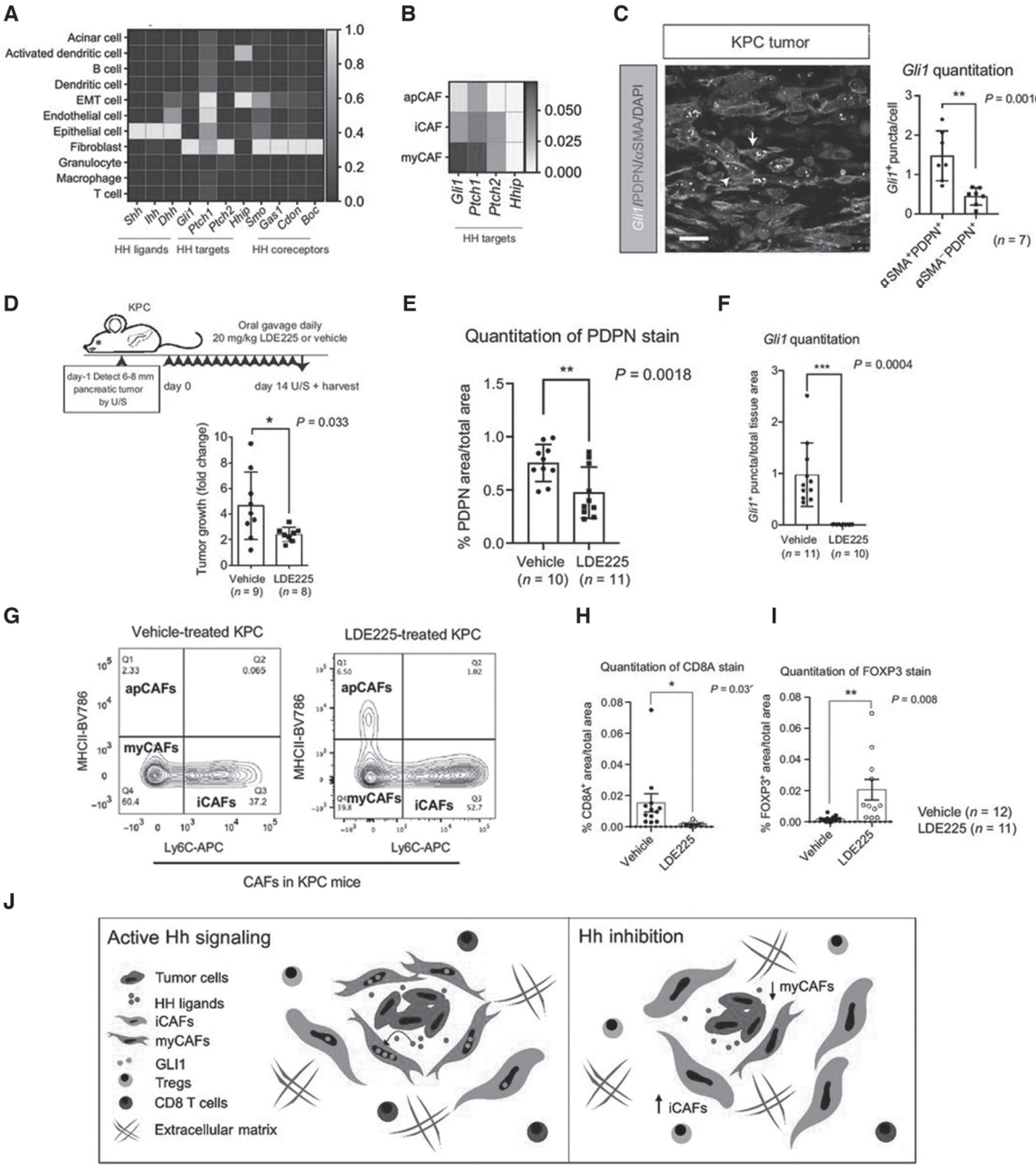


Figure 2. Inhibition of Hedgehog signaling reshapes cancer-associated fibroblast heterogeneity in pancreatic cancer. (A) Heatmap of scaled expression of Hh ligands (*Shh*, *Ihh*, *Dhh*), Hh targets (*Gli1*, *Ptch1*, *Ptch2, Hhip*), and HH receptor (*Smo*) and co-receptors (*Gas1*, *Cdon*, *Boc*) in different cell populations of pancreatic tumors of the KPC (*Kras*$^{LSL\text{-}G12D/+}$; *Trp53*$^{LSL\text{-}R172H/+}$; *Pdx1-Cre*) mouse model of PDAC ($n = 4$). Data are scaled to values between 0 (the lowest average expression) and 1 (the highest) for each gene. Data are from Elyada et al. (*Cancer Discovery*, 2020). (*B*) Heatmaps of normalized expression of Hh targets (*Gli1*, *Ptch1*, *Ptch2*, *Hhip*) in each fibroblast cluster from Elyada et al. (*Cancer Discovery* **8:** 1102 [2020]). Colors indicate log-scale gene counts. (*C*) Representative RNA ISH of *Gli1* (white) and co-IF of PDPN (green) and alpha smooth muscle actin (αSMA, red) in a KPC tumor (*left* panel). Counterstain, DAPI (blue). Scale bar, 20 μm. The arrowhead points at a PDPN$^+$αSMA$^+$ cell with lower *Gli1* expression; solid arrow points at a PDPN$^+$αSMA$^+$ cell with higher *Gli1* expression. Quantitation of *Gli1* stain in αSMA$^+$ PDPN$^+$ (myCAFs) and αSMA$^-$ PDPN$^+$ (non-myCAFs) cells in KPC tumors (*right* panel). (*Legend continues on following page.*)

drug AMG 510 sensitized cells to myr-RSK-induced growth suppression (data not shown). Our findings support a model in which oncogenic KRAS elicits negative feedback on wild-type RAS through NF1/RSK1. Taken together, our results, which were published in *Proc Natl Acad Sci* (Cheng et al. 2021), have identified the novel KRAS interactors RSK1/NF1 as inhibitors of wild-type KRAS signaling in pancreatic cancer, providing a mechanism through which cancer cells can escape therapies designed to block oncogenic KRAS. These findings provide a new avenue to the development of combination strategies targeting both oncogenic and wild-type RAS.

Inhibition of Hedgehog Signaling Reshapes Cancer-Associated Fibroblast Heterogeneity in Pancreatic Cancer

This work was done in collaboration with H.C. Crawford, B.L. Allen, C.A. Lyssiotis, and M.P. Magliano (Univ. of Michigan, Ann Arbor).

PDAC is characterized by the presence of a dense desmoplastic stroma consisting of extracellular matrix (ECM) and stromal cells. Cancer-associated fibroblasts (CAFs), one of the primary stromal cell types, have been reported to produce ECM proteins that impede drug delivery and serve tumor-supportive roles. Therefore, therapeutic strategies targeting these stromal cells emerged as a promising tactic to improve patient outcomes. However, these approaches have proven controversial. One strategy depleted stromal cells by inhibiting the Hedgehog (Hh) signaling pathway, which plays a key role in stromal activation, ECM deposition, and tumorigenesis. Although stromal targeting with a Hh pathway inhibitor (Smo inhibitor, IPI-926) increased drug (Gemcitabine) delivery and survival of KPC (*Kras*$^{LSL\text{-}G12D/+}$; *Trp53*$^{LSL\text{-}R172H/+}$; *Pdx1-Cre*) mice, clinical trials based on this approach were prematurely halted because of the poor patient outcomes. Furthermore, follow-up studies reported that genetic and long-term pharmacological inhibition of Hh signaling accelerated pancreatic cancer tumorigenesis and shortened survival in mice, suggesting that some CAFs may have tumor-restraining roles in this context. In support of this concept, our laboratory recently identified heterogeneity within the population of CAFs, each with their own distinct functions and active pathways. These fibroblasts include myofibroblastic (myCAFs), inflammatory (iCAFs) and antigen-presenting (apCAFs) CAFs. Therefore, understanding the underlying mechanisms of their active pathways is necessary for the development of appropriate therapeutic strategies to ablate tumor-promoting fibroblasts. Furthermore, we demonstrated that targeting JAK/STAT signaling shifted iCAFs to a myofibroblastic phenotype and resulted in tumor regression. We reasoned that heterogeneity and a high degree of CAF plasticity may account for the conflicting results found with previous stromal targeting attempts. To address this, we have revisited our studies targeting the Hh pathway in PDAC to investigate how Hh pathway inhibition reprograms the PDAC microenvironment.

We first evaluated the expression of paracrine Hh pathway components in different cellular compartments in PDAC by analyzing our previously published single-cell RNA-sequencing (scRNA-seq) data from human PDAC and mouse tumor samples from the KPC mouse model of PDAC. Consistent with previous studies, Hh ligands (*Shh* and *Ihh*) were highly

Figure 2. (*Continued from preceding page.*) Results show mean ± SEM of 7 biological replicates. $^{**}P < 0.01$, unpaired Student *t* test. (*D*) Schematic of 2-week treatment of tumor-bearing KPC mice with 20 mg/kg LDE225 or vehicle by daily oral gavage. (U/S) Ultrasound (*top* panel). Tumor volume as measured by U/S of vehicle- ($n = 9$) and LDE225- ($n = 8$) treated KPC tumors from the *upper* panel (*bottom* panel). Results show mean ± SEM. $^{*}P < 0.05$, unpaired Student *t* test. (*E*) Quantitation of PDPN stain in vehicle- ($n = 10$) and LDE225- ($n = 11$) treated KPC tumors. Results show mean ± SEM. $^{**}P < 0.01$, unpaired Student *t* test. (*F*) Quantitation of *Gli1* stain by RNA ISH in vehicle- ($n = 11$) and LDE225- ($n = 10$) treated KPC tumors. Results show mean ± SEM. $^{***}P < 0.001$, unpaired Student *t* test. (*G*) Representative flow plots showing the gating strategy for the analysis of DAPI$^-$ CD45$^-$ CD31$^-$ EpCAM$^-$ PDPN$^+$ CAFs in 2-week vehicle- ($n = 7$) and LDE225- ($n = 6$) treated KPC tumors. (*H*) Quantitation of CD8A stain by IHC in 2-week vehicle- ($n = 12$) and LDE225- ($n = 11$) treated KPC tumors. Results show mean ± SEM. $^{*}P < 0.05$, unpaired Student's *t*-test. (*I*) Quantitation of FOXP3 stain by IHC in 2-week vehicle- ($n = 12$) and LDE225- ($n = 11$) treated KPC tumors. Results show mean ± SEM. $^{**}P < 0.01$, unpaired Student's *t*-test. (*J*) Model explaining the role of HH signaling and the effects of HH inhibition in the PDAC microenvironment. Cancer-secreted HH ligands, such as SHH and IHH, activate HH signaling in surrounding fibroblasts (arrow), especially in myCAFs (*left* panel). HH inhibition leads to a reduction in myCAFs and an increase in iCAFs and to decreased CD8$^+$ T cells and more abundant regulatory T cells (*right* panel).

expressed in epithelial cells, whereas the expression of Hh target genes and co-receptors was primarily restricted to fibroblasts in human (data not shown) and mouse PDAC (Fig. 2A). Given that CAFs are heterogeneous and composed of different CAF subtypes, we sought to investigate whether Hh signaling was differentially active in distinct CAF subtypes. Our scRNA-seq analysis revealed that myCAFs preferentially express target genes of Hh signaling in human (data not shown) and mouse PDAC (Fig. 2B), whereas both iCAF and myCAF subtypes express co-receptors required for paracrine Hh activation (data not shown). Accordingly, RNA in situ hybridization (ISH) of *Gli1* (a transcription factor downstream of Hh signaling) with co-IF of podoplanin (PDPN; a pan-fibroblast marker protein) and the myCAF marker alpha-smooth muscle actin (αSMA), showed higher *Gli1* expression in myCAFs (αSMA+ and PDPN+ cells) compared with non-myCAFs (aSMA- and PDPN- cells) in KPC tumors (Fig. 2C). Together, these data suggest that Hh signaling is upregulated in myCAFs, although both myCAF and iCAF can respond to Hh ligands.

Next, we sought to investigate whether Hh inhibition differentially affects CAF subtypes. To that end, we treated KPC mice with a Smo inhibitor (LDE225) that is undergoing clinical evaluation. The KPC mice were enrolled when tumors reached 6-8 mm in diameter for a 2-week short-term study. Intriguingly, short-term treatment with LDE225 resulted in impaired tumor growth (Fig. 2D) and decreased overall CAFs as assessed by PDPN staining (Fig. 2E), suggesting that Hh signaling generally promotes PDAC progression. In this setting, we confirmed that LDE treatment completely abolished the expression of Gli (Fig. 2F), unlike the previous partial inhibition of Gli expression by IPI-296. As both myCAFs and iCAFs demonstrated the capacity to respond to Hh ligands, we sought to assess the effect of LDE225 treatment on the distinct CAF subtypes. First, RNA-seq analysis revealed up-regulation of the myCAF gene signature and down-regulation of the iCAF gene signature in fluorescence-activated cell sorting (FACS)-sorted PDPN+ cells (pan-CAFs) from KPC mice treated with LDE225 compared with vehicle-treated controls (data not shown). Consistently, we found that LDE225 treatment induced myCAF depletion and increased iCAF populations by flow cytometry (Fig. 2G), suggesting that the Hh inhibition preferentially reshaped the tumor environment toward a more fibro-inflammatory stroma. Given that iCAFs produce inflammatory cytokines and chemokines that are known to result in immunosuppression, we hypothesize that the increased iCAF proportion following Hh pathway inhibition modulates immune infiltration to increase immunosuppression. To test this hypothesis, we analyzed the immune composition by immunohistochemistry. These results showed fewer cytotoxic $CD8^+$ T cells (Fig. 2H) and increased levels of $FOXP3^+$ regulatory T cells (Fig. 2I) in LDE-treated KPC tumors compared with vehicle-treated controls, suggesting that Hh pathway inhibition increased immunosuppression. These results suggest that the detrimental effect of long-term Hh inhibition might be attributed to a more immune-suppressive microenvironment potentiated by iCAF enrichment (Fig. 2J), leading to tumor progression. Taken together, our results, which were published in *Clinical Cancer Research* (Steele et al. 2021), suggest that Hh signaling is a key pathway in the maintenance of the myCAF subtype in pancreatic cancer, and Hh pathway inhibition decreases the ratio of myCAF/iCAF populations in PDAC, shifting the stroma toward a more immunosuppressive microenvironment.

PUBLICATIONS

Cheng DK, Oni TE, Thalappillil JS, Park Y, Ting HC, Alagesan B, Prasad NV, Addison K, Rivera KD, Pappin DJ, et al. 2021. Oncogenic KRAS engages an RSK1/NF1 pathway to inhibit wild-type RAS signaling in pancreatic cancer. *Proc Natl Acad Sci* **118:** e2016904118.

Steele NG, Biffi G, Kemp SB, Zhang Y, Drouillard D, Syu L, Hao Y, Oni TE, Brosnan E, Elyada E, et al. 2021. Inhibition of Hedgehog signaling alters fibroblast composition in pancreatic cancer. *Clin Cancer Res* **27:** 2023–2037.

Ras AND Rho REGULATORS IN DEVELOPMENT AND DISEASE

L. Van Aelst A. Berisha, F. Boato, N.B. Gallo, S. George, C. Gonzalez, E. Michalopoulou, M. Wang

Research in the Van Aelst 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, the laboratory has continued to define the functions of selected GTPases, their regulators, and effectors, using animal models of cancer and neurodevelopmental/neurological disorders. Below are highlighted our key projects.

Molecular and Cellular Mechanisms Governing Chandelier Cell Morphology and Connectivity

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 types, chandelier cells (ChCs), in particular, have a powerful influence over the output of excitatory PyNs because of their unique morphology and the 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, still 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 (Ango et al. 2021). Notably, though, 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 molecular determinants of ChC/PyN AIS innervation, we performed an in vivo RNAi screen of PyN-expressed axonal cell adhesion molecules (CAMs) and select Ephs/ephrins. Strikingly, we found the L1 family member L1CAM to be the only molecule required for ChC/PyN AIS innervation. We further showed that L1CAM is required during both the establishment and maintenance of innervation and that selective innervation of PyN AISs by ChCs requires AIS anchoring of L1CAM by the cytoskeletal ankyrin-G/βIV-spectrin complex. Thus, our findings identify PyN-expressed L1CAM as the first known CAM required for innervation of neocortical PyN AISs by ChCs in the neocortex (Ango et al. 2021; Fig. 1).

In addition to the above intrinsic molecular players, we more recently also uncovered a novel, previously unrecognized role for microglia in the regulation of PyN AIS synapse formation/maintenance by ChCs in the neocortex (Gallo et al. 2022). Specifically, we identified a synaptogenic/growth-promoting role for microglia in regulating PyN AIS synapse formation by ChCs. We showed that a subset of microglia contacts PyN AISs and ChC cartridges and that such tripartite interactions, which rely on the unique AIS

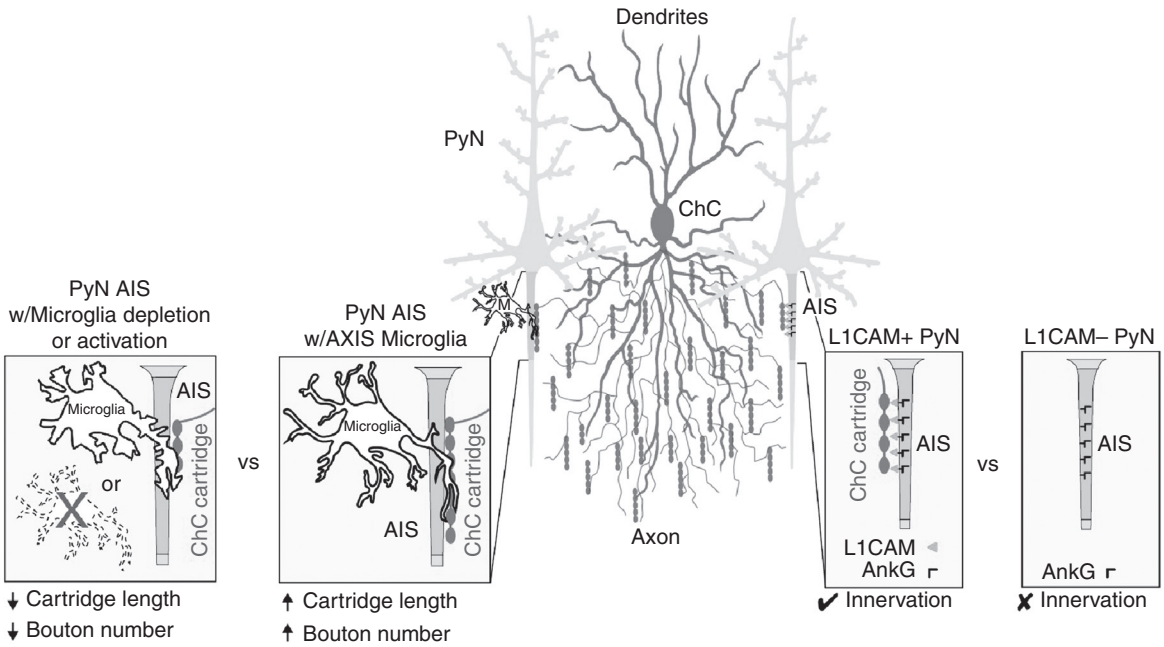


Figure 1. Chandelier cell (ChC) cartridge/bouton morphogenesis is regulated by microglia, whereas ChC/pyramidal neuron (PyN) axon initial segment (AIS) innervation is dependent on AnkG-clustered L1CAM. ChC connectivity to the AIS of neighboring PyNs in the neocortex is dependent on both cell-intrinsic and -extrinsic mechanisms. Proper subcellular innervation of PyN AISs by ChCs, a process starting at approximately P12 and maintained through adulthood, is dependent on cell–cell adhesion mediated by postsynaptic, AnkG-clustered L1CAM at the AIS (*right*). Loss of PyN expression of L1CAM or L1CAM binding to the unique AnkG- and βIV-spectrin-enriched AIS cytoskeleton impairs proper ChC/PyN AIS innervation. Apart from this, a fraction of microglia, the resident immune cells of the brain, target a subset of PyN AISs in the cortex, ultimately establishing a tripartite cellular complex necessary for ChC cartridge/bouton morphogenesis (*left*). Depletion or immunological activation of microglia during early postnatal development hinders their ChC synaptogenic function, resulting in PyN AISs being innervated by shorter ChC cartridges possessing fewer synaptic boutons.

cytoskeleton and microglial $GABA_{B1}$ receptors, are associated with increased ChC cartridge length and bouton number and AIS synaptogenesis. Conversely, microglia depletion or disease-induced aberrant microglia activation (e.g., under Alzheimer's disease–associated neuroinflammation conditions) impairs the proper development and maintenance of ChC cartridges and boutons as well as AIS synaptogenesis. These findings unveil key roles for homeostatic, AIS-associated microglia in regulating proper ChC axonal morphogenesis and synaptic connectivity in the neocortex (Gallo et al. 2022; Fig. 1). Importantly, given our findings that microglia are critical for proper ChC structure/function and because perturbations in both microglia and GABAergic ChCs have been reported in patients with neurological conditions associated with aberrant excitatory/inhibitory balance, our work also sheds new light on potential microglial-dependent interneuronopathies underlying brain disorders.

Rho-GAP Oligophrenin-1 Moderates Behavioral Responses to Stress by 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, including ours, 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. In line with this, *Ophn1*-deficient mice also exhibited maladaptive behavioral responses in a repeated social defeat stress model. Strikingly, *Ophn1* deletion exclusively in parvalbumin (PV) interneurons in the PL-mPFC was sufficient to induce helplessness. 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, suppressing PL-mPFC neuronal activity with an inhibitory DREADD (designer receptors exclusively activated by designer drugs) or by increasing the activity of PL-mPFC PV interneurons with an excitatory DREADD reversed the helpless behavioral phenotype of *Ophn1*-deficient mice. Furthermore, we uncovered that OPHN1's effect on neuronal activity/stress-related behavior is critically dependent upon 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. Notably, given that fasudil has been successfully used to treat cerebral vasospasm in Japan and does not appear to produce any major adverse side effects, it could be a promising drug for the stress-related behavioral problems of *OPHN1* patients. Taken together, our results identify OPHN1 as a critical regulator of adaptive behavioral responses to stress and shed light onto the mechanistic link between *OPHN1* genetic deficits, mPFC circuit dysfunction, and abnormalities in stress-related behaviors (Wang et al. 2021; Fig. 2).

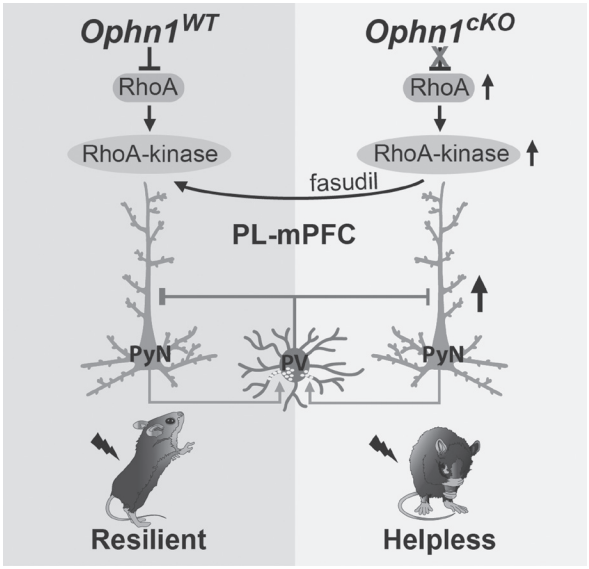


Figure 2. Deficiency of the intellectual disability gene *Ophn1* enhances stress-induced helpless/depressive-like behavior. This phenotype is mediated by a diminished excitatory drive onto *Ophn1*-deficient parvalbumin interneurons in the prelimbic medial prefrontal cortex, leading to hyperactivity in this region. Suppressing neuronal activity or RhoA/Rho-kinase signaling reverses helpless behavior.

Target Discovery for Organ-Specific and Multiorgan Metastasis

To gain insight into the mechanisms that mediate multiorgan metastases for lung adenocarcinoma (ADC), we implemented orthotopic, xenograft transplantation techniques to model lung cancer multiorgan 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 (CSHL), we initiated an in vivo RNAi screen to determine how 15 high-priority genes affect organ-specific metastasis. For the visualization of distant metastases, we took advantage of the NIS (sodium iodide symporter) system, implemented at CSHL by the Lyons laboratory, which gives the ability to resolve and dynamically follow the development of multiple individual and small metastatic lesions (~1 mm^3) in full 3D, irrespective of tissue depth or proximity to larger lesions in the body. Interestingly, our initial in vivo RNAi screen identified the EphA4 tyrosine kinase receptor and Rho-GAP Arhgap26 as potential candidate suppressor genes of brain and multiorgan metastases from lung ADC, respectively. 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 multiorgan 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. Furthermore, we began to investigate how stress promotes breast tumor growth and metastasis. We found that corticosterone, released as a response to stress, increased the levels of circulating neutrophils. These neutrophils are critical for the formation of metastasis, as corticosterone triggers these neutrophils to form neutrophil extracellular traps (NETs). Importantly, by targeting these stress-induced NETs, we were able to effectively prevent metastasis formation in the "stressed" mice.

In addition to the above studies, we collaborated with the Tuveson and Pappin laboratories (CSHL) to identify additional functionally relevant KRAS interactions in pancreatic ductal adenocarcinoma (PDAC) that may allow for a better understanding of feedback mechanisms and unveil new potential therapeutic targets (Cheng et al. 2021). Biotin ligase–mediated proximity labeling was applied to identify protein interactors of active KRAS in murine PDAC cells. Interestingly, mass spectrometry analysis revealed that RSK1 was enriched among proteins that selectively interacted with membrane-bound $KRAS^{G12D}$. We further found that RSK1 required the NF1 and SPRED proteins to interact with KRAS-GTP at the membrane. In both murine and human PDAC cell lines, membrane-targeted RSK1 was tolerated but inhibited cell proliferation following oncogenic KRAS abrogation to reveal a negative feedback role for membrane-localized RSK1 on wild-type KRAS. Inhibition of wild-type KRAS, which has been previously proposed to suppress KRAS oncogenesis, may partially explain how RSK1 has been identified as a dependency in some KRAS mutant cells and may provide an additional function for NF1 in paradoxically promoting tumorigenesis in KRAS mutant cells.

PUBLICATIONS

Ango F, Gallo NB, Van Aelst L. 2021. Molecular mechanisms of axo-axonic innervation. *Curr Opin Neurobiol* **69:** 105–112. doi:10.1016/j.conb.2021.03.002

Cheng DK, Oni TE, Thalappillil JS, Park Y, Ting H-C, Alagesan B, Prasad N, Addison K, Rivera KD, Pappin DJ, et al. 2021. Oncogenic KRAS engages an RSK1/NF1 pathway to inhibit wild-type RAS signaling in pancreatic cancer. *Proc Natl Acad Sci* **118:** e2016904118. doi:10.1073/pnas.2016904118

Wang M, Yang YT, Tai Y, Li B, Van Aelst L. 2021. Oligophrenin-1 moderates behavioral responses to stress by regulating parvalbumin interneuron activity in the medial prefrontal cortex. *Neuron* **109:** 1636–1656. doi:10.1016/j.neuron.2021.03.016

In Press

Chang C, Ge E, Bush A, Casini A, Cobine P, Cross J, DeNicola G, Dou QP, Franz K, Gohil V, et al. 2022. Connecting copper and cancer: from transition metal signaling to metalloplasia. *Nat Rev Cancer* **22:** 102–113. doi:10.1038/s41568-021-00417-2

Gallo NB, Berisha A, Van Aelst L. 2022. Microglia regulate chandelier cell axo-axonic synaptogenesis. *Proc Natl Acad Sci* **119:** e2114476119. doi:10.1073/pnas.2114476119

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, translated into neuronal signals by specific receptors in the nose, 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 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 cortex, make use of such information during behaviors.

Social animals must interact with each other to cooperate and compete. Using sounds for such interactions is common across many taxa. Humans engaged in conversation, for example, take rapid turns to go back and forth—a feat that most of us tend to perform effortlessly, but which breaks down during neuropsychiatric disorders. Our understanding of neural circuits that underlie vocal communication, especially in mammals, remains quite rudimentary. Recently, it has been discovered that a neotropical rodent, Alston's singing mouse, engages in fast vocal interactions, even under laboratory settings. The **Arkarup Banerjee** laboratory, using this novel model system, seeks to pursue two complementary questions. First, how does the auditory system interact with the motor system to generate the sensorimotor loop required for vocal communication? Second, what are the neural circuit modifications that allow behavioral novelty to emerge during evolution? Various rodent species exhibit marked differences in vocal behaviors. Genes that determine such behavioral differences (e.g., between the singing mouse and the lab mouse) must act via neural circuits within the brain. Yet, the structural and functional changes in the brain that specify the distinct vocal repertoires across related species remain unknown. Research in the laboratory combines cutting-edge systems neuroscience and comparative evolutionary analyses of neural circuitry across rodent species to bridge this knowledge gap.

The powerful influence of sensory experience on brain development has been appreciated since the 1960s. Yet, even today, the fundamental cellular and molecular mechanisms through which sensory input shapes developing neural circuits remain largely mysterious. The **Lucas Cheadle** laboratory recently discovered that sensory experience alters gene and protein expression in microglia, the resident immune cells of the brain. These sensory-induced changes allow microglia to interact with neighboring neurons to strengthen and maintain a subset of synaptic connections

and to eliminate others. These findings raise the exciting possibility that microglia, which are predominantly associated with immune responses to injury and disease, also decode salient features of the physical world and contribute to neural responses to the environment.

The Cheadle laboratory applies a multidisciplinary approach to the visual system of the mouse to investigate the contributions of microglia to sensory experience–dependent synapse development and plasticity. They further seek to identify the molecular mechanisms through which microglia effect changes at synapses and thereby exert control over brain function. To accomplish this, the Cheadle laboratory images microglial interactions with synapses in the brains of living mice, which allows the researchers to characterize the specific features of the environment to which microglia respond. In parallel, the research team uses cutting-edge single-cell transcriptomic and genomic strategies, such as single-cell RNA sequencing, to profile the molecular changes in microglia that are elicited by distinct sensory stimuli. With these combined approaches, the Cheadle laboratory is interrogating the ways in which environmental stimuli converge upon the microglial genome to shape neural circuit development and function.

The **Tatiana Engel** laboratory develops 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. 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 hundreds of thousands of neurons and thus are only vaguely glimpsed 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, the Engel laboratory develops and applies computational methods for discovering collective neural dynamics from sparse, high-dimensional spike-train data. In these endeavors, they employ and extend tools and ideas from diverse fields, including 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.

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 the **Bo Li** laboratory. They 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 to determine how these circuits participate in adaptive or maladaptive behavioral responses in various tasks.

Partha Mitra is interested in understanding intelligent machines that 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 carrying out 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. He uses 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 of 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 that were independently predicted by a model of odor learning developed in Alexei Koulakov's lab. 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 exhibits 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.

The **Jessica Tollkuhn** 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 utilize, 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 consequent 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 whose 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 made 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.

THE NEURAL SUBSTRATES OF OLFACTORY PERCEPTION AND SENSORIMOTOR PREDICTIONS

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

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 the 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 group is understanding how actions relate to perception.*

Wiring Logic of the Early Rodent Olfactory System Revealed by High-Throughput Sequencing of Single-Neuron Projections

The structure of intra- and interbrain region connectivity provides a scaffolding for neural computation. Neuronal connectivity across different sensory modalities (vision, audition, somatosensation, etc.) is organized into topographic maps of relevant stimulus features such as spatial location, orientation, sound frequency, or body location. This structured organization reflects the computations that a neural circuit performs and thereby provides a window into how each sensory system processes its inputs.

In this respect, olfaction has traditionally been considered fundamentally different from other sensory modalities. Despite a large body of work across several decades, no apparent structure has been found in mammalian olfactory cortex (piriform) connectivity. This is in striking contrast to the structured connectivity observed in all other sensory modalities. Thus, in the absence of evidence of structure, the olfactory cortex has been modeled as a tabula rasa. Because piriform circuits were presumed to be devoid of any innate structure, models of olfactory learning have assumed random connectivity and were proposed to rely entirely on plasticity to construct meaningful representations.

In collaboration with the Koulakov and Zador laboratories, we have revisited the question of the structure of olfactory sensory representations. We asked whether newer and more sensitive methods might uncover structure in the olfactory system where none had previously been observed. Enabled by unprecedented massive throughput of mapping connectivity using multiplexed analysis of projections by sequencing (MAPseq), we mapped the brain-wide projections of thousands of individual neurons originating from both the olfactory bulb (5,309) and piriform cortex (30,433) respectively (Fig. 1). The high yield of these methods and the resulting sample sizes—two to three orders of magnitude more neurons probed than in previous studies—reveal the wiring of the olfactory system at unparalleled scale and resolution.

Our findings challenge the prevailing belief of "random connectivity" in olfactory processing and, instead, reveal modularity and specificity of connections. We demonstrate a systematic relationship between the inputs from the olfactory bulb to the piriform cortex and the piriform cortex outputs. We identify nonrandom features of olfactory connectivity in the form of structured, matching input–output projection gradients along the anterior–posterior (A-P) axis of the piriform cortex with respect to connections to extra-piriform regions (Fig. 2), as well as local connectivity within the piriform cortex. Single neurons in the olfactory bulb targeting a specific region of piriform cortex also project with high probability to other

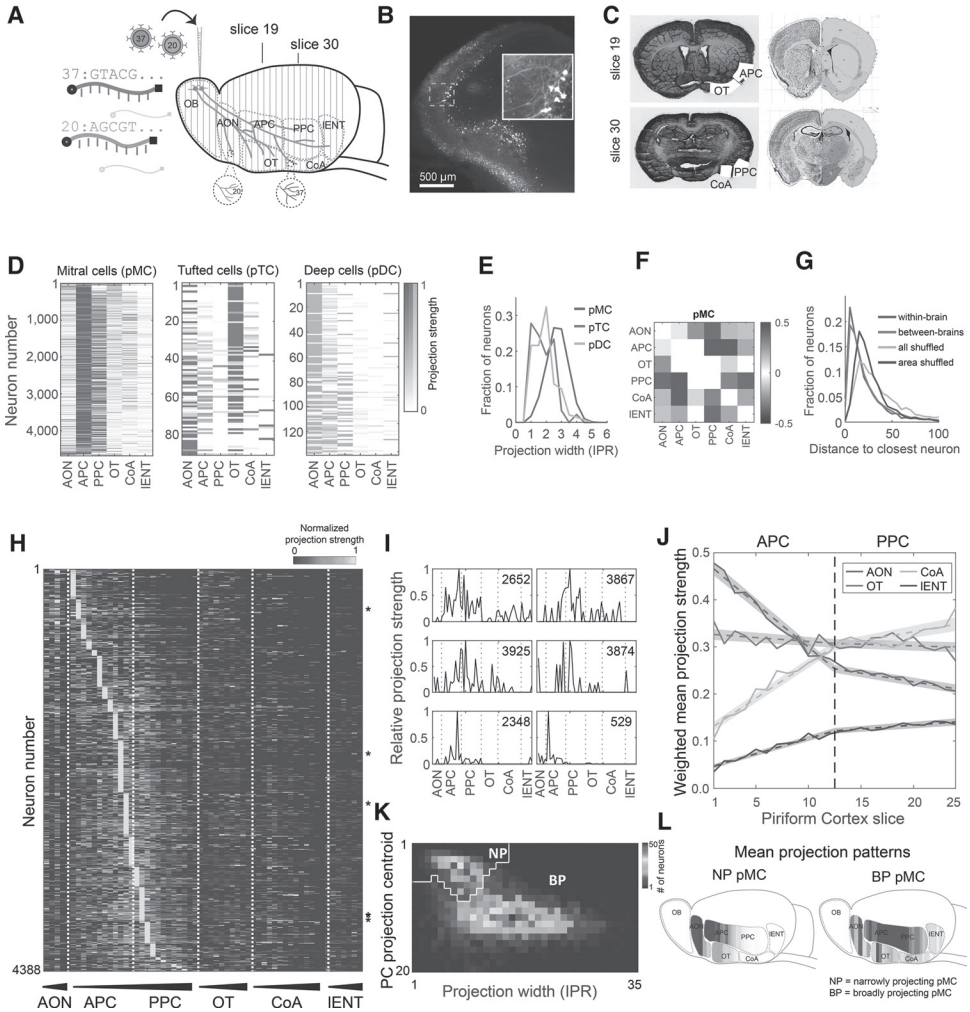


Figure 1. MAPseq and BARseq projection mapping of individual olfactory bulb neurons. (*A*) Schematics of the MAPseq strategy, which uses RNA barcodes to label neurons and map their brain-wide projections. (*B*) Infection of mitral and tufted cells by Sindbis virus carrying the barcodes and a fluorophore (GFP). (*C*) Laser capture micro-dissection of target brain regions from Nissl-stained coronal sections and corresponding sections registered to the Allen Brain reference atlas. (*D*) The projection patterns of all MAPseq-analyzed neurons; columns represent projection brain regions and rows indicate individual barcoded neurons. Barcoded neurons are sorted by probability of cell type classification based on running their projection patterns through the classifier. (*E*) Distribution of the broadness of projections, as measured by inverse participation ratio (IPR; *x*-axis) at brain region level. (*F*) Pearson correlation between putative mitral cell (pMC) projections to different target regions. Only correlations that passed statistical significance after Bonferroni correction are shown. (*G*) Distribution of the city block distance between the projection patterns of each pMC identified using the BARseq-based classifier and the most similarly projecting pMC within the same brain (blue), across different brains (red), across all brains (six) sampled after shuffling all elements in the projection matrix (yellow), or after shuffling the neuron identities for each area separately (purple). (*H*) Projection patterns of pMC neurons at 200 µm resolution along the A-P of piriform cortex. Rows represent neurons and each column represents a single 200-µm section in the brain region indicated at the bottom. Within each region, slices are arranged by A-P position, with the most anterior sections on the left. (*I*) Projection patterns of example pMC neurons. Projections are normalized by maximum barcode count for each neuron. Neuron numbers are indicated on top of each plot and correspond to the fiduciary marks shown in *H*. (*J*) Weighted mean projection strength for OB neurons to four major extra-piriform targets as a function of location of PC co-innervation (the conditional probability of co-innervation P(target|PC location), solid lines). Dashed lines/shaded areas show piecewise linear fits in anterior piriform cortex (APC) and posterior piriform cortex (PPC) with the 95% confidence interval obtained by bootstrap. (*K*) The distribution of neurons as a function of projection centroid and projection width (inverse participation ratio, IPR). Two types of neurons form distinct clusters, the narrowly and broadly projecting pMCs (NP and BP, respectively). Watershed clustering identifies a separatrix (white line) for the distributions of these two cell types. (*L*) Heatmaps of mean projection patterns of NP and BP neurons.

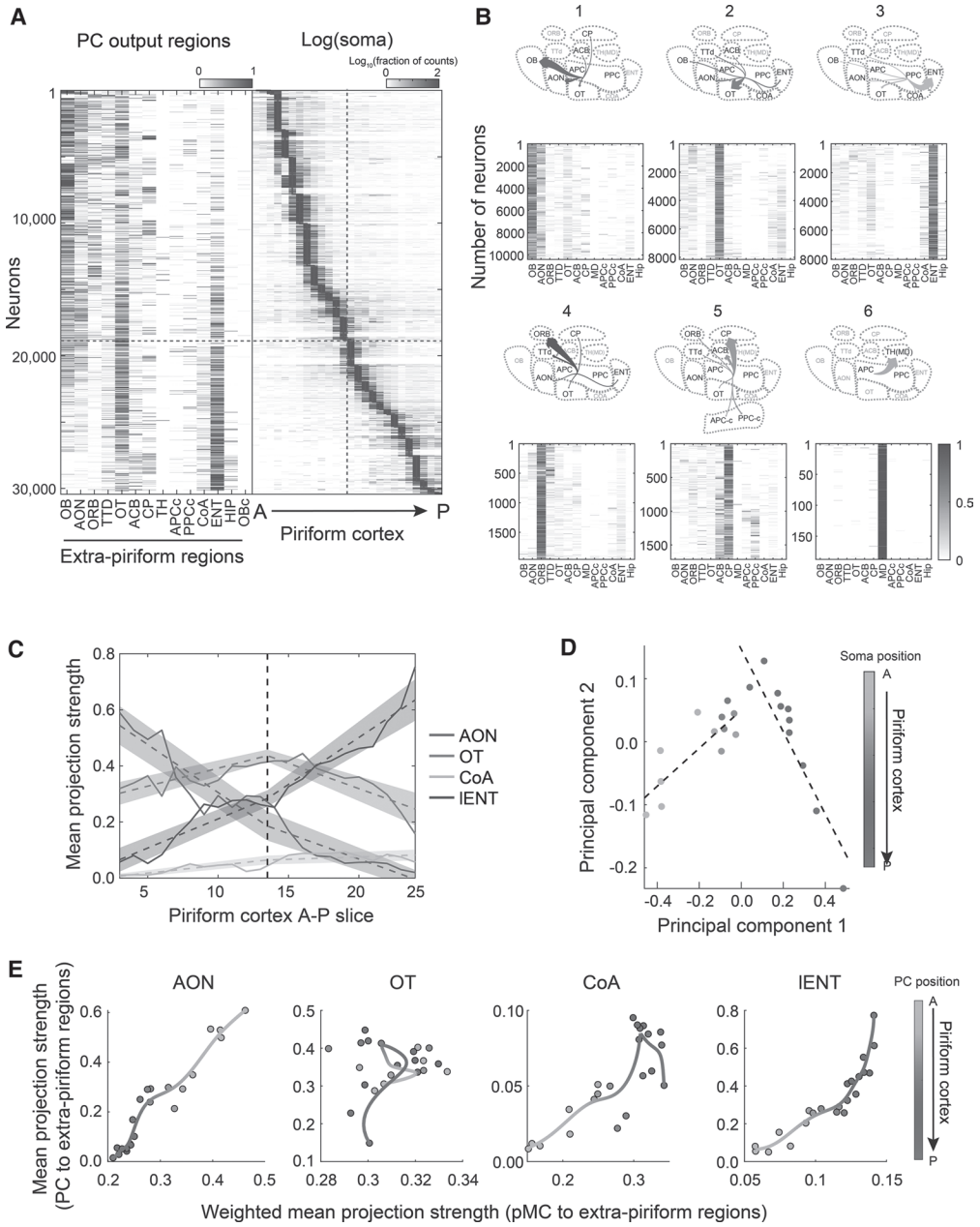


Figure 2. MAPseq projection mapping of individual piriform cortex neurons and matching input–output projection gradients along the A-P axis of the piriform cortex. (*A*) (*Left*) Projection patterns of piriform cortex (PC) output neurons to extra-piriform brain regions and (*right*) within the PC along the A-P axis. Projection density is color-coded on a log scale. In the PC, for a given barcoded neuron, the A-P position with the most barcode counts is taken as the location of the soma. (*B*) Mean projection patterns (*top*) and the projection patterns of individual neurons (*bottom*) of groups of PC output neurons to extra-PC target brain regions. (*C*) Mean projection patterns of PC output neurons at the indicated A-P position of barcoded somata in the PC. Dotted lines indicate linear fits and shaded areas the range of fits from bootstrap. (*D*) Mean loadings for the first two principal components of the mean projection strengths of PC output neurons to anterior olfactory nucleus (AON), cortical amygdaloid nucleus (CoA), lateral entorhinal cortex (lENT), and olfactory tubercle (OT) sampled at the indicated A-P positions in the piriform cortex. Dotted lines indicate linear fits for anterior piriform cortex (APC) and posterior piriform cortex (PPC). (*E*) Mean projection strengths of piriform projection neurons to extra-piriform target regions, organized by the location of their somata along the A-P axis of PC (*y*-axis) plotted against the mean projection strengths of putative mitral cell (pMC) neurons to extra-piriform OB target regions weighted by projections to a particular A-P position in the piriform cortex, P(target|PC location) (*x*-axis). Colors indicate A-P positions in the piriform cortex.

brain regions that are targeted by individual piriform neurons in that locality, forming triadic circuit motifs. Our results suggest that, as reported for other brain circuits (such as the "what/where" pathway in vision), olfactory information is routed to functionally diverse targets in a coordinated manner along parallel streams.

Impact and Significance: Our findings suggest that olfactory information is organized into several processing streams. Distinct processing streams have been observed in other sensory modalities, where they connect different sets of brain regions and carry different sensory information. In the visual system, where such streams were first identified, cortical information is organized into a ventral ("what") pathway involved in perception and a dorsal ("where") pathway involved in action. A similar two-stream what/where segregation has been proposed for auditory and somatosensory processing. The adaptive advantage of segregating information into several streams is that features such as the identity, size, shape, orientation, and relative position of visual objects can be computed in parallel. Neurons encoding these features are then organized in multiple neuroanatomical pathways spanning the parietal and temporal lobes. Our data supports the view that olfactory information leaving the olfactory bulb might similarly be segregated into perception (olfactory bulb [OB]-anterior piriform cortex [APC]-anterior olfactory nucleus [AON]), valence (OB-PPC-cortical amygdala [CoA]), and action (OB-PPC-lateral entorhinal cortex [lENT]) pathways, structured according to the triadic connectivity motifs identified here. Interestingly, we find that the piriform cortex can be viewed as a continuous spatial map representing shifting preferences among these three pathways. Within the perception stream, the triadic circuit consisting of direct OB → APC and OB → AON projections, combined with indirect APC → AON projections, may enable efficient computation of perceptual features such as the identity, intensity and relative timing of odorants. In this view, the OB carries raw sensory data, and the APC → AON pathway might carry contextual gating information needed to sift through different stimulus features based on their behavioral relevance. As AON neurons also strongly innervate the APC, the AON-APC bi-directional connectivity may further consolidate the computation of these features. Similarly, within the valence stream, the triadic circuit consisting of direct OB → PPC and OB→ CoA projections, combined with indirect PPC → CoA projections, is ideally placed to compute odor valence by mixing both innate and learned neural representations. Finally, within the action stream, the triadic circuit consisting of direct OB → PPC and OB → lENT projections, combined with indirect PPC → lENT projections, may support computations such as olfactory spatial navigation, as well as fast recall of olfactory object memories. These processing streams enable a given extra-piriform target region to compare a direct input from the olfactory bulb with a version of the same that has been processed by the piriform cortex.

In summary, our results reveal that the architecture of the olfactory system is structured and thus need not rely on algorithms that assume random connectivity. We suggest an alternative model in which odor stimuli are processed along parallel, spatially segregated, functionally distinct streams. By connecting the odorant receptor molecular identity of glomeruli to the brain-wide connectivity and responses of individual neurons within the olfactory bulb, piriform cortex, and extra-piriform regions, future work will determine the relationship between functional cell types, structured olfactory connectivity and the algorithms for processing olfactory information.

Neuronal Substrates of Olfactory Perception

This work was done in collaboration with the Zador and Koulakov laboratories (CSHL).

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 and 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, DNNs).

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 olfactory bulb (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 mitral/tufted cells (MTCs) in awake mice that are naïve 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. Altogether, we aim to identify the analogs of the red–green–blue basis set of odor perception.

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

We addressed a long-standing open question pertaining to the neural substrates of odor identification and invariant perception. In mammals, olfactory information is relayed from the main olfactory bulb to the rest of the brain by two classes of principal neurons: the mitral and tufted cells. Mitral cells innervate strongly the piriform cortex, whereas the much-less-studied tufted cells innervate instead primarily the anterior olfactory cortex (anterior olfactory nucleus, AON). Over the past decades, a rich body of experimental work and computational models have proposed that odor identity (independent or not of intensity) is computed by the piriform cortex drawing from mitral cell inputs. Our results, enabled by using multiphoton microscopy to specifically monitor the odor representations of mitral versus tufted cells, provide novel key insights on both the identity as well as the specificity of neural circuits supporting odor decoding. We found (1) unexpected superiority of the tufted cell over mitral cell ensembles in decoding both odor identity and intensity, acting largely in a feedforward manner; (2) specificity in the interplay of long-range feedforward and feedback signals between tufted, mitral cells and their respective major cortical targets; and (3) different functional roles of corticobulbar feedback from the anterior olfactory nucleus and piriform cortex in regulating the response gain versus odor tuning of their dominant feedforward inputs.

Our work identified the tufted cells, rather than mitral cells, as the primary players in decoding odor identity (both dependent and independent of stimulus intensity), as well as in sensing odor concentration (Fig. 3). Inspired by behavioral challenges faced by the olfactory system, we designed several decoding schemes and found that the performance of tufted cells, rather than mitral cells, was significantly better at (1) extracting odor identity in complex odor scenes, (2) identifying odors independent of concentration variations, and (3) reporting changes in the concentration of stimuli of interest. As tufted cells relay information mainly to brain regions other than the piriform cortex, such as the anterior olfactory nucleus and olfactory striatum (tubercle), these findings redefine our understanding of how and where odor identity is computed in the mammalian olfactory system.

Top-down feedback originating in the piriform cortex versus AON controls specifically the activity of its dominant feedforward input—that is, AON feedback acts preferentially on tufted cells, whereas piriform cortex feedback selectively controls the activity of mitral cells. This feedforward–feedback closed-loop architecture along multiple pathways is powerful because it enables both implementing different computations on inputs from the sensory periphery (i.e., olfactory bulb) via parallel functional streams and cross-talk and comparisons across streams.

Cortical feedback originating in the AON, acting specifically on tufted cells, proportionally suppresses

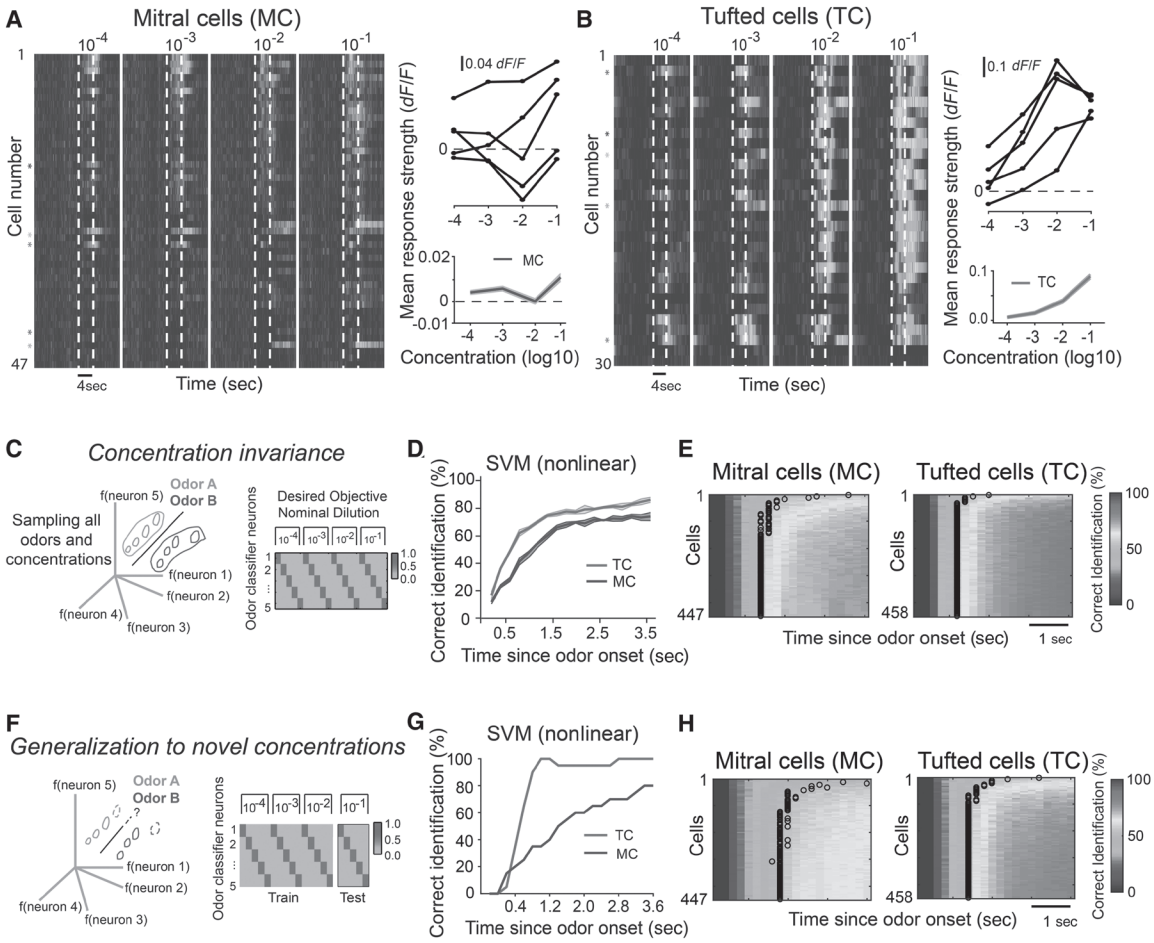


Figure 3. Tufted cells are superior to mitral cell ensembles in decoding concentration-invariant odor identity. (*A*) Mean peri-stimulus time histogram of 47 exemplar simultaneously recorded mitral cells to increasing concentrations of valeraldehyde within the same field of view (FOV) (*left*). Color 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 mitral cells (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 where 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 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).

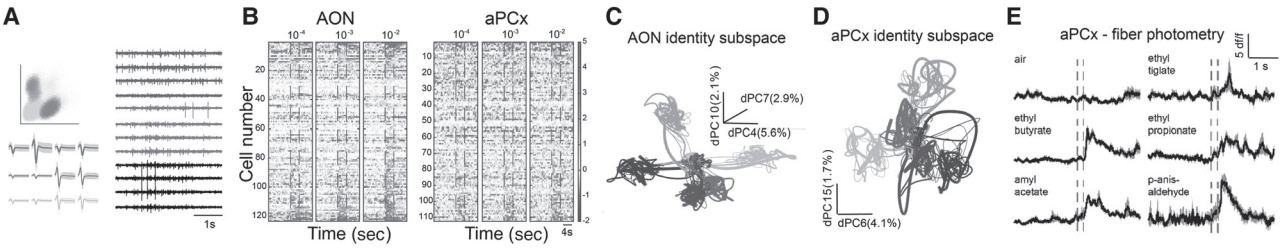


Figure 4. Anterior olfactory nucleus (AON) neurons outperform anterior piriform cortex (APC) ensembles in concentration-invariant decoding of odor identity. (*A*) Example single units and spontaneous activity on tetrodes from a microdrive implanted in AON. (*B*) Example responses of AON and anterior piriform cortex (aPCx) single units to odor concentration series. (*C,D*) Population trajectories in the neural state space defined by the top three identity principal components for AON (*C*) and APC (*D*). Colors, odor identity; thickness, concentration. (*E*) Example odor responses (GCaMP6f) using fiber photometry in aPCx (Emx-Cre x Ai148).

the feedforward input drive for each stimulus, thereby regulating the gain of their odor responses, without altering the odor tuning of tufted cells. In contrast, feedback from the piriform cortex specifically restructures mitral cell responses, modifying their odor tuning beyond simple scaling of response gain.

Tetrode micro-drives implanted in anterior piriform cortex (APC) and AON enabled monitoring individual neuron responses to five odors across the same concentration range as for the OB outputs. We used demixed principal component analysis (dPCA) to investigate whether odor identity and concentration information could be linearly separated from AON and APC responses. Overall, compared to APC (308 neurons, six mice), AON ensembles (360 units, five mice) appeared better-placed for decoding odor identity invariant of concentration (Fig. 4).

We have also determined how cortical feedback affects learning of odor associations via contextual signals. We tested the working hypotheses that (1) cortical feedback binds recent activity patterns in the piriform cortex to ongoing sensory inputs in the olfactory bulb and modulates mitral cell representations to facilitate the identification of behaviorally relevant odors, and (2) cortical feedback broadcasts contextual signals that are used by the olfactory bulb output to help identify rewarded stimuli.

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 the 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 cell 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

This work was done in collaboration with the Engel laboratory (CSHL).

Can we predict the perceptual similarity of two odorants if we know which olfactory receptors they activate? This seemingly simple question remains unsolved, as little is known about how the brain maps differences in odorant receptor (OR) activation into distinct percepts. This is due in part to difficulty in controlling olfactory stimuli at the level of individual receptor types during behavior, which makes it challenging to disentangle the contribution of specific ORs to shaping olfactory perception. To

overcome this limitation, we exploited the anatomical clustering of ORs to individual glomeruli that naturally occurs in the early olfactory system. Our experimental approach enabled identifying several dozens of glomeruli on the olfactory bulb surface and determining their responses to a large set of odorants (~130). We created synthetic olfactory stimuli by optogenetically activating specific combinations of identified glomeruli using patterned photostimulation. To obtain a metric of perceptual distance between stimulus pairs, we asked expert mice engaged in a go–no go task to report differences in identity between sets of photoactivated glomeruli. In this manner, we aim to quantify the contribution of odorant receptor identity, including its functional properties, and the corresponding glomerulus placement on the bulb surface to the perceptual similarity between synthetic olfactory stimuli. Preliminary results confirmed that the degree of overlap in glomerular activation between stimulus pairs partially explains their perceived similarity, consistent with previous results. Further, the identity and the odor response tuning of the glomeruli that composed each synthetic olfactory stimulus also contributed to the perceptual similarity between stimulus pairs. A model that combined these factors showed strong predictive power on the perceived similarity between the synthetic olfactory stimuli sampled. Our approach provides a framework to relate the perceptual similarity of olfactory stimuli to individual OR identity and odor response features.

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 methods via spatial light modulators (SLMs) to optogenetically control neurons of interest at the single-cell level and digital micromirror device (DMD)-based methods to control cell type–specific populations across large brain regions. 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 (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

Understanding the Neuronal Substrates of Internal Models of the World

We all know the startling feeling when settling into a chair that's just an inch lower than expected! This brief mismatch between reality and expectations reveals a fundamental process that occupies our brain: predicting the consequences of our actions and verifying whether current sensory observations match expectations. These computations require generative internal models that are learned and updated by minimizing prediction errors to accurately capture the causal relationship between movements and sensation. This continual tally of reality and prediction is at the core of sensory perception and adaptive motor control and may form the basis of thought. Indeed, studies in human patients suggest that nervous system dysfunctions may be interpreted as failures in sensorimotor inference (e.g., schizophrenia, overpredicting; autism, scarcity of predictions). However, little is known about how internal models of the world emerge from the basic blocks of computation at the level of cells, neural connectivity, and circuit function.

To understand the circuit-level mechanisms of internal models, we are leveraging a novel closed-loop behavioral framework to read out predictions in mice and state-of-the-art methods for monitoring and manipulating activity across the brain. Briefly, we link actions to precise sensory predictions across two different modalities (olfaction and vision), and further read out the learned predictions by observing what animals do when faced with sensory surprise. This is akin to inferring a driver's model of how the steering wheel controls the car by observing her compensatory maneuvers when the car unexpectedly skids. Using a rich repertoire of unexpected sensory feedback perturbations and computational analysis, we will reverse-engineer each animal's internal model. We hypothesize that sensorimotor perturbations trigger error

signals in those brain circuits that support sensorimotor predictions and, hence, will use them as handles to dissect the circuit-level substrates of internal models. Specifically, we aim to understand (1) the algorithms and logic of neural circuits that compute sensorimotor prediction and errors that form the basis of internal models, and (2) how sensorimotor errors are used to update internal models given changing sensorimotor relationships.

To study internal models both at the behavioral and circuit level, we developed a novel behavioral task where 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 learned 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 learned 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. 5–7).

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

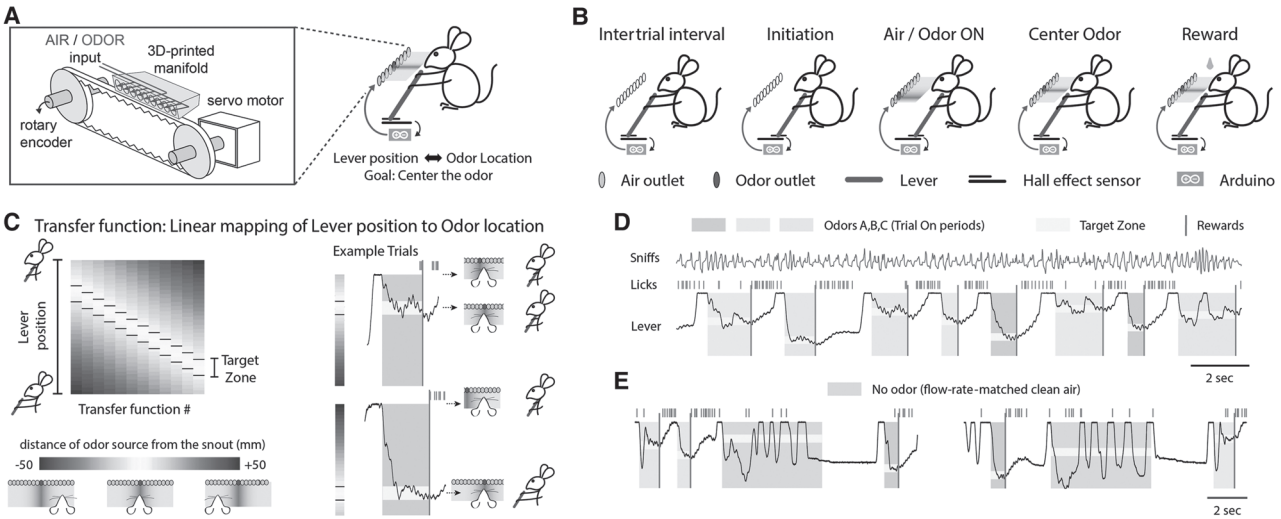


Figure 5. A novel method for fast, closed-loop control of odor source location. (*A*) Schematic of the system: 1D movement of the lever (read by a Hall effect sensor) is transformed to left–right displacement of an odor source. The odor location is manipulated by displacing a 3D-printed manifold affixed to a timing belt servo system. The manifold comprises a central odor outlet and 16 air outlets on either side. Flow rate (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, the set of lever positions that place the odor <3 mm from the snout. All transfer functions have the same gain, the unit displacement of 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.

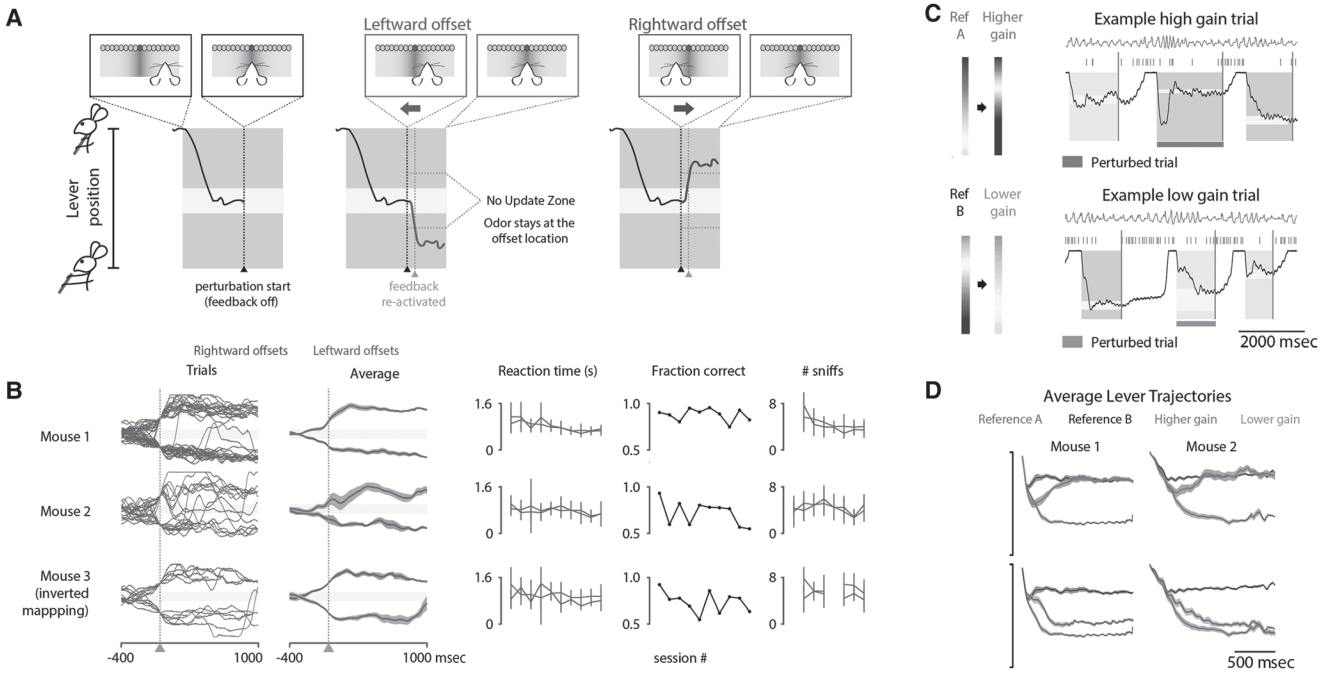


Figure 6. Feedback perturbations trigger corrective movements that reveal the animal's internal model. (*A*) Schematic of within-trial perturbations of odor location and expected corrective movements (red, leftward odor offset; blue, rightward odor offset). This example illustrates the scenario for a 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 (2× 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 and Mouse 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 (3×) 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 (1×) and then quickly corrects to successfully center the odor, despite the higher gain. (*Bottom*) Example trial with lower gain (0.4×) 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.

of the same session. Mice tended to follow paw/lever movement patterns similar to those observed in the closed-loop trials, matching their sensory experience 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 modulated 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.

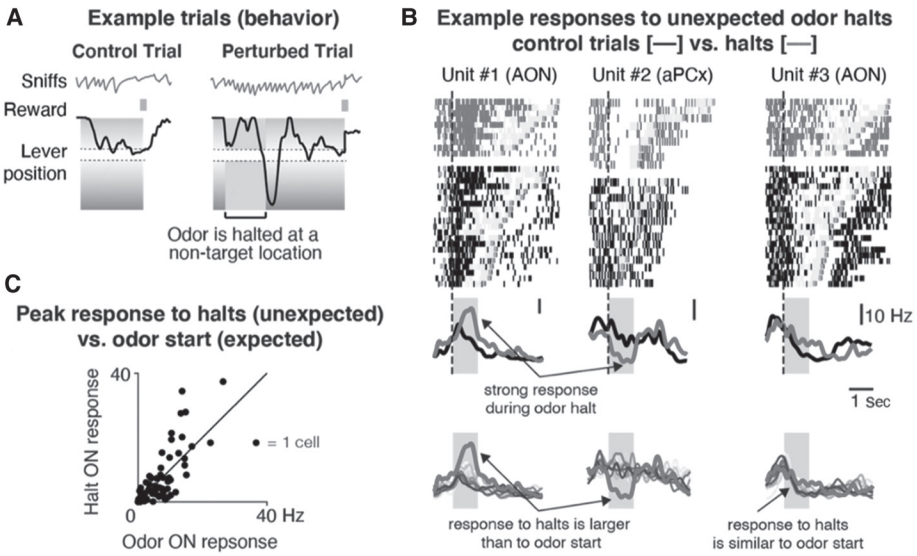


Figure 7. Transient sensory feedback perturbations trigger strong responses in olfactory cortex. (*A*) Example behavior upon within-trial perturbation in which odor is halted at an off-target location for 1 sec. (*B*) Example responses of olfactory cortex neurons (*top*, spike rasters; *bottom*, firing rate) aligned to perturbation start. Red and black traces show time-matched response in control trials (no perturbation) and in odor halt trials, respectively. Units 1 and 2 respond stronger upon odor halts than in time-matched control trials (*middle*), as well as for any average odor ON across all trials (*bottom*). Unit 3 is more sensory: Halt response is comparable to typical odor-evoked response. (*C*) Summary of the effect shown in *B* for 71 co-recorded AON units in one mouse. Halt responses of many neurons are larger than their typical odor ON responses.

Other Collaborative Projects with CSHL Groups

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

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

Janowitz: Optical monitoring of neuronal activity in induced models of cachexia during olfactory discrimination behaviors.

PUBLICATIONS

Chen Y, Chen X, Baserdem B, Zhan H, Li Y, Davis MB, Kebschull JM, Zador AM, Koulakov AA, Albeanu DF. 2021. Wiring logic of the early rodent olfactory system revealed by high-throughput sequencing of single neuron projections. bioRxiv doi:10.1101/2021.05.12.443929

In Press

Chae HG, Banerjee A, Dussauze MD, Albeanu DF. 2022. Long-range functional loops in the olfactory system and their roles in computing odor identity. *Neuron* (in press). doi:10.1016/j.neuron.2022.09.005

NEURAL CIRCUITS FOR VOCAL COMMUNICATION

A. Banerjee M. Davis E. Isko
C. Harpole M. Zheng

The primary goal of our laboratory is to investigate brain-wide neural circuits for vocal communication. We use the rich vocal behavior of the Alston's singing mice to pursue two complementary questions. First, how does the auditory system interact with the motor system to generate the fast sensorimotor loop required for vocal communication? Second, what are the neural circuit modifications that allow behavioral novelty to emerge during evolution? Using a comparative approach, we investigate brain-wide connectivity and neural circuitry differences between the Alston's singing mouse and other rodent species (e.g., the laboratory mouse) that have intermediate degrees of vocal behavior. These two lines of work combine cutting-edge neural circuit analysis of a natural behavior with comparative evolutionary analyses across species to gain insight into the function and evolution of neural circuits for vocal communication.

Motor Cortical Dynamics during Vocal Production

A. Banerjee [in collaboration with M. Long, NYU; S. Druckmann, Stanford University]

Social interactions often require behavioral flexibility to coordinate across individuals—for instance, conversations involving intricately entwining speech with minimal delays. Similarly, in the Alston's singing mouse (*Scotinomys teguina*), vocal partners can take part in fast and flexible counter-singing in which two animals can coordinate their ~10-s songs with silent gaps of ~500 ms. Recently, we discovered that a cortical locus in the singing mouse, the orofacial motor cortex (OMC), is crucial for coordinated vocal interactions; inactivation of this region abolished counter-singing behavior (Okobi et al., *Science* **263:** 983 [2019]; Banerjee et al., *Curr Biol* **29:** R190 [2019]). However, the circuit dynamics within the OMC and their impact on vocal production remained unclear.

To address this issue, we used chronic high-density silicon probe recordings in OMC ($n = 396$ neurons, 14 sessions, five animals) of male singing mice. During singing, we found that neural ensembles displayed reliable activity that was distinct from that recorded outside of the context of song. Further analysis revealed structured spiking activity on two behaviorally relevant timescales. About 35% of recorded neurons have slowly (approximately seconds) varying persistent dynamics that—as a population—track the initiation, progression, and termination of songs (Fig. 1). Leveraging the large trial-by-trial variability in motor output (songs), we found that this persistent activity in individual neurons "stretches" or "compresses" in accordance with the overall song durations. A second population (~30%) exhibited spiking that was temporally phase-tuned to individual song notes, tiling the entire respiratory cycle. Further analyses of spike times revealed that this fast-timescale modulation has a ~10-ms lag with respect to song notes, fitting the profile of sensory feedback signals. To understand the impact of motor cortical dynamics on song production, we next constructed two models in which the OMC outputs are either (1) completely determining the structure of the songs, including the articulation of constituent notes, or (2) working hierarchically with a downstream pattern generator capable of shaping note structure. Each model generates behavioral predictions that we compare against both normal songs and songs after focal cooling of OMC. From this analysis, we find support for the second model, in which the duration of vocal bouts is controlled by OMC using a temporal scaling mechanism, whereas the structure of individual notes is determined downstream.

Taken together, we describe a strategy by which motor cortical activity, acting via downstream pattern-generator circuits, generates behavioral (vocal) flexibility. These results provide a systems-level framework to study hierarchical motor control, a challenge faced by natural and artificial agents moving through the world. This work is being written up for publication.

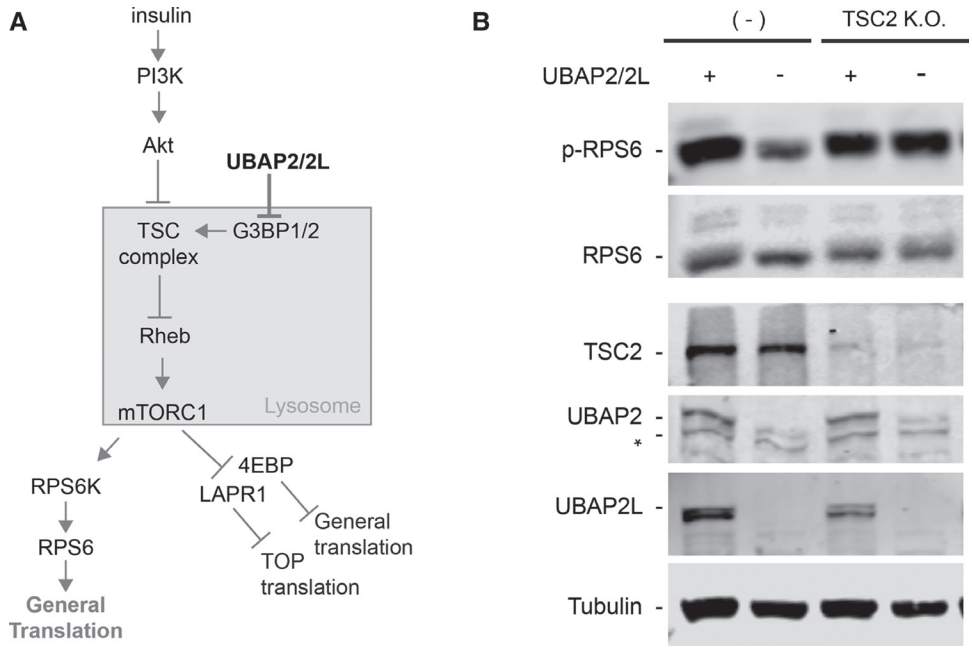


Figure 1. Chronic electrophysiology in orofacial cortex (OMC). (*A*) Extracellular recordings in freely moving singing mice, with high-density (128-channel) Si probe spanning all cortical layers. (*B*) Spiking activity of 29 layer-5 OMC neurons for an exemplar advertisement song. (*C*) Trial-averaged firing rates (mean ± SEM) of six example neurons in OMC, showing robust neural activity tracking song initiation, progression, and termination.

Neural Circuits for Context-Specific Switching of Vocalizations

M. Zheng, C. Harpole, M. Davis

Laboratory mice emit ultrasonic vocalizations (USVs) in social contexts, whereas singing mice produce both USVs and stereotyped advertisement songs. Moreover, we discovered that these mice are able to switch between these vocal modes flexibly depending upon interanimal distance (Fig. 2). The species-specific USVs are known to be controlled by periaqueductal gray (PAG) and anterior cingulate cortex (ACC). Previous work in the laboratory has shown that the advertisement songs in singing mice are modulated by OMC. This current study focuses on understanding the neural circuits underlying these two types of vocalizations. We aim to test the roles of the two cortical regions (ACC and OMC) during USVs in both species and during advertisement songs in singing mice. We hypothesize that ACC projections to PAG control USV production in both species and OMC projections to PAG control advertisement song in singing mice. Over the last year, we have built a behavioral rig suitable for high-speed videography to track the 3D poses of freely moving animals. Future analysis of the pose data will not only describe the pose dynamics during vocalizations, but also uncover the organization of behavior in singing mice. Currently, we are using optogenetics to test the role of OMC in song initiation and maintenance by transiently perturbing neural activity during different parts of the song. We will perform electrophysiological recordings subsequently to understand the neural circuits of vocal motor control across the two rodent species.

Modifiability of an Innate Behavior: How Social Hierarchy and Experience Influence Vocal Interactions in the Singing Mice

C. Harpole

In this project, we are testing the hypothesis that vocal behaviors in the singing mice, although innate, are modifiable by experience and social status. We have developed a behavioral paradigm to quantify the degree to

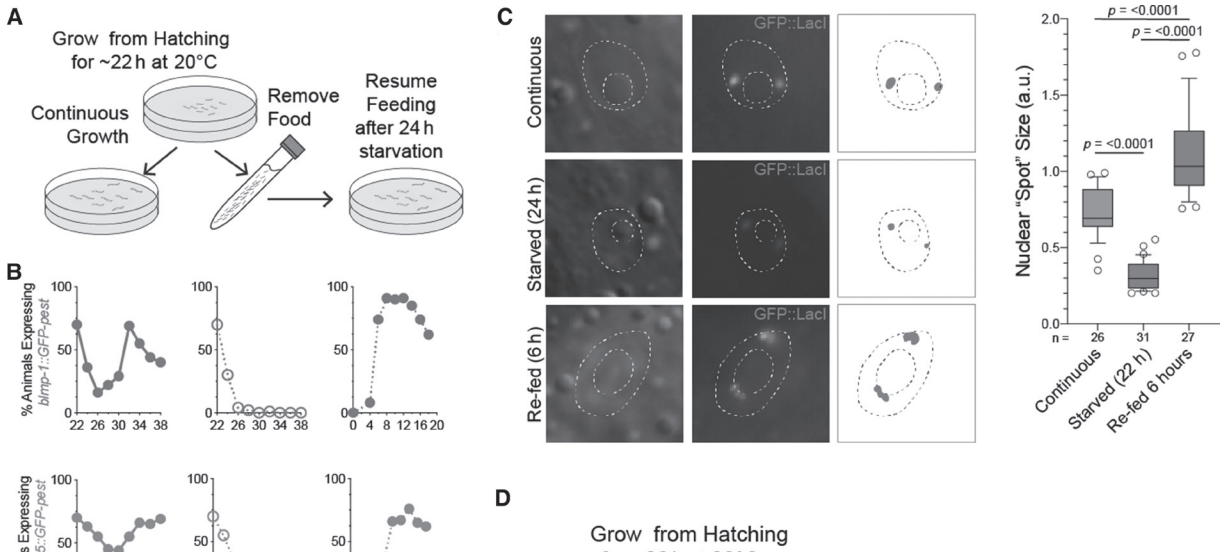


Figure 2. Fast and flexible switching between two vocal modes. (*A*) Behavioral arena with a 1/8-in. Plexiglas divider, with an array of tiny slits that allow the animals to physically interact. The arena is fit with an overhead high-speed camera and two asymmetrically placed microphones. (*B*) A single camera frame where a female (*left*) and male (*right*) *S. teguina* are interacting at the divider. (*C*) Interanimal distance distribution shows a clear bimodality. They either spend time interacting near the divider or far away from each other. 1 pixel = 0.54 mm. (*D*) One minute of vocalizations showing vocal mode switching—ultrasonic vocalization (USV) to song and back to USV. Interanimal distance is measured and plotted alongside (black trace). Two hundred and thirty-eight USVs were detected in 1 min. In 61 min of this example recording session, 10 songs were detected, interspersed with an average of 235 USVs per minute. (*E*) Exemplar spectrograms of two USVs and one song color-matched to arrows in *D*. Preliminary observations indicate that at long distances, songs are more likely (purple) and at short distances, USVs are more likely (blue and green).

which vocal interactions are influenced by social hierarchy. Preliminary results with several pairs of singing mice suggest that, for a given pair, the order in which they call back and forth is asymmetric and strongly influenced by novelty and social status. Moreover, each consecutive session with the same individuals has a reduced counter-singing response. We are using this behavioral output to determine whether singing mice can use sensory information encoded in songs to identify other individuals. Ultimately, this complex social behavior may form the basis to understand the neural circuits underlying this ability to have rapid conversational vocal exchanges. This will help us understand how neural circuits compute social hierarchy information and process auditory cues to flexibly modify vocal behaviors based on individual recognition.

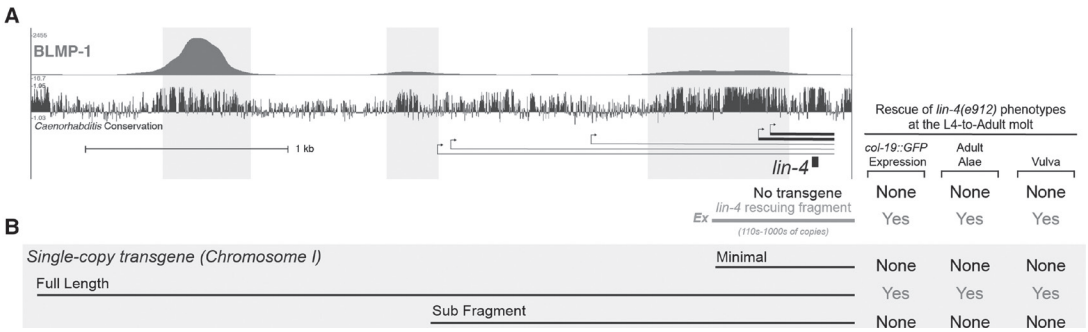


Figure 3. Divergent vocal behaviors between laboratory mice and singing mice. (*A*) Males of both species produce ultrasonic vocalizations (USVs) that last for ~200 ms. In addition to the USVs, only the singing mice produce a "song"—a series of approximately 100 notes that evolve predictably over ~10 s. The ordinate represents frequency from 0 to 100 kHz. (*B*) Representative brain slices from a laboratory mouse (*left*) and a singing mouse (*right*).

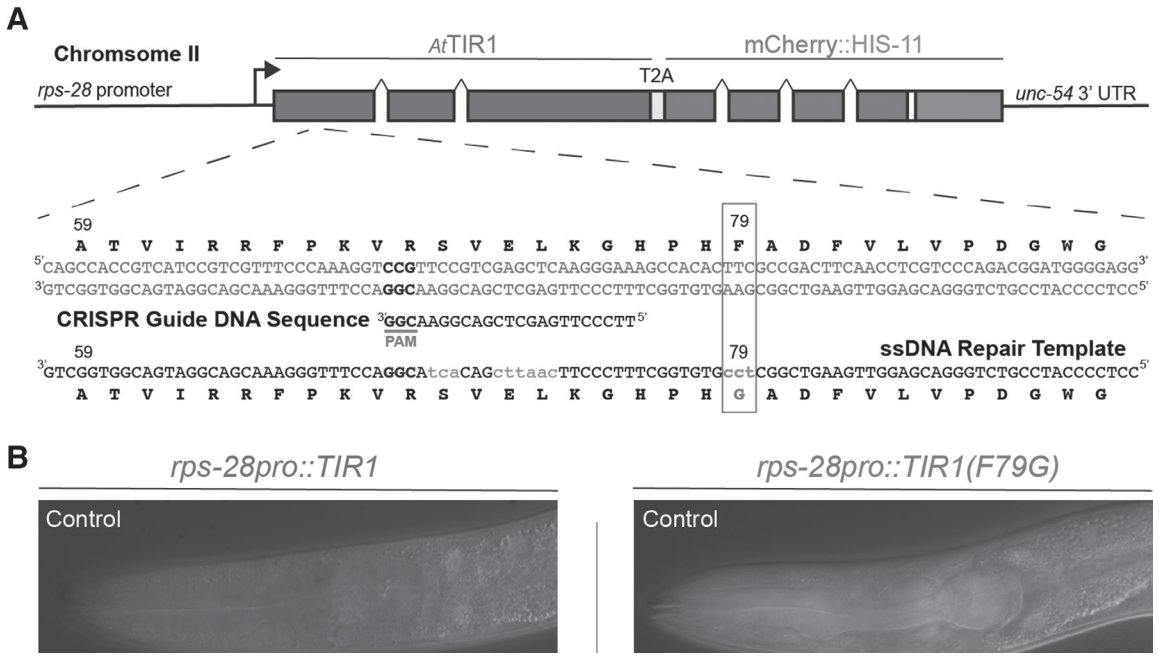


Figure 4. Workflow for MAPseq. (*A*) Anterograde tracing in the laboratory mouse showing orofacial cortex (OMC) projections to many subcortical brain areas. (*B*) MAPseq starts with the sindbis virus injections in OMC carrying RNA barcodes. By dissecting potential downstream brain regions (guided by the bulk tracing experiments) and then precisely identifying the projections of individual neurons via RNA sequencing, projection patterns of thousands of OMC neurons can be determined within a single animal in a single experiment. (*C*) Preliminary MAPseq data of projection patterns of 6207 OMC neurons from one laboratory mouse. OMC projections to 12 downstream brain areas are consistent with known results and are reproducible in two different brains (*top*).

Comparative Connectomics and Transcriptomics in Laboratory Mice and Singing Mice

E. Isko [in collaboration with the Zador laboratory, CSHL]

Despite a close evolutionary relationship, Alston's singing mice (*S. teguina*) and laboratory mice (*Mus musculus*) exhibit extremely divergent vocal behaviors (Fig. 3). This behavioral divergence must derive from differences in the underlying biology of neural cell types and/or neural circuits. To test this hypothesis, we are applying technology developed by the Zador laboratory to determine the neural circuits of vocal behavior in laboratory and singing mice. We have begun our connectomic comparison in the OMC, a brain area involved in the singing mouse song proven through electrical, pharmacological, and cooling experiments (Okobi et al., *Science* **263:** 983 [2019]). We are using both bulk methods (viral tracing followed by whole-brain two-photon tomography) and single-cell barcoding methods (MAPseq) to characterize the projection patterns of neurons in the OMC of the two species (Fig. 4). Our future directions include using single-nucleus RNA-seq (snRNAseq) and BARseq2 (an in situ sequencing technique that combines gene expression and projection data) to determine differences in cell type and the spatial location of these cells in the OMC of laboratory and singing mice. We hope to discover the differences in molecular identity, spatial distribution, and projection patterns of specific cell types in the OMC of the two species. Finding transcriptomic and neural circuit differences between laboratory and singing mice will give insights into the evolution of neural circuits underlying vocal behavior.

PUBLICATIONS

Banerjee A, Egger R, Long MA. 2021. Using focal cooling to link neural dynamics and behavior. *Neuron* **109:** 2508–2518. doi:10.1016/j.neuron.2021.05.029

In Press

Banerjee A, Vallentin D. 2022. Convergent behavioral strategies and neural computations during vocal turn-taking across diverse species. *Curr Opin Neurobiol* **73:** 102529. doi:10.1016/j.conb.2022.102529

ILLUMINATING NEUROIMMUNE AND NONNEURONAL MECHANISMS OF BRAIN DEVELOPMENT, PLASTICITY, AND FUNCTION

L. Cheadle Y. Auguste A-S. Nichitiu
J. Dixon U. Vrudhula
A. Ferro A. Xavier
J. Kahng

The establishment of the precise number and organization of synaptic connections between neurons in the brain is essential for mature neurological function. Whereas initial stages of nascent synapse formation are organized by intrinsic genetic mechanisms (nature), immature synapses are either strengthened or eliminated early in postnatal life in response to sensory experience (nurture). A critical but understudied component of late-stage brain development, the sensory-dependent elimination of excess synapses, reroutes computational power to the subset of remaining synapses that will ultimately mediate mature brain function. However, our understanding of the fundamental mechanisms underlying synapse elimination remains limited, particularly in comparison with what is known about the earliest stages of synapse assembly. Given mounting evidence that deficits in sensory-dependent synapse elimination contribute to a host of neurodevelopmental and psychiatric disorders, including autism and schizophrenia, addressing this gap in knowledge is critical for establishing a detailed understanding of brain development and plasticity in health and disease.

The overarching goal of our research program is to systematically characterize the cellular and molecular mechanisms through which sensory experience shapes the development and plasticity of the brain. Based on our discovery that interactions between neurons and a specialized class of brain-resident immune cells called microglia are coordinated by sensory experience to modify synaptic connections, our work focuses on understanding how communication between neurons and immune cells, and between neurons and other types of nonneuronal brain cells, orchestrates brain function. We take a unique multidisciplinary approach merging in vivo two-photon microscopy with genome-scale molecular assays to (1) characterize the spatial and temporal dynamics of sensory-dependent synapse elimination and remodeling in vivo; (2) investigate the downstream mechanisms through which cytokine signaling between microglia and neurons disassembles active synapses; and (3) discover the genetic regulatory machinery through which sensory experience controls gene expression in microglia. We predict that these initial projects will lay the foundation for our long-term goals of understanding how immune cells and other nonneuronal cell types contribute to sensory integration and neural coding in the heathy brain, and how impairments in neuroimmune interactions lead to neurodevelopmental, psychiatric, and neurodegenerative disorders.

Establishing Tools to Characterize the Spatial and Temporal Dynamics of Synapse Elimination In Vivo

A. Ferro

Our studies in fixed tissue indicate that microglia eliminate synapses through both phagocytic and nonphagocytic mechanisms during development. However, previous studies have been insufficient to define the precise spatial and temporal dynamics of these processes because of the relatively low spatial resolution of two-photon imaging and the widely appreciated difficulty of infecting microglia with fluorescent markers, among other significant technical challenges. Over the past year, we have begun to address these challenges. For example, we have developed a novel lentiviral strategy for selectively infecting microglia in vivo with a fluorescent sensor that is only visible when an infected microglial cell is in direct contact with a synapse. Additionally, we have built a custom, state-of-the-art two-photon microscope specialized to image these and similar viral tools in three separate channels in rapid succession in vivo and are now poised to detect and quantify direct interactions between microglia and synapses in the visual cortex of live, awake mice as they respond to physiologically relevant visual stimuli.

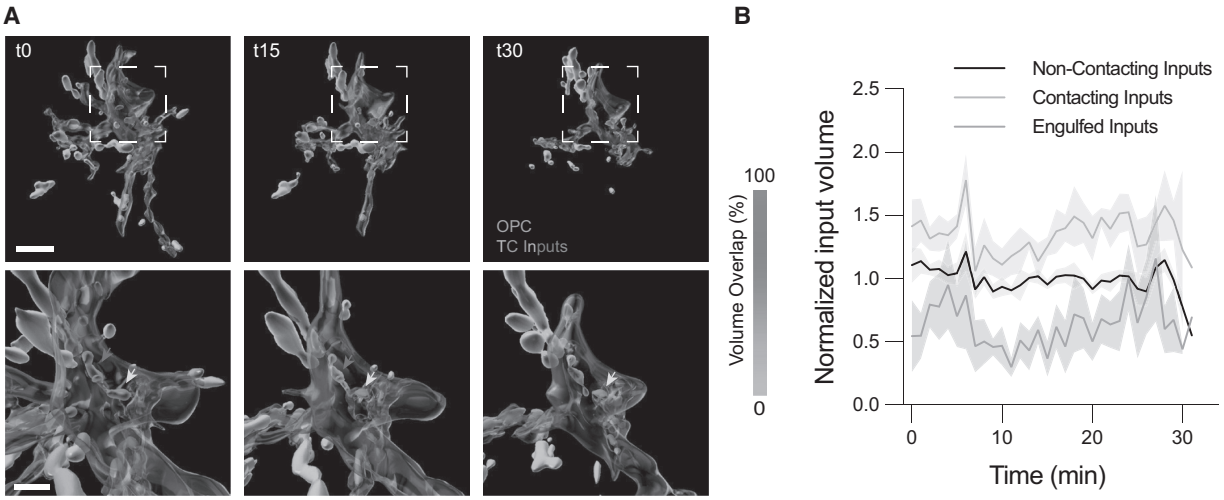


Figure 1. Oligodendrocyte precursor cells engulf synaptic inputs in mouse visual cortex. (*A*) Volumetric reconstructions of oligodendrocyte precursor cells (OPCs; green) interacting with synaptic inputs colored based on percentage overlap with the OPC (color legend to *right* of images). Inputs that are completely internalized are shown in magenta, whereas those that do not contact the OPC are in cyan. Representative images are taken from a 30-min time-lapse imaging session on a two-photon microscope. Scale bars, 10 µm (*top*); 5 µm (*bottom*). Yellow arrows show engulfed input that is present throughout imaging session. Magenta arrows show engulfed input that disappears during session. (*B*) Average volume of all inputs across the imaging period depending on whether the inputs are in contact with or engulfed by OPCs. Mixed effects analysis of synapse category, $p < 0.001$; time, $p < 0.05$. $n = 6$ videos taken from three mice.

Through these studies, we expect to uncover the dynamic structural basis of sensory-dependent synapse elimination by microglia in vivo for the first time.

Uncovering a New Role for Oligodendrocyte Precursor Cells in the Developing Brain

Y. Auguste, A. Ferro, J. Dixon, J. Kahng, A-S. Nichitiu

In developing visual circuits, microglia tend to engulf far more synapses before the onset of sensory experience at eye-opening than afterward, raising the possibility that alternative mechanisms may also contribute to synapse elimination during sensory-dependent phases of circuit development. One of the most exciting discoveries made in the Cheadle laboratory in 2021 was the discovery that oligodendrocyte precursor cells (OPCs), a relatively poorly understood class of non-neuronal progenitors in the brain, eliminate synapses in response to sensory experience through phagocytic engulfment. Applying a combination of standard confocal, super-resolution, and in vivo imaging using the laboratory's new two-photon microscope (described above), we have demonstrated that cortical OPCs internalize presynaptic inputs and digest them within acidic lysosomal compartments (Fig. 1). OPCs engulf more synapses when visual experience is heightened, and this experience-dependent increase in engulfment requires signaling between OPCs and microglia. A School of Biological Sciences graduate student in the laboratory, Jessica Kahng, is currently following up on these findings using a reductionist in vitro OPC-neuron co-culture system to uncover the detailed molecular mechanisms through which OPCs engulf synapses. These discoveries have opened up several new avenues in the lab focused on understanding the relevance of OPC-mediated engulfment for healthy brain development and function, as well as the contributions of synapse engulfment by OPCs to neurodevelopmental and neurodegenerative pathology.

Investigating the Downstream Mechanisms through Which TWEAK-Fn14 Cytokine Signaling Eliminates Active Synapses

A. Ferro, U. Vrudhula, Y. Auguste

Previously, we discovered the TWEAK-Fn14 cytokine signaling pathway as a mechanism through which microglia instruct synapse elimination in the developing brain. However, the downstream mechanisms through which the binding of the neuronal cell surface receptor

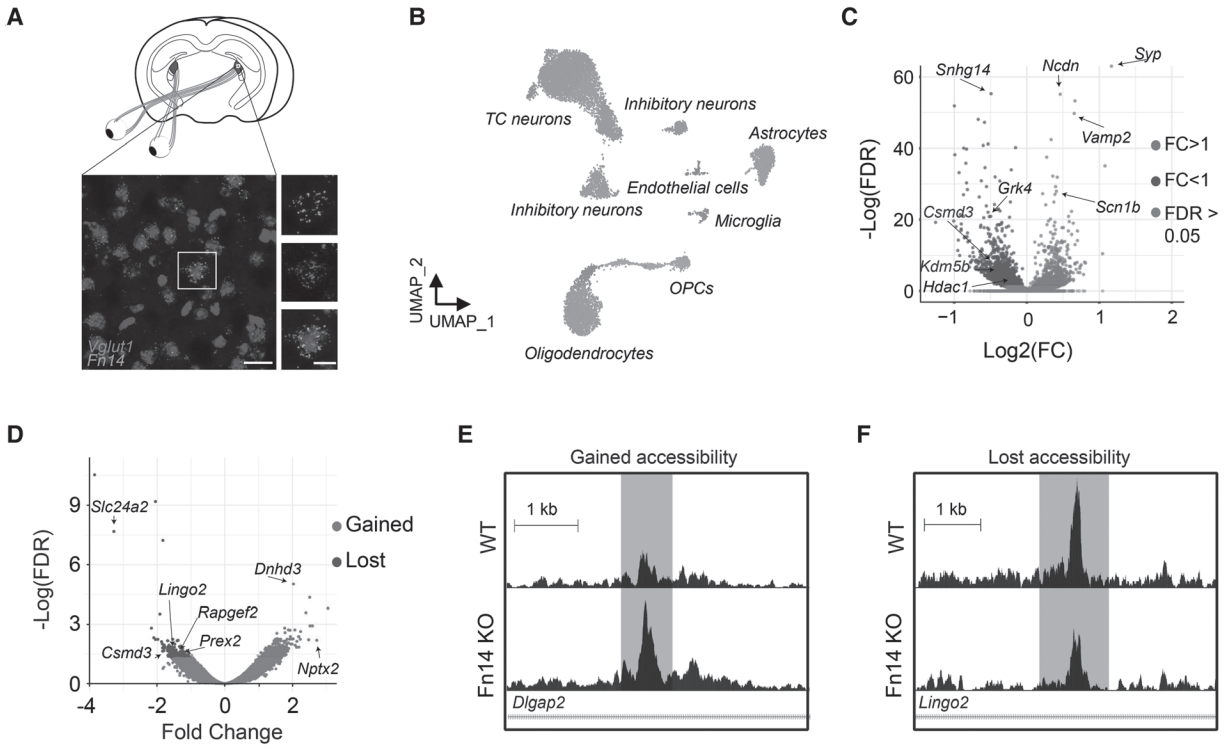


Figure 2. Fn14 regulates transcription and chromatin accessibility in developing neurons. (*A*) Fluorescence in situ hybridization of the thalamocortical neuron marker *Vglut1* and *Fn14* in the dorsolateral geniculate nucleus (dLGN). Scale bars, 20 µm. (*B*) UMAP (uniform manifold approximation and projection for dimension reduction) plot displaying cell clusters present in the single-nucleus RNA sequencing (snRNA-seq) data set. (*C*) Volcano plot of differentially expressed genes in thalamocortical cluster 1. Genes significantly down-regulated in the knockout (KO) shown in blue, nonsignificantly changed genes in gray, and significantly up-regulated genes in the KO shown in red. (*D*) Volcano plot displaying all ATAC peaks with red points representing genomic regions that are significantly more accessible in Fn14 KO tissue, and blue points representing regions that are significantly less accessible in the KO (FDR < 0.05). (*E*) Example of an intronic region within *Dlgap2* that is significantly more accessible in the Fn14 KO compared to WT. *x*-axis, chromosomal location; *y*-axis, average reads. (*F*) Example of an intronic region within *Lingo2* that is significantly less accessible in the Fn14 KO compared to WT. *x*-axis, chromosomal location; *y*-axis, average reads mapped.

Fn14 to the microglia-derived cytokine TWEAK promotes synapse elimination remained largely unknown. By applying single-nucleus RNA sequencing to the dorsolateral geniculate nuclei (dLGNs) of Fn14 knockout (KO) and wild-type (WT) littermate mice, we discovered that Fn14 coordinates robust transcriptional programs in developing neurons, including the up-regulation of genes encoding histone lysine demethylases (KDMs), a family of nuclear proteins that impact gene expression by modifying the histones that are in tight contact with genomic DNA (Fig. 2A–C). Consistent with Fn14 mediating chromatin accessibility via the regulation of KDM expression, assay for transposase-accessible chromatin followed by sequencing (ATAC-seq) revealed about 700 regions of the genome that were more or less accessible in the absence of Fn14, many of which were predicted to be *cis*-regulatory regions that control the expression of genes with key roles in synapse signaling, remodeling, and function (Fig. 2D–F). Coupling these molecular approaches with behavioral and physiological analyses in vivo, we have defined a role for Fn14 in memory and neuronal excitability that we believe may be mechanistically linked to the epigenomic and transcriptomic maturation of developing neurons. Importantly, these data highlight that neuronal receptors tuned to respond to cytokine signals from microglia not only mediate the local elimination of synapses, but may shape more global aspects of neuronal maturation as well (e.g., the chromatin landscape). These findings are reported in a manuscript that is currently under revision following an initial round of review, and a preliminary draft of the manuscript can be found on bioRxiv (Ferro et al. 2021a). We also published a review article focused on the roles of microglia and cytokine

signaling in brain development and function, which highlights our findings (Ferro et al. 2021b).

Elucidating the Genetic Regulatory Mechanisms through Which Sensory Experience Controls Gene Expression in Microglia

A. Xavier, Y. Auguste, J. Kahng

One of the most unexpected findings to arise from our early work is the observation that microglia mount robust transcriptional responses to sensory stimulation. Although experience-dependent transcription has been studied in neurons for many years, virtually nothing is known about the genetic regulatory machinery that controls this process in microglia. Following up on this finding, we performed ATAC-seq on microglia isolated from the visual cortex of control, sensory-deprived, and sensory-stimulated mice, identifying more than 1,500 genomic regions that become accessible to transcription factor binding following light stimulation. Remarkably, almost no regions lose accessibility following stimulation, indicating that experience selectively increases chromatin accessibility in microglia across the genome. To map microglial responses to experience in more detail, we are following up on these studies by performing single-cell multiomics (RNA + chromatin accessibility) analysis of microglia from stimulated and unstimulated mice and combining this analysis with spatial transcriptomics. We ultimately expect these studies to define the genetic regulatory machinery that coordinates experience-dependent transcription in microglia.

A New Partnership with Northwell Health Exploring the Molecular Basis of Temporal Lobe Epilepsy

A. Xavier

Epilepsy is the most common neurological disorder, impacting 1 in 26 people at some point during their life span. It is also a form of pathology that co-occurs with a large number of other neurological diseases ranging from autism to Alzheimer's disease. In 2021, we launched a collaboration with Dr. Ashesh Mehta and his neurosurgical team at Northwell Health to study the molecular basis of epilepsy. Dr. Mehta performs tissue resections on patients with intractable temporal lobe epilepsy, removing tissue that is epileptiform in nature as well as relatively healthy tissue surrounding the epileptic locus. We are using single-cell RNA sequencing to uncover the transcriptional programs that are altered across all cortical cell types in the epileptiform compared to the healthy control tissue, and applying histological approaches to assess interactions between microglia and synapses in the tissue as well. Ultimately, we hope to harness these insights to identify potential therapeutic targets for treating epilepsy.

PUBLICATIONS

Ferro A, Vrudhula U, Auguste YSS, Cheadle L. 2021a. Microglia and Fn14 regulate transcription and chromatin accessibility in developing neurons. bioRxiv doi:10.1101/2021.08.16.456505v2

Ferro A, Auguste YSS, Cheadle L. 2021b. Microglia, cytokines, and neural activity: unexpected interactions in brain development and function. *Front Immunol* **1:** 703527. doi:10.3389/fimmu.2021.703527

DYNAMICS OF BRAIN NETWORKS FOR PERCEPTION AND COGNITION

T. Engel C. Aghamohammadi, J. Couto, M. Genkin, C. Langdon, J. Roach, Y. Shi, P. Tolmachev, J. Wang

Core brain functions—perception, attention, decision-making—emerge from complex patterns of neural activity coordinated within local microcircuits and across brain regions. Recent advances in massively parallel neural recording technologies have enabled 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. We also take part in the NeuroAI program at CSHL to develop the next generation of artificial intelligence systems inspired by the brain (Bhattasali et al. 2022).

Flexible Identification of Neural Population Dynamics underlying Decision-Making

M. Genkin [in collaboration with C. Chandrasekaran, Boston University; K. Shenoy, Stanford University]

Behaviorally relevant signals are often represented in neural population dynamics, which evolve on a low-dimensional manifold embedded into a high-dimensional space of neural responses. Revealing population dynamics from spikes is challenging because the dynamics and embedding are nonlinear and obscured by diverse responses of individual neurons and spiking noise. For example, the decision-related activity of single neurons was hypothesized to arise from either gradual ramping or abrupt stepping dynamics on single trials, but selection between these alternatives is brittle because of diversity of neural responses. Moreover, ramping and stepping are impoverished hypotheses for heterogeneous decision-related neural populations. We need frameworks that can flexibly identify neural dynamics from data.

We developed a flexible framework for inferring neural population dynamics from spikes (Genkin et al. 2021). In our framework, population dynamics are modeled by a latent dynamical system defined by a potential function. The framework covers a continuous space of hypotheses represented by different potential shapes. The activity of individual neurons is related to the population dynamics through unique firing-rate functions, which account for the heterogeneity of neural responses. We incorporate the distribution of the population state at the trial start and boundary conditions modeling trial termination, such as commitment to a choice when reaching a decision boundary. The potential, firing-rate functions, and initial state distribution are continuous functions that are inferred from data.

On simulated neurons, our framework correctly recovered the ramping and stepping models, which correspond to linear and three-well potentials, respectively (Genkin et al. 2021). We applied the framework to neurons recorded simultaneously from the macaque premotor cortex during decision-making. The inferred potential revealed dynamics that evolve gradually toward the correct choice but have to overcome a potential barrier toward the incorrect choice. This discovery is inconsistent with the simple hypotheses proposed previously and in agreement with attractor dynamics in decision-making circuits (Genkin and Engel 2021). Our results demonstrate that a flexible approach can discover new hypotheses about population dynamics from data.

Latent Circuit Inference from Heterogeneous Neural Responses during Cognitive Tasks

C. Langdon

Higher cortical areas carry a wide range of sensory, cognitive, and motor signals supporting complex goal-directed behavior. These signals are mixed in heterogeneous responses of single neurons tuned to multiple task variables. Dimensionality reduction methods used to analyze neural responses rely merely on correlations, leaving unknown how heterogeneous neural activity arises from connectivity to drive behavior. We developed a framework for inferring a low-dimensional connectivity structure—the latent circuit—from high-dimensional neural response data (Langdon and Engel 2022). The latent circuit captures mechanistic interactions between task variables and their mixed representations in single neurons. We applied the latent circuit inference to recurrent neural networks trained to perform a context-dependent decision-making task and found a suppression mechanism in which contextual representations inhibit irrelevant sensory responses. We validated this mechanism by confirming the behavioral effects of patterned connectivity perturbations predicted by the latent circuit structure. Our approach can reveal interpretable and causally testable circuit mechanisms from heterogeneous neural responses during cognitive tasks.

Specificity of Inhibition in Cortical Decision-Making Circuits

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

During perceptual decision-making, the firing rates of cortical neurons reflect upcoming choices. Recent work showed that excitatory and inhibitory neurons are equally selective for choice. However, functional consequences of inhibitory choice selectivity in decision-making circuits are unknown. We developed a circuit model of decision-making that accounts for the specificity of inputs to and outputs from inhibitory neurons (Roach et al. 2022). We found that selective inhibition expands the space of circuits supporting decision-making, allowing for weaker or stronger recurrent excitation when connected in a competitive or feedback motif. The specificity of inhibitory outputs sets the trade-off between speed and accuracy of decisions by altering the attractor dynamics in the circuit. Recurrent neural networks trained to make decisions display the same dependence on inhibitory specificity and the strength of recurrent excitation. Our results reveal two concurrent roles for selective inhibition in decision-making circuits: stabilizing strongly connected excitatory populations and maximizing competition between oppositely selective populations.

Regional Heterogeneity of Sleep and Waking States and Their Alteration through Disease

J. Wang [in collaboration with I. Timofeev, Laval University; J. Borniger and T. Janowitz, CSHL]

Brain states are conventionally divided into wake, slow wave sleep (SWS), and rapid eye movement (REM) sleep based on distinct patterns of neural activity and muscle tone. However, recently available large-scale recordings indicate that this conventional division of brain states is insufficient to account for the rich heterogeneity of neural dynamics on the global scale. During sleep, neural activity in some brain regions can exhibit awake signatures and vice versa. While brain states provide the backdrop for any activity underlying behavioral functions, the spatiotemporal structure of multiregional brain states remains unexplored.

We analyzed simultaneous recordings of electromyogram (EMG) and local field potentials (LFPs) at 14 sites across the mouse cortex during the natural variation of sleep and wake cycles continuously over multiple days (Wang et al. 2022). To characterize the heterogeneity of brain states in these multiregional recordings, we developed an approach to uncover a low-dimensional manifold on which these states evolve. We use unsupervised dimensionality reduction based on a variational autoencoder (VAE) that predicts the next point in time. We trained the model on activity from an individual channel to uncover a local characterization of brain states. For single channels, the inferred manifold revealed three major clusters corresponding with human-expert labels of the basic wake, SWS, and REM states. Classical frequency bands, such as alpha, beta, and gamma, contributed nonlinearly to the inferred manifold. Applying the model to other electrodes, we found profound differences in the expression of states across cortical areas—particularly, the lack of REM-like activity in the lateral somatosensory cortex. We found that heterogeneity of states largely appears during transition periods between primary states, suggesting a more continuous

global manifold. Our work provides a framework for quantifying heterogeneous brains states and shows that the regional coexistence of wake and sleep states is a common feature of global brain activity. Finally, we apply the model to longitudinal EEG recordings in mice during cancer progression from start through final days. Our approach is a sensitive tool for identifying the robust components of the sleep–wake cycle, their reflection in heterogeneous spatiotemporal activity across the brain, and alteration through disease.

Variability and Correlations in Cortical Networks

Y. Shi, C. Aghamohammadi [in collaboration with N.A. Steinmetz, Washington University; T. Moore, Stanford University; A. Thiele, Newcastle University; A. Levina, University of Tübingen; C. Chandrasekaran, Boston University]

Neocortical activity fluctuates endogenously, with much variability shared among 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 develop network models to explain how this structure emerges from connectivity and inputs and how it relates to behavior.

We developed network models 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 in networks with spatial connectivity, correlated fluctuations at different spatial frequencies give rise to multiple intrinsic timescales. 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 and spatial correlations of endogenous activity fluctuations change during spatial attention in the primate visual cortex (Engel et al. 2021; Zeraati et al. 2021, 2022; Shi et al. 2022).

Further, we developed a theoretical framework for partitioning spiking variability of single neurons into sources due to the variability of spike generation (defined by the neuron's physiology) and variations of the firing rate (reflecting states of the surrounding network). Separating these different sources can reveal computations carried out by the network. We model spike trains as doubly stochastic renewal point processes, with variability of spike generation captured by a scalar parameter ϕ and firing-rate variability by a time-varying function $\lambda(t)$. We derived a decomposition of the spike-count variance and developed a method for estimating ϕ from spike data. We validated our framework using intracellular voltage recordings in the mouse cortex and then applied it to extracellular spike recordings from the primate cortex. We found that ϕ is a neuron-specific constant across different task conditions and difficulties, which provides a constraint for biophysical neural network models.

PUBLICATIONS

Engel TA, Schölvinck ML, Lewis CM. 2021. The diversity and specificity of functional connectivity across spatial and temporal scales. *NeuroImage* **245:** 118692.

Genkin M, Engel TA. 2021. Flexible identification of population dynamics from neural activity recordings. Keynote & Tutorial at the Conference on Cognitive Computational Neuroscience, 2021. YouTube video, September 25, 2021. https://www.youtube.com/watch?v=H7Ec3thwVkA

Genkin M, Hughes O, Engel TA. 2021. Learning non-stationary Langevin dynamics from stochastic observations of latent trajectories. *Nat Commun* **12:** 5986.

Zeraati R, Shi Y, Steinmetz NA, Gieselmann MA, Thiele A, Moore T, Levina A, Engel T. 2021. Attentional modulation of intrinsic timescales in visual cortex and spatial networks. bioRxiv doi:10.1101/2021.05.17.444537v1

In Press

Bhattasali NX, Zador AM, Engel T. 2022. Neural circuit architectural priors for embodied control. arXiv doi:10.48550/arxiv.2201.05242

Langdon C, Engel TA. 2022. Latent circuit inference from heterogeneous neural responses during cognitive tasks. bioRxiv doi:10.1101/2022.01.23.477431v1

Roach JP, Churchland AK, Engel TA. 2022. Two roles for choice selective inhibition in decision-making circuits. bioRxiv doi:10.1101/2022.01.24.477635v1

Shi Y, Steinmetz NA, Moore T, Boahen K, Engel TA. 2022. Cortical state dynamics and selective attention define the spatial pattern of correlated variability in neocortex. *Nat Commun* **13:** 44.

Wang J, Chuavette S, Kwapich R, Timofeev I, Engel TA. 2022. A manifold of heterogeneous vigilance states across cortical areas. Computational and Systems Neuroscience (Cosyne) poster.

Zeraati R, Engel TA, Levina A. 2022. A flexible Bayesian framework for unbiased estimation of timescales. *Nat Comput Sci* **2:** 193–204.

STRUCTURES AND REAGENTS OF NEURONAL RECEPTORS AND CHANNELS

H. Furukawa E. Chou R. Gómez K. Michalski J. Syrjanen
M. Epstein S. Kleeman N. Simorowski

Tight regulation of neurotransmission is crucial for brain development and functions, including learning and memory. These processes are largely facilitated by the transport of ions and substrates across the biological membranes through a number of molecular machines, including neurotransmitter-gated ion channels and large-pore channels. The research in the Furukawa laboratory works toward understanding neurotransmission and neuroplasticity at the molecular level with a scope to develop reagents with therapeutic potentials for neurological diseases and disorders including schizophrenia, depression, stroke, and Alzheimer's disease. To achieve our goals, we implement structural and functional studies on cell surface receptors and ion channels that regulate intracellular calcium signaling upon stimulation by voltage and/or neurotransmitters. These ion channels regulate the strength of neurotransmission, which comprises fundamental molecular events that control neuronal activities. Here, we describe two major projects that the Furukawa laboratory currently works on: (1) reagent development for ligand-gated ion channels called *N*-methyl-D-aspartate receptors (NMDARs); and (2) structural and functional studies on a large-pore channel called pannexin. The abnormal activation of these channels is implicated in neurodegeneration, neuroinflammation, and dysfunctional neuronal activities. The Furukawa laboratory employs diverse techniques including cutting-edge X-ray crystallography and single-particle electron cryomicroscopy (cryo-EM) and biochemical and biophysical techniques such as electrophysiology.

Development and Characterization of Antibody Reagents against NMDARs

To isolate functional antibodies against the GluN1-GluN2B NMDARs, we immunized mice with purified intact rat GluN1a-GluN2B NMDAR proteins prepared in lauryl maltose neopentyl glycol (LMNG) (Furukawa et al. 2021). We isolated approximately 30 monoclonal antibodies (mAbs) and specifically selected for ones that recognize folded protein rather than flexible loops or denatured proteins because "folding-specific" antibodies typically recognize the protein surface and have a higher tendency to alter functions of target proteins, as demonstrated previously (Gao et al., *Proc Natl Acad Sci* **106:** 3071 [2009]). Toward this end, we screened for immunoglobulin Gs (IgGs) that showed signal in an enzyme-linked immunosorbent assay (ELISA) using the intact rat GluN1a-GluN2B NMDAR proteins in the presence of 0.01% LMNG and no signal in western blotting executed in a denaturing condition (Fig. 1A). We identified four antibodies that satisfied the above criteria and found one of them, IgG2, that inhibits the activity of the GluN1-GluN2B NMDARs as measured by two-electrode voltage clamp (TEVC) on complementary RNA (cRNA)-injected *Xenopus laevis* oocytes (Fig. 1B). The inhibition occurs in a concentration-dependent manner (Fig. 1B,F; efficacy = 54.2% ± 5.5% of maximum current at 0.1 mg/mL, $n = 5$). Importantly, little or no effect was observed when IgG2 was applied to the oocytes expressing the GluN1-GluN2A, GluN1-GluN2C, and GluN1-GluN2D NMDARs, indicating that this inhibitory effect is specific to the GluN1-GluN2B NMDARs (Fig. 1C–E). Another "protein folding–specific" antibody, IgG5, has a minor potentiating effect rather than an inhibitory effect, implying that the approach to control NMDAR functions by antibodies may be applicable to both up-regulation and down-regulation (Fig. 1G).

We next sought to identify the binding site of Fab2 within the GluN1-GluN2B NMDARs in order to understand the potential mechanism of inhibition by Fab2. Toward this end, we purified the intact GluN1b-GluN2B NMDAR proteins (Chou et al., *Cell* **182:** 357 [2020]; Tajima et al., *Nature* **534:** 63 [2016]) in the presence of 1 mM glycine and 1 mM glutamate and complexed them with the purified Fab2 fragment.

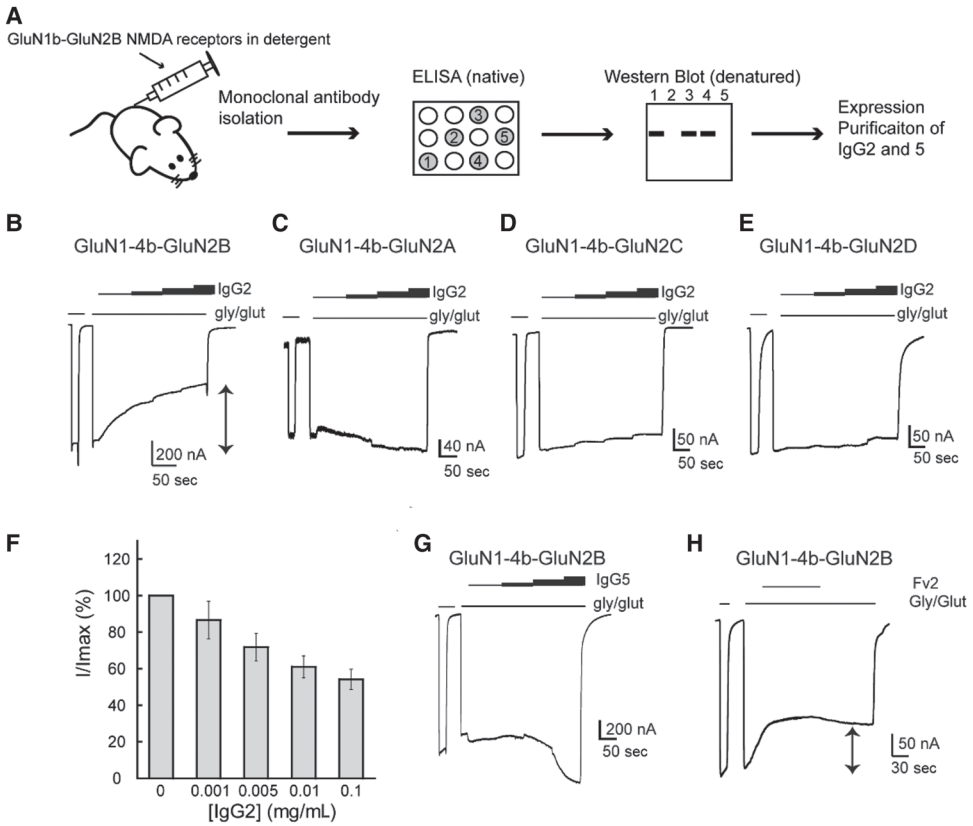


Figure 1. Isolation and characterization of anti-GluN1-GluN2B NMDA receptor IgGs. (*A*) Monoclonal antibodies were produced by mouse immunization by intact rat GluN1b-GluN2B NMDA receptors purified in lauryl maltose neopentyl glycol (LMNG). Clones that produced signal in enzyme-linked immunosorbent assay (ELISA) and no signal in western blot were isolated. (*B–F*) Glycine/glutamate-evoked currents measured by two-electrode voltage clamp (TEVC) on complementary RNA (cRNA)-injected *Xenopus* oocytes expressing rat GluN1b-2B, GluN1b-2A, GluN1b-2C, and GluN1b-2D in the presence of various concentrations (0.001-0.1 mg/mL) of purified IgG2. The specific inhibitory effect of IgG2 on the GluN1b-GluN2B NMDA receptors is dose-dependent. (*G*) Application of the various concentrations (0.001-0.1 mg/mL) of IgG5 has no inhibitory effect, but has a slight potentiating effect at 0.1 mg/mL (111 ± 7.5%; $n = 5$). (*H*) The Fv fragment of IgG2 (Fv2) retains an inhibitory capability. Shown here is the current recorded in the presence of 0.1 mg/mL of Fv2.

Single-particle analysis resulted in three 3D classes that differed from each other mainly in their amino-terminal domain (ATD) conformations (Fig. 2). The resolution of the 3D reconstructions ranged from 3.9 to 6.6 Å as estimated by Fourier shell correlation (FSC) curves; thus, the quality of the cryo-EM density maps was mostly sufficient to identify and trace secondary structural elements as well as the side chains of bulky residues. Furthermore, we solved an X-ray crystallographic structure of Fab2 at 2.5 Å to facilitate model building into the cryo-EM density.

The cryo-EM structures unambiguously showed that Fab2 binds to the R1 lobe of the GluN2B ATD (Fig. 2C). One heterotetrameric GluN1b-GluN2B NMDAR channel is capable of binding two Fab2 fragments at the equivalent region of the two GluN2B subunits. The single-particle cryo-EM showed density for the ATD, the ligand-binding domain (LBD), and the transmembrane domain (TMD) of the GluN1b-GluN2B NMDARs along with the density for the Fv portion of the Fab2 fragment, which interacts only with the GluN2B ATD. The majority of antibody binding involves the GluN2B residues Ser31, Glu55, Asp57-58, Phe59, His60, and Arg67 within the ATD and residues from the complementary determining region (CDR) 2 and CDR 3 from the heavy chain of IgG2 (Fig. 3C,D). The GluN2B residues involved in IgG2 binding are not conserved in the other GluN2 subtypes (A, C, and

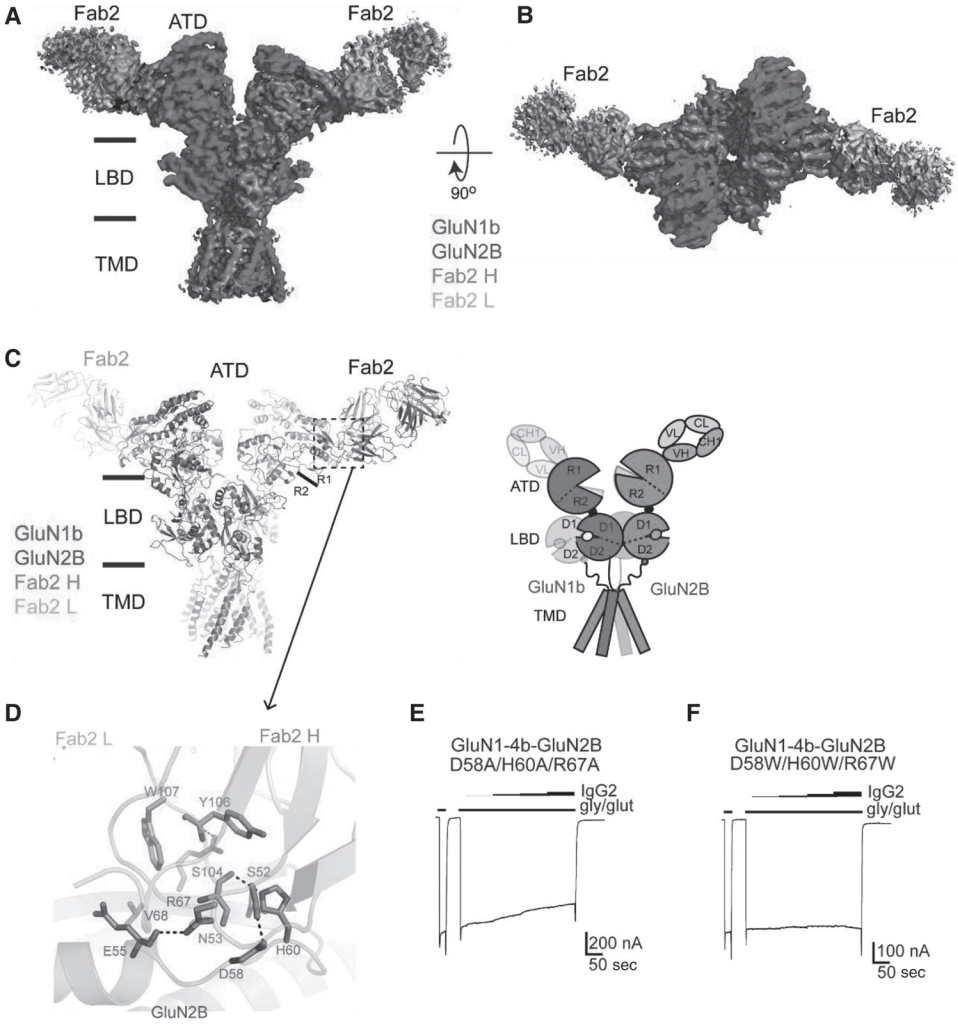


Figure 2. Structural analysis of agonist-bound GluN1b-GluN2B *N*-methyl-D-aspartate receptors (NMDARs) complexed to Fab2. (*A,B*) Cryo-EM density of 3D class 1 at overall resolution of 4.01 Å from the "side" (*A*) and "top" (*B*) of the amino terminus. Densities for the transmembrane domain (TMD), ligand-binding domain (LBD), and amino-terminal domain (ATD) of GluN1b (magenta), GluN2B (dark green), Fab2 heavy chain (orange), and light chain (light pink) were observed. The structure was solved in the presence of 1 mM glycine and glutamate. (*C*) The molecular model built based on the cryo-EM density in the same color code as in *A* and *B*. (*D*) Zoomed view of the ATD-Fab2 interaction site demonstrating residues from CDR 1 and 3 of the heavy chain are mediating polar and hydrophobic interactions. (*E,F*) Interacting residues, Asp58, His60, and Arg67, were mutated to Ala (*E*) or Trp (*F*) and measured for inhibitory effects as in Figure 1. The Ala triple mutant slightly retains the inhibitory effect, whereas the Trp triple mutant completely removes the effect.

D), consistent with the subtype-specific inhibitory effect shown in the electrophysiological experiments (Fig. 1). The binding residues are conserved among the mammalian GluN2B subunits, implying that the IgG2 will likely recognize and inhibit GluN1-GluN2B NMDARs from other mammalian species.

To understand the factors that contribute to the different effects exhibited by IgG2 and IgG5, we sought to determine the binding site for IgG5. For this, we implemented single-particle cryo-EM on the intact GluN1b-GluN2B NMDAR-Fab5 complex (Fig. 3) and X-ray crystallography on the GluN1b-GluN2B ATD complexed to Fab5. Together, these two structural analyses delineated the binding sites as well as protein conformational states in the context of the intact NMDAR channel. The cryo-EM structure of

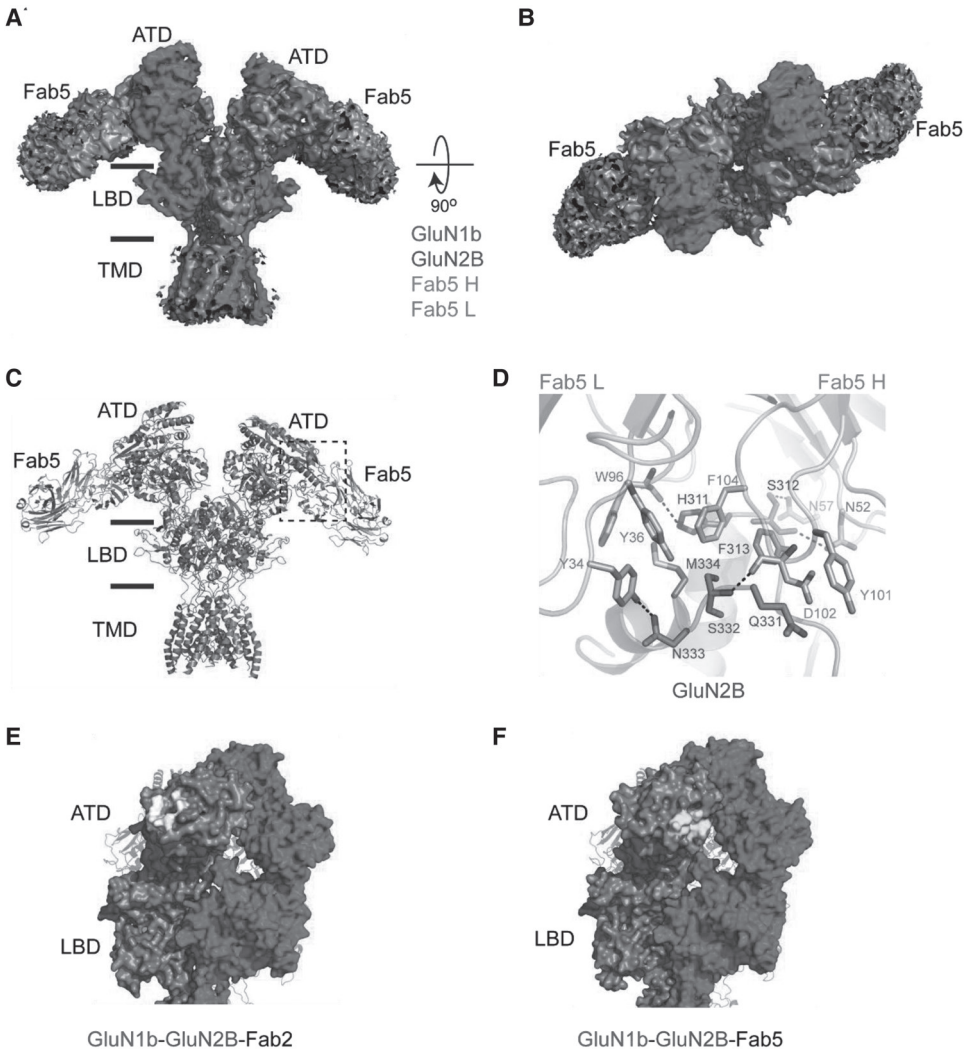


Figure 3. Structural analysis of agonist-bound GluN1b-GluN2B *N*-methyl-D-aspartate receptors (NMDARs) complexed to Fab5. (*A,B*) Cryo-EM density of 3D class 4 at overall resolution of 4.45 Å from the "side" (*A*) and "top" (*B*) of the amino terminus. Densities for the transmembrane domain (TMD), ligand-binding domain (LBD), and amino-terminal domain (ATD) of GluN1b and GluN2B and Fab5 heavy chain and light chain were observed. The structure was solved in the presence of 1 mM glycine and glutamate. (*C*) The molecular model built based on the cryo-EM density. (*D*) Zoomed view of the ATD–Fab5 interaction site. Shown here is the crystal structure of the GluN1b-GluN2B ATD complexed to Fab5 at 4.54 Å. The molecular model around the GluN2B-Fab5 interface fits well into the cryo-EM density. Binding of Fab5 involves residues from CDR 1 and 3 of the light chain and CDR 2 and 3 of the heavy chain. (*E,F*) Surface presentation of residues in GluN2B interacting with Fab2 (*E*, surface) and Fab5 (*F*, surface) illustrating that there is no overlap between the binding sites.

GluN1b-GluN2B NMDAR-Fab5 was obtained at resolutions ranging from 4.45 to 7.51 Å as estimated by FSC curves. Although the overall quality of the cryo-EM density is inferior to that of GluN1b-GluN2B NMDAR-Fab2, it is sufficient to capture patterns of conformational alteration. The specific residues involved in the interaction between GluN1b-GluN2B NMDAR and Fab5 were captured by the X-ray crystallographic structure of GluN1b-GluN2B ATD-Fab5 at 4.54 Å, which shows that the binding involves His311, Ser312, Phe313, Gln331, Ser332, Asn333, and Met334 of GluN2B in the R1 lobe and residues from CDR1 and CDR3 of the light and the heavy chains of IgG5, respectively. Although both IgG2 and

IgG5 bind the GluN2B ATD, they do so at different protein surfaces (Fig. 3E,F).

Extensive 3D classification in the single-particle analyses revealed discrete conformations that were also observed in our recent studies on GluN1b-GluN2B NMDARs with no bound antibodies (Chou et al., *Cell* **182:** 357 [2020]; Tajima et al., *Nature* **534:** 63 [2016]). In these studies, GluN1b-GluN2B NMDARs in the presence of glycine and glutamate reside in the three major conformations, non-active1, non-active2, and active. Non-active1 and non-active2 contain closed and open GluN2B ATD bilobes, respectively. In the active conformation, the GluN2B ATD bilobe is open, and the heterodimeric interface of the GluN1-GluN2B ATD is rearranged, which results in a rolling motion of the two GluN1-GluN2B LBD heterodimers to open the channel gate (Fig. 4). Thus, the prerequisite for activation of the GluN1-GluN2B NMDAR is opening of the GluN2B ATD bilobe. On the other hand, stabilization of the closed GluN2B ATD favors inhibition. Non-active1 is similar to the conformation of the receptor bound to an allosteric inhibitor such as ifenprodil that stabilizes the closed GluN2B ATD bilobe (Karakas et al., *Nature* **475:** 249 [2011]; Regan et al., *Nat Commun* **10:** 321 [2019]).

The cryo-EM data for the GluN1b-GluN2B NMDAR-Fab5 in the presence of glycine and glutamate was classified into four 3D classes (Fig. 4): One corresponds to active, another is similar to non-active1, and the other two are similar to non-active2. This conformational pattern is similar to that of the GluN1b-GluN2B NMDAR with no antibodies bound (Chou et al., *Cell* **182:** 357 [2020]; Tajima et al., *Nature* **534:** 63 [2016]); thus, it is consistent with the observation that IgG5 has little or no effect on the function of GluN1b-GluN2B NMDARs.

The cryo-EM data for the GluN1b-GluN2B NMDAR-Fab2 in the presence of glycine and glutamate was classified into three similar 3D classes, in which two of them correspond to non-active1 with the closed GluN2B ATD bilobe, and the other corresponds to non-active2-like but with the GluN2B ATD bilobe only slightly open (Fig. 5). The closed GluN2B ATD bilobe disallows sufficient reorientation of the GluN1b-GluN2B ATD interface to cause rolling of the GluN1b-GluN2B LBD heterodimers; thus, the channel gate is closed. There is no clear evidence for the presence of protein conformations representing the active and non-active2 conformations with wide-open GluN2B ATD bilobes. Therefore, we suggest that the mechanism of inhibition by Fab2 or IgG2 may involve an alteration of the free energy landscape that results in unfavorable transition from non-active1 to non-active2 and active conformations by stabilization of the closed GluN2B ATD bilobe.

Overall, the above study sets a lead for further investigation aimed at improving the design of the Fab2 (IgG2) aimed at fine-tuning its efficacy and potency for subtype-specific inhibition of GluN1-2B NMDARs.

Large-Pore Biology—Pannexin

Large-pore channels play important roles in cell to cell communication by responding to diverse stimuli and releasing signaling molecules like ATP and amino acids (Giaume et al., *Front Pharmacol* **4:** 88 [2013]; Ma et al., *Pflugers Arch* **468:** 395 [2016]; Okada et al., *Curr Top Membr* **81:** 125 [2018]; Osei-Owusu et al., *Curr Top Membr* **81:** 177 [2018]). Pannexins are a family of ubiquitously expressed large-pore channels that regulate nucleotide release during apoptosis (Chekeni et al., *Nature* **467:** 863 [2010]), blood pressure (Billaud et al., *Circ Res* **109:** 80 [2011]; Billaud et al., *Sci Signal* **8:** ra17 [2015]), and neuropathic pain (Bravo et al., *Pain* **155:** 2108 [2014]; Mousseau et al., *Sci Adv* **4:** eaas9846 [2018]; Weaver et al., *Sci Rep* **7:** 42550 [2017]). Although pannexins have limited sequence identity with innexins (~15% identity), they have virtually no sequence similarity to other large-pore channels (Panchin et al., *Curr Biol* **10:** R473 [2000]). Among the pannexin family, pannexin 1 (Panx1) has garnered the most attention for its role as a large-pore channel responsible for ATP release from a variety of cell types (Bao et al., *FEBS Lett* **572:** 65 [2004]; Dahl, *Philos Trans R Soc Lond B Biol Sci* **370:** 20140191 [2015]). Different kinds of stimuli have been reported to activate Panx1, including voltage, membrane stretch, increased intracellular calcium levels, and positive membrane potentials (Bao et al., *FEBS Lett* **572:** 65 [2004]; Bruzzone et al., *Proc Natl Acad Sci* **100:** 13644 [2003]; Chiu et al., *J Gen Physiol* **150:** 19 [2018]; Locovei et al., *FEBS Lett* **580:** 239 [2006]; Wang et al., *Sci Signal* **7:** ra69 [2014]). Panx1 is also targeted by signaling effectors, such as proteases and kinases, to permanently or temporarily stimulate channel activity (Billaud et al., *Sci Signal* **8:** ra17 [2015]; Lohman et al., *Nat Commun* **6:** 7965 [2015]; Pelegrin and Surprenant, *EMBO J* **25:**

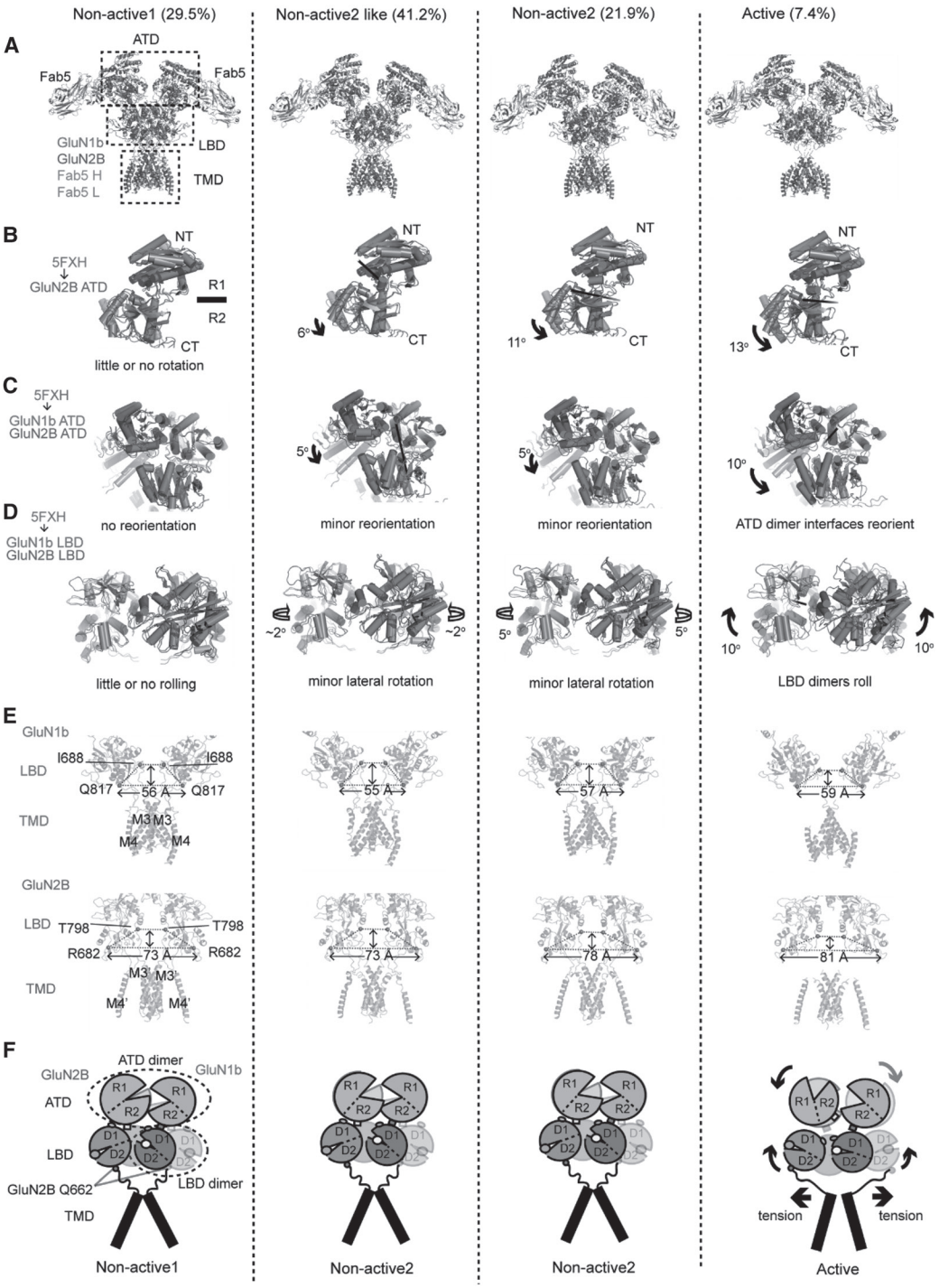


Figure 4. Conformational states of the GluN1b-GluN2B NMDAR-Fab5 complex. (*A*) 3D classes of the GluN1b-GluN2B NMDAR-Fab5 in the presence of glycine and glutamate. The four 3D classes belong to non-active1, non-active2-like, non-active2, and active conformations. (*B–D*) The structures of GluN2B ATD (*B*), GluN1b-GluN2B ATD heterodimer (*C*), and GluN1b-GluN2B LBD heterodimers (*D*) from GluN1b-GluN2B NMDAR-Fab5 are compared to those of GluN1b-GluN2B NMDAR in non-active1 (PDB code 5FXH; colored gray). (*E*) Side views of GluN1b (*upper* panel) and GluN2B (*lower* panel) showing residues around the channel gating ring (GluN1b Gln817 and GluN2B Arg682 in spheres). (*F*) Schematic presentation of interdomain and intersubunit movements.

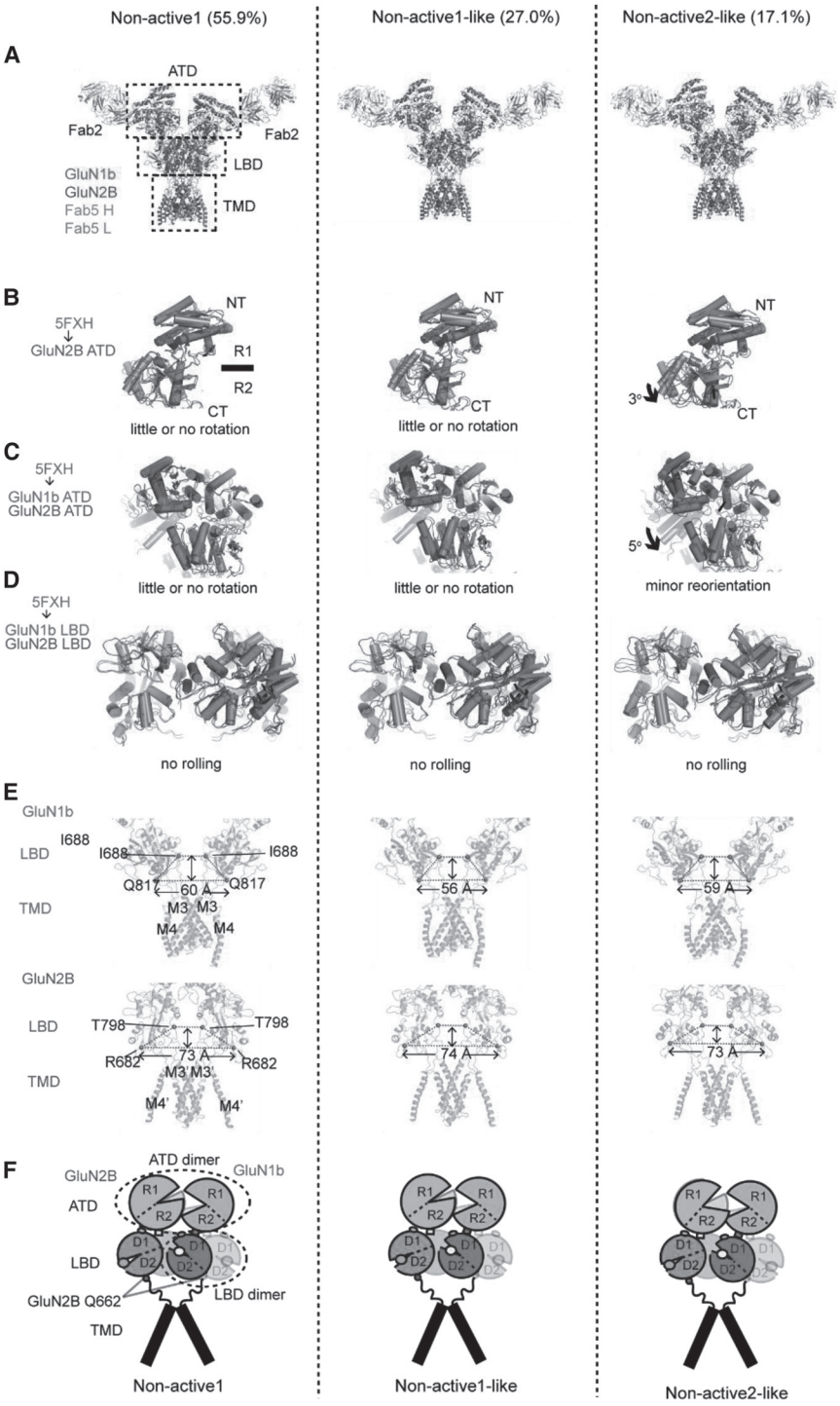


Figure 5. Conformational states of the GluN1b-GluN2B NMDAR-Fab2 complex. (*A*) 3D classes of the GluN1b-GluN2B NMDAR-Fab2 in the presence of glycine and glutamate. The four 3D classes belong to non-active1, non-active2-like, non-active2, and active conformations. (*B–D*) The structures of GluN2B ATD (*B*), GluN1b-GluN2B ATD heterodimer (*C*), and GluN1b-GluN2B LBD heterodimers (*D*) from GluN1b-GluN2B NMDAR-Fab5 are compared to those of GluN1b-GluN2B NMDAR in non-active1 (PDB code 5FXH; colored gray). (*E*) Side views of GluN1b (*upper* panel) and GluN2B (*lower* panel) showing residues around the channel gating ring (GluN1b Gln817 and GluN2B Arg682 in spheres). (*F*) Schematic presentation of interdomain and intersubunit movements.

5071 [2006]; Sandilos et al., *J Biol Chem* **287:** 11303 [2012]; Thompson et al., *Science* **322:** 1555 [2008]). The above evidence suggests that Panx1 has a capacity to integrate distinct stimuli into channel activation, leading to ATP release. Despite playing critical roles in a variety of biological processes, a mechanistic understanding of pannexin function has been largely limited because of the lack of a high-resolution structure. Therefore, we obtained the cryo-EM structure of Panx1, which revealed the pattern of heptameric assembly, pore lining residues, important residues for ion selection, and a putative carbenoxolone binding site for the first time.

Structure of Pannexin1 (Panx1)

The frPanx1-ΔLC structure revealed a heptameric assembly, which is unique among the known eukaryotic channels (Fig. 6). Other large-pore channels include hexameric connexins (Maeda et al., *Nature* **458:** 597 [2009]) and LRRC8s (Deneka et al., *Nature* **558:** 254 [2018]; Kasuya et al., *Nat Struct Mol Biol* **25:** 797 [2018]; Kefauver et al., *eLife* **7:** e38461 [2018]), and the octameric innexins (Oshima et al., *J Mol Biol* **428:** 1227 [2016]) and calcium homeostasis modulator (CALHM1) (Syrjanen et al., *Nat Struct Mol Biol* **27:** 150 [2020]). This unique heptameric assembly is established by intersubunit interactions at three locations: (1) ECL1s and the loop between β2 and β3; (2) TM1-TM1 and TM2-TM4 interfaces; and (3) α9-helix and the surrounding α3- and α4-helices, and the amino-terminal loop from the neighboring subunit (Fig. 7). Notably, the majority of residues mediating these interactions are highly conserved (e.g., Phe67 and Tyr111).

The overall protomer structure of Panx1 resembles that of other large-pore channels including connexins,

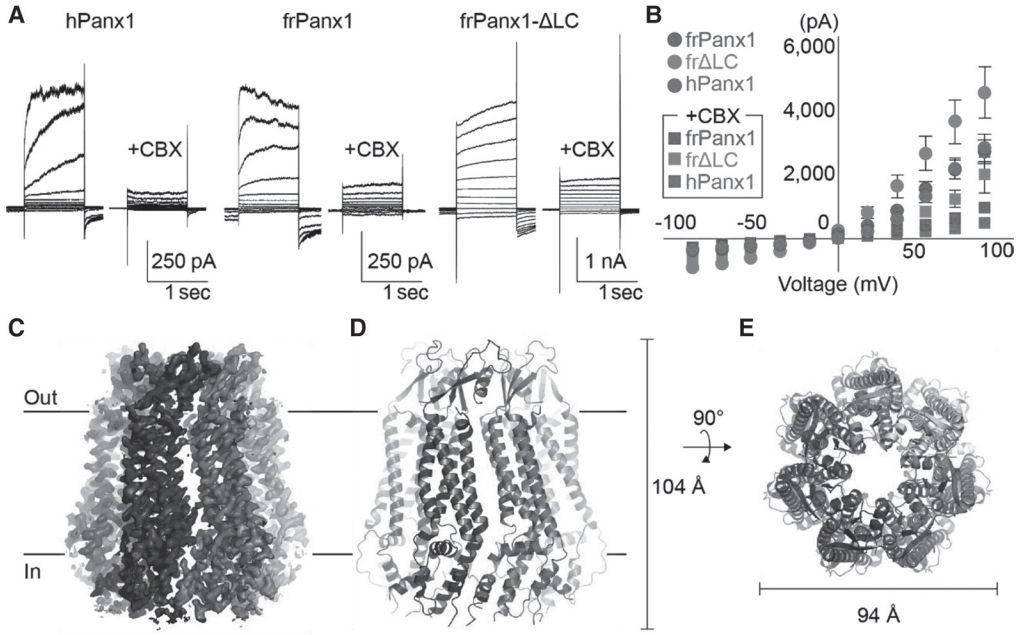


Figure 6. frPanx1 forms a heptameric ion channel. (*A*) Whole-cell patch-clamp recordings from HEK 293 cells expressing hPanx1, frPanx1, and frPanx1-ΔLC. Cells were clamped at –60 mV and stepped from –100 mV to +100 mV for 1 sec in 20-mV increments. To facilitate electrophysiological studies, we inserted a Gly-Ser motif immediately after the start Met to enhance Panx1 channel opening. Carbenoxolone (CBX) (100 μM) was applied through a rapid solution exchanger. (*B*) Current–voltage plot of the same channels shown in *A*. Recordings performed in normal external buffer are shown as circles, and those performed during CBX (100 μM) application are shown as squares. Each point represents the mean of at least three different recordings, and error bars represent the SEM. (*C*) Electron microscopy (EM) map of frPanx1-ΔLC shown from within the plane of the membrane. Each protomer is colored differently, with the extracellular side designated as "out" and the intracellular side as "in." (*D*) Overall structure of frPanx1-ΔLC viewed from within the lipid bilayer. (*E*) Structure of frPanx1 viewed from the extracellular face.

innexins, and LRRC8. Like other large-pore channels, each Panx1 protomer harbors four transmembrane helices (TM1-4), two extracellular loops (ECL1 and 2), two intracellular loops (ICL1 and 2), and an amino-terminal loop (Fig. 7A,B). The transmembrane helices of Panx1 are assembled as a bundle in which the overall helix lengths, angles, and positions strongly resemble the transmembrane arrangements observed in other large-pore channels (Fig. 7C). In contrast, Panx1 has no similarity in transmembrane arrangement to another group of large-pore channels—CALHMs, whose protomers also contain four transmembrane helices (Choi et al., *Nature* **576:** 163 [2019]; Syrjanen et al., *Nat Struct Mol Biol* **27:** 150 [2020]). Structural features in the Panx1 ECL1 and ECL2 domains are conserved among large-pore channels despite limited sequence similarity (Fig. 7D–G). For example, the Panx1 ECL1 and ECL2 are joined together by two conserved disulfide bonds (Cys66 with Cys267, Cys84 with Cys248) in addition to several β-strands. ECL1 also contains an α-helix that extends toward the central pore and forms an extracellular constriction of the permeation pathway. Although much of the transmembrane domains and extracellular loops show similarities to other large-pore channels, the Panx1 intracellular domains are structurally unique (Fig. 2). ICL1 and ICL2, for example, together form a bundle of helices that makes contact with the amino terminus. The amino-terminal loop of Panx1 forms a constriction of the permeation pathway and extends toward the intracellular region. The first ~10 amino acids of the amino terminus are disordered in our structure, but these residues might play a role in ion permeation or ion selectivity (Wang and Dahl, *J Gen Physiol* **136:** 515 [2010]).

Ion Permeation Pathway and Selectivity

The Panx1 permeation pathway spans a length of 104 Å, with constrictions formed by the amino-terminal loop, Ile58, and Trp74 (Fig. 8A,B). The narrowest constriction is surrounded by Trp74 located on ECL1 (Fig. 8C). Trp74 is highly conserved among species including hPanx1 (Fig. 6). Because Panx1 has been previously characterized as an anion-selective channel (Chiu et al., *Channels (Austin)* **8:** 103 [2014]; Ma et al., *Pflugers Arch* **463:** 395 [2012]; Romanov et al., *J Cell Sci* **125:** 5514 [2012]), we wondered if positively charged amino acids around the narrowest constriction formed by Trp74 may contribute to anion selectivity of the channel. Interestingly, Arg75 is situated nearest to the tightest constriction of the permeation pathway (Fig. 8D). We hypothesized that Arg75 might be a major determinant of anion selectivity of Panx1 channels in the open state. To assess whether Arg75 contributes to anion selectivity, we generated a series of point mutations at this position on hPanx1 and compared their reversal potentials (Erev) in asymmetric solutions using whole-cell patch-clamp electrophysiology (Fig. 8E). We kept sodium chloride

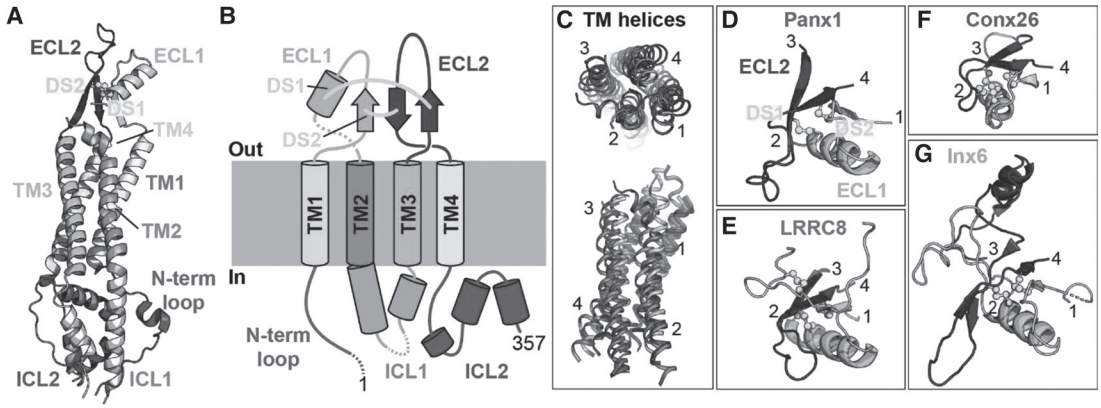


Figure 7. Subunit architecture of frPanx1. (*A*) Structure of the frPanx1 protomer. Each domain is colored according to the cartoon scheme presented in *B*. (*C*) Superimposition of the transmembrane helices from frPanx1, connexin-26, innexin-6, and LRRC8 shown top-down from the extracellular side (*top*) or from within the plane of the membrane (*bottom*). (*D–G*) Cartoon representation of the extracellular loops of large-pore channels. ECL1 is colored in light blue, ECL2 is colored in dark blue, and disulfide bridges are shown as yellow spheres. These domains are viewed from the same angle (from *top*) as shown in the *top* panel in *C*.

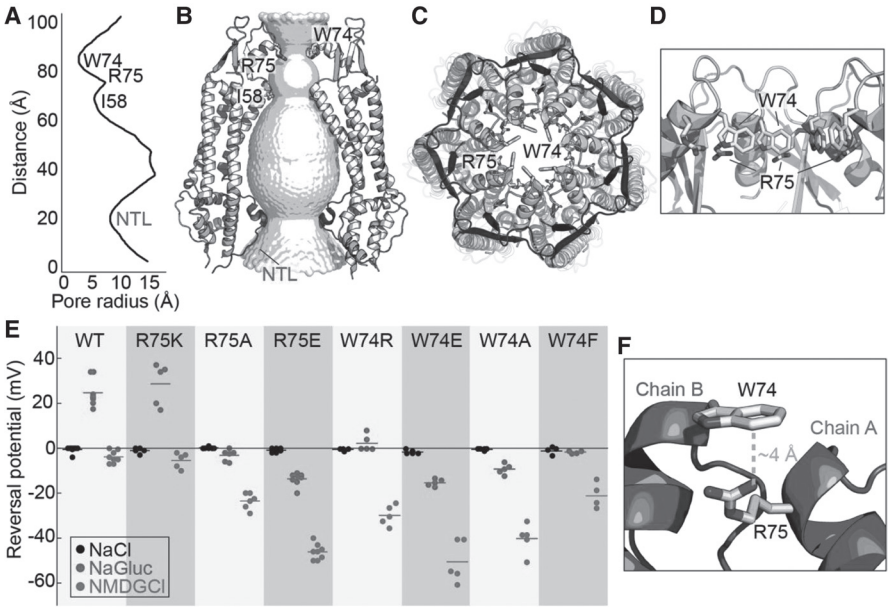


Figure 8. Permeation and ion selectivity of Panx1 channels. (*A*) HOLE (Smart et al., *J Mol Graphics* **4:** 354 [1996]) diagram demonstrating constrictions along the permeation pathway. (NTL) Amino-terminal loop. (*B*) Surface representation of the internal space along the molecular sevenfold axis running through the center of frPanx1. The surface was generated using HOLE. (*C,D*) Top view facing the extracellular side (*C*) or side view (*D*) of frPanx1, with ECL1 shown in light blue and ECL2 in dark blue. Trp74 and Arg75 are shown as sticks. (*E*) Reversal potentials of various hPanx1 ion selectivity mutants. Each point represents the Erev measured in NaCl (black), NaGluc (red), or NMDGCl (blue), and bars represent the mean values. Current–voltage characteristic (I-V) curves were obtained by a ramp protocol from –80 mV to +80 mV. (*F*) Close-up view of the Trp74-Arg75 interaction at the interface of protomer A and B.

(NaCl) constant in the pipette solution while varying the extracellular solution. When treated with the large anion, gluconate ($Gluc^-$), Erev shifted to +26 mV, suggesting the channel is more permeable to Cl^- than to $Gluc^-$. When exposed to the large cation, *N*-methyl-D-glucamine ($NMDG^+$), Erev remained close to 0 mV, suggesting that Na^+ and $NMDG^+$ equally (or do not) permeate Panx1. These results are consistent with Panx1 being an anion-selective channel. The Arg75Lys mutant maintains the positive charge of this position, and displayed Erev values are comparable to wild type (WT). Removing the positive charge at this position, as shown by the Arg75Ala mutant, diminished Cl^- selectivity as the Erev in NaGluc remained near 0 mV. Interestingly, the Erev in NMDGCl shifted to –22 mV, suggesting the channel had lost anion selectivity and Na^+ became more permeable than $NMDG^+$. A charge reversal mutant, Arg75Glu, shifted the Erev in NaGluc to –16 mV and in NMDGCl to –45 mV, indicating that $Gluc^-$ became more permeable to Cl^-. Overall, these results support the idea that the positively charged Arg75 plays a role in anion selectivity of Panx1.

We next wondered if introducing a charge at position 74 might alter ion selectivity of Panx1 channels. Interestingly, both Trp74Arg and Trp74Glu mutants become less selective to anions and more permeable to Na^+ (Fig. 8E). These results suggest that introducing a charge at this position disrupts the natural ion selectivity of Panx1 channels, but that position 74 itself does not control ion selectivity. We observed that the distance between the guanidyl group of Arg75 and the benzene ring of Trp74 from an adjacent subunit is ~4 Å, suggesting that these two residues likely participate in an intersubunit cation-π interaction key to Panx1 ion selectivity (Fig. 8F). To test this hypothesis, we generated Trp74Ala and Trp74Phe mutations and measured Erev potentials. Trp74Ala showed a marked decrease in Cl^- permeability and an increase in Na^+ permeability, despite preservation of the positive charge at Arg75. A more conservative mutation, Trp74Phe, still disrupted ion selectivity, suggesting that proper positioning of the benzene ring at position 74 is important for anion selection. Altogether, our data suggest that anion selectivity is only achieved when Trp74 and Arg75 form a cation-π interaction. Given that our

structure has disordered and truncated regions in the amino terminus, ICL1, and ICL2, it is possible that additional ion selectivity or gating regions exist in the full-length channel. For example, the amino termini of LRRC8 and connexins perform an important role in ion selectivity (Kefauver et al., *eLife* **7:** e48361 [2018]; Kronengold et al., *J Membr Biol* **245:** 453 [2012]; Kyle et al., *J Cell Sci* **121:** 2744 [2008]). It is possible that the amino terminus of Panx1 is mobile and may further constrict the permeation pathway. Another possibility is that the electrostatic potential along the pore pathway contributes to the ion selectivity. Interestingly, both cytoplasmic and extracellular entrances of the permeation pathway are mostly basic, suggesting that nonpermeant cations may be excluded from the pore (Fig. 8). In contrast, the region underneath the W74 constriction is highly acidic, supporting the idea that anions may be selected around this area.

In conclusion, our frPanx1-ΔLC structure provides an important atomic blueprint for dissecting functional mechanisms of Panx1. Although we did not observe a gate-like structure in the current cryo-EM map, the missing domains, especially the amino-terminal loop and the carboxy-terminal domain, may serve as a channel gate on the intracellular side of the channel. Further structure-based experiments such as cysteine accessibility and molecular dynamics simulations will facilitate our understanding of how this unique large-pore channel functions.

PUBLICATIONS

Furukawa H, Simorowski N, Michalski K. 2021. Effective production of oligomeric membrane proteins by EarlyBac-insect cell system. *Methods Enzymol* **653:** 3–19. doi:10.1016/bs.mie.2020.12.019

Hansen KB, Wollmuth LP, Bowie D, Furukawa H, Menniti FS, Sobolevsky AI, Swanson GT, Swanger SA, Greger IH, Nakagawa T, et al. 2021. Structure, function, and pharmacology of glutamate receptor ion channels. *Pharmacol Rev* **73:** 298–487. doi:10.1124/pharmrev.120.000131.PMID: 34753794

Syrjanen J, Michalski K, Kawate T, Furukawa H. 2021. On the molecular nature of large-pore channels. *J Mol Biol* **433:** 166994. doi:10.1016/j.jmb.2021.166994

In Press

Tajima N, Simorowski N, Yovanno RA, Regan MC, Michalski K, Gómez R, Lau AY, Furukawa H. 2022. Development and characterization of functional antibodies targeting NMDA receptors. *Nat Commun* **13:** 923.

GENETIC DISSECTION OF CORTICAL CIRCUITS UNDERLYING COGNITIVE PROCESSING

J. Huang X. An, S. W. Galbavy, J. Hatfield, D. Huilgol, Yi. Li, K. Matho, H. Mohan, Y. Qian, R. Raudales, S. Srivas, B. Wang

The overarching goal of the Huang laboratory is to apply innovative and multidisciplinary approaches to study the development and function of neural circuits of the cerebral cortex that enable motor control and cognitive processing. Our starting point is building molecular genetic tools to systematically identify and manipulate a diverse array of specific neuronal cell types (i.e., basic circuit elements) in the mouse brain; our recent invention of an RNA-programmable cell targeting and editing method pioneers a new generation of facile and scalable cell type technology, generalizable across animal species, including humans. To discover the developmental and molecular genetic basis of cortical cell types, we carry out state-of-the-art genetic fate mapping to track their developmental trajectories from progenitor origin and lineage progression to input–output connectivity in cortical circuits. We then apply cell type–targeted single-cell transcriptomics and epigenomics to reveal gene expression profiles and genetic programs underlying neuron type identities. To explore the cell type basis of neural circuit function, we start by designing a variety of (head-constrained as well as free-moving naturalistic) behavioral paradigms to probe the mouse's motor and cognitive capacities that engage cortical circuits. Guided by insights from behavioral analysis, we apply cell type genetic tools that integrate in vivo imaging, electrophysiology, optogenetic, and anatomic circuit tracing to discover the specific roles of distinct neuron types in cortical circuit operation. By leveraging experimentally accessible neuron types as a solid "middle ground" to navigate across levels and scales of brain organization, our research paradigm coherently integrates multidisciplinary approaches toward discovering the general principles of cerebral cortex architecture and function, with major implications in understanding and treatment of a variety of neuropsychiatric disorders.

Genetic Tools and Technologies for Cell Type Targeting and Editing

A key obstacle to understanding brain circuits is the diversity of neuronal cell types. Over the past two decades, we have pioneered a genetic dissection of cortical circuits through systematic targeting of neuronal cell types in the mouse. Leveraging molecular markers and developmental programs, we use gene targeting in embryonic stem cells to generate more than 60 driver and reporter mouse lines targeting major types and lineages of GABAergic inhibitory interneurons and glutamatergic excitatory projection neurons. We further design sophisticated combinatorial strategies that incorporate viral vectors to target increasingly specific subtypes. These genetic tools have transformed the study of cortical circuits by enabling experimental access and integration of a full range of modern techniques to an increasingly comprehensive set of neuron types. Dr. Huang has been directing a Comprehensive Center for Mouse Brain Cell Atlas as part of the U.S. BRAIN Initiative Cell Census Network (BICCN) since 2017. A major goal of our Center is to build genetic tools for projection neuron types across the cortico–striatal–thalamic system. These tools will greatly empower a large community of neuroscientists to study the cortico–striatal–thalamic system—prominent brain networks that mediate a wide range of sensory, motor, emotional, and cognitive functions.

Beyond germline genetic engineering in a few model organisms (e.g., mouse, *Drosophila*), advances in biomedical research require cell type technologies that are easy, facile, scalable, and generalizable across species including human. In the past year, we have recently invented a new class of cell type technology that leverages cell type–defining RNAs. CellREADR, cell access through RNA sensing by endogenous ADAR (adenosine deaminase acting on RNA), harnesses an

RNA sensing and editing mechanism ubiquitous to all animal cells for detecting the presence of specific cellular RNAs and switching on the translation of effector proteins to monitor and manipulate the cell type. Deployed as a single RNA molecule operating through Watson–Crick base pairing, CellREADR is inherently specific, easy, scalable, programmable, and general. As RNAs are the universal messengers of gene expression that underlie cell identities and states across organ systems and species, CellREADR will have a wide impact across biomedical research. We are devoting major efforts to the further innovation and optimization of CellREADR for cell type editing across organ systems and species including human tissues. With its capacity to sense and program cellular physiology, we will harness CellREADR for cell type precision diagnostics and therapeutics.

Genetic Programs and Developmental Trajectories of Neuronal Cell Types

Despite their immense complexity, a unique feature of brain networks, in contrast to engineered computational networks, is that they self-assemble during the course of development. Perhaps the most complex region of the nervous system, the cerebral cortex consists of dozens of cortical areas, each comprising multiple layers and diverse neuron types, that form intricate local circuits that are integrated across global networks. A fundamental problem in neuroscience is how genome information, shaped by evolutionary processes, directs the developmental construction of a set of stereotyped and species-typic network scaffolds and circuit templates, which are then customized through individual life experiences. Unraveling the cellular and molecular basis of cortical circuit assembly has major implications in understanding and treating brain disorders from autism to intellectual disability and schizophrenia.

Our basic tenet is that, across its nested levels of organization, the cortex is fundamentally built from a large set of cardinal neuron types, which assemble (1) local connectivity according to a "canonical microcircuit template" duplicated and modified across areas, (2) species-typic intracortical processing streams, and (3) subcortical-directed output channels. These cardinal neuron types are reliably generated through developmental programs embedded in neural progenitors and rooted in the genome. We focus our study on two key aspects of cortical development: the lineage origin and molecular genetic basis of neuronal cell types—basic building blocks of circuit assembly. Our overarching strategy is a systematic genetic dissection of neuron types, as well as fate mapping their developmental trajectories from neural progenitors. Using driver lines and combinatorial strategies targeting progenitor subpopulations, we have discovered that fate-restricted radial glial progenitors are defined by combinatorial transcription factor (TF) expression, and lineage progression and cell birth order contribute to the cardinal identity of major glutamatergic (GLU) projection neurons (PN) and GABA interneurons. For example, we are able to track the developmental trajectory of chandelier cells, a powerful inhibitory interneuron type that control PN spiking at the axon initial segment, from lineage origin to circuit connectivity. Using a series of custom-built driver lines targeting TF-defined progenitors, we are systematically fate-mapping major PN types that constitute intracortical processing streams (intratelencephalic; IT) and output channels (pyramidal track; PT). These studies will provide a road map for tracking the assembly of cortical circuits at the resolution of cell type building blocks.

In the past year, we have completed a major study demonstrating that direct and indirect neurogenesis generate a mosaic of distinct glutamatergic projection neuron types and cortical subnetworks. Variations in size and complexity of the cerebral cortex result from differences in neuron number and composition, which are rooted in evolutionary changes in direct and indirect neurogenesis (dNG and iNG) mediated by radial glial progenitors and intermediate progenitors, respectively. How dNG and iNG differentially contribute to cortical neuronal number, diversity, and connectivity is unknown. Establishing a genetic fate-mapping method to differentially visualize dNG and iNG in mice, we found that although both dNG and iNG contribute to all cortical structures, iNG contributes the largest relative proportions to the hippocampus and neocortex compared to insular and piriform cortex, claustrum, and the pallial amygdala. Within the neocortex, whereas dNG generates all major glutamatergic PN classes, iNG differentially amplifies and diversifies PNs within each class; the two neurogenic pathways generate distinct PN types and assemble fine mosaics of lineage-based cortical subnetworks. Our results provide

a ground-level lineage framework of cortical development and evolution by linking foundational progenitor types and neurogenic pathways to PN types.

Cortical Circuit Basis of Motor Control and Cognitive Processing

The neocortex comprises a constellation of functional areas that form a representation map of the external and internal world. These areas are strategically connected as processing networks to integrate multisensory information with internal goals, make decisions, and configure descending instructions that guide intelligent behaviors. Across cortical areas, the basic circuit architecture is a cortico–striatal–thalamic loop, which is duplicated and topographically organized along multiple functional systems according to the cortical representation map. We use forelimb motor control as a paradigm to explore the general principles of cortical circuit functional organization.

Among brain functions ranging from perception to cognition and action, the generation of complex movements enables animals to navigate and impact the world. In particular, the evolutionary elaboration of increasingly sophisticated forelimb behaviors, such as to reach, grasp, handle, and manipulate objects, ultimately leads to uniquely human abilities. Previous studies of forelimb movements have largely been carried out in humans and nonhuman primates, but the neural circuit mechanisms remain difficult to explore. Rodents such as mice are capable of complex forelimb and orofacial movements, driven by the necessity of their omnivorous feeding on diverse food sources that require dexterous hand–mouth maneuvering. Thus, mice present a powerful experimental system in which our cell type genetic tools can integrate the full range of modern technologies for exploring cortical circuits underlying motor control.

We have established several mouse forelimb behavior paradigms that engage cortical function. In head-constrained setups, mice are trained to reach for a water drop at multiple spatial locations or to retrieve and eat food pellets. In a freely behaving setup, mice use highly coordinated and dexterous oro-manual movements to retrieve, handle, manipulate, and eat various food items. We apply machine learning–based quantitative analysis to delineate ethograms of movement components and their organization. Guided by insights from behavior analysis, we then leverage our multiple genetic drive lines to carry out systematic screens that identify cortical areas and cell types involved in various aspects of sensorimotor control. We use wide-field mesoscopic imaging of genetically encoded calcium indicator (GCaMP6f) across dorsal cortex to monitor the spatiotemporal activity patterns of molecular- and projection-defined neuron types; we use optogenetic activation and inhibition to probe their roles in different cortical areas during sensorimotor behaviors. Identification of key cortical areas and cell types allows us to apply cellular-resolution two-photon imaging, optogenetic-guided electrophysiological recording, and input–output viral tracing to explore neural circuit operations.

In the past year, we have completed a study demonstrating that cortical glutamatergic projection neuron types contribute to distinct functional subnetworks. The cellular basis of cerebral cortex functional architecture remains not well understood. A major challenge is to monitor and decipher neural network dynamics across broad cortical areas yet with PN type resolution in real time during behavior. Combining genetic targeting and wide-field imaging, we monitored activity dynamics of subcortical-projecting (PTFezf2) and intratelencephalic-projecting (ITPlxnD1) types across dorsal cortex of mice during multiple brain states and behaviors. ITPlxnD1 and PTFezf2 showed distinct activation patterns during wakeful resting, spontaneous movements, and upon sensory stimulation. Distinct ITPlxnD1 and PTFezf2 subnetworks are dynamically tuned to different sensorimotor components of a naturalistic feeding behavior, and optogenetic inhibition of subnetwork nodes disrupted specific components of this behavior. ITPlxnD1 and PTFezf2 projection patterns supported their subnetwork activation patterns. Our results suggest that, in addition to the concept of columnar organization, dynamic areal and PN type–specific subnetworks are key features of cortical functional architecture linking microcircuit components to global brain networks.

PUBLICATIONS

BRAIN Initiative Cell Census Network (BICCN). 2021. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598:** 86–102. doi:10.1038/s41586-021-03950-0

Deng H, Xiao X, Yang T, Ritola K, Hantman A, Li Y, Huang ZJ, Li B. 2021. A genetically defined insula-brainstem circuit

selectively controls motivational vigor. *Cell* **184:** 6344–6360.e18. doi:10.1016/j.cell.2021.11.019

Dudok B, Szoboszlay M, Paul A, Klein PM, Liao Z, Hwaun E, Szabo GG, Geiller T, Vancura B, Wang BS, et al. 2021. Recruitment and inhibitory action of hippocampal axo-axonic cells during behavior. *Neuron* **109:** 3838–3850.e8. doi:10.1016/j.neuron.2021.09.033

Matho KS, Huilgol KD, Galbavy W, Kim G, He M, An X, Lu J, Wu P, Di Bella DJ, Shetty AS, et al. 2021. Genetic dissection of the glutamatergic neuron system in cerebral cortex. *Nature* **598:** 182–187. doi:10.1038/s41586-021-03955-9

Muñoz-Castañeda R, Zingg B, Matho KS, Chen X, Wang Q, Foster NN, Li A, Narasimhan A, Hirokawa KE, Huo B, et al. 2021. Cellular anatomy of the mouse primary motor cortex. *Nature* **598:** 159–166. doi:10.1038/s41586-021-03970-w

Wang BS, Bernardez Sarria MS, An X, He M, Alam NM, Prusky GT, Crair MC, Huang ZJ. 2021. Retinal and callosal activity–dependent chandelier cell elimination shapes binocularity in primary visual cortex. *Neuron* **109:** 502–515.e7. doi:10.1016/j.neuron.2020.11.004

THEORETICAL NEUROSCIENCE AND ARTIFICIAL INTELLIGENCE

A.A. Koulakov B. Baserdem S.A. Shuvaev
K. Samoylova N. Tran

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 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 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 for 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 (i.e., 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.

Neural Networks with Motivation

S.A. Shuvaev, N.B. Tran, A.A. Koulakov [in collaboration with M. Stephenson-Jones and B. Li, CSHL]

Animals rely on internal motivational states to make decisions. The overall role of motivational salience in decision-making is in early stages of mathematical understanding. Here, we propose a reinforcement learning framework that relies on neural networks to learn optimal ongoing behavior for dynamically changing motivation vectors. First, we show that neural networks implementing Q-learning with motivational salience can navigate in environment with dynamic rewards without adjustments in synaptic strengths when the needs of an agent shift. In this setting, our networks display elements of addictive behaviors. Second, we use a similar framework in a hierarchical manager–agent system to implement a reinforcement learning algorithm with motivation that both infers motivational states and behaves. Finally, we show that, when trained in the Pavlovian conditioning setting, the responses of the neurons in our model resemble previously published neuronal recordings in the ventral pallidum, a basal ganglia structure involved in motivated behaviors. We conclude that motivation allows Q-learning networks to quickly adapt their behavior to conditions when expected reward is modulated by their dynamic needs. Overall, our approach addresses the algorithmic rationale of motivation and takes a step toward better interpretability of behavioral data via inference of motivational dynamics in the brain.

Wiring Logic of the Early Rodent Olfactory System Revealed by High-Throughput Sequencing of Single-Neuron Projections

B. Baserdem and A.A. Koulakov [in collaboration with Y. Chen, X. Chen, H. Zhan, Y. Li, M.B. Davis, J.M. Kebschull, A.M. Zador, and D.F. Albeanu, CSHL]

The structure of neuronal connectivity often provides insights into the relevant stimulus features, such as spatial location, orientation, sound frequency, etc. The olfactory system, however, appears to lack structured connectivity as suggested by reports of broad and distributed connections both from the olfactory bulb to the pyriform cortex and within the cortex. These studies have inspired computational models of circuit function that rely on random connectivity. It remains, nonetheless, unclear whether the olfactory connectivity contains spatial structure. Here, we use high-throughput anatomical methods (MAPseq and BARseq) to analyze the projections of 5,309 bulb and 30,433 pyriform cortex output neurons in the mouse at single-cell resolution. We identify previously unrecognized spatial organization in connectivity along the anterior–posterior axis (A-P) of the pyriform cortex. We find that both the bulb projections to the cortex and the cortical outputs are not random, but rather form gradients along the A-P axis. Strikingly, these gradients are matched: bulb neurons targeting a given location within the pyriform cortex co-innervate extra-pyriform regions that receive strong inputs from neurons within that pyriform locus. We also identify signatures of local connectivity in the pyriform cortex. Our findings suggest an organizing principle of matched direct and indirect olfactory pathways that innervate extra-pyriform targets in a coordinated manner, thus supporting models of information processing that rely on structured connectivity within the olfactory system.

Encoding Innate Ability through a Genomic Bottleneck

A. Koulakov, S.A. Shuvaev [in collaboration with D. Lachi and A.M. Zador, CSHL]

Animals are born with extensive innate behavioral capabilities, which arise from neural circuits encoded in the genome. However, the information capacity of the genome is orders of magnitude smaller than that needed to specify the connectivity of an arbitrary brain circuit, indicating that the rules encoding circuit formation must fit through a "genomic bottleneck" as they pass from one generation to the next. Here we formulate the problem of innate behavioral capacity in the context of artificial neural networks in terms of lossy compression of the weight matrix (Fig. 1). We find that several standard network architectures can be compressed by several orders of magnitude, yielding pretraining performance that can approach that of the fully trained network. Interestingly, for complex but not for simple test problems, the genomic bottleneck algorithm also captures essential features of the circuit, leading to enhanced transfer learning to novel tasks and data sets. Our results suggest that compressing a neural circuit through the genomic bottleneck serves as a regularizer, enabling evolution to select simple circuits that can be readily adapted to important real-world tasks. The genomic bottleneck also suggests how innate priors can complement conventional approaches to learning in designing algorithms for artificial intelligence.

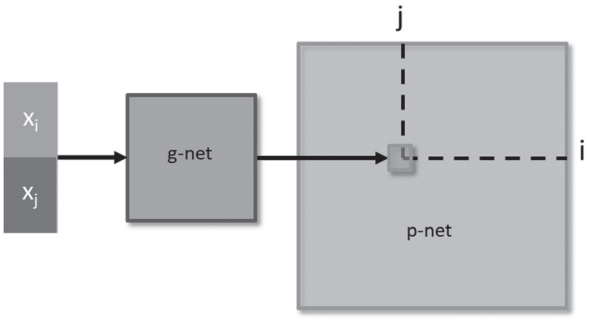


Figure 1. Compression of cortical connectivity through a genomic bottleneck. The input to the genomic network ("g-network") is a pair of sets of molecular labels defining two connecting neurons. The g-network represents connectivity rules specified in the genome, whereas the "phenotypic" p-network represents the final cortical connectome.

PUBLICATIONS

Chen Y, Chen X, Baserdem B, Zhan H, Li Y, Davis MB, Kebschull JM, Zador AM, Koulakov AA, Albeanu DF. 2021. Wiring logic of the early rodent olfactory system revealed by high-throughput sequencing of single neuron projections. bioRxiv doi: 0.1101/2021.05.12.443929

Koulakov AA, Shuvaev SA, Zador AM. 2021. Encoding innate ability through a genomic bottleneck. bioRxiv doi:10.1101/2021.03.16.435261

Shuvaev SA, Tran NB, Stephenson-Jones M, Li B, Koulakov AA. 2021. Neural networks with motivation. *Front Syst Neurosci* **14:** 609316. doi:10.3389/fnsys.2020.609316

THE FUNCTION AND PLASTICITY OF NEURAL CIRCUITS IN THE BRAIN IN MOTIVATED BEHAVIORS RELATED TO PSYCHIATRIC AND METABOLIC DISORDERS

B. Li S. Boyle, A. Furlan, W. Guan, A. Kaplan, L. Kimoto, L. Ramirez Sanchez, R. Sharma, Q. Sun, D. van de Lisdonk, T. Yang, K. Yu

The focus of research in the Li laboratory has been to understand the link between neural circuits and behaviors. We are particularly interested in studying the synaptic and circuit mechanisms underlying motivated behaviors and brain–body interactions, as well as synaptic and circuit dysfunctions that may underlie the pathophysiology of mental disorders and metabolic disorders, including anxiety disorders, depression, autism, drug addiction, obesity, and cancer-associated cachexia. 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 behaviors and metabolic processes. We have made the following major progress in 2021.

Genetic Identification of Amygdala–Striatal Circuits for Valence-Specific Behaviors

The basolateral amygdala (BLA) plays essential roles in behaviors motivated by stimuli with either positive or negative valence, but how it processes motivationally opposing information and participates in establishing valence-specific behaviors remains unclear. In Zhang et al. (2021), by targeting *Fezf2*-expressing neurons in the BLA, we identify and characterize two functionally distinct classes in behaving mice, the negative-valence neurons and the positive-valence neurons, which innately represent aversive and rewarding stimuli, respectively, and through learning acquire predictive responses that are essential for 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, 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 valence-specific motivated behaviors.

A Genetically Defined Insula-Brainstem Circuit Selectively Controls Motivational Vigor

The anterior insular cortex (aIC) plays a critical role in cognitive and motivational control of behavior, but the underlying neural mechanism remains elusive. In Deng et al. (2021) we show that aIC neurons expressing *Fezf2* (aIC^{Fezf2}), which are the pyramidal tract neurons, signal motivational vigor and invigorate need-seeking behavior through projections to the brainstem nucleus tractus solitarii (NTS). aIC^{Fezf2} neurons and their postsynaptic NTS neurons acquire anticipatory activity through learning, which encodes the perceived value and the vigor of actions to pursue homeostatic needs. Correspondingly, aIC → NTS circuit activity controls vigor, effort, and striatal dopamine release, but only if the action is learned and the outcome is needed. Notably, aIC^{Fezf2} neurons do not represent taste or valence. Moreover, aIC → NTS activity neither drives reinforcement nor influences total consumption. These results pinpoint specific functions of aIC → NTS circuit for selectively controlling motivational vigor and suggest that motivation is subserved, in part, by aIC's top-down regulation of dopamine signaling.

Valence- and Experience-Dependent Salience Signals in the Central Amygdala

The central amygdala (CeA) is involved in a range of mental processes, including attention, learning, memory formation, and extinction, and is known to have an important role in behaviors motivated by either aversive or appetitive stimuli. However, how the CeA participates in these divergent brain functions and contributes to both aversive and appetitive behaviors remains elusive. In Yang et al. (2022) we show that a major CeA population—the somatostatin-expressing (Sst^+) neurons—contains two distinct functional types that are selectively and predominantly excited by either innately aversive or innately appetitive stimuli (i.e., aversive or appetitive unconditioned stimuli [USs], respectively). In vivo imaging and optogenetics in behaving mice demonstrate that the valence-specific US-evoked responses in these neurons undergo pronounced strengthening through experience, encode the salience of the stimuli, and are required for aversive or reward learning. Notably, Sst^+ CeA neurons are capable of driving activation of dopamine neurons in the substantia nigra pars compacta (SNc), and the US responses of the projections from Sst^+ CeA neurons to the SNc are required for reward learning but not aversive learning. Thus, our results uncover valence-specific and experience-dependent salience signals in the CeA, which can control learning through, at least in part, the CeA → SNc projections regulating dopamine neurons.

Neurotensin Neurons in the Extended Amygdala Control Dietary Choice and Energy Homeostasis

Obesity is a global pandemic that is causally linked to many life-threatening diseases. Apart from some rare genetic conditions, the biological drivers of overeating and reduced activity are unclear. In Furlan et al. (2022) we show that neurotensin-expressing neurons in the mouse IPAC, a nucleus of the central extended amygdala, encode dietary preference for unhealthy energy-dense foods. Optogenetic activation of neurotensin-expressing neurons in the IPAC ($IPAC^{Nts}$ neurons) promotes obesogenic behaviors, such as hedonic eating and reduced locomotion. Conversely, inhibition of $IPAC^{Nts}$ neurons abolishes hedonic feeding and enhances locomotion. As a result, these leptogenic behaviors induce long-term weight loss, improve metabolic health, and prevent obesity. Thus, the activity of a single neuronal population bidirectionally regulates energy homeostasis. Our findings could lead to new therapeutic strategies to prevent and treat obesity.

PUBLICATIONS

Deng H, Xiao X, Yang T, Ritola K, Hantman A, Li Y, Huang ZJ, Li B. 2021. A genetically defined insula-brainstem circuit selectively controls motivational vigor. *Cell* **184:** 6344–6360.

Dong A, He K, Dudok B, Farrell JS, Guan W, Liput DJ, Puhl HL, Cai R, Wang H, Duan J, et al. 2021. A fluorescent sensor for spatiotemporally resolved imaging of endocannabinoid dynamics in vivo. *Nat Biotechnol* doi:10.1038/s41587-021-01074-4

Furlan A, Corona A, Boyle S, Sharma R, Rubino R, Habel J, Gablenz EC, Giovanniello J, Beyaz S, Janowitz T, et al. 2021. Neurotensin neurons in the extended amygdala control energy balance. bioRxiv doi:10.1101/2021.08.03.454970

Gschwend O, Yang T, van de Lisdonk D, Zhang X, Sharma R, Li B. 2021. Prefrontal top-down projections control context-dependent strategy selection. bioRxiv doi:10.1101/2021.12.14.472559v1

Zhang X, Guan W, Yang T, Furlan A, Xiao X, Yu K, An X, Galbavy W, Ramakrishnan C, Deisseroth K, et al. 2021. Genetically defined amygdala–striatal circuits for valence-specific behaviors. *Nat Neurosci* **24:** 1586–1600.

In Press

Yang T, Yu K, Zhang X, Xiao X, Li B. 2022. Valence- and experience-dependent salience signals in the central amygdala. *Nature* (in revision).

THE STUDY OF INTELLIGENT MACHINES

P.P. Mitra S. Abbu, S. Banerjee, P. Flannery, C. Hsu, B. Huo, L. Kimoto, X. Li, B. Lodato, L. Lodato, C. Mezias, J. O'Rourke, S. Savoia, F. Xu

BRAIN-WIDE MESOSCALE CIRCUIT MAPPING IN RODENTS AND PRIMATES

The core research thrust in the Mitra laboratory ("our laboratory") is to uncover the principles of brain architecture by mapping brain circuitry at the mesoscopic scale, using tracer injections, across vertebrate species. Our laboratory proposed and pioneered brain-wide mesoscale circuit mapping in vertebrates as an approach to understanding how brains work. We run an integrated neurohistology processing pipeline that spans the wet lab, computational analyses, and web dissemination and that has been continuously collecting and analyzing data sets across species, with a focus on the laboratory mouse, since 2010. The experimental part of the pipeline was successfully replicated in RIKEN, Japan, for mapping the marmoset brain mesoscale circuits. The computational part of the pipeline has been processing data across species, from both CSHL and RIKEN. More recently, our laboratory has been analyzing cell type–specific projections in the mouse through BRAIN Initiative Cell Census Network (BICCN) as well as in the marmoset through a collaboration with MIT.

Although the laboratory mouse is the most widely used model organism in neuroscience, there has been increasing concern that taxonomic differences may hinder the translation of findings in the mouse to primates, particularly to humans. This may be one of the obstacles to understanding the bases of major human neurological disorders and developing suitable therapies. Thus, it is of great interest to understand the similarities and differences between the brain circuits in rodents and primates. As a result of the unprecedented joint data set collected in our laboratory using uniform methodology (tape-transfer-assisted cryosectioning) across species, and the scalable computational pipeline, we are currently able to carry out an unprecedented study of comparative neuroanatomy, analyzing brain-wide mesoscale circuitry in the mouse and the marmoset—a model primate that is being increasingly adapted for biomedical research. Previously, we established a framework for reconciling different atlas hierarchies, which enabled cross-species comparisons by identifying tentative homologous regions. With only part of the marmoset brain injected with tracers (both anterograde and retrograde), we are carrying out this comparative study in selected brain subsystems, including the medial prefrontal cortex and the tectum, in collaboration with neuroanatomical experts who work on the relevant subsystems.

A Comparative Study of Tectal Projections in Marmoset and Mouse

This work was done in collaboration with T. Isa (Kyoto University).

The superior colliculus (SC) is an evolutionary conserved structure across mammalian brains. It is multisensory by nature and is the center for quick transduction of sensory inputs to behavioral responses such as orienting and defense. By comparing the mesoscale connectivity data in mouse and marmoset, we observed similarities in the connectivity patterns between SC and many cortical regions, including the (pre)frontal, visual, auditory, parietal, and temporal cortices in both species. Yet similar subregions of SC are likely to have evolved to have different connections with other brain regions, especially cortex, to serve species-specific functions. In rodents, the SC is important for active orientation and fear-related responses, whereas in primates, the SC also has cognitive roles such as attention and decision-making. Apparent

differences were observed in connection strengths and cortical lateralization. For example, retrograde injections in marmoset SC produce extensive, bilateral dorsal–ventral label in coronal sections at the level of proposed frontal eye fields (FEFs). In contrast, retrograde injections in mouse SC produce more localized and lateralized label at similar prefrontal coronal levels. This result raises doubts about previous homological assignments between rodent frontal cortex and primate prefrontal subregions such as FEFs. This study is still in progress and is expected to reach maturity in the next year or so.

A Comparative Study of Medial Prefrontal Cortical Projections in Marmoset and Mouse

This work was done in collaboration with A. Roberts (Cambridge University) and M. Fatih Yanik (ETH, Zurich).

The medial prefrontal cortex (mPFC) is involved in various neurological disorders such as anxiety disorders. In contrast with the SC, the PFC has undergone evolutionary expansions in the primate and therefore has less homology between the rodent and the primate. In previous research, discrepancies have been found between the cytoarchitecture-based homologous regions and the function-based homology. Additionally, the mPFC contains targets for deep brain stimulation–based therapeutic interventions for major depression in humans. This makes the mPFC an interesting target for a comparative connectivity study. Specifically, Brodmann area 25 (A25) in primates was deemed cytoarchitectonically homologous to infralimbic area (IL) in rodents. By analyzing projections into A25 in marmosets and IL in mice, we found marmoset-unique connections in cortical areas such as Brodmann areas 9, 14, and 46, which lack homologous regions in the mouse brain. Conversely, we also found mouse-unique connections in claustrum and piriform cortex. Within the homologous areas in both species projecting into A25/IL, the connectivity strength varies. For example, the paratenial nucleus of thalamus sends strong projections to A25 in marmoset but only weakly to IL in mouse. Similarly, we found qualitative and quantitative differences in projections into Area 24 in the marmoset, cytoarchitectonically homologous to anterior cingulate cortex in the mouse. These differences therefore raise cautionary notes about using mice as models of neurological disorders that involve mPFC.

Cell Type–Specific Mapping in the Marmoset Brain

This work was done in collaboration with G. Feng (MIT/Broad Institute).

The advent of high-throughput genetic screens and sequencing technologies has revolutionized neuroscience. These technologies have helped identify candidate genetic determinants underlying neurological function/disease and have led to the development of tools to enable cell type–specific interrogation of neurobiological processes in model organisms. The common marmoset has recently emerged as a new promising model for developing genetic tools as well as genetic engineering. In a new collaboration with Guoping Feng and Ian Wickersham at MIT and Steve McCaroll at Harvard, we are using a novel high-throughput approach that combines multiplex barcoding with single-nucleus RNA sequencing (snRNA-seq) readouts to test hundreds of putative enhancer sequences simultaneously within individual marmoset brains, identify cell type–specific functional enhancers, and further validate these enhancers by green fluorescent protein (GFP) expression and traditional histochemical methods. We are also establishing experimental methods and computational pipelines for mapping cell type–specific connectivity using existing and newly identified enhancers. Together, these novel tools, methods, and pipelines will greatly expand the use of the marmoset model in neuroscience research by facilitating cell type–specific circuit mapping and functional perturbation.

THE VIRTUAL NEUROANATOMIST

Our laboratory is fairly unique in combining experimental, computational, and theoretical work closely within the same research laboratory. Given the large data sizes, automated methods for analyzing data, including AI methods, are central to our research. We conceptualize this work as the virtualization of neuroanatomical expertise. The computational/informatics arm of our laboratory performs management, analysis, and web dissemination of the data gathered in the experimental arm of our 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 project as the development of a "Virtual Neuroanatomist." In addition, our laboratory maintains an advanced web portal for visualizing large-scale histology images that is unique in presenting whole-brain light-microscopic level data across multiple species.

Multimodal, Robust Brain to Atlas Registration

This work was done in collaboration with D. Tward (UCLA) and M. Miller (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, analogous in importance to mapping genomic data to a reference genome. We have previously developed a rigorous, nonparametric generative framework that registers multimodality data that learns unknown mappings between contrast mechanisms from data and infers missing data. The quantitative, scalable, and streamlined workflow 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. In a continuing effort of improving the performance of the pipeline, a shortest-path graph-based solution for transforming imaging data between any two members of a multimodal data set and standardization of data format have been developed. The method has especially assisted the development of a new reference brain in which transformation across different modalities of the same brain and registration across individuals, are essential.

Skeletonization of Neuronal Morphology Using Topological Methods

This work was done in collaboration with Y. Wang (UCSD).

We previously developed an automated method for extracting single-neuron morphology from sparsely labeled brains, as well as a method to summarize the projections arising from anterograde tracer injections using discrete Morse theory. In the last year we worked on making this approach robust and scalable. This work necessitates online proofreading of terabyte-sized data sets. We worked on adopting the Janelia workbench in collaboration with scientists at Janelia who set up such an online proofreading tool for brains from the MouseLight project. We are implementing this methodology to analyze tracer-injected data sets from the Brain Architecture project as well as fluorescence micro-optical sectioning tomography (fMOST) data sets from BICCN for tracing single-neuron morphology brain-wide at scale.

Informatics/Web Portal

To visualize all analyses' results together with the corresponding histology samples, the Brain Architecture data portal (www.brainarchitecture.org) continued to serve as the public-facing data visualization portal for high-resolution, multispecies brain architecture data.

BRAIN INITIATIVE CONSORTIA

This work was done in collaboration with neuroscientists at CSHL (J. Huang, P. Osten) as well as other Brain Initiative collaborators (M. Miller [Johns Hopkins University], D. Tward [UCLA], E. Macosko and T. Tickle [MIT], D. Kleinfeld [UCSD], A. Koretsky and L. Latour [NINDS], and BICCN consortium researchers).

The Mitra laboratory has contributed to several BRAIN Initiative collaborative efforts. In the Cell Census Network consortium, our laboratory has been leading the Data Core for the CSHL/Harvard BICCN U19 grant, providing data governance, analytics, and dissemination, as well as participating in the anatomy development of a new mouse brain reference. In the UCSD U19 project for studying rodent brainstem circuitry, our laboratory plays a major role in developing advanced histology techniques for mesoscale neuroanatomy.

Data Core for BRAIN Initiative Cell Census Network (BICCN)

The Mitra laboratory leads the Data Core in the CSHL/Harvard U19 project in BICCN. Our laboratory performed data ownership, stewardship, and management for cell type–specific projection, spatial distribution, and single-neuron morphology data analysis of different imaging modalities from participating laboratories. We continued to operate the data

processing pipeline for serial two-photon tomography (STPT) data and developed a similar pipeline for fMOST data, including image compression, quality control, and NIH-funded data archive submission, high-performance atlas registration, and high-resolution image visualization on the web portal.

New Mouse Reference Brain Atlas (Nissl CCF V4)

This work was done in collaboration with A. Koretsky and L. Latour (NINDS).

Reference brain images and atlases are of fundamental importance to neuroscience, as annotated reference genomes are to cellular and molecular biology. They provide the foundation that holds together information and knowledge across the full spectrum ranging from neuroscience research programs to clinical practice—and indeed provide the basic conceptual framework for our understanding of brains. Surprisingly, even for the laboratory mouse, by far the most widely used laboratory organism, the latest digital reference atlas (Common Coordinate Framework [CCF] Version 3 from the Allen Institute) has major deficits (e.g., it lacks cytoarchitectonic information or accompanying in vivo MRI data). To address these issues, our laboratory has taken on the task of creating a next-generation Nissl-based reference atlas (Nissl CCF V4) of the mouse brain as a part of the anatomy efforts in BICCN. We collected pilot data for this project to demonstrate feasibility and have now started processing the brains to be used for the atlas, including pre- and postmortem MRI, CT, and Nissl/Myelin alternating histology series.

Histology Core for UCSD Kleinfeld U19 project

This work was done in collaboration with David Kleinfeld (UCSD).

Despite the advances in whole-brain imaging techniques utilizing brain clearing and light-sheet-based volumetric microscopy, classical Nissl-stained thin section images remain the most accurate overall method for cytoarchitectural delineation. As the histology core for a BRAIN Initiative U19 project led by David Kleinfeld (UCSD), we set up and productionized an arm of the pipeline utilizing a custom-designed robotic stainer and a water-based coverslipper (as opposed to xylene-based media) for better preservation of fluorescent signal using NeuroTrace as a fluorescent Nissl-like stain. By retaining the brainstem during brain preparation and adding brainstem delineations, we brought our atlas-registration methods and cell/process detection algorithms to bear on brainstem. This ongoing work will complete our understanding of the brainstem neuroanatomy as well as its connectivity with the rest of the brain.

BUILDING A 3D MULTIMODAL MICRON-SCALE HUMAN BRAIN ATLAS

This work was done in collaboration with D. Nauen (Johns Hopkins University), J. Zhang, D. Novikov, E. Fiermans, and T. Shepherd (NYU), M. Sivaprakasam (IIT Madras), P. Hof (Icahn School of Medicine at Mount Sinai), D. Tward (UCLA), S. Pochareddy and N. Sestan (Yale), and G. Simic (U. Zagreb).

Digitized reference brains for humans lag far behind the corresponding CCFs for nonhuman model organisms such as the laboratory mouse. Existing data sets either have sections spaced relatively far apart or lack in-plane resolution down to the micron scale. We will meet this need by creating an unprecedented micron-scale 3D atlas that combines multiple MRI modalities as well as continuous serial section histology. This work is based on the scale-up of our tape transfer–based neurohistological pipeline, which was previously demonstrated in human brain samples with traumatic brain injury. In particular, the reference atlas will consist of Nissl, myelin, and hematoxylin and eosin (H&E) stains, with 20-μ contiguous serial sections. We have completed collecting two samples of one hippocampus and one amygdala of human brain, acquired multimodal MRI images including T1/T2, MTR, and DTI/DKI, and processed through the histology pipeline. A computational framework of reconstructing and calculating maps from these MRI modalities has been developed and will permit co-registration with the multimodal histology images. This project will contribute an important reference atlas that is missing from neuroscience: a multimodal, micron-scale 3D atlas of the human brain that links single-cell transcriptomic and epigenomic data sets to the clinically significant fields of neuroradiology and neuropathology.

In addition, there are no comparable digital histological series and atlases for the developing human brain. We recently started the project of generating and annotating a 4D (space-time) developing human brain reference data set, combining serial section histology (Nissl/myelin/H&E/immunohistochemistry) and microstructural MRI across 12 developmental stages. This will provide a Human Developmental Common Coordinate Framework (HDCCF) for human brain single-cell data sets. It will simultaneously provide crucial clinical links to neuroradiology and neuropathology and cross-register to existing developmental MRI brain atlases. To demonstrate the feasibility of building the HDCCF, we conducted preliminary analyses on a prenatal hippocampus, including both (1) cell body segmentation and Rakic developmental stage classification from Nissl images and (2) cross-modal registration of MR and Nissl image series.

TOPOLOGICAL DATA ANALYSIS METHODS FOR SINGLE-CELL 'OMICS DATA

This work was done in collaboration with M. Hawrylycz (American Institute of Biological Sciences) and Y. Wang (UCSD).

We have pioneered the usage of a specific method from topological data analysis (TDA), namely discrete Morse (DM) theory, for application to single-neuron reconstruction as well as for the analysis of single-cell 'omics data. Specifically, the framework takes a density function defined on the triangulation of the domain, and outputs the 1-stable manifold, which can be thought of as the mountain ridges, of the density function. Currently, biologists rely on 2D or 3D projections of the single-cell 'omics data (tSNE and UMAP being popular algorithms to do so), which heavily distorts the high-dimensional geometry of the data; and overreliance on clustering approaches may also prevent a full appreciation of the continuous structure of the data. Our goal is to obtain a graph structure, using the topological approach, in the original space of the single-cell 'omics data (roughly 4,000 dimensions) that better represents the nature of the data as compared to low-dimensional projections, by including continuous gradients in the high-dimensional space connecting the density peaks representing cell types. We are currently validating this approach on BICCN mouse brain scRNA-seq data sets and are also scaling the algorithms for large data sets.

VIRTUAL PATHOLOGIST

This work was done in collaboration with J. Crawford and M. Naseem (Northwell) and B. Gallas (FDA).

There are many current efforts, including commercially well-funded ones, to automate anatomic pathology diagnostics using machine-learning approaches. These efforts rely on data-driven deep-net approaches that are sensitive to changes in the input data domain and constitute black boxes that are not adequately transparent for regulatory purposes. We previously embarked on a nontraditional approach to this translational research problem. Our approach is to do behavioral psychophysical studies of practicing pathologists to observe how they carry out their diagnostic process (as opposed to simply dealing with labeled training data sets, as is the case for other approaches). This will allow us to build more transparent AI algorithms better informed by domain knowledge priors. To achieve this goal, we adapted the eeDAP apparatus developed by Brandon Gallas at the FDA. This apparatus allows the monitoring of the specific view of the microscope slide seen by the pathologist, as the pathologist zooms/pans the slide on the microscope stage. In addition, we are collecting input from the pathologist in the form of audio recordings and a questionnaire to be filled out during the diagnostic process.

The project encountered COVID-related delays as it became difficult for Northwell pathologists to come to CSHL for the study. As a result, in 2020 we decided to move the apparatus to Northwell for data collection. However, this triggered an additional IRB requirement from Northwell because of the new performance site. This took some time to obtain, but currently the workstation is at Northwell, the IRB permissions have been obtained, and we are in the process of acquiring the study data from participating pathologists.

THEORETICAL WORK ON MACHINE LEARNING

This work was done in collaboration with C. Sire (Université Paul Sabatier, Toulouse).

Dr. Mitra has been pursuing a specific theoretical thread in modern machine learning—namely, the highly overparameterized network models used for supervised learning are effectively performing data interpolation. This has been represented in a body of work

in the past few years with colleagues Misha Belkin (UCSD), Daniel Hsu (Columbia), and Clement Sire (Université Paul Sabatier, Toulouse). A key new idea is that of statistically consistent interpolation (SCI), in which noisy data is interpolated, but in contradiction to received wisdom from statistics textbooks, still leads to optimal generalization in the sense of statistical consistency, in the large data limit. In 2020 Dr. Mitra worked with Dr. Sire to prove a number of theorems pertaining to an interesting universal interpolating algorithm (the Hilbert Kernel estimator) that shows the SCI property.

Interestingly, this algorithm has no fitting parameters, in contrast to all other supervised-learning methods, and stands in contrast to the many-parameter fits performed by deep nets. In particular, it was possible to compute the convergence rate of the interpolating learner and show that convergence is logarithmic in sample size—in contrast to other estimators. Although statistical consistency was proven earlier, this is the first computation of the convergence rate for this universal SCI learner. Interesting power law behaviors were observed for this estimator, consistent with the scale-free nature of the estimator. The logarithmic convergence and power law behaviors have not been previously observed and hint at a new universality class of SCI learners. In contrast to black-box multiparameter deep nets, the interpolation algorithms are transparent and promise to shed light on the type of overfitted networks used in modern machine learning.

PUBLICATIONS

BRAIN Initiative Cell Census Network (BICCN). 2021. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598:** 86–102.

Matho KS, Huilgol D, Galbavy W, He M, Kim G, An X, Lu J, Wu P, Bella DJD, Shetty AS, et al. 2021. Genetic dissection of the glutamatergic neuron system in cerebral cortex. *Nature* **598:** 182–187.

Mitra PP. 2021. Fitting elephants in modern machine learning by statistically consistent interpolation. *Nat Mach Intell* **3:** 378–386.

Mitra PP, Sire C. 2021. Parameter-free statistically consistent interpolation: dimension-independent convergence rates for Hilbert kernel regression. arXiv 2106.03354.

Muñoz-Castañeda R, Zingg B, Matho KS, Chen X, Wang Q, Foster NN, Li A, Narasimhan A, Hirokawa KE, Huo X, et al. 2021. Cellular anatomy of the mouse primary motor cortex. *Nature* **598:** 159–166.

Sharma A, Jayakumar J, Sankaran N, Mitra PP, Chakraborti S, Kumar PS. 2021. Application of supervised machine learning to extract brain connectivity information from neuroscience research articles. *Interdiscip Sci Comput Life Sci* **13:** 731–750.

QUANTITATIVE ADVANCES FOR THE STUDY OF BRAIN ARCHITECTURE

P. Osten R. Drewes C. Elowsky R. Muñoz-Castañeda A. Narasimhan J. Palmer X. Qi

Technology Advances for a Comprehensive Study of Human Brain Architecture

Over the last eight years the Osten laboratory has been a part of the NIH BRAIN Initiative Cell Census Network (BICCN), a consortium tasked with mapping the cell type anatomy of the mouse and human brain at an unprecedented single-cell resolution across the entire brain structure. In 2021 this work culminated in a BICCN-dedicated *Nature* issue comprising 11 articles that together provide the most comprehensive view of mammalian brain structure to date, with detailed anatomical, physiological, and transcriptomic characterization of cell types in the mouse motor cortex and an initial mainly transcriptomic characterization of cell types in marmoset and human motor cortex. Our laboratory had a senior authorship on two and a first authorship by Rodrigo Muñoz-Castañeda on one of these publications (BRAIN Initiative Cell Census Network 2021; Muñoz-Castañeda et al. 2021), reflecting the leading role we played in building two mouse brain imaging technologies—serial two-photon tomography (STPT) and oblique light-sheet tomography (OLST)—that were used by our laboratory and the Allen Institute for Brain Science to collect the majority of the BICCN anatomical data, as well as in developing new computational methods for collaborative analysis and visualization of the motor cortex data sets comprising several hundred terabytes of data (Fig. 1).

Translating Methods from the Mouse to Human Brain Anatomy

Comprehensive mapping of the whole human brain at a cellular resolution has been a long-standing goal, ever since the drawings of Ramón y Cajal yielded the first glimpse of the immense complexity of brain cell types more than a century ago. Remarkably, even though this goal may have seemed unrealistic just a few years ago, the development of imaging and computational methods for mapping the mouse brain pioneered by our laboratory and others has progressed to the stage at which it is possible to tackle the considerable size of the human brain at a sufficient throughput and reasonable cost to enable data collection from many individuals, with the ultimate goal of determining the stereotyped features of neuronal circuits that are most likely to explain the remarkable capacity of the human brain that separates us from other species, including our closely related nonhuman primates. Our laboratory has recently teamed with Drs. Elizabeth Hillman (Columbia University) and Zhuhao Wu (Icahn School of Medicine at Mount Sinai) to build on and leverage the extensive expertise among us in method development, including the advanced engineering microscopy expertise of the Hillman laboratory, to build a truly synergistic research centered on new human brain–optimized light-sheet HOLiS microscopy that revolutionizes the most critical part of human brain anatomy profiling—the relationship between spatial resolution and acquisition speed—*to achieve sufficient cellular resolution across the entire human brain in only a few days of imaging*. The HOLiS platform is complemented by an optimized tissue preparation method for human brain, named HuB.Clear, and a computational data analysis pipeline for initial aligning of each imaged brain volume within its own 7T MRI scan, deep learning–based data analysis, and generation of a new human brain common coordinate framework with cellular diversity analysis based on multiplex protein profiling and precise spatial characterization. Once completed, the HOLiS platform, designed to be easily scalable and sharable with maximized benefit–cost ratio, will open the door to imaging hundreds or even thousands of human brains in the near future and greatly synergize with other efforts in BICCN

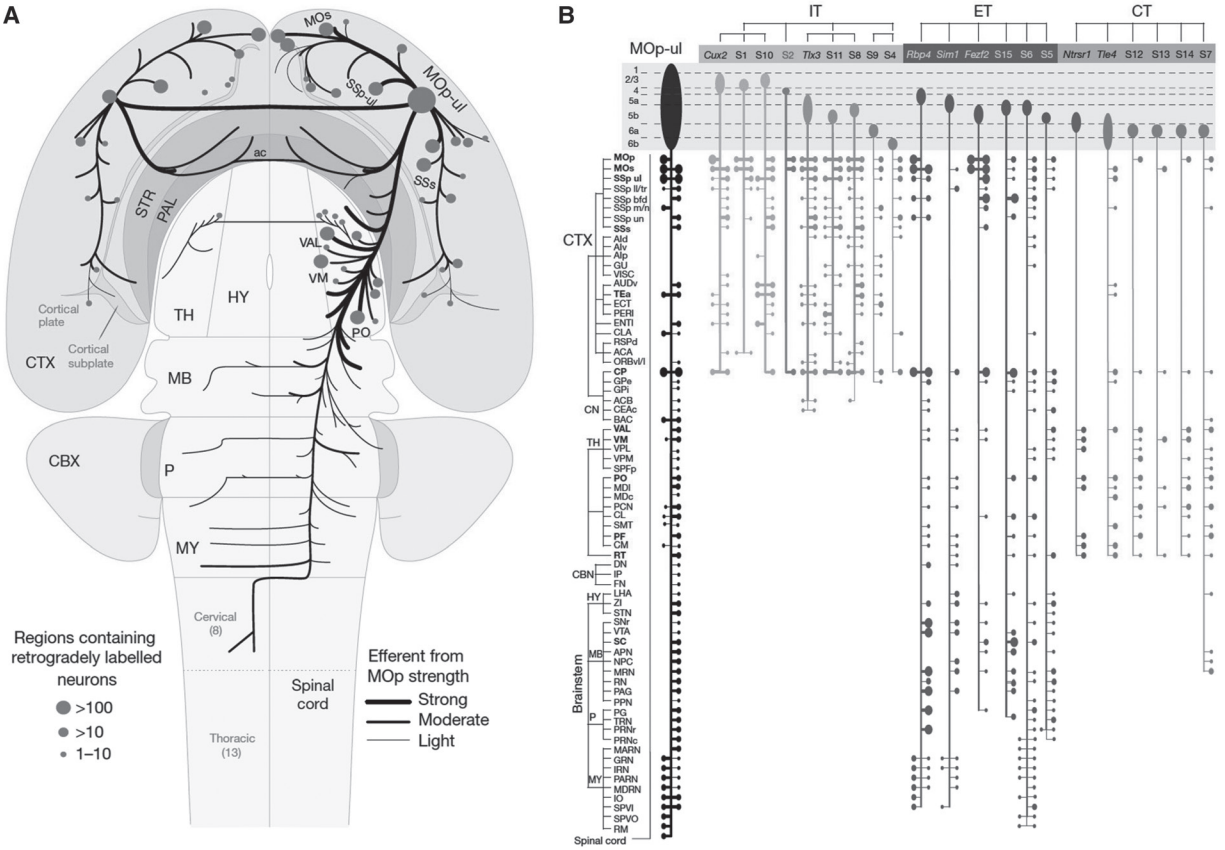


Figure 1. Global wiring diagram and anatomical characterization of motor cortex neuron types. (*A*) Flat map representation of the MOp-ul input–output wiring diagram. (*B*) Projection patterns arising from excitatory cell subclasses IT, ET, and CT, with corresponding Cre line assignment and somatic laminar location, compared with the overall projection pattern from the MOp-ul region (*left*, black). (Adapted from BRAIN Initiative Cell Census Network 2021.)

including large-scale single-cell profiling, delivering on a key premise of BICCN in linking cellular diversity and morphology to molecular signatures across the entire human brain at a sufficient cellular resolution, and facilitating further functional investigations and cross-species comparisons.

PUBLICATIONS

BRAIN Initiative Cell Census Network (BICCN). 2021. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598:** 86–102.

Muñoz-Castañeda R, Zingg B, Matho KS, Chen X, Wang Q, Foster NN, Li A, Narasimhan A, Hirokawa KE, Huo B. 2021. Cellular anatomy of the mouse primary motor cortex. *Nature* **598:** 159–166.

NEURAL CIRCUITRY FOR SOCIAL COMMUNICATION

S. Shea J. Choe R. Dvorkin A. Pagliaro L. Shen
A. Corona 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. Most organisms are endowed with a menu of species-typical behaviors, and they also possess flexible control over when to implement these behaviors and the choice of targets. Together, these elements can be used as building blocks to accelerate the acquisition of sophisticated adaptive behavior. Our group seeks to identify those building blocks and how they are used to devise flexible solutions for real-world problems. Importantly, this approach will ultimately bring us to a greater understanding of the antecedents of our own cognition.

More specifically, we are asking how neural activity and plasticity in olfactory and auditory brain circuits facilitate communication and social behavior. Our work has revealed some of the 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. We hope to understand how that information is combined and interfaced with behavioral decisions. Consequently, most experiments in our laboratory involve observing and manipulating brain activity in real time during ongoing free interactions between mice.

Proper interpretation of social signals is indispensable for survival and mating success. Moreover, 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. We hope to pinpoint and repair neural circuitry defects that impair appropriate use of social information.

We are currently addressing this question through a multilevel analysis of mouse maternal behavior, specifically "pup retrieval." In this behavior, altricial pups call for their mother with ultrasonic distress vocalizations (USVs) when they become separated from the nest. In response, seasoned mothers locate the pup and return it to the nest. Pup retrieval presents several methodological advantages. First, it is overt, quantifiable, and robust. Second, despite its apparent simplicity, we are discovering that the neurochemical events that underlie this behavior are highly rich and complex. Third, pup retrieval crucially depends on integration of multiple social cues, including auditory (USVs) and olfactory (the smell of pups). Fourth, as a social communication behavior, it is useful for probing deficits of social behavior in mouse models of ASDs. Fifth, and perhaps most importantly, mothers and other females cohoused with pups ("surrogates") improve dramatically in their performance with experience. Therefore, pup retrieval is an ethologically fundamental behavior with disease relevance that is discrete and easy to measure; it also has complex determinants and is shaped by an interplay of innate and experiential factors.

Quantitative Psychophysics of USV Perception

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? Our laboratory 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, female mice are taught 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, we "ask it questions" by periodically sneaking in a novel and ambiguous stimulus. This way we can learn whether the mouse responds to the stimulus as if it were a call or not. We are now incorporating optogenetic methods to disable specific neurons in the auditory cortex during this auditory perception task to test the hypothesis that those neurons are crucial for the mother to recognize and respond to the cries of her offspring. In the future, we plan to use this paradigm to quantitatively assess perceptual deficits in mutant mouse lines that exhibit impaired pup retrieval.

Social Cognition and Parental Care as a Model for Altruistic Behavior

A. Corona

We recently performed whole brain imaging (in collaboration with P. Osten) from male and female mice to screen for parental behavior–specific neural activity. These experiments identified a number of regions classically associated with social affiliation such as the bed nucleus of the stria terminalis (BNSt), the medial preoptic area (MPOA), and basal amygdala that exhibit activity specifically during retrieval behavior. One unexpected region to emerge from this screen was the anterior cingulate cortex (ACC), which has been linked to cost–benefit judgments, social decision-making, and vicarious fear learning. It is considered to be important for sensing distress in others. We made optical recordings from the ACC in freely behaving mice and found that this structure is activated during retrieval. Moreover, when we deactivate the ACC, we see disruption of both retrieval and investigatory response to pups trapped in a glass jar. We therefore see maternal care as a potentially useful model of altruistic behavior. We have designed an environment that includes a balance beam over a pit and we have begun using it while delivering looming stimuli—expanding dark circles above the mouse that simulate an aerial attack from a bird of prey. We also plan to incorporate predator odors, all in an effort to dissuade the mouse from retrieving the pup. In conjunction with this we will record and manipulate activity in the ACC to test our hypothesis that ACC integrates threat and distress cues and influences social decision-making accordingly.

The ACC is also interconnected with the noradrenergic brainstem nucleus locus coeruleus, another important regulator of parental behavior. By measuring activity in both regions and their connections with a noradrenaline sensor, we have shown that this circuit is activated in both mothers and fathers, but the dynamic pattern of activation differs between the two. This suggests that differences between maternal and paternal behavior arise from common circuits that are regulated in different ways.

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 ASDs, we are continuing a long-term project on the neural circuitry of the auditory cortex and how it contributes to retrieval learning in wild-type mice, as well as how it is disrupted in *Mecp2* mutant mice. *Mecp2* is a single gene that, when mutated, causes Rett syndrome (RTT) in humans. Therefore, we anticipate that understanding the biology of this gene in mice may point the way to novel therapeutics.

Previously, we established that (1) expression of *Mecp2* is acutely required in the auditory cortex for accurate pup retrieval learning; (2) this appears to be in large part a result of loss of *Mecp2* in parvalbumin-expressing inhibitory interneurons (PVns) and alterations in closely associated extracellular structures known as perineuronal nets (PNNs); in contrast, *Mecp2* is apparently dispensable for retrieval in other major inhibitory neuron classes (VIP and somatostatin); (3) PV neurons in wild-type (WT) mice disinhibit the auditory cortex in response to maternal experience, which may be important for experience-dependent plasticity, but this does not occur in *Mecp2* mutants; and (4) genetic and pharmacological manipulations that reversed abnormalities associated with *Mecp2* knockout also improved behavior and corrected neural activity patterns.

We are extending this work by examining the specific role of PV neurons in the auditory cortex during maternal learning in WT mice and in *Mecp2* mutants. We are doing this through two separate but related projects that have yielded converging lines of evidence placing PV neurons in a central role for maternal plasticity and its disruption in *Mecp2* mutants. First, we have performed electrophysiological and optical recordings, behavioral assays, and immunohistochemical staining in mice that sustained selective deletion of *Mecp2* in PVns. The data demonstrate that these mice exhibit the same impairments in behavioral performance, neuronal activity, and molecular patterns as the whole-brain mutants. Second, we are using optical recording methods to measure PV activity in mice while they are actively retrieving pups. We find very different and much more complex patterns of activity in PVns of awake interacting mice as compared with awake yet restrained mice. Surprisingly, despite the fact that these recordings are being made in a sensory structure, much of this activity appears to be largely unrelated to sensory stimuli, instead reflecting behavior state, attention, or other nonsensory influences.

Multisensory Integration in Maternal Retrieval Behavior

A. Nowlan

Successful navigation of social encounters requires integration of multimodal sensory events (e.g., sound, smell, touch) with emotional states. How our emotions may shape sensory responses to social cues remains unclear. We are working to determine how social odor cues activate the emotional circuitry of the basal amygdala to influence processing of social vocalizations by the auditory cortex. How pup retrieval is motivated is poorly understood, but interaction with pups is highly rewarding to mothers, suggesting that sensory experience with pups is a crucial trigger for its acquisition. In accordance with that idea, both the smell of pups and the sound of their vocalizations are jointly required for the performance of pup retrieval. Indeed, previous work has shown that the smell of pups can directly interact with the coding of vocalizations in the auditory cortex of maternally experienced females.

The following findings argue that pup odor and USVs are integrated via a pathway from odor-responsive neurons in the basal amygdala (BA) to the auditory cortex (AC). First, we find that a large, scattered population of excitatory neurons in the BA project directly to the AC. Second, using fiber photometry in awake mice, we find that AC-projecting BA neurons respond to pup odors and are active during search for pups in a retrieval assay. Third, we find that optogenetic activation of AC-projecting BA neurons elicits dramatic and sometimes complex restructuring of AC neurons' responses to sound. Finally, we show that the effects of activating this BA-AC circuit are magnified and switch from primarily inhibitory to primarily excitatory after maternal experience. These findings have two significant implications. First, the origin of the odor-signaling projection in the amygdala, a structure crucial for motivated behavior, raises the possibility that it carries affective information such as salience or reward. Second, our finding that the BA modulates responses in primary sensory cortex is unexpected and may constitute a novel mechanism for merging affective state and sensory processing. We have submitted a paper reporting these results.

Neural Activity Signaling Emotion, Arousal, and Reward during Social Encounters

R. Dvorkin, Y. Xie

The superficial simplicity of pup retrieval belies the complexity of the neural circuitry and neurochemical events that drive it. We view it as a whole-brain behavior. Two projects from our laboratory suggest that the neuromodulators noradrenaline (NA) and dopamine (DA) are important for motivating and reinforcing pup retrieval behavior.

The first project concerns a small nucleus in the brainstem called locus coeruleus (LC) that releases NA throughout the brain. 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, we made both electrical recordings of individual neurons and optical recordings of neuronal populations in LC during parental interactions. The data define common principles that govern LC's

participation in cognition and its role in social behavior. Contact with pups during retrieval events precisely coincided with phasic bursts in individual LC neurons and rapid, transient increases in optically detected bulk fluorescence that continued until the pup was dropped in the nest. The ubiquity of this response among LC neurons and its reliability and magnitude in fiber photometry recordings strongly suggest that these events are coordinated across LC and broadcast NA release throughout the brain. Retrieval-related LC bursts could not be explained merely by responses to sensory stimuli, general motor activity, or reward, and changes in tonic firing were not seen during highly similar, but nonsocial, motor activities. Analysis of the relationship between phasic events and retrieval behavior indicates that LC activity specifically correlates with impending behavior. A manuscript reporting these results was posted on bioRxiv in 2021 and is currently in press at the *Journal of Neuroscience*.

In the second project, we have performed 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" (RPE). What this means is that DA does not signal reward per se, but rather the difference between expected and encountered 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. RPE is important because it is theoretically optimally suited as a teaching signal to update the value of a reward in reinforcement learning. We developed a novel behavioral paradigm in which a sliding door opens to provide a maternal female mouse with access to a chamber. The task design allows us to manipulate the expectations of the subject for whether a pup will be available when the door opens, and then we can either meet or violate those expectations. We made the following observations: First, when the female mouse is introduced to a pup and performs more than 100 retrievals, we find that the change in approach velocity (update in value) is correlated with the previous trial's DA signal. Second, when the expectation of pup availability is violated, the VTA neurons show a below-average response, and when the expectation of pup unavailability is violated, the VTA neurons show a larger-than-average response. Third, we used temporally precise optogenetic inhibition of VTA triggered by pup proximity to cancel the DA response to pups, and we found that the mice were far slower to learn efficient pup retrieval. We conclude that maternal care is shaped by a dopamine RPE signal for which the pup itself is the unconditioned stimulus.

PUBLICATIONS

Dvorkin R, Shea SD. 2021. Precise and pervasive phasic bursting in locus coeruleus during maternal behavior in mice. bioRxiv doi:10.1101/2021.03.31.437751

Furlan A, Corona A, Boyle S, Sharma R, Rubino R, Habel J, Gablenz EC, Giovanniello J, Beyaz S, Janowitz T, Shea SD, Li B. 2021. Neurotensin neurons in the central extended amygdala control energy balance. bioRxiv doi:10.1101/2021.08.03.454970

Yoles-Frenkel M, Shea SD, Davison IG, Ben-Shaul Y. 2021. Balancing memory fidelity and representational stability in the female mouse accessory olfactory bulb. bioRxiv doi:10.1101/2021.10.27.466073

In Press

Dvorkin R, Shea SD. 2022. Precise and pervasive phasic bursting in locus coeruleus during maternal behavior in mice. *J Neurosci* (in press).

HORMONAL REGULATION OF GENE EXPRESSION IN THE BRAIN

J. Tollkuhn K. Denney S. Sun
B. Gegenhuber M. Wu

Our laboratory studies hormone receptor signaling in the brain. Gonadal steroid hormones such as estrogen and testosterone are the primary drivers of sex differences in neural physiology and behavior. These hormones principally act through their cognate nuclear receptors, which bind DNA to regulate gene expression. Our overarching goal is to identify the gene programs regulated by steroid hormones in the brain and determine how these genes influence brain development, behavior, and disease risk. We anticipate that hormone-responsive genes underlie sex differences in the incidence and etiology of psychiatric and neurological diseases.

Understanding the Actions of Selective Estrogen Receptor Modulators on the Brain

We have recently identified the first direct target genes of estrogen receptor alpha (ERα) in the brain, which include neurotransmitter receptors, ion channels, and growth factors. These findings close a major gap in understanding how estrogen exerts its diverse beneficial effects on mood, cognition, neuronal metabolism, and neurovascular health. We now aim to leverage this insight to identify novel biomarkers that will facilitate the development of brain-specific hormone therapeutics. A central challenge in harnessing the beneficial effects of estrogen for therapeutics arises from the broad actions of this hormone throughout the body. Treatment with estrogen alone increases risk of certain cancers, thromboembolic events, and stroke. Selective estrogen receptor modulators (SERMs) provide mixed agonist/antagonist activity against estrogen receptors, depending on the tissue of interest. Recent studies have found that the SERM raloxifene (commonly prescribed as Evista) is beneficial for bone and brain health, while maintaining antagonistic, antitumor effects in breast and reproductive organs. We have initiated studies to investigate how raloxifene alters ERα-mediated gene expression in the brain with the long-term goal of developing hormone replacement therapies that provide the positive cognitive, mood, and neuroprotective effects of estrogen without increasing risk of cancer or cardiovascular events.

Estrogen Regulation of Physical Activity

It has been known for decades that female mammals are more physically active when estrogen levels are high. In collaboration with the Ingraham laboratory at UCSF, we have found that the ERα target gene *Mc4r* is directly responsible for this phenomenon (Krause et al. 2021). *Mc4r* expression in the ventrolateral region of the ventromedial hypothalamus (VMHvl) peaks during proestrus (high-estrogen) and is almost undetectable in diestrus (low-estrogen) females or males. We determined that ERα binds the *Mc4r* promoter and an additional downstream enhancer region, and that the bound estrogen response element (ERE) sequences are evolutionarily conserved in humans (Fig. 1). Designing a guide RNA that directly binds the promoter ERE permitted CRISPR-mediated activation of *Mc4r* expression. Viral delivery of dCas9-VP64 and the *Mc4r* single-guide RNA (sgRNA) specifically into the VMHvl boosted *Mc4r* expression in ovariectomized females, thereby bypassing systemic hormone state. Females that received the *Mc4r* sgRNA traveled over twice the distance in a day compared to controls, and additionally showed a 10% increase in bone density 4 mo after infection. This work demonstrates that estrogen acts through *Mc4r* in the VMH to increase motivation to engage in physical activity. We expect that similar circuit-level dissections of other ERα target genes identified by our group will inform additional aspects of sex-differential physiology and behavior.

Evolution of the Social Brain

How does the brain form and maintain social bonds, and what happens when this ability is disrupted in

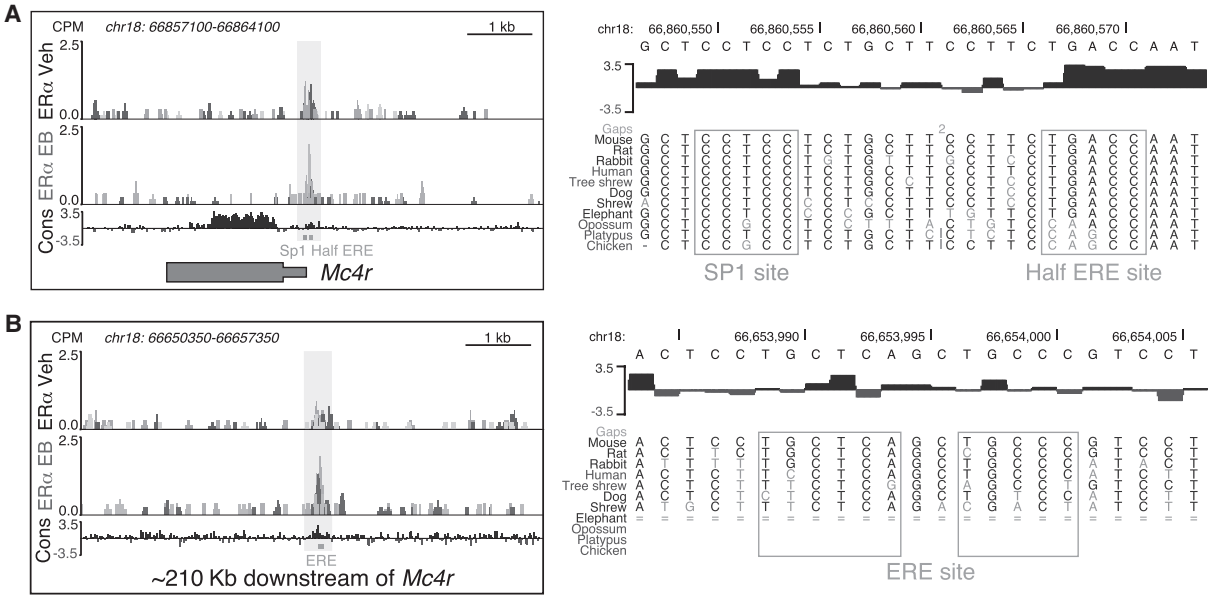


Figure 1. ERα genomic binding sites in *Mc4r* locus. CUT&RUN CPM-normalized coverage track showing estradiol-benzoate-specific ERα binding sites within *Mc4r* promoter (*A*) and ~210 kb downstream (*B*), in 400,000 subcortical brain nuclei collected from vehicle and estradiol benzoate (5 µg)-treated gonadectomized mice. Below each track the location and sequence conservation of (*A*) half SP1/ERE and (*B*) full ERE consensus sites in target gene loci are indicated by pink and green boxes.

psychiatric disorders? To address these fundamental questions, our laboratory is carrying out comparative studies of two vole species: prairie voles, which form lifelong pair bonds and show biparental care of offspring, and meadow voles, which like most rodents are promiscuous and uniparental. Studying voles is necessary because mice and rats do not form social attachments to other adult individuals. We hypothesize that (1) underlying genetic differences between the two vole species have contributed to altered gene functions or expression; (2) structural and sequence variation in prairie voles has produced species-specific *cis*-regulatory elements that specify the capacity to form social attachments; and (3) distinct genome variations between species alter regulatory regions to define changes in neural circuits that regulate distinct modules of social and attachment behaviors. Prior studies have implicated specific brain regions in social bonding. We are currently performing comparative epigenome analyses in these brain regions in prairie voles, meadow voles, and mice to identify putative prairie vole–specific enhancers that specify species-specific patterns of gene expression. Discovery of individual genes that directly relate to behavioral parameters has proved challenging in mammalian systems, and we have published a review highlighting technical advances, experimental design considerations, and conceptual framing that will aid future "sociogenomic" studies.

PUBLICATIONS

Krause WC, Rodriguez R, Gegenhuber B, Matharu N, Rodriguez AN, Padilla-Roger AM, Toma K, Herber CB, Correa SM, Duan X, et al. 2021. Oestrogen engages brain MC4R signalling to drive physical activity in female mice. *Nature* **599:** 131–135.

Ruiz-Ortiz J, Tollkuhn J. 2021. Specificity in sociogenomics: identifying causal relationships between genes and behavior. *Horm Behav* **127:** 104882.

CIRCUITS UNDERLYING DECISIONS IN AUDITORY CORTEX

A. Zador N. Bhattasali, X. Chen, Divyansha, G. Henry, L. Huang, E. Isko, C. Krasniak, S. Lu, D. Maharjan, K. Matho, C. Soitu, K. Vinehout, L. Yuan, A. Zhan, 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 a brain at single-neuron resolution rapidly and efficiently, at low cost.

Neural Circuit–Inspired Sensorimotor Control of Virtual Animals

N. Bhattasali

The fields of artificial intelligence and neuroscience have an intertwined history, and insights from the brain can be used to build better intelligent machines. For example, convolutional neural networks inspired by visual circuitry have become widely successful because they encode biases useful for vision tasks. However, the extent to which artificial neural network (ANN) architectures inspired by neural circuitry can yield useful inductive biases for other domains remains unknown. In our work, we asked what advantages biologically inspired network architecture can provide in the context of motor control. Specifically, we translated *Caenorhabditis elegans* circuits for locomotion into an ANN model applied to a simulated Swimmer agent. On a locomotion task, our architecture achieved asymptotic performance comparable to generic architectures, with significantly improved initial performance, data efficiency, parameter efficiency, interpretability, and transfer. Our work demonstrates several advantages of ANN architectures inspired by systems neuroscience and suggests a path toward modeling more complex animals.

Transcriptomic Basis of Areal Specialization of the Cortex

X. Chen

The cortex comprises cortical areas that function differently in behaving animals. These cortical areas are composed of neurons that are specialized in long-range connectivity and gene expression, both of which vary systematically across the whole cortex. However, whether differences in neuronal types defined by gene expression (i.e., transcriptomic types) can account for the specialized projections across cortical areas remains unclear. As a first step in resolving this question, we applied BARseq2 to interrogate the spatial distribution of transcriptomic types of neurons across the whole cortex. Clustering of neurons based on gene expression recovered transcriptomic types in both the cortex and subcortical areas. These transcriptomic types were consistent with previous single-cell RNA-seq studies and were distributed differentially across cortical areas, cortical layers, and noncortical structures. Surprisingly, we found that cortical neuronal types in the superficial layers were more predictive of cortical areas than those in the deeper layers. Our findings suggest that the superficial layers of the cortex are more specialized by areas than deep layers. We are now testing whether this difference in specialization is consistent with projection differences across areas by mapping projections of transcriptomic types across areas using BARseq2.

Encoding Innate Ability in Reinforcement Learning Agents through a Genomic Bottleneck

Divyansha [in collaboration with Koulakov laboratory, CSHL]

Many animals are born with elaborate behavioral capacities, the origins of which are the neural circuits

that are encoded in their genome. The rules required to form these neural circuits are several magnitudes higher than the capacity of the genome. This means that these rules pass through an "information bottleneck" while propagating from one generation to another. We build from previous work (Koulakov et al. 2021) in which the authors explored the innate behavior capacity of artificial neural networks in a supervised setting. Last year, we worked on exploring the problem of innate behavioral capacity in reinforcement learning models. We have worked on some classic reinforcement learning tasks like Atari games (Space Invaders and Beam Rider) for which we were able to preserve the model performance even after 1000x compression of model weights. We also tested our model on some locomotive benchmarks like the Half Cheetah environment of OpenAI gym. This environment has a continuous action space and is much more susceptible to perturbations in the policy weights, but we were still able to compress the policy and get high performance at the start of learning. Currently, we are running some transfer learning experiments to see how well these compressed weights can be transferred to different variations of the Half Cheetah—for example, scaling the size of its limbs or changing the gravity of the environment.

Barcode Sequencing in Hydrogel-Embedded Brain Slices

G. Henry

Sequencing of neuronal barcodes, delivered by a viral library, can be achieved by either bulk (MAPseq) or in situ (BARseq) methods. The latter is essential to the collection and correlation of projection and gene expression data. Furthermore, there is an interest in developing assays that will allow mapping the projections of neurons whose activity has been studied by in vivo two-photon imaging. Toward this end, we have optimized a hydrogel-based sequencing system that permits cDNA amplification and barcode sequencing in (300-micron) thick brain slices with the eventual goal of simplifying the registration of in vivo and ex vivo data collected from the same brain region. To do this, properties of the hydrogel were manipulated to enhance access of molecular reagents during cDNA amplification. These include shorter individual polymer lengths, increased porosity achieved through the use of a novel PEG-based linker, and an acute slice preparation that avoids the typical collapse of extracellular space observed in aldehyde-fixed brain slices. We are currently developing the tools required to register data obtained from multiple rounds of sequencing.

Highly Nonrandom Projection Patterns Revealed by Brain-Wide Projection Mapping at Single-Cell Resolution

L. Huang

Using further optimized BRICseq (brain-wide individual animal connectome mapping by sequencing), we have determined the projection patterns from more than 200,000 single neurons spanning an entire cortical hemisphere in a single animal. We found that although bulk corticocortical projections are dense, most of these projections are weak, and the top 20% of the strongest projections constitute 80% of total projection strengths. In strong contrast to the dense bulk connectivity, individual neurons only project to a few targets in a distance-dependent manner, and bulk projection strength is mostly determined by single-cell projection probability rather than single-cell projection strength. Moreover, we have shown that bifurcating and trifurcating projection motifs from single cells are highly nonrandom, and such nonrandomness can be partially explained by the minimize-wiring-length principle. Thus, our study provides novel insights into the principles that organize single-cell projection patterns on a brain-wide scale.

Comparative Connectomics and Transcriptomics in Lab and Singing Mice

E. Isko

Despite a close evolutionary relationship, Alston's singing mice (*Scotinomys teguina*) and lab mice (*Mus musculus*) exhibit extremely divergent vocal behaviors. This behavioral divergence must derive from differences in the underlying biology of neural cell types and/or neural circuits. To determine these differences, we are collaborating with the Banerjee laboratory to apply techniques developed in the Zador laboratory to determine the neural circuits of vocal behavior in lab and singing mice. We have begun our connectomic comparison in the orofacial motor cortex (OMC), a

brain area involved in the singing mouse song proven through electrical, pharmacological, and cooling experiments (Okobi et al., *Science* **363:** 983 [2019]). We are using both bulk methods (viral tracing) and single-cell methods (MAPseq, developed by the Zador laboratory) to characterize the projection patterns of neurons in the OMC of the two species. Our future directions include using single-nucleus RNA-seq (snRNA-seq) and BARseq2 (an in situ sequencing technique that combines gene expression and projection data developed by the Zador laboratory) to determine differences in cell type and the spatial location of these cells in the OMC of lab and singing mice. Finding transcriptomic and neural circuitry differences between lab and singing mice will give insights into the neural substrates of vocal behavior.

Mesoscale Neural Correlates of Stimulus Prior in Cortex as Seen with Wide-Field Imaging

C. Krasniak

To make decisions, animals must combine sensory information with prior beliefs to form a motor output. Although the representation of sensory information has been widely studied for decades, far less is known about how prior beliefs are represented in the brain. To study the cortical representation of this prior, we trained mice to perform the International Brain Lab (IBL) visual decision-making task, in which the stimulus prior is manipulated in blocks of trials. We then used wide-field calcium imaging to monitor neural activity during behavior from the entire dorsal cortical surface and applied a logistic regression model to decode the prior from single imaging frames.

From this mesoscale activity we were able to decode the task-generated prior from single imaging frames with a logistic regression model. Importantly, we can perform this decoding both before stimulus onset, as well as after stimulus onset, on catch trials in which no stimulus appears. Surprisingly, we find that before stimulus onset, the weights of the decoder are opposite the expected direction: Strong positive weights in visual and motor areas promote decoding the ipsilateral block. Even more surprisingly, shortly following the stimulus onset the representation flips to the expected sign—to a contralateral representation of block prior. This representation is not only in more frontal decision areas, but is also present as early as primary visual cortex. The flip in the sign of the decoder, and the widespread distribution of the bias signal throughout the cortex, strongly constrain models of neural Bayesian inference.

Assessing the Replicability of Spatial Gene Expression Using Atlas Data from the Adult Mouse Brain

S. Lu

What is the relationship between gene expression and brain areas defined by conventional neuroanatomy? To study this relationship, we are using data from a spatially resolved gene expression assay, Spatial Transcriptomics (ST), in conjunction with the Allen Brain Atlas adult mouse in situ hybridization data. Using linear models in a supervised learning framework, we found that brain areas are classifiable using only expression, but that generalizability across the two data sets was not bidirectional. ST generalized better to Allen than the reverse. Further, cross–data set performance is partly explained by distance in the phylogeny of brain areas of the reference atlas, but not by Euclidean distance in the brain. This suggests that gene expression is meaningful in brain parcellations and not merely capturing large gradients.

Investigation of Cell Type–Specific Corticostriatal Strengthening during Auditory Decision-Making

D. Maharjan

Animals can respond to environmental changes with appropriate actions in order to improve their chances of acquiring food and avoiding predation. To investigate how these stimulus-action associations are represented in the brain, our laboratory has developed an auditory decision-making task, called the "cloud of tones" task, in which rodents are trained to associate making rightward and leftward movements with stimuli containing primarily high or low tones. Previous studies from our laboratory have shown that learning of this association involves synaptic strengthening of connections between two regions of the brain involved in auditory decision-making:

auditory cortex and auditory striatum. Furthermore, the striatum is primarily composed of Drd1a receptor expressing (D1) medium spiny neurons and Drd2a receptor expressing (D2) medium spiny neurons. We are currently investigating whether this aforementioned strengthening of connections takes place in a cell type–specific manner onto D1 or D2 neurons of striatum.

Developmental Mechanisms Define Neuronal Diversification and Cell Type in the Mature Cerebral Cortex

K. Matho

Temporally specific transcription factors and effector genes play a key role during development in determining cortical glutamatergic pyramidal neuron (PyN) identity. Recent advances in mouse genetics have resulted in an extensive set of genetic strategies based on mouse knock-in driver lines for targeting and fate mapping PyN types and present an opportunity to examine the cell fate of genetically targeted progenitor pools, and to determine connectivity of lineage-defined and genetically defined PyNs, as they assemble into functional circuits (Muñoz-Castañeda et al. 2021). Evaluating the influence of developmental programs on cell type identity requires associating genetic strategies that define key aspects of developmental processes leading to PyN diversity with transcriptomic cell types and projection patterns—systematically and at cellular resolution. Recent developments from our laboratory in sequencing of gene panels in situ and neuroanatomical projection mapping have resulted in BARseq2, a novel technique that simultaneously detects multigene expression and maps single-cell long-range projections based on RNA barcode sequencing (Sun et al. 2021). We are working to integrate genetic strategies targeting progenitor pools and PyN subsets with BARseq2 to associate developmental trajectories with transcriptomic signatures and projection patterns at cellular resolution. We have validated a panel of genes to profile cortical cell types based on their developmental trajectories. In tandem, we have established mouse lines to investigate the link between developmental, molecular signature, and projection targets. We are currently combining these two elements to link developmental past with cell type identity.

High-Throughput Integration of Neuronal Activity, Connectivity, and Gene Expression

C. Soitu

The flow of information between different regions of the brain during decision-making is a largely unknown process. This is mainly because there are no tools available to disentangle the complex interaction of myriad cell types in the brain, governed by gene expression, connectivity, spatial organization, and other properties. This work aims to investigate the anatomical substrate of interregional communication at the single-neuron level. Tools developed in our group, BARseq and MAPseq, as well as our previous work on the topic, make this question accessible for investigation. We are using two-photon imaging of activity in single neurons to identify privileged populations of neurons. Successful completion of this project would generate unprecedented data sets that bridge information at all levels—anatomical, genetic, physiological, and behavioral. Gaining knowledge of the flow of information in the brain would push the frontiers of what it is one of the grand challenges of our time—to understand how behavior arises from neural circuits.

Linking Functional Properties and Transcriptomic Cell Types

K. Vinehout

Only by integrating data sets across experimental modalities—anatomical, genetic, physiological, and behavioral—can we develop an understanding of how the brain functions. We currently lack high-throughput methods and pipelines to combine these technologies in the same animal to characterize neurons in multiple ways (e.g., measuring their activity, connectivity, and transcriptome). To achieve this goal, we further developed a pipeline for the registration of light microscopy images and in vivo two-photon images.

After the alignment of in vitro slices to a 3D volume, we ran into issues with volume-to-volume registration. We moved away from advanced normalization tools (ANTs) registration and instead implemented Elastix registration. Some of the issues we encountered were with noise. We explored a wide range of options here and ultimately implemented a machine-learning-based approach called Noise2Void to remove noise from the images. This method yielded a better registration.

Currently we are trying to slice the brain in a way that allows for surface blood vessels to appear in the sliced light microscopy images. These are necessary for volume-to-volume registration.

Identifying Cell Type with Genomic Bottleneck

K. Vinehout

The only proof we have of intelligence is the brain. We believe taking principles from the proof of intelligence will help create artificial general intelligence. One small way to do this is to look at wiring rules in the brain and extrapolate these to machine learning models. Unfortunately, we do not and will not have the cellular connectome of anything except *C. elegans* for the foreseeable future. Regardless, we can take advantage of the idea that there are wiring rules in the brain to inform machine-learning architecture. We can achieve this by compressing machine learning models through a rule network to ensure compressible models arise. A challenge with this framework is to identify connection motifs that can arise in this framework.

This year we reworked the original code for the genomic bottleneck to allow for a more modular design. We also outlined requirements and methods to achieve wiring rules and connection motifs. In brief, our framework involves having simple rules, generalizable rules, and a large and complex P-net. Having simple rules involves a sparsity constraint on the G-net. Having generalizable rules involves one compression net for the whole P-net and data-agnostic training. Having a large P-net involves network growing methods and a complex P-net involves weight pruning. For motif identification we are looking at using graph neural networks. We recently started to implement some of these ideas into code.

Axonal BARseq, a Novel Technique for High-Throughput Mapping Single-Cell Projections In Situ

L. Yuan, X. Chen

A brain is an extremely complex structure and understanding how it is wired provides a road map for understanding brain function and behavior. Previously our laboratory developed MAPseq (multiplexed analysis of projections by sequencing) to barcode individual neurons and read out barcode in the projection targets using bulk sequencing results of high-throughput target mapping at single-cell resolution. Further, BARseq has been developed to sequence somatic barcode in situ. In the past several years, we have been combining both techniques into a novel technique called axonal BARseq to sequence axonal barcode in situ. The goal of axonal BARseq is to map single-cell projection from thousands of neurons within a single brain. As a proof of principle, we utilized this technique to map neural projection from mouse auditory cortex. The test was successful, and it generated a large amount of data. Currently, we are developing analysis tools to extract useful information from the data to answer biological questions.

In summary, we are continuing to develop and improve our axonal BARseq technique and its analysis. Hopefully this technique can be used to answer interesting biological questions in the coming year.

MAPseq Technique Development

H. Zhan

In the MAPseq technique, a modified presynaptic protein, MAPP-nl, was designed to specifically bind to and transport barcode mRNA into axon terminals. However, little work has been to done to identify the optimal presynaptic protein that can transport barcode mRNA efficiently into axons. We have selected 13 presynaptic proteins and compared their efficiency in transporting barcode mRNA. In 2020, we have found that snap25 and vamp2 have much higher transporting efficiency compared with the original MAPP-nl protein. This year, we compared the barcode transferring efficiency between 4xnl and 1xnl, and found that 1xnl is more efficient in binding box-B, which is connected with barcodes.

H. Zhan is also running the MAPseq Core facility. One of the projects in technique development is to make MAPseq work in the AAV virus. This year, we tried to make barcode RNA circulate in the AAV expression system and, in theory, circulated RNA should have a longer life span than linear RNA, and circulated RNA should be transferred into the axon terminal at a higher level in the AAV expression system. We have successfully made and been able to detect the circulated barcode RNA in the AAV virus–injected

area but unfortunately, we did not observe increased expression of RNA from the injected tissue area. More work has to be done to make MAPseq work in the AAV expression system.

Representations in Primary Visual Cortex underlying Visual Discrimination

A. Zhang

The neural circuits and representations underlying visual decision-making in rodents remain poorly understood. We set out to investigate the features represented in 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. We controlled stimulus viewing angle by implementing virtual head position control using online video tracking (Bonsai software; Lopes et al., *Front Neuroinform* **9:** 7 [2015]). Subjects readily learned to perform this task at high levels of accuracy. We used tetrode recordings to interrogate neuronal responses in primary visual cortex (V1) during behavior. In addition to classically spatially selective visual neurons, we found overlapping subpopulations of V1 single neurons that were selective for choice side and outcome from both current and past trials. Further, we found that representations were distributed across the population, with little correlation between tuning across features. We compared these activity patterns to those recorded in animals trained to view the visual stimulus, but who base their decision on a nonvisual (auditory) stream of information. We found that although stimulus and choice representations were preserved in the visually independent choice animals, other task features such as outcome and past trial parameters were less well represented. Finally, we used behavioral manipulations to study the underlying decision strategy used by the animals to solve this visual discrimination task and found a baseline tendency to use just the lower half of the full stimulus to solve the task. Interestingly, animals' strategy depended on the statistics of the stimulus distribution over trials. When we increased the number of trials presenting stimuli that depleted the success rate of animals' preferred strategy, they switched to an alternative strategy of comparing the stimulus content across the full stimulus space. Thus, it appears that rats integrate statistics of the environment over trials to guide selection of a discrimination strategy.

PUBLICATIONS

Chen S, Loper J, Chen X, Vaughan A, Zador AM, Paninski L. 2021. BARcode DEmixing through non-negative spatial regression (BarDensr). *PLoS Comput Biol* **17:** e1008256. doi:10.1371/journal.pcbi.108256

Ghosh S, Zador A.M. 2021. Corticostriatal plasticity established by initial learning persists after behavioral reversal. *eNeuro* **8:** ENEURO.0209-0220.202

Koulakov A, Shuvaev S, Lachi D, Zador A. 2021. Encoding innate ability through a genomic bottleneck. bioRxiv doi:10.1101/2021.03.16.435261

Muñoz-Castañeda R, Zingg B, Matho KS, Chen X, Wang Q, Foster NN, Li A, Narasimhan A, Hirokawa KE, Huo B, et al. 2021. Cellular anatomy of the mouse primary motor cortex. *Nature* **598:** 159–166. doi:10.1038/s41586-021-03970-w

Sun YC, Chen X, Fischer S, Lu S, Zhan H, Gillis J, Zador AM. 2021. Integrating barcoded neuroanatomy with spatial transcriptional profiling enables identification of gene correlates of projections. *Nat Neurosci* **24:** 873–885. doi:10.1038/s41593-021-00842-4

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 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 animals and plants, 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. This year, they also demonstrated 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-gen" profiling and chromatin immunoprecipitation methods that have revealed many new hypotheses in 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 process 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 Lippman laboratory 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 on 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 RNA responsible that guide 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 relocates resources to better adapt.

The **Ullas Pedmale** laboratory's research goals seek 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, S. Daniel Iohannes, M. Kitagawa, K. Lam, J-H. Lee, P. Lindsay, L. Liu, T. Petrizzo, A. Shen, T. Skopelitis, K. Swentowsky, T. Tran, M. Venezia, Z'D. Williams, P. Wu, X. Xu

Our research asks how the growth and shape of plants is controlled, with the long-term goal of improving crop yields. We aim to identify genes, signals, and pathways that regulate plant architecture and development. Plants grow and develop by controlling the flow 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 using CRISPR-Cas9 genome editing. We also completed a "single-cell" atlas of maize ear development, providing a rich resource for research into yield traits. We continue to characterize enzymes that function in sugar metabolism, but also may play a "moonlighting" role in the cell nucleus, and discovered how a signal important for stem cell fate in plants travels as an RNA message between cells.

The Control of Meristem Size in Maize

P. Lindsay, K. Swentowsky [in collaboration with F. Xu, Shandong University, China]

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.

We 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 meristem 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 IP-MS (immunoprecipitation-mass spectrometry). We also transformed constructs into maize for proximity labeling to assess transient interactions between FEA3 and interacting proteins. In this technique, proteins of interest are tagged with TurboID, a biotin ligase that has been engineered to biotinylate nearby proteins. This technique may prove particularly useful for identifying transient receptor interactions.

In addition to assessing protein–protein interactions, we tested the ability of FEA3 to bind to its predicted ligand, FCP1. In collaboration with Professor Yoshikatsu Matsubayashi at Nagoya University, we found that FEA3 does not directly bind FCP1, but BARELY ANY MERISTEM1 (BAM1), a leucine-rich repeat (LRR) receptor-like kinase that also controls meristem size, can bind FCP1. BAMs are part of the meristem signaling pathway in *Arabidopsis*, but their function in maize has not yet been evaluated. Intriguingly, one maize *BAM* gene, *BAM1D*, is up-regulated

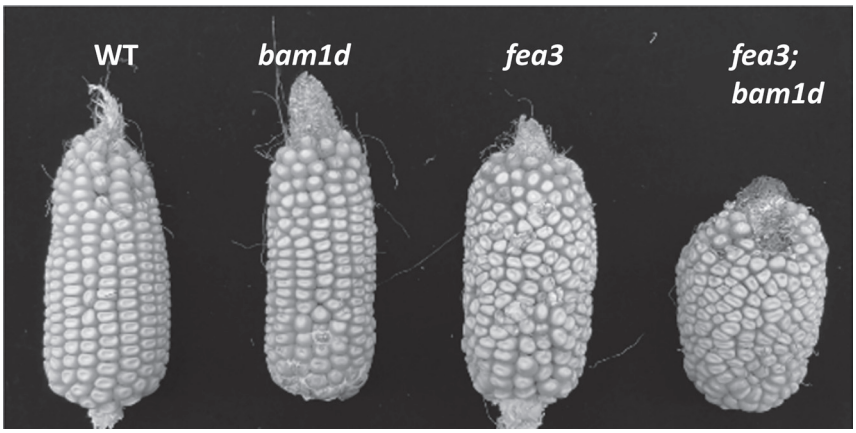


Figure 1. Double mutants suggest a genetic interaction between *fea3* and *bam1d*.

in the *fea3* mutant, and their expression overlaps, indicating these two proteins may form a receptor–co-receptor pair. Furthermore, a preliminary proximity labeling assay shows that FEA3 and BAM1D associate with one another in ear primordia. In addition to this physical interaction, *fea3; bam1d* double mutant ears are smaller than those of the *fea3* single mutant, indicating that *FEA3* and *BAM1D* interact genetically (Fig. 1). Because *BAM1D* has six paralogs in maize, other BAMs may compensate for a lack of *BAM1D*, so future work will assess compensation through gene expression and higher-order mutant analysis.

Characterization of additional factors that control inflorescence architecture will improve our understanding of this process. We have identified and rough-mapped two new fasciated ear mutants: *fea*-Leo* and *fea*-GN111*. Using quantitative trait locus sequencing (QTL-seq) from fasciated and normal ears, we mapped these two mutations to chromosome 6 and chromosome 4, respectively. Efforts to fine-map these two mutants and identify the causal genes are ongoing.

Genetic Redundancy in Circuits Controlling Meristem Development

L. Liu, P. Lindsay, X. Xu, T. Skopelitis, R. Chen [in collaboration with E. Demesa-Arévalo, Heinrich-Heine University, Germany; M. Bartlett, University of Massachusetts Amherst; Z. Lippman, CSHL]

The CLE (CLAVATA3/Endosperm surrounding region-related) peptides are fundamental players in meristem maintenance in plants, as mobile signals that establish feedback signaling to control the balance of stem cell division and differentiation. Weak alleles of fasciated ear mutants improve maize yield traits, such as kernel row number, by increasing meristem size and number of primordia, while maintaining structural integrity of the meristem. As *cle* mutants are fasciated, we asked if we could create weak alleles by mutating their promoters using CRISPR-Cas9 editing, and if these alleles could also enhance yield traits. *ZmFCP1* and *ZmCLE7* were targeted for CRISPR using multiplex guide RNAs, and a variety of edited alleles with deletions in accessible chromatin regions of their promoters were created. These alleles decreased expression, by ~45% to ~65%, and as hypothesized, they enhanced grain yield–related traits. This CRISPR-based promoter fine-tuning approach could be applied to diverse traits and crops.

Natural Variation in Inflorescence Architecture

P. Lindsay, S. Daniel Iohannes, R. Chen [in collaboration with M. Passalacqua, CSHL; K. Duncan and C. Topp, Donald Danforth Plant Science Center; J. Gillis, CSHL]

Maize inflorescence architecture has been a target for extensive selection by breeders, and the maize genome is highly diverse; hence, different inbred lines vary greatly. We are investigating gene expression differences in the inflorescence meristems of a set of diverse maize inbred lines called NAM. We are particularly interested in genes that modulate inflorescence meristem size because it correlates with kernel row number, an agronomically important trait. We performed 3′ RNA-seq on 16 of the NAM lines, and measured their inflorescence meristem (IM) sizes. After correlating gene expression profiles with meristem size, we

identified a candidate for CRISPR mutagenesis and will evaluate knockouts, as well as promoter alleles, to assess how this gene functions in inflorescence development. We will also perform RNA-seq on the remaining NAM founders to expand our data set. In collaboration with Dr. Christopher Topp at the Donald Danforth Center, we are also using X-ray microscopy to obtain high-resolution volumetric information about ear primordia in the NAM lines.

In addition to bulk changes in gene expression in the IM, we are interested in whether cell- or tissue-specific gene expression changes contribute to changes in IM size. Using our single-cell (sc) RNA-seq atlas as a reference point, we are deconvoluting our bulk RNAseq expression data. We will also perform scRNA-seq on a subset of the NAM lines to benchmark this approach. In the future, we hope to integrate gene expression profiles with single-cell assay for transposase-accessible chromatin with high-throughput sequencing (scATAC-seq), to pinpoint how chromatin accessibility impacts gene expression.

Identification and Functional Dissection of Shared *cis*-Regulatory Elements Controlling Quantitative Trait Variation across Angiosperms

K. Swentowsky [in collaboration with A. Hendelman and Z. Lippman, CSHL; M. Bartlett, University of Massachusetts Amherst; I. Efroni, The Hebrew University of Jerusalem]

Genomic *cis*-regulatory elements (CREs) mediate transcription factor binding and gene expression. Quantitative control of gene expression affects agronomically important traits and can be modulated using CRISPR-Cas9 editing of CREs. Although editing has become routine in many species, it has been difficult to identify CREs because of the substantial size and variation of noncoding genomic sequences. We hypothesize that many genes with conserved function and expression may contain deeply conserved CREs that can be identified through comparative genomic analysis. In collaboration with Idan Efroni at the Hebrew University of Jerusalem, we have constructed a computational pipeline for identifying deeply conserved CREs called Conservatory. Conservatory first identifies closely related genes between species within a plant family and finds conserved CREs within the family. Then, alignments are performed between families to identify more deeply conserved CREs.

In maize, we have designed CRISPR guide RNAs (gRNAs) to target CREs identified by Conservatory in the promoter of *INDETERMINATE1* (*ID1*). *ID1* encodes a transcription factor that promotes flowering, and *id1* mutant plants are severely delayed in flowering and have reduced determinacy. We identified conserved CREs in two regions of the *ID1* promoter: one in the first ~400 bp upstream of the start codon and a distal element about 2.5 kb upstream (Fig. 2). These regions overlap with accessible chromatin and are likely to be involved in regulation of *ID1* gene expression. Deletion of CREs important for *ID1* gene regulation should impact its expression and produce more subtle phenotypes compared to the null mutations.

In the 1960s, maize breeders sought to improve agricultural sustainability by breeding perennial maize

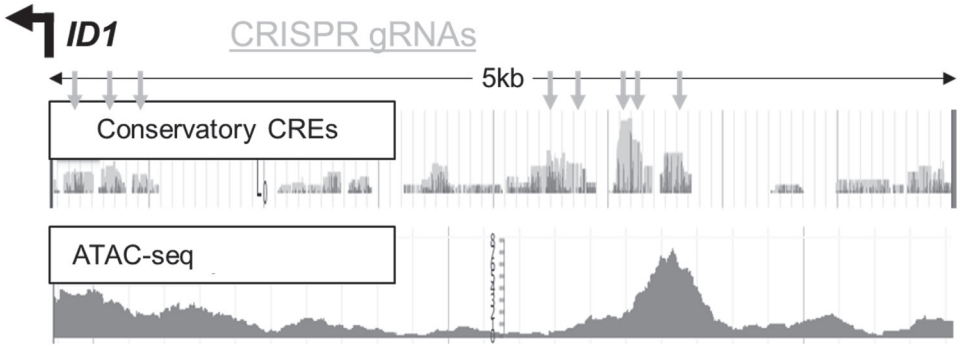


Figure 2. *cis*-regulatory elements in the maize *ID1* promoter identified by Conservatory. (*Top* track) Coverage of conserved sequences between grass species upstream of *ZmID1* identified by Conservatory; (*bottom* track) coverage of single-cell assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) reads showing accessible chromatin in bundle-sheath cells. Gray arrows indicate sites where CRISPR guide RNAs (gRNAs) have been designed for genome editing.

plants that can grow for multiple years. They combined the *id1* null mutant with the highly branched *grassy tillers1* (*gt1*) mutant and a perennial allele from a wild species. These perennial plants had undesirable characteristics, such as severely delayed flowering and extreme branching. Once we generate weak *id1* alleles by editing CREs, we will combine them with weak *gt1* alleles generated by Madelaine Bartlett's laboratory at University of Massachusetts Amherst and a locus for perennial regrowth that we recently mapped. We hypothesize that weak alleles of *id1* and *gt1* will allow plants to achieve perenniality without the undesirable phenotypes associated with the null alleles.

Functional Annotation of the Maize and Teosinte Reference Genomes

X. Xu [in collaboration with F. Li, J. Cahn, L. Wang, X. Wang, J. Drenkow, T. Gingeras, R. Martienssen, and D. Ware, CSHL]

Maize was domesticated from teosinte ~9000 years ago. This process was critical for adoption of maize as a productive crop and is largely due to selection of gene regulatory variation. To identify such variation, we performed chromatin immunoprecipitation assays with sequencing (ChIP-seq) for domestication-related transcription factors, such as GRASSY TILLERS1 (GT1) and TUNICATE (TU). We also conducted RNA-seq of developing inflorescences and pollen to functionally annotate the teosinte reference genome. By mapping these data sets to assembled teosinte and maize reference genomes and integrating them with histone ChIP-seq data sets generated by the National Science Foundation (NSF) MaizeCODE project, we identified variable regulatory regions associated with differentially expressed genes underlying the trajectories of maize domestication.

A High-Resolution Single-Cell Atlas of Shoot Meristems at Cross-Species and Multimodal Levels

X. Xu, D. Jackson [in collaboration with A.P. Marand and R.J. Schmitz, Department of Genetics, University of Georgia, Athens; M. Passalacqua, and J. Gillis, CSHL]

Single-cell technology has enhanced our understanding of plant development in recent years. However, it faces many challenges, such as capturing rare stem cells, cross-species comparisons, and integrating multimodal scRNA-seq and scATAC-seq data. In this study, we constructed a comprehensive single-cell atlas of plant shoot meristems across diverse species, including *Arabidopsis* and maize, to tackle these challenges.

We finely dissected developing apices, including the inflorescence meristems, of both *Arabidopsis* and maize and captured stem cell populations that were largely missed in previous single-cell studies (Fig. 3). We found conserved cell types, such as stem cells, epidermis, vasculature, and lateral organ primordia, between the two species. Our cross-species coexpression analysis at single-cell resolution found a significant overlap of stem cell markers between *Arabidopsis* and maize, and uncovered novel stem cell regulators.

We also profiled single cells from the inflorescence apices of mutants of the CLV-WUS pathway. We found hundreds of differentially expressed genes (DEGs) in the stem cells by comparing these mutants with wild type. We used CRISPR-Cas9 to knock out selected DEGs in a family of sugar kinases and found a striking meristem termination phenotype, validating the predictive power of our single-cell atlas.

Maize makes male and female inflorescences, the tassel and ear, and they develop similarly, except that the tassel makes branch primordia that are suppressed in the ear. To understand the developmental basis of these differences, we profiled developing maize ears and tassels using both scRNA-seq and scATAC-seq. We found that major cell types of the developing ear and tassel were broadly conserved, and the tassel branch primordia cell types resembled those in the finely dissected ear tips, suggesting that the same developmental program was reused in different meristem types across the two organs. We are integrating scRNA-seq and scATAC-seq to query chromatin accessibility and gene expression, and we uncovered accessible chromatin regions significantly associated with gene expression differences across cells in both ear and tassel.

Control of Shoot Branching and Determinacy

J-H. Lee, T. Tran, X. Xu, T. Skopelitis, A. Shen, Z'D. Williams [in collaboration with H. Claeys, Inari Agriculture, Belgium; E. Demesa-Arévalo, Heinrich-Heine University, Germany; R. Brennan, Duke University; H. Furukawa, CSHL]

Shoot branching is important for crop productivity. The *RAMOSA* (*RA*) genes impose determinacy on axillary

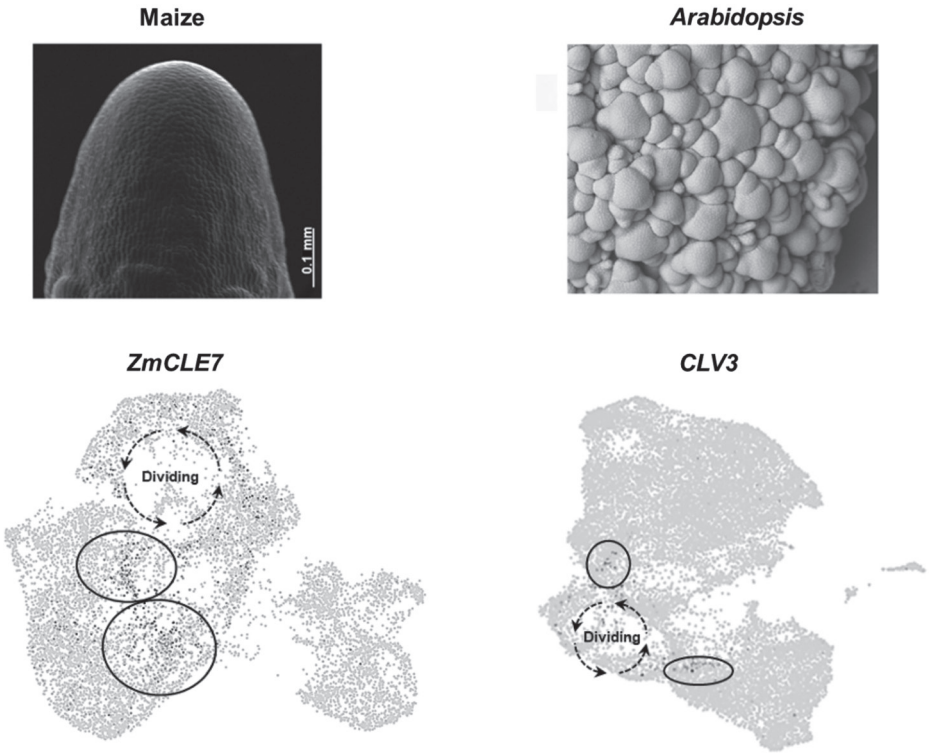


Figure 3. Stem cell markers are detected in both maize and *Arabidopsis*. (*Top*) Representative developing maize (*left*) and *Arabidopsis* inflorescence (*right*) used for scRNA-seq profiling. (*Bottom*) Expression maps of stem cell markers, *ZmCLAVATA3/EMBRYO SURROUNDING REGION-RELATED7* (*ZmCLE7*; *left*) in maize and *CLAVATA3* (*CLV3*; *right*) in *Arabidopsis*. Each dot indicates the profile of a single cell, and black dots are cells expressing *ZmCLE7* or *CLV3*. The black circles indicate groups of stem cells enriched in these genes, and dividing cells are marked by dashed circles with arrows.

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 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 may serve as mediators of signals—maybe a sugar signal—originating at the boundary domain and regulating determinacy. We also found that thiamine pyrophosphate (TPP) enzymes' activity could be uncoupled from their mutant phenotype, suggesting that TPP proteins have an alternative or "moonlighting" function.

The RA3 nuclear speckles colocalize with transcriptional markers. Such speckles can be formed by proteins with intrinsically disordered regions (IDRs), which mediate liquid-liquid phase separation. We found three candidate IDRs in RA3, two in the amino-terminal region and one at the carboxyl terminus (Fig. 4A). We purified full-length RA3 protein and an amino-terminal region as HIS-GFP-tag fusions from *Escherichia coli*. Both RA3 and its amino-terminal region could phase-separate (Fig. 4C,D), similar to a positive control protein, FUS (Fig. 4B), whereas a yellow fluorescent protein (YFP) control had no such activity (Fig. 4E). These results suggest that the in vivo RA3 nuclear puncta may be phase-separated condensates.

A catalytically dead mutant version of RA3 partially complements *ra3* mutants, supporting our moonlighting hypothesis. We used CRISPR-mediated base editing to refine this experiment by mutating endogenous RA3 and TPP4 loci to catalytically dead versions. We recovered a point mutation in TPP4, D294N, which replaces a conserved aspartic acid

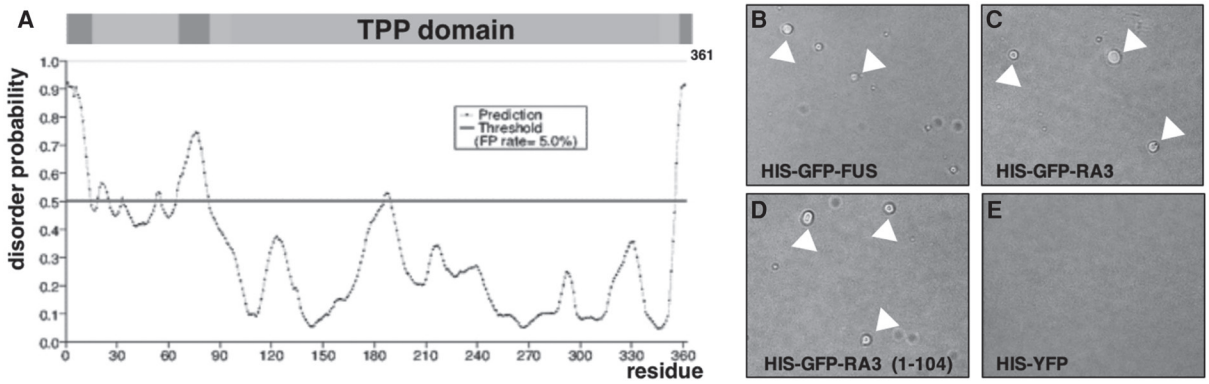


Figure 4. RA3 phase separates in vitro. (*A*) Map of RA3 protein and prediction of disordered regions (marked in dark gray in the schematic). (*B–E*) Phase separation of RA3 protein and its amino-terminal region into droplets (arrowheads), similar to a positive control, FUS.

(D) residue into asparagine (N). Collaboration with structural biologist Dick Brennan predicted that this single amino acid substitution does not affect TPP4 folding, and a purified TPP4 D294N protein from *E. coli* had almost undetectable activity in a phosphate release assay. However, our enzymatically null base edited mutant failed to complement *ra3;tpp4* double mutants. Our further analysis aims to uncover the biological significance of the enzymatic and nonenzymatic function of TPPs in modulating meristem activity.

We also screened for RA3 interactors using IP-MS. Of particular interest given the nuclear speckle localization of RA3, we found two predicted RNA binding proteins (RBPs). Bimolecular florescence complementation assays confirmed their interaction with RA3 in the nucleus, forming speckles, implying a potential regulatory function. Phylogenetic analysis found that the RBPs are in a group of four paralog genes. We designed multiplex CRISPR-Cas9 constructs to knock out all RBPs, and are combining with *ra3* mutations to test the functional significance of these interactions.

We also improved the sensitivity of our RA3 IP-MS and identified additional candidate interactors. Using co-immunoprecipitation experiments, we confirmed that RA3 interacts with a polyadenylate binding factor 2 (PABP2), which triggers the binding of eukaryotic initiation factor 4 complex directly to the poly(A) tail of mRNAs. We are characterizing transposon insertion lines for *PABP2* and crossed these mutants with *ra3* to dissect their genetic interactions.

Our IP-MS also found that RA3 interacts with ZmTPS1 and ZmTPS12, trehalose phosphate synthase proteins that naturally lack catalytic activity. *zmtps1* and *zmtps12* mutants enhance *ra3* phenotypes (Fig. 5), suggesting their physical interaction is biologically significant. We found that RA3 and ZmTPS1 were colocalized by immunolocalization. Interestingly, in yeast two-hybrid experiments, we found that ZmTPS1 also interacts with the two maize catalytically active TPSs. We mutated these two genes using CRISPR-Cas9 and found that the double mutants fail to complete embryogenesis, as in *Arabidopsis*. Next, we asked whether these interactions might affect enzymatic activity. We expressed and purified RA3 from *E. coli*; however, ZmTPS1 and ZmTPS14 are large proteins and did not express well, so we expressed and purified them from insect cells. We then designed an enzyme assay and found that the enzymatically inactive ZmTPS1 protein stimulated the activity of RA3+ ZmTPS14. We also expressed and purified RA3, ZmTPS1, and ZmTPS14 in insect cells, and found that they could form a complex. Next, we will visualize structural interactions and complex stoichiometry through cryo-electron microscopy.

In summary, our results suggest that plant TPPs function in a complex with both enzymatically inactive and enzymatically active TPSs, and the enzymatically inactive TPSs stimulate activity of the active enzymes. Our results provide insights for the first time into the combined activity of the two major trehalose gene classes, TPSs and TPPs, in plant development.

RA3 homologs are present in *Arabidopsis*; however, we found that single mutants have no phenotype. To overcome this redundancy, we used CRISPR-Cas9 to create a complete gene family knockout of the

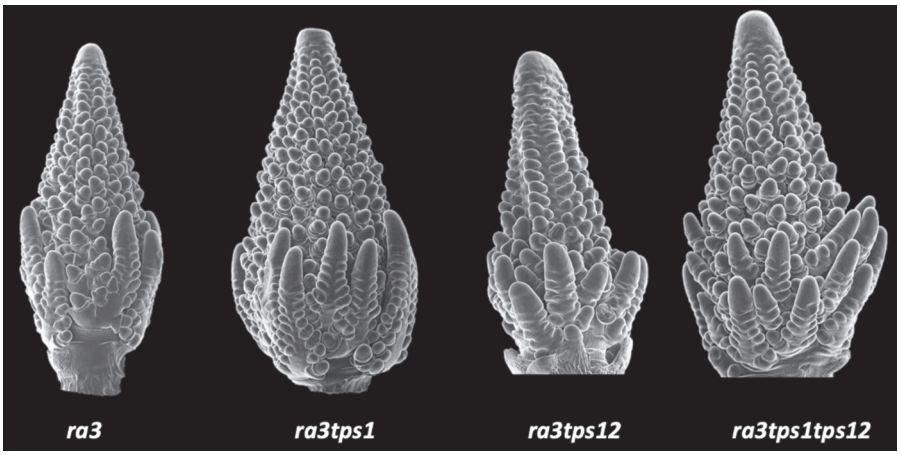


Figure 5. *zmtps1* and *zmtps12* mutants enhance *ra3* phenotypes. Images are scanning electron micrographs of developing ears.

10-member TPP gene family. These mutants enhance branching, reminiscent of the maize phenotype. We are characterizing the TPP knockout lines by RNA-seq analysis, sugar metabolite and nutrient analysis, and meristem measurements. We plan to use this multiplex knockout to test the enzymatic and nonenzymatic roles of the TPP gene family and the significance of its nuclear localization.

Uncovering the Nonenzymatic Function of RAMOSA3 in Maize Inflorescence Branching

X. Xu, D. Jackson [in collaboration with H. Claeys, Inari, Belgium; S.L. Vi, Agricultural Genetics Institute, Hanoi, Vietnam; George Chuck, Plant Gene Expression Center, UC Berkeley]

To tackle the mystery of the nonenzymatic function of RA3, we performed ethyl methanesulfonate (EMS) screening of *ra3* mutant and identified an enhancer, *indeterminate spikelet1* (*ids1*). IDS1 is an AP2-type transcription factor, and is the ortholog of the major wheat domestication gene *Q* that controls spikelet density and yield. By carefully examining the early developmental stages of *ra3;ids1* double mutants, we found that floral meristems were transformed into branches, rather than forming florets. We confirmed these findings by crossing *ra3* with additional *ids1* alleles and conducting allelism tests. Using in situ hybridization, we found RA3 and IDS1 were coexpressed around the boundary regions between floral meristems, which was also supported by our single-cell transcriptomic profiling data. To further examine whether IDS1 might be involved in the hypothetical transcriptional regulatory function of RA3, we tested physical interactions between RA3 and IDS1 *in planta* by bimolecular fluorescence complementation assays. Indeed, we found that RA3 and IDS1 proteins interact in nuclear speckles, reminiscent of the nuclear speckle localization of RA3. Our ongoing genetic and biochemical analyses will uncover the downstream targets of the RA3-IDS1 complex, and validate their interactions in vivo. Together, our data suggest that RA3 controls inflorescence branching by interacting with the transcription factor IDS1.

Mechanisms of Ear Development by a Maize RNA Binding Protein

P. Wu, M. Kitagawa, T. Tran, R. Chen, T. Skopelitis [in collaboration with H. Claeys, Inari Agriculture, Belgium; L. Liu, Huazhong Agriculture University, China; E. Demesa-Arévalo, Heinrich-Heine University, Germany]

RBPs govern a variety of important plant developmental processes by interacting with RNA targets and controlling posttranscriptional RNA metabolism. The maize RBP TERMINAL EAR 1 (TE1) regulates shoot apical meristem size, timing and position of leaf initiation, plant height, and ear development. However, how TE1 regulates these processes has been a mystery. We performed RNA immunoprecipitation and sequencing (RIP-seq) of TE1-YFP ear primordia, using nontransgenic siblings as a negative control. We found that TE1 bound preferentially to the 3′ UTR

and 5′ UTR regions of mRNAs, and many of them are transcription factors and development-related genes. Specifically, TE1 bound to its own mRNA and mRNAs of several *SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE* (*SPL)* genes, including *UNBRANCHED 2* (*UB2*) and *UNBRANCHED 3* (*UB3*). *UB2* positively regulates the expression of *UB3* by directly binding to its promoter, to control meristem size and kernel row number (KRN). We hypothesize that TE1 protein directly binds to *UB2* and *UB3* mRNAs to inhibit their translation. To test this hypothesis, we are purifying TE1 protein and will perform RNA electrophoretic mobility shift assay (EMSA) to confirm binding of TE1 to *UB2* and *UB3* mRNAs. We will also use translating ribosome affinity purification followed by RNA sequencing (TRAP-seq) to assess the translatome of WT and *te1* mutant ear primordia. To test the significance of the molecular interactions, we are also characterizing *te1* phenotypes, including triple mutants and quadruple mutants with maize *spl* genes, and are identifying TE1 interactors by IP-MS. Together these experiments could elucidate a new regulatory pathway of relevance to maize yield traits.

Mechanism of Active Transport through Plasmodesmata

M. Kitagawa, P. Wu, T. Skopelitis

Some transcription factors (TFs) and their mRNAs are actively and selectively transported between plant cells and function in cell fate determination. These molecules are transported through plasmodesmata (PD), membrane-lined channels traversing the cell wall. However, the mechanism underlying the active and selective transport through PD has been largely unknown. Previously, we established a system for evaluating transport in *Arabidopsis* seedlings using a mobile homeodomain TF, KNOTTED1 (KN1). Using this system, we isolated two trafficking-defective mutants that are mutated in the same gene, encoding *A. thaliana* ribosomal RNA processing protein (AtRRP) 44A, so we renamed the mutants as *atrrp44a-4* and *atrrp44a-5*. RRP44 is a subunit of the RNA exosome complex and functions in the processing and degradation of a variety of RNAs in eukaryotes. Both mutants had amino acid substitutions in the conserved catalytic domain. A native *AtRRP44A-RFP* construct was expressed in meristem tissues, overlapping with the *Arabidopsis KN1* homolog *SHOOT MERISTEMLESS* (*STM*). Double *atrrp44a; stm* mutants had much smaller meristems (Fig. 6). Collectively, these data support a hypothesis that AtRRP44A regulates cell-to-cell trafficking of KN1/STM signals to regulate stem cells in the shoot meristem.

AtRRP44A participates in RNA degradation and processing, but our RNA decay assays indicated no change in *STM* mRNA stability in our *atrrp44a* mutants. We also found that a noncatalytic AtRRP44A compensated for the defect in KN1 transport, suggesting that AtRRP44A function in RNA metabolism can be uncoupled from its role in KN1/STM transport. AtRRP44A is an RNA-binding protein, and it has been suggested that KN1 protein and mRNA interact when they traffic through PD. Therefore, to ask whether AtRRP44A directly controls *KN1/STM* mRNA trafficking, we visualized the *KN1* mRNA using the MS2 system. We found that *KN1* mRNA traffics between

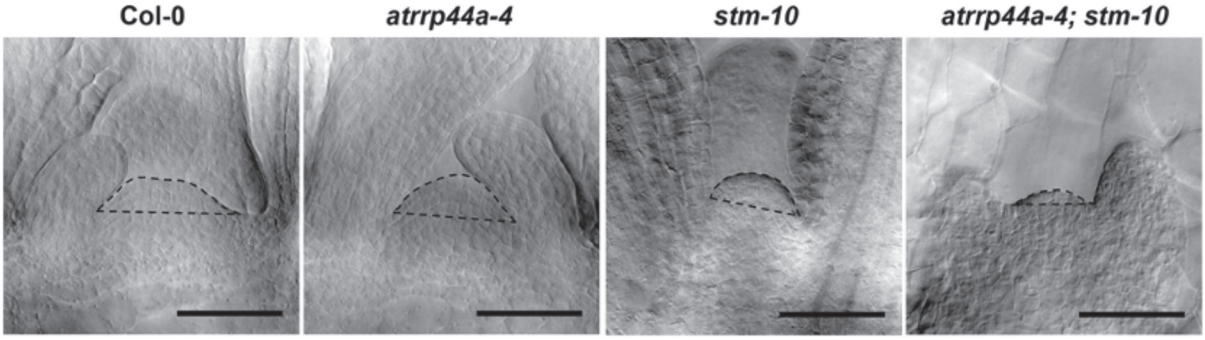


Figure 6. AtRRP44A is required for stem cell maintenance in shoot meristems through an STM-dependent pathway. *stm-10* mutants had smaller meristems, and the meristem size was even smaller in the *atrrp44a-4; stm-10* double mutants. Meristems are outlined by dotted lines. Scale bars, 50 µm.

cells, but this is impaired in *atrrp44a* mutants, and time-lapse imaging found that *KN1* mRNA transiently targeted PDs. RNA-immunoprecipitation reverse transcription polymerase chain reaction (RT-PCR) also revealed that AtRRP44A binds to *STM* mRNA. Collectively, these data support the idea that AtRRP44A binds to *KN1/STM* mRNA and mediates their trafficking between cells through PD to control stem cell fate.

PUBLICATIONS

Demesa-Arévalo E, Abraham-Juárez M, Xu X, Bartlett M. 2021. Maize RAMOSA3 accumulates in nuclear condensates enriched in RNA POLYMERASE II isoforms during the establishment of axillary meristem determinacy. bioRxiv doi:10.1101/2021.04.06.438639

Du Y, Lunde C, Li Y, Jackson D, Hake S, Zhang Z. 2021. Gene duplication at the *Fascicled ear1* locus controls the fate of inflorescence meristem cells in maize. *Proc Natl Acad Sci* **118:** e2019218118.

Liu L, Chen R, Fugina CJ, Siegel B, Jackson D. 2021. High-throughput and low-cost genotyping method for plant genome editing. *Curr Protoc* **1:** e100. doi:10.1002/cpz1.100

Liu L, Gallagher J, Demesa-Arévalo E, Chen R, Skopelitis T, Wu Q, Bartlett M, Jackson D. 2021. Enhancing grain-yield-related traits by CRISPR-Cas9 promoter editing of maize CLE genes. *Nat Plants* **7:** 287–294.

Liu L, Lindsay P, Jackson D. 2021. Next generation cereal crop yield enhancement: from knowledge of inflorescence development to practical engineering by genome editing. *Intl J Mol Sci* **22:** 5167.

Ning Q, Jian Y, Du Y, Li Y, Shen X, Jia H, Zhao R, Zhan J, Yang F, Jackson D, et al. 2021. An ethylene biosynthesis enzyme controls quantitative variation in maize ear length and kernel yield. *Nat Commun* **12:** 5832. doi:10.1038/s41467-021-26123-z

Ortiz-Ramírez C, Guillotin B, Xu X, Rahni R, Zhang S, Yan Z, Coqueiro Dias Araujo P, Demesa-Arévalo E, Lee L, Van Eck J, Gingeras TR, et al. 2021. Ground tissue circuitry regulates organ complexity in maize and *Setaria*. *Science* **374:** 1247–1252. doi:10.1126/science.abj2327

Rodríguez-Zapata F, Barnes AC, Blöcher-Juárez KA, Gates D, Kur A, Wang L, Janzen GM, Jensen S, Estévez-Palmas JM, Crow T, et al. 2021. Teosinte introgression modulates phosphatidylcholine levels and induces early maize flowering time. bioRxiv doi:10.1101/2021.01.25.426574

Tran TM, Demesa-Arévalo E, Kitagawa M, Garcia-Aguilar M, Grimanelli D, Jackson D. 2021. An optimized whole-mount immunofluorescence method for shoot apices. *Curr Protoc* **1:** e101. doi:10.1002/cpz1.101

Xu X, Crow M, Rice BR, Li F, Harris B, Liu L, Demesa-Arévalo E, Lu Z, Wang L, Fox N, et al. 2021. Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev Cell* **56:** 557–568. doi:10.1016/j.devcel.2020.12.015

Yang S, Xu F, Wang YM, Zhong WS, Dong L, Shi YN, Tang TJ, Sheng HJ, Jackson D, Yang F. 2021. Glutaredoxins regulate maize inflorescence meristem development via redox control of TGA transcriptional activity. *Nat Plants* **7:** 1589–1601. doi:10.1038/s41477-021-01029-2

Zhu C, Liu L, Crowell O, Zhao H, Brutnell TP, Jackson D, Kellogg EA. 2021. The CLV3 homolog in *Setaria viridis* selectively controls inflorescence meristem size. *Front Plant Sci* **12:** 636749.

In Press

Jackson D, Tian F, Zhang Z. 2022. Maize genetics, genomics, and sustainable improvement. *Mol Breeding* **42:** 2. doi:10.1007/s11032-021-01266-5

Kim DE, Jeong J-h, Kang YM, Park Y-H, Lee Y-J, Kang J-s, Choi Y-W, Son B-G, Kim ST, Jackson D, Je BI. 2022. The impact of fasciation on maize inflorescence architecture. *J Plant Biol* doi:0.1007/s12374-021-09342-1

Kitagawa M, Wu P, Balkunde R, Cunniff P, Jackson D. 2022. An RNA exosome subunit mediates cell-to-cell trafficking of a homeobox mRNA via plasmodesmata. *Science* **14:** 177–182. doi:10.1126/science.abm0840

Klein H, Gallagher J, Demesa-Arévalo E, Abraham-Juárez MJ, Heeney M, Feil R, Lunn JE, Xiao Y, Chuck G, Whipple C, et al. 2022. Recruitment of an ancient branching program to suppress carpel development in maize flowers. *Proc Natl Acad Sci* **119:** e2115871119. doi:10.1073/pnas.2115871119

GENETIC AND MOLECULAR MECHANISMS OF STEM CELL CONTROL AND QUANTITATIVE TRAIT VARIATION IN PLANT DEVELOPMENT AND AGRICULTURE

Z.B. Lippman L. Aguirre, M. Benoit, D. Ciren, B. Fitzgerald, I. Gentile, J. He, A. Hendelman, C-T. Kwon, A. Lanctot, S. Qiao, G. Robitaille, R. Santos, B. Seman, X. Wang, S. Zebell

Our laboratory uses the biology of shoot branching, flowering, and flower production to elucidate the genetic and molecular mechanisms underlying quantitative trait variation. Our emphasis is on understanding relationships between genes and their variants, and how individual and different combinations of natural and engineering mutations influence the control of gene expression and epistatic interactions between related genes. We use tomato and related species in the Solanaceae family as our primary model systems, supplemented with the classical model *Arabidopsis thaliana*. Using our continuing establishment of chromosome-scale reference genomes in tomato and related Solanaceae species as a foundation, we use a range of genome-editing technologies to generate libraries of targeted mutations in both coding and *cis*-regulatory sequences of gene families that control the production and maturation of shoot meristems. Regenerating stem cell populations in shoot meristems is the basis for continuous development of new vegetative and reproductive shoots, flowers, and fruits. Our work seeks to identify these stem cell–regulating genes and their downstream targets, and then uses genome editing in the context of natural genetic variation to modify the function of these genes—individually and together—to engineer continuums of quantitative trait variation. Beyond our fundamental questions, we are translating our discoveries to agriculture by showing how natural and engineered genetic diversity can be used in crop breeding, including to elevate poorly studied orphan crops that are relatives of the few major crops that feed the world. By dissecting the complexity of gene–gene interactions and gene expression control, we are expanding our understanding of genotype-to-phenotype relationships in a major plant family, thereby providing new approaches and principles to allow the rational design of engineered trait variation in other plant taxa of both fundamental and agricultural importance.

cis-Regulatory Sequences, Their Interactions, and Their Evolutionary Relationships in the Control of Gene Expression and Trait Variation

D. Ciren, A. Hendelman

Multicellular organisms carry the same DNA blueprint in every cell and yet consistently develop a wide diversity of cell types. This is possible due to the differential regulation of genes. Gene regulation is often controlled by combinations of *cis*-regulatory elements (CREs) located upstream of and downstream from genes, within introns and untranslated regions (UTRs), and at sites far away. These various elements likely interact genetically and/or physically to fine-tune the pattern, timing, and level of gene expression. The main goal of this project is to study the contribution of various types of DNA elements to the *cis*-regulation of genes. Utilizing the CRISPR-Cas9 genome, we are generating targeted mutations in regions of open chromatin and sequences of deep conservation. In particular, we are interested in studying how such predicted regulatory regions located directly downstream from genes contribute to variation in gene expression, and how they may interact with upstream regulatory elements.

We are taking advantage of *CLAVATA3* (*CLV3*), a signaling peptide gene that negatively regulates meristem size and flower organ number in many plants by repressing the stem cell–promoting homeobox transcription gene *WUSCHEL* (*WUS*). We previously used CRISPR-Cas9 multiplex targeting of the promoter of

tomato *CLV3* (*SlCLV3*) to engineer more than 30 alleles that provided a continuum of variation for fruit locule number (Wang et al. 2021). To explore the contribution of downstream regulatory elements, we used the same approach on the 3′ region of *SlCLV3.* Unlike promoter targeting, only a few of these alleles resulted in subtle changes in locule number. We are currently studying the consequences of combining upstream and downstream mutations by targeting specific promoter regions in the downstream alleles, and vice versa.

In related work, we are addressing the challenge of comparing the functions of typically highly divergent *cis*-regulatory sequences between orthologous genes in different plant families. Again taking advantage of CRISPR-Cas9, we have engineered dozens of individual and combined promoter and 3′ alleles in *Arabidopsis CLV3* (*AtCLV3*). Notably, unlike in tomato, the *AtCLV3* promoter was highly buffered to large sequence perturbations, whereas deletion of specific regions downstream resulted in weak increases in carpel number. Promoter-downstream combinations may have synergistic effects on carpel number, and we are currently exploring potential epistatic interactions among our alleles and other genes involved in meristem size regulation.

In a similar approach, we have targeted *cis*-regulatory elements upstream of and downstream from *SELF PRUNING 5G* (*SP5G*), a flowering time regulator in tomato in the florigen hormone gene family. Wild tomato species are short day–sensitive, flowering faster under longer nights, whereas shorter nights at higher latitudes delay flowering. Domesticated tomato is largely insensitive to day length, a trait that was key to expanding cultivation to northern latitudes. We previously found that this loss of day length sensitivity is based on a reduction in *SP5G* expression. To study the *cis*-regulatory elements involved in *SP5G* expression, we are using an introgression line carrying the day length–sensitive *SP5G* allele from the wild tomato *Solanum pennellii*. Exploiting CRISPR-Cas9 again, we targeted multiple regions upstream of and downstream from *SP5G,* defined by open chromatin, sequence conservation, and a previously published enhancer sequence located downstream. Phenotyping of the resulting alleles in long-day conditions revealed a spectrum of flowering time variation. Future work will reveal specific elements and their interactions involved in both shared and differential regulation of *SP5G* in domesticated and wild tomato species.

Revealing and Characterizing Deep Conservation of *cis*-Regulatory Sequences in Angiosperms

A. Lanctot, A. Hendelman

We recently developed (in collaboration with I. Efroni, Hebrew University) Conservatory, an algorithm that identifies putative functional *cis*-regulatory regions based on both microsynteny and sequence conservation (Hendelman et al. 2021). Conventional approaches use whole-genome alignments to identify conserved sequence surrounding orthologous genes. However, such methods are sensitive to both lineage-specific whole-genome or segmental duplications and variable repeat content, which are rampant in plants. Conservatory addresses these issues using a gene-centric approach and generates reliable orthogroups and *cis*-regulatory comparisons within and between taxa as outputs. Briefly, Conservatory first identifies the most closely related genes among related species (e.g., genus or family) using BLASTp hits, maintaining one-to-many gene relationships, wherein a gene in one species can have one or more close homologs in another. These homologous groups are filtered to identify homologs with similarity in both their coding and *cis*-regulatory sequences based on repeat-masked local alignments. Conserved noncoding sequences among these filtered genes are detected using phyloP, an established model-based tool that identifies regions of evolutionary acceleration and conservation. In this way, Conservatory finds conservation in extended regulatory sequences and accommodates incomplete assemblies, more genomes, and polyploid complexity compared to whole-genome alignments.

We are leveraging our first data from Conservatory to study the evolution of *cis*-regulatory control of developmental genes and pathways whose functions are conserved across deep evolutionary time. One example is floral specification, which is mediated by a deeply conserved pair of genes across flowering plants, the transcription factor *LEAFY* (*LFY*) and its interacting partner F-box gene *UNUSUAL FLORAL ORGANS* (*UFO*). The functions of *LFY* and *UFO* coding sequences are interchangeable across orthologs in distantly related species, indicative of their deeply conserved functions. For example, this conserved function is present between *Arabidopsis*, petunia, and tomato, member species of two plant families, the Brassicaceae and the Solanaceae, whose evolutionary distance is

similar to that between humans and mice. Interestingly, despite this conserved function, the *cis*-regulation of these two genes is fundamentally different in the Brassicaceae and Solanaceae. In Brassicaceae species, *LFY* is not expressed until the reproductive meristem transition, whereas *UFO* is expressed throughout vegetative and reproductive meristem development. By contrast, in Solanaceae species, including tomato, the *LFY* ortholog is expressed throughout meristematic development, and *UFO* expression only begins during the transition to reproductive development to drive floral specification. In this project, we are determining how floral specification, mediated by such evolutionarily conserved genes, is subject to family-specific *cis*-regulation that causes opposite temporal expression patterns.

To address this question, we are using Conservatory to find regions of conserved noncoding sequence (CNSs) across large evolutionary distances. By identifying in-family and cross-family CNSs in the *LFY* and *UFO* promoters, we can compare how the extent of sequence conservation impacts gene expression and function. We are targeting these conserved regions for CRISPR-mediated disruption in tomato and *Arabidopsis*. By characterizing the impacts of these disruptions on floral specification between the two species, we can show the extent to which conservation (in-family vs. cross-family) affects molecular and developmental phenotypes. Additionally, we will use these CNSs to construct putative "minimal" promoters and assay these promoters' ability to complement *lfy* and *ufo* mutants. We will expand this analysis to other species within the families and to other genes involved in flower development. For instance, as described next, we have already shown the effect of disrupting CNSs of tomato *WUSCHEL HOMEOBOX 9* (*WOX9*), another floral regulator. We are currently expanding this analysis by disrupting the corresponding CNSs in *Arabidopsis*. This project will increase our knowledge of how CNSs and their divergence regulate conserved developmental pathways, despite moderating disparate expression patterns.

Conservation and Divergence of Paralog Redundancy and *cis*-Regulatory Control in the *WOX* Gene Family

A. Hendelman

A profound knowledge of the genome is essential to understanding the relationship between genotypic and phenotypic variability. Until recently, high-cost error-prone sequencing technology focused on identifying and characterizing coding regions, shedding light on conservation of gene molecular functions and their roles in growth and development. The ongoing technological revolutions in genomics are providing more and more high-quality genomes across plant taxa, allowing for the first time a more comprehensive and nuanced evaluation of conservation and divergence in the noncoding sequences of genomes, particularly *cis*-regulatory regions. Within these regions, CREs are short noncoding sequences enriched for transcription factor binding sites (TFBSs) and may be within proximal promoters or in distal upstream, downstream, or intronic regions. These CREs likely act individually and together (e.g., as part of enhancers) to specify a complex regulatory code that dictates gene expression patterns and levels. As in coding sequences, deeply conserved noncoding sequences (CNSs) and the CREs within them may reflect functional constraint through purifying selection. To identify key functional regions, we collaborated with I. Efroni (Hebrew University) to develop a new bioinformatic tool called Conservatory (Hendelman et al. 2021). Conservatory was specifically designed to address the challenge of identifying putative functional *cis*-regulatory regions based on deep sequence conservation surrounding genes.

The *WOX* family of homeobox transcription factor genes play key roles in meristem development. In our previous work, we used a multiplex CRISPR-Cas9 drive system to generate a repertoire of promoter alleles in tomato *SlWOX9,* which, along with data in *Arabidopsis*, revealed previously hidden, deeply conserved pleiotropic functions in both the *Brassicaceae* and *Solanaceae* (Hendelman et al. 2021). We further associated specific *cis*-regulatory sequences and overlapping open chromatin with specific developmental functions. In *Arabidopsis*, multiple *WOX* genes act separately and, in some cases, redundantly to control early stages of embryo development. This includes *WOX9*, its paralog *WOX8*, and a homolog *WOX2*, which function together to establish the apical–basal embryo axis. Interestingly, tomato and all other species in the *Solanaceae* family lack a *WOX8* ortholog, and, consistently, tomato *SlWOX9* CRISPR null alleles are embryonic lethal like *Arabidopsis wox9* null mutants and *wox9* weak and *wox8* strong allele combinations. Thus, whereas *WOX9* function in embryogenesis is conserved, apical–basal axis formation in *Arabidopsis wox2* and

wox8 null mutant embryos is only partially compromised, reflecting differential evolution of redundancies between the two species. Notably, we have found that CRISPR-Cas9 targeting of tomato *SlWOX2* results in embryonic lethality. Moreover, Conservatory analysis of *SlWOX2* revealed multiple CNSs in the upstream and downstream regions, and CRISPR-Cas9 multiplex targeting of specific CNSs causes embryonic lethality. We are currently comparing Conservatory analysis between *WOX* genes in the Solanaceae and Brassicaceae to reveal additional targets to study taxa-specific evolution of redundancy and *cis*-regulatory control.

Regulation of Dosage-Dependent Gene Expression and Phenotypes in *cis* and *trans*

Sophia Zebell

Improving the predictability and tunability of gene expression is crucial to engineering quantitative phenotypes. However, the complex relationship between *cis*-regulatory sequence variation, *trans*-factor binding, and phenotypic output is only beginning to be understood. To address this relationship, we have established tomato inflorescence patterning as a highly quantitative assay for expression of the dosage-dependent developmental regulator *ENHANCER OF JOINTLESS 2 (EJ2)*. Using CRISPR-Cas9 multiplex targeting of evolutionarily conserved regions of open chromatin in the promoter of the *EJ2* MADS-box gene, we have identified sequences essential for its proper expression. Alleles that perturb these sequences give rise to a quantitative spectrum of inflorescence branching phenotypes in a sensitized background carrying a null allele of a paralogous gene, and, importantly, the order of allelic severity is robust over multiple experiments.

Through a combination of motif enrichment and gene expression analysis, we identified candidate transcription factor regulators of these sequences. We have cloned more than 40 candidate transcription factors, and yeast-one-hybrid assays demonstrate that several of these transcription factors bind the sequence of interest. We are now utilizing these transcription factors in heterologous DNA-binding and transcription activation assays, as well as in vitro kinetic DNA-binding assays with our various CRISPR alleles. To increase the allelic diversity of these experiments, we performed CRISPR-Cas9 editing with single-guide RNAs to target individual motifs, as well as editing with SPRY-Cas9, which can more flexibly target the genome due to fewer sequence dependencies, and SPRY-base editing with A-to-G and C-to-T base editors. We have also identified natural alleles present in wild relatives of tomato at the same genomic loci as our edited promoter variants, which we have crossed into our variants to test for complementation in a sensitized background. The presence of possible cryptic variation in these *cis*-regulatory regions highlights the power of tomato for studying the poorly understood relationships between changes in gene regulation and quantitative trait variation. Through this work, we will draw mechanistic links between quantitative changes in DNA binding, transcriptional regulation, and phenotype.

Dissecting Mechanisms of Quantitative Epistasis in Plants

L. Aguirre, M. Benoit

Introducing genetic variation to any biological system often results in unpredictable epistatic interactions—both positive and negative—that influence phenotypic outcomes. Epistasis can be defined as a statistical interaction between two or more genetic mutations with quantitative phenotypes that leads to a deviation from strictly additive effects. However, the study of epistasis has classically been limited to pairwise interactions of genes using null mutations in genes of interest. This has led to a lack of nuance in understanding potentially widespread quantitative aspects of epistatic interactions between genes and their allelic variants.

The genetic program regulating shoot meristem development is an excellent model system in which to interrogate quantitative epistasis. As described above, stem cell proliferation in meristems is regulated by a deeply conserved and dosage-sensitive negative feedback loop between *CLV3* and *WUS*. We previously generated quantitative variation in meristem size (and flower and fruit organ number) via CRISPR-Cas9 multiplex targeting of the promoter of tomato *CLV3*. Although the fact that *slwus* null mutants act epistatically to *slclv3* might seem to indicate that the relationship between these two genes is rather straightforward, evidence to the contrary can be seen in the interactions between weak alleles of both genes, revealing the *SlCLV3* and *SlWUS* relationship is likely also dosage-dependent and more broadly sensitive

to quantitative epistastic interactions between hypomorphic alleles. In support, our previous results showed that the moderate *SlCLV3* domestication promoter inversion allele *fas* is enhanced by the presence of a weak *SlWUS cis*-regulatory domestication allele, *lc*. Taking advantage of our unique collection of *SlCLV3* promoter alleles, we have assayed quantitative epistatic relationships between these two genes and their associated natural and engineered variants. By analyzing which *SlCLV3* promoter/CR-*lc* allele combinations show divergence in phenotypic strength from single mutants (determined by quantification of locule number), we determined the presence and degree of epistatic interactions across 15 *slclv3* promoter alleles. Over two experiments (in collaboration with S. Hutton at the University of Florida), we obtained robust data that show the relationship of positive epistasis between *SlCLV3* promoter single and double mutant lines.

In addition to the canonical *SlCLV3-SlWUS* feedback circuit, meristem size in tomato is also influenced by compensation from another *CLV3-EMBRYO SURROUNDING REGION* (*CLE*) family member, *SlCLE9*. This gene is a close paralog of *SlCLV3* and is actively up-regulated to buffer stem cell homeostasis in a *slclv3* mutant background. By leveraging the same suite of *SlCLV3* promoter alleles, we determined the allelic strength at which compensation from *SlCLE9* initiates and showed that this compensation increases linearly as *slcvl3* allele severity increases. Expression analyses of all single and double mutants revealed a consistent trend of increased *SlCLE9* expression (i.e., transcriptional compensation) as *SlCLV3* expression decreases. Dissection of these interlocus responses during active compensation and other regulatory epistatic interactions has provided the opportunity to understand the quantitative essence of a robust and deeply conserved developmental program. Moreover, deeper analysis of these data (in collaboration with CSHL colleague D. McCandlish) are providing a new understanding of the polygenic, dose-dependent nature of epistatic interactions and their impact on plant evolution and crop domestication.

Dynamic Evolution of Small Signaling Peptide Compensation in Plant Stem Cell Control

C-T. Kwon, X. Wang, I. Gentile, A. Hendelman, G. Robitaille

A major goal of plant biology is to understand how interactions between genes and their variants shape phenotypic diversity. One of the most widespread forms of genetic interaction is redundancy, which in its simplest state involves one gene substituting for the function of another closely related mutated gene. Although redundancy provides robustness against genetic or environmental change, it also allows the relaxation of selection pressure, which facilitates the accumulation of mutations, and thus divergence of gene function. Despite hundreds of studies in many plant systems and biological processes, we still do not understand the dynamics of how redundancy relationships evolve in lineage-specific ways, and how such evolution impacts species-specific phenotypes.

We uncovered and dissected dynamic evolution of redundancy within a prominent family of signaling peptide genes (Kwon et al. 2022) (Fig. 1). Although the most well-studied member of this family, *CLAVATA3* (*CLV3*), is deeply conserved in its control of stem cell proliferation and crop improvement, ancestral duplications in the lineages leading to many crop species have occurred. How redundant relationships have been shaped among related species within

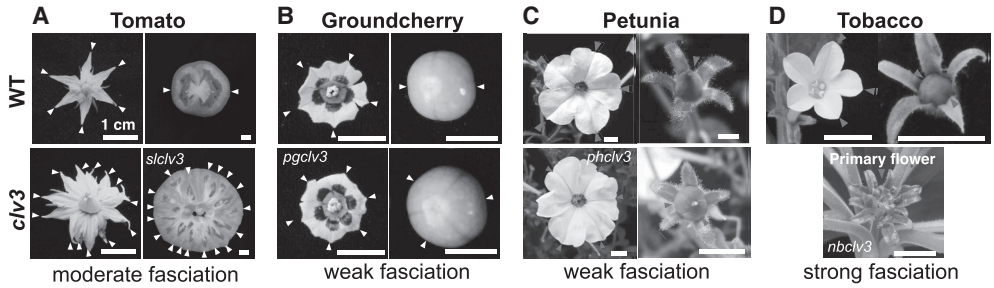


Figure 1. Dynamic evolution of CLE paralog redundancy causes species-specific clv3 phenotypes in the Solanaceae plants tomato (*A*), groundcherry (*B*), petunia (*C*), and tobacco (*D*).

these lineages, and whether they affect species-specific variation, has never been studied. We found that in tomato and its relatives in the Solanaceae family, an ancient duplication gave rise to a closely related *CLV3* paralog. We found that whereas redundancy from the paralog compensates for mutations in *CLV3,* there is remarkable variation in the potency of compensation between different Solanaceae species. Our genomic analysis of more than 25 Solanaceae revealed differential paralog loss and retention. Using genome editing in four species spanning 30 million years of Solanaceae evolution, we found (in collaboration with C. Xu, Chinese Academy of Sciences) that compensation was abolished in species that lost their paralogs. Surprisingly, we discovered that compensation varies quantitatively among species that retained their paralogs, and transcriptional and coding sequence differences in the paralogs explain these differences in compensation.

Pan-genomes are unveiling a remarkable level of paralog diversity both within and between plant species beyond presence-absence variation. Our study established a foundation for dissecting paralog relationships in other gene and plant families, which will provide insights into mechanisms of species-specific phenotypic variation. Beyond these fundamentals, our findings have immediate implications for genome editing in agriculture by showing how understanding dynamically evolving paralogs in related crop genotypes can improve predictability in trait engineering.

Establishing Two *Physalis* Species as New Solanaceae Systems Enables Genetic Reevaluation of the Inflated Calyx Syndrome

J. He

The evolution of morphological novelties has been a long-standing interest for biologists. The diverse floral organ traits among Solanaceae species have been attracting generations of scientists. 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 exhibit dramatic growth after anthesis, forming balloon-like structures that encapsulate fruits. However, studying the molecular and evolutionary mechanisms of evolutionary novelties like ICS has been confounded by the lack of genomic resources and genetic tools. This reflects a larger issue in plant biology: A limited number of genetically tractable model systems dominate and limit biological interrogation. Developing new reference species, especially relatives of established systems, is essential to understanding how evolution has shaped species-specific phenotypes.

To elevate *Physalis* as a new reference system among other Solanaceae species and to enable molecular and genetic dissection of ICS, we built (in collaboration with M. Schatz, Johns Hopkins University) high-quality chromosomal-scale reference genomes of two diploid *Physalis* species and demonstrated their value in studying multiple traits, including ICS. We first used the genomes of *P. grisea* and *P. pruinosa* to rapidly identify the gene underlying the most prominent phenotypic distinction between these species: purple versus yellow flower nectar guides. By combining mapping-by-sequencing and structural variant analysis, we discovered in *P. pruinosa* a 43-bp deletion in the basic helix-loop-helix (bHLH) transcription factor gene *AN1*, a regulator of anthocyanin biosynthetic genes first studied in petunia. This deletion results in intron inclusion and a premature stop codon. We next turned our attention to a much-needed and long-overdue reevaluation of the genes proposed to underlie ICS. Using CRISPR-Cas9 genome editing, we engineered (in collaboration with J. Van Eck, The Boyce Thompson Institute) multiple knockout mutations in two MADS-box genes previously concluded to be "master regulators" of ICS. Strikingly, although some floral defects in these mutants were observed, none prevented calyx inflation. Even more, we knocked out an additional nine MADS-box genes, and none of these affected ICS. These results provide the first concrete conclusion that this broad set of MADS-box genes are not core regulators of ICS. Finally, we carried out a forward genetics mutagenesis screen and identified a new mutant, *huskless* (*hu*), which fails to develop an inflated calyx because of the merging of sepals and petals into a single organ. Again leveraging our new genomes, we pinpointed the responsible gene and mutation to a transcriptional regulator previously shown to control floral development in *Arabidopsis*.

In summary, in addition to our continued sequencing of chromosome-scale tomato genomes (Alonge et al. 2021), we have established two *Physalis* species as new Solanaceae systems. In addition to enabling a more robust genetic dissection of ICS, our resources and tools open the door to the study of many other *Physalis*

traits of fundamental interest—for example, secondary metabolism and disease resistance—that will also help accelerate the improvement of this orphan crop.

PUBLICATIONS

Alonge M, Lebeigle L, Kirsche M, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. 2021. Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing. bioRxiv doi:10.1101/2021.11.18.469135

Hendelman A, Zebell S, Rodríguez-Leal D, Dukler N, Robitaille G, Wu X, Kostyun J, Tal L, Wang P, Bartlett ME, et al. 2021. Conserved pleiotropy of an ancient plant homeobox gene uncovered by *cis*-regulatory dissection. *Cell* **184:** 1–16. doi:10.1016/j.cell.2021.02.001

Wang X, Aguirre LM Rodríguez-Leal D, Hendelman A, Benoit M, Lippman ZB. 2021. Dissecting *cis*-regulatory control of quantitative trait variation in a plant stem cell circuit. *Nat Plants* **7:** 419–427. doi:10.1038/s41477-021-00898-x

In Press

Kwon C, Tang L, Wang X, Iacopo G, Hendelman A, Robitaille G, Van Eck J, Xu C, Lippman ZB. 2022. Dynamic evolution of small signaling peptide compensation in plant stem cell control. bioRxiv doi:10.1101/2022.01.03.474791

EPIGENETIC MECHANISMS OF GENE REGULATION AND INHERITANCE

R. Martienssen B. Berube, J. Cahn, K.H. (T.) Cheng, C. de Santis Alves Rosa, C.M. Elizalde, E. Ernst, H.S. Kim, A.A. Lakhani, S.C. Lee, J. Lynn, U. Ramu, M. Regulski, B. Roche, A. Shimada, J. Simorowski, L. Spielmann, J. Steinberg

Plants and fission yeast provide excellent models for epigenetic mechanisms of transposon regulation, gene silencing, and imprinting, important for both plant breeding and human health. We are investigating the role of RNA interference (RNAi) in epigenetic inheritance in plants, yeast, and mammalian cells. In fission yeast, we have found that RNAi promotes histone modification but is also required to resolve replication-induced transcription stress. In plants, we have shown that small RNAs in the embryo reprogram histone modifications using a histone methyltransferase previously thought to be inactive and also mediate paramutation in maize via the same Argonaute genes. We continue to investigate the regulatory landscape underlying the domestication of maize and have discovered an RNAi-based mechanism of selfish inheritance that has the potential to drive domestication as a consequence of hybridization. Our long-standing interest in the epigenetic inheritance of DNA methylation, particularly at centromeres, is beginning to yield some important findings as the DDM1 protein gives up its secrets. 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 japonica* (the humble duckweed). This year we said farewell to Atsushi Shimada, who left for a position at the Okinawa Institute of Technology in Japan. We welcomed Cristian Mateo Elizalde from Spain, who joins the duckweed team as a postdoctoral fellow, as well as visiting graduate student Asad Lakhani.

The Genetic and Epigenetic Landscape of *Arabidopsis* Centromeres

A. Shimada, E. Ernst, R.A. Martienssen [in collaboration with T. Kakutani, National Institute of Genetics, Japan; K. Schneeberger, Max Planck Institute, Köln, Germany; T. Michael, Salk Institute; M. Schatz, Johns Hopkins University; I. Henderson, Cambridge University]

Centromeres attach chromosomes to spindle microtubules during cell division and, despite this conserved role, show paradoxically rapid evolution and are typified by complex repeats. We used long-read sequencing to generate the Col-CEN genome assembly that resolves all five centromeres in *Arabidopsis thaliana*. The centromeres consist of megabase-scale tandemly repeated satellite arrays, which support CENTROMERE SPECIFIC HISTONE H3 (CENH3) occupancy and are densely DNA-methylated, with satellite variants specific to each chromosome. CENH3 preferentially occupies satellites that show the least amount of divergence and occur in higher-order repeats. The centromeres are invaded by *ATHILA* retrotransposons, which disrupt genetic and epigenetic organization. Centromeric crossover recombination is suppressed, yet low levels of meiotic DNA double-strand breaks occur that are regulated by DNA methylation. We propose that *Arabidopsis* centromeres are evolving through cycles of satellite homogenization and retrotransposon-driven diversification.

Decrease in DNA Methylation 1 (DDM1) Controls Epigenetic Inheritance by Remodeling Histone H3 Variants

S.C. Lee, B. Berube, J. Cahn, J. Lynn, R. Martienssen [in collaboration with P. Voigt, Babraham Institute, Cambridge, U.K.; D. Grimanelli, IRD Montpellier, France; L. Joshua-Tor, CSHL]

Epigenetic inheritance involves the faithful replication of DNA methylation and histone modification independent of the underlying DNA sequence and occurs during cell division and from one generation to the next. The Swi2/Snf2 chromatin remodeler DDM1 in plants, and its ortholog LSH/HELLS in mammals, is a master regulator of epigenetic inheritance that is also required for centromere function. Here, we show that DDM1 maintains DNA methylation by removal of the replication-independent histone H3.3 from heterochromatin and replacement with the replication-dependent histone H3.1. In *ddm1* mutants, the H3.3 chaperone HIRA inhibits DNA methylation, whereas the H3.1

chaperone CAF-1 becomes essential for viability. Similar to the chromatin remodeler ISWI/Snf2h, we show that DDM1 features an amino-terminal autoinhibitory domain, and that this domain promotes specific remodeling of H3.3 octamers in vitro. Live imaging of DDM1 revealed colocalization with H3 variants during the cell cycle and with the DNA methyltransferase MET1 (Dnmt1), thus promoting DNA methylation and its inheritance in subsequent generations.

Loss of DNA Methylation in the Cell Cycle Promotes Germline Reprogramming and Somaclonal Variation

F. Borges,* M. Donoghue,** C. LeBlanc,*** B. Berube, R. Martienssen [in collaboration with M. Tanurdžić, University of Queensland, Australia; W.F. Thompson and L. Hanley-Bowdoin, North Carolina State University; A. Schnittger, University of Hamburg; D. Bouyer, University of Lyon; V. Colot, École Normale Supérieure, Paris, France]; present address: * INRA, Versailles, France; **MSKCC, New York; ***Yale University

5-Methyl cytosine is widespread in plant genomes in both CG and non-CG contexts. During replication, hemimethylation on parental DNA strands guides symmetric CG methylation on nascent strands, but non-CG methylation requires modified histones and small RNA guides. We used immortalized *Arabidopsis* cell suspensions to sort replicating nuclei and determine genome-wide cytosine methylation dynamics during the plant cell cycle. We find that symmetric mCG and mCHG are selectively retained in actively dividing cells in culture, whereas mCHH is depleted. mCG becomes transiently asymmetric during S phase but is rapidly restored in G_2, whereas mCHG remains asymmetric throughout the cell cycle. Hundreds of loci gain ectopic CHG methylation, as well as 24-nt small interfering RNAs (siRNAs) and histone H3 lysine dimethylation (H3K9me2), without gaining CHH methylation. This suggests that spontaneous epialleles that arise in plant cell cultures are stably maintained by siRNA and H3K9me2 independent of the canonical RNA-directed DNA methylation (RdDM) pathway. In contrast, loci that fail to produce siRNA may be targeted for demethylation by the DNA glycosylases DEMETER (DME) and REPRESSOR OF SILENCING 1 (ROS1) when the cell cycle arrests. Comparative analysis with methylomes of various cell types in pollen and embryos suggests that loss of small-RNA-directed non-CG methylation during DNA replication promotes germline reprogramming and epigenetic variation in plants propagated as clones. We have now shown that ROS1 and DME1 act semiredundantly in the vegetative cell of pollen to demethylate DNA after cell cycle arrest and ensure proper pollen tube progression. Moreover, we have identified six pollen-specific genes with increased DNA methylation as well as reduced expression in *dme* and *dme;ros1* mutant pollen. For four of these genes, reinstating their expression individually in mutant pollen improves male fertility. Our findings demonstrate an essential role of active DNA demethylation in regulating genes involved in pollen function.

Small RNAs Guide Histone Methylation in *Arabidopsis* Embryos and Paramutation in Maize

J-S. Parent,* J. Cahn, R. Herridge,** R. Martienssen [in collaboration with J. Aubert, University of Montpellier; F. Bellegarde, Nagoya University; M.A. Arteaga-Vázquez, Universidad Veracruzana, Mexico; D. Grimanelli, IRD Montpellier, France]; present address: *Agriculture Canada, Ottawa; **University of Otago, New Zealand

Epigenetic reprogramming occurs during gametogenesis as well as during embryogenesis to reset the genome for early development. In flowering plants, many heterochromatic marks are maintained in sperm, but asymmetric DNA methylation is mostly lost. Asymmetric DNA methylation is dependent on small RNA, but the reestablishment of silencing in embryo is not well understood. We have shown that 24-nt small RNAs, bound to Argonautes AGO4, AGO6, and AGO9, direct histone H3 lysine 9 dimethylation during *A. thaliana* embryonic development (Fig. 1), resembling the fission yeast *Schizosaccharomyces pombe* in this respect, together with asymmetric DNA methylation. This de novo silencing mechanism depends on the catalytic domain of SUVH9, a Su(Var)3-9 homolog previously thought to be catalytically inactive. Paramutation is an epigenetic phenomenon in which epigenetic information is conserved over many generations through mitosis and meiosis. It requires components of the RdDM pathway in the early embryo. However, no mediators of paramutation have been identified in the effector complex of RdDM. Through a combination of reverse genetics, immunolocalization, and immunoprecipitation (siRNA-IP) we found that AGO104, AGO105, and AGO119 (homologs of AGO9, AGO6, and AGO4 in *Arabidopsis*) are members of the RdDM effector complex in maize and bind siRNAs produced

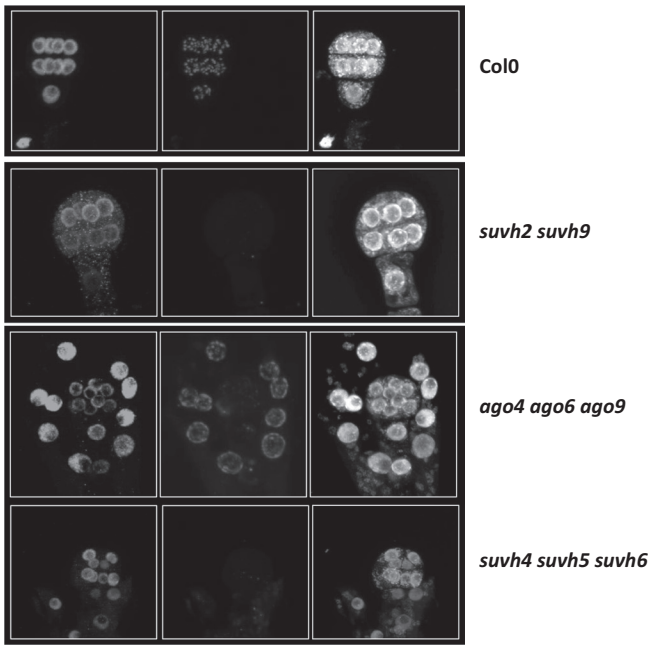


Figure 1. Immunofluorescent confocal scans of eight-cell *Arabidopsis* embryos for the indicated genotypes stained with antibodies to H3K4me1 (*left*), H3K9me2 (*center*), and with DAPI (DNA, *right*). Signal level of each image has been adjusted to the H3K4me1 channel. H3K9me2 is lost from *suvh2 suvh9* double mutants, *suvh4 suvh5 suvh6* triple mutants, and the embryo proper in *ago4 ago6 ago9* triple mutants, but not from somatic cells. (Modified from Parent et al. 2021.)

from the tandem repeats required for paramutation at the *b1* locus. Genetic analysis revealed that AGO104 is also required to enable paramutation.

Maizecode: DNA Regulatory Elements in Maize and Teosinte Provide Insight Into Maize Domestication

J. Cahn, M. Regulski, J. Lynn, E. Ernst, B. Berube, R. Martienssen [in collaboration with K. Birnbaum, NYU; M. Schatz, Johns Hopkins University; D. Micklos, W.R. McCombie, D. Ware, D. Jackson, and T. Gingeras, CSHL]

Early maize lines were domesticated from *Teosinte parviglumis (Zea mays parviglumis)*, with subsequent introgressions from neighboring *Teosinte mexicana*. Domestication traits in modern maize include increased kernel row number, loss of the hard fruit case, and dissociation from the cob upon maturity, as well as fewer tillers. Molecular approaches have identified several transcription factors involved in the development of these traits. However, these studies have also shown that a more complex regulatory network is responsible for these strong morphological differences than originally hypothesized, and our understanding of the tissue-specific regulation as well as of its variability across inbred lines is still lacking. We are investigating the transcriptional regulation that resulted from the domestication process, focusing on conservation and variability across multiple tissues and inbred lines. We generated histone-modification and transcription factor chromatin immunoprecipitation sequencing (ChIP-seq) in parallel with transcriptomics data sets in up to five different tissues of three inbred lines that span the phenotypic diversity of maize inbreds, as well as the teosinte inbred TIL11. We developed an automated computational pipeline to integrate these data sets as well as publicly available data. This pipeline generates metrics and outputs for both quality control and functional analyses, and it can also be applied to other species. We identified regulatory regions that emerged during the domestication process and are responsible for the tissue-specific expression of developmental genes. We show that, even though pollen grains are the most differentiated tissue on a transcriptomic level, and especially with respect to the regulation of transposable elements, ears show the least conservation between maize and teosinte, corroborating the very distinct morphological and physiological features. We hope to provide the maize community with a framework for a collaborative effort that follows the footsteps of the ENCODE project in

order to better understand and potentially improve the regulatory landscape of the maize genome.

RNA Interference and the Drive for Maize Domestication.

B. Berube, J. Cahn, R. Martienssen [in collaboration with J. Kermicle, University of Wisconsin, Madison; J. Ross-Ibarra, University of California at Davis]

Meiotic drivers subvert Mendelian heredity through targeted exploitation of the haploid stage, biasing their own transmission at the expense of competing gametes. The majority of drive complexes are hypothesized to exist in cryptic states, making it difficult to assess the full extent to which these selfish genetic elements impact natural populations. However, hybridization can lead to the production of naïve individuals where drive is unleashed. Consequently, cryptic drivers likely play important roles in reinforcing reproductive isolation and influencing patterns of gene flow between populations. We report the discovery and characterization of teosinte pollen drive (TPD), a selfish genetic system that occurs in hybrids between maize (*Z. mays ssp. mays*) and teosinte (*Z. mays ssp. mexicana*). Using direct sequencing of individual pollen grains as well as long-read de novo genome assembly, we define a *mexicana* inversion haplotype that causes severe transmission ratio distortion. Furthermore, we demonstrate that promiscuous RNAi plays a critical role in mediating this phenotype through ectopic silencing of a lipase gene that is essential for male fertility. The rapidly evolving nature of germline noncoding RNAs and the blurring of "self" versus "non-self" in outcrossing populations suggest that RNAi plays a broad and underappreciated role in potentiating genetic conflict.

Biological Design of *Lemnaceae* Aquatic Plants for Biodiesel Production

E. Ernst, C.M. Elizalde, J. Simorowski, U. Ramu, J. Lynn, R. Martienssen [in collaboration with T. Michael, Salk Institute; J. Birchler, U Missouri; E. Lam, Rutgers University; J. Schwender and John Shanklin, Brookhaven National Laboratory]

Lemnaceae species (commonly called duckweeds) are the world's smallest aquatic flowering plants. We have produced eight new reference quality *Lemnaceae* genome assemblies of *Wolffia australiana*, *Lemna gibba*, *Lemna minor, Lemna turionifera*, and three allotriploid hybrid *L. japonica (L. minor x L. turionifera)* clones using Oxford Nanopore long reads and Hi-C contact maps. Comparisons of the chromosome-scale assemblies demonstrate a high degree of synteny across all 21 chromosomes within the *Lemna* genus, with only a single translocation consistently identified specifically in *L. turionifera*, while *W. australiana* has 20 chromosomes like *Spirodela polyrhiza*, yet with significant architectural differences. Antibodies against centromeric histone H3 were raised for four species to determine centromere organization and identify centromeric repeat sequences across the duckweeds via CUT&RUN. The sequenced species encode between 14,000 and 20,000 genes, significantly fewer than terrestrial monocots such as rice and *Brachypodium*, and comparable to the algae *Chlamydomonas reinhardtii*. Methylome and small RNA sequencing revealed dramatic differences between the three genera, consistent with known pathways of RNA-directed DNA methylation. Analysis of missing and diverged orthogroups revealed variations that likely account for reduced morphology, aquatic habitat, clonal reproduction, dormancy, high photosynthetic rate, and lipid production.

To produce feedstocks for biofuel production, we expressed WRINKLED1 (WRI1), the master transcriptional activator of fatty acid synthesis in *L. japonica* line Lj8627. This resulted in <1% of DW of TAG along with large reduction in growth rate and significant developmental abnormalities. Next, we constructed a cyan fluorescent protein (CFP) amino-terminally tagged version of the *Arabidopsis* WRI1 under the control of estradiol inducible XVE promoter, coexpressed with a sesame Oleosin 1 gene variant in which its degradation signals had been minimized to optimize its TAG protective function, along with a very strong mammalian DGAT2 to create Lj8627-33 (ODW) transgenics. Growth of OWD transgenics cultured in the presence of 100 μM estradiol for four days resulted in the accumulation of 16.4% total fatty acid by DW compared to 5.2% in the parental line and 8.7% TAG per DW compared to 0.07 (124-fold increase). Subsequent detailed analyses confirmed accumulation of very long chain fatty acids in TAG, but revealed that estradiol induction of WRI1 in ODW lines results in an allocation tradeoff of starch to TAG and a reduction in growth rate. We will evaluate the extent of lipid futile cycling and model biomass synthesis using physiological, biochemical, and transcriptomic data to understand the growth rate reduction and inform strategies for mitigation.

Living Fossils: The *Wollemia nobilis* MegaGenome

C. de Santis Alves, R. Martienssen [in collaboration with G. Coruzzi, NYU; B. Ambrose and D. Stevenson, NYBG; S.O. Kolokotronis, SUNY Downstate, Brooklyn, New York; L. Croft and M. Richardson, Deakin University, Victoria, Australia; M. Schatz, Johns Hopkins University; W.R. McCombie, CSHL]

Gymnosperms were the first seed plants, and the "living fossil" *Wollemia nobilis*, a member of the *Araucariaceae* family, is a critically endangered gymnosperm that appeared in the Cretaceous period, remaining morphologically unchanged until its rediscovery in the Blue Mountains of Australia in 1994. With the advent of long-read sequencing to improve the assembly of large genomes with a high content of repetitive elements, the New York Plant Genome Consortium has generated a reference-quality genome of *Wollemia* (~12 Gb, N50 ~ 9 Mb). We used RNA-seq and small RNA-seq from different tissues to comprehensively annotate approximately 48,000 genes, approximately 62,000 transposable elements, and approximately 500,000 small RNAs. In pollen, 75% of the sRNA content is composed of 24-nt siRNAs, with 277 microRNAs, of which 38 are conserved, but in leaves and ovules, 21-nt *R*-gene phasiRNA overaccumulates because of miRNA targeting. There are eight *DICER-LIKE* and 11 *ARGONAUTE* genes. We have generated whole-genome bisulfite data from ovules and pollen and found 78% (CG), 65% (CHG), and 2.6% (CHH) levels of methylation in both tissues, mostly in transposable elements that have abundant small RNA, indicating deep conservation of the RNA-dependent DNA methylation pathway in the most ancient extant pollen and seeds. Our efforts to integrate the data will allow a better understanding of the species' genomic resilience to environmental changes.

Clr4^{SUV39H1} and Bdf2^{BRD4} Ubiquitination Mediate RNAi Transcriptional Silencing via Heterochromatic Phase Transitions

H-S. Kim, M. Gutbrod,* B. Roche, R. Martienssen [in collaboration with A. Schorn and C. Hammell, CSHL; A. Verdel, University of Grenoble, France]; present address: *MIT

Heterochromatin has crucial roles in genome stability by silencing repetitive DNA and by suppressing recombination. In the fission yeast *S. pombe*, Clr4 methylates histone H3 lysine 9 (H3K9) to maintain centromeric heterochromatin during the cell cycle, guided by RNAi. Clr4 interacts with a specialized ubiquitination complex called CLRC, which is recruited in part by the RNAi transcriptional silencing complex (RITS), but otherwise the role of CLRC is unknown. We have found that the E2 ubiquitin conjugating enzyme Ubc4 interacts with CLRC and targets Clr4 for mono-ubiquitination at an intrinsically disordered region. Mono-ubiquitination is essential for the transition from cotranscriptional (H3K9me2) to transcriptional (H3K9me3) gene silencing by transiently releasing Clr4 from centromeric heterochromatin and is reversed by the deubiquitinating enzyme Ubp3. Addition of RNA changes the liquid–liquid phase separation (LLPS) of Clr4 to enhance its dynamic activity. Ubc4-CLRC also polyubiquitinates Bdf2, the homolog of BET family double bromodomain protein BRD4, for degradation, which otherwise accumulates at centromeres to promote transcription and production of Dicer-dependent small RNAs. In mammals, Dicer-deficient embryonic stem cells have strong proliferation and chromosome segregation defects as well as increased transcription of centromeric satellite repeats, which triggers the interferon response. We conducted a CRISPR-Cas9 genetic screen to restore viability and identified transcriptional activators, histone H3K9 methyltransferases, and chromosome segregation factors as suppressors, resembling Dicer suppressors we previously identified in fission yeast. The strongest suppressors were mutations in the transcriptional coactivator Brd4, which reversed the strand-specific transcription of major satellite repeats and thereby suppressed the interferon response. We show that identical mutations in the second bromodomain of Brd4 rescue Dicer-dependent silencing and chromosome segregation defects in both mammalian cells and fission yeast. This remarkable conservation demonstrates that RNA interference has an ancient role in transcriptional silencing of satellite repeats, which is essential for cell cycle progression and proper chromosome segregation. Our results have pharmacological implications for cancer and autoimmune diseases characterized by unregulated transcription of satellite repeats.

Nuclear RNAi in *S. pombe* Mediates Transcription-Induced Replication Stress

K.H. Cheng, B. Roche, R. Martienssen

Gene regulation by RNA interference is usually attributed to microRNA, but RNAi has a more ancient

and fundamental role in heterochromatic silencing and genome stability. Using *S. pombe*—which lacks miRNA—as a model organism, we discovered a chromatin function of Dcr1 (Dicer) in mediating transcription-replication conflicts. We found that in cellular environments where R-loops accumulate, *dcr1*Δ mutants display severe sensitivity to low doses of genotoxic drugs. Deletion of genes known to suppress the phenotype of *dcr1*Δ, many being involved in transcriptional processes, also rescue the replication stress phenotype. We also performed a mutagenesis screen and identified additional alleles linked to the regulation of the early steps of the transcriptional cycle. Alleles of *med20*, encoding a subunit of the Mediator complex, stand out as the strongest suppressors. Deletion of *Med20* is known to cause a mild RNAi-like phenotype, such as loss of centromeric H3K9 methylation and therefore silencing. The *med20* alleles isolated, however, do not compromise centromere function while still rescuing *dcr1*Δ phenotypes in replication stress. We performed RNA-seq and precision run-on sequencing (PRO-seq), and the results implicate Dcr1 as having a direct chromatin role beyond the centromere, modulating transcriptional firing and termination, especially in highly expressed genes. Ongoing work includes the identification of causal Dcr1-dependent genome instability hotspots and further characterization of the type(s) of DNA damage mediated by Dcr1.

PUBLICATIONS

Borges F, Donoghue MTA, LeBlanc C, Wear EE, Tanurdžić M, Berube B, Brooks A, Thompson WF, Hanley-Bowdoin L, Martienssen RA. 2021. Loss of small-RNA-directed DNA methylation in the plant cell cycle promotes germline reprogramming and somaclonal variation. *Curr Biol* **31:** 591–600.e4. doi:10.1016/j.cub.2020.10.098

Khouider S, Borges F, LeBlanc C, Ungru A, Schnittger A, Martienssen R, Colot V, Bouyer D. 2021. Male fertility in *Arabidopsis* requires active DNA demethylation of genes that control pollen tube function. *Nat Commun* **12:** 410. doi:10.1038/s41467-020-20606-1

Kim HS, Roche B, Bhattacharjee S, Todeschini T, Chang AY, Hammell C, Verdel A, Martienssen RA. 2021. Clr4^{SUV39H1} and Bdf2^{BRD4} ubiquitination mediate transcriptional silencing via heterochromatic phase transitions. bioRxiv doi:10.1101/2021.01.08.425919

Lee SC, Martienssen R. 2021. Phase separation in plant miRNA processing. *Nat Cell Biol* **23:** 5–6. doi:10.1038/s41556-020-00614-5

Lee SC, Martienssen RA. 2021. Regulation of retrotransposition in *Arabidopsis*. *Biochem Soc Trans* **49:** 2241–2251.

Mo Z, Scheben A, Steinberg J, Siepel A, Martienssen RA. 2021. Circadian immunity, sunrise time and the seasonality of respiratory infections. medRxiv doi:10.1101/2021.03.29.21254556

Naish M, Alonge M, Wlodzimierz P, Tock AJ, Abramson BW, Schmücker A, Mandáková T, Jamge B, Lambing C, Kuo P, et al. 2021. The genetic and epigenetic landscape of the *Arabidopsis* centromeres. *Science* **374:** eabi7489.

Parent JS, Cahn J, Herridge RP, Grimanelli D, Martienssen RA. 2021. Small RNAs guide histone methylation in *Arabidopsis* embryos. *Genes Dev* 35: 841–846.

In Press

Aubert J, Bellegarde F, Oltehua-Lopez O, Leblanc O, Arteaga-Vázquez MA, Martienssen RA, Grimanelli D. 2022. AGO104 is an RdDM effector of paramutation at the maize *b1* locus. *PloS ONE* (in press).

Gutbrod MJ, Roche B, Steinberg JI, Lakhani AA, Chang K, Schorn AJ, Martienssen RA. 2022. Dicer promotes genome stability via the bromodomain transcriptional co-activator BRD4. *Nat Commun* (in press).

Liang Y, Yu X, Anaokar S, Shi H, Dahl W, Cai Y, Lo G, Chai J, Cai Y, Molla-Morales A, et al. 2022. Engineering triacylglycerol accumulation in duckweed (*Lemna japonica*). *Plant Biotech* (in press).

MOLECULAR SIGNALING EVENTS UNDERLYING ENVIRONMENTAL CONTROL OF PLANT GROWTH

U. Pedmale A. Ackermann, J. Arber, O. Artz, Y. Hu, J. Micko, D. Rosado, O. Spassibojko

All organisms grow by increasing in size or by cell division during their lifetime. Growth depends on the input from various exterior and interior signals perceived by the organism. Unchecked growth often leads to cancer and other developmental defects, severely affecting an organism's fitness. Furthermore, a fundamental question in biology, which remains unanswered, is how the environment or exogenous signals regulate the growth and development of an organism. Both plants and animals interact with their environment. However, plants grow postembryonically as they are incapable of moving around. Unlike animals, plants have no specific organs that see or hear various stimuli, yet plants are sensitive to their surrounding environment and modify their growth according to different external and internal signals. Plants regularly face variability in growth conditions—temperature, light quality and quantity, herbivores, pathogens, water availability, etc. Plants respond to these biotic and abiotic factors and survive substantial environmental fluctuations. Plants also must balance the range of potential threats and benefits confronting them and make appropriate resource allocation decisions. Remarkably, lacking a brain, plants can successfully integrate various cues and make the right 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, productivity decreases (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 many environmental signals could help develop crops that cope with unfavorable growth conditions without significant changes in yield. Our laboratory primarily studies the effect of light environments 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 and 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. For example, 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) because of absorption of red light by chlorophyll and reflection of far-red light by the neighboring foliage. Simultaneously, there is also a 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 also impacts human health, which could make this field appealing to diverse funding agencies. CRYs regulate growth and development and provide circadian entrainment to 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 essential as they are integral to the growth of plants but also can have an impact on human health.

How Does the Above-Ground Shoot Control Below-Ground Roots?

D. Rosado, L.N. Lindbäck, K. Schwartz [in collaboration with Joseph Noel, Salk Institute; Joanne Chory, Salk Institute and HHMI; Jose Dinneny, Stanford University; Magdalena Rossi, Universidade de São Paulo, Brazil]

Roots and shoots live in different environments, yet the root knows when the shoot faces adverse environmental conditions. During vegetational shading many aerial organs elongate rapidly, whereas root growth is reduced with the delay in emergence of the lateral roots (Fig. 1). 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, which in turn leads to unsupported roots and thereby 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. Additionally, it gives us the ability to explore 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* and tomato seedlings exposed to 30 min to 5 days of shade. In addition, we generated a time-course single-cell RNA sequencing (scRNA-seq) on shade-exposed roots and mock-shaded plants (Fig. 2). We found that stress-induced genes—specifically genes induced by defense response against pathogens—were upregulated in shaded roots compared to those grown under nonshading conditions. To validate this observation, we monitored genes induced by defense responses against pathogens, which did confirm induction to shade. Furthermore, we discovered that mitogen-activated protein kinase (MAPK) signaling was activated in shade, which usually responds to abiotic and biotic stress. 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

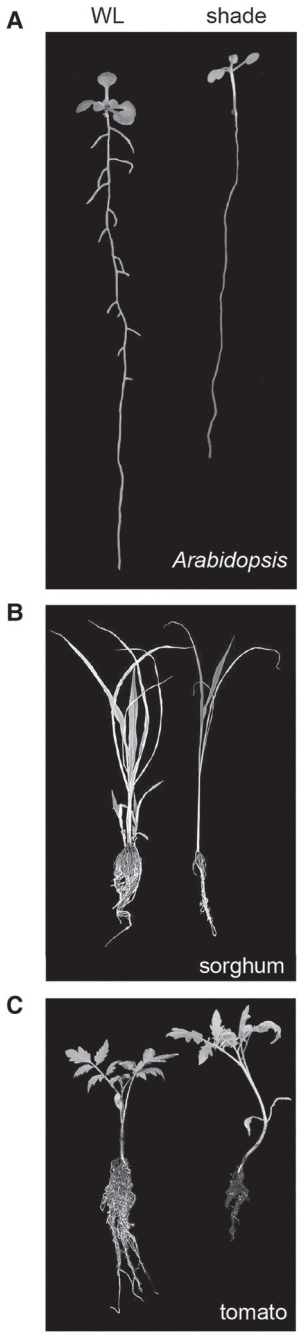


Figure 1. Vegetative shade perceived by the above-ground shoots restricts below-ground root growth in *Arabidopsis* (*A*), sorghum (*B*), and tomato (*C*) and decreases overall fitness of the plants. (WL) white light.

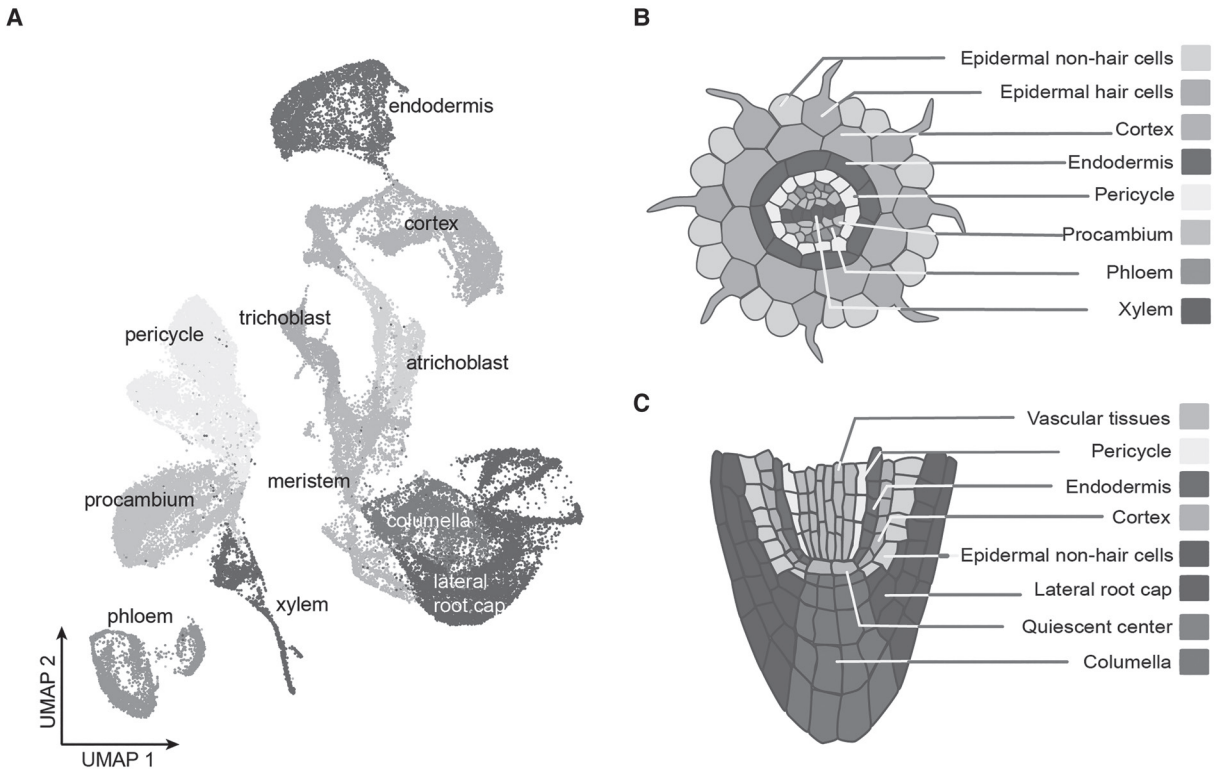


Figure 2. Identification of major cell types in *Arabidopsis* roots under shade and control light with single-cell RNA-seq (scRNA-seq). (*A*) UMAP dimensionality reduction plot of 57,649 root cells using scRNA-seq expression data. Each dot represents a unique cell and is colored according to its cell type, which in turn corresponds to one or more Louvain clusters. (*B,C*) Cartoon representing major cell types of the *Arabidopsis* root colored according to *A*, as seen in longitudinal (*B*) and transversal (*C*) views. Illustrations were adapted from the Plant Illustrations Repository (2017) (https://figshare.com/authors/Plant_Illustrations/3773596).

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. Ethyl methanesulfonate (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, which does 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 in order 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.

WRKY Transcription Factors and Ethylene Signaling Modify Root Growth during the Shade-Avoidance Response

D. Rosado, A. Ackermann, O. Spassibojko [in collaboration with Dr. Magdalena Rossi, Universidade de São Paulo, Brazil]

Plants are able to detect whether they are in the presence of neighboring vegetation or under a canopy that could threaten their access to sunlight. To avoid shading by neighbors, plants activate morphological responses such as fast stem and petiole elongation in order to outgrow competitors and better access sunlight. To allow rapid elongation of aerial organs, resources are reallocated from other parts of the plant, such as roots and harvestable organs, to the stems, resulting in lower crop yield and overall biomass. This reallocation to support stem and shoot elongation does help the plant to be competitive, but drastically

reduces and alters the growth of the roots. Roots of shaded plants are shorter and present fewer lateral roots, characteristics that limit plants' ability to assimilate nutrients and water from the soil. As consequence, the poorly developed root system is unable to support whole-plant growth, which might result in yield losses for crops cultivated in high density. At the same time, allocation of resources toward growth reduces shoot investment in defense, leaving plants more susceptible to diseases and herbivory. For this reason, shade-avoidance response is one of the biggest barriers to how densely crops can be planted in the field and has a major impact on food security when the population is increasing and the arable land is decreasing.

Although the mechanisms that govern shoot growth in response to shade are well known, little is known about the regulation of root development under such conditions. In the shoot, photoreceptor-mediated cascades activate growth-promoting genes, such as those involved in auxin hormone biosynthesis and cell-wall remodeling enzymes, which are both required for rapid stem elongation. How roots can sense when the shoots are exposed to suboptimal conditions and adapt their growth rhythms accordingly remains elusive. A time-course RNA-seq analysis performed by our group demonstrated that roots of tomato and *Arabidopsis* do not activate auxin and other growth pathways. Instead, stress- and defense-related genes are highly induced in response to shade. Indeed, further analysis of the transcriptomic data revealed that a group of 1,175 genes that are coexpressed and induced along the time course have promoters enriched in W-box motifs. Such elements are recognized by WRKY transcription factors, which have been characterized as regulators of defense and stress responses. In our data set, a group of 12 WRKY genes were found to be consistently upregulated in shaded roots throughout our time course, raising the question of whether these transcription factors are mediators of shade responses in roots. To address this possibility, we are generating loss-of-function mutants via CRISPR-Cas9 and overexpression lines. Our current research shows that overexpression of *WRKY26* and *WRKY45* results in constitutively short roots, phenocopying WT *Arabidopsis* seedlings exposed to shade. In shoots, on the other hand, shade normally induced hypocotyl elongation in the overexpression lines, suggesting that the role of these WRKYs is limited to the roots. Further analyses of the new lines and WRKY-regulated pathways will bring us closer to understanding how root development is regulated under shade perceived by the shoots.

The Role of RNA-Binding Proteins in Cryptochrome-Mediated Signaling

O. Artz, A. Ackermann [in collaboration with P. Koo, CSHL]

CRY2 elicits developmental responses through direct interaction with target proteins such as transcription factors. We recently discovered that CRY2 also interacts with various RNA-binding proteins. Among those is a protein that specifically binds to RNA modified by N6-methyl-adenosine (m6A). m6A is the most abundant internal mRNA modification and is highly conserved among eukaryotes. Transcript m6A modification greatly influences diverse aspects of RNA metabolism such as RNA structure, maturation, processing, export, and translation. Many reports have documented the relationship of human diseases such as cancer or obesity to m6A; also, in plants m6A has been associated with various developmental processes. Our genetic analyses in *Arabidopsis* mutants provide strong evidence that CRY2 and the m6A reader act in the same pathway, thus suggesting that CRY2 might control transcript m6A modification directly. We have performed transcriptome-wide profiling of m6A-modified RNA (MeRIP-seq) to identify differentially modified transcripts (Fig. 3). Our preliminary data suggest that CRY photoreceptors might play an important role in controlling transcript methylation. In our data set, m6A was mainly deposited within the context of a UGUA motif, which is

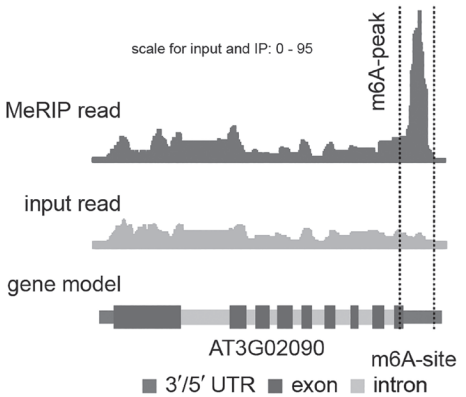


Figure 3. MeRIP-seq reads showing enrichment of m6A modification on a representative gene in *Arabidopsis* and the control input reads.

important for alternative polyadenylation. Interestingly, we find that m6A-modified transcripts show a higher preference for proximal polyadenylation sites. We provide first evidence that CRY-mediated light signaling might influence transcript m6A modification, thus controlling alternative polyadenylation. This might present a novel mechanism for gene regulation by environmental signals. Because of the high conservation of m6A, our discoveries will likely make an impact on our understanding of this newly emerging, dynamic RNA modification beyond the field of plant biology.

UBP12/UBP13 Deubiquitinases Destabilize CRY2 to Regulate Growth

L. Lindbäck, O. Artz, A. Ackermann

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 (BL). Therefore, it is obvious that the CRY2 protein level and activity are tightly regulated by light in order 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 the CRY signaling pathway, our laboratory has purified CRY2-containing protein complexes from *Arabidopsis thaliana* seedlings exposed to low intensity of 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.

Some 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 (DUB) mutants and in plants where they are overexpressed. Surprisingly, we found that CRY2 protein levels were very high in the deubiquitinase mutant and lower when overexpressed. This matches with the physiological response exhibited by the seedling stem length of these genetic backgrounds. The deubiquitinase mutant seedlings had a short hypocotyl when compared to the wild type and the *cry2* mutant, and the overexpression transgenic lines had a longer hypocotyl similar to *cry2.* Unexpectedly and differing from the current documented belief, the critical function of UBP12/13 DUBs in this process, however, turned out not to depend on its deubiquitination activity to stabilize and prevent CRY2 from degradation. But instead, these DUBs used their influence to decrease CRY2 protein level by ubiquitin-mediated degradation. Hypocotyl growth was disrupted in seedlings lacking UBP12 and UBP13 or when they were overexpressed, specifically in blue light. Our combined genetic and molecular data support a mechanistic model in which photoactivated CRY2 interacts directly with UBP12/13 DUBs, and this complex recruits constitutive

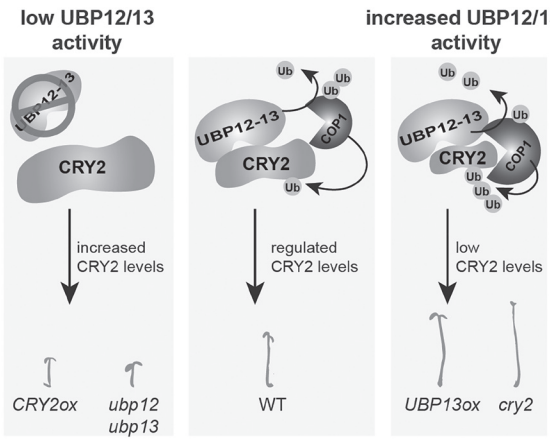


Figure 4. Model illustrating the role of UBP12 and UBP13 in regulating CRY2 protein and hypocotyl growth in blue light. During optimal photomorphogenesis (*middle* panel), CRY2-UBP12/13 complex recruits COP1 ubiquitin ligase and leads to its stabilization. This stabilized COP1 then targets CRY2 for ubiquitin-mediated proteasomal degradation, leading to optimal hypocotyl growth. Low UBP12/13 activity seen in the *ubp12ubp13* mutant increases CRY2 levels because of lack of proteasomal degradation. Therefore, the hypocotyl phenotype of *ubp12ubp13* resembles *CRY2ox* (*left* panel). Accordingly, increased UBP12/13 activity results in enhanced ubiquitination and degradation of CRY2, as *UBP13ox* seedlings have a blue light–insensitive phenotype similar to the *cry2* mutant (*right* panel).

phetomorphogenic1 (COP1) by direct contact (Fig. 4). Furthermore, UBP12/13-mediated deubiquitination of COP1 leads to its stabilization, which then promotes ubiquitination and degradation of CRY2 in blue light. This mechanism of attenuation of CRY2 is unusual among the reported mechanisms, but likely typifies the receptor's mitigation in an ever-changing light environment of the plant. Such regulation by the CRY2-UBP12/13-COP1 axis is particularly essential to optimize hypocotyl growth during photomorphogenesis.

Unlike animals, substrates for the large number (~64) 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.

Cryptochromes Mediate Repair of Damaged DNA

Y. Hu, L. Lindbäck, J. Micko

Cryptochromes evolved from DNA photolyase, which uses the energy from blue/UV-A light to repair DNA damage. However, present-day cryptochromes have lost the ability to repair DNA damage directly. In *Arabidopsis*, we found that the *cry1cry2* mutant was more sensitive to genotoxic stress (such as UV-C) than the wild-type (WT) plants (Fig. 5). This indicates that CRYs might positively affect DNA damage response indirectly or directly. We discovered that CRY2 formed punctate nuclear speckles in UV-C, similar to its function in blue light. To further explore the underlying mechanism of how CRYs regulate plant UV response, we performed a time-course whole-genome transcriptome study in the WT and *cry1cry2* double mutant, with and without UV-C treatment. From the analysis of the transcriptome data, we found that the flavonoid biosynthesis pathway was down-regulated in *cry1cry2* before and after UV-C exposure. Flavonoids function as a sunscreen by absorbing harmful UV radiation and also scavenge reactive oxygen species (ROS), which are generally generated by genotoxic stress. The *cry1cry2* double mutant contains less of these flavonoid compounds as they accumulate more cyclobutane pyrimidine dimers (CPDs) formed during the DNA damage. This down-regulation of the flavonoid biosynthesis pathway in *cry1cry2* likely contributes to the UV-C sensitivity phenotype of *cry1cry2*.

Further analysis of the transcriptome data showed that the activation of DNA damage–responsive genes after UV-C exposure was delayed in *cry1cry2* compared to the WT. This further indicates that the *cry1-cry2* mutant is deficient in activating the transcription of DNA damage–responsive genes after UV-C

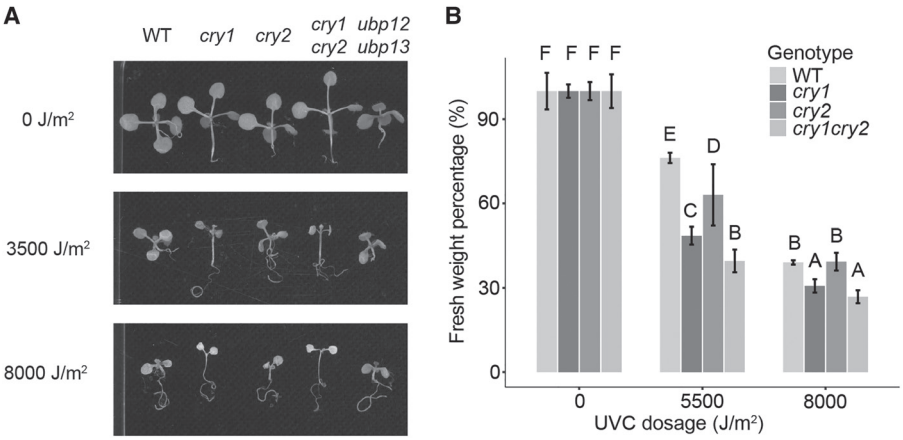


Figure 5. *Arabidopsis* plants lacking CRY1 and CRY2 are more sensitive to UV-C-induced DNA damage as measured by fresh weight postexposure (*A,B*).

exposure. Further analysis of our RNA-seq data showed that the HY5 and other calcium-binding transcription factors are essential for activating a subset of the UV-C-responsive genes. Experiments are being performed to study whether HY5 acts downstream of CRYs to regulate the transcription of DNA damage–responsive genes after UV-C exposure.

PUBLICATIONS

Lindbäck LN, Artz O, Ackermann A, Pedmale UV. 2021. UBP12 and UBP13 deubiquitinases destabilize the CRY2 blue-light receptor to regulate growth. bioRxiv doi:10.1101/2021.04.29.441934

Rosado D, Ackermann A, Spassibojko O, Rossi M, Pedmale UV. 2021. WRKY transcription factors and ethylene signaling modify root growth during the shade-avoidance response. *Plant Physiol* **188:** 1294–1311. doi:10.1093/plphys/kiab493

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, these advances also created a set of unique bioinformatics challenges for processing, integrating and interpreting the 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. They 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 the group is functional annotation of the noncoding genome via integration of multi-omics data generated by ENCODE, Roadmap Epigenomics and GTeX consortia, essential for deciphering gene regulation mechanisms, interpreting of disease-associated variants in genome-wide association studies (GWASs), and understanding epigenetic effects in cancer biology.

A dominant interest within computational biology and the **Jesse Gillis** laboratory 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. 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. 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 understanding cell biology.

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 RNAs, 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 maps of where and what types of RNA are transcribed serve as a foundation for many areas of disease diagnosis and treatment. In particular, non–protein coding

(nc)RNAs are responsible for the expression of a regulatory gene called *PTEN*. In endometrial cancer dysregulation of PTEN via overexpression of a lncRNA (HOTAIR) that binds to PTEN occurs in obese individuals. This dysregulation results in adipogenesis and inappropriate lipid metabolism leading to endometrial cancer.

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 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 laboratory 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 laboratory 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 laboratory 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.

NOVEL COMPUTATIONAL ALGORITHMS FOR SINGLE-CELL FUNCTIONAL GENOMICS

A. Dobin R. Assiimwe J. Hinds B. Kaminow

Heterogeneity of Cancer-Associated Fibroblasts in Pancreatic Tumors

This work was done in collaboration with Y. Park, J. Thalappillil, G. Caligiuri, M. Shakiba, and S. Nadella at the Tuveson laboratory (CSHL).

We analyzed the human patient and mouse model pancreatic ductal adenocarcinoma (PDAC) single-cell RNA sequencing (scRNA-seq) data generated by the Tuveson laboratory using STARsolo and identified various cell types in the tumor microenvironment, as well as subtypes of cancer-associated fibroblasts. One of the most crucial components of scRNA-seq data analysis is the classification of each cell into its characteristic type. This is typically done via unsupervised cell clustering and differential expression of marker genes, which is a manual time-consuming process with poor reproducibility. For the cancer-associated fibroblast (CAF) subtypes, this approach is especially problematic because of the sparsity of well-established marker genes, scarcity of the data sets that contain CAFs, and considerable batch effects. Moreover, multiple human and murine pancreas states (normal, preneoplastic lesions, primary tumors, and metastases) will be investigated in this proposal. Deciphering the CAF subtypes in such varying conditions requires a robust machine-learning technique for cell type classification. To address these challenges we are developing ProtoCell, a deep-learning algorithm that uses the prototypical network model from the domain of few-shot deep learning to enable accurate cell type ascertainment from limited training examples in scRNA-seq data. We applied the ProtoCell model to the scRNA-seq data sets generated in this project to robustly identify the CAF subtypes and reproducibly annotate other cell clusters. We tailored ProtoCell hyperparameters and training algorithms to specifically accommodate the transcriptomic profiles of the PDAC microenvironment. We trained the model on our previously published PDA data sets: scRNA-seq data sets from human tumors and KPC mouse model ("all viable cells" and "fibroblast-enriched"), and RNA-seq from fluorescence-activated cell sorting fibroblast subpopulations. We explored the latent space of the ProtoCell model to characterize the changes in the CAF transcriptomic programs from normal to various PDA progression stages. Implementing this powerful tool promises to be a robust method for identifying fibroblast subtypes and other cell types in PDA scRNA-seq data sets.

STAR-SpliceGraph

Rapid advances in long-read sequencing technologies offer unprecedented opportunities to generate individual reads that cover the full length of the transcripts. However, long reads come at a cost of a much higher sequencing error rate, which creates significant hindrances to the read alignment. To improve the spliced alignment accuracy, we developed a hybrid approach, STAR-SpliceGraph, which combines precise splice junction information from short high-fidelity RNA-seq reads (Illumina) and long connectivity information of the long, noisy reads (ONT, PacBio).

STAR-SpliceGraph builds a collection of splice graphs by combining annotated exons and splice junctions with the novel exons and junctions from the short-read alignments. The splice graphs are represented as contiguous super-transcript sequences with additional splice junctions. The splice graph representation eliminates repeating sequences owing to overlapping exons and is thus advantageous over the standard approach of mapping to the transcriptome. Furthermore, the advantage of the splice graph alignment over the transcriptome alignment is that the former can be built from short-read RNA-seq alignments without the knowledge of the full-length transcript structures, allowing for the detection of novel isoforms and genes.

For mapping to the splice graphs, we first find the candidate super-transcripts to which a read is most likely to align, using a seed-and-rank strategy. The super-transcripts

that provide the best seed coverage of a read are selected for the local alignment step: a dynamic programming algorithm (similar to the Smith–Waterman local alignment) on the splice graph that allows splice "jumps" between the columns of the scoring matrix.

We use both simulated and real data to compare STAR-SpliceGraph with popular long-read mappers—minimap2, Graphmap2, and deSALT—showing significantly higher precision and recall of our algorithm for both annotated and novel isoforms, exons, and splice junctions.

STARsolo

We continued working on *STARsolo*, a comprehensive turnkey solution for quantifying gene expression in single-cell/nucleus RNA-seq data, built into RNA-seq aligner *STAR*.

We released a preprint on bioRxiv describing this tool as well as benchmarking results (Kaminow et al. 2021). Using simulated data that closely resembles realistic scRNA-seq, we demonstrate that *STARsolo* is highly accurate and significantly outperforms pseudoalignment-to-transcriptome tools. *STARsolo* can replicate the results of, but is considerably faster than, *CellRanger*, currently the most widely used tool for preprocessing scRNA-seq data. In addition to uniquely mapped reads, *STARsolo* takes account of multigene reads, necessary to detect certain classes of biologically important genes. It has a flexible cell barcode processing scheme, compatible with many established scRNA-seq protocols and extendable to emerging technologies. *STARsolo* can quantify transcriptomic features beyond gene expression, which we illustrate by analyzing cell type–specific alternative splicing in the Tabula Muris project.

STAR-Consensus

We continued developing STAR-Consensus, the extension of our popular RNA-seq aligner that improves the mapping and quantification of RNA-seq data by replacing minor alleles in the human reference genome with the major alleles. We exemplified the effects of replacing the reference with a consensus genome on gene expression in a single-cell RNA-seq data set. The highly popular droplet-based single-cell sequencing technologies allow studying the differential transcriptomic programs between cell types. Because a large proportion of reads generated by these technologies originate from untranslated regions (UTRs) and introns, they are especially susceptible to incorrect mapping due to minor alleles in the reference. In this example, we used the peripheral blood mononuclear cell data set generated by the 10X Chromium v3 protocol. Although the gene expression changes are small for the majority of the genes, several genes exhibit a significant change in expression when minor alleles in the reference are replaced with major alleles. The genes that show increased expression (23 genes, 12 protein-coding, seven disease-associated) in the pan-human consensus represent an improvement in sensitivity. On the other hand, the genes whose expression is higher in the reference (51 genes, 31 protein-coding, 22 disease-associated) are false positives that are eliminated in the consensus genome. This effect can also be observed in the differential gene expression between the different clusters. Although as before, only several genes are impacted, the biological interpretation of such genes will be significantly altered by the consensus genome. Given these observations, we can conjecture that other gene-expression-based analyses, such as expression quantitative trait locus (eQTL) and transcriptome-wide association studies (TWASs), can also be improved by replacing the reference genome with the pan-human consensus. Furthermore, we found that analyses that go beyond gene expression, such as alternative splicing and differential isoform expression, are also noticeably affected by the reference replacement with the consensus.

The STAR-Consensus manuscript was accepted for publication in *Genome Research* (Kaminow et al. 2022).

PUBLICATIONS

Bakken TE, Jorstad NL, Hu Q, Lake BB, Tian W, Kalmback BE, Crow M, Hodge RD, Krienen FM, Sorenson SA, et al. 2021. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598:** 111–119. doi:10.1038/s41586-021-03465-8

Haas B, Dobin A, Ghandi M, Van Arsdale A, Tickle T, Robinson JT, Gillani R, Kasif S, Regev A. 2021. Targeted in silico characterization of fusion transcripts in tumor and normal tissues via FusionInspector. bioRxiv doi:10.1101/2021.08.02.454639

Kaminow B, Yunusov D, Dobin A. 2021. *STARsolo*: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. bioRxiv doi:10.1101/2021.05.05.442755

Rozowsky J, Drenkow J, Yang YT, Gursoy G, Galeev T, Borsari B, Epstein CB, Xiong K, Xu J, Gao J, et al. 2021. Multitissue integrative analysis of personal epigenomes. bioRxiv doi:10.1101/2021.04.26.441442

In Press

Kaminow B, Ballouz S, Gillis J, Dobin A. 2022. Pan-human consensus genome significantly improves the accuracy of RNA-seq analyses. *Gen Res* doi:10.1101/gr.275613.121

CELL TYPES, GENE EXPRESSION, AND META-ANALYSIS

J. Gillis S. Fischer N. Fox B. Harris J. Hover R. Kawaguchi J. Lee R. Lohia S. Lu S. Muzumdar M. Passalacqua H. Suresh J. Werner

Neuronal Cell Types

Characterizing neuronal cell types is a crucial step toward understanding how neurons work together and how neurological disorders arise. To this end, the National Institutes of Health's (NIH's) BRAIN Initiative Cell Census Network (BICCN) consortium published a collection of articles providing an unprecedented atlas of neuronal cell types across a wide variety of modalities and species (single-cell transcriptomics, epigenomics, spatial transcriptomics, connectivity, morphology, and electrophysiology).

One of the Gillis laboratory's interests within the BICCN is to quantify the replicability of these novel cell types across laboratories and modalities. We developed a statistical tool called MetaNeighbor that uses a neighbor voting system to identify cell types with consistent expression signatures, suggesting broad support across a variety of experiments and computational pipelines. Cell types with high MetaNeighbor scores are more likely to replicate in independent studies; thus, they are more likely to be observed and easier to target in follow-up experiments. Thanks to its flexibility and scalability, MetaNeighbor was successfully used in a variety of contexts, from the characterization of novel cell types in maize to the understanding of evolutionary convergence and divergence of neurons across mammals.

Signatures of Individuality

Cells in a fully developed organism share life histories traced back through their divisions, defining lineages. Marks left on a cell early in the lineage can be inherited down through cell divisions, leaving shared features across cells, barcoding their lineages. Attempts to ascertain the existence of these permanent shared markings in previous work have mainly focused on the strongest events—monoallelism—in the simplest systems—cell lines—with mixed results. However, like X-inactivation, lineage specification occurring via autosomal epigenetic marks may have a very broad impact across both the genome and cell populations.

To more systematically assess their organismal impact, we turned to *Dasypus novemcinctus* (the nine-banded armadillo), which has a polyembryonic reproductive strategy, producing litters of identical quadruplets. This system enables an unusual level of environmental and genetic control, allowing us to assess cellular barcoding arising as noise in the assignment of epigenetic marks on alleles. In combination, these allelic imbalances progressively barcode trajectories across cell lineages as they move forward during development. We show that autosomal allelic ratios varying between individuals, and consistent with an early developmental original, are also enriched for stability over time, suggesting this is likely to have an important impact on disease variability by rendering otherwise haplosufficient genetic variation penetrant.

Cross-Kingdom Coexpression

During evolution, species often acquire new functions and structural features that help them adapt to their surroundings. Changes in gene regulation across species are common mechanisms underlying this functional evolution. Since coexpressed genes reflect shared function, comparative cross-species analysis of gene coexpression is useful for inferring essential, evolutionarily conserved regulatory modules, as well as species-specific expression changes driving morphological novelty.

Using meta-analytic coexpression networks, we investigate the evolution of gene regulation in 37 species across the eukaryotic tree of life. Ancient genes are conserved across kingdoms, whereas younger genes show

cell type– and species-specific expression profiles, suggesting new function acquisition through specialized expression signatures. This framework has also enabled us to analyze and compare the evolution of functional modules in organisms with poor genetic annotation, elucidating essential systems and separating them from niche metabolic pathways. By leveraging this powerful method and applying it across the plant kingdom, we can efficiently evaluate thousands of gene modules for their conservation, identifying deeply conserved metabolic modules and revealing modules being rapidly rewired toward new functions. We find that secondary metabolism modules are frequently repurposed for novel functions across species, and that in comparison to animals, plant primary metabolism genes develop novel functions much more rapidly.

X-Chromosome Variability

Mammalian females are effectively composed of two genetically distinct cellular populations that vary in expression from the X chromosome via X-chromosome Inactivation (XCI). XCI is a stochastic, permanent, and developmentally early epigenetic decision made in every female cell to transcriptionally silence a single X allele to match the single X chromosome in males. Taking into account the inherent stochasticity and permanence of XCI, shared variance in XCI mosaicism across cellular populations is indicative of shared developmental lineage. We have developed methods to model X-linked allelic imbalance from transcriptomics data, allowing us to quantify variance in X-linked mosaicism across human tissues and individuals and across mammalian species. By extracting and comparing variance in X-linked mosaicism across tissues, individuals, and species, we aim to identify fundamental features of lineage and conserved developmental principles.

Variance in XCI additionally plays an important role in heterozygous X-linked disease phenotypes in females, in which the direction and degree of allelic inactivation can directly influence disease phenotypes. X-linked chronic granulomatous disease (X-CGD) is a rare, monogenetic primary immunodeficiency characterized by recurrent, life-threatening infections and the formation of phagocyte-derived granulomas at infection loci. Comparing single-cell expression profiles between cells expressing the wild-type or mutated allele for X-CGD within a single carrier tightly controls for the impacts of genetic variation in disease pathogenesis. We aim to apply single-cell transcriptomic approaches to X-CGD carriers to identify important cell autonomous signaling in X-CGD pathogenesis.

Gene Function Learning

Because correlated gene expression implies involvement in shared processes, coexpression is a powerful, albeit noisy, source of information about function. This informs several aspects of our laboratory's activities. To address inter-/intra-experiment variability and noise, we aggregate many data sets to build coexpression networks (Lee et al., *Nucl Acids Res* **48:** W566 [2020]). To make expression data as easy to use and interpretable as sequence-based data, we develop efficient network algorithms to perform various kinds of inference and to test replicability (Ballouz et al., *Bioinformatics* **33:** 612 [2017]; Fischer et al. 2021). As part of our effort to demonstrate the utility of aggregate expression networks, we participated in the fourth Critical Assessment of Function Annotation (CAFA), demonstrating that the performance of a combination of expression and sequence-based data is better than either alone.

Coexpression data is especially useful when combined with other data modalities and when generated cross-species. Our laboratory has generated aggregate Hi-C chromatin contact networks combining up to 100 data sets per species. Hi-C and expression networks were then probed in combination, demonstrating improved performance of trans-interactions measuring evolutionary conservation and divergence. Aggregate meta-networks can be used as input, individually or in combination, to more complex machine-learning frameworks to explore gene function, regulation, and evolution. Planned work along these lines will leverage expression data with Google DeepMind's AlphaFold to probe protein–protein interactions.

PUBLICATIONS

Bakken TE, Jorstad NL, Hu Q, Lake BB, Tian W, Kalmbach BE, Crow M, Hodge RD, Krienen FM, Sorensen SA, et al. 2021. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598:** 111–119. doi:10.1038/s41586-021-03465-8

BRAIN Initiative Cell Census Network (BICCN). 2021. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598:** 86–102. doi:10.1038/s41586-021-03950-0

Fischer S, Gillis J. 2021. How many markers are needed to robustly determine a cell's type? *iScience* **24:** 103292. doi:10.1016/j.isci.2021.103292

Fischer S, Crow M, Harris BD, Gillis J. 2021. Scaling up reproducible research for single-cell transcriptomics using MetaNeighbor. *Nat Prot* **16:** 4031–4067. doi:10.1038/s41596-021-00575-5

Harris BD, Crow M, Fischer S, Gillis J. 2021. Single-cell co-expression analysis reveals that transcriptional modules are shared across cell types in the brain. *Cell Systems* **12:** 748–756. doi:10.1016/j.cels.2021.04.010

Lu S, Fürth D, Gillis J. 2021. Integrative analysis methods for spatial transcriptomics. *Nat Methods* **18:** 1282–1283. doi:10.1038/s41592-021-01272-7

Lu S, Ortiz C, Fürth D, Fischer S, Meletis K, Zador A, Gillis J. 2021. Assessing the replicability of spatial gene expression using atlas data from the adult mouse brain. *PLoS Biol* **19:** e3001341. doi:10.1371/journal.pbio.3001341

Muñoz-Castañeda R, Zingg B, Matho KS, Chen X, Wang Q, Foster NN, Li A, Narasimhan A, Hirokawa KE, Huo B, et al. 2021. Cellular anatomy of the mouse primary motor cortex. *Nature* **598:** 159–166. doi:10.1038/s41586-021-03970-w

Sun Y-C, Chen X, Fischer S, Lu S, Zhan H, Gillis J, Zador AM. 2021. Integrating barcoded neuroanatomy with spatial transcriptional profiling enables identification of gene correlates of projections. *Nat Neurosci* **24:** 873–885. doi:10.1038/s41593-021-00842-4

Yao Z, Liu H, Xie F, Fischer S, Adkins RS, Aldridge AI, Ament SA, Bartlett A, Behrens MM, Van den Berge K, et al. 2021. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598:** 103–110. doi:10.1038/s41586-021-03500-8

GENOME ORGANIZATION, REGULATION, AND FUNCTIONAL ROLES OF NONCODING RNAs

T.R. Gingeras J. Drenkow

The analysis of genomes often centers on finding mutations at both the single-nucleotide (single-nucleotide polymorphisms [SNPs]) or larger segments of DNA levels (structural variations [SVs]). These variations are often, but not exclusively, the cause of changes in a cell's endophenotypes (transcription, chromatin modifications and susceptibility, translational capability) leading to disease states. It is important to recall that such variations are most often located outside the coding regions of genes, particularly in intergenic regions, and are present in only one of the two alleles of a genome. This is the case for both plant and animal genomes.

The studies described in this year's report focus on both human and plant systems. The human projects consist of (1) assembling personal genome sequences that locate SNPs and SVs and using this information in developing predictive models that identify whether a gene's expression will be allele-specific and having the capability to predict if a gene is exhibiting allele-specific expression in a tissue that cannot be analyzed directly (EN-TEx project), and (2) identifying the significant issues and insights in evaluating transcriptional profiles of postmortem brain tissues versus freshly collected normal tissue samples (from epilepsy patients). This later subject is of particular importance because many studies rely on postmortem samples for identifying causes of disease.

The plant studies consists of (1) using single-cell RNA sequencing to rapidly generate a single-cell-resolution map of the maize root, (2) investigating the transcriptional profiles of single cells from developing maize ears, resulting in a developmental atlas providing a single-cell RNA sequencing (scRNA-seq) map of an inflorescence, and (3) validating of the 5′ total soluble solid (TSS) annotations present in the genome assemblies of two lines of *Sorghum* and RNA annotation and mapping of promoters for analysis of gene expression using RAMPAGE to identify the TSSs in roots and shoots.

Single-Cell Maize Root Maps

This work was performed in collaboration with the K. Birnbaum (NYU) and T.R. Gingeras and D. Jackson (CSHL) laboratories.

Roots are radially symmetrical organs composed of three fundamental tissue types: the epidermis on the outside, the ground tissue at the middle, and a core of vascular elements plus pericycle that lie in a central cylinder known as the stele. Variations in ground tissue patterning—particularly the number of cortex cell layers—are common across species and represent one of the defining features giving rise to interspecies root morphological diversity. This diversity allows plants to cope with biotic and abiotic stress and adapt to challenging environments. For example, maize cortex plays a role in drought and flood tolerance and hosts colonization of beneficial mycorrhizal associations that reduce stress and improve nutrient uptake. Therefore, ongoing questions in root biology are how tissue patterning is adjusted to produce divergent root morphologies and what alterations in the genetic networks control developmental differences among species.

To construct a single-cell map of the maize root, we took advantage of the concentric arrangement of tissues in maize roots. The approach allowed for enrichment of cells in each different concentric tissue layer using blue/green ratios in fluorescence-activated cell sorting (FACS). Using scRNA-seq, performed by our laboratory, of cells from each layer, a high-resolution map of the maize root was constructed, revealing an alternative configuration of the tissue formative transcription factor SHORTROOT (SHR) adjacent to an expanded cortex. We observed that maize SHR protein is hypermobile, moving at least eight cell layers into the cortex during the course of root construction. SHR mutants in both maize and *Setaria* have reduced numbers of cortical layers, showing that the SHR pathway controls expansion of cortical tissue to elaborate anatomical complexity. The results show that SHR has a role in monocots in controlling the expansion of cortex, which sets up many traits for

environmental acclimation. This illustrates how subtle divergence of a conserved developmental regulator can mediate anatomical complexity that has given rise to specialized functions. Related to the complexity of the root, we identify four distinct cortical cellular subtypes in our bioinformatic analysis, although further work is needed to verify their spatial relationship. Finally, the results show that rapid transcriptome mapping using single-cell dissection can provide insights into the mechanisms that mediate anatomical diversity. The use of dye labeling to generate a scaffold locational map, together with scRNA-seq, provides a maize root tissue map that can be used as a reference in maize and related plants.

Single-Cell RNA Sequencing of Developing Maize Ears

This work was done in collaboration with the T.R. Gingeras and D. Jackson laboratories (CSHL).

Many key transcription factor (TF) regulators (*KNOTTED1, BARREN STALK1, BRANCHED SILKLESS1, RAMOSA*) have reshaped inflorescence architecture during evolution or domestication, and their discovery was enabled by the availability of mutants that block specific aspects of development. However, such insights are limited by genetic redundancy and pleiotropy, so a high-resolution expression atlas of specific cell types and domains is needed to gain further insights into the gene networks that control development.

To generate a single-cell atlas from developing maize ears, we used the 5- to 10-mm stage, in which major developmental and architectural decisions, including meristem initiation, maintenance and determinacy, organ specification, and differentiation of vascular and ground tissues, are being made. A cell wall digestion method was optimized, taking into account the different composition of grass cell walls that allowed us to isolate ear protoplasts. However, developing ear protoplasts were fragile, and we removed small debris and organelles from broken cells by filtration followed by FACS before loading into the 10x Genomics Chromium Controller. Then, scRNA-seq libraries were generated and sequenced on the Illumina platform. The transcriptional profiles of 12,525 single cells from each sample of developing maize ears resulted in the development of a single-cell atlas map of an inflorescence. We validated our results by mRNA in situ hybridization and by FACS RNA-seq, and we show how these data may facilitate genetic studies by predicting genetic redundancy, integrating transcriptional networks, and identifying candidate genes associated with crop yield traits.

Genomic Variation Correlated with Sorghum and Sugarcane Aphid Resistance Genes

This work was done in collaboration with the D. Ware and T.R. Gingeras laboratories (CSHL).

Sorghum bicolor, one of the most important grass crops around the world, harbors a high degree of genetic diversity. A chromosome-level genome assembly was constructed for two important sorghum inbred lines, Tx2783 and RTx436. Genes were annotated in these assembled genomes using evidence-based and de novo gene predictors. Transcriptional start sites (TSSs) identified in the annotations were evaluated using RAMPAGE and found to be consistently confirmed. Together with other public sorghum genomes, BTx623, RTx430, and Rio, extensive SVs of various sizes were characterized using Tx2783 as a reference. Genome-wide scanning for disease resistance (R) genes revealed high levels of diversity among these five sorghum accessions. To characterize sugarcane aphid (SCA) resistance in Tx2783, the resistance region on chromosome 6 was mapped using a recombinant inbred line (RIL) population and revealed an SV of 191 kb containing a cluster of R genes in Tx2783. Using Tx2783 as a backbone, along with the SVs, we constructed a pan-genome to support alignment of resequencing data from 62 sorghum accessions and then identified core and dispensable genes using this population. This study provides the first overview of the extent of genomic structural variations and R genes in the sorghum population and reveals potential targets for breeding of SCA resistance.

Issues and Insights in Evaluating Transcriptional Profiles of Postmortem Brain Samples

This work was done in collaboration with the J.A. Loeb (University of Illinois), S.E. Celniker and J.B. Brown (Lawrence Berkeley National Laboratory, University of California), N. Boley (University of California, Berkeley), and T.R. Gingeras (CSHL) laboratories.

The transcription patterns and histological features of postmortem brains were compared to fresh

human neocortex isolated immediately following surgical removal. The fresh human brain transcriptome had an entirely unique transcriptional pattern compared to the postmortem neocortex samples (postmortem interval [PMI] = 29 h). To better understand this transcriptional difference, we measured genome-wide RNA expression as a function of time after fresh tissue removal to mimic the PMI. Within a few hours, a selective reduction in the number of neuronal activity–dependent transcripts occurred with relative preservation of housekeeping genes commonly used as a reference for RNA normalization. Gene clustering indicated a rapid reduction in neuronal gene expression with a reciprocal time-dependent increase in astroglia and microglial gene expression that continued to increase for at least 24 h after tissue resection. Predicted transcriptional changes were confirmed histologically on the same tissue, demonstrating that although neurons were degenerating, glial cells underwent an outgrowth of their processes. The rapid loss of neuronal genes and reciprocal expression of glial genes highlight highly dynamic transcriptional and cellular changes that occur during the PMI. Understanding these time-dependent changes in gene expression in postmortem brain samples is critical for the interpretation of research studies on human brain disorders and research in animal models that aim to understand fundamental disease processes. The fresh versus postmortem transcriptional profile differences have yet to be studied in other human tissues or single cells. These differences in transcriptional profiles are needed to provide reliable interpretations of the effects of environmental influences on human tissues/cells.

Multitissue Integrative Analysis of Personal Epigenomes (ENTEx Project)

Here, to better connect personal genomes and functional genomics, we initiated the EN-TEx study (ENCODE assays applied to GTEx samples). In particular, we created a uniformly processed data set of more than 10 functional genomic assays, consistently collected from approximately 25 tissues from four individuals. It is coupled with long-read genome assemblies containing comprehensive sets of SVs. Compared to what was previously possible, mapping reads from the assays directly to diploid genomes allows for more precise quantification of differential expression and regulatory-element activity and for directly visualizing the impact of SVs on chromatin; the uniform nature of the data set makes possible more precise ascertainment of interindividual versus intertissue differences; and the scale of the data enables the creation of the largest catalog of noncoding AS variants, an order of magnitude more than what was available previously.

We leveraged this catalog to build generalized models of variant impact. In particular, we created a model that predicts the AS imbalance resulting from a single-nucleotide variant (SNV) just from the extended sequence context around a site (i.e., within an ~250-bp window around an SNV with activity in a particular assay). It highlights the importance of ~100 TF motifs we term AS-sensitive. Moreover, we can use these to disentangle the complicated relationship between the AS expression of a gene and the AS activity of its promoter. In addition, we can relate the EN-TEx resource to external genome annotations—expression quantitative trait loci (eQTLs) and regulatory elements already known for the human genome (Fig. 1).

We also built generalized models that transfer eQTLs from a source tissue to a different target one, leveraging the fact that the EN-TEx resource represents a uniform collection of epigenetics data in hard-to-obtain tissues. This is practically quite useful given it is typically much easier to measure eQTLs directly in blood than other tissues, such as the heart, especially when using large cohorts of individuals.

We also show that the EN-TEx data can "decorate" existing regulatory elements, identifying subsets that are much more highly enriched with eQTLs and genome-wide association study (GWAS) SNVs than had been previously possible and illuminating broad relationships between conservation, AS activity, and tissue specificity.

Overall, EN-TEx provides a powerful resource of multi-omic information and describes a more accurate approach for future functional genomics. Such information will be crucial for the interpretation of whole genomes and for future applications of them in precision medicine.

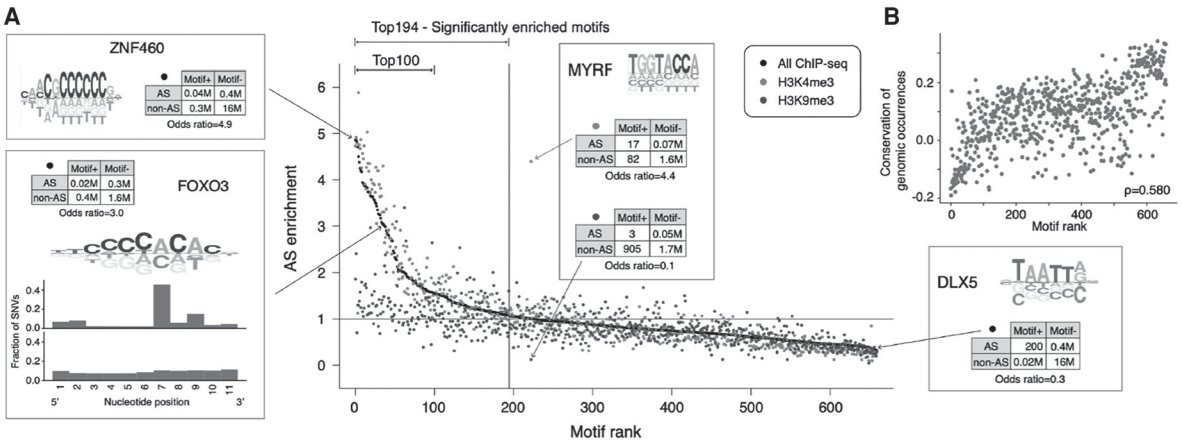


Figure 1. Relating transcription factor (TF) motifs to AS activity. (*A*) Motifs ranked by enrichment of AS single-nucleotide variants (SNVs). We calculated the enrichment of AS SNVs for the motif of each TF. Taking FOXO3 as an example for the calculation, we made a 2-by-2 contingency table: motif or nonmotif regions and AS SNVs or non-AS SNVs. The odds ratio of the table indicates the enrichment of AS SNVs for the FOXO3 motif. In the scatter plot, each dot corresponds to the motif of a TF, and the dots are ranked by their AS enrichment. Different colors indicate that the AS enrichments are calculated using different histone modification assays: using all accessible SNVs of all chromatin immunoprecipitation sequencing (ChIP-seq) (black), using accessible SNVs from H3K4me3 ChIP-seq (red), and using H3K9me3 ChIP-seq (blue). (*B*) Motif ranking is correlated with conservation. To calculate the conservation of a motif, we first calculated the phyloP score of each motif region in the genome, and then averaged these scores over the many genomic regions of this motif.

PUBLICATIONS

Dachet F, Brown JB, Valyi-Nagy T, Narayan KD, Serafini A, Boley N, Gingeras TR, Celniker SE, Mohapatra G, Loeb JA. 2021. Selective time-dependent changes in activity and cell-specific gene expression in human postmortem brain. *Sci Rep* **11:** 6078. doi:10.1038/s41598-021-85801-6

Grapotte M, Saraswat M, Bessière C, Menichelli C, Ramilowski JA, Severin J, Hayashizaki Y, Itoh M, Tagami M, Murata M, et al. 2021 Discovery of widespread transcription initiation at microsatellites predictable by sequence-based deep neural network. *Nat Commun* **12:** 3297. doi:10.1038/s41467-021-23143-7

Ortiz-Ramírez C, Guillotin B, Xu X, Rahni R, Zhang S, Yan Z, Coqueiro Dias Araujo P, Demesa-Arevalo E, Lee L, Van Eck J, et al. 2021. Ground tissue circuitry regulates organ complexity in maize and *Setaria*. *Science* **374:** 1248–1252. doi:10.1126/science.abj2327

Rozowsky J, Drenkow J, Yang Y, Gursoy G, Galeev T, Borsari B, Epstein C, Xiong K, Xu J, Gao J, et al. 2021 Multitissue integrative analysis of personal epigenomes. bioRxiv doi:10.1101/2021.04.26.441442

Wang B, Jiao Y, Chougule K, Olson A, Huang J, Llaca V, Fengler K, Wei X, Wang L, Wang X, et al. 2021 Pan-genome analysis in sorghum highlights the extent of genomic variation and sugarcane aphid resistance genes. bioRxiv doi:10.1101/2021.01.03.424980

Xu X, Crow M, Rice BR, Li F, Harris B, Liu L, Demesa-Arevalo E, Lu Z, Wang L, Fox N, et al. 2021. Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene. *Dev Cell* **56:** 557–568. doi:10.1016/j.devcel.2020.12.015

LEVERAGING LONG-READ SEQUENCING TECHNOLOGIES TO FACILITATE GENOMIC DISCOVERY

W.R. McCombie M. Anzaldi, E. Ghiban, S. Goodwin, S. Iyer, M. Kramer, S. Mavruk Eskipehlivan, O. Mendivil Ramos, S. Muller, A. Qui, R. Wappel

In much of 2020 we optimized the technology required for long-read sequencing of complex genomes. In 2021 we used those developments to dive more deeply into a number of biologically significant questions. These included examining the genetics of human diseases, including cancer and autism, and projects in plant genomics. We also continued to work on significantly advancing the technology of long-read sequencing, including applying it to detection of methylation in very large plant genomes and cancer genomes. This has been using both whole-genome sequencing of DNA and a targeted method of sequencing with very long reads that we have developed.

Long-Read Assembly and Genomic Analysis 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, B. Ambrose, S. Frangos, D. Little, and S. Wilson, NYBG; M. Schatz and S. Ramakrishnan, Johns Hopkins University; K. Varala, Purdue; S. Kolokotronis, SUNY Downstate Medical Center]

Living fossils is a term coined by Darwin to describe species of plants that, based on the morphology seen in fossils, have not evolved significantly over a long time span (up to hundreds of millions of years). This project's goals are to sequence some of these genomes and compare them to those of related species that have evolved over similar time spans in order to attempt to find characteristics of the living fossil genomes that have enabled them to survive multiple mass extinction events. In the last year we worked primarily on the species *Wollemia nobilis*, which is a critically endangered species of tree with a very large genome (12 gigabases). It exists in the wild only in a small number of areas that are close together in New South Wales, Australia.

We had previously sequenced this genome with the Oxford Nanopore (ONT) sequencer, but new base-calling software for the ONT sequencer became available, and we chose to re-call the bases from the raw data bases to improve accuracy. We did this, and the improved accuracy of the base calls improved the resulting genome assembly significantly. We began using this as our reference assembly for this genome. We also began calling CpG methylation within the genome using the same raw sequence data.

In addition to the genome sequencing with ONT chemistry, which was done from leaf tissue, we did methylation analysis (using bisulfite sequencing on an Illumina instrument) as well as small RNA sequencing from reproductive tissues. This had been delayed due to a combination of the extremely endangered nature of the species and COVID restrictions, which limited our ability to obtain the necessary tissue. We were able to obtain these tissues in 2021 and began to analyze the resulting sequence data in collaboration with the Martienssen laboratory.

We are also collaborating with a group of investigators in Australia who have done single-nucleotide polymorphism (SNP) sampling of many of the remaining number, ~200, of *W. nobilis* that survive in the wild. We will use these data in comparison with our reference to better understand the population diversity of the species in the wild. We have also done first pass sequencing and assembly on remaining genomes for the project including *Araucaria*, *Gnetum*, *Juniperus*, and *Metasequoia*.

Interrogating Bat Genomes: Insight into Mechanisms of Immunity and Cancer Resistance

O. Mendivil Ramos, M. Kramer, S. Goodwin, R. Wappel, W.R. McCombie [in collaboration with A. Siepel and A. Scheben, CSHL; N. Simmons and S. Oppenheim, American Museum of Natural History; M. Schatz, Johns Hopkins University]

In collaboration with the Siepel laboratory, we continued our study of the genomes of two bat species (*Artibeus jamaicensis* and *Pteronotus mesoamericanus*), taking advantage of the improved quality of the ONT assemblies to assess positive selection and gene expansions within the bat lineage. Our collaborator Armin Scheben's analysis highlighted positive selection in genes involved in immunity and cancer-related genes. He noted a large contraction of the type I interferon gene family and an expansion of the IFITM gene family in bats compared to other mammals, which may be linked to increased viral response in bats. Our copy number analysis identified a large expansion of the gene *PRDM9*, specific to the Phyllostomid clade of bats. *PRDM9* may have multiple functions, as it is involved in meiotic recombination and is also unregulated during viral infection. These changes may be associated with the long lifespan and low cancer incidence of bats.

These findings are summarized in a revised and updated bioRxiv manuscript (Scheben et al. 2021). In addition, in 2022 we plan to expand upon the findings in a collaboration with the Mills laboratory to functionally assess these candidates using constructs of these changes in cellular assays.

ACME: An Affinity-Based Cas9-Mediated Enrichment Method for Targeted Nanopore Sequencing

S. Iyer, M. Kramer, S. Goodwin, W.R. McCombie

Having established ACME (Affinity-based Cas9-Mediated Enrichment) as an effective targeted long-read method, over the past year we assessed ACME's ability to generate whole-gene spanning reads and worked on optimizing ACME for a wide range of applications.

A major advantage that ACME offers over other long-read targeting approaches is the ability to get multiple overlapping reads that span the entire length of the region of interest. For all genes on our panel that were ≤ 100 kb in size, we obtained three to seven times as many end-to-end target spanning reads in comparison to the same regions captured using Nanopore Cas9 Targeted Sequencing (nCATS), a method developed by Gilpatrick et al. (*Nat Biotechnol* **38:** 433 [2020]) (Table 1). End-to-end reads help reduce mapping errors that could arise due to the presence of structural variants. Even for larger targets, like *BRCA2* (90 kb) and *PAX7* (120 kb), we obtained two to 20 reads spanning the entire 90- to 120-kb targets, whereas these counts were between zero and two reads without ACME. We clearly see that without ACME we lose the ability to capture reads that span the full length of genes in targets >30 kb, thereby limiting our scope for targeting these genes.

Table 1. Number of reads spanning X% of gene for each targeted cancer panel gene across non-ACME and ACME pooled library sequencing runs

	MCF10A No ACME					MCF10A ACME[a]				SKBR3 ACME			
Gene	**Size**	**100%**	**80%**	**60%**	**40%**	**100%**	**80%**	**60%**	**40%**	**100%**	**80%**	**60%**	**40%**
MYC	12565	129	140	148	153	993	1036	1055	1075	2057	2210	2300	2424
HOXA9	18506	61	63	68	76	425	444	471	489	47	106	118	133
FGFR4	19916	36	39	41	43	233	242	253	263	186	206	227	239
STK11	30286	29	30	35	40	152	161	168	183	64	79	89	102
CDKN2A	30774	NA	NA	NA	NA	NA	NA	NA	NA	35	41	54	60
TERT	44787	7	9	9	14	61	74	87	119	0	6	7	106
KRAS	50955	9	13	19	26	111	123	142	168	27	37	43	60
BRCA2	91218	1	1	2	3	20	26	36	54	3	10	21	34
PAX7	120934	1	1	1	4	2	4	7	26	1	3	7	18
APC	144265	0	0	0	3	0	1	2	16	0	1	1	7

For some conditions, more than one run was performed during the testing phase. For these runs, average values are represented in the table. (ACME) Affinity-based Cas9-mediated enrichment, (NA) not available.

[a]For one of the three MCF10A ACME runs, only three samples were pooled together instead of four. All other runs had four samples pooled together for each run.

ACME—Further Optimization

For the work described below, we downsized the targets to the eight genes from the panel that were <100 kb in size, because of the coverage drop we observed in our >100-kb targets (*PAX7* and *APC*). In late 2020 ONT released a stand-alone Cas9 kit (CS9109) to encourage a switch from purchasing individual reagents from NEB. We tested the ONT kit and compared its performance with NEB reagents for non-ACME and ACME runs as both pooled and single-library preps using MCF 10A HMW DNA. Although we observed the expected trend of ACME preps outperforming the non-ACME preps, we noticed that both the NEB and ONT Cas9-based kits were comparable in performance for the pooled runs. For the single-sample runs, however, the ONT Cas9 kit outperformed the NEB reagents. We have since switched to using the ONT Cas9 kit for all our subsequent ACME optimization tests.

Native Barcoding

When performing single-sample ACME, we observed low pore occupancy on the flow cells, prompting us to look into multiplexing options to better utilize each flow cell's sequencing capacity. Multiplexing will also bring down per sample sequencing costs, making it an important avenue to explore. We tested the native barcoding kit (NBD104) that is a polymerase chain reaction (PCR)-free barcoding approach and will therefore retain DNA modifications when combined with ACME, which is also amplification-free. Our initial test of the NBD104 using MCF 10A gDNA gave promising results and good barcode classification, with <5% of reads unclassified. We next decided to try native barcoding with Cas9-based targeting. After trying different combinations of barcode ligation times and incubation temperatures, for both nCATS and ACME, we found that 20 min incubation at room temperature (RT) gave us the best results. Again, we observed the trend of ACME outperforming nCATS in number of on-target reads and coverage. However, close to 40%–50% of reads generated when combining Cas9-based targeting with native barcoding could not be classified under the barcodes used. This number improved to 35%–40% unclassified reads on rebasecalling the data with the newest basecaller (GuppyV6), but it is still far from ideal. We are now looking into further optimizing the protocol to improve barcode ligation and classification.

Table 2. Genes targeted, read counts, and coverage obtained from MCF 10A pooled ACME runs

	GridION		PromethION	
Gene	Reads	Coverage	Reads	Coverage
MYC	592	527	2584	2325
HOXA9	303	290	1221	1191
FGFR4	611	477	2109	1711
STK11	326	222	1276	936
TERT	223	115	1069	658
KRAS	240	139	1020	625
BRCA2	158	59	518	165
Total target	2453	171	9797	709

ACME on PromethION

Although Cas9-based targeting is currently supported only on the MinION/GridION, we tested ACME on the PromethION high-throughput instrument to evaluate compatibility and resultant throughput. Because the PromethION flow cells have a much higher pore count than the MinION cells, we did a pooled ACME run (starting with a total of 20 µg MCF 10A DNA) using the eight-gene panel described above. As seen in Table 2, the pooled ACME run on PromethION did phenomenally well when compared with a pooled run on the GridION, giving us a much higher coverage for genes as large as 90 kb (*BRCA2*). Because a pooled run needs 20 µg of DNA, this makes our efforts to benchmark the ACME + barcoding approach even more important. Being able to run ACME on the PromethION will allow us to potentially pool together more samples, perhaps being able to go as high as 15 samples per flow cell.

Circulomics Ultra-Long DNA Extraction

Through the various tests run while optimizing ACME, we observed a decrease in coverage with an increase in target size in our panel (e.g., *PAX7* and *APC* as mentioned above). To improve coverage of larger genes on our panel, we tried switching from our HMW spooling method to the ultra-long DNA extraction method from Circulomics (Pacific Biosciences) for MCF 10A. This did not show a significant difference in per gene coverage when applied to our panel of mixed-sized genes (10–150 kb). However, it has been observed that smaller targets tend to outcompete larger ones for pore occupancy on the flow cell. So, we grouped genes of ≥45 kb to form a separate gene panel and tested the spooled versus ultra-long approach. As seen in Table 3, we found that switching

Table 3. Coverage of long genes (≥ 45 kb) targeted using ACME on spooled HMW versus Circulomics UL DNA

		Coverage	
Gene	**Target size**	**Spooling**	**Circulomics**
TERT	44787	15	96
ERBB2	47211	15	74
KRAS	50955	9	109
BRCA1	90700	3	32
BRCA2	91218	4	35
APAF1	101790	8	44
PAX7	120934	5	32
APC	144265	2	22
Total target size	691860	6	45

to ultra-long DNA shows a promising improvement in coverage of long genes. As we expand our gene panel to include more targets of interest, we will continue testing various size combinations for these gene panels to identify one that would help maximize coverage of targets irrespective of size.

Variant Prioritization and Analysis in Early-Onset Cancer Pedigrees

W.R. McCombie, M. Kramer, S. Goodwin, R. Wappel, S. Muller [in collaboration with Z. Stadler Z. Patel, S. Lowe, and F. Barriga, MSKCC]

Continuing our collaboration with the Stadler group at MSKCC, we refined our analysis of early-onset cancer pedigrees, focusing on two colon cancer trios. We reprocessed the ONT data to take advantage of improved basecalling models with an updated version of Guppy to increase raw read accuracy. We then used our previously reported variant filtering method to reduce the number of genomic alterations by prioritizing those found in the affected member but not the unaffected relatives. We streamlined the structural variant (SV) filtering pipeline using updated versions of gnomA (Collins et al., *Nature* **590:** E55 [2021]) and dbVar (Lappalainen et al., *Nucl Acids Res* **41:** D936 [2013]) to reduce overlap with known common or benign variants. We are also actively working on improving detection of de novo variants by reducing missed inheritance. Because many of our findings are in intronic regions of cancer-related genes, we further prioritized variants by overlap with ENCODE (ENCODE Project Consortium, *Nature* **489:** 57 [2012]) regulatory elements such as histone marks, DNAse hypersensitivity regions, promoters, or enhancers to increase likelihood of identifying events with possible functional impact. We also worked to improve our SNP calling accuracy using the new PEPPER-Margin-DeepVariant pipeline (Shafin et al., *Nat Methods* **18:** 1322 [2021]), which uses haplotype awareness to decrease noise. In addition, we have begun implementing tools such as LongPhase (Lin et al., *Bioinformatics* btac058 [2022]) and NanoMethPhase (Akbari et al., *Genome Biol* **22:** 68 [2021]) to phase the SNVs and structural variants as well as the methylation calls in order to create a more informative context of these variants across haplotypes. The trio structure allows us to validate these events in the parental genomes.

In addition, we have begun analysis of the trios using the new CHM13 long-read reference genome (Nurk et al., bioRxiv 2021.05.26.445798 [2021]), which has shown significant improvements in terms of error correction, additional sequence, and improved variant calling over GRch38. We will continue our analyses in 2022 and are currently working on a manuscript describing our methods, as well as applying for a grant to expand upon this work in a larger cohort of early-onset colon trios, which, in collaboration with the Lowe laboratory, would allow us to use CRISPR-Cas9 modification and genome knock-in techniques to explore the effect of these variants on gene expression.

Sequencing and Assembly of the Wild Buckwheat *Erigonum umbellatum*

W.R. McCombie, S. Goodwin, M. Kramer, O. Mendivil Ramos [in collaboration with P.B. Pearman, Universidad del País Vasco/Euskal Herriko Unibertsitatea/Basque Center for Climate Change; J. Travis Columbus, California Botanic Garden]

In 2021 we continued our work on the whole-genome assembly of the wild buckwheat *E. umbellatum*, which is able to survive in multiple habitats and may be an important model for climate change. We investigated the Illumina short-read data generated from this sample using GenomeScope (Ranallo-Benavidez et al., *Nat Commun* **11:** 1432 [2020]) with kmer analysis to estimate genome size, repeat content, ploidy, and heterozygosity of this species about which little is known. According to these analyses, the best fit data model indicates that this species is most likely a highly repetitive autotetraploid.

We then focused on our improved Flye assembly (Freire et al. IEEE/ACM Trans Comput Biol Bioinform TCBB.2021.3108843 [2021]) of the Oxford

Nanopore long-read data, using two rounds of polytomous variable latent class analysis (poLCA) to polish the assembly with short-read Illumina data, enabling us to achieve an estimated mean quality of >99.9 (Q30). We used BUSCO (Seppey et al., *Methods Mol Biol* **1962:** 227 [2019]) to estimate genome completeness by conserved orthologs, and we found our assembly to be very complete, but highly duplicated. After reducing redundancy with Purge haplotigs (Seppey et al., *Methods Mol Biol* **1962:** 227 [2019]; Roach et al., *BMC Bioinformatics* **19:** 460 [2018]), we were able to maintain a high level of completeness (>90% of orthologs from the eudicot, embryophyte, and viridiplantae databases) while reducing duplications. We are continuing work on this genome and have mapped our collaborator's leaf transcriptome data from *E. umbellatum* to the assembly. Gene, repeat, and transposon annotation are ongoing to complete this work for publication.

Detection of BrdU in Long Nanopore Reads to Map Replication Forks in Yeast

M. Kramer, S. Goodwin, W.R. McCombie [in collaboration with Y. Hu and B. Stillman, CSHL]

Oxford Nanopore raw signal data allows analysis of base modifications by assessing the current shift of modified bases versus unmodified bases through the pore. In collaboration with the Stillman laboratory, we began a project to test the ability of Oxford Nanopore long reads to detect incorporation of the analog Bromodeoxyuridine (BrdU) instead of thymidine to detect origins of replication in *Saccharomyces cerevisiae* yeast strains. We first tested the ForkSeq pipeline (Hennion et al., *Genome Biol* **21:** 125 [2020]), which uses the RepNano tool (Boemo, *BMC Genomics* **22:** 430 [2021]) to align the raw signal of nanopore reads to the reference genome and apply convolutional neural network (CNN) and transition matrix (TM) models to determine BrdU/T ratio content along a growing nascent strand. These models detect increasing and decreasing BrdU/T ratios to determine fork directionality, and then determine the position of replication initiation and termination events using divergent and convergent forks, respectively. We were able to detect BrdU signal in the initial experiments; however, the signal to noise was not ideal, and the CNN/TM concordance was lower than expected, likely due to a difference in the efficiency of BrdU incorporation for the selected strain. Consequently, the Stillman laboratory used multiple laboratory methods to increase efficiency of BrdU incorporation for their strains of interest. We are now assessing the utility of another tool, DNAscent2 (Hennion et al., *Genome Biol* **21:** 125 [2020]; Boemo, *BMC Genomics* **22:** 430 [2021]), which allows detection of BrdU at single-base resolution and can be used with both synchronous and asynchronous cells.

Uncovering De Novo Mutations in Autism Quads Using Long-Read Sequencing

S. Goodwin, M. Kramer, W.R. McCombie [in collaboration with M.D. Noyes, W.T. Harvey, D. Porubsky, A. Sulovari, R. Li, N.R. Rose, P.A. Audano, K.M. Munson, A.P. Lewis, K. Hoekzema, and E.E. Eichler, University of Washington, Seattle; T. Mantere and A. Hoischen, Radboud University Medical Center, the Netherlands; T.A. Graves-Lindsay, Washington University in St. Louis; J.O. Korbel and A.D. Sanders, EMBL, Germany; Y. Mokrab, Weill Cornell Medicine, Qatar; M.C. Zody, NYGC]

In collaboration with the Eichler laboratory and the New York Genome Center (NYGC), we performed Oxford Nanopore whole-genome long-read sequencing of four family members—one female child affected with autism as well as two unaffected parents and an unaffected sibling. Analysis of sequence data from multiple platforms including Illumina, PacBio, and Oxford Nanopore revealed that long-read sequencing improved the ability to detect de novo variants that could contribute to autism, by allowing interrogation of complex regions that are not mappable with short reads. Long reads, combined with the improved reference assembly from CHM1 (Nurk et al., bioRxiv 2021.05.26.445798 [2021]), increased the number of detected de novo mutations by >25%. A manuscript describing this work has been accepted for publication at the *American Journal of Human Genetics*.

Profiling Targeted Cancer Genes in Breast Tumor Organoids

W.R. McCombie, S. Goodwin, M. Kramer, S. Muller, E. Ghiban [in collaboration with D. Tuveson, D. Spector, P. Sridevi, D. Plenker, G. Arun, S. Bhatia, S. Beyaz, and B. Yueh, CSHL; V. Corbo, University of Verona]

In collaboration with the Spector laboratory, we characterized breast tumor organoid models using panel capture and Illumina sequencing of targeted cancer driver genes that we had developed for the previously

reported Leidos project. Using our pipeline to prioritize pathogenic mutations and identify possible drivers, we assessed both germline and somatic mutations in the organoids as part of the multi-omic characterization of these organoids. Our work, along with analyses by our collaborators, showed that these models were highly informative and reflected cellular, genomic, and transcriptomic tumor characteristics. Therefore, these models could provide critical methods to test cancer treatment response and aid in personalized medicine. This work was published on bioRxiv (Bhatia et al. 2022) and has been submitted to *Cancer Discovery* for publication.

Optimizing Full-Length RNA-seq without Amplification on Oxford Nanopore

W.R. McCombie, S. Goodwin, M. Kramer, S. Muller, R. Wappel [in collaboration with A. Krainer, K. Lin, L. Wang, and A. Kral, CSHL; D. Garald, Oxford Nanopore]

The McCombie laboratory has continued to optimize and develop long-read sequencing for a wide range of different applications. One important area of optimization has been to adapt long-read technology, and Oxford Nanopore Technologies (ONT) long reads in particular, to RNA-seq applications.

In collaboration with the Krainer laboratory, the core has been comparing the results of direct RNA (dRNA) sequencing with complementary DNA (cDNA) sequencing for organoid samples. Although cDNA work at CSHL has generated as many as 80M reads, dRNA work has been more limited. One of the main hurdles for dRNA work on organoids is a limited availability of RNA. The official dRNA protocol from ONT requires 500 ng of enriched poly(A) RNA. Previously we have found that 5 µg of total RNA can be used in the direct RNA protocol, generating >4M reads on a PromethION cell. Although the large number of reads was superior to using pre-enriched RNA, analysis showed that ~50% of the reads were ribosomal. Generating 500 ng of poly(A) RNA has remained challenging, and to date we continue to use the total RNA approach despite its limitations. Conversely, we found poly(A) enrichment direct cDNA approaches to be effective. Like direct RNA, when total RNA is used as starting material without poly(A) enrichment, ~50% of all reads map to the ribosome. When up to 1200 ng of total RNA is treated with RiboZero (Illumina), the recovery is ~10%. This enriched material is sequenced on a single PromethION cell, yielding ~15M full-length cDNA reads with ~5% mapping to the ribosome. Additional work to optimize both direct RNA and direct cDNA methods will continue in 2022.

Expanding Long-Read Sequencing Capabilities to Diverse Samples

W.R. McCombie, S. Goodwin, M. Kramer, S. Muller, R. Wappel, S. Mavruk Eskipehlivan, E. Ghiban [in collaboration with Z. Lippman, M. Benoit, R. Santos, X. Wang, M. Alonge, S. Ramakrishnan, K. Jenike, CSHL; M. Schatz, Johns Hopkins; M. Ebbert, J. Brandon, and B. Heberle, University of Kentucky; D. Garald, Oxford Nanopore; R. Giordano and K. Donthu, University of Illinois; G. Concepcion, Pacific Biosciences]

In 2021, the McCombie laboratory was active in the development of varied strategies in long-read sequencing for a wide range of applications.

The first project attempted to validate brain RNA-seq on Oxford Nanopore. Brain tissue presents a unique but important challenge because of the fragile nature of brain-derived RNA. Previously, in collaboration with Mark Ebbert at the Mayo Clinic, we have shown that <25M ONT reads can detect the same number of genes as 100M short reads. Because of Dr. Ebbert's move from the Mayo Clinic to the University of Kentucky, progress in 2021 was slow. However, the University of Kentucky has an extensive brain bank that will be a good asset for this work. In 2021 the Ebbert group provided mouse brains for testing to extract high RNA integrity number (RIN) RNA. Using a brain RNA–specific method, the RIN scores were all >8. However, when the same method was used for human samples, the RIN was <6. The Ebbert group has been working to optimize the human brain RNA extraction. Both CSHL and the Ebbert group have also received new cDNA-PCR kits from ONT with a more stable motor protein that is expected to improve RNA yield. Testing will continue in 2022.

The second project aimed to extract and perform single individual long-read sequencing from aphids and springtails in collaboration with Rosanna Giordanno. Aphids and springtails are important pests and environmental decomposers, respectively. The mechanisms that drive aphids to be better pests and springtails to inhabit diverse environments are not well understood. Although bulk sequencing has been done on each specimen, single-individual sequencing can provide superior assemblies and methods, especially for the

parthenogenetic springtail. CSHL found that liquid nitrogen grinding of single aphids in a 1.5-mL tube with a glass pestle followed by standard phenol chloroform extraction yielded >500 ng of DNA from a single aphid. For springtails the mass was <300 ng. Initial sequencing of single springtails was promising, with many reads mapping to the existing references for both the insect and their known symbiotes. Unfortunately, the coverage was not high enough for a quality assembly. The PacBio gDNA amplification protocol was used to increase the mass of the springtail DNA samples. CSHL found that overamplification (causing many duplicate reads) was very problematic and affected half of all preps. Careful optimization of PCR cycles along with parallel reactions reduced this issue. The resulting data contained palindrome sequences as an artifact of amplification and subread generation. Work is ongoing to determine whether Pacasus (Warris et al., *BMC Genomics* **19:** 798[2018]) can remove the palindromes prior to assembly. Modifications to the HiFi protocol did allow adequate coverage of one of the two aphids without prior gDNA amplification. These modifications involve smaller shearing of the DNA (~10 kb) and bead-based size selection to remove fragments <6 kb. The second aphid samples were too degraded to be sequenced. Work on assembly and additional sequencing is ongoing.

The third project had been an extension of a long-term project in collaboration with Zach Lippman at CSHL. Previously, CSHL carried out high-coverage sequencing on 100 tomato strains with Oxford Nanopore. That work has since been published in *Cell* (Alonge et al., *Cell* **182:** 145 [2020]). The Lippman laboratory then switched to PacBio technology to take advantage of the higher accuracy of HiFi data (99.9% compared to ~90% on ONT). Initially, CSHL was using the standard PacBio HiFi protocol, which provided a mean read length of ~12–15 kb. These lengths proved acceptable for most applications, but longer read lengths were requested to improve assembly contiguity. The McCombie laboratory validated options (higher DNA input, more accurate shearing, more aggressive size selection) to achieve mean read lengths for HiFi data up to 20 kb. The increased read length also leads to greater circular consensus sequencing (CCS) yield—up to 30 Gb per flow cell. At this time, 10 tomato samples have been processed via the longer-read HiFi method, and we will continue this work in 2022.

High-Resolution Copy Number Inference in Cancer Using Short-Molecule Nanopore Sequencing

S. Goodwin, R. Wappel, W.R. McCombie [in collaboration with T. Baslan, Y. Zhang, S. Lowe, MSKCC; S. Kovaka, M. Schatz, CSHL/Johns Hopkins; F.J. Sedlazeck, Baylor College of Medicine]

Four years ago, the next-generation sequencing (NGS) core began exploring how the speed and simplicity of nanopore sequencing can be used to detect copy number variations in acute myeloid leukemia (AML). Many changes in the technology required updates to the project, and completion was delayed. As of 2020 ONT has become a more stable technology, allowing for a final analysis of the results. We have shown that short reads (<1 kb) on the MinION device can generate millions of reads with and without barcodes. These millions of reads represent <1× coverage of the genome and are sufficient to recapitulate all clinically relevant copy number variation detected by Illumina sequencing of the same sample. Using the ultra-low-cost ONT Flongle we have shown that <200,000 reads are sufficient for detection of major copy number variations (CNVs). This number of reads is typically generated within the first 2 hours of a sequencing run. Library prep and analysis each take ~2 hours. Altogether, the time from DNA to clinically actionable results is <1 day. These results show that low cost, high portability, and rapid turnaround time make the ONT MinION an attractive option for the detection of important CNVs in AML. In particular, this could be implemented in underserved populations, in which karyotyping or traditional sequencing technologies may not be readily available. This work was published in *Nucleic Acids Research* (Baslan et al. 2021).

PUBLICATIONS

Baslan T, Kovaka S, Sedlazeck FJ, Zhang Y, Wappel R, Tian S, Schatz MC. 2021. High resolution copy number inference in cancer using short-molecule nanopore sequencing. *Nucl Acids Res* **49:** e124.

Chou HC, Bhalla K, Demerdesh OE, Klingbeil O, Hanington K, Aganezov S, Andrews P, Alsudani H, Chang K, Vakoc CR, et al. 2021. The human origin recognition complex is essential for pre-RC assembly, mitosis, and maintenance of nuclear structure. *eLife* **10:** e61797. doi:10.7554/eLife.61797

Jia X, Goes FS, Locke AE, Palmer D, Wang W, Cohen-Woods S, Genovese G, Jackson AU, Jiang C, Kvale M, et al. 2021. Investigating rare pathogenic/likely pathogenic exonic variation

in bipolar disorder. *Mol Psychiatry* **26:** 5239–5250. doi:10.1038/s41380-020-01006-9

Malik SA, Zhu C, Li J, LaComb JF, Denoya PI, Kravets I, Miller JD, Yang J, Kramer M, McCombie WR, et al. 2021. Impact of preoperative antibiotics and other variables on integrated microbiome-host transcriptomic data generated from colorectal cancer resections. *World J Gastroenterol* **27:** 1465–1482. doi:10.3748/wjg.v27.i14.1465

Scheben A, Mendivil Ramos O, Kramer M, Goodwin S, Oppenheim SJ, Becker DJ, Schatz MC, Simmons NB, Siepel A, McCombie WR. 2021. Long-read sequencing reveals rapid evolution of immunity- and cancer-related genes in bats. bioRxiv doi:10.1101/2020.09.09.290502

Yang Z, Wu XS, Wei Y, Polyanskaya SA, Iyer SV, Jung M, Lach FP, Adelman ER, Klingbeil O, Milazzo JP, et al. 2021. Transcriptional silencing of *ALDH2* confers a dependency on Fanconi anemia proteins in acute myeloid leukemia. *Cancer Discov* **11:** 2300–2315. doi:10.1158/2159-8290.CD-20-1542

In Press

Bhatia S, Kramer M, Russo S, Naik P, Arun G, Brophy K, Andrews P, Fan C, Perou CM, Preall J, et al. 2022. Patient-derived triple-negative breast cancer organoids provide robust model systems that recapitulate tumor intrinsic characteristics. bioRxiv doi:10.1101/2021.08.09.455691

PLANT GENOME ORGANIZATION, EVOLUTION, AND FUNCTION

D. Ware J. Braynen, K. Chougule, A. Fahey, N. Gladman, F. Hu, C. Kim, V. Kumar, S. Kumari, F. Li, Z. Lu, I. Meijs, An. Olson, Au. Olson, M. Regulski, M.K. Tello-Ruiz, P. Van Buren, B. Wang, L. Wang, X. Wang, S. Wei, L. Zhang

The Ware laboratory has two primary goals: (1) understanding the genomic architecture and molecular pathways underlying traits of interest associated with the development and environmental response in agriculturally important crop plants; and (2) development of infrastructure, including standards, tools, and data sources, 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. Sustained efforts in many disparate disciplines are required to generate and annotate 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 scaffolds. The scaffolds, in turn, discern the position, relative order, and orientation of contigs within the chromosomes. The next step is annotation—the discovery and description of genes and other functional elements, followed by inference of homologies (evolutionary relationships) with other genomes. This information must be faithfully communicated and visualized in open-science, web-based platforms such as Gramene that promote FAIR (findable, accessible, interoperable, reusable) data usage.

All 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 reference genomes for individual species and supporting the development of multispecies representation as a "pan-genome." Ongoing projects within the maize, rice, sorghum, grape, and *Arabidopsis* 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 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. Such DNA modifications can be studied using new sequencing technologies and analytical methods.

Developing Pan-Genome Reference Resources for Crops

K. Chougule, B. Wang, S. Wei, A. Olson, Z. Lu, M.K. Tello-Ruiz [in collaboration with K. Fengler and V. Llaca, Corteva Agriscience; C. Hirsch, University of Minnesota; M. Hufford, Iowa State University; K. Dawe, University of Georgia; R. Wing, King Abdullah University of Science and Technology; E. Mace, University of Queensland; L. Cooper, University of North Carolina at Charlotte; D. Cantu, UC Davis; C. Rustenholz, INRAE COLMAR Université de Strasbourg; J. Vaughn, USDA]

Complete and accurate reference genomes are imperative for sustained progress in understanding the

genetic basis of trait variation and crop improvement. Continued advances in sequencing and assembly technologies are generating an abundance of high-quality reference assemblies within crop species, ushering a transition from single-genome to pan-genome research approaches. With this transition, communities will need ready access to precomputed comparisons of genome assemblies to identify and characterize common and variable regions. The Ware laboratory has played a leading role in the development and stewardship of several commodity crops, including rice, sorghum, maize, sugarcane, and grapevine, for more than a decade. To productionalize this initiative, the Gramene comparative genomics project supported by the Ware laboratory has launched a number of pan-genome subsites targeting these commodity crops.

A major effort during 2021 has been to support gene structural annotation of subpopulations of Asian cultivated rice. Asian cultivated rice is a staple food for half of the world population. With the planet's population expected to reach 10 billion by 2050, farmers must increase production by at least 100 million metric tons per year. To address this need, future rice cultivars should support higher yields, better nutrition quality, lower environmental footprint, and increased resilience to a variety of biotic and abiotic stresses. In the last year, a major effort was to annotate the rice MAGIC 16 population. The MAGIC 16 population includes 12 recently sequenced subpopulations of cultivated Asian rice for which no high-quality reference genomes existed, combined with four previously published genomes (Minghui 63 [MH 63], Zhenshan 97 [ZS 97], N 22, and IRGSP RefSeq). The 12 newly assembled genomes were generated with more than 100x genome coverage PacBio long-read sequence data and then validated with Bionano optical maps. The contig N50 value, which represents the sequence length of the shortest contig when half of the genome is represented in contigs, ranged from 7.35 Mb to 31.91 Mb. In the last year, the Ware laboratory generated uniform structural and functional annotations for all 16 accessions. The gene structural annotations were generated using transcript information, with MAKER-P using accession-specific transcript information and *ab initio* gene predictions using SNAP, Augustus, and Fgenesh. The gene models were further improved using full-length transcripts and EST data using PASA. On average, ~9000 protein-coding models were updated using the transcript data with the majority of improvement in the extension of the untranslated regions (UTRs) of the transcripts and the characterization of novel transcripts. Functional domain annotation was carried out using InterProScan, and TRaCE (Olson and Ware 2022) was used to assign the most represented/common transcript based on domain coverage, protein length, and similarity to expressed transcripts. On average, each rice accession contained 37,000 protein-coding and 200 noncoding gene models per genome, with well over 624,860 independent gene models generated across the 16 lines. Using 26 *Oryza* genomes, including the 13 previously published and reference Nipponbare genomes and six outgroup species, ~37,000 protein-based gene trees were constructed using a representative transcript model (single "canonical protein" coding model). To complement the uniform gene annotations, uniform repeat annotation was generated across the 26 *Oryza* genomes using the Ensembl repeat masking pipeline based on repeatmasker. On average, 44% of the *Oryza* genomes were repeatmasked, with long terminal repeats (LTRs) and type II transposons representing the dominant repeat classes. The *Oryza* annotation data and genomes were made available in October 2021 during release 3 of the *Oryza* pan-genome subsite (https://oryza.gramene.org).

Our work in grapevine included the second release of the Vitis pan-genome subsite (https://vitis.gramene.org/) and support for improved reference gene annotations for the PN40024 reference assembly, first released in 2007 (Jaillon et al., *Nature* **449:** 463 [2007]) as part of the French–Italian Public Consortium of Grapevine Genome Characterization. Vitis pan-genome release 2 provides access to 11 grape reference genomes, together with seven other species selected for comparative analysis, resulting in more than 24,000 protein-coding gene family trees. Although the grapevine reference assembly and gene structural annotations have been updated in the past, there have been limited resources such as expression data associated with the PN40024 clone to support and improve the gene structural annotations. This year, we worked with INTEGRAPE, a project funded by European Cooperation in Science and Technology (EU-COST Action) to support improved gene structural annotations using Nanopore sequencing from Oxford

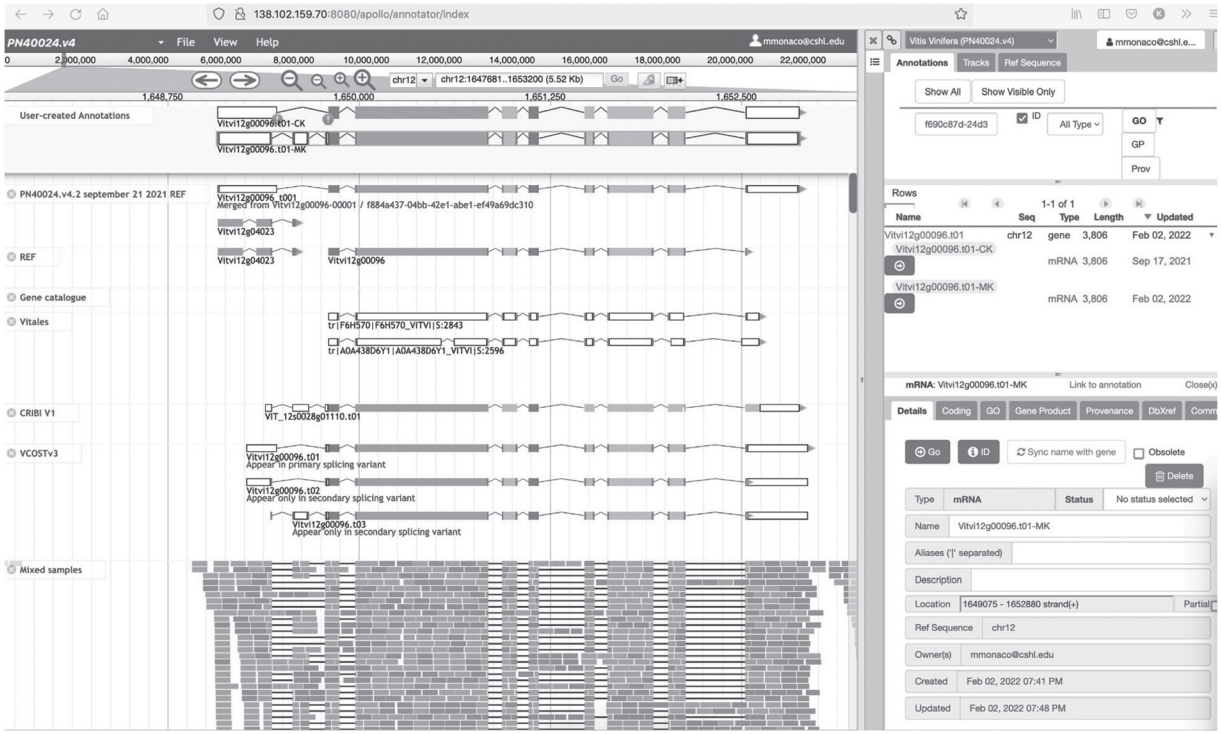


Figure 1. Improved grapevine gene model resulting from merging two incorrectly annotated gene models (*Vitvi12g04023 and Vitvi12g00096*) and extending the 5′ UTR using experimental (mRNA and orthologous protein) evidence overlaid in the Apollo genome editor.

Nanopore Technologies (ONT) and Pacific Bio-Sciences (PacBio) single-molecule real-time (SMRT) long-read isoform sequencing (Iso-seq). These methods offer many advantages over widely used high-throughput short-read RNA-seq approaches and allow a comprehensive analysis of transcriptomes in identifying full-length splice isoforms and several other posttranscriptional events. In our preliminary work using the new full-length transcripts from seven tissues, we have been able to update ~12,000 protein-coding models (the majority of updates are in the UTRs) and characterizations of novel isoforms of existing genes. Analysis of plant transcriptomes with these new powerful methods that require minimum sample processing is likely to become the norm and is expected to uncover novel co-/posttranscriptional gene regulatory mechanisms that control biological outcomes during plant development and in response to various stresses. In addition to using computational workflows to update the gene annotations, we are working with the community to support "gold standard" curated models for the "grapevine gene catalog." The catalog is enriched for genes associated with fruit ripening. As part of the collaboration, 356 grapevine genes in six transcription factor families were hand-curated, and training materials were developed to support community adoption that complies with the COST standards regarding metadata and FAIRness (Fig. 1). These will be used as part of the benchmarking data sets to compare the automated updated annotations using the full-length UTRs.

Last year, we released the second version of the Maize pan-genome web portal (https://maize-pangenome.gramene.org). This builds on our previous work focused on refining the gene annotations using a hybrid annotation workflow for 26 parents of the maize nested association mapping (NAM) panel that includes approximately 5000 recombinant inbred lines distributed across the globe. The release includes 35 assembled genomes, including the 25 maize NAM founders and the B73 reference (Hufford et al. 2021). Across the 26 maize genomes, there are a total of 1,056,152 genes. Most genes share orthologs with the grass (Poaceae) family and species in the Andropogoneae tribe of grasses, which includes maize and sorghum—of which 673,520 are found in both sorghum and maize,

and 197,280 specifically in maize. We are currently working on automating workflows to support defining the core/near-core portion of the pan-genome. The core genes are generally conserved in all higher plants (i.e., Viridiplantae and Poaceae), whereas those in the near-core and dispensable sets either share orthologs only with closely related species or are maize-specific. Previous work has found that the genes expressed in flowers are more likely to include genes that are more specific to sorghum and maize in comparison to other tissues (Wang et al., *Genome Res* **28:** 921 [2018]).

PLANT GENETICS AND SYSTEMS BIOLOGY

The global challenges confronting agricultural security are coming into sharper focus: 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 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 achieved 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.

Developmental Networks Controlling Root System Architecture during Limiting Nutrient Conditions

N. Gladman, M. Regulski, S. Kumari, J. Braynen, X. Wang, L. Wang, K. Chougule, D. Ware [in collaboration with R. Martienssen, CSHL; B. Hufnagel, Centre National de la Recherche Scientifique, Montpellier, Languedoc-Roussillon, France; Z. Liu and L. Kochian, Global Institute for Food Security, University of Saskatchewan, Saskatoon, Canada; and Jura Magalhães, Embrapa Milho e Sorgo, Sete Lagoas, Brazil]

Micro- and macronutrient content within the soil is a common limiting factor for plant growth and overall yield in agricultural crops. Understanding how the root system architecture (RSA) develops and adapts to marginal conditions is important for optimizing germplasms for growth on diverse soils and varying farm management practices. To study this, we have further developed hydroponic screening methods for the purposes of precisely controlling nutrient conditions of sorghum plants in normal, limiting, or excess levels of zinc, iron, nitrogen, and phosphorus. Hydroponic screening permits us to phenotype plant response at both the root and shoot level—visualizing RSA changes as well as leaf and stem health and growth (Fig. 2). We see the canonical effect of nutrient limitation and excess conditions through growth retardation, leaf chlorosis, and expansion or contraction of particular regions of the RSA. Using hydroponics also allows us to sample an array of important tissues for transcriptomic and epigenetic profiling, as we have performed with limiting phosphorus conditions and now continuing for iron and zinc. We have screened different germplasms based on a combination of their importance to the sorghum community (natural variant populations, mapping parental lines, reference genome lines, etc.) and existing public data about nutrient use efficiency.

Developmental Networks Controlling Inflorescence Architecture and Grain Yield in Grasses

F. Li, L. Wang, X. Wang, N. Gladman, and D. Ware [in collaboration with D. Jackson (PI), T. Gingeras, and J. Gillis, CSHL; and Z. Xin, USDA ARS]

Crop productivity depends on the activity of meristems that produce optimized plant architectures. The objective of this work is to integrate genetics and genomics data sets to find molecular networks that influence

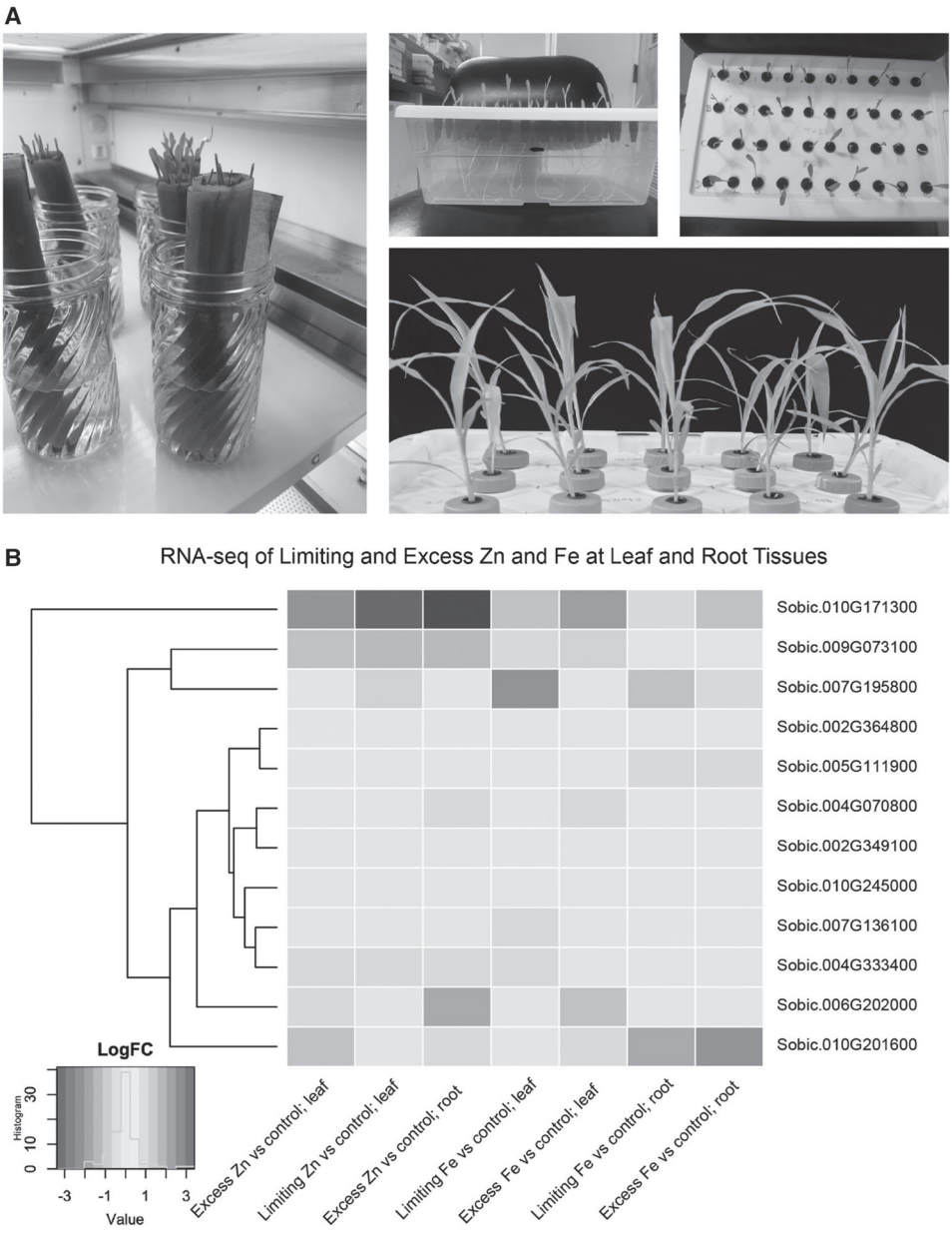


Figure 2. Micro- and macronutrient screening in sorghum using hydroponics. (*A*) Photographs of the hydroponic growing procedure: seedlings are germinated on brown germination paper in mason jars for 4–5 d (*left*); similar-sized seedlings are transplanted to minichambers in black 1.5-mL tubes and grown for 7 d (*top right*); and, finally, seedlings are moved into full-scale hydroponic tanks for the duration of treatment or until a particular stage of growth (*bottom right*). Nutrient media can be modified to contain any level of micro- or macronutrients as an abiotic stress condition. (*B*) Heatmap of RNA-seq data derived from roots and shoots that were sampled from hydroponically grown leaves and roots; conditions were limiting or excess micronutrients zinc or iron.

the morphology (architecture) of grass inflorescences (flowers), with a focus on *Zea mays* and *Sorghum bicolor* [L.] 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. A comprehensive understanding

of development requires insight into the full diversity of cell types and developmental domains, and the gene networks required to specify them. Until now, these were primarily classified by morphology and insights from classical genetics, which are limited by genetic redundancy and pleiotropy.

In collaboration with Dave Jackson's and Jesse Gillis' group at CSHL, we generated a high-resolution gene expression atlas of the maize meristem and demonstrated that distinct sets of genes govern the regulation and identity of stem cells. By using known markers, our collaborators identified meta-clusters of genes that identified spatial regulators of ear development, including discrete domains that control branching, vasculature growth, and determinate lateral organs. We further showed that the combined resolution of single-cell RNA-seq (scRNA-seq) and transcription factor chromatin immunoprecipitation sequencing (ChIP-seq) binding experiments could predict and validate directly modulated targets of transcription factors (TFs), such as the MADS networks controlling inflorescence development and epidermal differentiation. These transcriptomic networks can better predict genetic redundancy and identify candidate genes associated with beneficial crop yield traits (Wu et al., *Nat Genet* **52:** 118 [2020]).

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. However, only the sessile spikelet is fertile and will produce seeds. Previously, using a publicly available ethyl methanesulfonate (EMS) population, we identified independent multiseeded (*msd*) mutants that manifest both fertile sessile and pedicellate spikelets throughout the inflorescence. We are extending this work to generate transcriptome data from a single cell of developing meristem. Access to data from both maize and sorghum will allow us to compare what modules are the same and which differ between these two important grass species.

Characterization of Regulatory Networks Associated with Drought Response in Sugarcane

K. Chougule, A. Olson, S. Kumari, F. Li [in collaboration with G.M. Souza (PI) and A.L. Diniz, University of São Paulo]

Sugarcane is used for the production of sugar, biofuels, and biopolymers used in many products, including makeup, perfume, and pharmaceuticals. Modern sugarcane cultivars are polyploids, with genome sizes estimated to be 10 GB, containing more than 350,000 genes. In collaboration with Glaucia M. Souza from the University of São Paulo in Brazil, we are contributing to the development of improved gene predictions and characterization of the transcriptional landscape of differentially and coexpressed genes among genotypes contrasting for biomass production and response to drought. As a starting point for the work, in the last year, we were able to assign sugarcane genes to putative metabolic pathways using *Saccharum spontaneum* and *Sorghum bicolor*. Using gene expression differences among sugarcane genotypes contrasting for biomass production in drought conditions, we have been able to identify groups of genes associated with an increase in biomass under drought (Diniz et al., *Int J Mol Sci* **21:** 9124 [2020]). From the transcriptomic analysis of sugarcane genotypes contrasting for biomass production, we have revealed 73,946 novel transcripts not previously described and detected changes in the expression level of genes involved in amino acid, lipid, and carbohydrate metabolism. These changes are potentially regulated by phytohormones, such as ethylene, jasmonic acid, gibberellin, and abscisic acid, because these pathways are also enriched among differentially expressed genes. In addition, coexpression analysis revealed a potential role for ethylene as a regulator of high biomass production in sugarcane. Highly connected genes, up-regulated in upper internodes of high-biomass genotypes and previously not functionally annotated, were discovered. This work will aid in identifying additional genes for improved biomass and drought resistance in sugarcane.

Characterizing Master Regulators of Primary and Lateral Root Development and Response to the Environment

L. Zhang, Andrew Olson, S. Kumari, V. Kumar, A. Fahey, D. Ware [in collaboration with Y.K. Lee and C.C. Hu, Korea Institute of Fusion Energy, Jeollabuk-do, South Korea; 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 both a gene-centric and global

approach to characterize upstream regulators, TFs of microRNA (miRNA) families known to be involved in root development in the model plant *Arabidopsis thaliana*. Our approach uses a nearly complete root 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 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 to prioritize a set of TFs. From this data, we have prioritized 14 TF gene families. To evaluate their impact on root development, we are using genetically perturbed lines, mutants, that result in a complete loss or decrease in functions. The preliminary screen of more than 165 loss-of-function TFs found that ~15% exhibit a change in the primary root length. Using the TF mutants, we have generated new germplasm that contains multiple mutations to address the functional redundancy of the different family members. We are currently screening these germplasms for changes in root development.

Characterization of a Novel Zinc Chaperone in *Arabidopsis*

L. Zhang, F. Hu, A. Fahey, D. Ware [in collaboration with C. Blaby-Haas, M. Pasquini, and M. Xie, Brookhaven National Laboratory]

Transition metals occupy numerous and often essential positions within the biochemical framework of plant metabolism. As protein cofactors, these elements enable pivotal energy-intensive reactions and have expanded the breadth of protein-catalyzed reactions. To use metal ions as catalysts, the cell must balance a fundamental dichotomy: nutrient and toxin. Metal homeostasis has evolved to tightly modulate the availability of metals within the cell, avoiding cytotoxic interactions due to excess and protein inactivity due to deficiency. Even in the presence of homeostasis processes, however, low bioavailability of these essential metal nutrients in soils can negatively impact crop health and yield. Whereas research has largely focused on how plants assimilate metals, acclimation to metal-limited environments, such as marginal soils, requires a suite of strategies that are not necessarily involved in metal transport. The identification of these assimilation-independent mechanisms provides an opportunity to improve metal-use efficiency and optimize feedstock in low-nutrient soils without supplementing with expensive and environmentally damaging fertilizers.

Zinc is a vital micronutrient for plants. It is a constituent of approximately 2400 *Arabidopsis* proteins and an essential cofactor of enzymes and many regulatory proteins. However, how zinc-dependent proteins are bound to zinc was previously unknown. By leveraging phylogenomic and data-mining analyses combined with an interdisciplinary experimental approach, we have discovered a novel metal chaperone that delivers zinc to an essential zinc-dependent enzyme during zinc deficiency. We provide evidence that this function is universally conserved from fungi to plants. In plants, we propose that duplication has resulted in analogous zinc-trafficking pathways in the cytosol and chloroplast. In *Arabidopsis*, there is one zinc chaperone (ZNG1) in the cytoplasm and two paralogous chaperones (ZNG2A1 and ZNG2A2) localized to the chloroplast. We have obtained and characterized corresponding mutants for ZNG1, ZNG2A1, ZNG2A2, and MAPs (the target zinc-dependent proteins) and made several crosses among these mutants. Here we present preliminary data on the characterization of plant ZNG1. Using yeast two-hybrid (Y2H) and bimolecular fluorescence complementation (BiFc) studies, we show that ZNG1 can interact with the cytosolic target methionine aminopeptidase (MAP1A), as we have found in yeast. Our hypothesis is that plant ZNG1 is an activator of cytosolic MAP1, and cytosolic MAP2 can function as a backup enzyme when cytosolic MAP1 activity is compromised. We are in the process of confirming this hypothesis.

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 1000 Genomes to catalog human genetic variations, 1000 Plant Genomes to sequence expressed genes of 1000 different plant species, 1001 Genomes to discover whole-genome sequence variations in more than 1001 strains of 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 has continued to

be a challenge. As a result, the scale and complexity of genomics research have advanced from studying a 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 cyber-infrastructure platforms that adhere to the FAIR 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, C. Kim, V. Kumar, S. Kumari, A. Olson, M.K. Tello-Ruiz, B. Wang, S. Wei, L. Zhang [in collaboration with P. Jaiswal, Oregon State University; B. Contreras-Moreira and I. Papatheodorou, 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 2021, the project accomplished several major milestones, culminating in our 64th data release (September 2020), which included 114 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., *Nucl Acids Res* **48:** D1093 [2020]; Gillespie et al., *Nucl Acids Res* **50:** D687 [2022]).

The Gramene project is actively engaging 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—providing training and the community's feedback on our current tools and user suggestions for new functionality. In 2021, 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. User guides were generated for each of these sites (see, e.g., https://vitis.gramene.org/guides). Although generating a reference assembly has progressively become easier, there are still major challenges to developing 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.

SorghumBase: A Web-Based Portal for Sorghum Genetic Information and Community Advancement

N. Gladman, A. Olson, S. Wei, K. Chougule, Z. Lu, M. K. Tello-Ruiz, I. Meijs, P. Van Buren, B. Wang, V. Kumar, S. Kumari, L. Zhang, D. Ware [in collaboration with Y. Jiao, Texas Tech University; J. Burke, J. Chen, G. Burow, C. Hayes, Y. Emendack, and Z. Xin, USDA-ARS Cropping System Research Laboratory, Lubbock, Texas]

Many agriculturally important crops have dedicated web resources for their genomic and phenotypic

data (MaizeGDB, SoyBase, GrainGenes, T3: Triticeae toolbox). To provide a similar resource for *Sorghum bicolor*, we created SorghumBase.org, a modular web-based utility that houses genetic, genomic, pan-genomic, transcriptomic, and molecular biology data for more than a dozen different important sorghum cultivars. The currently active SorghumBase site contains data, external links, and visualizations for gene trees, syntenic regions, gene expression, molecular pathways, single-nucleotide polymorphism (SNP) variants, and sequence data for the primary reference genome BTx623, as well as 17 additional cultivars (Fig. 3). SorghumBase offers BLAST and genome browser tools, as well as synteny queries and assembly statistics for genomes. There is also an education and training section for the site, and we are actively adding and improving current content and functionality. Moving forward, we intend to incorporate additional use cases and services to benefit the larger sorghum research community, but with notable emphasis for breeders. The site was officially launched in July of 2021, and a subsequent release was done in November. For its third release, we began developing materials (standard operating procedures, templates for emails, publications, events, news posts, etc.) to gather and incorporate relevant information into the site and thus improve its usability.

SciApps: A Cloud-Based Reproducible Workflow Platform

L. Wang, Z. Lu, X. Wang, P. Van Buren [in collaboration with CSHL principal investigators (PIs) T. Gingeras, K. Birnbaum, D. Jackson, R. Martienssen, W.R. McCombie, D. Micklos, M. Schatz, D. Ware]

SciApps is 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. 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. A SciApps workflow captures the 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. SciApps has been used to host the workflows and data sets for the maize and sorghum community during the past two years.

The MaizeCODE project aims to characterize functional elements in the maize genome by monitoring gene expression and DNA modifications, including methylation or the binding of histones and transcription factors in various tissues. MaizeCODE is also tasked with disseminating the primary and analyzed data to the wider community of plant biologists worldwide. In its initial phase, MaizeCODE assayed five tissues of four maize accessions (B73, NC350, W22, TIL11) for RNA-seq, ChIP-seq, Rampage, small RNA, and MNase (outside collaboration). In support of this effort, we further improved SciApps by adding a RESTful API for automating batch processing of the MaizeCODE data and metadata management. To improve the accessibility of the data, a searchable data page was developed for the MaizeCODE data. The page is powered by a relational database and supports access to several analysis workflows and Genome Browser tracks, which are automatically generated from unique workflow identifiers via the RESTful API.

Working with the sorghum community, we added bulk segregant analyses (BSAs) to SciApps. BSAseq is not only a powerful tool for mapping quantitative trait loci, but also a useful way to identify causal gene mutations underlying phenotypes of interest. However, because of the presence of background mutations and errors in sequencing, genotyping, and reference assembly, it is often difficult to distinguish true causal mutations from background mutations. The BSAseq workflow in SciApps includes an automated bioinformatics analysis pipeline with a probabilistic model for estimating the linked region (the region linked to the causal mutation) and an interactive web application for visualizing the results. To demonstrate the utility, we deeply sequenced a sorghum male-sterile parental line (ms8) to capture the majority of background mutations in our bulked F_2 data. We applied the workflow to 11 bulked sorghum F_2 populations and one rice F_2 population and identified the true causal mutation in each population. The workflow is intuitive and straightforward, facilitating its adoption by users without bioinformatics skills for the identification of causal mutations for many phenotypes of interest.

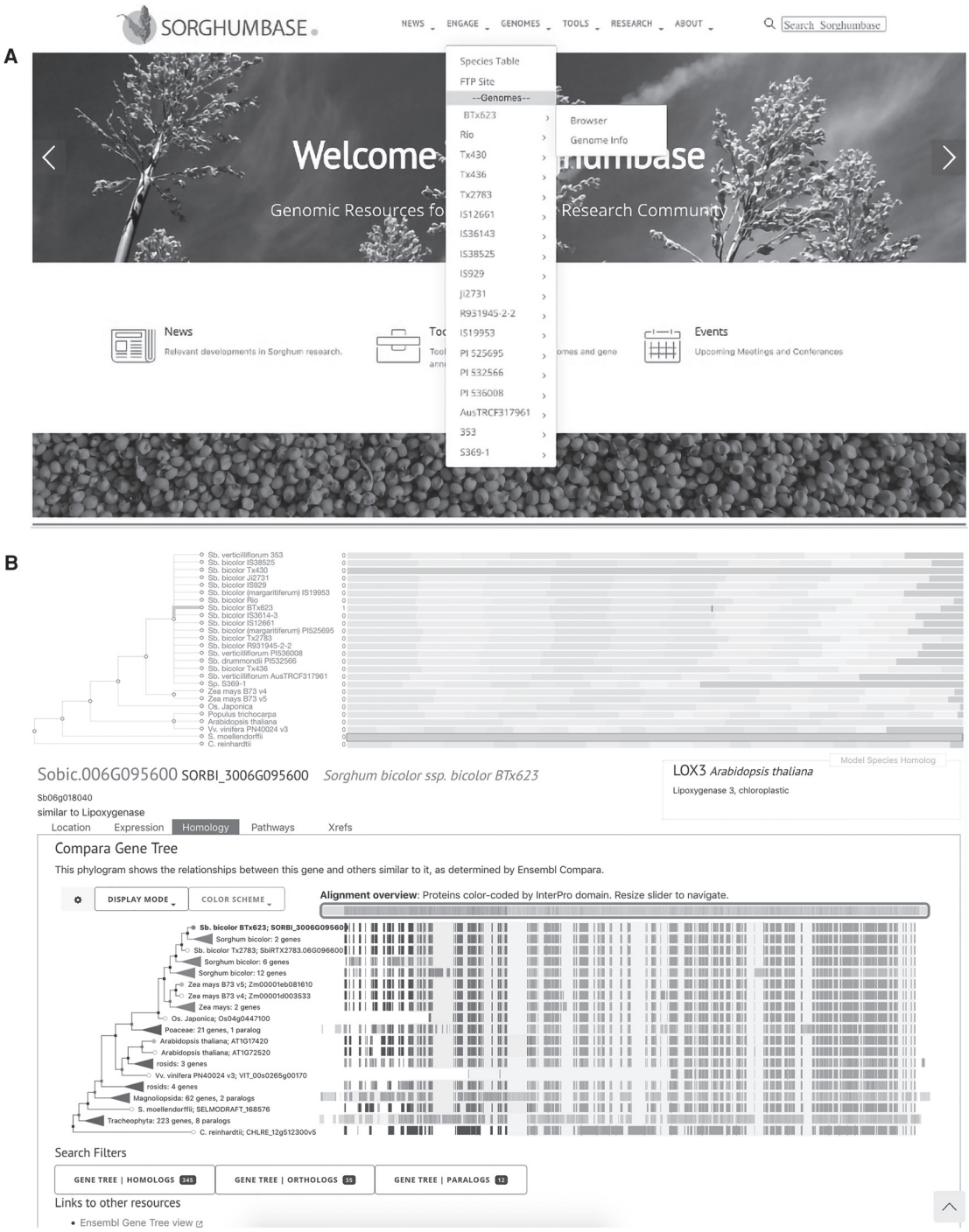


Figure 3. SorghumBase Web Portal. (*A*) Landing page of the SorghumBase.org website. Users can navigate to gene/keyword search options, FTP sites, and germplasm information for the 18 genomes for which data is currently housed on the site. (*B*) Gene search result for a lipoxygenase enzyme (SORBI_3006G09560) displaying homology views of related genes within sorghum and across other species, including a total-chromosome layout of all species for rapid assessment (red box). Other features are accessible from the results page, including gene expression data, genomic location information, and metabolic pathway relevance (if applicable).

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, 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. KBase enables secure sharing of data, tools, methods, and conclusions in a unified, extensible system where researchers collaboratively generate, test, and share hypotheses about biological functions; perform large-scale analyses on scalable computing infrastructure; combine multiple lines of evidence to accurately model plant and microbial physiology and community dynamics, and ultimately "publish" their work in FAIR ways. 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.

In 2021, KBase continued to invest in improving the stability and usability of the platform. KBase released the sample functionality that combines innovations in generic data representations, improvement in our ontology systems, and development of a novel sample representation, editing, access control, and search framework. In addition, significant improvements were made in metabolic reconstruction tools for plants, fungi, and microbes including the release of new metabolic map visualizations. KBase also added operational taxonomic unit (OTU) abundance and chemical abundance datatypes for metabolomics and amplicon data along with samples.

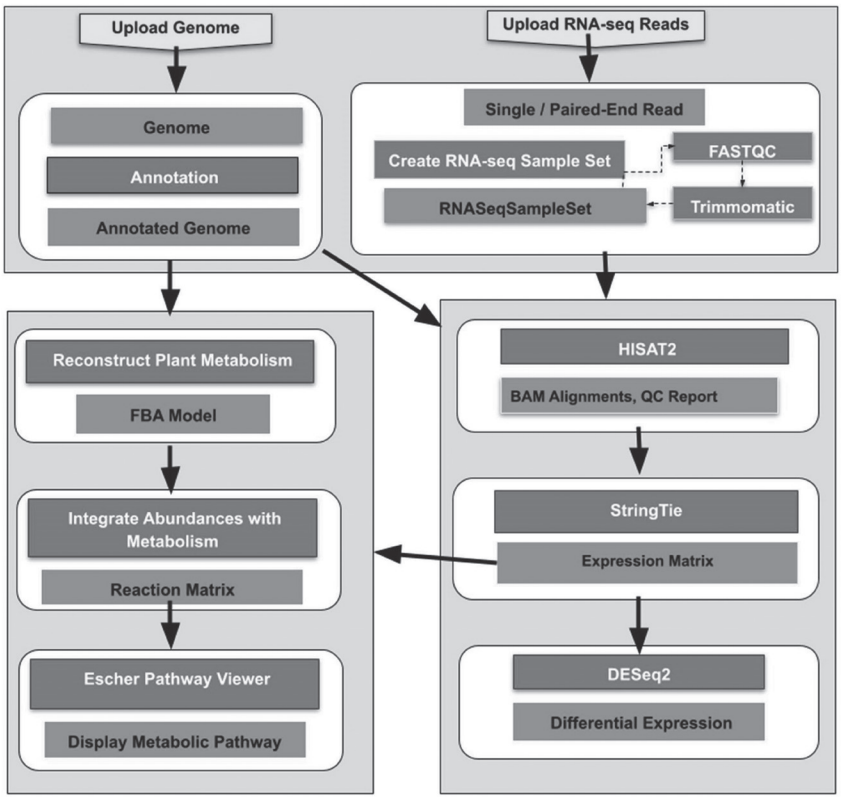


Figure 4. A schematic overview of KBase workflow on gene expression profiling and metabolic mapping with primary metabolism organized in three modules: module 1 (*top*)—import and preparation of genome and sequencing reads; module 2 (*bottom right*)—alignment, assembly, and quantification; module 3 (*bottom left*)—metabolic reconstruction, integration, and visualization of abundance data.

Although KBase aims to accelerate scientific discoveries and enable more advanced analyses, we also strive to empower users by expanding their knowledge and reach through training; sharing of their data, tools, and discoveries; and enabling the publication of their analyses. User working groups (UWGs) are a key part of our outreach strategy because they focus on areas of strategic importance and opportunity as a way to more intensely engage with specific, mutually interested individuals, projects, and groups who are capable of contributing resources and expertise to further develop KBase capabilities toward their science objectives. During 2021, the Ware laboratory continued and extended engagements with Epigenetic Control of Drought Response in *Sorghum bicolor* (EPICON) and SorghumSysBio (SSB) teams. In addition, we had additional engagements at KBase with two more research projects, Quantitative Plant Science Initiative (QPSI) with Brookhaven National Laboratory, and Plant–Pathogen Interaction project with Argonne National Laboratory (ANL). Our work with QPSI was presented at the Genome Science Planning Meeting (DOE GSP 2021) in a poster titled "Optimizing hydroponic growth system for metal stress studies of bioenergy crops." As a part of outreach efforts in collaboration with the ANL team, we worked on the planning and preparation of the workshop at the American Society of Plant Biologists (ASPB) meeting. We also presented at the Biology of Genomes meeting during 2021 Q2. Finally, in collaboration with the KBase team, we published our work featuring transcriptomics and plant metabolic modeling in KBase (Fig. 4) as a paper in *Current Plant Biology* titled "A KBase case study on genome-wide transcriptomics and plant primary metabolism in response to drought stress in Sorghum." We demonstrated in the manuscript how biologists can use the current RNA-seq tools in KBase, applicable to both plants and microbes, to assemble and quantify long transcripts and identify differentially expressed genes effectively. We then showed how biologists can use other KBase integrated tools such as prediction of metabolic pathway membership to find coherent differential and tissue-specific regulation of key plant subsystems, such as lignin biogenesis, that are differentially deployed under drought conditions.

PUBLICATIONS

Gladman NP, Hufnagel B, Regulski M, Liu Z, Wang X, Chougule K, Kochian L, Magalhaes J, Ware D. 2021. Sorghum root epigenetic landscape during limiting phosphorus conditions. bioRxiv doi:10.1101/2021.05.25.445633

Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, Liu J, Ricci WA, Guo T, Olson A, Qiu Y, et al. 2021. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373:** 655–662.

Kumari S, Kumar V, Beilsmith K, Seaver SMD, Canon S, Dehal P, Gu T, Joachimiak M, Lerma-Ortiz C, Liu F, et al. 2021. A KBase case study on genome-wide transcriptomics and plant primary metabolism in response to drought stress in Sorghum. *Curr Plant Biol* **28:** 100229.

Tello-Ruiz MK, Naithani S, Gupta P, Olson A, Wei S, Preece P, Jiao Y, Wang B, Chougule K, Garg P, et al. 2021. Gramene 2021: harnessing the power of comparative genomics and pathways for plant research. *Nucl Acids Res* **49:** D1452–D1463.

Tello-Ruiz MK, Wei S, Olson A, Gupta P, Naithani S, Chougule K, Kumari S, Kumar V, D'Eustachio C, Papatheodorou I, et al. 2021. Gramene's reference genomes and comparative resources for horticultural species. *Hortscience* **56:** S118.

Vaughn IN, Korani W, Stein JC, Edwards JD, Peterson DG, Simpson SA, Youngblood RC, Grimwood J, Chougule K, Ware DH, et al. 2021. Gene disruption by structural mutations drives selection in US rice breeding over the last century. *PLoS Genet* **17:** e1009389.

Wang B, Jiao Y, Chougule K, Olson A, Huang J, Llaca V, Fengler K, Wei X, Wang L, Wang X, et al. 2021. Pan-genome analysis in sorghum highlights the extent of genomic variation and sugarcane aphid resistance genes. bioRxiv doi:10.1101/2021.01.03.424980

Wang L, Lipzen A, Lu Z, Chen J, Wang X, Tello-Ruiz MK, Barry K, Mortimer J, Ware D, Xin Z. 2021. Comparative analysis of sorghum EMS mutants and natural populations. bioRxiv doi:10.1101/2021.06.06.447271

Wang L, Lu Z, Regulski M, Jiao Y, Chen J, Ware D, Xin Z. 2021. BSAseq: an interactive and integrated web-based workflow for identification of causal mutations in bulked F_2 populations. *Bioinformatics* **37:** 382–387.

Xu X, Crow M, Rice BR, Li F, Harris B, Liu L, Demesa-Arevalo E, Lu Z, Wang L, Fox N, et al. 2021. Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev Cell* **56:** 557–568.

In Press

Moreno P, Fexova S, George N, Manning JR, Miao Z, Mohammed S, Muñoz-Pomer A, Fullgrabe A, Bi Y, Bush N, et al. 2022. Expression Atlas update: gene and protein expression in multiple species. *Nucl Acids Res* **50:** D129–D140.

Olson AJ, Ware D. 2022. Ranked choice voting for representative transcripts with TRaCE. *Bioinformatics* **38:** 261–264. doi:10.1093/bioinformatics/btab542

Yates AD, Allen J, Amode RM, Azov AG, Barba M, Becerra A, Bhai J, Campbell LI, Carbajo Martinez M, et al. 2022. Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. *Nucl Acids Res* **50:** D996–D1003.

QUANTITATIVE BIOLOGY

The **Molly Gale Hammell** group studies transposons, viral-like parasites that lay dormant within our genomes. When active, transposons have the capacity to hop into new genomic locations and cause mutations as they break the surrounding DNA sequence. 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. Active transposons jeopardize the integrity of our genomes and interfere with normal cellular function. Mounting evidence has implicated transposon activity in a host of human diseases, with particular evidence for activation in neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS), Alzheimer's disease (AD), and frontotemporal dementia (FTD).

Ivan Iossifov's laboratory 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. His laboratory 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 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 at CSHL in 2010. His research focuses on developing next-generation DNA sequencing as a tool for dissecting the biophysical mechanisms of gene regulation. As a graduate student, Kinney co-invented a widely used technique now known as the massively parallel reporter assay (MPRA). Kinney and colleagues further showed how, using ideas from information theory, such experiments could be used to infer quantitative biophysical models describing how cells regulate gene expression. The Kinney laboratory continues to leverage a tightly knit combination of mathematical theory, machine learning, and experiments in order to illuminate the biophysics of gene regulation in two diverse contexts: bacterial transcriptional regulation and alternative mRNA splicing in humans. This latter context is highly relevant to understanding and treating human diseases like spinal muscular atrophy and cancer. The Kinney laboratory also develops algorithms and software for the analysis of MPRAs and other multiplex assays of variant effect (MAVEs).

The **Peter Koo** group develops methods to interpret high-performing deep-learning models to distill knowledge that they learn from big and noisy biological sequence data. Deep learning is being applied rapidly in many areas of genomics, demonstrating improved performance over previous methods on benchmark data sets. Despite the promise of deep learning, it remains unclear whether improved predictions will translate to new biological discoveries because of their low interpretability, which has earned them a reputation as a black box. Understanding the reasons underlying a deep-learning model's prediction may reveal new biological insights not captured by previous methods. Their goal is to elucidate biological mechanisms that underlie sequence–function relationships for gene regulation and protein (dys)function. Recently, they have teamed up with other members of the CSHL Cancer Center to investigate the sequence basis of epigenomic differences across healthy and cancer cells.

Alexander Krasnitz and colleagues develop mathematical and statistical tools to investigate population structure of cells constituting a malignant tumor and to reconstruct evolutionary processes leading up 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 genomic 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.

The **David McCandlish** 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. McCandlish and colleagues 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 drug resistance phenotypes in both populations of cancer cells and rapidly evolving microbial pathogens.

The thymus generates and selects a highly variable yet specific T-cell repertoire that discriminates between healthy and nonhealthy self and dangerous non-self antigens. **Hannah Meyer**'s research group uses a systems immunology approach to dissect the mechanisms crucial to the selection processes in the thymus. Her team develops experimental techniques and combines the resulting data with innovative computational models to generate accurate and testable hypotheses about tissue-level organ physiology.

Studying thymus physiology from a qualitative and quantitative perspective will provide us with a more fine-grained understanding of the selection processes and their downstream consequences such as autoimmunity, cancer immunosurveillance, and immune deficiency.

Saket Navlakha's laboratory studies "algorithms in nature" (i.e., how collections of molecules, cells, and organisms process information and solve interesting computational problems critical for survival). Indeed, there are many shared goals and constraints faced by biological and engineered systems, including (1) the use of distributed networks as a backbone for information

processing and communication; (2) trade-offs between optimization criteria, including efficiency, robustness, and adaptability; and (3) the need to develop low-cost, scalable solutions that conserve important metabolic or physical resources. An algorithmic perspective on biological problem-solving can lead to two ends: (1) new biological algorithms that are simple, flexible, and robust for use in computer science applications, and (2) quantitative frameworks to predict behavior, raise testable hypotheses, and guide experiments. The Navlakha laboratory has most recently focused on studying neural circuit computation and plant architecture optimization from this perspective.

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. **Adam Siepel**'s group focuses on a diverse collection of research questions in this interdisciplinary area, spanning applications in cancer biology, basic molecular biology, evolutionary genetics, infectious diseases, and agriculture. 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, and genetics.

Over the years, the Siepel group's 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, the identification of noncoding mutations important in cancer, and the discovery of ancient gene flow from humans to Neandertals. 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. They collaborate closely with experimentalists in cancer biology, transcriptional regulation, plant breeding, and many other areas.

THE CONTRIBUTION OF TRANSPOSABLE ELEMENTS TO NEURODEGENERATIVE DISEASE

M. Gale Hammell I. Bolger, C. Marshall, R. Shaw, T. Forcier, K. Natarajan, O. Tam, Y. Jin, K. O'Neill, C. Wunderlich

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. In addition, the byproducts of TE activity, such as double-stranded RNA and cytosolic DNA, can alert the host innate immune system and provide a trigger for inflammation. 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), frontotemporal dementia (FTD), and Parkinson's disease (PD).

Characterization of Genes That Control Transposable Element Activity

I. Bolger, C. Marshall, O. Tam

Human cells devote extensive resources to controlling the activity of TEs in order 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 TEs. 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, we have 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 fly, mouse, and human cells, that TDP-43 binding to TEs is lost in human patients diagnosed with FTLD, and that TDP-43 pathology is associated with elevated TE expression in human patients diagnosed with ALS. Although this firmly places TDP-43 in the list of genes that contribute to regulation of TE 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. Specifically, we have identified several TDP-43 co-factors that work with TDP-43 and are part of the known pathways that regulate TE expression, including repressive transcription factors that block transcription from TE loci and small RNA complex components that contribute to degrading TE transcripts.

Transposable Elements Mark a Subtype of ALS

K. O'Neill, R. Shaw, O. Tam [in collaboration with H. Phatnani, New York Genome Center and Columbia University]

Given the identification of TDP-43 protein as a regulator of TE activity, our laboratory has sought to

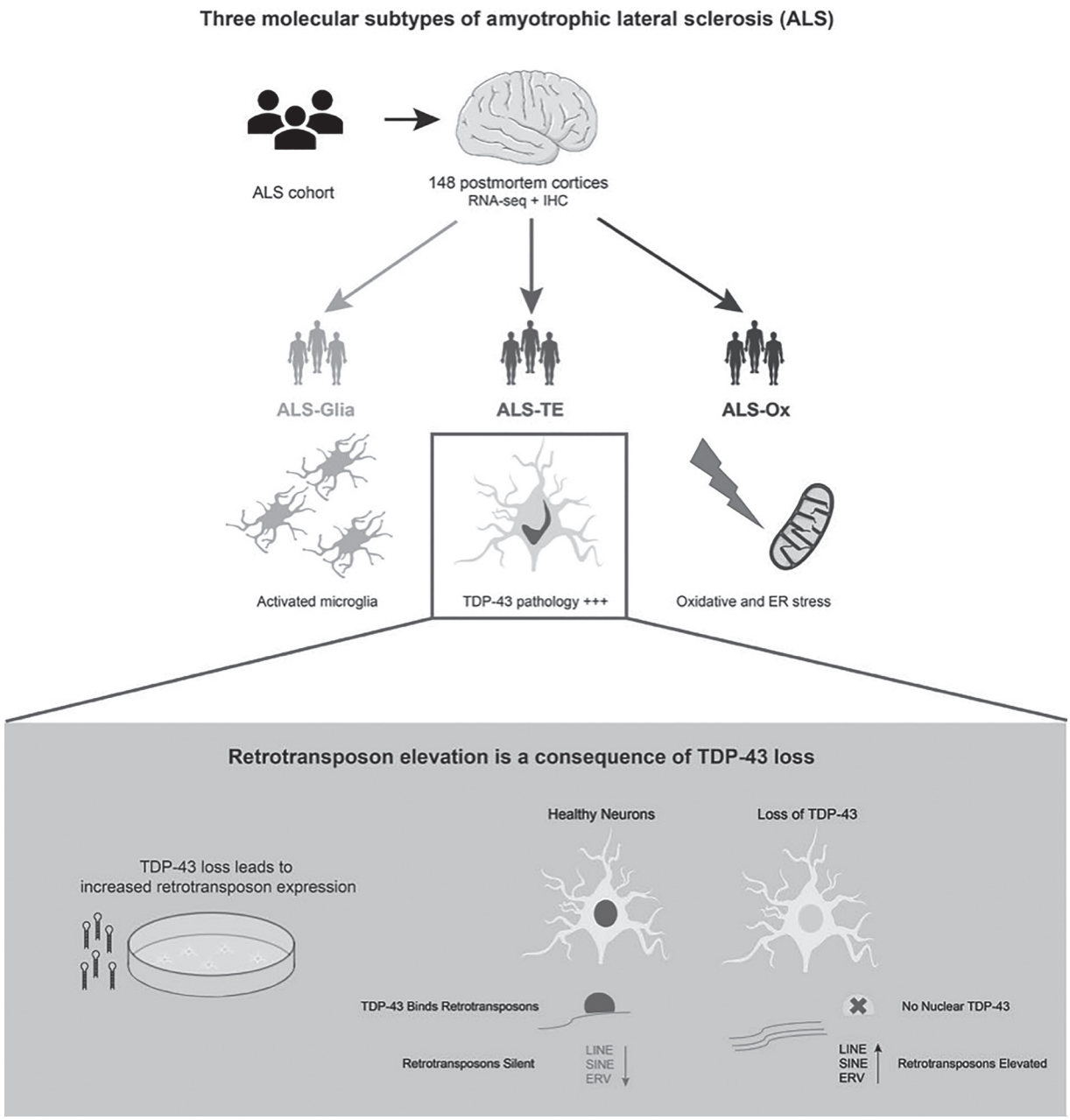


Figure 1. Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease characterized by the progressive loss of motor neurons. Although several pathogenic mutations have been identified, the vast majority of ALS cases have no family history of disease. Thus, for most ALS cases, the disease may be a product of multiple pathways contributing to varying degrees in each patient. Using machine-learning algorithms, we stratified the transcriptomes of 148 ALS postmortem cortex samples into three distinct molecular subtypes. The largest cluster, identified in 61% of patient samples, displayed hallmarks of oxidative and proteotoxic stress. Another 19% of the samples showed predominant signatures of glial activation. Finally, a third group (20%) exhibited high levels of retrotransposon expression and signatures of TARDBP/TDP-43 dysfunction. We further demonstrated that TDP-43 directly regulates retrotransposon transcripts and that TDP-43 aggregation pathology strongly correlates with retrotransposon elevation in ALS patient tissues (Tam et al., *Cell Rep* **29:** 1164 [2019]). Ongoing work in the Gale Hammell laboratory is following up on these results to identify biomarkers of the three ALS subtypes in peripheral tissues and biofluids.

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. 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, the laboratory next sought to understand why the ALS patient samples seemed to display such heterogeneity. The laboratory developed sophisticated machine learning methods based on nonnegative matrix factorization (NMF) algorithms in order 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 ALS samples could be delineated into three different subtypes characterized by three distinct types of cellular dysfunction (Fig. 1). 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. Ongoing work will explore methods to stratify patient populations such that targeted therapeutic strategies can be tailored to these ALS subtypes. Additional work is devoted to exploring the contribution of TDP-43 pathology and subsequent TE activity to other neurodegenerative diseases, such as AD, FTD, and PD.

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 genome, including within other gene sequences. Although most of these TE 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, TElocal for locus-specific RNA-seq analysis, TEsmall for small-RNA-seq, TEpeaks for ChIP-seq, and TEsingle for single-cell RNA-seq and nuc-seq data. All of 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 within the Gale Hammell laboratory and in several collaborative studies. Moreover, these algorithms form the most highly cited software suite for transposon genomics analysis. Together, these efforts will provide the computational infrastructure with which to determine the extent of TE activity in human development and disease.

Transposable Elements as a Trigger for Neuroinflammation in Parkinson's Disease

T. Forcier, O. Tam, C. Wunderlich [in collaboration with J. Jakobsson, Lund University; A. Kirkeby, the University of Copenhagen; and R. Barker, Cambridge University]

Inflammation is implicated in many neurodegenerative disorders, but the underlying cause of

inflammation, as well as its role in the disease process, remains unclear. Here we propose a role for aberrant activation of TEs as a cause of neuroinflammation in PD. Expression of some TEs has previously been seen to alert innate immune complexes in certain autoimmune diseases, suggesting a mechanism for activating inflammatory pathways. In the adult brain, TEs are normally transcriptionally silent but can become aberrantly activated in tissues from patients with neurodegenerative disease. Animal models expressing TE-derived proteins implicate these as a cause of neuroinflammation and neurotoxicity. However, there are conflicting studies on this topic, and more work is needed to clarify the connection between TEs, neurodegeneration, and PD.

In this project, we hypothesize that neurodegenerative disease pathology as well as general aging-associated mechanisms contribute to transcriptional activation of TEs—via chromatin relaxation and loss of transposon control factors. General elevation of TEs has previously been seen in the aging brain but can be accelerated by processes known to contribute to neurodegeneration. Aberrant expression of TEs results in the formation of double-stranded RNAs, reverse-transcribed cDNA molecules, and TE-derived peptides that induce a "viral mimicry," in which cells of the central nervous system respond as though infected. Although this hypothesis is attractive, more detailed investigations of this phenomenon are needed. In particular, interactions between adjacent cells are likely key to understanding how each cell type contributes to both cell-intrinsic and cell-extrinsic mechanisms of triggering inflammation. This project seeks to use single-cell sequencing of the substantia nigra, prefrontal cortex, and amygdala from PD patients and healthy controls in order to determine whether transposable elements are elevated in PD and could serve as triggers for inflammatory cascades in PD. Preliminary results indicate that certain TEs are specifically elevated in PD patient tissues and colocalize with markers of pro-inflammatory astrocytes and microglia. Ongoing work will determine whether TEs are driving these inflammatory cascades using induced pluripotent stem cell (iPSC) models to drive TE expression in either induced dopaminergic neurons (iDA) or induced glial cells (astrocytes and microglia).

In Press

Phatnani H, Kwan H, Sateen D, Broach JR, Simmons Z, Arcila-Londono X, Lee EB, Van Deerlin VM, Shneider NA, Fraenkel E, et al. 2022. An integrated multi-omic analysis of iPSC-derived motor neurons from C9orf72 ALS patients. *iScience* **24:** 103221. doi:10.1016/j.isci.2021.103221

GENETIC VARIANTS LINKED TO AUTISM TRAITS

I. Iossifov Y-h. Lee A. Munoz C. Yoon
S. Marks B. Yamrom

In 2020, the Iossifov laboratory included the following members: Adriana Munoz, Boris Yamrom, Steven Marks, Yoon-ha Lee, and Chris Yoon. The bulk of our work was in analyzing the large data set of whole-genome sequencing (WGS) data generated from approximately 2,400 of the Simons Simplex Collection (SSC) families and approximately 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 approximately 20,000 of the SPARK families have been released and SPARK is expected to grow to approximately 50,000 families in a couple of years. In addition, we initiated a research program to explore the potential of RNA sequencing in family collections like SSC. These data are a rich resource that we utilize in numerous projects.

Below are the abstracts of four projects we either finalized in 2021 or are still actively working on. These projects demonstrate our current efforts in studying the role of de novo noncoding variants, rare structural re-arrangements, and common variants in autism's etiology.

De Novo Disruption of Introns Contributes to Autism

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 simplex families than in their unaffected siblings. But the actual rate in low-risk families cannot be determined solely from simplex families, which are a mixture of low and high risk. The rate of de novo mutation in nearly pure populations of high-risk families, the multiplex families, has not previously been rigorously determined. Moreover, de novo mutation rates have been underestimated from studies based on low-resolution microarrays and WES.

In September of 2021, we published a paper that addressed these limitations by analyzing the WGS of both simplex families from the SSC and multiplex families from the AGRE. After removing the multiplex samples with excessive genomic drift, we find that the contribution of de novo mutation in multiplex autism is significantly smaller than the contribution in simplex. We use WGS to provide high-resolution copy number variant (CNV) profiles, analyze more than coding regions, and revise the rate upward in simplex due to 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 (Yoon et al. 2021).

A Platform for Access and Analysis of Genetic Variants in Phenotype-Rich Family Collections

We witnessed impressive success in autism genetics in recent years by analyzing sequence data sets generated from deeply phenotyped family collections like SSC. An enormous amount of work needs to follow the early success to develop effective treatment and early diagnostic strategies. A variety of future research projects will detail the effects of hundreds of genetic variants and genes at molecular, cellular, and organismic levels. Such projects will benefit from the accumulated data sets, but their large and complex structures create a significant obstacle to their efficient use.

We addressed such difficulties by building the Genotype and Phenotype in Families (GPF) system that enabled intuitive and straightforward interaction with such data sets. The system allowed our group to successfully integrate and analyze diverse phenotypic and genotypic data and efficiently distribute the valuable resources to the broader scientific community through the GPF's intuitive web interface. We developed the GPF as an open-source project (https://github.com/iossifovlab/gpf) so that different

groups could use it to operate on their data. We also deployed the GPF system at the Simons Foundation (https://gpf.sfari.org), where it manages several large genotypic and phenotypic data sets built through the Simons Foundation's support. These data sets include the phenotypic data gathered from SSC, VIP, SPARK, and AGRE and the genotypic data from the same collections generated through whole-exome, whole-genome sequencing, and high-density chip hybridization. To date, the system handles genotypes derived from approximately 10,000 individuals with whole-genome and approximately 60,000 individuals with whole-exome data. The system's distributed architecture allows us to plan deployments that manage genotypes for one million whole-genome samples.

RNA-seq of SSC

Our analysis of the whole-genome data from the approximately 2,000 quad families from the SSC demonstrated convincingly that de novo variants in the noncoding regions contribute to the incidence of autism (Yoon et al. 2021). Specifically, we observed a significantly increased rate of de novo intronic indels in affected children relative to their unaffected siblings when we restricted the rate observation to the autism genes previously implicated by WES. The increase in the rate is consistent with de novo intronic indels contributing to ~5% of the autism diagnosis.

We do not observe a similar increase in the rate of de novo intronic substitutions. Still, the study's size is insufficient to detect that signal, given the much higher background noise rate for substitutions. Nevertheless, we expect that de novo intronic substitutions have a contribution of similar magnitude to the de novo intronic indels' contribution. Others have reported an increased de novo mutation rate in affected versus unaffected children within the intergenic space's control regions. In total, we expect that the noncoding de novo mutations' contribution is close to 13%, perhaps only slightly less than the contribution from de novo coding mutation.

Despite the noncoding variants' significant contribution, we have no suitable purely analytic method to distinguish the specific causal sequence variation from the many random ones. We proposed to address that through a study of the RNA. We expect that for most of the causal de novo noncoding variants, the immediate effect would be on the expression of nearby genes. We can detect such changes in expression through RNA-seq data by comparing the expression of the affected gene allele to that of the unaffected allele, a method called allele-specific expression (ASE), or by identifying abnormal splicing patterns. We performed a pilot project to test this approach's feasibility using deep RNA-seq profiles of the lymphoblastoid cell lines (LCLs) from 202 individuals from 48 of the SSC families. In close collaboration with Kristin Baldwin (Genetics & Development at Columbia University) and Michael Wigler (CSHL), we also demonstrated that we could transform LCLs into induced pluripotent stem cells (iPSCs) and the iPSCs further into induced neurons (iNs). The iNs express nearly 90% of the autism genes identified by exome sequencing, whereas the LCLs express ~70% of these genes. Moreover, when both cell types express a gene, the allele-specific expression measures are usually similar.

The pilots' results were encouraging enough to help us convince the Simons Foundation to fund a large-scale project for generating 4,000 RNA profiles from the LCLs from all the SSC children. We have generated approximately 2,500 of these profiles to date, and we expect to complete the data generation by August 2021. The project will produce 100–200 de novo noncoding variants associated with the nearby genes' perturbed transcription. We are currently trying to fund a validation effort to use CRISPR-Cas9 and cell transformation and determine whether these candidate variants are responsible for the observed transcriptional abnormality and whether the abnormal transcription is preserved in the relevant iN cells.

Genome Sharing in Siblings Concordant or Discordant for Autism

We developed a method to measure the extent to which siblings, concordant and discordant for autism, share their parental genomes using genotypes derived from whole-genome data or chips. We applied the method to approximately 1,300 pairs of concordant and approximately 4,500 pairs of discordant siblings from the SSC, AGRE, and SPARK collections. Surprisingly, we observed that the fathers have an increased sharing consistent with them carrying causal determinants in at least half of the multiplex families, more than the families in which the mother has a causal variant. With lesser significance, the discordant

siblings from the simplex families corroborate the observation: the antisharing tended to be stronger for the paternal genome. The more extensive sharing of paternal than maternal genomes contradicted our expectations that mothers will be the primary source of damaging variants in the high-risk multiplex families and forced us to rethink our unified hypothesis of the genetic contribution to autism.

This work is a collaboration with Michael Wigler, Kenny Ye, and Dan Levy from CSHL and Abba Kreiger and Andreas Buja from the Department of Statistics at Wharton, UPenn. Our first manuscript on the topic is now under review in *Genome Research*. The manuscript describes the method, its first application, the surprising observation, and our hypotheses for what could cause it. One possible explanation is that the father may carry strong protective alleles; another is that maternal–fetal immune incompatibility may cause autism.

Exploring Maternal–Fetal Incompatibility

Our observation of paternal transmission in multiplex families suggests that maternal–fetal incompatibility caused by paternal alleles' expression in the fetus may contribute to autism. This immunological hypothesis is compatible with unexplained features of autism:

- Sibling risk diminishes over time.
- Increased risk in prematurely born offspring.
- Lack of multigenerational pedigrees.
- Higher prevalence in men.

Others have reported that maternal–fetal immune incompatibility contributes to the risk of neurological conditions, such as neuropsychiatric diseases and learning disabilities. In principle, the corresponding immune reaction could be monitored and, importantly, mitigated by intervention.

Last year, we initiated the exploration of this hypothesis following two separate paths. In an effort led by Tobias Janowitz (CSHL), we are attempting to develop mouse models of maternal–fetal incompatibility. We will use a combination of inbreeding and interbreeding of inbred mouse strains and immune modulation in the mothers, hoping to demonstrate that such incompatibility can lead to abnormal neurobiological development and function. Second, we apply computation approaches to identify traces of maternal–fetal incompatibility using the deep whole-genome data from the SSC. We use the mothers' human leukocyte antigen (HLA) types and missense variants carried by the fathers, but not by their spouses, together with HLA-peptide binding prediction tools to identify likely mother-immunogenic variants carried by the fathers and transmitted to autistic children or not transmitted at all. The computational approach is a collaboration between our laboratory and those of Alex Krasnitz, Hannah Meyer, and Tobias Janowitz.

PUBLICATIONS

Yoon S, Munoz A, Yamrom B, Lee Y-h, Andrews P, Marks S, Wang Z, Reeves C, Winterkorn L, Krieger AM, et al. 2021. Rates of contributory de novo mutation in high- and low-risk autism families. *Commun Biol* **4:** 1026. doi:10.1038/s42003-021-02533-z

MASSIVELY PARALLEL REPORTER ASSAYS, MACHINE LEARNING, AND THE BIOPHYSICS OF GENE REGULATION

J.B. Kinney A. Ayaz A. Tareen
W-C. Chen M.S. Wong
A. Posfai

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 relatively simple experiments. At first, he pursued dry laboratory research focused on developing 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 what has since become known as the massively parallel reporter assay (MPRA). He also showed that, when combined with techniques from machine learning, data from MPRAs can illuminate the mechanisms of gene regulation in ways that no other technology can. As an independent investigator, he continues to pursue a tightly knit combination of experiment, computation, and theory focused on using MPRAs to decipher gene regulatory mechanisms through the measurement and quantitative modeling of sequence–function relationships.

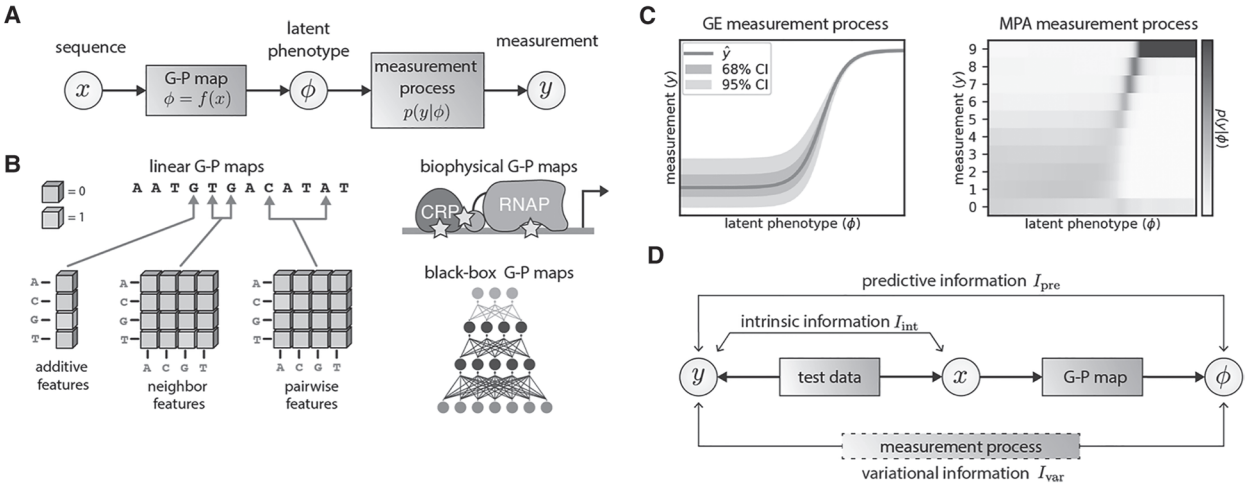


Figure 1. The MAVE-NN modeling strategy proposed in Tareen et al. (2022). (*A*) Structure of latent phenotype models. A deterministic genotype–phenotype (G-P) map assigns a latent phenotype ϕ to each input sequence x. This latent phenotype is then mapped to a distribution across possible measurement values y by a stochastic measurement process $p(y|\phi)$. (*B*) MAVE-NN supports a variety of G-P maps, including linear models based on fixed sequence features (additive, neighbor, or pairwise), biophysical models, and black-box neural network models. (*C*) MAVE-NN also supports two types of measurement processes. Global epistasis (GE) measurement processes model continuous measurements as noisy readouts of some nonlinear function $\hat{y}(\phi)$. Measurement process–agnostic (MPA) measurement processes can represent arbitrary conditional distributions $p(y|\phi)$ when y values are discrete. (*D*) MAVE-NN quantifies model performance using an information-theoretic framework. Intrinsic information, I_{int}, is the mutual information between measurements and assayed sequences. Predictive information, I_{pre}, is the mutual information between measurements and predicted latent phenotype values. Variational information, I_{var}, is a linear transformation of log likelihood that provides a variational lower bound on I_{pre}. Model performance is judged in light of the model performance inequality $I_{int} \geq I_{pre} \geq I_{var}$. Here, $I_{pre} = I_{int}$ obtains only when the G-P map is correct, and $I_{var} = I_{pre}$ obtains only when the measurement process is correct (assuming the G-P map is correct).

MAVE-NN: Learning Genotype–Phenotype Maps from Multiplex Assays of Variant Effect

The Kinney laboratory has developed a unified conceptual and computational framework for quantitatively modeling data from multiplex assays of variant effect (MAVEs). MAVEs, a class of technologies that the Kinney laboratory has helped to pioneer, encompass a broad range of experimental techniques that include MPRAs for studying transcriptional regulation or alternative mRNA splicing, deep mutational scanning (DMS) experiments on proteins, and more. Our approach is implemented in an easy-to-use Python package called MAVE-NN (https://mavenn.readthedocs.io). MAVE-NN represents MAVE experiments using latent phenotype models. These models consist of a genotype–phenotype (G-P) map, which maps each input sequence to an unobserved latent phenotype of biological interest, and a measurement process that maps this latent phenotype to an experimental measurement. MAVE-NN also implements an innovative information-theoretic framework for quantifying model performance and completeness, one that is applicable to all MAVE data sets. This strategy, illustrated in Figure 1, is motivated by the Kinney laboratory's previous theoretical work concerning an important but underappreciated connection between mutual information and likelihood.

PUBLICATIONS

Chen W-C, Zhou J, Sheltzer JM, Kinney JB, McCandlish DM. 2021. Field-theoretic density estimation for biological sequence space with applications to 5′ splice site diversity and aneuploidy in cancer. *Proc Natl Acad Sci* **118:** e2025782118.

Skalenko KS, Li L, Zhang Y, Vvedenskaya IO, Winkelman JT, Cope A, Taylor DM, Shah P, Ebright RH, Kinney JB, Zhang Y, Nickels BE. 2021. Promoter-sequence determinants and structural basis of primer-dependent transcription initiation in *Escherichia coli*. *Proc Natl Acad Sci* **118:** e2106388118.

In Press

Pukhrambam C, Molodtsov V, Kooshkbaghi M, Tareen A, Vu H, Skalenko KS, Su M, Zhou Y, Winkelman JT, Kinney JB, Ebright RH, Nickels BE. 2022. Structural and mechanistic basis of σ-dependent transcriptional pausing. *Proc Natl Acad Sci* **119:** e2201301119.

Tareen A, Kooshkbaghi M, Posfai A, Ireland WT, McCandlish DM, Kinney JB. 2022. MAVE-NN: learning genotype–phenotype maps from multiplex assays of variant effect. *Genome Biol* **23:** 98.

TOWARD TRUSTWORTHY AND INTERPRETABLE DEEP NEURAL NETWORKS FOR REGULATORY GENOMICS

P.K. Koo R. Ghotra N. Keone Lee Z.A. Tang
J. Kaczmarzyk A. Majdandzic S. Toneyan
E. Labelson C. Rajesh R. Tripathy

Our research studies the functional impact of genomic mutations through a computational lens using data-driven machine-learning solutions. We are broadly interested in applications of studying gene regulation and protein (dys)function. Our approach develops methods to interpret high-performing deep-learning models to distill knowledge that they learn from big, noisy biological data. Our goal is to elucidate biological mechanisms that underlie sequence–function relationships, with a broader aim of advancing precision medicine for complex diseases, including cancer.

Understanding What Drives Neural Network Predictions

A. Majdandzic, J. Kaczmarzyk, P.K. Koo

Although deep neural networks (DNNs) are becoming widely applied in regulatory genomics, it remains unclear why they make a given prediction. Advances in explainable artificial intelligence (AI) now make it possible to identify individual positions in each sequence that are important for model predictions via attribution methods. In practice, the resultant attribution scores are noisy, requiring humans to deduce what (if any) patterns emerge. Thus, although attribution methods are helpful to generate hypotheses, they do not inform what underlying biology drives model predictions.

The gold standard for measuring causal effects is the randomized controlled trial (RCT). In an RCT, individuals are randomly assigned to a treatment or control group. The individuals in the treatment group are given an intervention (i.e., a drug), whereas the control group receives a placebo. This approach ensures that the causal effect size of the intervention is measured and the effect of confounders in any individual are averaged over. Leveraging this approach, we have developed a new interpretability method for biological sequence-based DNNs, called global importance analysis (GIA) (Koo et al. 2021). Similar to RCTs, GIA treats each biological sequence within a data set as an individual. An intervention is applied to the sequences in the treatment group by embedding a pattern of interest, whereas the control group remains untouched. The effect size of the treatment and control group is measured via the DNN's prediction. Hence, GIA measures the causal effect size that patterns have on model predictions across a population of sequences in a quantitative manner.

As a case study, we used GIA to interpret a DNN trained to predict RNA sequence specificities of RNA-binding proteins (RBPs). Using GIA, we found that in addition to sequence motifs, our DNN learns a model that considers the number of motifs, their spacing, and sequence context, such as RNA secondary structure and GC bias (Fig. 1).

Moving forward, GIA demonstrates that we can treat a trained DNN as an in silico laboratory, where we can perform in silico experiments to probe the sequence–function relationship learned by a DNN, using model predictions in lieu of experimental measurements. Alternative in silico experiments include knockout and knock-in CRISPR experiments or even in silico massively parallel reporter assay to quantitatively characterize motifs and motif interactions. Of course, any insight is through the lens of the DNN. Hence, GIA is fundamentally a model interpretability tool. Recently, we have extended GIA to computer vision to study the causal effects that different features in an image have on model prediction. We plan to further develop GIA to understand the features that drive patient survival predictions of DNNs trained on histology data.

Interpretable Deep Neural Networks for Functional Genomics

R. Ghotra, P.K. Koo

Identifying *cis*-regulatory elements (CREs) and their interactions with other CREs is critical for

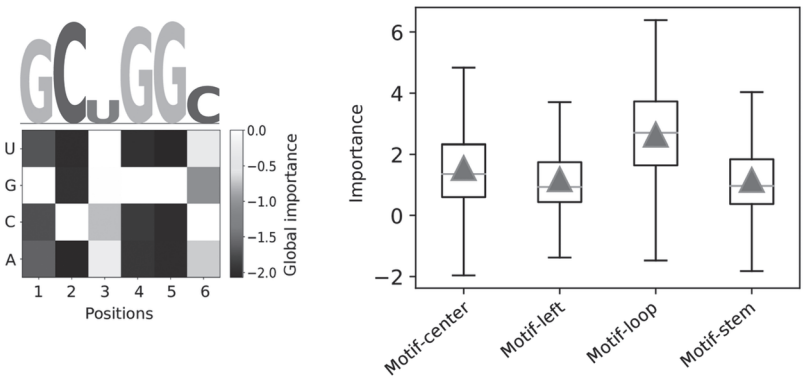


Figure 1. Analysis of ResidualBind trained on RNAcompete data for VTS1. (*Left*) Heatmap of the difference in the global importance analysis (GIA) scores for synthetic sequences embedded with single-nucleotide mutations of the canonical VTS1 motif from wild type with a sequence logo above. (*Right*) Box plot of importance for the VTS1 motif embedded in different regions of synthetic RNA sequences designed with a stem-loop structure and in the same positions in random RNA sequences. Green triangles represent the GIA scores.

understanding many important biological processes such as transcriptional regulation and alternative splicing. However, inferring CREs from genome-wide functional data is difficult because their activity in any given locus may depend on many other factors, including sequence context and/or the presence of other CREs nearby. Moreover, identifying the locations of individual CREs across the genome remains challenging because of the natural variability of CRE sequences. The recent success of genomic sequence analysis with convolutional neural networks (CNNs) across a wide range of functional genomic prediction tasks suggests that the CNNs are learning key DNA features that are highly correlated with function. However, CNNs are difficult to interpret, making it challenging to identify the factors that drive these performance gains.

To address their low interpretability, we demonstrated a new design principle that encourages the first layer parameters of a CNN to learn more interpretable representations of sequence motifs, while not sacrificing accuracy (Koo and Ploenzke 2021). One such example is the use of the exponential activation function in the first convolutional layer, instead of the common activation function called a rectified linear unit (ReLU). This new class of CNNs is interpretable by design and can reveal human-interpretable motif patterns that it has learned simply by visualizing the parameters of the model. Recently, we benchmarked various CNNs trained using sophisticated regularization methods and showed that exponential activations consistently improve the model's generalization performance as well as the efficacy of attribution maps (Fig. 2) (Labelson et al. 2021).

With our ongoing work, we continue to search for new design principles and training procedures to improve the interpretability of CNNs in the hope that they may provide new insights into the underlying biology of DNA–protein interactions and gene regulation, more broadly. Recently, we have begun exploring design principles in deeper layers that could identify motif interactions (Ghotra et al. 2021). Our initial analysis shows that there is promise with transformers, a relatively new type of neural network layer that consists of a self-attention mechanism. Self-attention introduces an intrinsically interpretable module that specifically performs pairwise interaction analysis and thus serves to capture interactions (if any) between motifs learned in the first layer. In practice, we found that the interactions identified by attention were quite noisy on an individual sequence basis (i.e., local attention). Hence, we developed a new approach that aggregates statistics to identify global trends of motif interactions, which we call global interactions of filter activity correlations (GLIFAC). We found that GLIFAC significantly improves our ability to identify motif interactions learned by attention-based models (Fig. 3). We plan to further develop design principles that ensure the attention maps do not require postprocessing but still can provide direct insights into motif interactions.

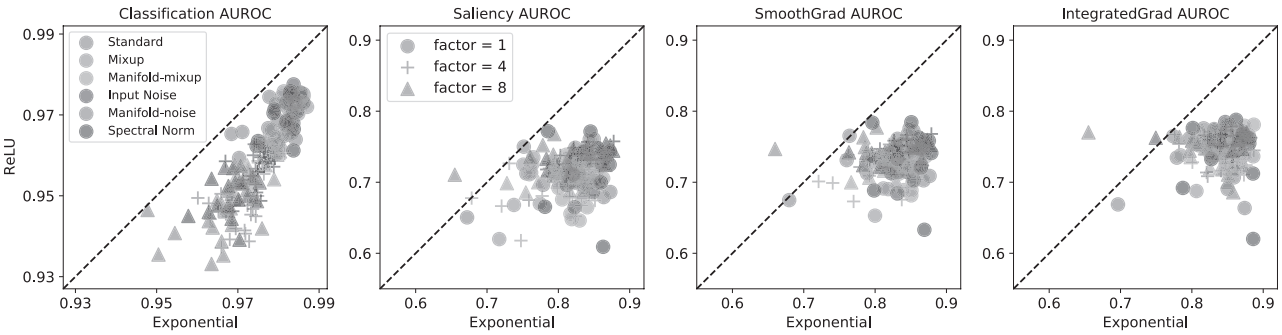


Figure 2. Performance comparison of models with ReLU activations versus exponential activations. Scatter plot of various performance metrics for convolutional neural networks (CNNs) trained with various regularization methods (shown in a different color) and with different model sizes (shown in a different marker) for ReLU activations versus exponential activations. Each dot represents the average of 10 trials with different random initializations for a given model training configuration.

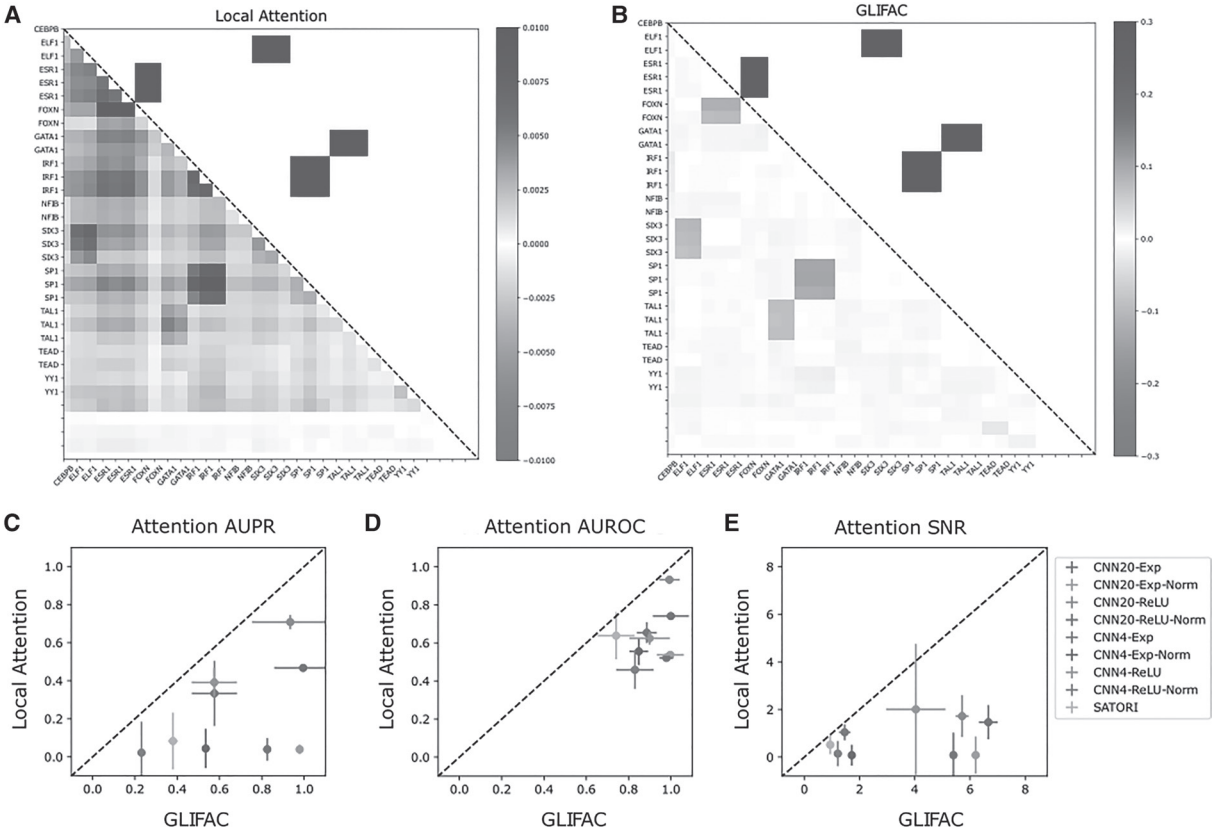


Figure 3. Performance comparison of extracting motif interactions. Example interaction maps generated by a convolutional-attention network using (*A*) aggregated local attention and (*B*) global interactions of filter activity correlations (GLIFAC). The upper triangle of each interaction map represents ground truth interactions. Performance comparison of aggregated local attention and GLIFAC-generated attention maps to capture ground truth motif interactions according to three metrics: (*C*) average precision recall (AUPR), (*D*) average under the receiver operating characteristic (ROC) curve (AUROC), and (*E*) signal-to-noise ratio (SNR).

Prioritizing Noncoding Variants with Neural Networks

A. Majdandzic, R. Tripathy, E. Labelson, C. Rajesh, P.K. Koo

One promising application of DNNs that are trained on genome-wide functional genomics data is to repurpose them to predict the functional effect of noncoding variants. Indeed, DNNs have provided mechanistic insights into various human diseases, including autism spectrum disorder, congenital heart disease, and autoimmune diseases. Using a DNN for this purpose requires trust in its predictions, especially on sequences with mutations, which the model was never trained on. This so-called in silico mutagenesis is one of many interpretability methods that falls under the umbrella of attribution methods. Attribution methods have been very insightful to reveal motif-like representations with base resolution along a given sequence. They have helped to elucidate key motifs enriched in cell type–specific enhancers in mouse cortex from the analysis of single-cell assay for transposase-accessible chromatin using sequencing (ATAC-seq) data (Kawaguchi et al. 2021).

In practice, attribution maps are noisy. Noise in attribution maps can lead to random predictions of variant effects. The origins of all noise sources in attribution maps are not well understood. Recently, we discovered a new source of noise in attribution maps that was not previously known (Majdandzic and Koo 2021). The origins of this noise, which we have termed off-manifold noise, arise because of the nature of how DNNs model biological sequence data. By deriving a correction for this noise, we showed that the efficacy of the attribution maps could be greatly improved.

The standard pipeline for scientific discovery with DNNs is to train them on training data, optimize hyperparameters of the network on validation data to identify the best-performing models, and evaluate the generalization capabilities on held-out test data. Thus, the model that will be utilized for downstream interpretability analysis is based on best performance on the validation set. Surprisingly, we found DNNs that yield better performance on held-out test data do not necessarily lead to more trustworthy predictions of mutations (Fig. 4) (Labelson et al. 2021). This work raises the red flag that we should not simply evaluate DNNs solely on model prediction performance on held-out test data, but rather should probe the network to validate that it has robustly learned biologically meaningful representations. Our ongoing research is focused on investigating ways to optimally select which DNN should be employed for post hoc model interpretability with attribution methods, thereby providing more reliable predictions of variant effects.

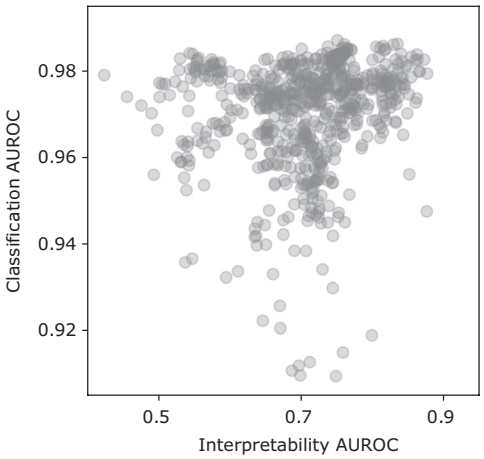


Figure 4. Lack of correlation between classification and interpretability performance. Scatter plot of classification performance versus interpretability performance for different models (each dot). Best-performing models (at the *top*) can either have a high or low interpretability. Thus, classification performance is not a good metric to select models for downstream model interpretability.

Improving Methods to Evaluate Neural Networks

S. Toneyan, Z.A. Tang, P.K. Koo

Over the past several years, the variety of DNNs being proposed to address regulatory genomic prediction tasks has increased substantially. The wide variations of proposed models, the data sets they are trained on, how the data sets are processed, and the tricks used to train the models make it challenging to assess which innovations are driving performance gains. A direct comparison cannot be made facilely because of the variation in how the prediction tasks are framed. As the number of applications continues to grow each year, a bottleneck of claims of modeling innovations is forming as we lack the ability to perform a critical assessment of newly proposed models with an evaluation that is appropriate and fair. Toward this, we have begun to establish a unified evaluation framework that enables direct comparisons of different models

irrespective of how the prediction task is framed or across variations in data processing choices. Our framework enables a direct comparison of not only the model's predictions, but also the robustness of the predictions and model interpretability. By enabling a fair comparison, our evaluation framework makes it possible to identify which modeling innovations directly lead to better overall performance.

PUBLICATIONS

Ghotra R, Lee N, Koo PK. 2021. Uncovering motif interactions from convolutional-attention networks for genomics. *NeurIPS 2021 AI for Science Workshop*.

Kawaguchi RK, Tang Z, Fischer S, Tripathy R, Koo PK, Gillis J. 2021. Exploiting marker genes for robust classification and characterization of single-cell chromatin accessibility. bioRxiv doi:10.1101/2021.04.01.438068

Koo PK, Ploenzke M. 2021. Improving representations of genomic sequence motifs in convolutional networks with exponential activations. *Nat Mach Intell* **3:** 258–266. doi:10.1038/s42256-020-00291-x

Koo PK, Majdandzic A, Ploenzke M, Anand P, Paul S. 2021. Global importance analysis: a method to quantify importance of genomic features in deep neural networks. *PLoS Comput Biol* doi:10.1371/journal.pcbi.1008925

Labelson EL, Tripathy R, Koo PK. 2021. Towards trustworthy explanations with gradient-based attribution methods. *NeurIPS 2021 AI for Science Workshop*.

Majdandzic A, Koo PK. 2021. Statistical correction of input gradients for black box models trained with categorical input features. *2021 ICML Workshop on Computational Biology*.

LEARNING FROM CANCER-DERIVED MOLECULAR DATA

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 (a) the potential impact of our work on clinical research and practice at present and in the foreseeable future, (b) the need to maximize the utility of emerging molecular technologies and research platforms in cancer biology, and (c) 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 (a) examination of intratumor genomic heterogeneity, its origin in cancer evolution, and its predictive value for aggressive and invasive potential of cancer; (b) reducing the complexity of genomic data for better interpretability while retaining their biological content; (c) derivation of clinically relevant molecular subtypes of the disease; and (d) design of predictive models for response to pharmacological interventions. As pursuit of these goals often reveals the inadequacy of the existing, and necessitates the development of novel computational tools, toolmaking is an important component of our activity. This report is a representative snapshot, rather than an exhaustive compendium, of our research, illustrating our activity in 2021.

Accurate and Robust Inference of Genetic Ancestry from Cancer-Derived Molecular Data

There is ample epidemiological evidence that race and/or ethnicity are important determinants of incidence, clinical course, and outcome in multiple types of cancer. Thus, these categories must be taken into account in the analysis of molecular data derived from cancer. A number of recently published large-scale genomic studies of cancer point to differences in the molecular makeup of the disease among groups of different ancestral background and to the need for more molecular data to power discovery of such differences.

Ancestry annotation of cancer-derived data largely draws on two sources. One is a patient's self-identified race and/or ethnicity (SIRE). SIRE is often missing, sometimes inaccurate, and usually incomplete. Furthermore, SIRE is not always consistent with genetic ancestry. Finally, a self-declaring patient is often given a choice from a small number of broad racial or ethnic categories that fail to capture complete ancestral information, especially in cases of mixed ancestry.

A far more accurate and detailed ancestral characterization may be obtained by genotyping a patient's DNA from a cancer-free tissue. Powerful methods exist for ancestry inference from germline DNA. However, genotyping of DNA from patient-matched cancer-free specimens is not part of standard clinical practice. As a result, it is not performed routinely outside academic clinical centers or major research projects. There also are studies yielding sequence data from tumors, whose purpose does not require germline profiling. RNA sequencing (RNA-seq) for expression quantification is in this category. Finally, peripheral blood is most often the source of germline DNA in the clinic, but this is not always the case for diseases of the hematopoietic system, such as leukemia, wherein cancer cells are massively present in circulation. In summary, matched germline DNA sequence is not universally available for cancer-derived molecular data. In such cases, it is necessary to infer ancestry from the nucleic acid sequence of the tumor itself.

Standard methods of ancestry inference commonly rely on population specificity of germline single-nucleotide variants (SNVs). Whole-genome sequencing (WGS) or whole-exome sequencing (WES), at depths sufficient for reliably calling SNVs, and readouts from genotyping microarrays, are therefore data types most suitable for this purpose. However, such detailed DNA profiling is often not performed in molecular studies of cancer. In such cases, it is necessary to infer ancestry from other types of tumor-derived data, including RNA sequence and DNA sequence for a small panel of genes (e.g., FoundationOne CDx®).

For all types of tumor-derived sequence, accurate inference of ancestry is a potential challenge. Tumor genome is often replete with somatic alterations, including loss of heterozygosity (LOH), copy number variants (CNVs), translocations, microsatellite instabilities, and SNVs. Of these, structural variants, especially LOH and CNVs, are the most likely to affect the genetic ancestry calls, but other types of alterations also are, to various degrees, potential obstacles to accurate ancestry inference. Tumor-derived RNA-seq presents additional challenges—namely, extremely uneven coverage of the transcript due to a broad range of RNA expression levels and distortions due to allele-specific expression. Gene panels represent a very small fraction of the genome, whose sufficiency for ancestry inference is not clear and may vary from panel to panel. In addition, cancer gene panels are enriched in cancer driver genes, which tend to undergo somatic alteration more frequently than other parts of the genome. Reliable and accurate ancestry inference from tumor-derived nucleic acids thus represents an unmet need, which we set out to address.

For this purpose, we designed an inference procedure having in mind a scenario, likely to occur in studies of existing data or of archived tissue specimens, with an input molecular profile of a tumor from a single patient, and no matching cancer-free sequence available. The profile in question may have its unique set of sequence properties. These include the target sequence and uniformity of its coverage, depth, read length, and sequencing quality. These profile-specific properties may be vastly dissimilar from those in the available public data sets with reliably known genetic ancestry of the patients. Furthermore, not all ancestries are equally easy to infer. This profile specificity would make it impossible to confidently assess the accuracy of the inference procedure for the input profile from its performance with the public cancer-derived data in aggregate. To overcome this difficulty, we develop a computational technique, which is described schematically in Figure 1, wherein the ancestral background of the patient is supplanted in the input profile by one of an unrelated individual with known ancestry. We next apply established methods of ancestry inference to this synthetic profile and compare the result to that known ancestry. Generating multiple such synthetic profiles allows us to assess how accurate the ancestry inference is for the patient, both overall and as a function of the profile's

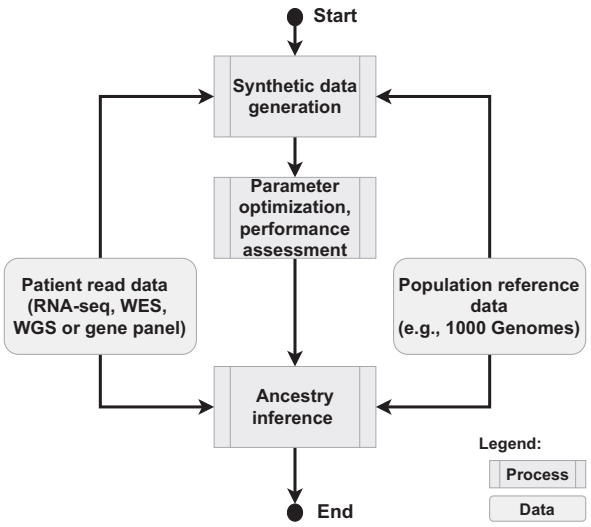


Figure 1. Overview of the procedure for inference of genetic ancestry.

continental-level ancestry. Furthermore, using synthetic data, we are able to optimize the inference procedure with respect to parameters on which it depends. Importantly, this assessment and optimization procedure does not require the profile in question to be part of a larger data set from a cohort of patients with a similar diagnosis. Very often in public cancer-derived data, such cohorts do not provide statistically meaningful representation of non-European ancestries. This insufficiency is not an impediment to the application of our methodology.

In 2021 we assessed the performance of genetic ancestry inference from three genomic data types: whole exomes, gene panels targeting exomes of several hundred cancer-related genes each, and RNA sequences. Throughout the study, we used the 1000 Genomes (1KG) data set as reference, against which patient molecular data were compared to infer continental-level global ancestry. We thereby classified each patient into one of five superpopulations of origin: African (AFR), East Asian (EAS), European (EUR), Mixed American (AMR), and South Asian (SAS).

We assessed the accuracy of global ancestry calls from tumor-derived data, in comparison to such calls from matching germline genotypes, as profiled by exome sequencing or single-nucleotide polymorphism (SNP) microarrays. We did so for three cancer types—namely, pancreatic ductal adenocarcinoma (PDAC) and ovarian cystadenocarcinoma (OV) as representative types of epithelial tumors and acute

Table 1. Comparison of genetic ancestry calls in The Cancer Genome Atlas ovarian cystadenocarcinoma (OV) patient cohort from cancer-derived data obtained using the three molecular platforms as indicated (columns) to those from the matching cancer-free genomes of the patients (rows)

	Whole exome					FoundationOne CDx®					RNA-seq				
	EAS	EUR	AFR	AMR	SAS	EAS	EUR	AFR	AMR	SAS	EAS	EUR	AFR	AMR	SAS
EAS	10	0	0	0	0	10	0	0	0	0	4	0	0	0	0
EUR	0	378	0	0	0	0	376	0	2	0	0	242	0	0	0
AFR	0	0	29	0	0	0	0	28	1	0	0	0	21	0	0
AMR	0	1	0	16	0	0	4	0	13	0	1	1	0	9	0
SAS	0	0	0	0	7	0	0	0	0	7	0	0	0	0	4

As an example, of the 11 patients identified as mixed Americans (AMR) using cancer-free genomes, nine were identified as AMR, while one each were identified as East Asian (EAS) and European (EUR) using cancer-derived RNA-seq. FoundationOne CDx® sequence data were simulated from the whole-exome sequences of each patient.

myeloid leukemia (AML) as an example of hematopoietic malignancy. Each of these data sets represents a unique challenge for patients' ancestry inference. OV is characterized by massive copy number alterations, often spanning much of the genome. Our PDAC data originate from patient-derived organoid (PDO) models of the disease. In PDO, near-100% tumor purity is achieved, exacerbating effects of copy number loss and loss of heterozygosity on the sequence. In AML the peripheral blood, the usual source of cancer-free DNA, may be severely contaminated by the cancer.

Our results demonstrate consistently high accuracy of global, continental-level ancestry inference from cancer-derived data for all possible combinations of the three cancer types, the three molecular profiling platforms, and the five super-populations examined in our study. For illustration, Table 1 is a summary of our ancestry calls for The Cancer Genome Atlas (TCGA) OV patient cohort, in comparison to the patient ancestry inferred from the matching cancer-free genomes. The table clearly demonstrates close agreement between the two sets of calls. Importantly, our analytic tools permit us to go beyond this cohort-level analysis of performance and to assess the ancestry call accuracy for each individual patient. We observed consistently high performance of our inference procedure across patient ancestral backgrounds, profiling platforms, sequence depths, and cancer types (data not shown).

With this work, we introduce a systematic approach to ancestry inference from cancer-derived molecular data. We anticipate this approach to have a major, twofold, impact on investigation of links between ancestry and cancer. First, it will become possible to massively boost the statistical power of such studies by leveraging existing tumor-derived molecular data sets without matching germline sequences or ancestry annotation.

Our search of the Gene Expression Omnibus (GEO) database alone has identified more than 1,250 such data sets, containing RNA expression data for nearly 48,000 cancer tissue specimens. Such resources dwarf those of fully annotated repositories, such as TCGA and International Cancer Genome Consortium (ICGC). Other molecular data repositories are likely to contain resources of this category on a similar order of magnitude. Second, hundreds of thousands of tumor tissue specimens stored at multiple clinical centers constitute another major resource for ancestry-aware molecular studies of cancer. Here again, matching normal tissue specimens are often absent and so is ethnic or racial annotation for the patients. According to a recent estimate, such annotation is missing in electronic health records of >50% of patients. Inferential tools we have designed will make these massive resources of archival tissues available for ancestry-oriented cancer research.

Multiple directions of exploratory and correlative analysis are open to pursuit with the accurate ancestry annotation made possible by the methods we have developed, even in the absence of matching cancer-free molecular data. Single-nucleotide and other small-scale somatic alterations may be identified in cancer-only exomes, both whole and restricted to specialized gene panels, using methods developed for this purpose alongside databases of frequent somatic variants in cancer and of frequent germline variants like gnomAD and 1KG. Copy number variants and losses of heterozygosity in cancer exomes are overwhelmingly somatic and may be determined computationally. Cancer RNA expression quantification is feasible in the absence of the germline genotype of the patient, including allele- and isoform-specific analysis. These

and similar genomic and transcriptional properties may be explored for associations with ancestral background of the patients.

Copy Number Analysis of Cancer Genomes

Somatic DNA copy number variation (SCNV) is a ubiquitous phenomenon in cancer, with important phenotypic and clinical consequences. As such, it has been and remains a major research focus of our group. In 2021, we aimed at developing a novel method for segmentation of DNA copy number profiles derived from low-coverage DNA sequencing data. In recent years low-coverage sequencing has largely displaced DNA microarrays as a platform for copy number analysis.

For this data type, the standard computational workflow has been based on partitioning the genome into intervals. Genomic extent for each interval was chosen to ensure approximately equal sequence read count per interval for DNA from normal diploid cells. Following data normalization, copy number profiles are derived from the resulting noisy signal using segmentation tools developed for this purpose in the microarray era. These profiles are piecewise-constant functions of the genomic position, approximating the noisy signal. Genomic resolution of such a procedure is no better than the user-chosen bin size.

We sought to develop an alternative segmentation method, better suited to low-coverage sequencing data, wherein the resolution is dictated by the data themselves. Our initial focus was on low-coverage sequences originating from DNA library protocols that use digestion by restriction enzymes. The resulting sequence reads map to a well-defined set of locations in the genome. In combination with low-coverage sequencing, these protocols yield a readout equivalent to a set of Bernoulli random variables, each corresponding to a restriction site. For this data type, powerful methods of statistical analysis exist, dating back to the work of Nicolaas Kuiper in the late 1950s. For our purposes these methods had to be adapted to reflect the inhomogeneity of Bernoulli variables constituting the input data. With this adaptation, we developed a fast and accurate segmentation algorithm, called Kuiperinha in Kuiper's honor. Its principles are illustrated in Figure 2 for the case of human chromosome 1. From the input read data, we compute a function devised by Kuiper and determine the difference between its global maximum and minimum. This difference is then subjected to the Kuiper statistical test and, if found significant, the interval is divided into segments, with segment ends given by the locations of the maximum and the minimum. We follow up by computing the copy number in each of the

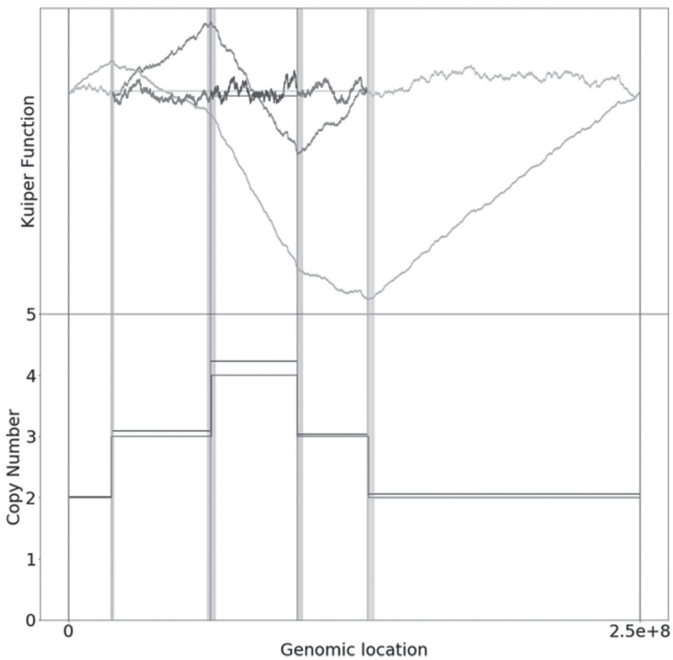


Figure 2. The Kuiperinha segmentation tool as applied to low-coverage sequence of human chromosome 1 (simulated data). The procedure starts with the whole chromosome and proceeds iteratively to each newly detected segment, until no new segments are detected. (*Top*) for each interval analyzed, a function defined by Kuiper is computed and used to conduct the Kuiper statistical test. If the test results in a significant finding, the interval is partitioned into segments, with the end points indicated by vertical bands with matching colors. The width of each band is given by the 95% confidence interval for the end-point location. (*Bottom*) The inferred copy number profile (black) compared to the ground truth (red). The genomic locations are given in base pair units.

segments and the uncertainties (confidence intervals) in the end-point positions. This process is then iterated within each newly discovered segment, until no new segments are found.

Although our method was originally designed for enzyme-digested DNA libraries, it can be readily generalized to detect copy number changes in all low-coverage genome sequences. The work toward this generalization is currently in progress. We expect the resulting computational tool to be applicable to both bulk- and single cell–derived DNA sequences.

PUBLICATION

In Press

Belleau P, Deschênes A, Tuveson DA, Krasnitz A. 2022. Accurate and robust inference of genetic ancestry from cancer-derived molecular data across genomic platforms. bioRxiv doi:10.1101/2022.02.01.478737

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 of <1 per 100 bases. These errors determine the lower limits of variant detection: Using standard sequencing, it is impossible to distinguish a variant frequency of <1% from machine error. We would like to measure variants at frequencies of 0.0001% or one part in a million.

Together with the Wigler laboratory, 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 sequencer errors. Taking a sequence consensus from all reads with the same tag corrects for sparse error. By tagging the initial molecule, the MASQ (multiplex accurate sensitive quantitation) protocol overcomes sequence error, enabling us to observe DNA variants occurring at very low frequencies. Importantly, this can be done on 50 to 100 different loci and millions of templates per reaction, obtaining extremely sensitive measurements from precious and limited clinical samples.

We have applied MASQ to measure residual disease in patients treated for acute myeloid leukemia (AML). Using custom, patient-specific probes, we were able to observe residual tumor load at levels as low as 1 part in 100,000. We are presently applying a similar pipeline to study a larger number of AML patients, tracking residual load in samples taken before, during, and after treatment, to explore the predictive value of measurement dynamics.

We are also exploring the application of MASQ in the context of solid tumors. Together with the Wigler and Krasnitz laboratories, we are partnering with physicians to measure tumor signal in the blood, where 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 1 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 employing enrichment strategies based on selection for epithelial cells, depletion of normal blood elements, or some combination of the two.

The second blood compartment that may harbor 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 1 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.

Preliminary results show robust signal from cell-free DNA in the plasma using our latest techniques. We are presently collecting and processing samples in sufficient numbers to generate relevant observations within and across tumor types.

Mutational Sequencing

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 to the present generation of short-read sequencers, these long-read platforms are expensive and error-prone.

We developed a method called muSeq (mutational sequencing) for obtaining long-read information from short reads by embedding a unique molecular identity throughout each template molecule by random mutation. We first implemented this idea in practice by 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. Unfortunately, even partial conversion with sodium bisulfite is harsh and limits the length of templates we can amplify to 5 kb. Recently, new enzymatic methods for cytosine deamination have become available. We developed a modified approach and can now obtain mutation-patterned templates longer than 10 kb.

The initial informatic methods for handling muSeq data required a known reference genome and used standard bisulfite mapping methods. However, long-read assembly is most useful when the reference genome is not available or is unreliable. Together with Siran Li of the Wigler laboratory, we developed a protocol and informatics for targeted and phased de novo muSeq assembly. The informatics pipeline first assembles mutated template molecules using graph-based methods augmented with a fast mapping algorithm. We then use a prior unmutated library to remove the mutation pattern wherever possible. Another subroutine divides the mutated templates into subgroups from a common haplotype, which are finally collapsed into individual sequence assemblies. These assemblies are of high quality with exceptionally low error rates.

Our latest protocols and informatics allow haplotype-phased assembly over genomic regions 10 kb or longer (Li et al., in press).

Accurate Measurement of Microsatellites

MASQ is excellent for accurately measuring a patient-specific set of tumor variants. However, each distinct variant requires a custom probe and almost all variants are unique to a specific patient. Therefore, reagents are custom-made at significant expense in time and money. That is the nature of single-nucleotide variants (SNVs): They occur, to a first order, sporadically throughout the genome.

In contrast to SNVs, we can instead consider microsatellite loci. A microsatellite is a short sequence of DNA repeated some number of times. For example, ACACACAC is an eight-base repeat consisting of four two-base units. These genomic positions are particularly prone to mutation, wherein the polymerase adds or deletes a single unit. The mutation rates at microsatellites are much higher than in the rest of the genome and, because they occur at well specified positions, we can develop a generic assay to track any patient and any tumor.

Unfortunately, the same copy-mediated mutability that makes microsatellites excellent markers for lineage also makes them difficult to measure. Any procedure that amplifies DNA using a polymerase will introduce the same type of errors, making the measurement of the true length difficult if not impossible.

Enter partial mutation. Just like with muSeq, we partially convert the sequence of the microsatellite *prior* to amplification. For microsatellites containing C bases, partial bisulfite mutation disrupts the repeat structure and therefore no longer slips during amplification. Provided the microsatellite is sufficiently disrupted, we can be confident of the sequence measure. This method allows the measurement of a single microsatellite read with an error rate near 1 in 10^3. Using multiple reads from the same template DNA enables error rates of <1 part in 10^6.

Together with members of the Wigler laboratory, we have demonstrated that disrupting microsatellites with partial bisulfite conversion works using synthetic DNA templates. We have since developed panels of capture reagents capable of enriching from approximately 1,200 microsatellite loci in the human genome by targeting the sequences flanking known microsatellite loci. Initial results show excellent enrichment and demonstrate that biologically derived DNA behaves as expected from our experiments with synthetic templates.

Our new method enables observing a form of genomic variation previously unmeasurable using standard sequencing methods. Its applications to cancer and human genetics are many, including tumor detection, measuring residual disease, determining microsatellite instability, genotyping populations, etc.

PUBLICATIONS

In Press

Li S, Park S, Ye C, Danyko C, Wroten M, Andrews P, Wigler M, Levy D. 2022. Targeted de novo phasing and long-range assembly by template mutagenesis. *Nucl Acids Res* **50:** e103. doi:10.1093/nar/gkac592

Wang Z, Moffitt AB, Andrews P, Wigler M, Levy D. 2022. Accurate measurement of microsatellite length by disrupting its tandem repeat structure. *Nucl Acids Res* doi:10.1093/nar/gkac723

PREDICTING EFFECTS OF MUTATIONS FROM HIGH-THROUGHPUT DATA

D. McCandlish B. Gitschlag A. Posfai
C. Martí-Gómez 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 somatic evolution in cancer and the evolution of drug resistance in human pathogens, 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 (McCandlish 2021).

This year we also started two new collaborations. The first is with the Lippman laboratory to study genetic interactions between mutations that influence fruit morphology in tomato. The second is with the Kinney and Krainer laboratories to investigate the mechanism of action for small-molecule drugs that modulate pre-mRNA splicing.

Genetic Interactions and High-Throughput Data

Modern high-throughput experimental techniques are for the first time providing large-scale measurements for not only how pairs of mutations interact, but higher-order interactions involving three to 20 mutations. However, how to model and understand these higher-order interactions remains an open question. Previously, we developed a method called empirical variance component regression that attempts to create a reconstruction of the full genotype–phenotype map that best matches the form and magnitude of genetic interactions present among the observed sequences. Although this method displays state-of-the-art performance on many high-throughput data sets and the underlying mathematical formalism provides insight into the predictability of mutational effects and the practical significance of higher-order interactions, it still has some important limitations. One such limitation is that the existing method only applies to haploid sequences, whereas—particularly for morphological and disease traits—it is important to be able to predict the effects of combinations of alleles at different sites in diploid organisms. This year postdoc Juannan Zhou was able to extend our empirical variance component regression technique to apply to diploid genotype–phenotype maps, including the effects of dominance and how it can be modified by changes at other loci, a phenomenon that is important in many applications. A second important limitation is the number of loci and alleles that our techniques can accommodate. Using graphics processing unit (GPU) acceleration, Juannan has also been working to scale this technique to genome-scale data. Many congratulations are also due to Juannan, who left this summer to start his own group at the University of Florida.

Postdoc Anna Posfai (jointly advised by Justin Kinney) has been continuing her work on "gauge freedoms" in models of the genotype–phenotype map. This work addresses the key issue that for typical models of the genotype–phenotype map there are many different sets of model parameters that produce identical predictions—a type of structural nonidentifiability referred to as "gauge freedoms" in the physics literature. It is essential to account for these gauge freedoms when interpreting model parameters or comparing different models, but it has previously not been clear how to do so. This year Anna worked on several aspects of this

problem. One area of progress was showing that gauge freedoms must arise in any model that obeys certain apparently reasonable criteria relating to model interpretability. Another area was in developing a theory of gauge freedoms for global epistasis models (in which one assumes that the observed phenotype is a nonlinear function of a latent additive trait and attempts to simultaneously infer the coefficients of the additive model and the form of the nonlinearity). Finally, she derived a natural one-dimensional family of methods for addressing gauge freedoms that differ in their tendency to ascribe phenotypic contributions to additive effects and lower-order genetic interactions.

Another key challenge in understanding genetic interactions is how to integrate and compare multiple sources of information on the form of genetic interactions, such as direct high-throughput measurements, observed patterns of genetic variation both within and across species, and the output of biophysical models. This year, postdoc Carlos Martí-Gómez joined the group and has been working to come to this type of integrative understanding for the Shine–Dalgarno sequence, a motif involved in translation initiation in many prokaryotic species that functions to recruit the ribosome to mRNA transcripts via its sequence complementarity with the 3′ end of the 18S rRNA. Importantly, combinatorially complete data sets in which translational activity is measured when the Shine–Dalgarno sequence is replaced by every possible RNA 9-mer are now available, which can be combined with genome-wide usage data from a variety of bacterial species as well as thermodynamic models of RNA binding. One key conclusion from his work so far is the importance of specific Shine–Dalgarno sequences that allow multiple physically distinct binding modes of the 18S rRNA. Specifically, these sequences not only appear to have a strong effect on translation initiation, but during long-term sequence evolution can also act as bridges that allow the Shine–Dalgarno sequence to change position relative to the translation start site while maintaining functionality.

Genetic Interactions in Cancer

Patterns of association between genetic alterations in cancer can provide clues as to the functional relationships between genes during cancer progression and can be useful in identifying driver mutations and therapeutic targets. One important pattern of association often observed among driver mutations is mutual exclusivity, in which the presence of a mutation in one gene predicts the absence of mutations in another gene and vice versa. The McCandlish laboratory has been collaborating with Justin Pritchard's group at Penn State on the evolution of tyrosine kinases and their role in cancer, and this year published a first paper from this collaboration describing a method for detecting mutually exclusive gene pairs when one or both of the mutations is only observed at low to moderate frequency (Iman et al. 2021).

Specifically, we developed a statistical model to assess whether a candidate gene pair has mutual exclusivity comparable to positive control gene pairs while controlling for the overall frequency of mutations in the candidate genes. As a case study for this approach, we considered ALK^{ATI}, a mutant form of the receptor tyrosine kinase ALK that results in a protein product consisting solely of the kinase domain, which has been proposed as a possible driver mutation and drug target for single-agent ALK inhibition therapy. Using our method, we showed that in melanoma ALK^{ATI} is not as mutually exclusive with BRAF or NRAS as BRAF and NRAS are with each other, even once we control for the low prevalence of ALK^{ATI} (which is observed in 2%–10% of melanoma patients), suggesting that ALK^{ATI} is not in fact acting as a driver mutation. Follow-up experiments then indicated several other lines of evidence, such as the inability of ALK^{ATI} to rescue growth in melanoma cell lines, suggesting that ALK^{ATI} is unlikely to provide a productive target for anti-ALK monotherapy in melanoma. The McCandlish laboratory is also collaborating with the Pritchard laboratory on several other projects, including high-throughput mutagenesis and drug sensitivity assays on the ABL kinase domain.

Although observed associations between pairs of mutations have long been useful in motivating mechanistic hypotheses, it is also interesting to ask what can be learned by looking at higher-order associations between three or more mutations. To identify such combinations of mutations, we collaborated with the Kinney and Sheltzer laboratories to develop a flexible Bayesian method of estimating probability distributions over large combinatorial spaces of genetic configurations. The method is called SeqDEFT (sequence density estimation using field theory), and this year we both further developed our understanding of its theoretical properties and published a first manuscript describing the method (Chen et al. 2021).

As an application, we used our method to study patterns of aneuploidy in cancer. Interestingly, although aneuploidy is associated with increased cellular proliferation and poor prognosis, altering the copy number of individual chromosomes in a euploid background tends to inhibit proliferation. This suggests that although most chromosomal alterations are individually deleterious, certain combinations of chromosomal abnormalities may act synergistically to increase proliferation. Using data on chromosomal abnormalities from 10,522 tumors drawn from the Cancer Genome Atlas, we scored each autosome in a tumor as being euploid or aneuploid and used SeqDEFT to estimate the joint distribution over the corresponding space of possible karyotypes, which distinguishes a total of $2^{22} = 4{,}194{,}304$ different patterns of chromosomal alteration. Using this method, we found several sets of chromosomes that tend to exhibit simultaneous copy number alterations, including up to seven-way interactions, observing such pattern in gliomas as well as multiple types of kidney cancers.

In ongoing work, we have been following up on several combinations of co-occurring chromosomal alterations in gliomas. These combinations differ between lower-grade gliomas and glioblastomas and also appear to distinguish different groups of glioblastomas that evolve toward distinct combinations of chromosomal abnormalities. This suggests the possibility of tracking the evolution of glioblastomas to identify the key decision points that lead to one fate or another. To further this goal, we have also been modeling survival time as a function of chromosomal state at diagnosis, finding substantial differences in patient survival time for these different sets of co-occurring chromosomal abnormalities and more broadly estimating the "survival landscape" over this large space of possible genetic configurations.

Influence of Mutational Biases on Molecular Adaptation

Evolutionary adaptation often occurs by the fixation of beneficial mutations. However, because of the specific mechanisms of DNA damage and repair at play in any given species, different beneficial mutations appear within an evolving species at different rates. Whether and to what extent these mutational biases influence the genetic basis of adaptive evolution has been a long-term subject of interest for the McCandlish laboratory.

This year, we continued our collaboration with Joshua Payne at ETH Zurich and Arlin Stoltzfus at the National Institute for Standards and Technology to develop a more quantitative statistical framework for measuring the influence of mutational biases on adaptation. Our framework uses negative binomial regression, a type of generalized linear model, to compare the spectrum of mutational types observed to contribute to adaptive evolution to independent estimates of mutation rates obtained from mutation accumulation experiments or patterns of segregating variation at putatively neutral sites. We have now applied this framework to analyze collections of adaptive substitutions in three species, *Saccharomyces cerevisiae*, *Escherichia coli*, and *Mycobacterium tuberculosis*. Our results indicate a directly proportional influence of the mutational spectrum on the spectrum of adaptive substitutions, which is the strongest effect compatible with current population-genetic theory.

Although the above study is based on large sets of adaptive mutations in three specific microbial species, a complementary question is whether similar patterns hold if we consider known examples of adaptive mutations from a broader array of species from across the tree of life. To address this question, postdoc Bryan Gitschlag joined the McCandlish laboratory this year in order to assemble a data set consisting of a large number of putative adaptive substitutions in protein coding sequences whose biochemical or fitness effect has been experimentally validated. Bryan is currently assembling this large data set. He is developing new mathematical theories to understand better the patterns of sequence and functional diversity among adaptive substitution that would be expected because of the influence of mutational biases.

PUBLICATIONS

Chen WC, Zhou J, Sheltzer JM, Kinney JB, McCandlish DM. 2021. Field-theoretic density estimation for biological sequence space with applications to 5′ splice site diversity and aneuploidy in cancer. *Proc Natl Acad Sci* **118:** e2025782118.

Inam H, Sokirniy I, Rao Y, Shah A, Naeemikia F, O'Brien E, Dong C, McCandlish DM, Pritchard JR. 2021. Genomic and experimental evidence that ALK^{ATI} does not predict single agent sensitivity to ALK inhibitors. *iScience* **24:** 103343.

McCandlish DM. 2021. System-specificity of genotype–phenotype map structure. Comment on "From genotypes to organisms: state-of-the-art and perspectives of a cornerstone in evolutionary dynamics," by Susanna Manrubia et al. *Phys Life Rev* **39:** 73–75.

UNCOVERING ALGORITHMS IN THE NATURAL WORLD

S. Navlakha J. Arber J. Bhandari E. Blum Y. She S. Srinivasan J. Suen X. Zheng I. Ziamtsov

Projecting COVID-19 Disease Severity in Cancer Patients Using Purposefully Designed Machine Learning

Accurately predicting outcomes for cancer patients with COVID-19 has been clinically challenging. Numerous clinical variables have been retrospectively associated with disease severity, but the predictive value of these variables, and how multiple variables interact to increase risk, remains unclear. We used machine-learning algorithms to predict COVID-19 severity in 348 cancer patients at Memorial Sloan Kettering Cancer Center in New York City. Using only clinical variables collected on or before a patient's COVID-19 positive date (time zero), we sought to classify patients into one of three possible future outcomes: severe-early (the patient required high levels of oxygen support within 3 days of being tested positive for COVID-19), severe-late (the patient required high levels of oxygen after 3 days), and nonsevere (the patient never required oxygen support). Our algorithm classified patients into these classes with an area under the receiver operating characteristic curve (AUROC) ranging from 70% to 85%, significantly outperforming prior methods and univariate analyses. Critically, classification accuracy is highest when using a potpourri of clinical variables—including basic patient information, preexisting diagnoses, laboratory and radiological work, and underlying cancer type—suggesting that COVID-19 in cancer patients comes with numerous combinatorial risk factors. Overall, we provide a computational tool that can identify high-risk patients early in their disease progression, which could aid in clinical decision-making and selecting treatment options. See Figure 1 and Navlakha et al. (2021).

A Correspondence between Normalization Strategies in Artificial and Biological Neural Networks

A fundamental challenge at the interface of machine learning and neuroscience is to uncover computational principles that are shared between artificial and biological neural networks. In deep learning, normalization methods such as batch normalization, weight normalization, and their many variants help to stabilize hidden unit activity and accelerate network training, and these methods have been called one of the most important recent innovations for optimizing deep networks. In the brain, homeostatic plasticity represents a set of mechanisms that also stabilize and normalize network activity to lie within certain ranges, and these mechanisms are critical for maintaining normal brain function. In Shen et al. (2021), we discuss parallels between artificial and biological normalization methods at four spatial scales: normalization of a single neuron's activity, normalization of synaptic weights of a neuron, normalization of a layer of neurons, and normalization of a network of neurons. We argue that both types of methods are functionally equivalent—that is, both push activation patterns of hidden units toward a homeostatic state, where all neurons are equally used—and we argue that such representations can improve coding capacity, discrimination, and regularization. As a proof of concept, we develop an algorithm, inspired by a neural normalization technique called synaptic scaling, and show that this algorithm performs competitively against existing normalization methods on several data sets. Overall, we hope this bidirectional connection will inspire neuroscientists and machine learners in three ways: to uncover new normalization algorithms based on established neurobiological principles; to help quantify the trade-offs of different

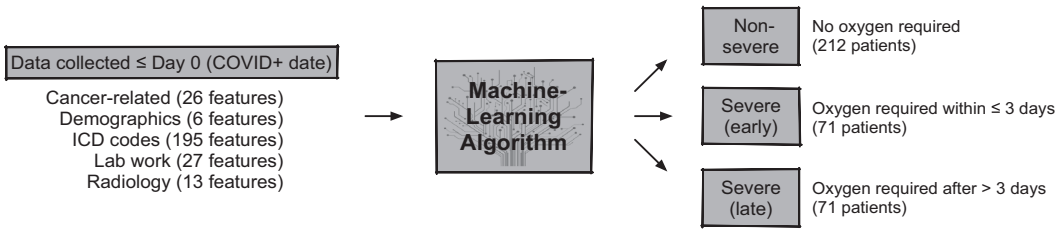


Figure 1. Overview of the study. (*Left*) Data for 348 inpatients at Memorial Sloan Kettering Cancer Center were analyzed. For each patient, up to 267 clinical variables were collected, including basic patient information, cancer history, ICD medical history, laboratory work, and radiology work. Variables were only collected up to the patient's COVID-19+ date (time zero). (*Middle*) Variables are inputted into a machine learning algorithm (a random forest classifier), which learns to predict patient outcomes based on interactions between multiple variables. (*Right*) Three possible patient outcomes. Of the 348 patients, 206 did not require high levels of oxygen support, 71 required oxygen support within three days of being tested positive for COVID-19, and 71 patients required oxygen support after three days.

homeostatic plasticity mechanisms used in the brain; and to offer insights about how stability may not hinder, but may actually promote, plasticity. See Figure 2.

Branch-Pipe: Improving Graph Skeletonization around Branch Points in 3D Point Clouds

Modern plant phenotyping requires tools that are robust to noise and missing data while being able to efficiently process large numbers of plants. Here, we studied the skeletonization of plant architectures from 3D point clouds, which is critical for many downstream tasks, including analyses of plant shape, morphology, and branching angles. Specifically, we developed an algorithm to improve skeletonization at branch points (forks) by leveraging the geometric properties of cylinders around branch points. We tested this algorithm on a diverse set of high-resolution 3D point clouds of tomato and tobacco plants grown in five environments and across multiple developmental timepoints. Compared to existing methods for 3D skeletonization, our method efficiently and more accurately estimated branching angles even in areas with noisy, missing, or nonuniformly sampled data. Our method is also applicable to inorganic data sets, such as scans of industrial pipes or urban scenes containing networks of complex cylindrical shapes. See Figure 3 and Ziamtsov et al. (2021).

Better Tired Than Lost: Turtle Ant Trail Networks Favor Coherence over Short Edges

Creating a routing backbone is a fundamental problem in both biology and engineering. The routing backbone of the trail networks of arboreal turtle ants (*Cephalotes goniodontus*) connects many nests and food sources using trail pheromone deposited by ants as they walk. Unlike species that forage on the ground, the trail networks of arboreal ants are constrained by the vegetation. We examined what objectives the trail networks meet by comparing the observed ant trail networks with networks of random, hypothetical trail networks in the same surrounding vegetation and with trails optimized for four objectives: minimizing path length, minimizing average edge length, minimizing number of nodes, and minimizing opportunities to get lost. The ants' trails minimized path length by minimizing the number of nodes traversed rather than choosing short edges. In addition, the ants' trails reduced the opportunity for ants to get lost at each node, favoring nodes with 3D configurations most likely to be reinforced by pheromone. Thus, rather than finding the shortest edges, turtle ant trail networks take advantage of natural variation in the environment to favor coherence, keeping the ants together on the trails. See Figure 4 and Chandrasekhar et al. (2021).

Neural Network Features Distinguish Chemosensory Stimuli in *Caenorhabditis elegans*

Nervous systems extract and process information from the environment to alter animal behavior and physiology. Despite progress in understanding how different stimuli are represented by changes in neuronal activity, less is known about how they affect broader neural network properties. We developed a framework for using graph-theoretic features of neural network activity to predict ecologically relevant stimulus properties—in particular, stimulus identity.

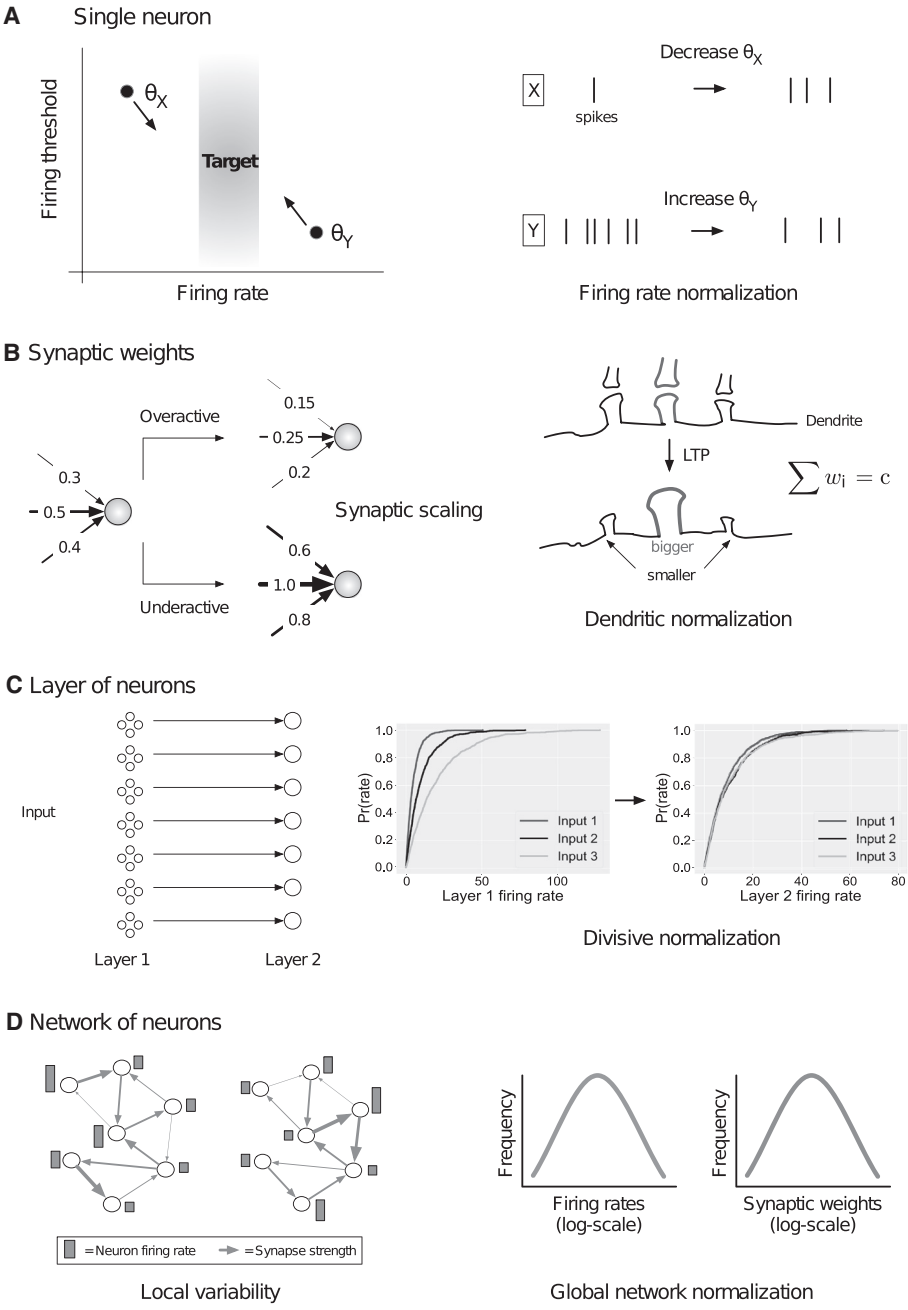


Figure 2. Neural homeostatic plasticity mechanisms across four spatial scales. (*A*) Normalization of a single neuron's activity. (*Left*) Neuron X has a relatively low firing rate and a high firing threshold, θ_X, and vice versa for neuron Y. (*Right*) Both neurons can be brought closer to their target firing rate by decreasing θ_X and increasing θ_Y. (*B*) Normalization of synaptic weights. (*Left*) (synaptic scaling): If a neuron is firing above its target rate, its synapses are multiplicatively decreased, and vice versa if the neuron is firing below its target rate. (*Right*) (dendritic normalization): If a synapse size increases due to strong LTP, its neighboring synapses decrease their size. (*C*) Normalization of a layer of neurons. (*Left*) Two layers of neurons with feedforward connections and other feedback inhibitory connections (not shown). (*Right*) The cumulative distribution of firing rates for neurons in the first layer is exponential with a different mean for different inputs. The activity of neurons in the second layer is normalized such that the means of the three exponentials are approximately the same. (*D*) (*Left*) Example of a neural circuit with the same units and connections but different activity levels for neurons (purple bars) and different weights (pink arrow thickness) under two different conditions. (*Right*) Despite local variability, the global distributions of firing rates and synaptic weights for the network remains stable (log-normally distributed) under both conditions.

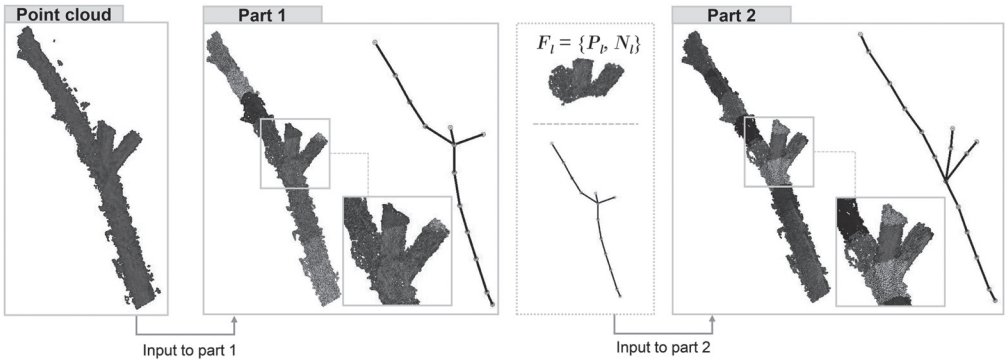


Figure 3. Overview. As input, we are provided a 3D point cloud of a plant architecture. A partial point cloud of the region around a branch point (fork) is shown. In part 1, we compute its skeleton. Each color represents points and normal vectors that belong to a node in the graph. In part 2, the geometry and branch angles around a fork are refined to better match the underlying plant shape.

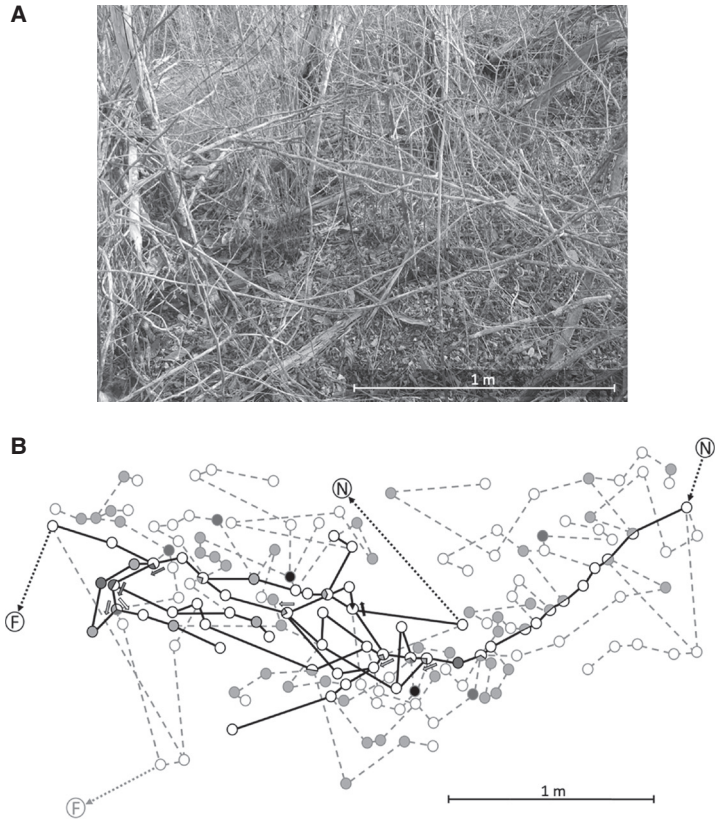


Figure 4. Used and available networks in the vegetation. (*A*) Vegetation in which the trail network mapped in B was made, photographed in the dry season before the branches have leaves. (*B*) Illustration of part of the trail network and network of surrounding vegetation for Tejon 189 on day 9. The figure shows 166 of the 217 nodes mapped in the surrounding vegetation. For many nodes, not all edges are shown to simplify the illustration. Edge lengths are scaled to measured distance, but actual location is not represented here. N represents a nest; F represents a food source. Circles represent nodes. Solid lines represent edges used on that day; dashed lines represent edges not used that day. The color of a node represents the transition index (TI) from the preceding edge to the following one: TI–1, open circles; TI–2, blue; TI–3, red; TI–4, black. At a node where there is a choice of more than one edge in the indicated direction, so that there could be more than one transition taken through a given node, a TI was assigned to each possible transition. For such nodes with more than one transition index, the TI is represented graphically with a pie chart, and arrows show which transition has the TI represented by the arrow's color.

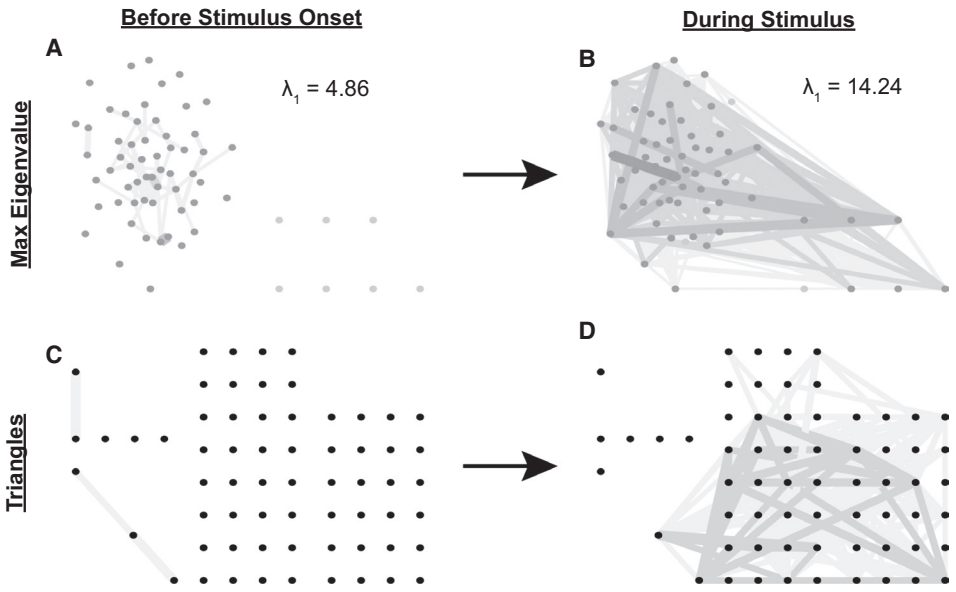


Figure 5. NaCl induces an increase in max eigenvalue and number of (strong) triangles. The networks depicted here, for one example worm, show neurons (circles) connected by lines (edges). In *A* and *B*, the normalized mutual information (NMI) between any two neurons is indicated by both edge shading and thickness, where darker and thicker lines indicate larger NMI. *A* and *B* show the network's max eigenvalue (λ_1), and *C* and *D* depict the strongest (i.e., NMI ≥ 0.4) interactions in the network (hence, why many edges are missing), with those in green indicating the presence of a triangle (i.e., a triplet of connected neurons whose total weight ≥ 1.8) and those in blue its absence. Before the worm is exposed to 200 mM NaCl, the neural network is composed of weak interactions, producing a low max eigenvalue (*A*) and no strong triangles (*C*). Once exposed to 200 mM NaCl, the worm's neural network has many stronger interactions, increasing its max eigenvalue (B) and forming several strong triangles (*D*). Edges in *A* and *B* were assigned to 0.2-wide bins from 0 to 1. All edges less than 0.4 were removed from *C* and *D* for the purpose of visualization. Neuron positions were fixed between *A* and *B*, and *C* and *D*, such that *A* can be directly compared to *B*, and *C* can be directly compared to *D*.

We used the transparent nematode, *C. elegans*, with its small nervous system to define neural network features associated with various chemosensory stimuli. We first immobilized animals using a microfluidic device and exposed their noses to chemical stimuli while monitoring changes in neural activity of more than 50 neurons in the head region. We found that graph-theoretic features, which capture patterns of interactions between neurons, are modulated by stimulus identity. Further, we show that a simple machine-learning classifier trained using graph-theoretic features alone, or in combination with neural activity features, can accurately predict salt stimulus. Moreover, by focusing on putative causal interactions between neurons, the graph-theoretic features were almost twice as predictive as the neural activity features. These results reveal that stimulus identity modulates the broad, network-level organization of the nervous system, and that graph theory can be used to characterize these changes. See Figure 5 and How et al. (2021).

PUBLICATIONS

Chandrasekhar A, Marshall JAR, Austin C, Navlakha S, Gordon DM. 2021. Better tired than lost: turtle ant trail networks favor coherence over short edges. *PLoS Comput Biol* **17:** e1009523.

How JJ, Navlakha S, Chalasani SH. 2021. Neural network features distinguish chemosensory stimuli in *Caenorhabditis elegans*. *PLoS Comput Biol* **17:** e1009591.

Navlakha S, Morjaria S, Perez-Johnston R, Zhang A, Taur Y. 2021. Projecting COVID-19 disease severity in cancer patients using purposefully-designed machine learning. *BMC Infect Dis* **21:** 391.

Shen S, Wang J, Navlakha S. 2021. A correspondence between normalization strategies in artificial and biological neural networks. *Neural Comput* **33:** 3179–3203.

Ziamtsov I, Faizi K, Navlakha S. 2021. Branch-Pipe: improving graph skeletonization around branch points in 3D point clouds. *Remote Sens* **13:** 3802.

POPULATION GENETICS AND TRANSCRIPTIONAL REGULATION

A. Siepel L. Liu M. Mughal A. Scheben Y. Zhao
Z. Mo R. Ramani A. Xue

Our research focuses on two major areas: human population genetics and the evolution of transcriptional regulation. The research in population genetics is done mostly with publicly available genomic sequence data. In contrast, most of the work on transcriptional regulation is done with our collaborator Charles Danko at Cornell. We also have smaller collaborative projects on topics ranging from prediction of the fitness consequences of mutations in maize and other crops (with Ed Buckler, Cornell), to comparative genomics of bats (with Dick McCombie, CSHL), to epidemiological modeling of COVID-19 (with Rob Martienssen, CSHL). We focus on theoretical and computational research and do not generate our own data, but we often work closely with experimental collaborators on projects with substantial experimental and computational components. We are broadly interested in molecular evolution, population genetics, gene regulation, 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 one staff member (Ritika Ramani) having departed in 2021 and one new postdoctoral associate soon to arrive.

Below we describe recent progress in three main research areas.

Inference of Demographic History and Natural Selection from Complete Genome Sequences

We have a long-standing interest in reconstructing the demographic history of complex, structured populations from DNA sequence data. Our group developed the first scalable method for genome-wide inference of the "ancestral recombination graph" or ARG. We used this method, called *ARGweaver*, to detect gene flow from modern humans into Altai Neandertal genome sequences and provided evidence for an earlier migration of modern humans out of Africa than indicated by most current estimates. Recently, we extended the *ARGweaver* algorithm to support a full multipopulation demographic model, including population divergence times and migration events. We used this method, called *ARGweaver-D*, to show that ~3% of Neandertal DNA—and possibly as much as 6%—came from modern humans who mated with Neandertals 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 separate collaborative study, we applied *ARGweaver* together with newly developed machine-learning methods to study speciation in the finch-like southern capuchino seedeaters of South America. We use these new methods to show signatures of recent selective sweeps around pigmentation genes, including many soft sweeps that acted on standing variation. Our observations help fill in the picture of how selection, recombination, and drift act together to shape the genomes of distinct species, with broad applications for the tree of life.

In a related project led by Ph.D. student Ziyi Mo and former postdoc Hussein Hejase, we have developed a deep-learning framework called selection inference using the ARG (SIA) for detecting selection based on inferred ARGs. SIA aims to improve selection inference by taking advantage of both the richness of information in ARGs and the flexibility and scalability of deep-learning methods (Hejase et al. 2021). The SIA framework can be trained to infer a full range of selection coefficients, the allele frequency trajectory, and time of selection onset. We benchmarked SIA extensively on simulations under a European human demographic model. We found that it performs as

well or better than some of the best available methods, including state-of-the-art machine-learning and ARG-based techniques. In addition, we used SIA to estimate selection coefficients at several loci associated with human phenotypes of interest. SIA detected novel signals of selection particular to the European (CEU) population at the *MC1R* and *ABCC11* loci. In addition, SIA recapitulated signals of selection at the *LCT* locus and several pigmentation-related genes. Overall, SIA uses deep learning to leverage the ARG, providing new insight into how selective sweeps shape genomic diversity.

Genome sequencing of tens of thousands of humans has enabled the measurement of significant selective effects for mutations to protein-coding genes. Mehreen Mughal and former postdoc Noah Dukler developed a new method called extremely rare INSIGHT (*ExtRaINSIGHT*) for measuring similar selective effects in noncoding and coding regions of the human genome (Dukler et al. 2022). *ExtRaINSIGHT* estimates the prevalence of strong purifying selection, or "ultraselection," as the fractional depletion of rare single-nucleotide variants in target genomic sites relative to matched sites that are putatively free from selection, after controlling for local variation neighbor dependence in mutation rate. We applied *ExtRaINSIGHT* to 71,702 whole-genome sequences from gnomAD v3 and found strong evidence of ultraselection in evolutionarily ancient miRNAs and neuronal protein-coding genes, as well as at splice sites. By contrast, we found much less ultraselection in other noncoding RNAs and transcription factor binding sites and only modest evidence of ultraselection in ultraconserved elements and human accelerated regions. Overall, this study shed new light on the genome-wide distribution of fitness effects for new point mutations by combining deep new sequencing data sets and classical theory from population genetics.

Alexander Xue in the group has been developing a model to estimate the strength of polygenic selection underlying a complex trait by exploiting data from genome-wide association studies (GWASs). He has recently developed a program called *ASSESS* that exploits the Poisson random field (PRF) framework to model selection coefficients from genome-wide allele frequency data. His method simultaneously models GWAS summary statistics and the true effect size of each genetic variant on a given trait to infer signatures of natural selection. The likelihood function is optimized with an expectation-maximization algorithm to jointly infer a trait's polygenicity, effect size distribution, and distribution of fitness effects (DFE). Xander has recently shown in simulation experiments that *ASSESS* can recover true parameter values under a range of demographic, selection, effect size, genomic, and ascertainment regimes. Work is under way to apply *ASSESS* to various publicly available human data sets, including ones derived from the UK Biobank and others for non-European populations. As GWAS results accumulate, we expect *ASSESS* to serve as a powerful yet convenient population genomic inference framework for the human population genomics community.

Analysis of Essentiality of Human Genes Using CRISPR

We developed a new probabilistic inference method, called analysis of CRISPR-based essentiality (*ACE*), for detecting and quantifying gene essentiality from high-throughput CRISPR-Cas9 knockout screens (Hutton et al. 2021). *ACE* uses a hierarchical model to account for the uncertainty associated with each stage of a CRISPR screen and enable maximum-likelihood estimation of gene-level essentiality. The modularity of our framework permits the analysis of a variety of CRISPR-screen data sets while accounting for differences in experimental design. Using simulations, we showed that *ACE* is competitive with the best available methods in predicting both absolute and differential gene essentiality. *ACE* identified known and novel candidates for genotype-specific essentiality when applied to publicly available CRISPR-screen data. Specifically, we showed that RNA methyltransferases exhibit enhanced essentiality in the presence of inactivating *TP53* mutations. Overall, *ACE* provides a robust framework for identifying genes responsive to subtype-specific therapeutic targeting.

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 (precision nuclear run-on sequencing). 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.

We developed a method for estimating relative RNA half-lives based on PRO-seq and RNA-seq data (Blumberg et al. 2021). 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. Together, this estimation method and systemic analysis shed light on the pervasive impacts of RNA stability on cellular RNA concentrations. In a more recent project, Yixin Zhao developed a new software tool called deconvolution of expression for nascent RNA (DENR) to address the unique challenges of quantifying newly transcribed RNA (Zhao et al. 2021). DENR uses a hybrid machine-learning/statistical-modeling approach to estimate expression at both the gene and transcript levels. We applied DENR to previously published PRO-seq data and found it reveals high levels of pre-RNA isoform diversity in human cells.

We recently introduced a unified probabilistic modeling framework to analyze nascent RNA-sequencing data (Siepel 2021) that attempts to address limitations in current methods that are generally narrowly focused on specific prediction tasks. Our generative model enables estimation of separate rates of initiation, pause release, and termination, up to a proportionality constant, and supports a variety of other analyses. This framework not only enables the estimation of key kinetic parameters but, if applied to time-course data in a nonequilibrium setting, can also be used to estimate elongation rates. Furthermore, the model naturally leads to likelihood ratio tests for differences between genes, conditions, or species in various rates of interest. A version of the model in which read counts are assumed to be Poisson-distributed leads to convenient, closed-form solutions for parameter estimates and likelihood ratio tests. Extensions are available to Bayesian inference and a generalized linear model that can be used to discover genomic features associated with elongation rates. This modeling framework enables a unified treatment of many common tasks in analyzing nascent RNA sequencing data.

Recent Collaborative Studies

In a collaborative study with the Martienssen laboratory at CSHL led by Ph.D. student Ziyi Mo, we explored whether circadian immunity contributes significantly to the seasonality of respiratory viruses, including influenza and SARS-CoV-2. Following the general susceptibility-infection-recovery-susceptibility (SIRS) paradigm, we developed models for influenza and COVID-19 (Mo et al. 2021). We fitted them to public data for infections (in influenza) and hospitalizations and deaths (in COVID-19). These models suggest that local sunrise time is a better predictor of the basic reproductive number (R0) than climate, even when day length is considered. Moreover, these models predict a window of susceptibility when local sunrise time corresponds to the morning commute, and the contact rate is expected to be high. Counterfactual modeling suggests that retaining daylight savings time in the fall would reduce the length of this window and substantially reduce seasonal waves of respiratory infections.

Molecular analyses suggest that SARS-CoV-2 crossed into humans from an animal species, most likely a bat. In a study with the McCombie laboratory at CSHL, postdoc Armin Scheben investigated the genomic underpinnings of unique bat adaptations by sequencing the genomes of the Jamaican fruit bat (*Artibeus jamaicensis*) and the Mesoamerican mustached bat (*Pteronotus mesoamericanus*) and comparing them to a diverse collection of 13 additional bat species and other mammals (Scheben et al. 2021). This study revealed a striking contraction of IFN-α at the type I interferon locus in bats, resulting in a major shift in relative IFN-ω and IFN-α copy numbers. This shift could contribute to the increased viral tolerance that has made bats a reservoir for viruses such as progenitors of SARS-CoV-1 and SARS-CoV-2. Antiviral genes stimulated by type I interferons also showed evidence of adaptation, including a lineage-specific duplication of interferon-induced transmembrane genes and positive selection in *IFIT2*. In addition, 33 tumor suppressors and six DNA-repair genes showed signs of positive selection, perhaps contributing to increased longevity and reduced cancer rates in bats. Our study provides valuable new genomic resources for bats and sheds new light on the extraordinary molecular adaptations in this critically important group of mammals.

PUBLICATIONS

Blumberg A, Zhao Y, Huang Y-F, Dukler N, Rice EJ, Krumholz K, Danko CG, Siepel A. 2021. Characterizing RNA stability genome-wide through combined analysis of PRO-seq and RNA-seq data. *BMC Biol* **19:** 30.

Hejase HA, Mo Z, Campagna L, Siepel A. 2021. A deep-learning approach for inference of selective sweeps from Ancestral Recombination Graph. *Mol Biol Evol* **39:** msab332.

Henry S, Trousdell MC, Cyrill SL, Zhao Y, Feigman MJ, Bouhuis JM, Aylard DA, Siepel A, dos Santos CO. 2021. Characterization of gene expression signatures for the identification of cellular heterogeneity in the developing mammary gland. *J Mammary Gland Biol Neoplasia* **26:** 43–66.

Hutton ER, Vakoc CR, Siepel A. 2021. ACE: a probabilistic model for characterizing gene-level essentiality in CRISPR screens. *Genome Biol* **22:** 278.

Mo Z, Scheben A, Steinberg J, Siepel A, Martienssen R. 2021. Circadian immunity, sunrise time and the seasonality of respiratory infections. medRxiv doi:10.1101/2021.03.29.21254556

Scheben A, Mendivil Ramos O, Kramer M, Goodwin S, Oppenheim S, Becker DR, Schatz MC, Simmons NB, Siepel A, McCombie WR. 2021. Long-read sequencing reveals rapid evolution of immunity- and cancer-related genes in bats. bioRxiv doi:10.1101/2020.09.09.290502

Siepel A. A unified probabilistic modeling framework for eukaryotic transcription based on nascent RNA sequencing data. bioRxiv doi:10.1101/2021.01.12.426408

Zhao Y, Dukler N, Barshad G, Toneyan S, Danko CG, Siepel A. 2021. Expression for nascent RNA sequencing data (DENR) highlights pre-RNA isoform diversity in human cells. *Bioinformatics* **37:** 4727–4736.

In Press

Dukler N, Mughal MR, Ramani R, Huang Y-F, Siepel A. 2022. Extreme purifying selection against point mutations in the human genome. *Nat Commun* **13:** 4312. doi:10.1038/s41467-022-31872-6

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 a Distinguished Professor at the University of California, Santa Cruz. 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** has been a CSHL Fellow since 2019. She combines genomics and mathematical modeling to study how the immune system discriminates self from non-self in order to effectively fight infection.

NUTRIENT–GENE INTERACTIONS IN REGENERATION, IMMUNITY, AND CANCER

S. Beyaz C. Chung A. Katcher
I. Ergin K. Ozler
O. Eskiocak V. Shah
A. Garipcan B. Yueh

The Beyaz laboratory is focused on understanding the mechanistic basis for nutritional regulation of cell fate. Although it is becoming increasingly evident that nutrients may directly regulate cell fate and function, little is known about the underlying causal mechanisms. The main challenge we are addressing in the field is the fact that nutrients vary in types, abundance, and patterns of diet and perturb interconnected modules of cells, cell networks in tissues, organs, and organismal states that reciprocally influence one another. This complexity makes it very challenging to uncover causal cellular and molecular mechanisms that link nutrients to regulation of cell fate and function. Currently there are two paradigms that posit a metabolic code in cell fate and function. The first paradigm is based on the identification of cell state–specific metabolic programs that associate with cell fate and function. The second paradigm is based on the observations that organismal metabolic states or diets influence cell fate and function in tissues and have significant impact on physiology and disease. Although these paradigms suggest a key role for metabolism in cell fate and function, underlying causal mechanisms in physiology and disease are not well understood.

We implement a modular and systematic strategy to address these challenges, fill the gaps in existing paradigms, and uncover the mechanistic basis for how nutrient metabolism, in particular fatty acid (FA) metabolism, regulates cell fate and function. The conceptual frameworks of this strategy are based on our ability to study how dietary FAs and their metabolism affect cell fate and function through cell-intrinsic epigenetic mechanisms or by perturbing the interactions between distinct modules of cell networks including stem cells, immune cells, and microbiome in the intestine. These modules have reciprocal interactions with physiology in the context of regeneration or disease states such as cancer.

Nutritional 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 discovered a causal epigenetic mechanism that links high-fat diet (HFD) and obesity to increased cancer risk through FA-induced activation of the lipid-sensing transcription factor Ppard in stem cells, which promotes stemness and contributes to tumor initiation when coupled with oncogenic mutations such as loss of the tumor suppressor *Apc* (Beyaz et al., *Nature* **531:** 53 [2016]). To further assess the precise role of Ppard in mediating stemness-enhancing effects of pro-obesity HFD in the intestine, we generated new conditional and inducible genetic models that produce a transcriptionally active form of PPAR (LSL-CAG-VP16-PPAR) specifically in intestinal epithelial cells (IECs; Villin-CreERt2) or intestinal stem cells (ISCs; Lgr5-CreERt2). Using these models, we definitively demonstrated that activation of PPAR in IECs or ISCs enhances stemness (Beyaz et al. 2021a).

To define how diverse dietary FAs influence stem cell regeneration in the intestine, we developed a phenotypic organoid screening strategy using dietary FAs. We identified arachidonic acid (AA) as a regenerative nutrient that effects stem cell reprogramming in mouse and human organoids. AA is a bioactive lipid that plays essential structural and functional roles in mammalian cells and tissues, including intestinal epithelium. However, little is known about the functional significance of dietary AA supplementation in intestinal homeostasis. By utilizing isocaloric diets that vary in AA abundance, we demonstrated that AA promotes both homeostatic and damage-induced intestinal stem cell regeneration in vivo. Cross-species gene expression analysis revealed induction of conserved repair-associated stem cell reprogramming signatures in response to AA treatment. Using single-cell

RNA sequencing (scRNA-seq), we identified AA-induced de novo stem cell states and dedifferentiation programs in vivo and in vitro. Mechanistically, AA-derived PGE2 signals through the PTGER4-cAMP-PKA axis to promote stemness in mice and humans. Finally, we found that AA evokes epigenetic reprogramming around stem cell regeneration–associated genes in a PTGER4-dependent manner. These findings demonstrate that dietary AA is a conserved promoter of stem cell regeneration that mimics the repair response to tissue injury through PGE2-PTGER4 signaling and downstream epigenetic reprogramming.

Nutritional Regulation of Microbe–Stem Cell–Immune Cell Interactions in the Context of Cancer

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. We uncovered causal mechanisms of how a lard-based pro-obesity HFD impairs immune–epithelial stem cell cross-talk and contributes to tumor initiation in the intestine. In this study, we defined a causal role for epithelial stem cell MHC-II expression in tumor initiation in the intestine. Importantly, we found that HFD-mediated perturbations in intestinal microbiome and immune cell states lead to dampened MHC-II expression in stem cells, which, when coupled with an oncogenic event such as APC loss, promoted tumor initiation (Beyaz et al. 2021b).

Development of Human Organoid Models to Study Nutrient–Gene Interactions in Humans

To study the mechanistic basis of how perturbations in nutrient metabolism influence human physiology and disease states, we established comprehensive human tissue and patient-derived organoid (PDO) biobanks. Given their significant association with cancer in the context of obesity, we focused on endometrium (between two- and sevenfold increase in cancer risk in obesity) and colon (~40% increase in cancer risk in obesity). We formed a collaborative network between scientists and clinicians at CSHL-Northwell-NYGC to contribute to the NYGC Polyethnic-1000 initiative with the aim of understanding the mechanistic basis of cancer health disparities in colorectal cancer and endometrial cancer, including potential genetic or environmental risk factors such as obesity. In addition, our collaborative team recently became part of the Chan Zuckerberg Initiative Human Cell Atlas Network to develop the most comprehensive cellular map of endometrium across ancestries with lean and obese states. To study the mechanisms of nutrient–gene interactions in endometrial cancer, we developed PDO models (for more than 150 patients) across endometrial cancer subtypes, including rare but very aggressive cancers such as carcinosarcoma. We successfully contributed to the collaborative CSHL/Human Cancer Models Initiative (HCMI) Leidos Project that aims to develop PDO models for the scientific community with 15 of our endometrial cancer PDO models.

PUBLICATIONS

Beyaz S, Mana MD, Yilmaz ÖH. 2021a. High-fat diet activates a PPAR-δ program to enhance intestinal stem cell function. *Cell Stem Cell* **28:** 598–599. doi:10.1016/j.stem.2021.03.001

Beyaz S, Chung C, Mou H, Bauer-Rowe KE, Xifaras ME, Ergin I, Dohnalova L, Biton M, Shekhar K, Eskiocak O, et al. 2021b. Dietary suppression of MHC class II expression in intestinal epithelial cells enhances intestinal tumorigenesis. *Cell Stem Cell* **28:** 1922–1935. doi:10.1016/j.stem.2021.08.007

Burr AHP, Ji J, Ozler K, Eskiocak O, Yueh B, Menk AV, Costa ASH, Rittenhouse N, Marshall CW, Chiaranunt P, et al. 2021. Excess dietary sugar impairs colonic epithelial regeneration in response to damage. bioRxiv doi:10.1101/2021.08.18.456840

Chu TH, Khairallah C, Shieh J, Cho R, Qiu Z, Zhang Y, Eskiocak O, Thanassi DG, Kaplan MH, Beyaz S, et al. 2021. γδ T cell IFNγ production is directly subverted by *Yersinia* pseudotuberculosis outer protein YopJ in mice and humans. *PLoS Pathog* **17:** e1010103. doi:10.1371/journal.ppat.1010103

Furlan A, Corona A, Boyle S, Sharma R, Rubino R, Habel J, Gablenz EC, Giovanniello J, Beyaz S, Janowitz T, et al. 2021. Neurotensin neurons in the central extended amygdala control energy balance. bioRxiv doi:10.1101/2021.08.03.454970

Gaudino SJ, Beaupre M, Lin X, Joshi P, Rathi S, McLaughlin PA, Kempen C, Mehta N, Eskiocak O, Yueh B, et al. 2021. IL-22 receptor signaling in Paneth cells is critical for their maturation, microbiota colonization, Th17-related immune responses, and anti-Salmonella immunity. *Mucosal Immunol* **14:** 389–401. doi:10.1038/s41385-020-00348-5

Hanasoge Somasundara AV, Moss MA, Feigman MJ, Chen C, Cyrill SL, Ciccone MF, Trousdell MC, Vollbrecht M, Li S, Kendall J, et al. Parity-induced changes to mammary epithelial cells control NKT cell expansion and mammary oncogenesis. *Cell Rep* **37:** 110099. doi:10.1016/j.celrep.2021.110099

MECHANISMS OF TOLERANCE INDUCTION IN THE THYMUS

H.V. Meyer S. Carcy A. Garipcan
J. Carter E. Mullé
S. Chapin K. Papciak
S. Fumagali

Mechanisms of Tolerance Induction in the Thymus

The thymus plays a key role in teaching T cells the ability to distinguish self from non-self. Flaws in the teaching process can lead to autoimmune diseases or immunodeficiency. The Meyer laboratory combines genomics and mathematical modeling to understand the mechanisms of healthy and pathological thymus function. 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.

In the past year, we investigated the rules that determine how epithelial cells in the thymus regulate the level of teaching they can provide and how that influences a T cell's ability to distinguish self from non-self. To do so, we investigate gene expression and regulation in human thymus samples, develop methods to analyze gene expression on a spatially resolved level, and design an in silico model of T-cell development to test and refine hypotheses about the teaching processes. A summary of our recent findings on transcriptional regulation in the human thymus is provided below. In addition, we have collaborated with other investigators at CSHL; a brief summary of these collaborative projects follows at the end.

Transcriptional Regulation in the Human Thymus

Antigen epitopes (i.e., the part of a protein recognized by the immune system) missing in the thymus can result in autoimmunity by allowing autoreactive T cells to escape into the periphery. To understand which antigens are available for T-cell education we need to uncover the transcriptional mechanisms underlying promiscuous gene expression in medullary thymic epithelial cells (mTECs).

We answered this question by conducting a genome-wide investigation of transcription start site usage, splicing patterns and endogenous retroelement (ERE)-induced gene expression using 5′ Cap-sequencing and RNA-seq in two differentiation states of human mTECs—immature, major histocompatibility complex II (MHCII) low-expressing mTECs^{lo}, and mature, MHCII high-expressing mTECs^{hi}. Previous research using targeted analyses of transcription initiation in mTECs had identified misinitiation events that led to escape of autoreactive T cells during education. For the first time, we conducted a genome-wide screen of transcription initiation events in human mTECs and discovered that misinitiation events are linked to the transcriptional regulator AIRE (Fig. 1A,B). In addition, prior studies lacked the depth or technique to conduct transcript-specific gene expression analysis in human mTECs. We report the first comprehensive splicing analysis in human mTECs and propose non-promiscuous usage of peripheral splicing factors as the mechanism for the increased splicing entropy that we discovered (Fig. 1C,D). Furthermore, we discovered linked expression of endogenous retroelements and protein coding genes in mTECs (Fig. 1E–G), a phenomenon that we propose is driven by epigenetic remodeling that enables promiscuous gene expression. Knowledge of epitopes that are missing during human T tolerance induction will be crucial to identify potential targets for cancer immunotherapy or serve as biomarkers in autoimmune diseases. We provide the epitope databases of human mTECs and peripheral tissues to the wider scientific community as an interactive web resource: http://epitopediversity.cshl.edu. A manuscript describing these findings is currently under review.

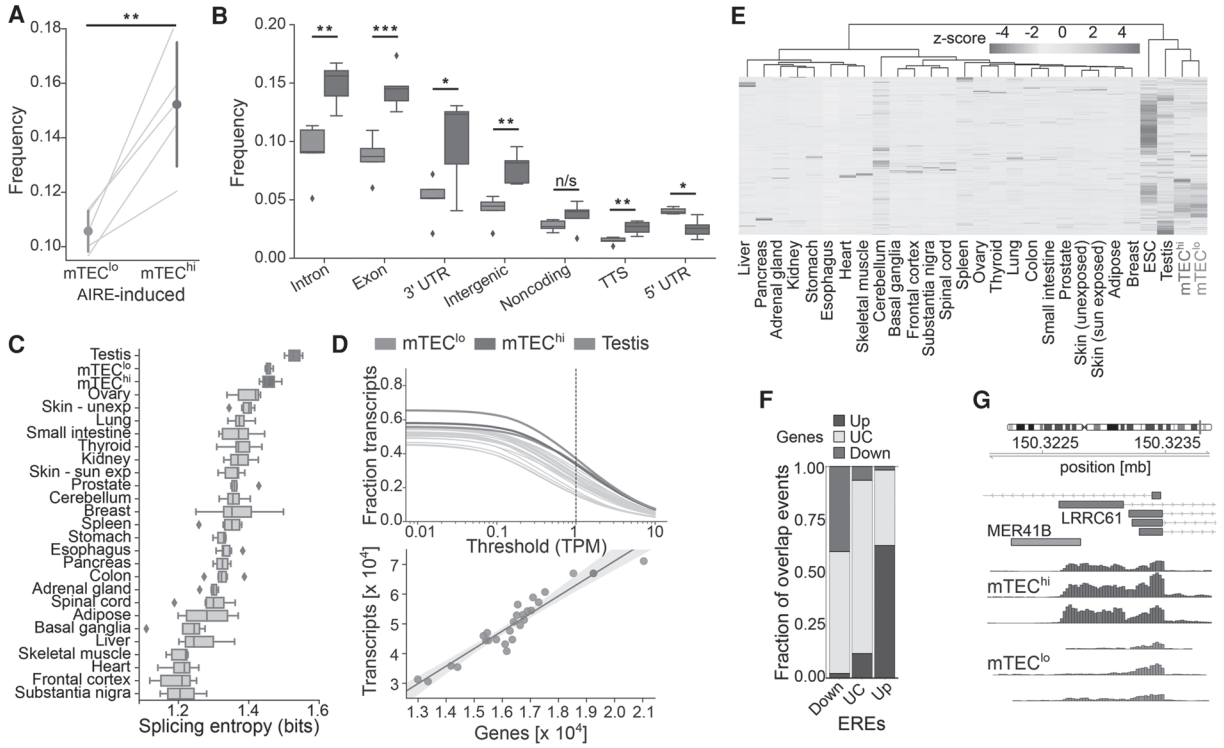


Figure 1. Epitope diversity in human thymic epithelial cells. (*A*) Comparison of the frequency with which transcription start regions (TSRs) are associated with autoimmune regulator (AIRE) dependent genes between paired (gray lines) human mTEC^{lo} (orange) and mTEC^{hi} (blue) samples. (*B*) Distribution of genomic location annotations for TSRs unique to either the mTEC^{hi} or mTEC^{lo} populations. Boxplots show median values with interquartile ranges. (TTS) Transcription termination site, (UTR) untranslated region. (*C*) Splicing entropy calculated for mTEC^{hi} and mTEC^{lo} samples as well as 25 healthy peripheral tissue samples from the Genotype-Tissue Expression (GTEx) project. (*D*) Fraction of transcriptome expressed at varying transcripts per million (TPM) thresholds for mTEC and peripheral tissue samples (*top*) and linear regression fitting the number of expressed transcripts as a function of the number of expressed genes in healthy GTEx samples predicts number of transcripts in mTEC^{hi} and mTEC^{lo} samples (*bottom*). (*E*) Expression (in *z*-scored TPM counts) for endogenous retroelement (ERE) subfamilies in mTECs, embryonic stem cells (ESCs), and 25 GTEx tissues. (*F*) Overlap of expression status between genes and EREs with a transcription start site within the gene body ± 1000 bp. The gene expression status is color-coded by the differential expression between mTEC^{hi} and mTEC^{lo}. (Up) Up-regulated, (Down) down-regulated, (UC) unchanged; corresponding categorization for ERE expression on the *x*-axis. (*G*) MTEC^{hi}-specific initiation event from the MER41B promoter (ERE) into the protein-coding gene *LRRC61*. Expression tracks show counts per million in the mTEC populations.

Collaborative Projects

In addition to our group's work on thymus physiology, we have worked closely with colleagues at CSHL. In these collaborations, we have applied our expertise in studying tissue physiology and high-dimensional data analysis to intestinal stem cell tumorigenesis (in collaboration with the Beyaz laboratory; Beyaz et al. 2021) and cancer immunotherapy (in collaboration with the Janowitz laboratory; Kleeman et al. 2021).

PUBLICATIONS

Beyaz S, Roper J, Bauer-Rowe KE, Xifaras ME, Ergin I, Dohnalova L, Biton M, Shekar K, Mou H, Eskiocak O, et al. 2021. Dietary suppression of MHC-II expression in intestinal stem cells enhances intestinal tumorigenesis. *Cell Stem Cell* **28:** 1922.e5–1935.e5. doi:10.1016/j.stem.2021.08.007

Kleeman SO, Ferrer M, Demestichas B, Bankier S, Lee H, Heywood T, Ruusalepp A, Bjorkegren JLM, Walker BR, Meyer HV, et al. 2021. Cystatin C is a glucocorticoid response gene predictive of cancer immunotherapy failure. medRxiv doi:10.1101/2021.08.17.21261668

THE GENETICS OF CANCER DEVELOPMENT AND PROGRESSION

J. Sheltzer V. Girish K. Long E. Sausville D. Shen P. Suri
A. Lakhani D. Lukow K. Schukken J. Smith 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.

Chromosomal Instability and Aneuploidy in Cancer

D. Lukow, A. Vasudevan, K. Schukken, A. Lakhani, V. Girish, P. Suri

Aneuploidy is a ubiquitous feature of human tumors, but the acquisition of aneuploidy is typically detrimental to cellular fitness. To investigate how aneuploidy could contribute to tumor growth, we triggered periods of chromosomal instability (CIN) in human cells and then exposed them to a variety of different culture environments. Although chromosomal instability was universally detrimental under normal growth conditions, we discovered that transient CIN reproducibly accelerated the ability of cells to adapt and thrive in the presence of anticancer therapeutic agents. Single-cell sequencing revealed that these drug-resistant populations recurrently developed specific whole-chromosome gains and losses. We independently derived one aneuploidy that was frequently recovered in cells exposed to paclitaxel, and we found that this chromosome loss event was sufficient to decrease paclitaxel sensitivity. Finally, we demonstrated that intrinsic levels of CIN correlate with poor responses to a variety of systemic therapies in a collection of patient-derived xenografts. In total, our results show that although chromosomal instability generally antagonizes cell fitness, it also provides phenotypic plasticity to cancer cells that can allow them to adapt to diverse stressful environments. Moreover, our findings suggest that aneuploidy may function as an underexplored cause of therapy failure in human tumors.

High levels of cancer aneuploidy are frequently associated with poor prognosis. To examine the relationship between aneuploidy and cancer progression, we analyzed a series of congenic cell lines that harbor single extra chromosomes. We found that across 13 different trisomic cell lines, 12 trisomies suppressed invasiveness or were largely neutral, whereas a single trisomy increased metastatic behavior by triggering a partial epithelial–mesenchymal transition. In contrast, we discovered that chromosomal instability activates cyclic GMP–AMP synthase-stimulator of interferon genes (cGAS-STING) signaling, but strongly suppresses invasiveness. By analyzing patient copy number data, we demonstrate that specific aneuploidies are associated with distinct outcomes, and the acquisition of certain aneuploidies is in fact linked to a favorable prognosis. Thus, aneuploidy is not a uniform driver of malignancy, and different aneuploidies can uniquely influence tumor progression. At the same time, the gain of a single chromosome is capable of inducing a profound cell state transition, thereby linking genomic plasticity, phenotypic plasticity, and metastasis.

PUBLICATIONS

Chen W, Zhou J, Sheltzer JM, Kinney JB, McCandlish DM. 2021. Non-parametric Bayesian density estimation for biological sequence space with applications to pre-mRNA splicing and the karyotypic diversity of human cancer. *Proc Natl Acad Sci* **18:** e2025782118. doi:10.1073/pnas.2025782118

Ghosh A, Girish V, Yuan ML, Coakley RD, Wrennal JA, Alexis NE, Sausville EL, Vasudevan A, Chait AR, Sheltzer JM, et al. 2021. Combustible and electronic cigarette exposures increase ACE2 activity and SARS-CoV-2 spike binding. *Am J Respir Crit Care Med* **205:** 129–133. doi:10.1164/rccm.202106-1377LE

Hoffmann M, Hofmann-Winkler H, Smith JC, Krüger N, Sørensen LK, Søgaard OS, Hasselstrøm JB, Winkler M, Hempel T, Raich L, et al. 2021. Camostat mesylate inhibits SARS-CoV-2 activation by TMPRSS2-related proteases and its metabolite GBPA exerts antiviral activity. *EBioMedicine* **65:** 103255.

Lukow DA, Sheltzer JM. 2021. Chromosomal instability and aneuploidy as causes of cancer drug resistance. *Trends Cancer* **8:** 43–53. doi:10.1016/j.trecan.2021.09.002

Lukow DA, Sausville EL, Suri P, Chunduri NK, Weiland A, Leu J, Smith JC, Girish V, Kumar AA, Kendall J, et al. 2021. Chromosomal instability accelerates the evolution of resistance to anti-cancer therapies. *Dev Cell* **56:** 2427–2439. doi:10.1016/j.devcel.2021.07.009

Schukken KM, Sheltzer JM. 2021. Extensive protein dosage compensation in aneuploid human cancers. bioRxiv doi:10.1101/2021.06.18.449005

Smith JC, Sheltzer JM. 2021. Genome-wide identification and analysis of prognostic features in human cancers. bioRxiv doi:10.1101/2021.06.01.446243

Vasudevan A, Schukken KM, Sausville EL, Girish V, Adebambo OA, Sheltzer JM. 2021. Aneuploidy as a promoter and suppressor of malignant growth. *Nat Rev Cancer* **21:** 89–103. doi:10.1038/s41568-020-00321-1

DECODING METABOLITES IN TUMOR MICROENVIRONMENT

L. Zhang A. Boyle W. Mei A. Thakkar
B. Li S. Mishra E. Wong
G. Mazzeo P. Pofali L. Zhu

The research in the Zhang laboratory centers on normal and malignant stem and progenitor cells in the hematopoietic system. Our research focuses on decoding the role of metabolites, including diets, nutrients and neurotransmitters, in the tumor microenvironment of hematologic malignancies. We utilize a combination of CRISPR-Cas9 functional genomics and metabolomics approaches to systematically uncover these metabolites and their genetic effectors, including receptors and metabolic enzymes, to reveal mechanistic insights into and identify key drug targets for hematologic malignancies.

Together, our laboratory recently uncovered a series of critical metabolites in the tumor microenvironment, including acetylcholine and pyridoxal, and their genetic effectors as novel regulators of hematologic malignancies. We identified cholinergic receptor muscarinic 4 (CHRM4) as a novel regulator of early erythroid progenitor differentiation and a therapeutic target for myelodysplastic syndrome (MDS). Our research uncovered the hematopoietic arc (HematopoArc) as a novel nervous system activity–based regulatory mechanism of hematopoietic stem and progenitor cell differentiation. We also identified the vitamin B6 pathway as a nutritional and metabolic dependency in acute myeloid leukemia (AML) that coordinates nucleotides and putrescine metabolism specifically required for leukemia maintenance. Our research uncovered the vitamin B6 pathway as a pharmacologically actionable target to treat leukemia with minimal bone marrow suppression effect.

Through collaborations with medicinal chemists and industry partners, we are building pharmacological approaches to target these novel regulators, and our findings will help to treat devastating hematologic malignancies, including refractory anemia, MDS, and leukemia. Our laboratory is currently working on an investigational new drug (IND)-enabling study to prepare the lead compounds of a first-in-class therapy for MDS—a lethal hematologic malignancy into clinical development, and our research will have a direct impact on disease treatment.

Decoding CHRM4 Pathway and Nervous System and Brain Activities in Regulating Hematopoiesis

Erythropoiesis is a multistage developmental process that results in erythrocyte production. Burst forming unit-erythroid (BFU-E) and colony forming unit-erythroid (CFU-E) are two progenitor cell types of erythroid lineage. BFU-E is an earlier progenitor capable of extensive differentiation, whereas the later progenitor CFU-E is only able to undergo approximately four to five divisions to generate approximately 30 erythrocytes. Clinically, bone marrow BFU-E cell levels predict responsiveness of MDS patients to erythropoietin (EPO) treatment. MDS patients with BFU-E numbers similar to normal individuals respond to EPO, whereas patients with inadequate BFU-E fail to respond. However, our understanding of molecular mechanisms underlying BFU-E differentiation and proliferation, especially druggable regulators and corresponding small chemical compounds targeting these regulators, is extremely limited, preventing us from targeting this process to treat MDS.

We have recently identified muscarinic acetylcholine receptor CHRM4 as a novel regulator of BFU-E differentiation and a druggable target to overcome BFU-E deficiency and EPO refractoriness in MDS through collaboration with Dr. Omar Abdel-Wahab's laboratory at Memorial Sloan Kettering Cancer Center. Together, these results demonstrated that targeting CHRM4 corrects BFU-E insufficiency and reduces abnormal endogenous serum EPO level in MDS, and we are translating these exciting findings into a clinical trial for MDS resistant to standard EPO, lenalidomide, and luspatercept treatments. As CHRM4 is a cell membrane receptor, it is crucial to understand how CHRM4 communicates with extracellular signals and is regulated by nervous system activity in the hematopoietic tissue, in both normal physiological and diseased conditions. We have characterized CHRM4 as a cell surface marker for early erythroid progenitor, and uncovered its regulation by nervous

system. Our research will delineate a novel regulatory mechanism modulated by the nervous system through collaboration with Dr. Bo Li's laboratory at CSHL.

Uncovering Micronutrient Addictions in Hematologic Malignancies

In comparison to macronutrient, the role of micronutrient in regulating hematologic malignancies is much less understood. The vitamin B6 family comprises six pyridine ring–containing compounds—pyridoxal, pyridoxamine, pyridoxine, and their phosphorylated forms correspondingly. Pyridoxal kinase (PDXK) phosphorylates the pyridoxal into its bioactive form pyridoxal 5′-phosphate (PLP), which acts as a cofactor of biochemical reactions. Epidemiological investigations demonstrated the abnormal plasma PLP levels in leukemia patients, implicating a dysregulated vitamin B6 metabolism in leukemia. Through collaboration with Dr. Scott Lowe's laboratory at Memorial Sloan Kettering Cancer Center, our research showed that genetic blockage of PDXK inhibited proliferation of mouse and human AML cells. Additionally, PDXK depletion resulted in the reduction of PLP levels in leukemia cells. Of note, PDXK depletion did not trigger leukemia cell differentiation but instead impaired cell cycle progression and increased apoptosis to exert both cytostatic and cytotoxic effects.

Pharmacological inhibition of vitamin B6 pathway exerts anti-leukemia activity with minimal bone marrow suppression effects. Consistent with pharmacological results, feeding mice with a vitamin B6–deficient diet also impaired the disease progression without exhibiting toxicity. We have further showed that pharmacological inhibition of the vitamin B6 pathway exerts no effects on normal hematopoiesis and is well tolerated. Altogether, pharmacological inhibition of the vitamin B6 pathway is well tolerated and exerts no effects on normal bone marrow hematopoietic stem and progenitor and immune cell growth. In addition, as isoniazid has been widely used as a therapeutic approach in tuberculosis patients clinically, isoniazid is well tolerated. Together, our results highlighted the therapeutic potential of the vitamin B6 pathway as an anti-leukemia target with minimal bone marrow suppression effects.

Developing Novel Therapeutics for Myelodysplastic Syndrome

MDS is a lethal hematopoietic malignancy with very limited therapeutic options (Fig. 1). Only a small proportion of MDS patients benefit from standard EPO treatment, and recently approved compound lenalidomide only targets a limited portion of MDS patients who carry a chromosome 5q deletion. Luspatercept targets the very terminal stage of erythropoiesis and

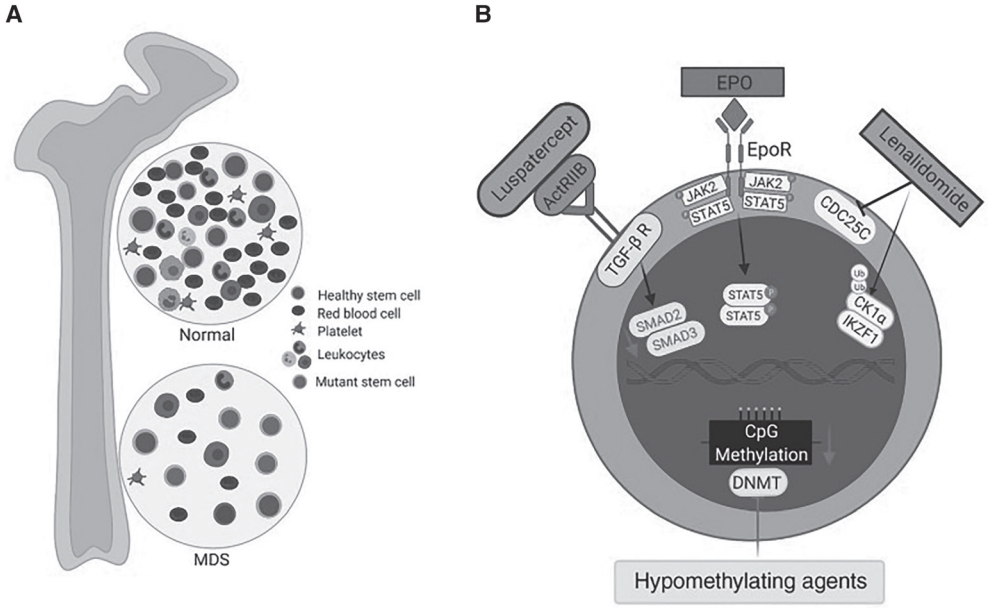


Figure 1. Current therapeutics for myelodysplastic syndrome (MDS). (*A*) Cytopenia including anemia is a major challenging for MDS treatment. (*B*) Current therapeutic options for MDS are limited. (From Trivedi et al. 2021.)

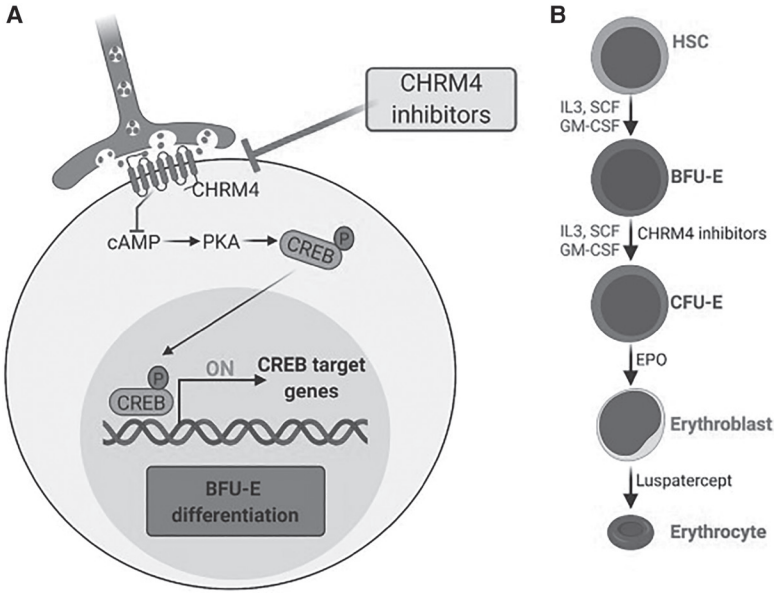


Figure 2. Targeting the CHRM4 pathway represents a novel therapeutic strategy for myelodysplastic syndrome (MDS). (*A*) The CHRM4 pathway regulates early erythroid progenitor differentiation. (*B*) CHRM4 inhibitors target early erythroid progenitor stage. (From Trivedi et al. 2021.)

exhibits limited clinical efficacy, particularly for MDS patients with high serum EPO levels, in whom the reagent achieves transfusion independence in limited populations. In addition, many of the initial responders do not have long-term response. Currently, the only option for patients who do not respond to EPO and lenalidomide is red blood cell transfusion, which exposes patients to insufficient correction of anemia, alloimmunization, and organ failure secondary to iron overload.

Our research addresses this challenge by specifically targeting early erythroid progenitors whose insufficiency contributes to refractoriness of MDS patients to current therapies. We have demonstrated that CHRM4 selective antagonists completely corrected anemia of MDS through correcting the early erythroid progenitor insufficiency. We have also showed that CHRM4 antagonists lowered the abnormally elevated plasma EPO level, a clinical predictive marker of refractoriness of MDS patients to EPO treatment (Fig. 2). Together with industry partners and Dr. Yousef Al-Abed's laboratory at Northwell Health, we have initiated a program to develop CHRM selective antagonists to treat MDS. The ultimate goal of our research is to benefit MDS patients and cancer patients who are undergoing chemotherapy and radiation therapy to become anemic.

In summary, recent evidence suggested that diets and nutrients, as well as neurotransmitters, are two classes of critical but underinvestigated metabolites in the tumor microenvironment. In the hematopoietic system, these metabolites regulate hematologic cancers through interaction with their genetic effectors expressed in hematopoietic stem and progenitor cells, the function of which are even far less understood. Our research aiming to identify both metabolites in the tumor microenvironment and their genetic effectors in hematopoietic stem and progenitor cells will provide mechanistic insights toward a systematic understanding of metabolites in the tumor microenvironment of hematologic cancers. Clinically, our identified druggable targets of the genetic effectors of these metabolites in tumor microenvironment will provide a basis to develop pharmacological approaches to therapeutically target cancers. Through collaborations with industry partners, we will move these discoveries into preclinical/clinical development. Together, our proposed research will have a significant impact on both basic understanding of the biology of metabolites and tumor microenvironment and a direct influence on the treatment of hematologic cancers.

PUBLICATION

Trivedi G, Inoue D, Zhang L. 2021. Targeting low-risk myelodysplastic syndrome with novel therapeutic strategies. *Trends Mol Med* **27:** 990–999.

AUTHOR INDEX

SCHOOL OF BIOLOGICAL SCIENCES

SCHOOL OF BIOLOGICAL SCIENCES

ADMINISTRATION

Alexander Gann, Ph.D., Professor and Dean
Monn Monn Myat, Ph.D., Associate Dean
Alyson Kass-Eisler, Ph.D., Director of Academic Programs and Registrar
Kimberly Creteur, M.Ed., M.S.Ed., Admissions and Recruitment Director
Kimberley Graham, Administrative Assistant

EXECUTIVE COMMITTEE

Chair
Alexander Gann

Members
Arkarup Banerjee
Camila dos Santos
Jesse Gillis
John Inglis
David L. Spector (Director of Research)
Lloyd Trotman

Student Representatives
Alexa Pagliaro, SBS
Danilo Segovia, SBU

Secretary
Monn Monn Myat

ADMISSIONS COMMITTEE

Chair
Adrian Krainer

Members
Alexander Gann (*ex officio*)
Mikala Egeblad
Leemor Joshua-Tor
Bo Li
Zachary Lippman
David McCandlish
W. Richard McCombie
Stephen Shea
Christopher Vakoc
Linda Van Aelst

Secretary
Kimberly Creteur

QUALIFYING EXAM COMMITTEE

Chair
Linda Van Aelst

Members
Molly Hammell
Zachary Lippman
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

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

SCHOOL OF BIOLOGICAL SCIENCES DEAN'S REPORT

Despite the ongoing COVID-19 pandemic, School activities have been taking place normally, with only minimal disruptions as described below. A new class of students arrived on campus in fall 2021 in time to start the semester, and all classes were held in person. It has been a challenging year in many regards, but we are grateful that the students, faculty, and support from CSHL have allowed our program to succeed and grow despite these obstacles.

Entering Class of 2021

On August 16, 2021, the School welcomed the 23rd incoming class, which comprised eight new students: Hoda Ansari, Paul Bunk, Nikolas Holland, Sessen Daniel Iohannes, Emmanuella (Ella) Nnuji-John, Rachel Polfer, Yihan (Leonie) Qin, and Lucía Téllez Pérez.

ENTERING CLASS OF 2021

Hoda Ansari, Sarah Lawrence College: B.A. in Psychology (2016); University College London: M.Sc. in Cognitive Neuroscience (2017)
Academic Mentor: Semir Beyaz

Paul Bunk, University College London: B.Sc. in Biomedical Sciences (2021)
Academic Mentor: Christopher Vakoc

Nikolas Holland, New York University: B.Sc. in Neural Science (2020)
Academic Mentor: Jeremy Borniger

Sessen Daniel Iohannes, Università di Pisa: M.Sc. in Molecular Biotechnology/B.S. in Biological Sciences (2020/2018); Scuola Superiore Sant'Anna: M.Sc. Honours Degree in Molecular Biotechnology (2021)
Academic Mentor: Jessica Tollkuhn

Emmanuella (Ella) Nnuji-John, Alcorn State University: B.S. in Biochemistry (2021)
Academic Mentor: Molly Gale Hammell

Rachel Polfer, Mount Holyoke College: B.A. in Biochemistry and Philosophy (2018); University of Cambridge: M.Phil. in Technology Policy (2019)
Academic Mentor: David Jackson

Yihan (Leonie) Qin, The College of William & Mary: B.S. in Biology and Chemistry (2021)
Academic Mentor: Hiro Furukawa

Lucía Téllez Pérez, Universidad Autonoma de Madrid: B.S. in Biochemistry (2020); University of Edinburgh: M.Sc. in Bioinformatics (2021)
Academic Mentor: John Inglis

2021 Entering Class: (*From left to right*) Paul Bunk, Nikolas Holland, Leonie Qin, Lucía Téllez Pérez, Hoda Ansari, Sessen Daniel Iohannes, Ella Nnuji-John, and Rachel Polfer.

Graduation

This year's graduation ceremony, originally pushed back to August 22, 2021, was canceled primarily because of Hurricane Henri, which was due to hit Long Island that weekend. However, we still celebrated the achievements of the 18th graduating class, some of whom were able to take photos in their caps and gowns with the Dean. Nine students were awarded Ph.D. degrees this year: Benjamin Berube, Matt Lee, Katarina Meze, Alexandra Nowlan, Sofya Polyanskaya, and Ngoc (Tumi) Tran from the Entering Class of 2015; Brianna Bibel and Shaina Lu from the Entering Class of 2016; and Benjamin Harris from the Entering Class of 2017.

Dean Alex Gann with graduates.

2021 DOCTORAL RECIPIENTS

Student	Thesis advisor	Academic mentor	Current position
Benjamin Berube	Robert Martienssen	Zachary Lippman	Graduate Student Post, CSHL
Brianna Bibel	Leemor Joshua-Tor	Hiro Furukawa	Postdoctoral Fellow, UCSF (Advisor: Danica Fujimori)
Benjamin Harris	Jesse Gillis	W. Richard McCombie	Computational Biologist, Lyell Immunopharma, CA
Matt Lee	Lloyd Trotman	Nicholas Tonks	Clinical Science Liaison, TG Therapeutics
Shaina Lu	Jesse Gillis and Anthony Zador	Leemor Joshua-Tor	Bioinformatics Scientist, Eagle Genomics
Katarina Meze	Leemor Joshua-Tor	Je Hyuk Lee	Postdoctoral Fellow, University of Michigan (Advisor: Nils Walters)
Alexandra Nowlan	Stephen Shea	Jessica Tollkuhn	Postdoctoral Fellow, UNC Chapel Hill (Advisor: Zoe McElligott)
Sofya Polyanskaya	Christopher Vakoc	Alexander Krasnitz	Consultant, Scitaris, Berlin, Germany
Ngoc (Tumi) Tran	Alexei Koulakov	Leemor Joshua-Tor	Machine Learning Scientist, Relay Therapeutics, MA

2021 THESIS DISSERTATION DEFENSES

ENTERING CLASS OF 2015

Matt Lee, February 8, 2021

Endocytic regulation of PTEN

Thesis Examining Committee

Chair:	**Christopher Vakoc**
Research Mentor:	**Lloyd Trotman**
Academic Mentor:	**Nicholas Tonks**
Committee Member:	**David Spector**
External Examiner:	**Ramon Parsons**
	Icahn School of Medicine at Mount Sinai

Alexandra Nowlan, April 28, 2021

An amygdalo-cortical circuit for multisensory processing in maternal behavior

Thesis Examining Committee

Chair:	**Bo Li**
Research Mentor:	**Stephen Shea**
Academic Mentor:	**Jessica Tollkuhn**
Committee Member:	**Anthony Zador**
External Examiner:	**Dayu Lin**
	NYU Grossman School of Medicine

Ngoc Tumi Tran, May 6, 2021

DeepNose: Using artificial neural networks to represent the space of odorants

Thesis Examining Committee

Chair:	**Pavel Osten**
Research Mentor:	**Alexei Koulakov**
Academic Mentor:	**Leemor Joshua-Tor**
Committee Member:	**Justin Kinney**
External Examiner:	**Dmitry Rinberg**
	New York University

Sofya Polyanskaya, June 15, 2021

A nuclear phosphatase-kinase signaling complex that supports acute myeloid leukemia

Thesis Examining Committee

Chair:	**Lloyd Trotman**
Research Mentor:	**Christopher Vakoc**
Academic Mentor:	**Alexander Krasnitz**
Committee Member:	**Nicholas Tonks**
Committee Member:	**David Tuveson**
External Examiner:	**Alex Kentsis**
	Memorial Sloan Kettering Cancer Center

Katarina Meze, July 9, 2021

Mechanism of structured RNA degradation by the human exonuclease Dis3L2

Thesis Examining Committee

Chair:	**Christopher Hammell**
Research Mentor:	**Leemor Joshua-Tor**
Academic Mentor:	**Je Hyuk Lee**
Committee Member:	**Hiro Furukawa**
Committee Member:	**Adrian Krainer**
External Examiner:	**Daniel Herschlag**
	Stanford University

Benjamin Berube, August 16, 2021

Epigenetic control of meiotic recombination and drive

Thesis Examining Committee

Chair:	**Ullas Pedmale**
Research Mentor:	**Robert Martienssen**
Academic Mentor:	**Zachary Lippman**
Committee Member:	**David Jackson**
External Examiner:	**James A. Birchler**
	University of Missouri

ENTERING CLASS OF 2016

Shaina Lu, September 23, 2021

The replicability of spatially resolved transcriptomics for modern neuroscience

Thesis Examining Committee

Chair:	**Molly Gale Hammell**
Co-Research Mentor:	**Jesse Gillis**
Co-Research Mentor:	**Anthony Zador**
Academic Mentor:	**Leemor Joshua-Tor**
Committee Member:	**Jessica Tollkuhn**
Committee Member:	**Dimitri Rinberg** (*NYU*)
External Examiner:	**Adam Hantman**
	HHMI Janelia Research Campus

Brianna Bibel, October 5, 2021

Regulating the regulator: phosphorylation-mediated regulation of the RNAi effector protein Argonaute

Thesis Examining Committee

Chair:	**Christopher Hammell**
Research Mentor:	**Leemor Joshua-Tor**
Academic Mentor:	**Hiro Furukawa**
Committee Member:	**Adrian Krainer**
Committee Member:	**Nicholas Tonks**
External Examiner:	**Rachel Green**
	Johns Hopkins University

ENTERING CLASS OF 2017

Benjamin Harris, October 12, 2021

Atlas level scRNAseq analysis reveals the functional landscape of cell types

Thesis Examining Committee

Chair:	**David McCandlish**
Research Mentor:	**Jesse Gillis**
Academic Mentor:	**W. Richard McCombie**
Committee Member:	**David Jackson**
External Examiner:	**Ziv Bar-Joseph**
	Carnegie Mellon

DOCTORAL THESIS RESEARCH

Student	Academic mentor	Research mentor	Thesis research
ENTERING CLASS OF 2016			
Alberto Corona *NIH Predoctoral Trainee* *Hearst Foundation Fellow*	David Jackson	Stephen Shea	Identification of neural circuitry underlying paternal behaviors
David Johnson *Gilliam Fellow* *National Science Foundation Fellow* *Hearst Foundation Scholar*	Zachary Lippman	Alea Mills	Elucidating the role of BRPF1 in human glioblastoma multiforme
Christopher Krasniak *NIH Predoctoral Trainee*	Jan Witkowski	Anthony Zador	The role of cholinergic input to visual cortex in mouse spatial visual attention
Kathryn O'Neill *National Science Foundation Fellow* *NIH Predoctoral Trainee* *Hearst Foundation Fellow*	Camila dos Santos	Molly Hammell	Investigations into TDP-43-mediated effects on sRNA biology
Luqun Shen *Edward & Martha Gerry Fellow*	David Stewart	Stephen Shea	Neural mechanisms of vocal discrimination in maternally experienced mice
Martyna Sroka *George and Marjorie Anderson Fellow*	Molly Hammell	Christopher Vakoc	Molecular dissection of the PAX3-FOXO1 fusion oncoprotein pathway in rhabdomyosarcoma
Ran Yan *George and Marjorie Anderson Fellow*	David Tuveson	Douglas Fearon	Identification of endogenous antigen-specific T cells in pancreatic cancer metastasis
Chengxiang (Charlie) Yuan *A*STAR Fellow*	Nicholas Tonks	Jay Lee	Linking the cell cycle and developmental fate specification
ENTERING CLASS OF 2017			
Lyndsey Aguirre *National Science Foundation Fellow* *Hearst Foundation Fellow*	Ullas Pedmale	Zachary Lippman	Decoding *cis*-regulatory control of quantitative trait variation in tomato
Sara Boyle *NIH Predoctoral Trainee*	Jessica Tollkuhn	Bo Li	Can the central amygdala's interaction with midbrain dopamine areas control motivated behavior?
Jordan (Bruno) Gegenhuber *Charles A. Dana Fellow* *John & Amy Phelan Fellow* *NIH Predoctoral Trainee* *NRSA Predoctoral Fellow*	John Inglis	Jessica Tollkuhn	Gene regulatory mechanisms underlying brain organization by perinatal estradiol
Yuzhao (Richard) Hu *George and Marjorie Anderson Fellow*	Justin Kinney	Ullas Pedmale	Characterizing cryptochrome 2 downstream targets in *Arabidopsis*
Dennis Maharjan *John & Amy Phelan Fellow*	Florin Albeanu	Anthony Zador	Role of corticostriatal potentiation in auditory decision-making
Diogo Maia e Silva *Robert and Theresa Lindsay Fellow* *Boehringer Ingelheim Fonds Fellow*	Bruce Stillman	Christopher Vakoc	Defining the molecular origins of squamous pancreatic ductal adenocarcinoma
Cole Wunderlich *NIH Predoctoral Trainee*	Adam Siepel	Molly Hammell	Exploring transposable element expression in ALS at single-cell resolution
ENTERING CLASS OF 2018			
King Hei (Teri) Cheng *Robert and Theresa Lindsay Fellow*	Adrian Krainer	Robert Martienssen	Investigating the molecular mechanism of RNAi in resolving transcription–replication conflicts
Danielle Ciren *NSERC Scholar* *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

(*continued*)

DOCTORAL THESIS RESEARCH (*continued*)

Student	Academic mentor	Research mentor	Thesis research
Marie Dussauze *Florence Gould Fellow* *Annette Kade Fellow*	Stephen Shea	Florin Albeanu	Understanding algorithms and underlying neuronal substrates of sensorimotor integration in closed-loop olfaction
Ilgin Ergin *Starr Centennial Scholar*	Thomas Gingeras	Semir Beyaz	Mechanistic dissection of dietary regulation of T-cell function
Connor Fitzpatrick *Ainslie Family Fellow*	Robert Martienssen	Christopher Vakoc	Elucidating the mechanism of transcriptional activation by achaete-scute homolog 1
Amritha Varshini Hanasoge Somasundara *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 *Starr Centennial Scholar*	David Tuveson	Jason Sheltzer	Dissecting the role of recurrent hyperploidy in tumorigenesis
Ziyi Mo *Gladys and Roland Harriman Foundation Fellow*	David McCandlish	Adam Siepel	A flexible deep learning framework for inferring parameters of selection based on the ancestral recombination graph
Alexa Pagliaro *Jordan and Thomas A. Saunders III Neuroscience Fellow* *NIH Predoctoral Trainee*	John Inglis	Stephen Shea	Parvalbumin-positive interneuron activity during maternal behavior in a mouse model of Rett syndrome
Jenelys Ruiz *NIH Predoctoral Trainee* *Hearst Foundation Fellow*	Molly Gale Hammell	Camila dos Santos	Developing models to study cellular composition changes across pregnancy cycles in the human mammary gland
Jonathan Werner *National Science Foundation Fellow*	Adam Siepel	Jesse Gillis	Revealing cellular lineage and sex-specific gene expression through transcriptional analysis of the X-chromosome
ENTERING CLASS OF 2019			
Leah Braviner *Elisabeth Sloan Livingston Fellow*	Linda Van Aelst	Leemor Joshua-Tor	Structural and biochemical characterization of CSR-1 and C04F12.1, two closely related WAGO clade Argonautes in *Caenorhabditis elegans*
Patrick Cunniff *NIH Predoctoral Trainee*	David Jackson	Christopher Vakoc	Targeting of Krüppel-like factor 5 (KLF5) in pancreatic ductal adenocarcinoma
Michael Passalacqua *Hearst Foundation Scholar*	Thomas Gingeras	Jesse Gillis	Using plant co-expression to understand gene sub-functionalization and neo-functionalization
Leonardo Jared Ramirez Sanchez *Gonzalo Rio Arronte Fellow*	Christopher Hammell	Bo Li	Understanding the neural representations of positive and negative reinforcement in the striosome compartment
Nicole Sivetz *NIH Predoctoral Trainee*	Camila dos Santos	Mikala Egeblad	Investigating stromal interactions and tissue reprogramming during pancreatitis and PDAC progression
Ziqi (Amber) Tang *Charles A. Dana Fellow*	Alea Mills	Peter Koo	Uncovering *cis*-regulatory codes in alternative splicing with deep neural network
Shushan Toneyan *Crick-Clay Fellow*	David Stewart	Peter Koo	Modeling gene regulation using an integrative and interpretable neural network
Julia Wang *Jenny and Jeff Kelter Neuroscience Fellow*	Saket Navlakha	Tatiana Engel	Multidimensional brain states and their role in meta-learning and continual learning

(*continued*)

DOCTORAL THESIS RESEARCH (*continued*)

Student	Academic mentor	Research mentor	Thesis research
ENTERING CLASS OF 2020			
Salomé Carcy *Starr Centennial Scholar*	John Inglis	Hannah Meyer	Development of invariant natural killer T cells in the murine and human thymus
Jed de Ruiter-Swain *Leslie C. Quick, Jr. Fellow*	Tobias Janowitz	Michael Lukey	Astrocyte–cancer cell metabolic coupling as a novel therapeutic target in brain-metastatic breast cancer
Iacopo Gentile *George and Marjorie Anderson Fellow*	Adam Siepel	Zachary Lippman	Genetic and mechanistic study of pollen polysiphony in tetraploid *Physalis grisea*
Emily Isko *NIH Predoctoral Trainee*	Stephen Shea	Arkarup Banerjee and Anthony Zador	Evolving neural circuits underlying vocal behaviors in closely related murine species
Jessica Kahng *NIH Predoctoral Trainee*	Rebecca Leshan	Lucas Cheadle	A structural and molecular characterization of sensory-induced oligodendrocyte precursors cell (OPC) engulfment of thalamocortical synapses in the visual cortex
Sam Kleeman *David H. Koch Fellow*	Adrian Krainer	Tobias Janowitz	Molecular mechanisms of anti-NMDA receptor autoantibodies
Matty Peacey *George and Marjorie Anderson Fellow*	Hiro Furukawa	Andrea Schorn	3′ tRNA fragment-mediated silencing of active and domesticated LTR-retrotransposons
Xiaoyue (Mike) Zheng *Goldberg-Lindsay Fellow*	Saket Navlakha	Arkarup Banerjee	Neural circuits of context-specific vocalizations in rodents
Xingyu (CiCi) Zheng *William R. Miller Fellow*	Justin Kinney	Saket Navlakha	Statistical analysis and modeling of spatial networks in plants

Teaching Award

Dr. Jason Sheltzer, the co-instructor of the Scientific Exposition and Ethics course, was chosen by the first-year students for the 16th Winship Herr Award for Excellence in Teaching. Here is what one student said in their nomination: "Jason's lectures were the most clearly explained, and also going to the point, allowing us to easily identify the take-home messages … he was always finishing on time or a little early, which allowed to have some discussion time."

Faculty Changes

There were no new faculty members appointed in 2021, but two faculty members, Jesse Gillis and Jason Sheltzer, left the School. Jesse moved his laboratory to the University of Toronto and Jason moved his laboratory to Yale University.

Academic Mentoring

The 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.

Recruiting Efforts

Although there was no recruitment travel this year, we did attend and host a number of virtual events. We received 683 applications for the Entering Class of 2022, the highest number to date, and it appears that many outstanding candidates have once again applied to the program. Additionally, emails 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 Laboratory. We are grateful to these departments for sharing this contact list.

The School hosted its second Diversity Recruitment Events on September 17 and 18, 2021, at which approximately 50 undergraduate, post-baccalaureate, and graduate students belonging to minority groups underrepresented in science from across the nation attended via Zoom. They learned of CSHL's Undergraduate Research Program, the Ph.D. graduate program and the postdoctoral program from the School's Dean and staff, faculty, and administrators. They also heard faculty research talks, personal stories of our current students' scientific journeys, and attended a workshop on applying to graduate school. The recruitment event was a success as indicated by a postevent survey.

SCHOOL OF BIOLOGICAL SCIENCES 2021 RECRUITMENT SCHEDULE

Event	Location	Date
Society for Neuroscience Global Connectome Graduate School Fair	Virtual Forum	January 11–13
University of Maryland McNair Scholars Research Conference Graduate School Fair	Virtual Conference	March 10–12
Wisconsin Association of Educational Opportunity Program Personnel (WAEOPP) McNair and SSS Graduate School Retreat Graduate School Fair	Virtual Forum	April 9
National Conference on Undergraduate Research (NCUR) Annual Conference	Virtual Conference	April 12–14 April 26
Southeastern Association of Educational Opportunity Program Personnel (SAEOPP) McNair/SSS Scholars Research Conference Annual Conference	Virtual Conference	June 23–25
NIH Graduate & Professional School Fair Annual Meeting	Virtual Forum	July 19–22
University of California, Los Angeles National McNair Conference Graduate School Fair	Virtual Conference	July 27–30
Baylor University McNair Research Conference Graduate School Fair	Virtual Forum	August 20
CSHL Diversity Recruitment Events Open House	Virtual Forum	September 17–18
Rose-Hulman Institute of Technology Graduate School Fair	Virtual Forum	September 21
WAEOPP McNair/SSS Graduate School Retreat Graduate School Fair	Virtual Forum	September 24–25
MKN McNair Heartland Research Conference Graduate School Fair	Virtual Forum	September 24–26
McNair and other Underrepresented populations (Tribal Colleges, Hispanic-Serving Institutions, and HBCUs) Graduate School/Summer Research Fair	Virtual Forum	October 1
Atlanta University Center Consortium: Clark Atlanta University, Morehouse College, and Spelman College Graduate & Professional School Fair	Virtual Forum	October 20
Hunter College MARC and RISE/MBRS Information Session	Virtual Presentation	October 20
National Association of African American Honors Programs (NAAAHP) Conference Graduate School Fair	Virtual Conference	October 20–24
University of Maryland, Baltimore County Meyerhoff Scholars Program, Campus Connection Information Session	Virtual Presentation	October 22
California Forum for Diversity in Graduate Education Graduate School Fair	Virtual Forum	October 22–23
Louis Stokes Midwest Regional Center for Excellence Resource Fair	Virtual Conference	October 22–24
Society for Advancement of Chicanos and Native Americans in Science (SACNAS) Conference National Conference	Virtual Conference	October 25–29
University College London, Life Sciences Society Information Session	Virtual Presentation	October 26
Alcorn State University Information Session	Virtual Presentation	November 2
Sigma Xi Annual Meeting and Student Research Conference Graduate School Fair	Virtual Conference	November 4–7
Big 10+ Graduate School Expo/Midwest Graduate School Summit Graduate School Fair	Virtual Forum	November 7–8

(*continued*)

SCHOOL OF BIOLOGICAL SCIENCES 2021 RECRUITMENT SCHEDULE (*continued*)

Event	Location	Date
Society for Neuroscience Annual Meeting Graduate School Fair	Virtual Conference	November 8–11
Annual Biomedical Research Conference for Minority Students (ABRCMS) National Conference	Virtual Conference	November 10–13
The State University of New York, College at Old Westbury Women in Science & Engineering and Equality for All Information Session	Virtual Presentation	November 15
Hunter College, QuBi Student Group Information Session	Virtual Presentation	November 17

Students from Other Institutions

Students enrolled in the School of Biological Sciences account for approximately half of the total graduate student population here at CSHL; the other half comprises 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 these students is 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 a visiting student from the Netherlands. The 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 students listed in the box below joined us from SBU this year.

STUDENT	CSHL RESEARCH MENTOR	STONY BROOK UNIVERSITY PROGRAM
Sergey Alekseev	Alexei Koulakov	Physics
Santiago Espinosa	Christopher Vakoc	Genetics
Seung Tea Kim	Lloyd Trotman	Molecular and Cell Biology
Alexander Kral	Adrian Krainer	Microbiology and Immunology
Steven Lewis	Camila dos Santos	Genetics
Jordan Pearson	Douglas Fearon	Genetics
Lucia Yang	Adrian Krainer	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. Typically, the Symposium consists of senior students giving short talks, while coffee breaks and lunch provide opportunities for more informal interactions. In October 2021 we modified the format to accommodate a virtual program and had the students do 5-minute "lightning talks." This new format is designed to help the students develop skills for distilling the salient points of their research and is common at digital conferences.

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 2021:

POSTDOCTORAL FELLOWS

Xu An	Dhananjay Huilgol	Sung Rye Park	Xinyi Wang
Oliver Artz	Stellar Hur	Klaske Schukken	Zhikai Wang

Matthias Benoit
Stephanie Yvonn Blockhuys
Hanfei Deng
Sarah Fumagalli
Olivier Gschwend
Longwen Huang
Yuhan Huang
Alessandra Inguscio
Gukhan Kim
Yi Li
Lei Liu
Grinu Mathew
Haiwei Mou
Farhana Nasrin
Donglai Shen
Atsushi Shimada
James Sturgill
Michael Tramantano
Bo Wang
Minghui Wang
Xingang Wang
Xiong Xiao
Yunyao Xie
Zhaolin Yang
Juannan Zhou

GRADUATE STUDENTS

Benjamin Berube
Brianna Bibel
Kung-Chi Chang
Hsiang-Chen Chou
Eva Carlotta Gablenz
Nicholas Gallo
Benjamin Harris
Matthew Lee
Shaina Lu
Devon Lukow
Katarina Meze
Alexandra Nowlan
Sofya Polyanskaya
Ngoc Tran

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 Arkarup Banerjee, Camila dos Santos, Jesse Gillis, John Inglis, David Spector, and Lloyd Trotman for their service in 2021. I would also like to thank the student representatives Alexa Pagliaro and Danilo Segovia (SBU), who contributed to discussions and provided useful suggestions and feedback from their colleagues.

Student and Alumni Achievements

To date, 127 students have received their Ph.D. degree from the School. Thirty-nine of our graduates have secured tenure track faculty positions (though three have now left these positions for industry), and two others are in independent research positions within academia. Nine have been promoted to Associate Professor (often conferring tenure) and six are full Professors. Our graduates have also moved into influential positions in administration, publishing, consulting, science communication, and industry. In 2021, Elena Ezkhova and Elizabeth Murchison were promoted to full Professor; Santanu Chakraborty became a Program Director for Sustainable Life Sciences and an Associate Professor at Atria University in India; Jeremy Wilsuz moved his lab to Baylor College of Medicine, where he is an Associate Professor; Zhenxun Wang became an Adjunct Junior Principal Investigator at the Singapore Eye Research Institute and an Adjunct Assistant Professor at the Duke-NUS Medical School in Singapore; Felix Schlesinger joined Scale Bio as a Bioinformatics Lead, Megan (Bodnar) Hogan became a Senior Scientist at Neochromosome; Marek Kudla is a Software Developer at Design Everest; Robert Aboukhalil became a Senior Software Engineer at the Chan Zuckerberg Initiative; Jack Walleshauser is a Senior Scientist at Revolution Medicines; Elizabeth Hutton joined ArtisanBio as a Computational Biologist; Colleen Carlston graduated from medical school and is now a Clinical Fellow in Pediatrics at Boston Children's Hospital; Brittany Cazakoff graduated from law school and is now an Associate Lawyer at Cooley; Lital Chartarifsky is an Associate Medical Director at FCB Cure, while Emilis Bružas is a Medical Director at FCB Health; Michael Gutbrod joined SANTÉ as an Associate.

In 2021, our current students were successful in receiving the following prestigious awards and fellowships:

- **Sessen Daniel Iohannes** was awarded a P.E.O. International Peace Scholarship.
- **Lucía Telléz Peréz** was awarded a Fulbright Fellowship.

2021 STUDENT (CURRENT OR PREVIOUS) PUBLICATIONS

Beyaz S, Chung C, Mou H, Bauer-Rowe KE, Xifaras ME, **Ergin I**, Dohnalova L, Biton M, Shekhar K, Eskiocak O, et al. 2021. Dietary suppression of MHC class II expression in intestinal epithelial cells enhances intestinal tumorigenesis. *Cell Stem Cell* **28:** 1922–1935.

Fischer S, Crow M, **Harris BD**, Gillis J. 2021. Scaling up reproducible research for single-cell transcriptomics using MetaNeighbor. *Nat Protoc* **16:** 4031–4067.

Ghosh S, Zador AM. 2021. Corticostriatal plasticity established by initial learning persists after behavioral reversal. *eNeuro* **8**: 1–10.

Hanasoge Somasundara AV, Moss MA, Feigman MJ, Chen C, Cyrill SL, Ciccone MF, Trousdell MC, Vollbrecht M, Li S, Kendall J, et al. 2021. Parity-induced changes to mammary epithelial cells control NKT cell expansion and mammary oncogenesis. *Cell Rep* **37:** 110099.

Hulse-Kemp AM, Bostan H, Chen S, Ashrafi H, Stoffel K, Sanseverino W, Li L, Cheng S, Schatz MC, **Garvin T**, et al. 2021. An anchored chromosome-scale genome assembly of spinach improves annotation and reveals extensive gene rearrangements in euasterids. *Plant Genome* **14:** e20101.

Hutton ER, Vakoc CR, Siepel A. 2021. ACE: a probabilistic model for characterizing gene-level essentiality in CRISPR screens. *Genome Biol* **22:** 278.

International Brain Laboratory, Aguillon-Rodriguez V, Angelaki D, Bayer H, Bonacchi N, Carandini M, Cazettes F, Chapuis G, Churchland AK, … , **Krasniak C**, et al. 2021. Standardized and reproducible measurement of decision-making in mice. *Elife* **10:** e63711.

Janowitz T, **Kleeman S**, Vonderheide RH. 2021. Reconsidering dexamethasone for antiemesis when combining chemotherapy and immunotherapy. *Oncologist* **26:** 269–273.

Krause WC, Rodriguez R, **Gegenhuber B**, Matharu N, Rodriguez AN, Padilla-Roger AM, Toma K, Herber CB, Correa SM, Duan X, et al. 2021. Oestrogen engages brain MC4R signalling to drive physical activity in female mice. *Nature* **599:** 131–135.

Lu S, Fürth D, Gillis J. 2021. Integrative analysis methods for spatial transcriptomics. *Nat Methods* **18:** 1282–1283.

Lu S, Ortiz C, Fürth D, Fischer S, Meletis K, Zador A, Gillis J. 2021. Assessing the replicability of spatial gene expression using atlas data from the adult mouse brain. *PLoS Biol* **19:** e3001341.

Maiorino L, Daßler-Plenker J, Sun L, Egeblad M. 2021. Innate immunity and cancer pathophysiology. *Annu Rev Pathol* **7:** 425–457.

Nattestad M, **Aboukhalil R**, Chin CS, Schat M.C. 2021. Ribbon: intuitive visualization for complex genomic variation. *Bioinformatics* **37:** 413–415.

Pisupati S*, **Chartarifsky-Lynn L***, Khanal A, Churchland AK. 2021. Lapses in perceptual decisions reflect exploration. *Elife* **10:** e55490.

Rosado D, Ackermann A, **Spassibojko O**, Rossi M, Pedmale UV. 2021. WRKY transcription factors and ethylene signaling modify root growth during the shade avoidance response. *Plant Physiol* **188:** 1294–1311.

Safaric Tepes P, Pal D, Lindsted T, Ibarra I, Lujambio A, Jimenez Sabinina V, Senturk S, Miller M, Korimerla N, **Huang J**, et al. 2021. An epigenetic switch regulates the ontogeny of AXL-positive/EGFR-TKi-resistant cells by modulating miR-335 expression. *Elife* **10:** e66109.

Shen Y, **Wang J**, Navlakha S. 2021. A correspondence between normalization strategies in artificial and biological neural networks. *Neural Comput* **30:** 1–25.

Shuvaev SA, **Tran NB**, Stephenson-Jones M, Li B, Koulakov AA. 2021. Neural networks with motivation. *Front Syst Neurosci* **14:** 609316.

Sroka MW, Vakoc CR. 2021. An epigenetic tipping point in cancer comes under the microscope. *Nature* **590:** 399–400.

Sun YC, Chen X, Fischer S, **Lu S**, Zhan H, Gillis J, Zador AM. 2021. Integrating barcoded neuroanatomy with spatial transcriptional profiling enables identification of gene correlates of projections. *Nat Neurosci* **24:** 873–885.

Wang X, **Aguirre L**, Rodríguez-Leal D, Hendelman A, Benoit M, Lippman ZB. 2021. Dissecting *cis*-regulatory control of quantitative trait variation in a plant stem cell circuit. *Nat Plants* **7**: 419–427.

Wei Y, Huang YH, Skopelitis DS, Iyer SV, Costa ASH, Yang Z, Kramer M, Adelman ER, Klingbeil O, … , **Polyanskaya SA**, et al. 2021. SLC5A3-dependent myo-inositol auxotrophy in acute myeloid leukemia. *Cancer Discov* **12:** 450–467.

Xu X, Crow M, Rice BR, Li F, **Harris B**, Liu L, Demesa-Arevalo E, Lu Z, Wang L, Fox N, et al. 2021. Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev Cell* **56:** 557–568.

Yang Z, Wu XS, Wei Y, **Polyanskaya SA**, Iyer SV, Jung M, Lach FP, Adelman ER, Klingbeil O, Milazzo JP, et al. 2021. Transcriptional silencing of ALDH2 confers a dependency on Fanconi anemia proteins in acute myeloid leukemia. *Cancer Discov* **11:** 2300–2315.

Zhao Y, Dukler N, Barshad G, **Toneyan S**, Danko CG, Siepel A. 2021. Deconvolution of expression for nascent RNA sequencing data (DENR) highlights pre-RNA isoform diversity in human cells. *Bioinformatics* **37:** 4727–4736.

*Authors contributed equally to the work. **Boldface indicates School of Biological Sciences student.**

SBS GRADUATES IN FACULTY OR INDEPENDENT POSITIONS (IN ORDER OF COMPLETION)

Name	Faculty Position
Ira Hall	Professor, Yale University, Connecticut
Niraj Tolia	Senior Investigator, NIAID, National Institutes of Health
Patrick Paddison	Professor, Fred Hutchinson Cancer Research Center, Washington
Elizabeth Bartom (*nee* Thomas)	Assistant Professor, Northwestern University, Illinois
Michelle Heck (*nee* Cilia)	Research Molecular Biologist, USDA-ARS and Adjunct Associate Professor, Cornell University
Zachary Lippman	Professor, Cold Spring Harbor Laboratory and Investigator, Howard Hughes Medical Institute
Ji-Joon Song	Professor, Korea Advanced Institute of Science and Technology (KAIST), South Korea
Elena Ezhkova	Professor, Mount Sinai School of Medicine, New York, New York
Masafumi Muratani	Professor, University of Tsukuba, Japan
Santanu Chakraborty	Program Director for Sustainable Life Sciences and Associate Professor, Atria University, India
Claudia Feierstein	Research Associate, Champalimaud Neuroscience Programme, Portugal
Gowan Tervo	Laboratory Head & Principal Scientist, HHMI, Janelia Farm Research Campus, Virginia
Marco Mangone	Associate Professor, Arizona State University, Arizona
Elizabeth Murchison	Professor, Cambridge University, United Kingdom
Hiroki Asari	Group leader, EMBL Monterotondo, Rome
François Bolduc	Associate Professor, University of Alberta, Canada
Wei Wei	Associate Professor, University of Chicago, Illinois
Christopher Harvey	Associate Professor, Harvard University, Massachusetts
Tomas Hromadka	Group Leader, Institute of Neuroimmunology, Slovak Academy of Sciences
Monica Dus	Assistant Professor, University of Michigan, Ann Arbor
Shu-Ling Chiu	Assistant Professor, Academia Sinica, Taiwan
Daniel Chitwood	Assistant Professor, Michigan State University, East Lansing
Jeremy Wilsuz	Associate Professor, Baylor College of Medicine, Texas
Shraddha Pai	Investigator, Ontario Institute for Cancer Research, Toronto
Oliver Fregoso	Assistant Professor, University of California, Los Angeles
Hiroshi Makino	Assistant Professor, Nanyang Technological University, Singapore
Katherine McJunkin	Stadtman Tenure Track Investigator, National Institutes of Health
Zhenxun Wang	Adjunct Junior Principal Investigator, Singapore Eye Research Institute and Adjunct Assistant Professor Duke-NUS Medical School, Singapore
Petr Znamenskiy	Assistant Professor, The Crick Institute, United Kingdom
Michael Pautler	Head, Genomics Services, Platform Genetics & Vineland Research and Innovation Centre, Lincoln, Ontario, Canada
Melanie Eckersley-Maslin	Assistant Professor, Peter Mac Cancer Centre at University of Melbourne
Elvin Wagenblast	Assistant Professor, Mount Sinai School of Medicine, New York, New York
Kristen Delevich	Assistant Professor, Washington State University, Pullman
Cinthya Zepeda Mendoza	Medical Director, Cytogenetics and Genomic Microarray at ARUP Laboratories; Assistant Professor, Department of Pathology, University of Utah, Salt Lake City
Wee Siong Goh	Assistant Professor Shenzhen Bay Laboratory, China
Arkarup Bandyopadhyay	Assistant Professor, Cold Spring Harbor Laboratory
Charles Underwood	Group Leader, Max Planck Institute for Plant Breeding Research, Cologne
Justus Kebschull	Assistant Professor, The Johns Hopkins University, Baltimore, Maryland

SBS GRADUATES IN INDUSTRY POSITIONS (IN ORDER OF COMPLETION)

Student	Current Position
Amy Caudy	Principal, Scientific Operations, Maple Flavored Solutions, Stony Brook, New York*
Emiliano Rial-Verde	Vice President, Food & Ingredients Strategy, Bunge Limited, New York
Rebecca Ewald	International Business Leader, Ventana Medical Systems/Roche, Tucson, Arizona

(*continued*)

SBS GRADUATES IN INDUSTRY POSITIONS (IN ORDER OF COMPLETION) (*continued*)

Student	Current Position
Catherine Seiler (*nee* Cormier)	Head, Biomarker Operations, H3 Medicine, Massachusetts
Yaniv Erlich	Chief Science Officer, My Heritage, Israel*
Colin Malone	Co-Founder & Head of Biology at VNV NewCo, New York, New York*
Amy Leung	Solutions Engineer, DNAnexus, California*
Amy Rappaport	Senior Scientist, Gristone Oncology, Emeryville, California
Frederick Rollins	Director, Competitive Intelligence, Oncology R&D, AstraZeneca, Washington DC
Patrick Finigan	Associate Director, Regulatory Affairs CMC, Gilead Sciences, Foster City, California
Kyle Honegger	Data Scientist, Ann & Robert H. Lurie Children's Hospital of Chicago, Illinois
Maria Pineda	Co-Founder, CEO, Envisagenics, New York, New York
Felix Schlesinger	Bioinformatics Lead, Scale Bio, California
Megan Hogan (*nee* Bodnar)	Senior Scientist, Neochromosome, Inc., New York
Paloma Guzzardo	Head, Target Biology, Variant Bio, Seattle, Washington
Saya Ebbesen	Associate Director, Medical + Scientific Strategy at BluPrint Oncology, London, United Kingdom
Hassana Oyibo	Senior Analytical Scientist, Ring Therapeutics, Massachusetts
Marek Kudla	Software Developer, Design Everest
Joshua Sanders	Founder and C.E.O., Sanworks, L.L.C., Rochester, New York
Katie Liberatore	Senior Manager, Trait Discovery & Targeting, Calyxt, Minneapolis, Minnesota
Kaja Wasik	Co-Founder and Chief Scientific Officer, Gencove & Variant Bio, New York, New York
Stephane Castel	Co-Founder at Variant Bio and Senior Research Fellow at New York Genome Center
Mitchell Bekritsky	Senior Bioinformatics Scientist, Illumina, Inc., Cambridge, United Kingdom
Sang-Geol Koh	Scientist and Entrepreneur, {Mind}, South Korea
Ozlem Aksoy (*nee* Mert)	Senior Scientist, Pfizer, California
Susann Weissmueller	Strategic Partnering Associate, Roche, Switzerland
Nilgun Tasdemir	Senior Scientist, Pfizer, New York
Silvia Fenoglio	Senior Scientist, Tango Therapeutics, Cambridge, Massachusetts
Jack Walleshauser	Senior Scientist Revolution Medicines, California
Lisa Krug	Scientist, Kallyope, New York, New York
Robert Aboukhalil	Senior Software Engineer, Chan Zuckerberg Initiative
Anja Hohmann	Associate Director, Discovery Biology, Constellation Pharmaceuticals, Cambridge, Massachusetts
Matt Koh	Research Engineer, ASAAP, New York, New York
Annabel Romero Hernandez	Director of AI for Drug Discovery and Molecular Modeling, SFL Scientific, Massachusetts
Maria Nattestad	Senior Software Engineer, Google, Palo Alto, California
Onyekachi Odoemene	Principal Data Scientist, Capital One, Baltimore, Maryland
Daniel Kepple	Senior Machine Learning Engineer, Samsung Artificial Intelligence Center, New York, New York
Kristina Grigaityte	Senior Data Scientist, Prometheus Biosicences, Inc.
Elizabeth Hutton	Computational Biologist, ArtisanBio, Colorado
Ngoc (Tumi) Tran	Machine Learning Scientist, Relay Therapeutics, Massachusetts
Shaina Lu	Bioinformatics Scientist, Eagle Genomics
Benjamin Harris	Computational Biologist, Lyell Immunopharma, California

*Left a faculty position

SBS GRADUATES IN POSTDOCTORAL OR ACADEMIC RESEARCH POSITIONS (IN ORDER OF COMPLETION)

Name	Faculty Position
Charles Kopec	Associate Professional Specialist, Princeton University (Advisor: Dr. Carlos Brody)
Oliver Tam	Computational Science Analyst, Cold Spring Harbor Laboratory, New York
Galen Collins	Postdoctoral Fellow, Harvard Medical School (Advisor: Dr. Alfred Goldberg)

(*continued*)

SBS GRADUATES IN POSTDOCTORAL OR ACADEMIC RESEARCH POSITIONS (IN ORDER OF COMPLETION) (*continued*)

Name	Faculty Position
David Simpson	Postdoctoral Fellow, Stanford University (Advisor: Dr. Alejandro Sweet-Cordero)
Claudio Scuoppo	Instructor, Columbia University (Advisor: Riccardo Dalla-Favera)
Elizabeth Nakasone	Heme-Onc Fellow, Fred Hutchinson Cancer Research Center
Eyal Gruntman	Research Scientist, Janelia Farm (Advisor: Michael Reiser)
Yevgeniy Plavskin	Postdoctoral Fellow, New York University (Advisor: Mark Siegal)
Zinaida Perova	Project Lead, Cancer Models, European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom
Dario Bressan	Head of IMAXT Laboratory, CRUK, Cambridge Institute, United Kingdom
Philippe Batut	Postdoctoral Fellow, Princeton University (Advisor: Michael Levine)
Joaquina Delas Vives	Postdoctoral Fellow, Francis Crick Institute, United Kingdom (Advisor: James Briscoe)
Abram Santana	Postdoctoral Fellow, Harvard University (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 (Advisor: Scott Lowe)
Paul Masset	Postdoctoral Fellow, Harvard University (Advisor: Venkatesh Murthy and Naoshige Uchida)
Talitha Forcier	Postdoctoral Fellow, Cold Spring Harbor Laboratory (Advisor: Molly Hammell)
Laura Maiorino	Postdoctoral Fellow, Koch Institute, MIT (Advisor: Darrell Irvine)
Giorgia Battistoni	Research Associate, Cancer Research, UK (Advisor: Gregory Hannon)
Sashank Pisupati	Postdoctoral Fellow, Princeton University (Advisor: Yael Niv)
Jacqueline Giovanniello	Postdoctoral Fellow, UCLA (Advisor: Kate Wassum)
Hamza Giaffar	Postdoctoral Fellow, UCSD (Advisor: Mikio Aoi)
Anqi Zhang	Postdoctoral Fellow, Harvard University (Advisor: Florian Engert)
Alexandra Nowlan	Postdoctoral Fellow, UNC Chapel Hill (Advisor: Zoe McElligott)
Katie Meze	Postdoctoral Fellow, University of Michigan (Advisor: Nils Walters)
Brianna Bibel	Postdoctoral Fellow, UCSF (Advisor: Danica Fujimori)

SBS GRADUATES IN NONRESEARCH POSITIONS (IN ORDER OF COMPLETION)

Student	Current Position
Ahmet M. Denli	Associate Editor, *Genome Research*, CSHL, New York
Beth Chen	Operations Manager at Homer Scientific Holdings Inc.
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	Managing Director, Scientific Communications, LifeSci Advisors, New York, New York
Keisha John	Associate Dean for Diversity, Equity and Inclusion, University of Virginia, Charlottesville
Elizabeth Nakasone	Heme-Onc Fellow, Fred Hutchinson Cancer Research Center
Ian Peikon	Venture Partner, Lux Capital, New York, New York
Colleen Carlston	Clinical Fellow in Pediatrics, Boston Children's Hospital, Massachusetts
Tyler Garvin	VP Operations, Underground Cellar, California
Brittany Cazakoff	Associate Lawyer, Cooley, California
Lital Chartarifsky	Associate Medical Director, FCB CURE, New York
Michael Gutbrod	Associate, SANTÉ, Massachusetts
Emilis Bružas	Medical Director, FCB Health, New York
Sanchari Ghosh	Content Specialist for bioRxiv, New York
Georgi Yordanov	Equity Research Associate, Cowen and Company, New York, New York
Jue Xiang Wang	Consultant, Boston Consulting Group, New York, New York
Matt Lee	Clinical Science Liaison, TG Therapeutics
Sofya Polyanskaya	Consultant, Scitaris, Berlin, Germany

*Left a faculty position.

The School Continues to Benefit from Generous Benefactors

It is the support of generous donors and benefactors that allows us to run a unique and successful graduate program. These gifts, which have over the years been essential to establish and maintain our program, include 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, Jenny and Jeff Kelter, Mr. David H. Koch, Guru Krupa Foundation, Inc., 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, 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. A new competitive application was submitted in May 2021.

Alexander Gann
Professor and Dean

SPRING CURRICULUM

TOPICS IN BIOLOGY

The spring Topics in Biology courses were suspended in 2021 because of COVID-19. The students will have the opportunity to take these classes in future years.

Teaching Experience at the Dolan DNA Learning Center

The teaching at the Dolan DNA Learning Center was suspended in 2021 because of COVID-19. Students will have the opportunity to teach in spring 2022.

Laboratory Rotations

Entering Class of 2020

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, 23 faculty members served as rotation mentors, some mentoring more than one student.

ROTATION MENTORS	Florin Albeanu	Hiro Furukawa	Zachary Lippman	Adam Siepel
	Arkarup Banerjee	Jesse Gillis	Michael Lukey	Jessica Tollkuhn
	Jeremy Borniger	Tobias Janowitz	Robert Martienssen	Lloyd Trotman
	Lucas Cheadle	Justin Kinney	Hannah Meyer	Christopher Vakoc
	Mikala Egeblad	Peter Koo	Saket Navlakha	Anthony Zador
	Tatiana Engel	Adrian Krainer	Andrea Schorn	

FALL CURRICULUM

Entering Class of 2021

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 lectures from Semir Beyaz, Hiro Furukawa, Dick McCombie, Monn Monn Myat, and Lloyd Trotman, while the Quantitative Biology boot camp was taught by Justin Kinney.

CORE COURSES

The Leslie C. Quick, Jr. Core Course on Scientific Reasoning and Logic

INSTRUCTORS	**Linda Van Aelst** (Lead) **Alexander Gann** **Christopher Hammell** **Leemor Joshua-Tor** **Jessica Tollkuhn**
GUEST LECTURERS	**Hiro Furukawa** **Adrian Krainer** **Bo Li** **Christopher Vakoc**
TEACHING ASSISTANTS	**Ankur Garg** **Nicholas Gladman** **Matt Jaremko** **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, and, in each, students read an assigned set of research articles and, at the end of the module, provide 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 funded National Institutes of Health R01 grants are reviewed and critiqued. This allows students to evaluate the research questions before discoveries are made, 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 2021, 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

INSTRUCTORS
David Jackson (Lead)
Sydney Gary
Rebecca Leshan
Hannah Meyer

GUEST LECTURERS (CSHL)
Semir Beyaz
Jeremy Borniger
Camila dos Santos
Molly Hammell
John Inglis
Alyson Kass-Eisler
Justin Kinney
Charla Lambert
Saket Navlakha
Rachel Rubino
Richard Sever
Adam Siepel

VISITING LECTURERS
Onyinye Balogun, Weill Cornell Medicine
Melissa Davis, Weill Cornell Medicine
Susan Friedman, The Innocence Project
Jaclyn Jansen, Weill Cornell Medicine
Emanuel Moss, Cornell Tech
Kendra Sirak, Harvard University

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 is essential 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 to be integral aspects of scientific research.

Research Topics Series

ORGANIZERS **Kimberley Graham**
Alyson Kass-Eisler

This series 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, methods of investigation, and mentorship styles each Wednesday evening. The students learned how to approach important problems in biology. These seminars provided students with a basis for selecting laboratories in which to do rotations.

SPECIALIZED DISCIPLINES COURSES

The students in the Entering Class of 2021 took a total of four Specialized Disciplines courses this fall: Quantitative Biology, Genetics, Cancer, and Systems Neuroscience.

Quantitative Biology

Throughout the semester

INSTRUCTOR **Justin Kinney** (Lead)

CSHL GUEST LECTURERS **Alex Dobin**
Peter Koo
David McCandlish
Hannah Meyer
Jon Preall
Adam Siepel

TEACHING ASSISTANTS **Mahdi Kooshkbaghi**
Anna Posfai

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

September 7–October 5

INSTRUCTOR	**Ullas Pedmale** (Lead)
GUEST LECTURERS	**David Jackson** **Zachary Lippman** **Robert Martienssen** **Benjamin Roche** **Sophie Zebell**
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 catalogued, organized, and mined? These questions and concepts were fleshed out using examples from the literature.

Cancer

October 1–22

INSTRUCTORS	**Mikala Egeblad** (Co-lead) **Christopher Vakoc** (Co-lead)
GUEST LECTURERS	**Semir Beyaz** **Jeremy Borniger** **Camila dos Santos** **Tobias Janowitz** **Michael Lukey** **David Tuveson** **Linda Van Aelst**

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 U.S. 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

November 1–December 3

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 BioXcel Therapeutics; Flare Therapeutics; Fred and Pamela Buffett Cancer Center, University of Nebraska Medical Center; the Hong Kong University of Science and Technology, Guangzhou; Mission Bio; National Chung Cheng University, Taiwan; National Research Institute for Agriculture, Food and Environment (INRAE), Laboratory of Plants, Microbes, and Environment Interactions (LIPME), Toulouse, France; Ohalo Genetics; University of Florida; and University of Puerto Rico School of Medicine.

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 Samantha Cyrill, Mary Doherty, Nikita Francis, Penelope Lindsay, and Jason Lynn.

CSHL endeavors to prepare postdocs to be competitive for the 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. In 2021 their events included the following talks and workshops: "Biotech spin-outs from academia," with Gregg Sando, the founder of Cell Medica Ltd; a Business Consulting Virtual Workshop: "How to crack the McKinsey case interview," with Manu Lakshmanan, author of *The Elements of the Case Interview*; a chat with CSHL alum, Dr. Emilis Bružas, Medical Director at FCB Health; and "Intellectual property law as an alternative career for biomedical researchers," with CSHL alum, Dr. Fatih Mercan.

Academic Career Training

Academic career training includes courses, lectures, and workshops on scientific enrichment, career exploration, and transferable skills, like leadership, mentorship, and communication. This year the series included "The Chalk Talk," "Interviewing for Academic Positions," "Navigating the International Job Search," and "Preparing Research, Diversity and Mentoring, and Teaching Statements." The CSHL faculty did one-on-one reviews of personal statements and CVs and conducted mock interviews for those postdocs who were on the job market in 2021. We also sent out a survey to the faculty asking them about what their likes and dislikes are in reviewing

applicants. This feedback was shared in the session on chalk talks. In addition, the Writing Resource Center held a Coffee Chat with Sydney Gary at which she shared tips on preparing CVs and cover letters.

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 or the United States and do not have a built-in social network. While the weather was good, and COVID transmission low, we returned to our monthly ice cream socials, and in the early fall we had a donuts and apple cider social. The PDLC also hosted an outdoor mini retreat in September. The retreat included a poster session; a Keynote talk on "Teaming up to investigate meiotic differentiation," by Gloria Brar and Elçin Ünal, who co-run a laboratory at UC Berkeley; and a BBQ picnic and mingle with special guest Bruce Stillman. We also distributed tote bags with the CSHL postdocs logo.

WiSE and DIAS

There are two affinity groups on campus, largely run by postdocs and students, dedicated to promoting diversity, inclusion, and equity in science: WiSE (Women in Science & Engineering) and DIAS (Diversity Initiative for the Advancement of STEM).

In addition to hosting prominent women and underrepresented minority scientists during the weekly CSHL seminar series, the groups held special events throughout the year. In February, BEC and WiSE collaborated on a "Women STEM Entrepreneurs—Panel Discussion," with invited speakers Nancy Thornberry, CEO of Kallyope, and Kaja Wasik, Co-founder and CSO at Variant Bio (and a graduate of our program). WiSE also hosted a virtual professional development workshop in partnership with the Committee on the Advancement of Women Chemists (COACh), called "Career Launch & Acceleration."

DIAS hosted a talk, "Towards Equity: Intersections of Policy and Higher Education Pre, During, and Post the Covid-19 Era," presented by Yvonne Muñoz as part of Black History Month in February 2021. Along with PDLC, they also held a screening of the *Coded Bias* documentary.

The WiSE, DIAS, and PDLC affinity groups held a leadership retreat in March 2021. The retreat provided an opportunity for affinity groups to hear from CSHL administrators, share feedback and suggestions, talk to each other about activities that could be implemented collaboratively, and start developing their proposals for institutional funds to run their programming.

CSHL Cancer Center and CSHL Research Operations

Beyond the career development opportunities organized by the School, there is additional labwide programming for students and postdocs including the Core Knowledge Series; Writing Resource Center Coffee Chats; the Career Directions series; Biostatistics Office Hours; and one-on-one assistance at the Writing Resource Center. CSHL's annual four-week Science Writing Course took place virtually in June. This course is designed to help researchers at all career stages improve their written communication skills. Topics in the 2021 course include Best practices for strong science writing and editing, Communicating with audiences outside your field, Writing a compelling fellowship or grant, and Drafting a manuscript.

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 metropolitan area. The Alliance's aim is to provide career and professional development programming for postdoctoral fellows and graduate students in science and engineering. The NYAS offered a number of virtual events this year, including a Career Paths series, which covered careers in Clinical Research and Management; Faculty Route; Life Science Consulting; Academic Workforce Development; PepsiCo (Data Science); and Policy. They also held workshops on how to Craft Your STEM Elevator Pitch and How to Write an Inclusive Letter of Recommendation. And they held the annual course on Scientists Teaching Science. The biannual conference "What Can You Be with a PhD," which is co-hosted by the NYAS and NYU, and supported by local institutions, including CSHL, was held on October 16–17, 2021.

OITE

The NIH Office of Intramural Training & Education (OITE) has opened up its programming to trainees at institutions outside the NIH. Our students and postdocs took advantage of these virtual offerings, which include a careers blog, various career workshops, and a series on "Becoming a Resilient Scientist."

UNDERGRADUATE RESEARCH PROGRAM

PROGRAM DIRECTORS **Jesse Gillis**
Christopher Hammell

PROGRAM ADMINISTRATOR **Kimberly Creteur**

Established 62 years ago, the CSHL Undergraduate Research Program (URP) provides undergraduates from around the world with hands-on undergraduate research training in biological sciences. The 10-week program begins the first week of June. Because of the ongoing COVID-19 pandemic, the program was held virtually in the summer of 2021. The modified program included training in scientific research, science communication, responsible conduct of research, mentoring-up, career preparation, and bioinformatics and computational biology. 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 their 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.

The following 12 students, selected from 724 applicants, took part in the 2021 program:

Vandana Agarwala
Advisor: **Peter Koo**
Funding: **National Science Foundation Scholar**
Representation learning of genomic sequence motifs via generative adversarial network model.

Jessica Dixon
Advisor: **Lucas Cheadle**
Funding: **National Science Foundation Scholar**
Defining the interactions between microglia and oligodendrocyte precursor cells in synapse elimination.

Matias Enriquez
Advisor: **Camila dos Santos**
Funding: **National Science Foundation Scholar**
Exploring parity-induced effects of EZH2-inhibited cells grown with EPP through CUT&RUN histone modification analysis.

Rajee Ganesan
Advisor: **Hannah Meyer**
Funding: **Garfield Fellow and Robert H.P. Olney Fellow**
Integrative analysis of single-cell expression and chromatin states in medullary thymic epithelial cells.

2021 Undergraduate Research Program Participants

Emily Guernsey
Advisor: **Zhuang Laboratory, University of Chicago**
Funding: **Libby Fellowship, Von Stade Fellowship and William Shakespeare Fellowship**
Facilitating inhibitory learning in the ventral striatum.

Catherine Kim
Advisor: **Doreen Ware**
Funding: **National Science Foundation Scholar**
Plant comparative genomics.

Caleb Mallery
Advisor: **Jon Preall**
Funding: **National Science Foundation Scholar**
Deep profiling of single-cell transcriptomes for detection of RNA degradation and miRNA biogenesis.

Lucas Melo
Advisor: **Alexander Krasnitz**
Funding: **National Science Foundation Scholar**
Integrative patterns of copy number variation in acute myeloid leukemia.

Elliot Meyers
Advisor: **Zachary Lippman**
Funding: **Alfred L. Goldberg Fellow, Burroughs Wellcome Fellow, and Joan Redmond Read Fellow**
Analyzing shared intergenic sequences to find putative regulatory elements.

Shoshana Novik
Advisor: **Tatiana Engel**
Funding: **National Science Foundation Scholar**
Comparing biological data to models of slow wave cortical neural spiking dynamics.

Noah Sobel
Advisor: **Jesse Gillis**
Funding: **Dorcas Cummings Scholar, Former URP Fund Scholar, and 30th Anniversary URP Scholar**
Single-cell co-methylation network construction and analysis.

Jess Stone
Advisor: **Ullas Pedmale**
Funding: **Katya H. Davey Fellow**
An exploratory analysis of the repression of root growth in response to shade.

SUMMER RESEARCH INTERNSHIP FOR MEDICAL STUDENTS (SRIMS)

PROGRAM DIRECTOR **Priya Sridevi**, Ph.D.

PROGRAM ADMINISTRATIVE COORDINATOR **Joanie O'Connor**

Through the CSHL and Northwell Health affiliation, a summer internship program, SRIMS (Summer Research Internship for Medical Students) was created to give first-year Zucker School of Medicine–Hofstra University students basic research experience and the opportunity to spend a summer working in a CSHL laboratory. To date, 16 students have been offered positions in mainly cancer and neuroscience laboratories at CSHL. Students commit 8–10 weeks (roughly July–September) to work full-time in a CSHL research laboratory during the program. The students work with their host PI to design a research project and present their work both at CSHL and at the annual "Medical Student Research Day" at Hofstra University, the following fall.

The following students took part in the 2021 program:

STUDENT	CSHL MENTOR
Shyam Bhagat	Chris Vakoc
Kyle Shutkind	Jeremy C. Borniger
Carol Wang	Camila dos Santos

PARTNERS FOR THE FUTURE

PROGRAM DIRECTOR **David Jackson**

PROGRAM ADMINISTRATOR **Lynn Carmen and Bridget Shanley**

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 biological and biomedical research at Cold Spring Harbor Laboratory. Applications to this highly competitive program are 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 for 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. While 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 2021–2022 Partners for the Future program students were chosen from ~70 nominations:

NAME	SCHOOL	MENTOR	LAB
Diana Benedicto-Jimenez	Schreiber HS	Olaf Klingbeil	Christopher Vakoc
Aurrel Bhatia	Bethpage High School	Julia Wang	Tatiana Engel
Kevin Biggiani	Oyster Bay East Norwich Schools	Amy Lanctot	Zachary Lippman
Kaitlin Bondi	Sanford H Calhoun HS	Hyun Soo Kim	Robert Martienssen
Aneesh Despande	Rocky Point High School	Arun Narasimhan	Pavel Osten
Kate Dooling	Lynbrook Senior High School	Nicholas Gladman/Vivek Kumar	Doreen Ware
Rohan Ghotra	Syosset High School	Peter Koo	Peter Koo
Ethan Labelson	Friends Academy	Nick Lee	Peter Koo
Jason Long	Half Hollow Hills High School East	Krasnitz Laboratory	Alexander Krasnitz
Isadora Luce	Rocky Point High School	Talitha Forcier	Molly Gale Hammell
Ashley Moon	Commack High School	Katherine Denney	Jessica Tollkuhn
Anne-Sarah Nichitiu	The Stony Brook School	Lucas Cheadle	Lucas Cheadle
Chelsea Russo	Seaford UFSD	May (Yue) Wu	Jeremy Borniger
Annabel Shen	Cold Spring Harbor High School	Jae Hyung Lee	David Jackson
Sophia Shen	Cold Spring Harbor High School	Min Yao	Douglas Fearon
Jeremiah Vargese	New Hyde Park Memorial HS	Mary Doherty	Lloyd Trotman

MEETINGS & COURSES PROGRAM

ACADEMIC AFFAIRS

With the COVID-19 pandemic continuing into its second year, the meetings and courses division continued to offer virtual meetings using Zoom videoconferencing, virtual poster sessions, and Slack channels for individual discussion. By year's end, the program had hosted 28 virtual meetings, including two further COVID/SARS CoV-2 Rapid Research Reports and five virtual courses.

Virtual Meeting	*To*	*From*	*Count*
COVID/SARS CoV-2 Rapid Research Reports #5	1/26	1/27	271
Network Biology	3/16	3/19	328
Nucleic Acid Therapies	3/24	3/26	288
Biology and Genomics of Social Insects	3/30	4/1	232
Brain Barriers	4/7	4/9	552
Probabilistic Modeling in Genomics	4/14	4/16	489
Systems Immunology	4/20	4/23	594
Ubiquitin, Autophagy, and Disease	4/27	4/30	552
Starr Cancer Consortium	5/3	5/4	327
The Biology of Genomes	5/11	5/14	922
Cellular Dynamics and Models	5/19	5/21	125
Retroviruses	5/25	5/28	603
85th Symposium: Biological Time Keeping	6/1	6/5	580
COVID/SARS CoV-2 Rapid Research Reports #6	7/7	7/8	157
ISHPSSB Biennial Meeting 2021	7/13	7/20	451
Plant Photobiology	7/22	7/25	260
Genome Engineering: CRISPR Frontiers	8/18	8/20	852
Eukaryotic mRNA Processing	8/24	8/27	439
Mechanisms of Eukaryotic Transcription	8/31	9/3	670
Eukaryotic DNA Replication and Genome Maintenance	9/8	9/12	531
Microbial Pathogenesis and Host Response	9/21	9/24	392
Neurobiology of *Drosophila*	10/5	10/8	638
Biology of Cancer: Microenvironment and Metastasis	10/12	10/15	484
Mechanisms of Metabolic Signaling	10/26	10/29	311
Genome Informatics	11/3	11/5	408
Single-Cell Analyses	11/10	11/12	582
Plant Genomes, Systems Biology, and Engineering	12/1	12/3	202
Telomeres and Telomerase	12/14	12/17	469

However, several courses were offered in virtual formats in 2021, and in November two courses were successfully offered in-person with no mishaps, and shortly before the appearance of the omicron variant.

2021 Courses	*To*	*From*	*Virtual Count*	*In-Person Count*
Tutorials in Genomics and Bioinformatics (Virtual)	2/17	2/19	55	
Ion Channels (Virtual)	6/7	6/18	40	
Mouse Engineering Minicourse (Virtual)	6/15	6/18	75	
Proteomics (Virtual)	8/10	8/13	45	
Antibody Engineering and Display Technologies (Hybrid)	11/9	11/22	6	26
Advanced Sequencing Technologies and Bioinformatics Analysis (Virtual)	11/11	11/19	47	
Scientific Writing Retreat (In-Person)	11/17	11/21		19
Tutorials in Genomics and Bioinformatics (Virtual)	11/30	12/2	44	

Renovated cabin and Dolan Hall room.

Institutional Investment in Campus Housing

With absolutely no visitors on campus in 2021, CSHL took advantage of the hiatus to continue various infrastructure improvements—including the refurbishment of the cabins to create two family units for instructors and providing every bedroom in the cabins with an en suite bathroom. On the Banbury campus, Sammis Hall also underwent renovation with the result being the elimination of shared bathrooms. These improvements will be keenly appreciated by our visiting scientists as they transition back to in-person conferences and courses.

Nobel News

The winners of the 2021 Nobel Prize in Physiology or Medicine both participated in our programs. David Julius was an instructor (1994–1995) in the Molecular Cloning of Neural Genes course and also lectured in several other years of the course in the 1990s. Ardem Patapoutian lectured in the same course (now Advanced Techniques in Molecular Neuroscience) (2002–2008) and has been an invited speaker at several CSHL meetings. Most importantly, Ardem was a student in the 1995 Banbury course on Developmental Neurobiology, bringing to 11 the total number of Nobelists who have taken a CSHL course in the modern era.

Cold Spring Harbor Asia

After a relatively quiet 2020, the COVID-19 situation in mainland China was such that we were able to offer a number of hybrid as well as virtual events in summer and fall, although travel restrictions into China were such that the in-person audiences were purely domestic, with international speakers and participants attending remotely.

The auditorium in the newly refurbished Cold Spring Harbor Asia Academic Center located in the Dushu Lake Pavilion (Dushu Ge).

2021 Cold Spring Harbor Asia Conferences	*To*	*From*	*Virtual Count*	*In-Person Count*
Frontiers of Biomedicine Empowered by Computational Approaches and Biophysical Principles (Virtual)	3/15	3/17	78	
Immunoreceptor Signaling: From Bench to Bed (Virtual)	4/20	4/22	240	
Yeast and Life Sciences (Hybrid)	6/21	6/24	54	57
Synthetic Biology (Hybrid)	10/25	10/29	61	34
Mitochondria and Metabolism in Health and Disease (Hybrid)	11/15	11/19	58	25
DNA Metabolism, Genomic Stability, and Human Disease (Hybrid)	11/22	11/26	126	20
Liver Development, Metabolism, Disease, and Cancer (Virtual)	12/7	12/10	228	
Integrative Epigenetics in Plants (Virtual)	12/14	12/14	196	
TOTAL			**1041**	**136**

Program Staff

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. The COVID-19 pandemic posed a host of new challenges, and the entire Meetings & Courses team worked diligently through a second year of largely virtual events. Their hard work, enthusiasm, and willingness to develop and adapt to new and untried virtual operations was truly remarkable, all the more so because so much was done while working remotely. In particular, the audiovisual and digital arts director, Ed Campodonico, and his audiovisual team not only worked tirelessly to support the virtual conferences and courses, but also supported many other Laboratory activities, from Board of Trustees meetings and major Development events to the frequent presidential town halls and the many scientific seminars and lectures, all of which were conducted remotely. We said goodbye to a number of staff, including Catherine Carr, Alicia/Vanessa Franco, Rachel Lopez, and Andrea Newell, whom we collectively thank for their hard work on behalf of the Laboratory.

David Stewart
Executive Director, Meetings & Courses Program/
President, Cold Spring Harbor Asia

Terri Grodzicker
Academic Guidance, Dean of Academic Affairs

Program Funding

In 2021, the National Institutes of Health provided multiple grants for individual meetings, whereas Regeneron, the Chan Zuckerberg Initiative, and the International Brain Research Organization provided targeted support to broaden access to certain meetings for scientists at minority-serving institutions in the United States and Canada and from low- and middle-income countries.

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
Biogen
Bristol-Myers Squibb Company
Calico Life Sciences LLC
Merck & Co., Inc.
New England BioLabs

Corporate Partners
Alexandria Real Estate
Enzo Biochem
Gilead
Novartis
Pacific Biosciences

We appreciate the ongoing major financial support for our courses from the following: the Helmsley Charitable Trust, the Howard Hughes Medical Institute, the National Institutes of Health, and the National Science Foundation. The course program is normally supported by major equipment and reagent companies and we appreciate their flexibility during the COVID-19 pandemic.

85TH COLD SPRING HARBOR LABORATORY SYMPOSIUM ON QUANTITATIVE BIOLOGY

Biological Time Keeping (Virtual)

June 1–5 580 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 85th year (the 2020 Symposium was postponed to 2022 because of the COVID-19 pandemic). The series was 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 at which general and intensive scrutiny and review is warranted. Several previous Cold Spring Harbor Symposia have addressed different aspects of the theme of chronobiology, including the 25th Symposium on Biological Clocks (1960) and the 72nd Symposium on Clocks and Rhythms (2007). The enormous progress in the field in the past 15 years led us to conclude that the time is past due for another Symposium on this fascinating topic.

Abstract book cover art

The scope of the Symposium on Biological Time Keeping covered a broad range of topics including Structures, Biophysics, and Dynamics of Clocks; Genetics, Genomics, and Systems Biology; Clock Complexes and Regulation; Regulation of Downstream Targets; Cell Biology of Time Keeping; Control of Cell Cycle and Cell Proliferation; Developmental Timing; Metabolic Rhythms; Reproductive Cycles and Sex Differences; Sleep and Sleep Disorders; Brain and Neural Rhythms; Seasonal and Circannual Clocks; Clocks and Aging; and Disorders of Time Keeping.

The Symposium attracted 500 virtual participants and provided an extraordinary four-day synthesis of current understanding in the field. Opening talks in the introductory session setting the scene for later sessions included Joseph Takahashi, who provided a masterful introduction to the field in his talk on "Circadian clocks and their impact on metabolism, aging, and longevity"; Carrie Partch on "Origins of cooperativity in the cyanobacterial post-translational oscillator"; Mitchell Lazar, who addressed "Tissue-specific control of rhythms and metabolism by circadian REV-ERBs"; and Amita Sehgal on "Cellular insights into why we sleep." 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, New York*

Metabolism
S. Masri, *University of California, Irvine*; R. Evans, *Salk Institute, La Jolla, California*

Clock Components
U. Schibler, *University of Geneva, Switzerland*; C. Partch, *University of California, Santa Cruz*

Clocks and Disease
J. Bass, *Northwestern University Feinberg School of Medicine, Chicago, Illinois*; C. Dibner, *University of Geneva, Switzerland*

Coordinating Clocks
C. Green, *UT Southwestern, Dallas, Texas*; S. Golden, *University of California, San Diego*

Neural Clocks
M. Shirasu-Hiza, *Columbia University Medical Center, New York, New York*; J. Takahashi, *HHMI/UT Southwestern Medical Center, Dallas, Texas*

Sleep
M. Rosbash, *HHMI/Brandeis University, Waltham, Massachusetts*; A. Sehgal, *HHMI/University of Pennsylvania, Philadelphia*

Seasonal Rhythms/Aging
A. Brunet, *Stanford University, California*; C. Johnson, *Vanderbilt University, Nashville, Tennessee*

MEETINGS

COVID/SARS CoV-2 Rapid Research Reports #5 (Virtual)

January 26–27 271 Participants

ARRANGED BY **Hung Fan**, University of California, Irvine
Adrian Hayday, King's College London, United Kingdom
Volker Thiel, University of Bern, Switzerland
Susan Weiss, University of Pennsylvania Perelman School of Medicine, Philadelphia

COVID/SARS-CoV-2 Rapid Research Reports series brings together expert scientists on the forefront of COVID research into viral origins and variation, virus biology, host response, and antivirals and vaccines. Previous meetings in the series were held in June, July, August, and November 2020. The series is designed to stimulate ideas and collaborations among scientists with the goal of hastening a solution to vanquish this worldwide disease. The full program for this fifth meeting is listed below, and a meeting blog by CSHL graduate student Brianna Bibel is available online (https://currentexchange.cshl.edu/blog/2021/5/covid-meeting-series-part-4?rq=covid).

Support for this meeting was provided by the Corporate Sponsor Program.

PROGRAM

Session I: Host–Virus Interactions/Structure

Co-Chairpersons: M. Kikkert, *Leiden University Medical Center, the Netherlands;* T. Schountz, *Colorado State University, Fort Collins*

N. Ferguson, *Imperial College London, United Kingdom:* Epidemiology of B.1.1.7.

H. Zeberg, *Karolinska Institute, Sweden:* Neandertal gene variants and risk of severe COVID-19.

B. Schneider, *The Rockefeller University, New York*: CRISPR screens identify TMEM41B as a pan-coronavirus host factor.

A. Kratzel, *Universität Bern, Switzerland:* A genome-wide CRISPR screen identifies interactors of the autophagy pathway as conserved coronavirus targets.

N. Ban, *ETH Zürich, Switzerland*: Structural basis of SARS-CoV-2 translational shutdown and programmed ribosomal frameshifting.

Y. Xiong, *Yale University, New Haven, Connecticut*: Nonstructural protein 1 of SARS-CoV-2 is a potent pathogenicity factor redirecting host protein synthesis machinery toward viral RNA.

A. Wrobel, *The Francis Crick Institute, United Kingdom:* Furin cleavage and receptor binding enhances SARS-CoV-2 spike opening and primes it for membrane fusion.

W. Vuong, *University of Alberta, Canada:* A feline coronavirus drug inhibits SARS-CoV-2 Mpro and halts viral replication.

Session II: Coronavirus Biology

Co-Chairpersons: V. Menachery, *University of Texas Medical Branch;* S. Pfaender, *Ruhr-Universität Bochum, Germany*

L. Zaeck, *Friedrich Loeffler Institute, Germany:* 3D reconstruction of SARS-CoV-2 infection in ferrets emphasizes focal infection pattern in the upper respiratory tract.

W. Mothes, *Yale University, New Haven, Connecticut:* Imaging SARS-CoV-2 across multiple scales from animals to single molecules.

N. Altan-Bonnet, *National Institutes of Health, Bethesda, Maryland:* Sneaking out with the trash: How coronaviruses exit cells.

R. Baric, *University of North Carolina, Chapel Hill:* SARS-CoV-2 and D614G emergence, transmission, and evolution.

B. Zhou, Centers for Disease Control and Prevention, Atlanta, Georgia: SARS-CoV-2 spike 614G variant confers enhanced replication and transmission.

L. Chakrabarti, *Institut Pasteur, France:* SARS-CoV-2 Infection induces the dedifferentiation of multi-ciliated cells and impairs mucociliary clearance.

A. Hume, *Boston University, Massachusetts:* SARS-CoV-2 infection of pluripotent stem cell-derived human lung alveolar type 2 cells reveals actionable cytopathogenic host responses.

Session III: Pathogenesis and the Immune Response

Co-Chairpersons: P. Klenerman, *Medawar/University of Oxford, United Kingdom*; F. Di Rosa, *Institute of Molecular Biology & Pathology (CNR), Italy*

Y. Hamazaki, *Kyoto University, Japan:* The impact of aging on SARS-CoV-2 reactive T cells.

F. Krammer, *Icahn School of Medicine at Mount Sinai, New York*: Antibody responses to the SARS-CoV-2 spike.

A. Bertoletti, *Duke-NUS Medical School, Singapore*: Profile of SARS-CoV2 T cell response in asymptomatic infection.

P. Khatri, Stanford University, California: Learning across viruses and preparing for the next pandemic.

J. Heath, *Institute for Systems Biology, Seattle, Washington*: Deep, longitudinal analyses of 200 COVID-19 patients from clinical diagnosis to disease recovery.

A. Ring, *Yale University, New Haven, Connecticut:* Autoantibodies in COVID-19.

A.-K. Reuschl, *University College London, United Kingdom:* RIG-I and MDA5 sensing of SARS-CoV-2 RNA links epithelial infection to macrophage inflammation.

Session IV: Vaccines/Antivirals

Co-Chairpersons: M. Frieman, *University of Maryland, Baltimore;* L. Gralinski, *University of North Carolina, Chapel Hill*

F. Lund, *University of Alabama, Birmingham*: Development of mucosal vaccines for SARS-CoV-2.

S. Whelan, *Washington University, St. Louis, Missouri: Vesicular stomatitis* virus as a bullet against COVID.

J. Chodera, *Memorial Sloan Kettering Cancer Center, New York:* The COVID moonshot: Closing in on an orally bioavailable non-peptidomimetic small-molecule inhibitor of SARS-CoV-2 Mpro with an open science collaboration.

C. Schofield, *University of Oxford, United Kingdom:* Mechanistic and inhibition studies on Mpro from SARS-CoV-2.

M. Porotto, *Columbia University, New York:* Fusion inhibitory lipopeptide prevents SARS-CoV-2 infection in vivo.

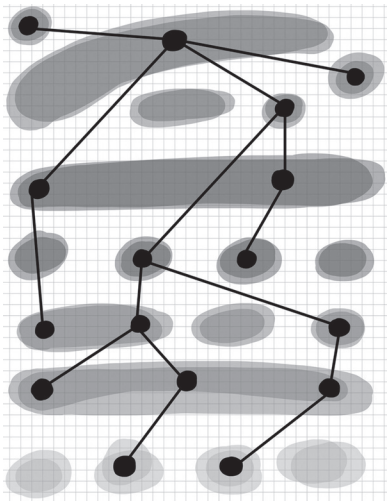
A. Hargrove, *Duke University, Durham, North Carolina:* Amilorides inhibit SARS-CoV-2 replication in vitro by targeting RNA structures.

T. Yaron, *Weill Cornell Medical College, New York*: The FDA-approved drug Alectinib compromises SARS-CoV-2 nucleocapsid phosphorylation and inhibits viral infection.

Network Biology (Virtual)

March 16–19 328 Participants

ARRANGED BY **Anne-Ruxandra Carvunis**, University of Pittsburgh, Pennsylvania
Pascal Falter-Braun, Helmholtz Zentrum Münich and Ludwig-Maximilian-University, Germany
Roded Sharan, Tel Aviv University, Israel
Michael Springer, Harvard Medical School, Boston, Massachusetts



Abstract book cover art

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 still 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, which has led to an explosion not just in the quantity, but in the types of data available. Two key goals of the Network Biology 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. The most recent meeting successfully achieved these goals and 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 previous meetings. This year's panel discussed the applicability and translatability of network biology for bettering human health. This highly interactive session involved panelists and plenty of audience participation, including a mix of scientific questions and career development advice. The discussion was sparked by presentations from the panelists illustrating their successes for COVID-19 and other diseases. The "meet the PI" lunch was replaced by small Zoom gatherings where pre-registered audience participants could spend time with PIs of their choice to discuss science and career development. This format was intended to facilitate and catalyze networking and facilitate young scientists getting in direct contact with senior scientists and PIs virtually. In addition, speakers, chairs, and audience members could meet in a Zoom "discussion zone" to elaborate and deepen the conversation after each session. Finally, Slack was used for continuing discussions, including poster feedback. These elements were praised by attendees and resulted in continued discussions throughout the virtual meeting.

The scientific program opened on the morning of March 16. There were 22 invited presentations and 26 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 (30% women

presenters). The talks covered a wide range of concepts, from 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 network biology brings together people from different fields of biology. New areas and areas of growth for the community included deep learning, single-cell networks, and dynamic network modeling. These presentations showcased recent advances and also 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. Naama Barkai opened the meeting with an overview of her lab's efforts to understand how network biology governs the evolutionary fate of gene duplicates. Stephen Michnick closed the meeting by highlighting biomolecular condensates as a next frontier in network biology.

Support for this meeting was provided in part by the National Human Genome Research Institute, a branch of the National Institutes of Health.

PROGRAM

Keynote Address
N. Barkai, *Weizmann Institute of Science, Israel*

Protein Networks
Chairperson: M. Springer, *Harvard Medical School, Boston, Massachusetts*

Regulatory Networks
Chairperson: A.-Ruxandra Carvunis, *University of Pittsburgh, Pennsylvania*

Signaling Networks
Chairperson: S. Gaudet, *Novartis Institute of Biomedical Research, Cambridge, Massachusetts*

Panel Discussion: Applicability and Translatability of Network Biology
Chairperson: P. Falter-Braun, *Helmholtz Zentrum Münich and Ludwig-Maximilian-University, Neuherberg, Germany*

Networks for Precision Medicine
Chairperson: C. Myers, *University of Minnesota-Twin Cities, Minneapolis*

Computational Methods for Network Biology
Chairperson: R. Sharan, *Tel Aviv University, Israel*

Deep Learning in Network Biology
Chairperson: P. Falter-Braun, *Helmholtz Zentrum Münich and Ludwig-Maximilian-University, Neuherberg, Germany*

Genetic Networks
Chairperson: F. Roth, *University of Toronto, Canada*

Network Evolution
Chairperson: A.-R. Carvunis, *University of Pittsburgh, Pennsylvania*

Single Cell Networks
Chairperson: M. Springer, *Harvard Medical School, Boston, Massachusetts*

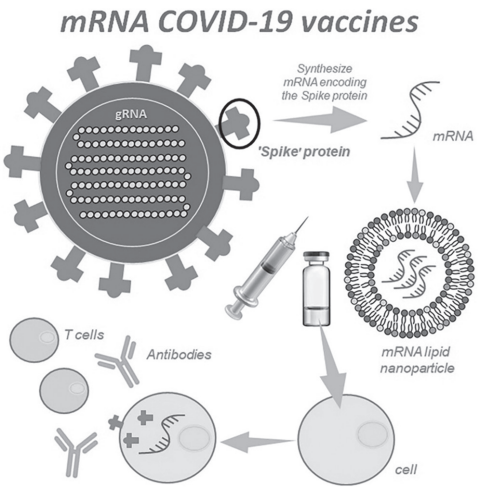
Keynote Address: The Dimensions of Gene Function
S.W. Michnick, *Université de Montréal, Montréal, Canada*

Nucleic Acid Therapies (Virtual)

March 24–26 288 Participants

ARRANGED BY **Shalini Andersson**, AstraZeneca, Sweden
Masad Damha, McGill University, Québec, Canada
Anastasia Khvorova, University of Massachusetts Medical School, Worcester
Matthew Stanton, Generation Bio, Cambridge, Massachusetts

This sixth Cold Spring Harbor conference focused on the 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, numerous nucleic acid therapies have received regulatory approval, including the dramatically rapid development of COVID-19 vaccines. The Keynote Address given by Pfizer's Dr. Philip Dormitzer described the clinical development and launch of their COVID-19 vaccine. Other sessions showed a diversity of nucleic acid modalities being developed preclinically and clinically. Furthermore, there were sessions on important aspects such as safety, chemistry and delivery, and gene editing.



Abstract book cover art

The participants came from 125 companies and from universities and research institutions from the United States and abroad. The five scientific sessions featured 33 platform talks, 66 posters, and a panel discussion and included 288 registered attendees. Animated and insightful exchanges continued throughout the sessions, during breaks, and at social events. Most participants indicated the meeting was excellent and expressed interest in attending the next edition.

Support for this meeting was provided in part by Intellia Therapeutics.

PROGRAM

Nucleic Acid Chemistry
Chairpersons: M. Damha, *McGill University Montréal, Canada;* M. Manoharan, *Alnylam Pharmaceuticals, Cambridge, Massachusetts*

Delivery
Chairpersons: M. Stanton, *Generation Bio, Cambridge, Massachusetts;* S. Dowdy, *University of California, San Diego*

Gene Editing
Chairpersons: L. Sepp-Lorenzino, *Intellia Therapeutics, Cambridge, Massachusetts;* M. Stanton, *Generation Bio, Cambridge, Massachusetts*

Keynote Address and Fireside Chat
P. Dormitzer, *Pfizer, New York*

Nucleic Acid Preclinical Programs
Chairpersons: A. Aartsma-Rus, *Leiden University Medical Center, the Netherlands;* S. Andersson, *AstraZeneca, Gothenburg, Sweden*

Nucleic Acid Clinical Programs
Chairperson: A. Khvorova, *University of Massachusetts Medical School, Worcester*

Biology and Genomics of Social Insects (Virtual)

March 30–April 1 288 Participants

ARRANGED BY **Sandra Rehan**, York University, Canada
Olav Rueppell, University of Alberta, Canada
Seirian Sumner, University College London, United Kingdom

Abstract book cover art

Enabled by its virtual format, this third meeting was attended by a record 232 participants from five continents. Continuing previous positive trends toward equal gender representation and diversifying the meeting, 47.5% of the participants were female, 27.8% were graduate students, and 18.7% were postdocs. Notably, 63 corporate scientists and four journal staff were also among the attendees. The meeting included 60 talks, organized into three keynote presentations, 10 oral sessions, and 130 posters, presented virtually during three poster sessions. A wide variety of specific topics were covered, ranging from long-standing topics, such as "Chemical Ecology and Communication in Social Insects," to emerging themes, such as "Novel Research Methods and Their Application to Social Insects" and "Beyond Classic Gene Regulation in Social Insects." Because of its practical importance, a well-attended session on applied honeybee health and stress research was also included. An immense amount of content was packed into three intense days that allowed attendees to update each other about the rapid progress in our systemic understanding of social evolution and its causes and consequences. Critical discussions were held with the potential of shaping the trajectory of the field's scientific progress, and new connections and collaborations were formed. This conference demonstrated how sophisticated and new genomic approaches can accelerate our understanding of complex biological systems. It also became once more apparent that social insect scientists are tackling cutting-edge topics at the forefront of more general biological disciplines, such a developmental biology and microbiome research, providing unique perspectives and valuable comparative model systems. Single-cell transcriptome sequencing and machine-learning approaches to big data analysis were emerging as two new intriguing approaches. Despite the rich diversity of topics, methodology, and investigated species, the meeting also highlighted our collaborative community of scientists who are ultimately unified in the quest to understand the biology and genomics of social insects.

Support for this meeting was provided by the Corporate Sponsor Program.

PROGRAM

Keynote Address
M. Sokolowski, *University of Toronto, Canada*

Genomics of the Social Nervous System
Chairperson: C. Rittshof, *University of Kentucky, Lexington*

Evolution of Social Genomes
Chairperson: G. Zhang, *University of Copenhagen, Denmark*

Beyond Classic Gene Regulation in Social Insects
Chairperson: M. Flenniken, *Montana State University, Bozeman*

Keynote Address
K. Matsuura, *Kyoto University, Japan*

Chemical Ecology and Communication of Social Insects
Chairperson: E. Amselem, *Pennsylvania State University, University Park*

Genomics of Social Transitions
Chairperson: E. Abouheif, *McGill University, Montréal, Canada*

Ecological Genomics of Social Insects
Chairperson: Y. Wurm, *Queen Mary University of London, United Kingdom*

Functions of Microbiomes in Social Insects
Chairperson: M. Poulson, *University of Copenhagen, Denmark*

Keynote Address
S. Foitzik, *Johannes Gutenberg University, Mainz, Germany*

Disease and Stress Interactions Compromising Social Insect Health
Chairperson: M. Spivak, *University of Minnesota, St. Paul*

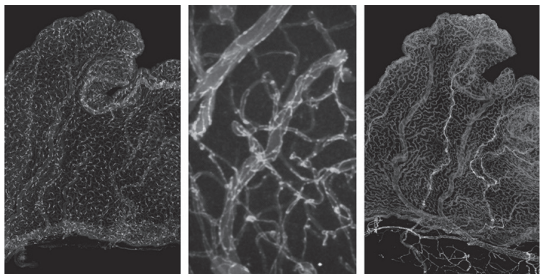
Complex Phenotype Genomics in Social Contexts
Chairperson: Z. Simoes, *University of São Paulo, Brazil*

Novel Research Methods and Their Application to Social Insects
Chairperson: R. Waterhouse, *University of Lausanne, Switzerland*

Brain Barriers (Virtual)

April 7–9 552 Participants

ARRANGED BY **Dritan Agalliu**, Columbia University, New York
Maria Lehtinen, Boston Children's Hospital, Massachusetts
Benoit Vanhollebeke, Université Libre de Bruxelles, Belgium



Abstract book cover art

This meeting highlighted the latest scientific advances linked to the biology of the central nervous system barriers, including the blood-brain barrier (BBB), the blood-retinal barrier, and the blood-cerebrospinal fluid (CSF) barrier. By virtue of its virtual format, the conference attendance was particularly high, with more than three times the participants of the previous meeting. Close to 70% of the submitted abstracts were from graduate students or postdoctoral fellows, who played an important role in giving oral presentations (44.2%). Moreover, women scientists were well represented at the meeting. There were 57.2% female attendees, and many of them were session chairs and oral presenters (52.4%).

Besides the significant time allocated to emerging technologies and model organisms, the conference highlighted more than in the past the important role played by CNS barriers other than the BBB in regulating molecular and cellular traffic in and out of the CNS. Another focus of the conference was the molecular definition of the human brain barriers and their comparison to the model systems structures. Time was also allocated to the clinical implications of the brain barriers, balancing between BBB dysfunction in disease and emerging therapeutic strategies for CNS drug delivery and BBB repair.

The Keynote Address was given by Chenghua Gu. The meeting oral sessions are listed below under Program. Selection of material for oral and poster presentation was made by the organizers on the basis of scientific merit.

Support for this meeting was provided in part by the National Institute of Neurological Disorders and Stroke, a branch of the National Institutes of Health; the Chan Zuckerberg Initiative; Brains for Brain Foundation; and IBRO.

PROGRAM

Keynote Address: Molecular Mechanisms Governing the Blood–Brain Barrier Function
C. Gu, *Harvard Medical School, Boston, Massachusetts*

Emerging Concepts and Hot Topics in CNS Barriers
Chairpersons: E. Monuki, *University of California, Irvine;* S. Liebner, *Goethe University Frankfurt, Germany*

Imaging Approaches to Study CNS Barriers
Chairpersons: R. Daneman, *University of California, San Diego;* C. Moore, *Brown University, Providence, Rhode Island*

Development, Plasticity, and Specialization of CNS Barriers
Chairpersons: P. Magnusson, *Uppsala University, Sweden;* A. Acker-Palmer, *Goethe University Frankfurt, Germany*

Panel Discussion: The Brilliance Barrier
A. Cimpian, *New York University, New York*

Systems, Computational, and Cell Biology of the CNS Barriers
Chairpersons: A. Eichmann, *Yale University, New Haven, Connecticut;* B. Engelhardt, *University of Bern, Switzerland*

Neuroimmune Interactions Across CNS Barriers
Chairpersons: T. Wyss-Coray, *Stanford University, California;* Z. Fabry, *University of Wisconsin, Madison*
Workshop: How to Frame and Write Constructive, Fair, Peer-Review

CNS Barriers in Aging, Senescence, and Neurodegeneration
Chairpersons: R. Klein, *Washington University School of Medicine, St. Louis, Missouri;* M. Campbell, *Trinity College Dublin, Ireland*

Probabilistic Modeling in Genomics (Virtual)

April 14–16 489 Participants

ARRANGED BY **Andrew Kern**, University of Oregon, Eugene
Sohini Ramachandran, Brown University, Providence, Rhode Island
Adam Siepel, Cold Spring Harbor Laboratory

Abstract book cover art

This meeting, which grew out of two ad hoc meetings on a similar topic in 2013 at Janelia Farm and in 2014 at Merton College, Oxford, provides a forum for presentation and exchange of ideas among researchers who are working in the general area of genomics, but are particularly focused on the development of new probabilistic and machine-learning models, algorithms, and methods for inference. These researchers come from a variety of backgrounds, including computer science, statistics, applied mathematics, and physics. Initially, the meeting strongly emphasized population genetics, but over time it has branched out to include topics such as functional genomics, systems biology, quantitative genetics, and cancer evolution. In this version of the meeting, we made a particular effort to emphasize emerging techniques in machine learning across various application areas.

The session topics are listed below under Program. There were six oral presentations per session for a total of 36 presentations. Two invited session chairs presented in each session, and four additional talks were selected from submitted abstracts. Talks were 20 minutes long plus five minutes for questions and answers. The quality of the presentations was very high overall, with considerable mathematical sophistication combined with a clear focus on biological relevance. We were honored to recruit two distinguished senior scientists to attend the meeting and present Keynote Addresses: Daphne Koller and Daniela Witten.

Notably, the diversity of the meeting has steadily improved over time. This time, slightly more than half of the speakers, and 38% of attendees, were women. The meeting continues to be well represented by trainees, with 38% of attendees being grad students and another 24% being postdocs. In addition, about a third (160) of attendees came from outside the United States, with representation from 27 countries across Europe, the Americas, the Middle East, Asia, and Africa.

Overall, the meeting appeared to be highly successful, with several attendees commenting on the Slack channel and on social media about the quality of the content, and about how well the online forum worked. The poster session was particularly appreciated, with some participants preferring it to the conventional protocol, particularly given the technical content of many of the posters. We anticipate the next meeting will be held in Europe in 2022, but a discussion is still under way about the organizers and location.

Support for this meeting was provided by the Corporate Sponsor Program.

PROGRAM

Quantitative Genetics and Association Mapping
Chairpersons: L. Crawford, *Microsoft Research New England, Cambridge, Massachusetts;* H. Finucane, *Broad Institute of MIT and Harvard, Cambridge, Massachusetts*

Machine Learning in Genomics
Chairpersons: P. Koo, *Cold Spring Harbor Laboratory;* S. Mathieson, *Haverford College, Pennsylvania*

Keynote Address
D. Koller, *insitro, South San Francisco, California*

Transcriptomics and Epigenomics
Chairpersons: A. Kundaje, *Stanford University, California;* T. Lappalainen, *New York Genome Center and Columbia University, New York*

Population Genetics I: Natural Selection
Chairpersons: E. Huerta-Sanchez, *Brown University, Providence, Rhode Island;* D. Schrider, *University of North Carolina, Chapel Hill*

Applications to Cancer and other Diseases
Chairpersons: E. Khurana, *Weill Cornell Medicine, New York;* E. Leffler, *University of Utah, Salt Lake City*

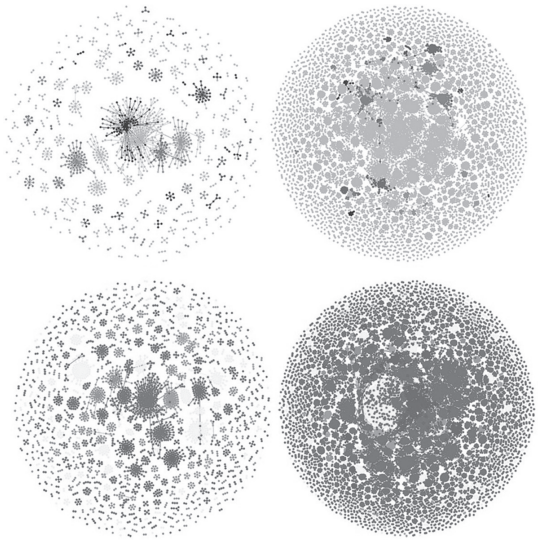
Keynote Address
D. Witten, *University of Washington, Seattle*

Population Genetics II: Mutation, Recombination, and Demography
Chairpersons: F. Jay, *Paris-Saclay University, CNRS, INRIA, Orsay, France;* F. Racimo, *University of Copenhagen, Denmark*

Systems Immunology (Virtual)

April 20–23 594 Participants

ARRANGED BY **Menna Clatworthy**, University of Cambridge, United Kingdom
Kathryn Miller-Jensen, Yale University, New Haven, Connecticut
Harinder Singh, University of Pittsburgh, Philadelphia
John Tsang, National Institute of Allergy and Infectious Diseases/NIH, Bethesda, Maryland



Abstract book cover art

This biennial meeting is designed to bring together scientists working at the interface of experimental immunology and computational and systems biology. There were 594 participants, of whom 47.5% were female, 27.8% were graduate students, and 18.7% were postdocs. Notably, 63 corporate scientists and four journal staff were also among the attendees. The meeting included 60 talks and 130 posters. Advances in single-cell genomic and proteomic profiling, as well as spatial and time-resolved imaging and their coupling to computational approaches, are making 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 cells, notably CAR T cells, have ushered in the dawn of systems immunology. This second meeting enhanced the nucleation and fostered the growth of a vibrant community of systems immunologists. The scientific program is listed below under Program. The format of the meeting included a large number of oral presentations that were selected from submitted abstracts, in addition to those from invited speakers. Based on 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 2023 to sustain and nurture this vital and rapidly evolving field. Major fundamental as well as translational advances in this field are anticipated in the coming years.

Support for this meeting was provided in part by the National Institute of Allergy and Infectious Diseases, a branch of the National Institutes of Health.

PROGRAM

Single-Cell Analysis I
Chairpersons: S. Teichmann, *Wellcome Sanger Institute, Cambridge, United Kingdom;* H. Singh, *University of Pittsburgh, Pennsylvania*

Immunoreceptors: Specificities and Signaling
Chairpersons: M. Ackerman, *Dartmouth College, Hanover, New Hampshire;* M. Davis, *Stanford University School of Medicine, California*

Cellular Dynamics, Interactions, and Communication
Chairpersons: R. Germain, *National Institute of Allergy and Infectious Diseases/NIH Bethesda, Maryland;* K. Miller-Jensen, *Yale University, New Haven, Connecticut*

Systems Human Immunology I
Chairpersons: B. Pulendran, *Stanford University School of Medicine, California;* M. Clatworthy, *University of Cambridge, United Kingdom*

Modeling of Immune Signaling and Gene Regulatory Networks
Chairpersons: A.S. Perelson, *Los Alamos National Laboratory, New Mexico;* U. Alon, *Weizmann Institute of Science, Rehovot, Israel*

Single-Cell Analysis II
Chairpersons: G. Trynka, *Wellcome Sanger Institute, Cambridge, United Kingdom;* S. Quake, *Stanford*

University, California; C. Zuckerberg, *Biohub, Mission Bay, California*

Evolution of Immune Systems
Chairpersons: T. Boehm, *Max Planck Institute of Immunology and Epigenetics, Freiburg, Germany;* R. Medzhitov, *HHMI/Yale University, New Haven, Connecticut*

Synthetic Immunology and Engineering Systems
Chairperson: W. Lim, *University of California, San Francisco*

Systems Human Immunology II: COVID-19 Pandemic
Chairpersons: C. Blish, *Stanford University, California;* J. Tsang, *National Institute of Allergy and Infectious Diseases/NIH, Bethesda, Maryland*

Ubiquitin, Autophagy, and Disease (Virtual)

April 27–30 552 Participants

ARRANGED BY **Anne Bertolotti**, MRC LMB, United Kingdom
Michael Rape, HHMI/University of California, Berkeley
Nicolas Thomä, Friedrich Miescher Institute for Biomedical Research, Switzerland

Abstract book cover art

This is the 12th meeting (previously titled 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. To pay tribute to the evolution of the ubiquitin field, this meeting focused more on autophagy and, particularly, on biomedical applications that alter the activity of ubiquitin-dependent signaling in disease: An entire session was dedicated to the new therapeutic modality of induced protein degradation. As in previous years, this meeting discussed fundamental questions of 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, the cell cycle, or development; or how aberrant ubiquitylation contributes to disease. These questions are being elegantly addressed using rapidly evolving structural biology, cutting-edge library screening, or 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—despite the virtual format of the meeting—fostered deeply informed and creative discussion.

Two highlights of this meeting were its exciting Keynote Addresses. The first was delivered by David Sabatini, an HHMI investigator who spearheaded discovery of central nutrient sensing pathways. The second Address was delivered by Cynthia Wolberger, a department head at Johns Hopkins University, who has produced groundbreaking work on the molecular workings of protein machines on chromatin. The online format and the increasing interest in the ubiquitin field in relation to drug discovery has led to a record number of 552 participants who 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, the structure and the dissection of the workings of the UBR1 and Gid1 ubiquitin ligases; the coating of *Salmonella* by bacterial lipopolysaccharide as catalyzed by RNF213; molecular mechanisms of the reductive stress sensing in stem cells; new insights into the complex interplay of autophagy receptors and their substrates; and new and exciting approaches to isolate molecular glues tethering pathological proteins to E3 ubiquitin ligases. The meeting attracted many members of the rapidly increasing community of ubiquitin researchers in 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 virtual poster session, discussion meet-ups, and icebreakers. Notably, many talks were presented by graduate students and postdoctoral researchers. The collaborative and interdisciplinary nature of this field, now deeply relevant for drug discovery, was particularly obvious throughout this newly designed meeting, and its relevance will only increase once we can move back to personal meetings.

Support for this meeting was provided by the Corporate Sponsor Program.

PROGRAM

Keynote Address
D. Sabatini, *Whitehead Institute for Biomedical Research*

Ubiquitin and Disease
Chairpersons: I. Wertz, *Genentech, Inc., South San Francisco, California;* E. Fischer, *Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts*

Drug Discovery
Chairpersons: J. Bradner, *Novartis Institutes for Biomedical Research, Cambridge, Massachusetts;* C. Crews, *Yale University, New Haven, Connecticut*

Ubiquitin-dependent Signaling I
Chairpersons: Y. Merbl, *Weizmann Institute of Science, Rehovot, Israel;* S. Hur, *Harvard Medical School, Boston, Massachusetts*

Ubiquitin-dependent Signaling II
Chairpersons: W. Harper, *Harvard Medical School, Boston, Massachusetts;* D. Komander, *Walter and Elisa Hall Institute, Melbourne, Australia*

Quality Control/Autophagy I
Chairpersons: C. Joazeiro, *Heidelberg University, Germany;* J. Gestwicki, *University of California, San Francisco*

Autophagy II
Chairpersons: A. Simonsen, *University of Oslo, Norway;* R. Perera, *University of California, San Francisco*

Structure I
Chairpersons: B. Schulman, *Max Planck Institute of Biochemistry, Martinsried, Germany;* C. Lima, *HHMI/Sloan Kettering Institute, New York*

Keynote Address
C. Wolberger, *Johns Hopkins University School of Medicine, Baltimore, Maryland*

Structure II
Chairpersons: L. Passmore, *MRC Laboratory of Molecular Biology, Cambridge, United Kingdom;* A. Martin, *HHMI/University of California, Berkeley, California*

The Biology of Genomes (Virtual)

May 11–14 922 Participants

ARRANGED BY **Christina Curtis**, Stanford University, California
Hopi Hoekstra, Harvard University, Boston, Massachusetts
John Marioni, European Bioinformatics Institute, United Kingdom
Jay Shendure, University of Washington, St Louis, Missouri

Abstract book cover art

Because of the continued challenges associated with the COVID-19 pandemic, this 33rd meeting was held virtually (rather than in-person at CSHL) for the second year in a row. Although in-person meetings are preferred, CSHL continues to set the standard for how to execute a scientific meeting over Zoom, with additional enhancements introduced since last year. Overall, the meeting went smoothly, with robust attendance and sustained engagement across the several days of the meeting. Questions were asked throughout, with preference consistently given to trainees by session chairs. The Slack channel and social media chatter was also active. In addition to being the largest meeting in history (enabled by its virtual format), this likely was the most diverse lineup of speakers and discussion leaders of any BoG meeting to date.

The meeting kicked off on Tuesday with talks in Computational Genomics. The dissection of *cis*-regulation was a notable theme across several talks in this session, both leveraging natural variation in such regulation across cell types (Stein Aerts) as well as synthetic approaches to dissecting regulatory grammar (Jennifer Mitchell). Following the first poster presentations, the next session featured genomic work in a wide range of species in the Evolutionary and Non-Human Genomics session. A particularly notable talk came from session chairperson Hunter Fraser, who described the generation of human–chimpanzee hybrid stem cell lines, together with their application to investigate factors underlying human evolution.

Wednesday's sessions included Cancer and Medical Genomics as well as Developmental and Single-Cell Genomics. In the former session, a particularly striking talk came from Nick Banovich, on applying single-cell profiling to understand the mechanisms underlying patients with glioblastoma multiforme's differential responses to CAR T cell therapy. In the latter session, notable talks included "whole organism" dissection of cell types in a jellyfish (Tara Chari), new approaches for modeling cell state dynamics (Fabian Theis), and atlases of perturbed gene expression in patients who succumbed to SARS-CoV-2 (Aviv Regev).

The next day's sessions started with the Complex Traits and Microbiome session. This was followed by a session on Population Genomics. Notable talks from Steve McCarroll on protein repeat polymorphisms underlying GWAS hits and another from Ben Vernot on recovery and analysis of ancient human DNA from sediment. The day concluded with the ELSI Panel and Discussion on the topic of "Re-identification in Genomics."

The final day started with two invited Keynote Addresses given by Edith Heard and Michael Elowitz. Dr. Heard spoke about her seminal work on the epigenetics of X chromosome inactivation, and Dr. Elowitz discussed a remarkable new system for achieving synthetic multistability in mammalian cells. Following a final poster session, the meeting concluded with a session on *Functional Genomics*, which included terrific talks from trainees Zachary Chiang on in situ genomic sequencing and Sudarshan Pinglay on synthetic reconstruction of the Hox locus.

Overall, it was an outstanding meeting. The shifting landscape of genomics continues to be evident, with new technologies enabling increasingly sophisticated exploration and modeling of gene regulation, as well as of ancient and recent evolution. We can also see the increasing penetration of synthetic biology into the genomics field, which is exciting. Finally, we note that although the meeting went about as well as a virtual meeting could go, we are hopeful to be able to return to an in-person format at CSHL next year.

Support for this meeting was provided in part by the National Human Genome Research Institute, a branch of the National Institutes of Health; Arima Genomics; Oxford Nanopore Technologies; and PacBio.

PROGRAM

Computational Genomics
Chairpersons: S. Aerts, *University of Leuven, Belgium;* A. Tanay, *Weizmann Institute, Rehovot, Israel*

Evolutionary and Non-Human Genomics
Chairpersons: S. Edwards, *Harvard University, Cambridge, Massachusetts;* H. Fraser, *Stanford University, California*

Cancer and Medical Genomics
Chairpersons: C. Rotimi, *National Institutes of Health, Bethesda, Maryland;* S. Shah, *Memorial Sloan Kettering Cancer Center, New York*

Developmental and Single-Cell Genomics
Chairpersons: N. King, *HHMI/University of California, Berkeley;* F. Theis, *Helmholtz Zentrum München, Germany*

Complex Traits and Microbiome
Chairpersons: R. Ley, *Max Planck Institute for Developmental Biology, Tübingen, Germany;* G. Trynka, *Wellcome Sanger Institute, Cambridge, United Kingdom*

Population Genomics
Chairpersons: K. Harris, *University of Washington, Seattle;* P. Moorjani, *University of California, Berkeley*

ELSI Panel and Discussion: Re-Identification in Genomics—Risks, Realities, and Remedies. D. Kaufman, *National Human Genome Research Institute/NIH, Bethesda, Maryland*

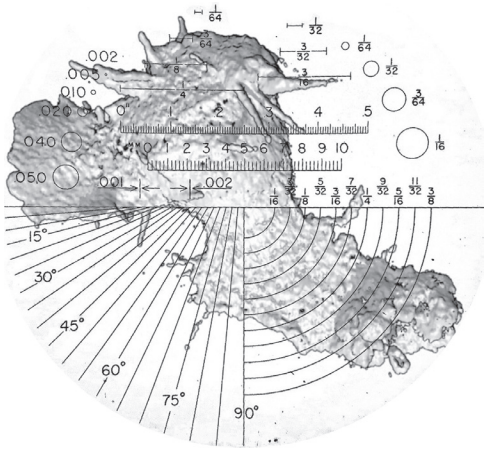
Guest Speakers
Chairpersons: E. Heard, *European Molecular Biology Laboratory, Heidelberg, Germany;* M. Elowitz, *California Institute of Technology, Pasadena*

Functional Genomics
Chairpersons: E. Furlong, *European Molecular Biology Laboratory (EMBL), Heidelberg, Germany;* J. Wysocka, *HHMI/Stanford University, California*

Cellular Dynamics and Models (Virtual)

May 19–21 125 Participants

ARRANGED BY **Hana El-Samad**, University of California, San Francisco
Carlos Lopez, Vanderbilt University, Nashville, Tennessee
Dyche Mullins, University of California, San Francisco
Susanne Rafelski, Allen Institute for Cell Biology, Seattle, Washington



Abstract book cover art

This meeting was held in virtual format and its structure adjusted slightly to account for the unique challenges posed by an online conference. Although the format changed slightly, the meeting maintained its historical emphasis on quantitative approaches to cell biology, including computational modeling, quantitative measurement, and mathematical analysis.

The meeting comprised two Keynote Addresses, seven themed plenary sessions, and three poster sessions. The themed sessions included 12 pre-invited speakers and several additional speakers selected based on their submitted abstracts. To encourage virtual participation at posters, the organizers included short "poster preview" talks in most of the plenary sessions, which gave every poster presenter who expressed interest in giving a talk the opportunity to share the highlights of their work with the full audience. These lively poster preview sessions, combined with CSHL's virtual poster session platform, enabled poster presenters to have productive direct interactions with other participants, even with the virtual format challenge.

The organizers were delighted at the extent to which conference participants interacted as a virtual community and the enthusiasm and collegiality that arose from this community, both in the oral presentation sessions and the poster sessions. Questions, discussions, and deep interest in the scientific content of the presentations, as well as support for trainee participation in this more constrained virtual format, were plentiful, and all participants were committed to making the most of an opportunity to be together to share in some great science.

The evolution of the "Computational Cell Biology" meeting, renamed several years ago to "Cellular Dynamics and Models," continued with significant representation from both the computational and experimental cell biologists and an ever-increasing proportion of the presented research being produced in a truly interdisciplinary way, integrating all parts of the computational-to-experimental spectrum. The Cellular Tradecraft session, launched in 2019, maintained its popularity and permitted participants to showcase some of the more technology- and method development–focused efforts that are so important for this interdisciplinary cell biology community. Because of the virtual format, several planned improvements to the in-person format from 2019 could not be incorporated, but the organizers look forward to the opportunity to do so in person at the next installment of this successful interdisciplinary meeting.

Support for this meeting was provided by the Corporate Sponsor Program.

PROGRAM

Cellular Engineering: Build to Understand
Chairpersons: C. Lopez, *Vanderbilt University, Nashville, Tennessee;* S. Rafelski, *Allen Institute for Cell Science, Seattle, Washington*

Keynote Address
M. Zernicka-Goetz, *University of Cambridge, United Kingdom*

Dynamic Control Cell and Tissue Organization
Chairpersons: S. Di Talia, *Stanford University, California;* S. Shvartsman, *Princeton University, New Jersey*

From Genes to Physiology: Multiscale Analysis and Modeling
Chairpersons: L. Bintu, *Stanford University, California;* A. Murugan, *Princeton University, New Jersey*

Evolutionary Cell Biology: Cells, Populations and Their Environment
Chairpersons: L. Laan, *Delft University of Technology, the Netherlands;* J. McCutcheon, *Arizona State University, Tempe*

Forces, Shapes, Movement (Cell Mechanics, Motility, and Cytoskeleton)
Chairpersons: E. Paluch, *University of Cambridge, United Kingdom;* G. Johnson, *Allen Institute for Cell Science, Seattle, Washington*

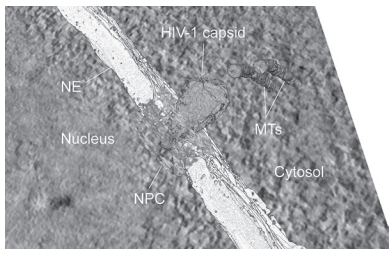
Keynote Address: Evolution of Morphogenesis
P. Tomancak, *Max Planck Institute of Molecular Cell Biology and Genetics, Germany*

Cellular Tradecraft (Workshop)
Chairperson: D. Mullins, *University of California, San Francisco*

Retroviruses (Virtual)

May 25–28 603 Participants

ARRANGED BY **Zandrea Ambrose**, University of Pittsburgh School of Medicine, Pennsylvania
Stuart Neil, King's College London, United Kingdom



Abstract book cover art

This meeting, considered the best for 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 been announced.

This year, eight sessions were grouped with an emphasis on the viral life cycle and key viral or host proteins important in retroviral infection. Groundbreaking presentations included the identification of the restriction factor SERINC3 as a lipid flippase that alters a specific lipid on the HIV-1 membrane to reduce its infectivity (featured on the back cover). The Keynote Addresses were given by Drs. Alice Telesnitsky and Hans-Georg Kräusslich. Dr. Telesnitsky provided a historical perspective on her work in retrovirology, which has focused largely on viral genome recombination, packaging, and transcription. Dr. Kräusslich gave an overview of his seminal imaging studies of retroviral capsids in cells, which was featured on the front cover. In addition, a special lunchtime symposium on the impact of COVID-19 on retrovirology research was well received by the audience. This included oral presentations by Dr. Ed Campbell on developing SARS-CoV-2 testing for his Illinois K-12 school district, Dr. Frank Maldarelli on COVID-19 impacts on HIV-1 patients and clinical research, and Dr. Richard Sever of Cold Spring Harbor Press on the growth in preprint server submissions over the past year.

The meeting was held virtually for the second year in a row. Similar to last year, the networking platform, Slack, was used to allow attendees to receive technical help, to facilitate introductions between scientists, to raise questions for oral presenters, and to direct messages. Each poster also had its own Slack channel to facilitate discussions between the presenters, many of whom were trainees, and other attendees. In addition, graduate students and postdoctoral fellows were offered small group Zoom meetings to connect with senior scientists attending the meeting for career advice. Modifications to the online format from last year included the ability of poster presenters to upload a short video presentation and availability of separate Zoom "Discussion Zones" after each session to allow face-to-face conversations between speakers and attendees.

The oral sessions were once again notable in the mix of presentations by both junior and senior researchers, an important feature of this meeting. As in past years, awards were presented to a distinguished postdoctoral scientist (Andy Kaplan Prize established in 2007), a distinguished senior graduate student (Uta von Schwedler Prize established in 2012), and best poster presentation of the meeting (Daniel Wolf Prize established in 2011). Attendees joined from 19 countries and 45% were female or nonbinary, reflecting the diversity of the retrovirology field.

Support for this meeting was provided in part by the National Institute of Allergy and Infectious Diseases, a branch of the National Institutes of Health.

PROGRAM

Entry and SERINCs
Chairpersons: R. Gummuluru, *Boston University School of Medicine, Massachusetts;* A. Kajaste-Rudnitski, *Ospedale San Raffaele, Milano, Italy*

Lunchtime Symposium: The "Other Virus"—Tales from a Year in COVID-19

Post-Entry/Capsid
Chairpersons: T. Hatziioannou, *The Rockefeller University, New York;* M. Naghavi, *Northwestern University, Chicago, Illinois*

Fifteenth Annual **Andy Kaplan Prize**
Presented by Z. Ambrose, *University of Pittsburgh School of Medicine, Pennsylvania*
Awarded to A. Dharan, *Loyola University, Chicago, Illinois*

Integration
Chairpersons: A. Engelman, *Dana-Farber Cancer Institute, Boston, Massachusetts;* Y-C. Ho, *Yale University School of Medicine, New Haven, Connecticut*

RNA and Vpr
Chairpersons: K. Boris-Lawrie, *University of Minnesota, St. Paul;* L. Wu, *University of Iowa, Iowa City*

Keynote Address
A. Telesnitsky, *University of Michigan, Ann Arbor*

Transcription and Latency
Chairpersons: M. Emerman, *Fred Hutchinson Cancer Research Center, Seattle, Washington;* N. Sherer, *University of Wisconsin, Madison*

Assembly and Release
Chairpersons: C. Carter, *Stony Brook University, New York;* A. Ono, *University of Michigan, Ann Arbor*

Keynote Address
H-G. Kräusslich, *University of Heidelberg, Germany*

Antiviral Factors and Accessory Proteins
Chairpersons: C. Berlioz-Torrent, *Institut Cochin-INSERM U1016, Paris, France;* M. Malim, *King's College London, United Kingdom*

Tenth Annual **Uta von Schwedler Prize for Retrovirology**
Presented by T. Hope, *Northwestern University, Evanston, Illinois;*
Awarded to L. Rheinemann, *University of Utah, Salt Lake City*

Eleventh Annual **Daniel Wolf Prize**
Awarded to best poster presentation

ERVs and Pathogenesis
Chairperson: W. Johnson, *Boston College, Massachusetts*

COVID/SARS CoV-2 Rapid Research Reports #6 (Virtual)

July 7–8 157 Participants

ARRANGED BY **Hung Fan**, University of California, Irvine
Adrian Hayday, King's College London, United Kingdom
Brenda Hogue, Arizona State University, Tucson
Volker Thiel, University of Bern, Switzerland

COVID/SARS CoV-2 Rapid Research Reports series brings together expert scientists on the forefront of COVID research into viral origins and variation, virus biology, host response, and antivirals and vaccines. Previous meetings in the series were held in June, July, August, and November 2020, and in January 2021. The series is designed to stimulate ideas and collaborations among scientists with the goal of hastening a solution to vanquish this worldwide disease. The full program for the sixth meeting is listed below.

Support for this meeting was provided by the Corporate Sponsor Program.

PROGRAM

Session I: Host–Virus Interactions/Structure

Co-Chairpersons: B. Hogue, *Arizona State University, Tucson*; L. Joshua-Tor, *Cold Spring Harbor Laboratory*

R. Neher, *University of Basel, Switzerland:* Tracking and predicting SARS-CoV-2 variants with Nextstrain.

D. Martin, *University of Cape Town, South Africa*: The emergence and ongoing convergent evolution of the N501Y lineages coincides with a major global shift in the SARS-CoV-2 selective landscape.

S. van der Werf, *Institut Pasteur, France*: SARS-CoV-2 evolution in humans extends host range to mice.

M. Parker, *The University of Sheffield, United Kingdom:* Transcriptomic changes in SARS-CoV-2 variants detected by large-scale subgenomic RNA analysis in clinical samples.

G. Towers, *University College London, United Kingdom*: Evolution of enhanced innate immune evasion by the SARS-CoV-2 B.1.1.7 UK variant.

P. Schuck, *National Institutes of Health, Bethesda, Maryland:* Energetic and structural features of SARS-CoV-2 N-protein co-assemblies with nucleic acids.

R. Martienssen, *Cold Spring Harbor Laboratory:* Circadian immunity at sunrise underlies the seasonality of Flu and COVID-19.

Abstract book cover art

Session II: Coronavirus Biology

Co-Chairpersons: L. Saif, *The Ohio State University, Columbus;* Y. Li, *Arizona State University, Tucson;* J.F. Drexler, *Charité–Universitätsmedizin Berlin, Germany:* Evolutionary origins of human coronaviruses.

L.J. Saif, *The Ohio State University, Columbus:* Origin and avian–mammalian interspecies transmission of porcine deltacoronavirus: A spillover scenario analogous to influenza A virus?

S. Schultz-Cherry, *St. Jude Children's Research Hospital:* Impact of obesity on influenza and SARS-CoV-2.

M. Beer, *Friedrich-Loeffler-Institut, Germany:* SARS-CoV-2 VOC characterization in animal models.

V. Thacker, *Ecole Polytechnique Federale de Lausanne, Switzerland*: A lung-on-chip model for SARS-CoV-2 endotheliitis.

E. Kastenhuber, *Weill Cornell Medical College*, *New York:* Coagulation factors and their effects on viral entry.

M. Santiago, *University of Colorado, Denver:* Interferon resistance of emerging SARS-CoV-2 variants.

Session III: Pathogenesis and the Immune Response

Co-Chairpersons: A. Hayday, *King's College London, United Kingdom;* T. Miura, *University of Idaho, Moscow*

J. Ordovas-Montañes, *Boston Children's Hospital, Massachusetts:* Single-cell biology of barrier tissues and COVID-19.

P.P. Cherepanov, *Francis Crick Institute, London, United Kingdom:* Interaction of SARS-CoV-2 spike with tetrapyrroles.

N.R. Roan, *J. David Gladstone Institutes/University of California, San Francisco:* T cell immunity to SARS-CoV-2.

A. Nath, *National Institute of Neurological Disorders and Stroke/NIH, Bethesda Maryland:* Long-haul COVID.

A. Jamieson, *Brown University, Providence, Rhode Island:* A novel role for the lung epithelium in driving SARS-CoV-2-specific coagulopathies.

J. Files, *University of Alabama, Birmingham:* Duration of post-COVID-19 symptoms are associated with sustained SARS-CoV-2-specific immune responses.

Session IV: Vaccines/Antivirals

Co-Chairpersons: V. Thiel, *University of Bern, Switzerland;* K. Jurado, *University of Pennsylvania, Perelman School of Medicine*

L. Tene, *Maccabi Research and Innovation Center, Israel:* Israel vaccine experience.

A. Walls, *University of Washington, St. Louis, Missouri:* Elicitation of broadly protective sarbecovirus immunity by receptor-binding domain nanoparticle vaccines.

P. Duprex, *University of Pittsburgh School of Medicine, Pennsylvania:* Inhalable nanobody.

B. Case, *Washington University, St. Louis, Missouri:* Ultrapotent miniproteins targeting the receptor-binding domain protect against SARS-CoV-2 infection and disease.

S. Byrareddy, *University of Nebraska Medical Center, Omaha:* Blockade of SARS-CoV-2 infection by targeting virus–host interaction using small molecule entry and PI3K-α/mTOR/BRD4 inhibitors.

K-O. Chang, *Kansas State University, Manhattan:* Postinfection treatment with a protease inhibitor increases survival of mice with a fatal SARS-CoV-2 infection.

F. Diaz-Griffero, *Albert Einstein College of Medicine, The Bronx, New York:* Glycosylated dyphillin as a broad-spectrum antiviral active against SARS-CoV-2.

ISHPSSB Biennial Meeting (Virtual)

July 13–20 451 participants

PROGRAM COMMITTEE CO-CHAIRS **Luis Campos**, University of New Mexico, Albuquerque
Roberta Millstein, University of California, Davis

VIRTUAL LOCAL ORGANIZATION COMMITTEE CHAIR **Matt Haber**, University of Utah, Salt Lake City

The International Society for the History, Philosophy, and Social Studies of Biology (ISHPSSB)'s biennial meeting was hosted virtually by Cold Spring Harbor Laboratory. ISHPSSB brings together scholars from diverse disciplinary backgrounds to discuss historical, conceptual, philosophical, political, social, cultural, institutional, and ethical issues of the life sciences in an open and informal setting.

Abstract book cover art

> A lot of what makes an ISH conference so successful lies, it seems to me, in the Society's endearingly awkward name. By being explicitly international, and also by embracing historians and philosophers and social scientists and all the scholars whose work lies across and between those categories, the Society thumbs its nose at borders, geographic and disciplinary. The spirit is one of welcome, wherever you come from. Yet within all the diversity there remains the unifying interest in biology. The result is a meeting which, at its best, offers unrivalled scope for sharing work in progress with fellow specialists while also maximizing chances for encountering new ideas and perspectives which are just the right distance from where you're at now to be surprising but stimulating. You leave the conference energized—and looking forward to the next one.
>
> *ISHPSSB President Greg Radick*

The conference consisted of familiar society events including organized sessions, individual papers, plenaries, and panels, and the society's general meeting. In response to the pandemical predicament, the conference also included a variety of "community-building" sessions, including icebreakers, happy office hours, professional development skill building sessions, and opportunities for network building.

Major sessions included:

CSHL WELCOME PLENARY PANEL

Greg Radick, Bruce Stillman, David Spector, Rebecca Leshan, Marsha Richmond, Nathaniel Comfort, Luis Campos

PLENARY SESSIONS

Open Science, Data Sharing, and Solidarity: Who Benefits?
Organizers: Ciara Staunton and Andrés Barragan

Revisiting Darwin's Descent (1871–2021)
Organizer and Chair: Thierry Hoquet

Viral Vulnerability: From Permissivity to Pessimism
Organizer and Chair: Eben Kirksey

INDIVIDUAL PAPERS SESSIONS

From "Adaptationism" to "Transmission, treatment, and therapy" (46 sessions)

TRADITIONAL SESSIONS

From "Actors and actors' categories in history, philosophy, and social studies of biology" (Chair: Robert Meunier) to "Wingspread birth and development of endocrine disruption science" (Chair: Ana Soto) (35 sessions)

DIVERSE FORMAT SESSIONS

From "A critical discussion of Darwin's argument by analogy: from artificial to natural selection" (Chair: Gregory Radick) to "'Your genitals don't lie!' An escorted encounter with the history and philosophy of phallic and cervical measuring" (Chair: Caterina Schürch) (21 sessions)

Plant Photobiology (Virtual)

July 22–25 260 Participants

ARRANGED BY **Beronda Montgomery**, Michigan State University, East Lansing
Dmitri Nusinow, Donald Danforth Plant Science Center, Olivette, Missouri
Ullas Pedmale, Cold Spring Harbor Laboratory
C. Robertson McClung, Dartmouth College, Hanover, New Hampshire

Abstract book cover art

The International Symposium on Plant Photobiology (ISPP) is a biennial meeting bringing together the brightest minds discussing breakthroughs into how light perception shapes plant form, development, and interorgan communication. ISPP, fondly known as the plant photobiology meeting, began 30 years ago and is solely organized on a volunteer basis without the backing of any scientific organization. Instead, the coordination of this meeting relies on volunteers from plant photobiology laboratories. This meeting showcased the latest developments and discoveries in a broad spectrum of areas ranging from the structural basis of light perception, environmental adaptation, light signaling, and optogenetics. Participants and presenters were a healthy mix of junior and senior lab heads, trainees, personnel from industries, and scientific journals. The organizers worked to ensure gender, nationality, and career-stage diversity in the invited speakers. The meeting stayed true to the vision and scope of ISPP and the spirit of Cold Spring Harbor Laboratory. As a result, presenters shared an exciting breadth of new, unpublished work illuminating recent discoveries, uncovered fundamental principles, and direct applications in the field.

The 2021 meeting brought researchers from North and South America, Europe, Asia, and Australasia to present their latest research. Invited presentations included a new understanding of the role of photoreceptors as thermosensors, how photoperception is linked to plant competition, seasonal and circadian rhythms, and new optogenetic tools that use light to manipulate cellular signaling.

The sessions that formed the basis of this meeting are listed below under Program. Across sessions, a central theme was the multifaceted integration of various signals of genes, proteins, and metabolites within organisms and between organisms and their environments. For example, some talks discussed the importance of communication between chloroplast and the nucleus and the nature of the signal. Some talks focused on carbon partitioning in algae and *Arabidopsis*. Other talks described the convergence of iron signaling and light signaling in plants during nutrient acquisition. Another main thread was related to engineering plants to improve the health of people and the planet. Examples included understanding how shade information from single leaves is transmitted throughout the plant and methods to suppress weeds in the field by muting shade responses. Additional topics included engineering plant phototropins and cryptochromes as tools for optogenetic applications in plant and animal systems.

Dr. Chandra Tucker and Dr. Richard Vierstra presented the Keynote Addresses. Dr. Chandra's talk focused on using plant photoreceptors as optogenetic tools with broad applications in cell biology, neurobiology, and the treatment of neuronal disorders. Dr. Vierstra's work presented the 3D structure of the plant phytochrome B red/far-red light photoreceptor, and how this structure sheds light on the control of many light and temperature responses. The work described has implications in plant engineering to mitigate harmful effects of global warming and to better understand downstream signaling.

Support for this meeting was provided in part by Agrisera; Percival; and Carl Zeiss Microscopy.

PROGRAM

Plenary Speaker
R.D. Vierstra, *Washington University, St. Louis, Missouri*

Chloroplast Biology
Chairperson: K. Dehesh, *University of California, Riverside*

Red Light Signaling I
Chairperson: U. Hoecker, *University of Cologne, Germany*

Red Light Signaling II
Chairperson: E. Huq, *University of Texas, Austin*

Circadian Rhythms
Chairperson: P. Mas, *Centre for Research in Agricultural Genomics (CRAG), Barcelona, Spain*

Blue and UV Light Signaling
Chairperson: H. Liu, *Chinese Academy of Sciences, Shanghai*

Light–Environment Interactions I
Chairperson: J. Casal, *Fundación Instituto Leloir, Buenos Aires, Argentina*

Light–Environment Interactions II
Chairperson: R. Ulm, *University of Geneva, Switzerland*

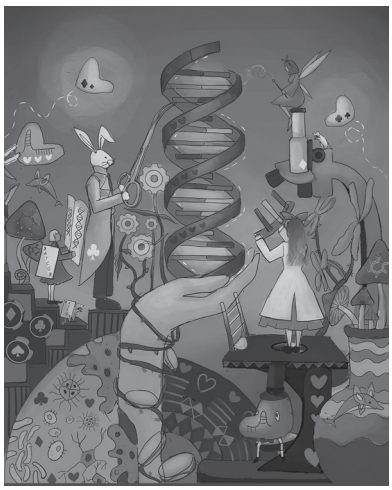
Plenary Speaker
C. Tucker, *University of Colorado Anschutz Medical Campus, Aurora*

Optogenics
Chairperson: M. Zurbriggen, *Heinrich Heine University, Düsseldorf, Germany*

Genome Engineering: CRISPR Frontiers (Virtual)

August 18–20 852 Participants

ARRANGED BY **Jennifer Doudna**, HHMI/University of California, Berkeley
Maria Jasin, Memorial Sloan Kettering Cancer Center, New York
David R. Liu, Broad Institute of Harvard and MIT, Cambridge, Massachusetts
Jonathan Weissman, Whitehead Institute and HHMI/MIT, Cambridge, Massachusetts



Abstract book cover art

This virtual meeting was the seventh consecutive conference in the series held at Cold Spring Harbor. The robust attendance continued to expand, reaching 852 participants. It was planned to hold the meeting again in 2022. This was a special year for the conference given that meeting co-organizer Jennifer Doudna was co-recipient of the 2020 Nobel Prize in Chemistry. She gave the Keynote Address, following which she and 2018 Nobelist in Chemistry Frances Arnold of CalTech, who is also part of President Biden's PCAST, had a broad-ranging discussion on a number of topics including using CRISPR to fight climate change and making CRISPR therapeutics affordable.

The goal of this meeting continued to be met in terms of bringing together researchers, virtually in this case, working in diverse fields to stimulate discussions and ideas to further exploit CRISPR and related technologies for biological discovery and medical applications. Session titles are listed below under Program. These diverse topics were covered by 23 speakers, including some who had not previously attended this meeting (e.g., Carol Gross, Kristin Knouse, Taekjip Ha, Lacra Bintu, and Lee Zou) and young faculty who had not previously spoken (e.g., Josh Modell and Neville Sanjana), and another approximately 33 speakers were chosen from submitted abstracts. In addition to talks using CRISPR nucleases, several talks discussed other CRISPR-derived approaches, including prime editing, base editing, and gene regulation (CRISPRa and CRISPi), as well as delivery options to select tissues (e.g., by lipid nanoparticles).

The percentage of female speakers has increased (43%) over previous years. Speakers represent diverse ethnic backgrounds, although there was only one African American, reflecting the low number in the field. The meeting continued to have a large number of corporate attendees. Included within the meeting were two workshops hosted by companies (Horizon Discovery and IDT).

One innovation was to solicit cover art for the conference abstract book. A number of potential covers that were submitted were of very high quality. The winning artwork had an Alice in Wonderland theme which incorporated CRISPR as enabling science that would previously have been thought to be unattainable or even unimagined.

Support for this meeting was provided in part by Integrated Data Technologies IDT; Horizon Discover; and Synthego.

PROGRAM

CRISPR Biology
Chairperson: C. Gross, *University of California, San Francisco*

Keynote Address
J. Doudna, *HHMI/Innovative Genomics Institute, Berkeley, California;* Interview with Walter Isaacson, author of *The Code Breaker*

Technology I
Chairperson: D. Liu, *Broad Institute of Harvard/MIT, Cambridge, Massachusetts*

Genomics/Screens
J. Weissman, *Whitehead Institute/MIT, Cambridge, Massachusetts*

Technology II
Chairpersons: N. Sanjana, *New York Genome Center, New York;* S. Sternberg, *Columbia University, New York*

DNA Repair
Chairpersons: B. Adamson, *Princeton University, New Jersey;* M. Jasin, *Memorial Sloan Kettering Cancer Center, New York*

Disease/Therapeutic
Chairperson: J.K. Joung, *Massachusetts General Hospital/ Harvard Medical School, Boston*

Eukaryotic mRNA Processing (Virtual)

August 24–27 439 Participants

ARRANGED BY **Tracy Johnson**, University of California, Los Angeles
Alberto Kornblihtt, University of Buenos Aires, Argentina
Karla Neugebauer, Yale University, New Haven, Connecticut
Yongsheng Shi, University of California, Irvine

This 13th meeting was held this summer to present and discuss recent developments in mRNA metabolism in eukaryotes. Because of the pandemic, the meeting was fully virtual. As per tradition for this meeting, oral presentations were solely selected from submitted abstracts focusing on unpublished work primarily from graduate students, postdoctoral researchers, and junior faculty. The sessions are listed below under Program. 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 159 posters, with a record 439 participants in total.

Abstract book cover art

Discussions through the different channels of the virtual platforms (Zoom Q&A, Zoom chat, Slack) were vibrant and fruitful. Organizers and participants felt that the use of Slack helped promote spontaneous 1:1 and group interactions that would otherwise have been lacking in a Zoom-only event.

The highlights in the Mechanisms of RNA Splicing session were studies to reveal the structure of the intron branch-point intermediates and the efforts to understand the mechanism and function of recursive splicing. Moreover, a number of exciting translational studies were presented that discussed a range of disorders associated with mutations in splicing factors. These were detailed in this session as well as the session on RNA Processing in Disease, providing tantalizing hints at opportunities for RNA therapeutics. In sessions on Alternative Splicing, there were presentations on mechanisms and networks of alternative splicing and alternative exon production as well as their biological consequences. Highlights include the new concept of hybrid exons, generated from alternative transcription start-site usage, backsplicing mechanisms that yield circular RNAs, and evidence for coordination of splicing and polyadenylation.

The 3′ End Processing session provided structural, molecular, and cellular insights into the mechanism and regulation of 3′ end formation of mRNAs and small noncoding RNAs. New, sophisticated techniques to globally identify RNA modifications and structures in cells were presented in the RNA Modification sessions. Sessions on Cotranscriptional RNA Processing focused on the timing and coordination between transcription and RNA processing events. There was important exploration of how strict the requirement is for splicing to be cotranscriptional and discussion of how this knowledge could contribute to splicing-correcting therapeutic strategies for human hereditary disease. 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 session of Viral RNAs focused on how viruses manipulate host and viral mRNAs to the viruses' advantage, and two sessions on RNA–RNA and RNA–Protein Interactions covered a number of areas of eukaryotic mRNA processing where RNA-protein complexes and RNA structures play central roles.

Research into RNA processing continues to bring important and surprising insights into cellular processes and disease!

Support for this meeting was provided in part by the National Cancer Institute, a branch of the National Institutes of Health.

PROGRAM

Mechanisms of RNA Splicing I
Chairperson: S. Fica, *University of Oxford, United Kingdom*

Mechanisms of RNA Splicing II
Chairperson: H. Madhani, *University of California, San Francisco*

Alternative Splicing I
Chairperson: K. Hertel, *University of California, Irvine*

Alternative Splicing II
Chairperson: A. Srebrow, *University of Buenos Aires, Argentina*

RNA Turnover and Quality Control I
Chairperson: S. Jagannathan, *University of Colorado Anschutz Medical Campus, Aurora*

RNA Turnover and Quality Control II
Chairperson: M. Muller McNicoll, *Goethe University Frankfurt, Germany*

Gene Regulation by RNA Modifications
Chairperson: K. Meyer, *Duke University School of Medicine, Durham, North Carolina*

RNA–RNA and RNA–Protein Interactions I
Chairperson: J. Cáceres, *MRC Human Genetics Unit, University of Edinburgh, United Kingdom*

RNA–RNA and RNA–Protein Interactions II
Chairperson: Y. Wan, *Genome Institute of Singapore, Singapore*

RNA–RNA and RNA–Protein Interactions III
Chairperson: O. Rissland, *University of Colorado School of Medicine, Aurora*

CoTranscriptional RNA Processing I
Chairperson: C. Fonseca, *University of Lisbon Medical School, Portugal*

CoTranscriptional RNA Processing II
Chairperson: H. Tilgner, *Weill Cornell Medicine, New York*

3′-End Processing I
Chairperson: L. Tong, *Columbia University, New York*

3′-End Processing II
Chairperson: C. Lima, *HHMI/Sloan Kettering Institute, New York*

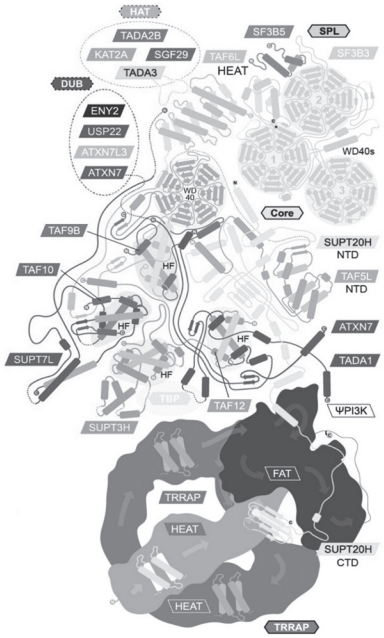
Viruses, Disease and RNA Processing I
Chairperson: A. Kalsotra, *University of Illinois, Urbana*

Viruses, Disease, and RNA Processing II
Chairperson: N. Kim, *Seoul National University, Korea*

Mechanisms of Eukaryotic Transcription (Virtual)

August 31–September 3 670 Participants

ARRANGED BY **Michael Levine**, Princeton University, New Jersey
Jane Mellor, University of Oxford, United Kingdom
Eva Nogales, University of California, Berkeley



Abstract book cover art

Transcription is the first step in gene expression, and its regulation plays a central role in the development and growth of eukaryotic organisms. The transcription field is very broad and requires a multitude of approaches, ranging from biochemistry, single-cell imaging, and structural biology, to functional genomics, stem cell research, and developmental biology. Advances from all of these areas were discussed at this meeting. Although many participants missed meeting in person, the virtual format enabled a more diverse group of scientists to contribute and participate in the meeting, which was widely appreciated and demonstrated the very strong interest in the field. It included eight live plenary sessions, recorded for future access, with eight talks each session via Zoom, complemented with Slack channels available for six weeks and dedicated to four poster sessions, poster chats, discussion forums, and answers to additional questions from the live post-talk Q&A sessions.

The meeting was structured around research areas (see below under Program) that are providing mechanistic insights into transcription in its broadest sense: Talks, including many from early-career scientists, illustrated the diversity of experimental approaches and range of organisms available to address mechanistic questions and allowed the audience to appreciate which areas in the field still challenge our understanding. Also demonstrated was how much insight can be gained from model organisms, challenging to the reluctance of some funders to support work in these systems or journals to publish such work. The Q&A format very successfully allowed for many more questions and elicited better audience engagement with each talk than in the traditional in-person meeting. This was reinforced by the discussion following each session, which enabled those in the audience to listen to and engage in dialog with and between the speakers. New areas addressed included what is meant by proximity in chromatin, the concept of topological operons, new insights into how weak interactions bring about precisely controlled transcriptional events, how kinetics influences all aspects of transcription, new functions for old favorites, new regulatory functions for RNA transcripts—all supported by remarkable structures, clever biochemistry, and nanomicroscopy in real time.

There were four poster sessions where participants had the opportunity to present their equally exiting unpublished research in transcription. Accessing the posters was clearly much more challenging for all participants, a common problem for most virtual meetings. However, the format does allow audience engagement for six weeks after the meeting, and it will be interesting to see what the feedback about these sessions ends up being. The meeting ran seamlessly, with (remarkably) only two online issues, otherwise quickly resolved by AV mission control and the Chairpersons. The Chairpersons should be thanked for their exemplary time keeping and handling of questions.

Support for this meeting was provided by the Corporate Sponsor Program.

PROGRAM

Higher-Order Chromatin
Chairperson: J. Mellor, *University of Oxford, United Kingdom*

Transcription Factors
Chairperson: A. Stark, *Research Institute of Molecular Pathology (IMP), Vienna, Austria*

Mediator/PIC/Coactivators
Chairperson: E. Nogales, *University of California, Berkeley*

Local Chromatin Environments
Chairperson: C. Wolberger, *Johns Hopkins University School of Medicine, Baltimore, Maryland*

Remodeling Chromatin
Chairperson: G. Narlikar, *University of California, San Francisco*

Early Transcriptional Events
Chairperson: P. Cramer, *Max Planck Institute for Biophysical Chemistry, Göttingen, Germany*

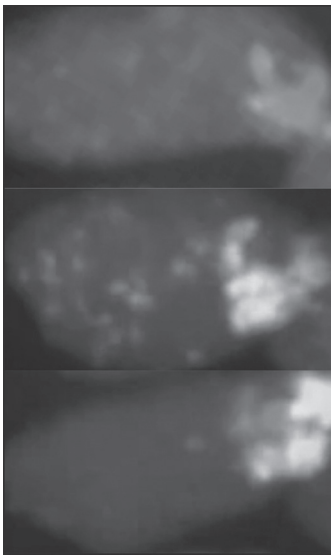
Disease and Development
Chairperson: S. Churchman, *Harvard Medical School, Boston, Massachusetts*

The 4D Nucleome
Chairperson: M. Levine, *Princeton University, New Jersey*

Eukaryotic DNA Replication and Genome Maintenance (Virtual)

September 8–12 531 Participants

ARRANGED BY **Karlene Cimprich**, Stanford University, California
John Diffley, The Francis Crick Institute, United Kingdom
Bob Duronio, University of North Carolina, Chapel Hill



Abstract book cover art

This 18th biennial meeting provided 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—and this year, the intersection of the fields was more evident than ever. In spite of the late stage of the pandemic and growing fatigue with Zoom meetings, we saw strong attendance with the number of participants 531. The percentage of female participants was 49% and participation of students and postdocs was also strong with 294 registered. The meeting featured 80 platform talks and more than 100 posters. The poster sessions were well attended in the virtual format and there was ample discussion. An added feature of the virtual format was an open discussion that followed each session and which allowed for continued interactions among participants. There were lively discussions at these sessions, and they were well attended and well received by presenters and viewers alike.

The ten platform sessions reflected the great breadth of the field, covering a wide range of topics from detailed molecular mechanisms to global regulation, genomic 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 that continued through the post-session discussions. Highlighted in the sessions covering Replication Origin Control in the Cell Cycle and Replication Initiation Factors and Origin Activation was a detailed understanding of how the replicative MCM helicase is loaded at origins using single-molecule approaches, cryo-EM, and ensemble biochemistry. High-resolution cryo-EM structures were presented showing how the initiating protein kinase, DDK, specifically targets the MCM double hexamer. There were several talks on how protein kinases, phosphatases and DNA binding proteins cooperate to produce the temporal program of origin firing, as well as several talks on exciting new technologies to map replication origins in mammalian genomes. The Developmental Control of Replication and Replication Timing and Chromatin and Genome/Nuclear Organization sessions highlighted how hormone signaling leads to developmentally programed gene amplification and examples of other mechanisms that alter canonical cell cycle control of replication. Advances in evaluating genomic data revealed how replication timing co-evolves with the emergence of new species, and advances in single-locus chromatin isolation is permitting a characterization of the different proteomes present at early and late replication regions of the genome. Molecular analyses of how chromatin impacts replication described how ORC, nucleosome remodeling complexes, and histone chaperones cooperate to organize nucleosomes at origins before and after replication initiation. Four sessions covered various aspects of processes critical for ensuring stable genome maintenance: Fork Stalling, Stabilization and Checkpoint Activation, Replication Stress and

Transcription, and Replication-Coupled Genome Instability and Disease. Among the highlights from these sessions was an increasingly clear view of mechanisms involved in fork protection and the pathways employed when protection is lost, and the emerging relationship between genome instability and the innate immune response. The ability to use a reconstituted replication system to study barriers to fork progression such as R-loops or secondary DNA structures was also elegantly demonstrated and reflected the continuing intersection of these fields and increased mechanistic understanding of that interface. The meeting concluded with a lively session on Fork Progression and Termination, which highlighted how DNA strand extrusion from the active CMG helicase blocks the recruitment of the E3 ligase that ubiquitylates MCM7 to trigger helicase unloading during replication termination, as well as reconstitution of a human replisome and initial studies on the control of replication fork progression.

This meeting continues to be the preeminent meeting in the field. The quality of the presentations and discussions indicates that this will continue to be the case going forward.

Support for this meeting was provided in part by the National Cancer Institute, a branch of the National Institutes of Health.

PROGRAM

Replication Origin Control in the Cell Cycle
Chairpersons: A. Donaldson, *University of Aberdeen, United Kingdom;* S. Prasanth, *University of Illinois, Urbana-Champaign*

Developmental Control of Replication and Replication Timing
Chairpersons: P. O'Farrell, *University of California, San Francisco;* J. Nordman, *Vanderbilt University, Nashville, Tennessee*

Chromatin and Genome/Nuclear Organization
Chairpersons: C. Alabert, *University of Dundee, United Kingdom;* F. Uhlmann, *The Francis Crick Institute, London, United Kingdom*

Replication Fork Stalling, Stabilization, and Checkpoint Activation
Chairpersons: A. Smogorzewska, *The Rockefeller University, New York;* A. van Oijen, *University of Wollongong, Australia*

Replication Initiation Factors and Origin Activation
Chairpersons: F. Bleichert, *Yale University School of Medicine, New Haven, Connecticut;* S. Bell, *Massachusetts Institute of Technology, Cambridge*

Replication-coupled Genome Instability, Cancer, and Disease
Chairpersons: R. Scully, *Beth Israel Deaconess Medical Center/ Harvard Medical School, Boston, Massachusetts;* X. Zhao, *Memorial Sloan Kettering Cancer Center, New York*

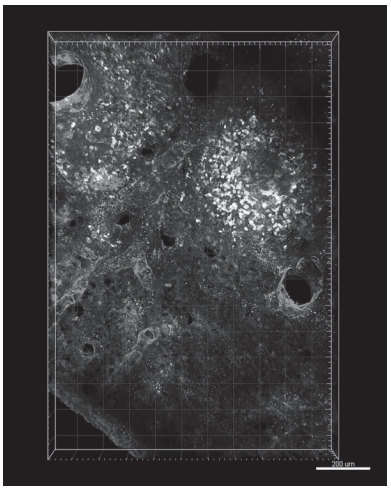
Replication Stress and Transcription
Chairpersons: H. Merrikh, *Vanderbilt University, Nashville, Tennessee;* P. Janscak, *University of Zürich, Switzerland*

Fork Progression and Termination
Chairpersons: X. Chen, *Johns Hopkins University, Baltimore, Maryland;* A. Gambus, *University of Birmingham, United Kingdom*

Microbial Pathogenesis and Host Response (Virtual)

September 21–24 392 Participants

ARRANGED BY **Denise Monack**, Stanford University, California
Anita Sil, University of California, San Francisco
Victor Torres, New York University School of Medicine, New York



Abstract book cover art

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 virtual 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 molecular microbiology, eukaryotic cell biology, immunology, and genomics, and 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. Areas covered included the evolution of microbial virulence, cell biology of pathogen infections, microbial toxins, regulation of virulence, mechanisms of innate immunity, and trans-kingdom interactions within hosts. A session entitled Phage Rage specifically focused on interactions between bacteria, mammalian hosts, and bacteriophage. Speakers for each session were a mixture of established leaders in the field and new investigators. Half of the speakers, including postdoctoral fellows and graduate students, were chosen from submitted abstracts. Dr. Joseph Heitman presented the Keynote Address. Dr. Heitman is an internationally recognized leader in molecular mechanisms of pathogenic fungi. His presentation provided an exciting story of how he has discovered a new mechanism conferring antifungal drug resistance in the human fungal pathogen *Mucor circinelloides*. He described how spontaneous resistance to the antifungal drug FK506 evolved via two distinct mechanisms. One involves Mendelian mutations and the other occurs via an epigenetic RNA interference–mediated pathway resulting in unstable, transient drug resistance. His findings suggest that the full impact of epimutations in fungal pathogens may have eluded discovery previously given their inherently unstable nature.

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 Q&A sessions via Zoom followed each oral presentation. The Zoom poster sessions were engaging with vibrant discussions. We strongly encouraged submission of abstracts by junior researchers in the field, and many young investigators were in attendance. These interactions have already produced fruitful scientific collaborations. This year also included opportunities for trainees to interact more formally with speakers and organizers at "Zoom P.I. Chats" and offered an informal gathering for trainees and junior faculty who are seeking to network for further training opportunities.

Support for this meeting was provided in part by funds from the National Institute of Allergy & Infectious Diseases, a branch of the National Institutes of Health.

PROGRAM

Phage Rage: Interactions between Bacteria, Mammalian Hosts, and Phage
Chairpersons: K. Maxwell, *University of Toronto, Canada;* A. Hryckowian, *University of Wisconsin, Madison*

Microbial Toxins and Effectors
Chairpersons: B. Lacy, *Vanderbilt University, Nashville, Tennessee;* J. Naglik, *King's College London, United Kingdom*

Host Adaptation and Pathogenesis
Chairpersons: P. Cotter, *University of North Carolina, Chapel Hill;* C. Hull, *University of Wisconsin, Madison*

Microbiome–Host Interactions in Development, Neurobiology, and Disease
Chairpersons: K. Guillemin, *University of Oregon, Eugene;* J. Oh, *the Jackson Laboratory, Farmington, Connecticut*

Cell Biology of Pathogen Infections
Chairpersons: H. Madhani, *University of California, San Francisco;* D. Soldati-Favre, *University of Geneva, Switzerland*

Innate Immunity during Infection
Chairpersons: I. Brodsky, *University of Pennsylvania, Philadelphia;* S. Shin, *University of Pennsylvania, Philadelphia*

Microbial Development and Signaling
Chairpersons: M. Kendall, *University of Virginia, Charlottesville;* C. Nobile, *University of California, Merced*

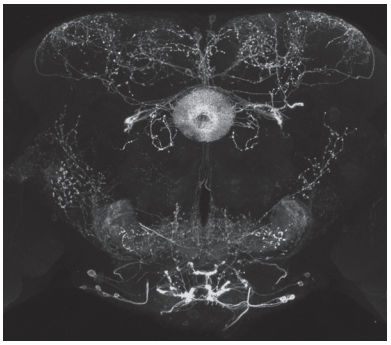
Trans-Kingdom Interactions within Hosts
Chairpersons: T. Hohl, *Memorial Sloan Kettering Cancer Center, New York;* K. Cadwell, *New York University Grossman School of Medicine, New York*

Keynote Address: RNA-dependent Epimutations Evoke Transient Antifungal Drug Resistance
J. Heitman, *Duke University, Durham, North Carolina.*

Neurobiology of *Drosophila* (Virtual)

October 5–8 638 Participants

ARRANGED BY **Adrian Rothenfluth**, University of Utah, Salt Lake City
Jill Wildonger, University of California, San Diego



Abstract book cover art

This meeting provided a forum for the discussion of new discoveries, techniques, and advances in *Drosophila* neurobiology. Fourteen sessions ran in series over four days with alternating platform and poster presentations and two professional development sessions. The virtual format of the meeting was optimally leveraged to provide platforms for both formal presentations and the informal interactions that are a highlight of this meeting. One way in which the online forum excelled was in speaker–audience interactions: Each speaker received an abundance of questions, likely because it was less intimidating to ask a question through a Zoom chat box than raising a hand in a large auditorium. The meeting's online Slack site had multiple channels that facilitated collaborations, exchange of reagents (e.g., antibodies, clones, mutants, and other stocks), methods (genetic, physiological, optical), and ideas among trainees and investigators. A virtual meet-up space hosted through wonder.me provided opportunities for one-on-one and small-group interactions. This combination of formal and informal interactions was especially important for scientists new to *Drosophila* and/or neurobiology and who benefit from learning of new advances in the field and opportunities to build their professional networks.

The eight platform session topics were chosen to reflect the areas in which cutting-edge advances are being made: the formation and function of circuits, synaptic transmission, the perception and integration of sensory cues to generate behavior, nervous system development and regeneration, neurological disease and injury, and emerging fields of neur-omics and neuroscience and evolution in non-*melanogaster* insects. Many people commented positively on the diversity of presentations and balance of research investigating the nervous system at different levels. Session chairs and the meeting organizers selected presenters for these platform sessions from submitted abstracts; the remaining abstracts were presented as posters. The majority of the speakers were graduate students and postdoctoral fellows. Demographically, slightly more than half the presenters were female, and approximately a third of abstracts submitted by members of historically underrepresented groups were selected for talks, which matches with the overall talk selection average. Overall, ~10% of the talks were presented by attendees who self-identified as Hispanic, Black, American Indian/Native American, and/or Native Hawaiian/Pacific Islander. The Seymour Benzer Lecture was presented by Leslie Vosshall (Robin Chemers Neustein Professor). She presented an engaging account of her seminal and ongoing research into the molecular neurobiology of mosquito host-seeking behavior. 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. Jenny Lu, who trained as a graduate student with Dr. Rachel Wilson at Harvard, and who presented her exceptional work on neural circuit mechanisms that relate body- and world-centric positions in navigating flies. This meeting was extremely well attended, with presentations spanning the breadth of modern neurobiology. The interactions and career development opportunities fostered by this meeting are sure to enhance this vibrant field.

Support for this meeting was provided by the National Institute of Neurological Disorders and Stroke, a branch of the National Institutes of Health.

PROGRAM

Circuit Formation and Function
Chairperson: A. Seeds, *University of Puerto Rico Medical Sciences Campus, San Juan*

Professional Development Session
Chairperson: F. Maderspacher, Senior Editor, *Current Biology*

(Re)Generation of Neurons and Glia
Chairperson: M. Kohwi, *Columbia University, New York*

Sensory Integration and Behavior
Chairperson: K. Nagel, *New York University School of Medicine, New York*

Technological Innovations
Chairperson: M. Louis, *University of California, Santa Barbara*

Elkins Award Memorial Lecture: Transforming Representations of Movement from Body to World-Centric Space
J. Lu, *Harvard Medical School, Boston, Massachusetts*

Neuroscience and Evolution in Non-*melanogaster* Insects
Chairperson: O. Riabinina, *Durham University, United Kingdom*

Neurological Disease and Injury
Chairperson: T. Lloyd, *Johns Hopkins University, Baltimore, Maryland*

Seymour Benzer Lecture
L. Vosshall, *The Rockefeller University, New York*

Synaptic Transmission and Plasticity
Chairperson: R. Carrillo, *University of Chicago, Illinois*

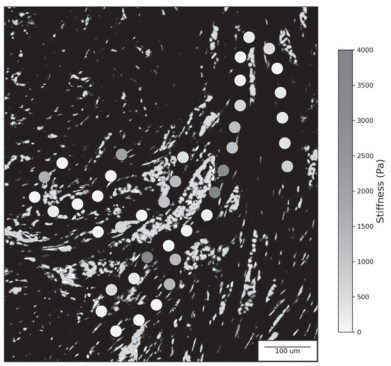
Professional Development Session
Chairperson: D. Miller, *Program Director, National Institute of Neurological Disorders and Stroke/NIH, Bethesda, Maryland*

Neur-Omics
Chairperson: A. Crocker, *Middlebury College, Vermont*

Biology of Cancer: Microenvironment and Metastasis (Virtual)

October 12–15 484 Participants

ARRANGED BY **Scott Lowe**, Memorial Sloan Kettering Cancer Center, New York
M. Celeste Simon, University of Pennsylvania Medical School, Philadelphia
Valerie Weaver, University of California, San Francisco



Abstract book cover art

Tumors are evolving organs composed of transformed cells surrounded by a complex tissue stroma. The tumor stroma consists of cellular (immune, stromal fibroblast, nerve, adipocyte, vascular endothelial cells, lymphatic cells) and a noncellular (extracellular matrix, exosome, hypoxia, soluble factors) components. Emerging studies demonstrate how the dynamic interaction between the transformed cells and the cellular and noncellular stroma promotes malignant transformation, drives tumor aggression and metastasis, and modifies treatment efficacy. Furthermore, there is a growing recognition that the primary tumor stroma and metastatic stroma exhibit distinct and heterogeneous properties that are predicted to govern metastatic progression. This meeting aimed to capture this phenotypic behavior of the tumor organ by considering the cancer program from an organ and systemic perspective. The conference highlighted emerging topics currently being investigated by cancer researchers around the world, bringing together cancer cell and molecular biology experts, immunologists and systems biologists to address a diverse array of topics including stress and metabolic adaptation, tumor hypoxia, tumor immunology, cancer dormancy, tumor evolution, tumor heterogeneity, metastasis, and tumor innervation. All of the talks were of extremely high caliber, and this prompted an extended and insightful discussion from the audience that was facilitated through both a Q&A online format and a subsequent speaker discussion forum. The first Keynote Address was dedicated to the memory of the tumor microenvironment pioneer Zena Werb, University of California, San Francisco. Lisa Coussens discussed Dr. Werb's groundbreaking studies implicating innate immunity in tumor progression and treatment focusing on current clinical trials and work demonstrating the impact of manipulating myeloid cells on antitumor immunity and therapy response. The second Keynote Address was given by the esteemed Nobel Prize awardee William G. Kaelin, Jr., who discussed the historical role demonstrated for tumor hypoxia and regulation of the key HIF transcription factors and elaborated on the field's current understanding of their role in malignancy and treatment response. Given the current interest in single-cell analysis to explore the role of cancer and stromal cell heterogeneity, an afternoon single-cell session chaired by Scott Lowe was incorporated into Wednesday's meeting with special lectures by Aviv Regev from Genentech and Dana Pe'er from MSKCC. In spite of the limitations imposed by the virtual meeting format, we are confident that, overall, the participants enjoyed the diversity, timeliness, and depth of the topics covered and were proud to be part of this vibrant community.

Support for this meeting was provided by the National Cancer Institute, a branch of the National Institutes of Health.

PROGRAM

Cross Talk between the Tumor and Microenvironment
Chairpersons: D. Tuveson, *Cold Spring Harbor Laboratory*; C. Abate-Shen, *Columbia University, New York*

Keynote Address
L. Coussens, *Oregon Health & Science University, Portland*

Immune Surveillance
Chairpersons: J. Powell, *Johns Hopkins University, Baltimore, Maryland;* T. Gajewski, *University of Chicago, Illinois*

Cancer Heterogeneity and Plasticity
Chairpersons: C. Curtis, *Stanford University, California;* E. Batlle, *IRB Barcelona, Spain*

Special Single-Cell Session
Chairperson: S. Lowe, *Memorial Sloan Kettering Cancer Center, New York*

Metastatic Niche
Chairpersons: R. Perera, *University of California, San Francisco;* P. Schedin, *Oregon Health & Science University, Portland*

Inflammation
Chairpersons: J. Rathmell, *Vanderbilt University School of Medicine, Nashville, Tennessee;* K. de Visser, *The Netherlands Cancer Institute, Amsterdam*

Keynote Address
W. Kaelin, *HHMI/Dana-Farber Cancer Institute/Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts*

Systemic Factors
Chairpersons: J. Borniger, *Cold Spring Harbor Laboratory;* B. Spiegelman, *Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts*

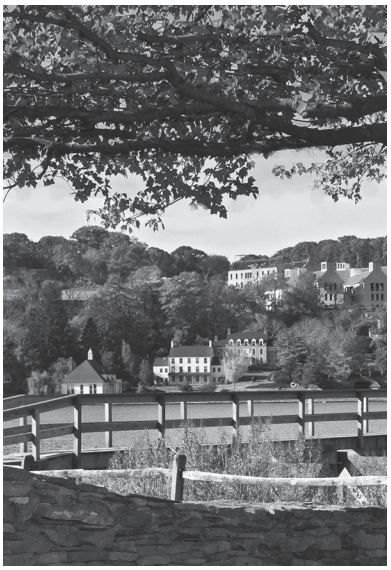
Stromal Control of Tumorigenesis
Chairpersons: M. Monje-Deisseroth, *Stanford University, California;* A. Kimmelman, *New York University Langone Health, New York*

The Stressful Tumor Microenvironment
Chairpersons: M. Vander Heiden, *Massachusetts Institute of Technology, Cambridge;* K. Vousden, *The Francis Crick Institute, London, United Kingdom*

Mechanisms of Metabolic Signaling (Virtual)

October 26–29 311 Participants

ARRANGED BY **Susanne Mandrup**, University of Southern Denmark, Odense
Jared Rutter, HHMI/University of Utah School of Medicine, Salt Lake City
Kathryn (Katy) Wellen, University of Pennsylvania, Philadelphia



Abstract book cover art

This virtual fifth meeting was held on the Zoom platform, and followed a highly successful Mechanisms of Metabolic Signaling meeting held in person 2.5 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 28 invited speakers are leaders in the various aspects of metabolic research from all over the world, and the majority had not been part of the 2017 or 2019 programs. The meeting included two inspiring Keynote Addresses by Aviv Regev and Karen Vousden. Dr. Regev described how her impressive single-cell multi-omics technologies can be applied to solve important questions in metabolism and cancer. Dr. Vousden told an impressive story about the importance of serine metabolism in cancer cells. The eight sessions featured oral presentations selected from the submitted abstracts, all highlighting unpublished research and focusing 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, as well as emerging technologies to study metabolism, 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 program. In all there were 51 talks by speakers from Canada, Europe, Australia, Asia, and the United States; 21 of the talks were given by women.

In addition to the eight oral sessions, an online poster session featured a total of 101 posters, using the Slack platform. There was engagement from most of the attendees for most of the posters. All eight sessions were characterized by open and wide-ranging discussions, both immediately after each talk and in the 30-minute Discussion Zone held after each session. Slack also facilitated further discussion opportunities. 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.

Support for this meeting was provided in part by the National Institute of Diabetes and Digestive and Kidney Diseases, a branch of the National Institutes of Health.

PROGRAM

Metabolism and Cell Proliferation
Chairperson: K. Wellen, *University of Pennsylvania, Philadelphia*

Emerging Technologies
Chairperson: J. Estall, *IRCM, Montréal, Canada*

Keynote Address and Workshop
S. Mandrup, *University of Southern Denmark, Odense*

Nutrient Utilization
Chairperson: E. Taylor, *University of Iowa, Iowa City*

Metabolic Cross Talk with Genome and Epigenome
Chairperson: L. Finley, *Memorial Sloan Kettering Cancer Center, New York*

Metabolism in Specialized Cell Types
Chairperson: A. Kelekar, *University of Minnesota, Minneapolis*

Inter-Organ Cross Talk
Chairperson: J. Rutter, *University of Utah School of Medicine, Salt Lake City*

Metabolite Signaling
Chairperson: J. Ye, *Stanford University, California*

Genome Informatics (Virtual)

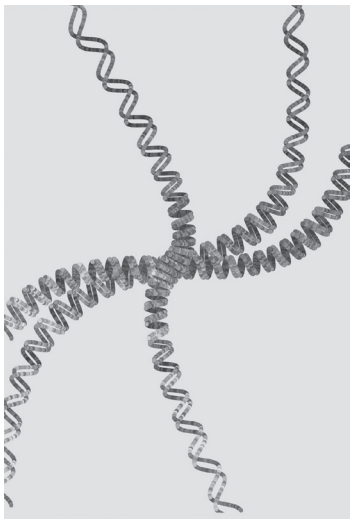
November 3–5 408 Participants

ARRANGED BY **Joanna Kelley**, Washington State University, Pullman
Pall Melsted, University of Iceland, Reykjavik
Oliver Stegle, German Cancer Research Center, Heidelberg

This 16th 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 400 registrants, presenting 171 abstracts, and offering a snapshot of the latest developments in the field. There were 10 invited talks by session chairs, and two Keynote Addresses. The remaining 26 talks were all selected for presentation by session chairs from openly submitted abstracts, and we also had 143 posters presented.

This year, abstracts covered a wide variety of genomic analyses, with a special emphasis on innovations in genetic variant discovery, functional genomics, and new insights gained from integrating large-scale genomics data sets. There was also an emphasis on new developments in metagenomics and single-cell analysis. A list of the sessions is given below under Program.

The first Keynote Address, delivered by Dr. Mark Blaxter, focused on the extensive genome sequencing and biological insights from the Sanger Institute's Tree of Life Programme. The second Keynote Address, delivered by Dr. Barbara Engelhardt, discussed new statistical analysis approaches, with an emphasis on applications in single-cell genomics. Both talks were extremely well received and led to extensive follow-up discussions.



Abstract book cover art

This meeting was very actively discussed on Twitter (using hashtag #gi2021), and the Slack channel was a success for additional discussion and poster presentations. More than 2000 messages were broadcast by users around the world to discuss and debate the ideas presented.

Support for this meeting was provided in part by the National Human Genome Research Institute, a branch of the National Institutes of Health.

PROGRAM

Sequencing Algorithms, Variant Discovery, and Genome Assembly
Chairpersons: A. Phillippy, *National Human Genome Research Institute, Bethesda, Maryland;* L. Salmela, *University of Helsinki, Finland*

Epigenetics and Genome Structure
Chairpersons: M. Colomé-Tatché, *Helmholtz Zentrum München, Germany;* T. Marschall, *Heinrich-Heine University, Düsseldorf, Germany*

Functional Genomics
Chairpersons: C. Leslie, *Memorial Sloan Kettering Cancer Center, New York;* J. Listgarten, *University of California, Berkeley*

Keynote Address
M. Blaxter, *Wellcome Sanger Institute, Hinxton, United Kingdom.*

Evolution, Complex Traits and Phylogenetics
Chairpersons: J. González, *Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain;* E. Holmes, *University of Sydney, Australia*

Microbial and Metagenomics
Chairpersons: Z. Iqbal, *European Bioinformatics Institute, Cambridge, United Kingdom;* D. Kempa, *Stony Brook University, New York*

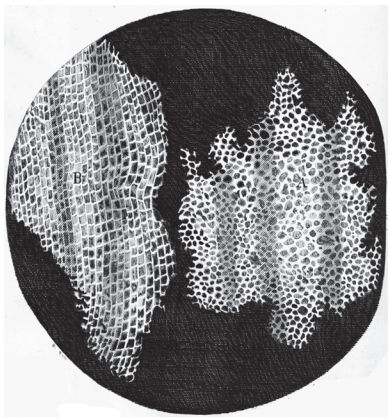
Keynote Address
B. Engelhardt, *Princeton University, New Jersey*

Single-Cell Genomics
Chairpersons: H. Tilgner, *Weill Cornell Medicine, New York;* A. Conesa, *Institute for Integrative Systems Biology, CSIC, Valencia, Spain*

Single-Cell Analyses (Virtual)

November 10–12 582 Participants

ARRANGED BY **Scott Fraser**, University of Southern California, Los Angeles
Junhyong Kim, University of Pennsylvania, Philadelphia
Tatjana Sauka-Spengler, University of Oxford, United Kingdom
Aaron Streets, University of California, Berkeley



Abstract book cover art

The goal of this sixth 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. Because of the pandemic, the meeting was held in fully virtual form—but the virtual format allowed more than 500 participants with a typical audience of 200–300 for every talk, more than doubling the participant numbers from prepandemic 2019. This year, we had two Keynote Addresses, 13 invited speakers, 22 contributed speakers, 18 two-minute flash-talks, and a total of 91 posters. The first Keynote Address, given by Dr. Long Cai, a pioneer of multiplexed FISH for spatial transcriptomics, spoke about the latest developments in spatial single-cell biology and his perspectives on key conceptual questions in single-cell biology. The second Keynote Address was given by Katy Borner, who leads the informatics groups for the NIH Human Biomolecular Atlas Project (HuBMAP). Dr. Borner gave a broad-scope talk on data models, visualizing data, and HuBMAP's data coordination efforts. As in previous years, much of the meeting concentrated on covering new technologies and methods loosely organized under the topics listed below under Program. The talks covered continued remarkable advances in single-cell methods, especially increasing the measurement modalities and spatial assays. Applications of single-cell techniques continued to expand into novel domains, including microbiology and plant development. This year also had a special invited speaker session with leaders from various cell atlas projects including Cori Bargmann from the Chan Zuckerberg Initiative, Sarah Teichmann from the Human Cell Atlas project, and John Ngai from NIH's BRAIN initiative. All speakers presented the current status of these large-scale collaborative projects and future prospects of these projects. As a virtual meeting, this year's meeting lacked one of the key characteristics of CSHL Single Cell Analysis meeting: that of catalyzing collaborative interactions among participants from very different fields. Nevertheless, the participants were still universally appreciative of the special nature of this meeting as a place for biomedical scientists, engineers, chemists, physicists, and computational researchers to come together. The meeting solidified single-cell biology as one of the most exciting areas of biomedical sciences with unabated growth in technologies, impact, and interest.

Support for this meeting was provided in part by 10X Genomics.

PROGRAM

Single-Cell Perturbations
Chairperson: S. Domcke, *University of Washington, Seattle*

Multimodal Measurement
Chairperson: F. Gaiti, *Weill Cornell Medicine, New York*

Keynote Address
L. Cai, *California Institute of Technology, Pasadena*

Data Visualizations and Single-Cell Applications
Chairperson: A. Gayoso, *University of California, Berkeley*

Spatial Omics I
Chairperson: R. Cotrim Chaves, *University of California, Berkeley*

Regulation and Development
Chairperson: S. Wiedenmann, *Helmholtz Zentrum Muenchen, Neuherberg, Germany*

Keynote Address
K. Börner, *Indiana University, Bloomington*

Single-Cell Phenomics
Chairperson: O. Arguello-Miranda, *UT Southwestern Medical Center, Dallas*

Spatial Omics II
Chairperson: A. Gupta, *University of California, Berkeley*

New Platforms I
Chairperson: A. Maslan, *University of California, Berkeley*

New Platforms II
Chairperson: Z. Tatarova, *Oregon Health & Science University, Portland*

Building the Cell Atlas
Chairperson: A. Streets, *University of California, Berkeley*

Plant Genomes, Systems Biology, and Engineering (Virtual)

December 1–3 202 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, DC

Abstract book cover art

This 13th 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 were presented.

The most notable trend at this conference relative to its predecessors was the cross-fertilization of concepts and approaches across the seven sessions. Several talks presented new work that links metabolism to defense and development. Others talked about the overlap between pathogenic and symbiotic interactions between plants and microbes. As the genes underlying the mechanisms of biological processes are revealed, we anticipate seeing more examples of such interdependencies.

Another prominent thread in the meeting was the power of emerging technologies for addressing long-standing problems in plant biology. Examples include talks on "transgenerational inheritance of DNA methylation during adaptation to low temperature in rice"; how epigenomic divergence between *Arabidopsis thaliana* paralogs underlies changes in mutation rates and sequence evolution; identifying insulator sequences with plant STARR-seq; comprehensive understanding of apical spikelet abortion in barley inflorescence by tissue-specific multi-omics analyses; high-resolution gene network analyses revealing the significance of developmental networks in directing cell type–specific immunity; insights to molecular mechanisms of plant defense priming, a genome-scale; context-specific gene regulatory network governing the balance of growth- and defense-associated metabolism in *A. thaliana;* interrogating plant hormone and stress responses; developing systems and synthetic biology tools to understand and engineer tomato transcriptional regulation of nitrate and phosphate response; interrogating plant hormone and stress responses at cellular resolution using time series single-cell transcriptomics; identifying *cis*-regulatory DNA using machine learning; decoding and recoding lateral root development using integrase switches; and multicomponent engineering of *Cyanobacteria* toward a photoproduction platform for plant terpenes.

Two Keynote Addresses were presented by Dr. Xumei Chen and Dr. Rob Jackson. Dr. Chen focused on new technologies and approaches that revealed NAD-based noncanonical RNA capping, an emerging posttranscriptional regulation that links small-molecule metabolism to gene expression and hormone signaling pathways. Dr. Jackson presented a timely and eye-opening overview of the global geochemical cycle and its relationship to the biosphere and in particular agriculture. He emphasized the growing importance of plants as terrestrial sinks of carbon dioxide and described the importance of incorporating methane and nitrous oxide in the global carbon cycling studies. The work described has implications on plant engineering for mitigating global warming. Notably, although the ongoing COVID-19 pandemic required a last-minute change

to a virtual meeting, participation was higher than prior in-person meetings, and the virtual format facilitated numerous opportunities for interactions between diverse scientists from across the globe.

Support for this meeting was provided by the Corporate Sponsor Program.

PROGRAM

Frontier Technologies and Synthetic Biology
Chairperson: M. Frank, *Cornell University, Ithaca, New York*

Metabolic Circuits
Chairperson: R. Lozano-Duran, *Eberhard Karls University, Tübingen, Germany*

Crop Biology and Trait Enhancement
Chairperson: L. Dixon, *University of Leeds, United Kingdom*

Plant–Microbe
S. Hogenhout, *John Innes Centre, Norwich, United Kingdom*

Keynote Address
X. Chen, *University of California, Riverside*

Development: Modules to Networks
Chairperson: N. Geldner, *University of Lausanne, Switzerland*

Biodiversity and Environmental Adaptation
Chairperson: A. Hancock, *Max Planck Institute for Plant Breeding Research, Cologne, Germany*

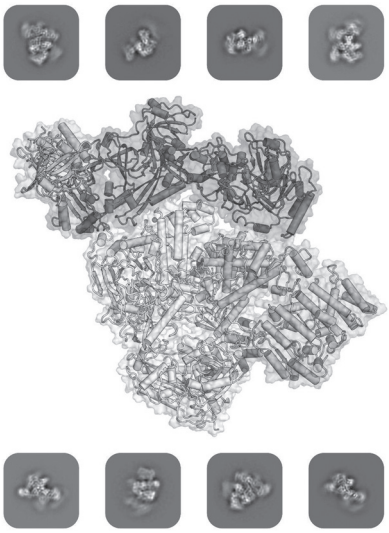
Keynote Address
R. Jackson, *Stanford University, California*

Genomes and Epigenomes
Chairperson: N. Stein, *Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany*

Telomeres and Telomerase (Virtual)

December 14–17 469 Participants

ARRANGED BY **Steven Artandi**, Stanford University, California
Jan Karlseder, The Salk Institute, California
Debbie Wuttke, University of Colorado, Boulder



Abstract book cover art

This 12th conference has been held every two years from 1999 onward. The conference consisted of eight sessions of talks and four poster sessions. Because of the COVID pandemic, we postponed this meeting from the normal spring timeframe to December and were forced to pivot to a virtual format. For session chairpersons, we chose a mix of established scientists and younger scientists who had already made their mark as independent investigators. Session chairs were given the choice between giving a 12-minute presentation themselves or having a member of their lab give a talk. The remainder of the talks (also 12 min), primarily by graduate students and postdoctoral fellows, were chosen from submitted abstracts, allowing as many presentations as possible. There were 64 talks and approximately 100 posters at the meeting. Each half-day program was followed by a 45-minute discussion session, and these were well attended, with up to 110 participants contributing. The discussion sessions were very productive and allowed participants to ask additional questions of the speakers while providing for a lively interaction on key concepts that emerged from the oral sessions. For the meeting overall, there were a record 469 participants, 53% of whom were female and 8.9% from under-represented minority groups.

The talks and posters covered diverse aspects of telomere and telomerase biology: chromosome and genome stability; telomerase biogenesis and regulation; telomere protection and DNA-damage signaling; telomere function during the cell cycle and cell death regulation; regulation of immortality by telomerase; regulation of immortality by ALT; and telomere biology disorders, premature aging and cancer predisposition.

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 2023.

Support for this meeting was provided in part by the National Institute on Aging, a branch of the National Institutes of Health.

PROGRAM

Chromosome and Genome Stability
Chairpersons: D. Baird, *Cardiff University, United Kingdom;* N. Arnoult, *University of Colorado, Boulder*

Replication Stress
Chairpersons: C. Greider, *University of California, Santa Cruz;* E. Fouquerel, *Thomas Jefferson University, Philadelphia, Pennsylvania*

Telomerase Biogenesis and Regulation
Chairpersons: P. Baumann, *Institute of Molecular Biology, Mainz, Germany;* M. Stone, *University of California, Santa Cruz*

Telomere Protection and DNA Damage Signaling
Chairpersons: J.K. Nandakumar, *University of Michigan, Ann Arbor;* E. Cusanelli, *University of Trento, Italy*

Telomere Function during the Cell Cycle and Cell Death Regulation
Chairpersons: T. de Lange, *The Rockefeller University, New York;* T. Cech, *HHMI/University of Colorado, Boulder*

Regulation of Immortality by Telomerase
Chairpersons: M.T. Teixeira, *CNRS, Sorbonne University, Paris, France;* J. Cooper, *University of Colorado Anschutz Medical Campus, Aurora*

Regulation of Immortality by ALT
Chairpersons: E. Lazzerini Denchi, *National Cancer Institute, Bethesda, Maryland;* C. Autexier, *McGill University, Montréal, Canada*

Telomeropathies, Premature Aging, and Cancer Predisposition
Chairpersons: S. Boulton, *The Francis Crick Institute, London, United Kingdom;* S. Savage, *National Cancer Institute, Bethesda, Maryland*

POSTGRADUATE COURSES

Ion Channels in Synaptic and Neural Circuit Physiology (Virtual)

June 7–18

INSTRUCTORS
C. Grienberger, Brandeis University, Waltham, Massachusetts
J. Grundemann, University of Basel, Switzerland
A. Lampert, RWTH Aachen University, Germany
A. Scimemi, University at Albany, New York
N. Wanaverbecq, Aix Marseille University, France

CO-INSTRUCTOR
S. Krabbe, Janelia Research Campus, Ashburn, Virginia/Friedrich Miescher Institute, Basel, Switzerland

ASSISTANTS
N. Affinnih, University at Albany, New York
R. Bott, RWTH Aachen University, Germany
A. Erickson, RWTH Aachen University, Germany
J. Korner, RWTH Aachen University, Germany
M. Petroccione, SUNY Albany, New York

Ion channels are the fundamental building blocks of excitability in the nervous system. The primary goal of this course was to provide an overview of the different biophysical properties of ion channels that enable neurons to perform unique physiological functions in a variety of systems. Areas of particular interest included voltage- and ligand-gated ion channels at central and peripheral synapses; synaptic integration and plasticity; neural circuit function in vitro and in vivo; in vivo recording using Neuropixel recordings; and optogenetic strategies for circuit manipulation.

A typical day consisted of lectures, a data blitz allowing students to present their own work, and one-on-one sessions with instructors to discuss research projects and methodological approaches and to give practical advice. The course provided students with conceptual and practical knowledge about the use of patch-clamp electrophysiology to examine single-channel activity in cultured cells, ion channel biophysics in acutely dissociated neurons and synaptic integration in slices, and plasticity and circuit dynamics in in vitro slices and in vivo preparations.

Discussed were different recording modes and configurations (e.g., cell-attached, whole-cell dendritic, and somatic patch clamp recordings) and the advantages and limitations of each method in relation to specific scientific questions. The course also covered circuit manipulation techniques, in vitro and in vivo (i.e., pharmacological, electrophysiological, optogenetics, Neuropixel probes). Admissions priority was given to students and postdocs who showed a clear interest and specific plans to apply these techniques to a defined scientific problem.

This course was supported in part by grants from Helmsley Charitable Trust and the Howard Hughes Medical Institute through the Science Education Program. Scholarship support was provided by Regeneron Pharmaceuticals and the International Brain Research Organization.

PARTICIPANTS

Acevedo-Triana, C., M.Sc., University of Alabama, Birmingham
Alexander, T., B.A., Thomas Jefferson University, Philadelphia, Pennsylvania
Anderson, A., B.S., University of California, Riverside
Chan, J., B.Sc., The Hospital for Sick Children, Toronto, Canada
Craig, T., B.S., University of Michigan, Ann Arbor
Do, Q., B.S., University of Oxford, United Kingdom
Echeverry Bautista, F., Ph.D., New York University School of Medicine, New York
Edwards, R., B.A., University of North Carolina, Chapel Hill
Fernando, K., B.S., Duke University School of Medicine, Durham, North Carolina
Fontanini, M., M.S., International School for Advanced Studies (ISAS), Trieste, Italy
Lin, S., B.A., Weill Cornell Medicine, New York
Prakash, M., Ph.D., Central Michigan University, Mount Pleasant

SEMINARS

Baranovic, J., University of Edinburgh, United Kingdom: Single channels.
Bean, B., Harvard Medical School, Boston, Massachusetts: Overview of ion channels.
Beeton, C., Baylor College of Medicine, Houston, Texas: Potassium channels, from patch-clamp to drug target discovery.
Branco, T., Sainsbury Wellcome Centre, London, United Kingdom: Synaptic plasticity and long-term potentiation using 2-photon optogenetics to study spike-timing-dependent plasticity.
Cohen, J., Johns Hopkins School of Medicine, Baltimore, Maryland: In vivo neurophysiology.
Creed, M., Washington University School of Medicine in St. Louis, Missouri: Circuit interrogation with optogenetics.
Deschenes, I., Oklahoma State University Medical Center, Tulsa: Sodium channels: Lessons learned from bedside to bench.
Dolphin, A., University College London, United Kingdom: Calcium channels: Function, pharmacology, and trafficking.
Duguid, I., University of Edinburgh, United Kingdom: Cerebellar-recipient motor thalamus drives behavioral context-specific movement initiation.
Grundemann, J., University of Basel, Switzerland: Signal conditioning.
Harris, T., Janelia Research Campus, Ashburn, Virginia: Neuropixels.
Kaeser, P., Harvard Medical School, Boston, Massachusetts: Molecular mechanisms of neurotransmitter release.
Kleinlogel, S., University of Bern, Hinterkappelen, Switzerland: Optogenetics. Native and engineered vertebrate opsins as next-generation optogenetic tools: Biophysics/design and targeted applications.
Luthi, A., Friedrich Miescher Institute, Basel, Switzerland: Amygdala circuits.

Overstreet-Waldiche, L., University of Alabama, Birmingham: Synaptic inhibition.
Schmidt-Hieber, C., Institut Pasteur, Paris, France: Series resistance.
Sjostrom, J., McGill University, Montreal, Canada: Synaptic integration.
Wanaverbecq, N., Aix Marseille University, Marseille, France: General membrane biophysics.
Xu-Friedman, M., University of Buffalo, New York: Synaptic transmission and short-term plasticity.

Mouse Engineering Minicourse (Virtual)

June 15–18

INSTRUCTORS **T. Caspary,** Emory University School of Medicine, Atlanta, Georgia
C. Forsberg, University of California, Santa Cruz
D. Laird, University of San Francisco, California
F. Mariani, University of South California School of Medicine, Altadena

ASSISTANTS **A. Chareunsouk,** University of Southern California, Los Angeles
E. Gaylord, University of California, San Francisco
A. Long, Emory University, Atlanta, Georgia
A. Rodriguez y Baena, University of California, Santa Cruz
M. Serowoky, University of Southern California, Pasadena
B. Soygur, University of California, San Francisco
R. Van Sciver, Emory University, Atlanta, Georgia
A. Worthington, University of California, Santa Cruz

This online mini-course consisted of a series of morning lectures and afternoon sessions on how to engineer mouse models for a variety of biological questions. The mini-course was two-tiered, with lecture-only participants attending only morning sessions (10 am to 1 pm EDT), and full participants attending both morning lectures and the afternoon workshops (10 am to 5 pm EDT). The morning sessions each day included lectures by prominent mouse engineers, and the afternoon workshop sessions included small group discussions of approaches with experts, problem-solving and other troubleshooting, and nuts-and-bolts discussions. This mini-course is targeted to graduate students, postdoctoral trainees, and other science professionals who are interested in using engineered mouse models in their research in areas such as stem cell biology, development, cancer, and human disease. The goal of the mini-course was to enable participants to:

- Become familiar with important online tools and databases for mouse genetics
- Compare different strategies of genetic engineering, including CRISPR
- Learn the principles of designing custom alleles
- Understand principles of germline transmission and the strategies for breeding, genotyping, and phenotyping
- Know how to use and design genetic lineage tracing models
- Learn the power and challenges of inducible transgenic systems
- Implement reporters, tags, and other modifications

Major support for this course was provided by the National Cancer Institute. This course was supported in part by grants from Helmsley Charitable Trust and the Howard Hughes Medical Institute through the Science Education Program. Scholarship support provided by Regeneron Pharmaceuticals.

PARTICIPANTS

Akrioti, E-K., M.Sc., Hellenic Pasteur Institute, Athens, Greece
Al-Hraishawi, H., Ph.D., Rutgers, The State University of New Jersey, New Brunswick
Batra, H., Ph.D., The Catholic University of America, Washington, D.C.
Bear, R.M., B.S., Emory University, Atlanta, Georgia
Bowen, C.A., B.S., Emory University, Atlanta, Georgia
Cheung, F., B.Sc., The University of Queensland, Brisbane, Australia
Chongtham, J., M.Sc., University of Delhi, South Campus, New Delhi, India
Clearman, K.R., M.S., University of Alabama, Birmingham
Cruz, J., B.A., Federal University of Rio de Janeiro, Brazil
da Cunha Menezes, E.M., Ph.D., Nathan Kline Institute, Orangeburg, New York
Fifield, B-A., Ph.D., University of Windsor, Canada
Fritz Garcia, J.H., B.Sc., International Laboratory for Human Genome Research, Queretaro, Mexico
Gatie, M., B.Sc., University of Western Ontario, London, Canada
Hutchins, E.J., Ph.D., California Institute of Technology, Pasadena
Javed, K., Ph.D., University of Cambridge, United Kingdom
Javier, J.L., B.A., Emory University, Atlanta, Georgia
Konjusha, D., M.S., Uppsala University, Sweden
Kulkarni, A., Ph.D., University of Chicago, Illinois
Leung, F-S., B.S., University of California, San Francisco
Lim, G., Ph.D., Agency for Science, Technology and Research, Singapore
Mainwaring, O.L., B.Sc., Uppsala University, Sweden
Marallano, V.J., B.S., Icahn School of Medicine at Mount Sinai, New York
Meyerink, B.L., B.S., University of South Dakota/Sanford School of Medicine, Sioux Falls
Moen, J.K., B.A., Icahn School of Medicine at Mount Sinai, New York
Mohamed, A.M., B.S., MD Anderson, Houston, Texas
Mohapatra, S., B.S.-M.S., The University of Texas, MD Anderson Cancer Center, Houston
Moore, E., B.S., Stowers Institute for Medical Research, Kansas City, Missouri
Mueller, K., B.S., ALS Therapy Development Institute, Watertown, Massachusetts
Nadella, S., M.D., Cold Spring Harbor Laboratory
Oldak, B., M.Sc., Weizmann Institute of Science, Rehovot, Israel
Osterholt, I., M.S., USC Keck School of Medicine, Los Angeles, California
Piven, O.O., Ph.D., Institute of Molecular Biology and Genetics, NASU, Kyiv, Ukraine
Qu, F., V.M.D.-Ph.D., Washington University, St. Louis, Missouri
Ranade, P., M.S., University of North Texas Health Science Center, Fort Worth
Reichelt, M., M.S., The University of Queensland, St. Lucia, Australia
Sanchez, E.S., Ph.D., Vasquez Caltech, La Cañada, California
Stewart, A.A., B.S. (in progress 2022), University of California, Santa Cruz
Suarez, C.S., Ph.D., Cabrera CIEMAT, Madrid, Spain
Tang, Y., Ph.D., Augusta University, Georgia
Tashrin Mimu, T.T., B.S., Max Planck Institute of Biochemistry, Münich, Germany
Thabet, E.H., B.A., Alexandria University, Egypt
Thalappillil, J.S., B.S., Cold Spring Harbor Laboratory
Torrente, L., Ph.D., Moffitt Cancer Center, Tampa, Florida
Trivedi, R., B.Tech., University of North Texas, Health Science Center, Fort Worth
Velasquez Escobar, A.M., B.S., Texas A&M University, College Station
Voth, S.B., Ph.D., Yale University, New Haven, Connecticut
Woappi, Y., Ph.D., Harvard Medical School/Brigham and Women's Hospital, Boston, Massachusetts
Yang, J., M.S., Stanford University, Palo Alto, California
Yaw, A.M., Ph.D., Michigan State University, East Lansing
Zhou, P., Ph.D., Tufts Medical Center Boston, Massachusetts

SEMINARS

Alonso Curbelo, D., Memorial Sloan Kettering Cancer Center, New York: Modeling and perturbing disease mechanisms in mice: Germline and non-germline approaches.
Behringer, R., University of Texas MD Anderson Cancer Center, Houston: Transgenesis; Phenotyping challenges.
Bush, J., University of California, San Francisco: CRISPR delivery approaches and their applications; Sequential editing.
Forsberg, C., University of California, Santa Cruz: Phenotype databases and IMPRESS protocol.
Hadjantonakis, K., Memorial Sloan Kettering Cancer Center, New York: Seeing is believing: Reporter strategies for imaging at the cutting edge.
Hanna, J., Weizmann Institute of Science, Israel: Embryo culture.
Hippenmeyer, S., IST Austria: Lineage tracing using mosaic analysis with double markers (MADM); Recombinant chromosomes.
Joyner, A., Memorial Sloan Kettering Cancer Center, New York: Inducible systems; Designing inducible gene expression systems.

Laird, D., University of San Francisco, California: How to use MGI and mouse databases.
Lewandoski, M., National Cancer Institute/NIH, Frederick, Maryland: Conditional alleles; Cre/Lox breeding.
Mariani, F., University of Southern California School of Medicine, Altadena: Mouse colony management.
Murray, S., The Jackson Laboratory, Bar Harbor, Maine: IMPC alleles; Assessing available alleles.
Parker-Thornburg, J., MD Anderson Cancer Center, Houston, Texas: ES cell targeting and homologous recombination; Conditional alleles by homologous recombination.
Rossant, J., The Hospital for Sick Children, Ontario, Canada: Challenges and ethics.
Saunders, T., University of Michigan, Ann Arbor: CRISPR editing; Conditional alleles and point mutations by CRISPR.
Zernicka-Goetz, M., University of Cambridge, United Kingdom: Embryo construction.

Proteomics (Virtual)

August 10–13

INSTRUCTORS **L. Anderson,** National High Magnetic Field Laboratory, Florida
R. Chalkley, University of California, San Francisco
D. Pappin, Cold Spring Harbor Laboratory

This course was conducted completely online. It was an interactive, but mostly lecture-based, course that focused on cutting-edge proteomic approaches and technologies.

The course started by providing a grounding in mass spectrometry and peptide fragmentation analysis. Bioinformatic strategies for data analysis were explained. Quantitative strategies were introduced, including isotopic labeling and label-free analysis, with a brief introduction to the Skyline software.

Day three focused on analysis of protein post-translational modifications. Approaches for enrichment using antibodies and other affinity strategies were surveyed and data analysis challenges, including modification site assignment, were covered.

On the final day more advanced applications were presented. Protein-level mass spectrometry for proteoform analysis was described and protein structural analysis, mostly through cross-linking, was explained. The afternoon focused on strategies for studying protein interactions and associations through affinity purification and proximity labeling.

The aim of the course was to provide each student with the fundamental knowledge to perform their own proteomic experiments. The overall goal is to train students to identify new opportunities and applications for proteomic approaches in their biological research.

Major support for this course was provided by the National Institute of Child Health & Human Development and in part by grants from Helmsley Charitable Trust and the Howard Hughes Medical Institute through the Science Education Program. Scholarship support was provided by Regeneron Pharmaceuticals.

PARTICIPANTS

Akhlaq, R., M.Sc., University of Karachi, Pakistan
Akoth, M., M.S., Egerton University, Nairobi, Kenya
Arumugam, S., Ph.D., NIPER Kolkata, India
Auger, S., B.S., University of Minnesota, Minneapolis
Azzam, S., Ph.D., Case Western Reserve University, Cleveland, Ohio
Bizzaro, G., M.S., University of Stavanger, Norway
Bowen, C., B.S., Emory University, Atlanta, Georgia
Caefer, D., B.S., University of Connecticut, Storrs
Chadha, M., M.Sc., The Institute of Cancer Research, London, United Kingdom
Dang, V., B.S., University of Texas, Austin
Darira, S., B.E., University of Maryland Baltimore County, Baltimore
Davies, C., Ph.D., U.S. Department of Agriculture-ARS, Beltsville, Maryland
Deng, J., S.B., Salk Institute, La Jolla, California
Ge, X., B.S., Yale University, New Haven, Connecticut
Hober, A., M.Sc., KTH Royal Institute of Technology, Solna, Sweden
Hunziker, L., Ph.D., Curtin University, Perth, Australia
Kim, S-Y., B.S., Massachusetts Institute of Technology, Cambridge
Lu, D., M.M., The Hong Kong Polytechnic University, Hong Kong
Lu, X., Ph.D., Medical University of South Carolina, Charleston
Lutsenko, S., Ph.D., Johns Hopkins University, Baltimore, Maryland
Michelis, G., B.S., National Eye Institute/NIH, Bethesda, Maryland
Mohamamdi, M., B.S., Royan Institute, Tehran, Iran
Mustor, E., B.S., University of South Florida, Tampa
Narayanan, B., M.S., Johns Hopkins University, Baltimore, Maryland
Poeschel, A., M.A., University of Zürich, Switzerland

Portelance, R., B.S., University of Virginia, Charlottesville
Pushalkar, S., Ph.D., New York University, New York
Quiñones Avilés, Y., B.S., Yale University, West Haven, Connecticut
Ramelow, C., Ph.D., Emory University, Atlanta, Georgia
Rigo, F., M.S., University of Rijeka, Croatia
Sagini, M., M.Sc., University of Heidelberg, Germany
Sturgess, W., M.S., Lund University, Sweden
Van den Ackerveken, P., Ph.D., Belgian Volition, Isnes, Belgium
Vega Trejo, R., Ph.D., The University of Oxford, United Kingdom
Wang, Y., Ph.D., Shenzhen Institutes of Advanced Technology, China
Yoon, J., B.A., Massachusetts Institute of Technology, Cambridge
Zhu, Q., Ph.D., University of North Carolina, Chapel Hill

SEMINARS

Anderson, L., NHMFL-Florida State University, Tallahassee: Strategies for analysis of intact proteins by FT-ICR mass spectrometry at 21 tesla.
Chalkley, R., University of California, San Francisco: Software for MS/MS-based protein identification; PTM analysis: searching and evaluating results; PTM enrichment: Other than antibodies.
Clauser, K., Broad Institute of MIT and Harvard, Cambridge, Massachusetts: Manual de novo peptide MS/MS interpretation for evaluating database search results.
Farnsworth, C., Cell Signaling Technology, Concord, Massachusetts: Antibody-based proteomics.
Pappin, D., Cold Spring Harbor Laboratory: Proteomics: Introduction to MS fundamentals; Proteomics: Protein quantitation with stable isotopes.
Samavarchi-Tehrani, P., Lunefeld-Tanenbaum Research Institute, Toronto, Canada: AP-MS and proximity labeling.
Soderblom, E., Duke University, Raleigh, North Carolina: Introduction to label-free quantitation and the Skyline data analysis tool.
Trnka, M., University of California, San Francisco: Crosslinking mass spectrometry for exploring protein architecture and interactions.

Antibody Engineering and Display Technologies

November 9–22

INSTRUCTORS **C. Rader,** Scripps Research Institute, Jupiter, Florida
G. Silverman, New York University School of Medicine, New York

CO-INSTRUCTOR **G. Veggiani,** University of Toronto, Canada

ASSISTANTS **M. Karlander,** Royal Institute of Technology, Stockholm, Sweden
G. Martyn, University of Toronto, 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 *E. coli* were also 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 in part by grants from Helmsley Charitable Trust and Howard Hughes Medical Institute through the Science Education Program. Partial scholarship support was provided by the Regeneron Scholars Fund.

PARTICIPANTS

Bachas, S., Ph.D., Absci, Vancouver, Washington
Cui, A., Ph.D., Broad Institute of MIT and Harvard, Cambridge, Massachusetts
Huang, W., Ph.D., Louisiana State University, Baton Rouge
Jonsson, M., M.Sc., Eng., Royal Institute of Technology, KTH, Stockholm, Sweden
Kang, J., M.D., West Virginia University, Morgantown
Kochman, D., M.Sc., Novoron, San Diego, California
Layer, J., Ph.D., Indiana University School of Medicine, Indianapolis
Lewicka, A., Ph.D., The University of Chicago, Illinois
Lloyd, H., M.S., Harvard University, Cambridge, Massachusetts
Miao, Z., A.B., Stanford University, Palo Alto, California
Najeeb, J., Ph.D., Stanford University, California
Ntumngia, F., Ph.D., University of South Florida, Tampa
Parks, L., B.Sc., KTH Royal Institute of Technology, Stockholm, Sweden
Sinha, A., Ph.D., University of Iowa, Iowa City
Trinh, T., Ph.D., Stanford University, California
Zambrello, M., D.M.D., Ph.D., University of Connecticut, Farmington

SEMINARS

Dekosky, B., Kansas University, Lawrence: High-throughput sequencing and functional analysis of immune receptor repertoires.
Derda, R., University of Alberta, Canada: Genetically encoded chemistry by phage, on phage.
Dreier, B., University of Zürich, Switzerland: In vitro evolution of proteins by ribosome display.
Koide, S., New York University School of Medicine, New York: Design of exquisite specificity in synthetic binding proteins.
Pohl, M.A., Molecular Templates, New York: Phage display for therapeutic antibody drug discovery.
Rader, C., Scripps Research Institute, Jupiter, Florida: From phage display to cancer immunotherapy.
Sidhu, S., University of Toronto, Canada: Large-scale development of biologics to target the human proteome
Siegel, D., University of Pennsylvania, Philadelphia: Project CARLOS.
Silverman, G., New York University School of Medicine, New York: B-cell immunobiology and antibody cloning by phage display.
Smith, G., University of Missouri, Columbia: Principles of affinity selection from phage-display libraries.
Stahl, S., Royal Swedish Institute (KTH), Stockholm, Sweden: Engineering of affibody molecules.

Advanced Sequencing Technologies and Bioinformatics Analysis (Virtual)

November 11–19

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
K. Cotto, Washington University School of Medicine in St. Louis, Missouri
A. Danos, Washington University School of Medicine in St. Louis, Missouri
H. Dashnow, University of Utah, Salt Lake City
S. Freshour, Washington University School of Medicine 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
M. Kramer, Cold Spring Harbor Laboratory
J. Kunisaki, University of Utah, Salt Lake City
C. Miller, Washington University, St. Louis, Missouri
J. Preall, Cold Spring Harbor Laboratory
M. Richters, Washington University School of Medicine 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, enabling 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 condensed seven-day virtual 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 (Illumina, PacBio, Nanopore, etc.), including library construction procedures, general data processing, and in-depth data analysis. Students were introduced to Unix command-line, important file formats, alignment, data visualization, basic scripting in R, bash and other program languages, cluster job submission, and bioinformatics pipeline development. Given the condensed curriculum, several types of biological questions enabled by massively parallel sequencing technologies were explored, including bulk transcriptome profiling, single-cell transcriptome profiling, and small variant discovery/interpretation, along with other approaches 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.

Major support for this course was provided by the National Human Genome Research Institute and in part by grants from Helmsley Charitable Trust and the Howard Hughes Medical Institute through the Science Education Program.

Partial scholarship support was provided by the Regeneron Scholars Fund.

Computational resources were provided by Amazon Web Services.

PARTICIPANTS

Brathwaite, K., M.D., Montefiore Medical Center, Bronx, New York
Castillo Guzman, D., B.S., University of California, Davis
Ellis, T., Ph.D., University of North Florida, Jacksonville
Frerich, C., Ph.D., UT Southwestern Medical Center, Dallas, Texas
Harlemon, M., Ph.D., Clark Atlanta University, Roswell, Georgia
Hedehus, L., B.S., Memorial Sloan Kettering Cancer Center, New York
Kaddoura, M., M.D., Mayo Clinic, Rochester, Minnesota
Knox, R., M.D./Ph.D., Nationwide Children's Hospital, Columbus, Ohio
Kwon, H-J., Ph.D., University at Buffalo, New York
Liu, Y., Ph.D., Johns Hopkins University School of Medicine, Baltimore, Maryland
Mankouski, A., M.D., University of Utah, Salt Lake City
Mao, B., Ph.D., National Institutes of Health, Bethesda, Maryland
McDeed, A., B.S., Georgetown University, Washington, D.C.
Mikhalchenko, A., Ph.D., Oregon Health & Science University, Portland
Mori, L., B.S., The Scripps Research Institute, Jupiter, Florida
Perez-Matos, M., M.D., Harvard TH Chan School of Public Health, Boston, Massachusetts
Peritore-Galve, C., Ph.D., Vanderbilt University Medical Center, Nashville, Tennessee
Qin, S., B.S., University of Rochester School of Medicine, New York
Ruiz, C., Ph.D., Yale University, West Haven, Connecticut
Shi, T., B.S., Cincinnati Children's Hospital Medical Center, Ohio
Soto, C., Ph.D., FDA, Silver Spring, Maryland
Trinh, L., B.S., Vanderbilt University, Nashville, Tennessee
Vasileva, E., Ph.D., Children's Hospital Los Angeles, Los Angeles, California
Wasen, C., Ph.D., Brigham and Women's Hospital, Boston, Massachusetts
Zieba, J., Ph.D., University of California, Los Angeles

SEMINARS

Griffith, O., Washington University School of Medicine in St. Louis, Missouri: Intro to RNA sequencing and analysis.
Handley, S., Washington University School of Medicine in St Louis, Missouri: Metagenomics and viromics.
Li, Y., University of California, San Diego, La Jolla: Comprehensive analysis of single cell epigenomic data.
Mardis, E., Nationwide Children's Hospital, Columbus, Ohio: Overview of next-generation short read sequencing technologies.
McCombie, R., Cold Spring Harbor Laboratory: PacBIO and Oxford Nanopore long read sequencing.
Miller, C., Washington University, St. Louis, Missouri: Intro to scRNA-seq analysis.
Quinlan, A., University of Utah, Salt Lake City: Statistics for count-based data and gene expression/differential expression analysis.
Scacheri, P., Case Western Reserve University, Cleveland, Ohio: Epigenomics and NGS.
Schatz, M., Johns Hopkins University, Baltimore, Maryland: A complete reference genome improves analysis of human genetic variation.
Wagner, A., Nationwide Children's Hospital, Columbus, Ohio: Variant representation, annotation, and interpretation.
Zody, M., New York Genome Center, New York: Population sequencing and analysis projects at NYGC.

Scientific Writing Retreat

November 17–21

INSTRUCTORS	**C. Lambert,** Cold Spring Harbor Laboratory **S. Matheson,** *Cell Reports*, Cambridge, Massachusetts
CO-INSTRUCTOR	**M. Bao,** Harvard Medical School, Boston, Massachusetts
WRITING COACHES	**S. Hindle,** Cold Spring Harbor Laboratory **J. Jansen,** Cold Spring Harbor Laboratory **J. Rubin,** Columbia University, New York **R. Leshan,** Cold Spring Harbor Laboratory

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 in part by grants from Helmsley Charitable Trust and the Howard Hughes Medical Institute through the Science Education Program. Partial scholarship support was provided by the Regeneron Scholars Fund.

PARTICIPANTS

Allen, J., Ph.D., National Institutes of Health, Bethesda, Maryland
Flyak, A., Ph.D., California Institute of Technology, Pasadena
Gajenthra Kumar, N., Ph.D., University of California, Berkeley
Gowripalan, A., Ph.D., Albert Einstein College of Medicine, Bronx, New York
Lorenzini, I., Ph.D., Barrow Neurological Institute, Phoenix, Arizona
Marullo, R., Ph.D., Weill Cornell Medicine, New York
Miao, L., Ph.D., Yale University, New Haven, Connecticut
Sanchez, E., Ph.D., San Francisco State University, California
Singh, P., Ph.D., Dana-Farber Cancer Institute, Boston, Massachusetts
Wang, S., Ph.D., Duke University, Durham, North Carolina
Wilt, L., Ph.D., St. Jude Children's Research Hospital, Memphis, Tennessee
Zhang, L., Ph.D., Cold Spring Harbor Laboratory

SEMINARS

Jansen, J., Cold Spring Harbor Laboratory: Grantwriting and grantmanship.
Lambert, C., Cold Spring Harbor Laboratory; Matheson, S., Cell Press, Cambridge, Massachusetts: Session on Top 10 tips.
Lambert, C., Cold Spring Harbor Laboratory: Writing for non-expert audiences.
Matheson, S., *Cell Press*, Cambridge, Massachusetts; Hindle, S., Cold Spring Harbor Laboratory: Publications and manuscripts.

Tutorials in Genomics and Bioinformatics (Virtual)

February 17–19 and November 30–December 2

INSTRUCTORS **D. Fagegaltier,** New York Genome Center
E. Hodges, Vanderbilt University School of Medicine
B. King, University of Maine
S. Munger, The Jackson Laboratory

The Tutorials in Genomics and Bioinformatics (TGB) is an intensive three-day introductory virtual course on genomics and bioinformatics. The course began at 10 am (eastern time) on Tuesday, November 30 and ended late afternoon (eastern time) on Thursday, December 2.

TGB is modeled on Cold Spring Harbor Laboratory's Genome Access Course, a two-day course normally offered in person at CSHL and other locations. TGB 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. Ample opportunities were designed for students to engage instructors virtually during the course with specific tasks or problems that pertain to their own research.

The core of the course 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. The course also features methods to assist the analysis and prioritization of gene lists from large-scale microarray gene expression and proteomics experiments.

This course was supported in part by grants from Helmsley Charitable Trust and the Howard Hughes Medical Institute through the Science Education Program.

Two iterations of TGB took place in 2021, in February and November.

February 17–19

Students: 41

PARTICIPANTS

Brammer, J., The Ohio State University, Columbus
Chang, C., Southern University of Science and Technology, Shenzhen, China
Chaubey, K., Case Western Reserve University, Cleveland, Ohio
Collier, A., University of Miami: Miller School of Medicine, Florida
Dong, S., University of Calgary, Canada
Encarnacion-Rosado, J., NYU School of Medicine, New York
Gallagher, D., Brandeis University, Walltham, Massachusetts
Gavin, A., University of Minnesota Twin Cities, Minneapolis
Ghosh, A., Memorial Sloan Kettering, New York
Goodman, C., Nationwide Children's Hospital, Columbus, Ohio
Hall, W., Medical College of Wisconsin, Milwaukee
Hodges, J., Sangamo Therapeutics, Richmond, California
Johnson, C., Dana-Farber Cancer Institute, Boston, Massachusetts
Juarez, C., Colorado State University, Fort Collins
Khan, A., Memorial Sloan Kettering, New York
Kurylo, K., Maze Therapeutics, South San Francisco, California
Leonaviciene, G., Vilnius University, Lithuania
Liaki, V., Spanish National Cancer Research Center, Madrid, Spain
Linney, M., University of Hawaii, Manoa, Honolulu
Mitchell, S., University of Minnesota Twin Cities, Minneapolis
Mori, L., The Scripps Research Institute, Jupiter, Florida
Nguyen, D., University Hospitals Cleveland Medical Center, Ohio
Oles, A., The Medical University of South Carolina, Charleston
Owens, D., University of Miami: Miller School of Medicine, Florida
Pandya, P., Indiana University, Indianapolis
Payan Parra, O., Northwestern University, Evanston, Illinois
Penumaka, A., Broad Institute, Boston, Massachusetts
Peterson, N., St. Jude Children's Research Hospital, Memphis, Tennessee
Pline, K., Weill Cornell Medicine, New York
Saito-Diaz, V., University of Georgia, Athens
Sharan, D., University of Minnesota Twin Cities, Minneapolis
Shersher, E., University of Miami: Miller School of Medicine, Florida
Singh, S., University of Miami: Miller School of Medicine, Florida
Siththanandan, V., Stanford University, San Francisco, California
Soliman, G., City University of New York, New York
Vorphal, M., Universidad de Concepcion, Concepcion, Chile
Wang, N., Columbia University, New York
Wojciechowski, A., Vanderbilt University, Nashville, Tennessee
Yu, Z., Cold Spring Harbor Laboratory
Zhang, H., Columbia University Medical Center, New York
Ziko, L., University of Hertfordshire, hosted by GAF, Cairo, Egypt

November 30–December 2

Students: 40

PARTICIPANTS

Bardales, C., Louisiana State University, Alexandria
Bashore, A., Columbia University, New York
Bedasee, D., City College, New York
Belanger, K., Augusta University, Georgia
Carlos dos Santos, J., University of Sao Paulo, Ribeirao Preto, Brazil,
Chen, J., Boston University, Massachusetts
Cilento, M., Emory University, Atlanta, Georgia
Denman, R., The University of Queensland Diamantina, Wooloongabba, Australia
Devotta, A., New York University College of Dentistry, New York
Do Carmo, M., Yale University, New Haven, Connecticut
Dutta Banik, D., Indiana University School of Medicine, Indianapolis
Fischer, U., University Düsseldorf, Germany
Gumber, H., Colgate-Palmolive, Piscataway, New Jersey
Hofbauer, L., Research Institute of Molecular Pathology, Vienna, Austria
Holden, R., Broad Institute, Cambridge, Massachusetts
Hung, S-H., University of Texas MD Anderson Cancer Center, Houston
Huttenlocher, A., University of Wisconsin, Madison
Jepsen, V., University Hospital Düsseldorf, Germany
Khan, A., Nationwide Children's Hospital, Columbus, Ohio

Layman, D., Estée Lauder Companies, Inc., Melville, New York
Mahiou, J., Cystic Fibrosis Foundation, Lexington, Massachusetts
Mateo Elizalde, C., Cold Spring Harbor Laboratory
Mavuluri, J., St. Jude Children's Research Hospital, Memphis, Tennessee
Murphy, M., University of Wisconsin, Madison
Ndambuki, D., Albany Medical College, New York
Nityanandam, A., St. Jude Children's Research Hospital, Memphis, Tennessee
Pandey, S., Albert Einstein College of Medicine, Bronx, New York
Pumiglia, K., Albany Medical College, New York
Rosser, F., University of Pittsburgh School of Medicine, Pennsylvania
Schwab, N., University of Toronto, Canada
Singh, J., JNCASR, Bangalore, India
Tedbury, P., Emory University, Atlanta, Georgia
Tolic, A., IBISS, University of Belgrade, Serbia
Trasser, M., Gregor Mendel Institute GmbH, Vienna, Austria
Truong, J., Emory University, Atlanta, Georgia
Vargas-Castillo, A., Dana-Farber Cancer Institute, Boston, Massachusetts
Wong, E., Cystic Fibrosis Foundation, Lexington, Massachusetts
Yeung, J., The Rockefeller University, New York
Zhang, L., University of California, Los Angeles
Zuloaga, K., Albany Medical College, New York

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 for the exchange of ideas in an informal setting. The program was limited during the COVID-19 pandemic.

Speaker	Title	Host
January		
Betty Diamond, M.D., Director and Professor, Institute of Molecular Medicine, Feinstein Institute for Medical Research	Antibodies, behavior, and cognition: a biomedical frontier	Bruce Stillman
Sohail Tavazoie, M.D., Ph.D., Professor, The Rockefeller University	Biology and therapy of metastatic disease	Camila dos Santos
Nadia Urbain, Ph.D., Chargé de Recherche, INSERM, Centre de Recherche en Neurosciences de Lyon (CRNL), Lyon, France	Thalamic neuronal dynamics across sleep and wake	Postdocs
Emily Bernstein, Ph.D., Professor, Mount Sinai School of Medicine	Remodeling chromatin in cancer	Camila dos Santos
February		
Guillaume Lettre, Professor of Medicine, Université de Montréal and Montreal Heart Institute	Defining the role of vascular endothelial cells in human coronary artery disease	Zach Lippman
Geeta Narlikar, Ph.D., Department of Biochemistry and Biophysics, University of California, San Francisco	Genome regulation by phase-separation and ATP-dependent chromatin remodeling	Terri Grodzicker
Magdalena Gotz, Ph.D., Research Professor, LMU Munich	New mechanisms of neurogenesis and neural repair	Linda Van Aelst
March		
Miriam Merad, Professor, Department of Oncological Sciences, Director of the Human Immuno-Monitoring Center, Endowed Chair in Cancer Immunology, Director, Precision Immunology Institute, Icahn School of Medicine at Mount Sinai	Macrophages contribution to health and disease	WiSE
Kirsten Bomblies, Swiss Federal Institute of Technology Zürich, Institute of Molecular Plant Biology	How to tango with four: meiotic adaptation to whole genome duplication	Zach Lippman
Joan Brugge, Director of the Ludwig Center, Louise Foote Pfeiffer Professor of Cell Biology, Harvard Medical School, Department of Cell Biology	Oxidative stresses and cancer	WiSE

Speaker	Title	Host
April		
Navdeep Chandel, David W. Cugell Professor of Medicine, Biochemistry and Molecular Genetics, Northwestern University	Linking cancer metabolism to neurological diseases	Semir Beyaz
Edith Heard, Ph.D., EMBL Director General and Collège de France Professor	Mechanisms of gene silencing and dosage regulation during X-chromosome inactivation	Graduate Students
Linda Richards, Ph.D., Professor, Department of Neuroscience, Washington University School of Medicine in St. Louis	Activity-dependent and molecular mechanisms regulating the development of cortical connectivity	Linda Van Aelst
October		
Richard Youle, Ph.D., Senior Investigator, Biochemistry Section, NINDS, NIH	Intertwined roles of two (or three) Parkinson's disease genes in mitochondria, autophagy and inflammation	Richard Sever
Neil Johnson, Associate Professor, Fox Chase Cancer Center	New insights into the role of *BRCA1* mutations in DNA repair and cancer	Jeff Boyd
Christopher Mason, Professor of Genomics, Physiology, and Biophysics, Weill Cornell Medicine, New York	Single-cell, city-scale, and inter-planetary genomics	Tom Gingeras
November		
Ralph Deberardinis, Howard Hughes Medical Institute and UT Southwestern Medical Center	Metabolic reprogramming in human disease	Richard Sever
Peter Tonge, Professor and Chair, Stony Brook University	Widening the therapeutic window: kinetic selectivity and target vulnerability	John Moses
December		
Michelle Monje, Professor of Neurology & Neurological Sciences, Stanford University	Neuron-glial interactions in health and disease: from cognition to cancer	Lucas Cheadle

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	
Florin Albeanu	Organizing principles of olfactory neural circuits
Katarina Meze	Molecular machines in time and space: how a nifty nuclease degrades structured RNAs
February	
Johanna Syrjanen	Exploring the molecular architectures of large-pore channels
Camila dos Santos	Mammary development, gene regulation, and oncogenesis—from epithelial cells and beyond
March	
David McCandlish	Understanding complex genetic interactions
Saket Navlakha	How not to forget: lessons from fruit flies
Hannah Meyer	Integrative approaches to study organ function
April	
Christopher Hammell	An epigenetic priming mechanism mediated by nutrient sensing regulates transcriptional output during *C. elegans* development
Arkarup Banerjee	Singing mice: what's all the fuss about?
Semir Beyaz	Mechanisms of environment–gene interactions
Lucas Cheadle	Exploring a new role for microglia in sensory-dependent circuit development
Pavel Osten	Imaging the brain and other tissues at sufficient resolution
October	
Zhikai Wang	Carcinomas assemble a filamentous CXCL12-keratin19 coating that suppresses T-cell-mediated immune attack
Nicholas Gallo	Shedding light on the cellular and molecular mechanisms governing chandelier cell connectivity in health and disease
Joshua Homer	Accelerated SuFEx click chemistry for modular synthesis
November	
Lucas Cheadle	Oligodendrocyte precursor cells (OPCs) engulf synaptic inputs in an experience- and microglia-dependent manner
Alexei Koulakov	Animal behavior as a machine-learning problem
December	
Xueqin Sun	An unprecedented role of BRD8 bromodomain in reprogramming TP53 network in glioblastoma

BANBURY CENTER

BANBURY CENTER
EXECUTIVE DIRECTOR'S REPORT

After an especially tumultuous 2020, COVID-19 vaccines brought optimism for in-person convening in 2021. However, as the year progressed, continued uncertainties and new viral variants led to further postponements. As a result, the Banbury Center remained closed to in-person meetings for the duration of 2021.

Activities

Virtual meetings continued to be unrealistic for most Banbury discussions, chiefly because the format limits informal engagement, which can be the most productive time for open discussion and building connections. However, we were once again able to bring together three groups remotely. We continued remote discussions with our *Environmental Consequences of Deep-sea Mining* expert group. Led by organizers Anna Metaxas (Dalhousie University) and Verena Tunnicliffe (University of Victoria), and facilitated by Maya Breitburg-Smith (RESOLVE), the two virtual meetings took on the difficult task of narrowing objectives for the eventual in-person meeting. As a result of the virtual efforts, working groups were formed that met semiregularly throughout 2021 to further discussions on three specific subtopics. In April, Banbury virtually convened the CSHL *Technology and Education Council* (TEC), an advisory group comprised of senior scientists from the pharmaceutical and biotechnology industries. The high-level discussions covered three main topics: diversity in clinical trials, indirect effects of COVID-19, and challenges of data. Insights from the day's conversations will provide themes for future Banbury Center meetings. Finally, in December, we virtually kicked off the planned 2022 meeting, *Making Career-spanning Learning in the Life Sciences Inclusive and Effective for All*. Expert participants from around the globe came together to review the meeting objectives, and to start work on case study challenges faced by short-format trainers in the life sciences.

In addition to virtual work, the Center was energized by three 2021 publications resulting from prior Banbury meetings. Participants in the 2019 in-person and 2020 virtual follow-up meetings on *Bridging the Research-to-Practice Chasm in Digital Mental Health* published the "Banbury Forum consensus statement on the path forward for digital mental health treatment" in *Psychiatric Services*. The 2020 *MAVEN Project* leadership team produced "Introducing the MAVEN Leadership Training Initiative to diversify the scientific workforce" in *eLife*. Finally, participants from Banbury's last in-person meeting before the COVID-19 shutdown (*Copper Cancer Consortium*) published "Connecting copper and cancer: from transition metal signalling to metalloplasia" in *Nature Reviews Cancer*.

Although COVID-19 prevented in-person convening, the estate was humming in 2021 with renovations to Sammis Hall. The newest of the estate's three residence buildings, Sammis Hall has

provided housing for meeting participants, summer courses students, and CSHL Meetings attendees since 1981. Renovations added en suite bathrooms to each room, removing the need for sharing bathrooms in a "post-COVID" world. The makeover also added to the estate's accessible guest rooms, improving our ability to accommodate participants with disabilities at future meetings.

Support

Funding is always a major hurdle to organizing Banbury meetings, as topics often fall at new intersections of science and technology or deal with delicate ethical or policy issues. We are ever grateful to the organizations and individuals that provide the financial support to enable Banbury to convene global leaders. The CSHL Corporate Sponsor Program remains a critical resource for cutting-edge meetings and provided the necessary funds to support virtual convenings. In 2021 we are especially grateful to our funding partners who graciously extended grants so that we could postpone meetings, rather than cancel or attempt virtual conversions.

The Team

The Center is successful thanks to a team of professionals who ensure the estate and programs are running at a high level. The pandemic, and subsequent lack of in-person meetings, meant that the office staff worked remotely for half of the year. We were sad to lose Michelle Corbeaux, who moved to Columbia University after expertly managing the Center's finances, co-managing the Corporate Sponsor Program, and supporting travel for thousands of meeting participants during her six years at Banbury. In the autumn the Center expanded rapidly, welcoming three new team members: Duncan Yates as Lodging Manager, Vanessa Franco as Finance and Development Coordinator, and Hannah Stewart as Communications and Special Projects Coordinator.

Despite the lack of meetings, housekeepers Miriam, Maria, Emma, and Terry, supervised by Claudia Schmid and Patricia McAdams, ensured our offices were kept clean and disinfected. Jose Peña-Corvera, Paulo Krizanovski, and Juan Colocho skillfully maintained 55 acres of impeccable grounds, keeping the estate beautiful and accessible throughout the entire pandemic. As always, we are grateful to the entire Facilities team who led the Sammis Hall renovations and kept us up-to-date on health and safety protocols.

Rebecca Leshan, Ph.D.
Executive Director

2021 Publications Resulting from Banbury Meetings

Ge EJ, Bush AI, Casini A, Cobine PA, Cross JR, DeNicola GM, Dou QP, Franz KJ, Gohil VM, Gupta S, et al. 2021. Connecting copper and cancer: from transition metal signalling to metalloplasia. *Nat Rev Cancer* **22:** 102–113. doi:10.1038/s41568-021-00417-2

Mohr DC, Azocar F, Bertagnolli A, Choudhury T, Chrisp P, Frank R, Harbin H, Histon T, Kaysen D, Nebeker C, et al., on behalf of the Banbury Forum on Digital Mental Health. 2021. Banbury Forum consensus statement on the path forward for digital mental health treatment. *Psychiatr Serv* **72:** 677–683. doi:10.1176/appi.ps.202000561

Wang YC, Brondolo E, Monane R, Kiernan M, Davidson KW, MAVEN Leadership Team. 2021. Introducing the MAVEN Leadership Training Initiative to diversify the scientific workforce. *eLife* **10:** e69063. doi:10.7554/eLife.69063

BANBURY CENTER MEETINGS

Dates	*Title*	*Organizer(s)*
January 14, 19	Environmental Consequences of Deep-sea Mining, a Comparison with Land-based Mining: Virtual Pre-meeting II	A. Metaxas, V. Tunnicliffe
February 8–10	Environmental Consequences of Deep-sea Mining, a Comparison with Land-based Mining: Virtual Pre-meeting III	A. Metaxas, V. Tunnicliffe
April 22	CSH Technology and Education Council (Virtual)	R. Leshan
December 6, 9	Making Career-spanning Learning in the Life Sciences Inclusive and Effective for All: Virtual Kickoff	R. Tractenberg, J. Williams

BANBURY CENTER MEETINGS

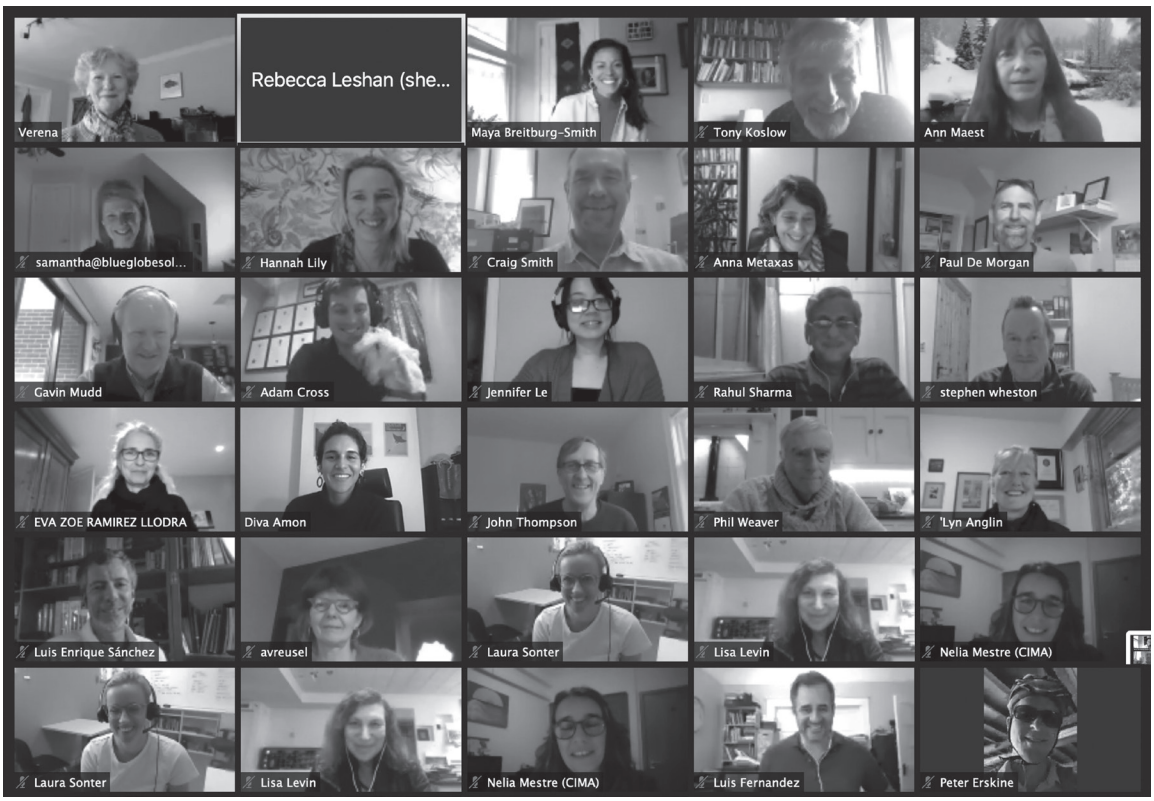
Environmental Consequences of Deep-sea Mining, a Comparison with Land-based Mining: Virtual Session II

January 14, 19

ARRANGED BY **A. Metaxas,** Dalhousie University, Halifax, Nova Scotia, Canada
V. Tunnicliffe, University of Victoria, British Columbia, Canada

FUNDED BY **Cold Spring Harbor Laboratory Corporate Sponsor Program**

The presence of metalliferous ores on the seabed is fueling speculation of greater access to metal supplies to support current and projected global demands. As international authorities finalize the regulations to enable exploitation, questions concerning the environmental consequences must be addressed despite large knowledge gaps. The challenge is to assess the risks in the context of alternative metal supplies. Building on virtual work in 2020, this second virtual workshop convened land-based mining experts to provide background on deep-sea mining, explore environmental effects of land-based mining, and suggest approaches to limit the scope for comparisons.



DAY ONE

Overview of deep-sea mining
Review of Virtual Pre-meeting I outcomes
Approaches to comparing environmental impacts of deep-sea and land-based mining

DAY TWO

Variability of land-based mining
Refining objectives for next meeting

PARTICIPANTS

C.D. ('Lyn) Anglin, Anglin & Associates, North Vancouver, Canada
M. Breitburg-Smith, RESOLVE, Washington, D.C.
B. Butler, U.S. Environmental Protection Agency/ORD, Cincinnati, Ohio
A. Cross, Centre for Mine Site, Perth, Western Australia, Australia
P. De Morgan, RESOLVE, Logan, Utah
P. Erskine, University of Queensland, Brisbane, Queensland, Australia
L.E. Fernández, Wake Forest University, Winston-Salem, North Carolina
A.S. Maest, Buka Environmental, Boulder, Colorado
A. Metaxas, Dalhousie University, Halifax, Nova Scotia, Canada
G. Mudd, RMIT University, Melbourne, Victoria, Australia
L.E. Sánchez, University of São Paulo, São Paulo, Brazil
L. Sonter, University of Queensland, Brisbane, Queensland, Australia
J.F.H. Thompson, PetraScience Consultants, Vancouver, British Columbia, Canada
V. Tunnicliffe, University of Victoria, British Columbia, Canada
S. Wheston, Tembusu Limited, Cashel, Ireland

Environmental Consequences of Deep-sea Mining, a Comparison with Land-based Mining: Virtual Session III

February 8–10

ARRANGED BY **A. Metaxas,** Dalhousie University, Halifax, Nova Scotia, Canada
V. Tunnicliffe, University of Victoria, British Columbia, Canada

FUNDED BY **Cold Spring Harbor Laboratory Corporate Sponsor Program**

The presence of metalliferous ores on the seabed is fueling speculation of greater access to metal supplies to support current and projected global demands. As international authorities finalize the regulations to enable exploitation, questions concerning the environmental consequences must be addressed despite large knowledge gaps. The challenge is to assess the risks in the context of alternative metal supplies. Our third virtual workshop convened the full group of land-based mining and deep-sea experts to consider the supposition that the environmental trade-off between deep-sea and terrestrial extraction favors opening this new ocean frontier.

DAY ONE

Opening Remarks: M. Breitburg-Smith, RESOLVE, Washington, D.C.
R. Leshan, Banbury Center, Cold Spring Harbor Laboratory, New York
A. Metaxas, Dalhousie University, Halifax, Nova Scotia, Canada
V. Tunnicliffe, University of Victoria, British Columbia, Canada

V. Tunnicliffe, University of Victoria, British Columbia, Canada: Pre-meeting outcomes

A. Metaxas, Dalhousie University, Halifax, Nova Scotia, Canada: The purpose of comparing environmental impacts of deep-sea mining with land-based mining

John Thompson, PetraScience Consultants, Vancouver, British Columbia, Canada: Comparing deep-sea mining and land-based mining

Group Discussion

DAY TWO

Group Discussion: Land-based mining scenarios

C. Smith, University of Hawaii at Mānoa, Honolulu, Hawaii (presenter); D. Amon, Natural History Museum, London, United Kingdom: Nodule Mining I

S. Smith, Blue Globe Solutions, Toronto, Ontario, Canada: Nodule Mining II

Group Discussion: Deep-sea mining scenarios

Group Discussion: Challenges and opportunities

DAY THREE

Breakout Groups: Comparing deep-sea mining and land-based mining

Breakout Group Reporting

Group Discussion

PARTICIPANTS

D.J. Amon, Natural History Museum, London, United Kingdom
C.D. ('Lyn) Anglin, Anglin & Associates, North Vancouver, Canada
M. Breitburg-Smith, RESOLVE, Washington, D.C.
B. Butler, U.S. Environmental Protection Agency/ORD, Cincinnati, Ohio
A. Cross, Centre for Mine Site, Perth, Western Australia, Australia
P. De Morgan, RESOLVE, Logan, Utah
P. Erskine, University of Queensland, Brisbane, Queensland, Australia
L.E. Fernández, Wake Forest University, Winston-Salem, North Carolina
T. Koslow, Scripps Institution of Oceanography, Sandford, Tasmania, Australia
J. Le, Silver Spring, Maryland
L. Levin, Scripps Institution of Oceanography, University of California, San Diego, La Jolla
H. Lily, London, United Kingdom
A.S. Maest, Buka Environmental, Boulder, Colorado
N. Mestre, CIMA, Universidade do Algarve, Faro, Portugal
A. Metaxas, Dalhousie University, Halifax, Nova Scotia, Canada
G. Mudd, RMIT University, Melbourne, Victoria, Australia
E. Ramirez-Llodra, REV Ocean, Oslo, Norway
L.E. Sánchez, University of São Paulo, São Paulo, Brazil
R. Sharma, National Institute of Oceanography, Goa, India
C. Smith, University of Hawaii at Mānoa, Honolulu, Hawaii
S. Smith, Blue Globe Solutions, Toronto, Ontario, Canada
L. Sonter, University of Queensland, Brisbane, Queensland, Australia
J.F.H. Thompson, PetraScience Consultants, Vancouver, British Columbia, Canada
V. Tunnicliffe, University of Victoria, British Columbia, Canada
A. Vanreusel, Ghent University, Belgium
P. Weaver, Seascape Consultants, Ltd., United Kingdom
S. Wheston, Tembusu Limited, Cashel, Ireland

CSHL Technology and Education Council (Virtual)

April 22

ARRANGED BY **R. Leshan,** Cold Spring Harbor Laboratory, New York

FUNDED BY **Cold Spring Harbor Laboratory Corporate Sponsor Program**

Cold Spring Harbor Laboratory's Technology and Education Council (TEC) is an advisory group composed of senior scientists from the pharmaceutical and biotechnology industries. This meeting convened members of TEC as well as global experts and thought leaders for high-level, interdisciplinary engagement around the challenges and opportunities in precision medicine. In addition to open discussions of new opportunities and strategies, the meeting identified topics for future Banbury meetings.

Welcoming Remarks: **R. Leshan,** Cold Spring Harbor Laboratory, New York
D. Stewart, Cold Spring Harbor Laboratory, New York
B. Stillman, Cold Spring Harbor Laboratory, New York

SESSION 1: Diversity in Testing and Clinical Trials

Chairperson: R. Leshan, Cold Spring Harbor Laboratory, New York

SESSION 2: Emerging (Indirect) Impacts of COVID-19

Chairperson: D. Stewart, Cold Spring Harbor Laboratory, New York

SESSION 3: Challenges of Data

Chair: A. Whiteley, Cold Spring Harbor Laboratory, New York

SESSION 4: Flash Topics

Chair: R. Leshan, Cold Spring Harbor Laboratory, New York

PARTICIPANTS

L. Brizuela, Agilent Technologies, Santa Clara, California
C. Donaldson, Cold Spring Harbor Laboratory, New York
T. Evans, New England BioLabs, Inc., Ipswich, Massachusetts
M. Gurtowski, Cold Spring Harbor Laboratory, New York
J. Herman, Northwell Health, New Hyde Park, New York
S. Korn, Merck Research Laboratories, Kenilworth, New Jersey
R. Leshan, Cold Spring Harbor Laboratory, New York
D. Rocco, Biogen, Cambridge, Massachusetts
A. Sachs, Thermo Fisher Scientific, San Francisco Bay Area, California
C. Saliou, The Estée Lauder Companies, New York, New York
A. Sireci, Loxo Oncology at Lilly, New York
D. Stewart, Cold Spring Harbor Laboratory, New York
B. Stillman, Cold Spring Harbor Laboratory, New York
A. Whiteley, Cold Spring Harbor Laboratory, New York

Making Career-spanning Learning in the Life Sciences Inclusive and Effective for All: Virtual Kickoff

December 6, 9

ARRANGED BY	**R. Tractenberg,** Georgetown University, Washington, D.C. **J. Williams,** Cold Spring Harbor Laboratory, New York
ORGANIZING COMMITTEE	**B. Batut,** University of Freiburg, Germany **S. Donovan,** BioQUEST Curriculum Consortium **K.L. Jordan,** The Carpentries, Oakland, California **T. Mourad,** Ecological Society of America, Washington, D.C. **T. Teal,** RStudio, Boston, Massachusetts **C. van Gelder,** Dutch Techcentre for Life Sciences, Utrecht, The Netherlands

FUNDED BY **Cold Spring Harbor Laboratory Corporate Sponsor Program**

Continual advances in life sciences methods and technologies have made it unrealistic to rely solely on university degrees for up-to-date training. Short-format training (SFT) has emerged to fill this gap, providing focused learning opportunities with the agility to adapt to the latest scientific developments. Studies of SFT have warned of challenges for the format, including potential disparities in accessibility for marginalized groups. A planned 2022 Banbury meeting will bring together experts on training, education, and life science research to explore challenges, foster a community motivated to improve, and develop a blueprint for a more inclusive and effective future of SFT. This virtual kickoff meeting convened the expert group to review objectives for the 2022 in-person meeting and to begin to tackle current challenges faced by SFT.

PARTICIPANTS

B. Batut, University of Freiburg, Germany
E. Becker, The Carpentries, Oakland, California
A. Brown, Virginia Tech, Blacksburg
M. Burke, Australian BioCommons/QCIF and RCC at the University of Queensland, Brisbane, Australia
B. Busby, DNAnexus
A. Clyburne-Sherin, Reproducibility for Everyone
A. Dillman, National Institutes of Health (NIH), Bethesda, Maryland
S. Donovan, BioQUEST Curriculum Consortium
N. Faya, Genomics Aotearoa, Dunedin, New Zealand
J. Guo, University of Arizona, Tucson
C. Hall, Australian BioCommons, The University of Melbourne, Parkville, Victoria, Australia
K. Hertweck, Chan Zuckerberg Initiative, Redwood City, California
K. Jordan, The Carpentries, Oakland, California
J.R. Jungck, University of Delaware, Newark
C. Lambert, Cold Spring Harbor Laboratory, New York
A. Latour, IDEA-STEM
J. Lindvall, ELIXIR-SE, Uppsala, Sweden
M. Lloret Llinares, European Molecular Biology Laboratory–European Bioinformatics Institute, Cambridge, United Kingdom
R. Morris, National Center for Biotechnology Information, NLM, NIH, Bethesda, Maryland
T. Mourad, Ecological Society of America, Washington, D.C.
A. Nisselle, Melbourne Genomics, The University of Melbourne, Parkville, Victoria, Australia
P. Ordóñez, University of Puerto Rico, Río Piedras
L. Paladin, The European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
P. Palagi, SIB Swiss Institute of Bioinformatics, Lausanne
M. Sukhai, Canadian National Institute for the Blind, Toronto, Ontario, Canada
T. Teal, RStudio, Boston, Massachusetts
R. Tractenberg, Georgetown University, Washington, D.C.
C. van Gelder, Dutch Techcentre for Life Sciences, Utrecht, The Netherlands
J. Williams, Cold Spring Harbor Laboratory DNA Learning Center, New York
L. Woodley, Community Initiatives/CSCCE, Oakland, California
H. Ye, University of Florida, Gainesville

DNA LEARNING CENTER

DNA LEARNING CENTER
EXECUTIVE DIRECTOR'S REPORT

ADMINISTRATION

Lauren Correri
Andréa Mahee
Valerie Meszaros
David Micklos
Donna Smith

INSTRUCTION

Kelsie Anson
Justin Burke
Megan Capobianco
Elna Carrasco
Kelly Eames
Arden Feil
Anna Feitzinger
Cristina Fernandez-Marco
Jennifer Hackett
Brittany Johnson
Jack Kellogg
Allison Mayle
Amanda McBrien
Bruce Nash
Sharon Pepenella
Jeffry Petracca
Xiaoqun Catherine Zhang

BIOMEDIA

Daniel Jacobs
Susan Lauter
Jason Williams
Chun-hua Yang

Carrying on with COVID

A year into the COVID-19 pandemic, the DNA Learning Center (DNALC) continued to find ways to be a hands-on science center "without hands." We converted all of our traditional lab field trips into virtual events, where students worked "live" with an online DNALC instructor, at home or in school. In the spring, we began to see a steady trickle of students coming for in-person instruction on school field trips, *Saturday DNA!*, and February winter break courses. To allow for socially distanced instruction, we initially operated labs at one-quarter capacity—with only one student per lab bench! This was part of an extensive "COVID-19 Safety Net" that included daily student temperature monitoring and health screening and twice-weekly testing for DNALC instructional staff.

Hoping to capitalize on pent-up demand for hands-on science experiences, we rebooted our in-person science camps for the summer season. This required new Internet infrastructure to securely collect required vaccine records or negative COVID-19 tests, 50% occupancy to allow for social distancing, and vaccination of all DNALC staff. This allowed us to operate the entire summer without any reported COVID-19 cases. Camps held at our new Brooklyn facility and with partner schools helped us to compensate for smaller classes and accommodate 751 in-person campers—so attendance actually equaled pre-COVID-19 numbers of summer 2019. With an additional 102 virtual campers, supported with mail-order kits, the total summer attendance topped 2019 by

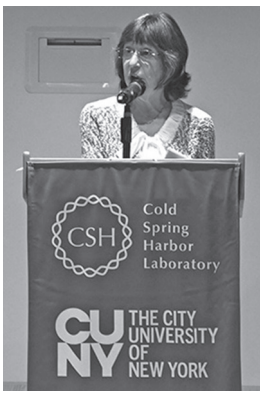
(*Left*) During one of our first in-person programs since March of 2020, Amanda McBrien leads a *Green Genes* Camp presented over consecutive Saturdays. (*Right*) Participants prep bacteria in an *Agar Art* spring break session.

14%! However, income just equaled 2019 because of significant numbers of scholarships for minority/disadvantaged students (85) and Cold Spring Harbor Laboratory (CSHL)/City University of New York (CUNY) employees (17).

So, it was clear that students were anxious to get back into the lab for the hands-on experiences they missed during the pandemic. However, with the advent of the Omicron variant, late fall proved disastrous, with most Long Island school districts halting out-of-school activities and NYC public students prohibited from using public transport for field trips. Our new facility at Regeneron Pharmaceuticals remained shuttered until late fall.

The Long Road to Brooklyn

Postponed because of the pandemic, *DNALC NYC* officially opened on September 24 on the campus of the New York City (NYC) College of Technology (City Tech) in downtown Brooklyn. With six teaching labs, two bioinformatics labs, exhibition, and lunchroom, the 17,500-square-foot facility is 25% larger than our flagship location in Cold Spring Harbor (CSH). At the opening ceremony, Bruce Stillman summarized CSHL's historical association with Brooklyn: "One hundred and thirty years ago, the Brooklyn Institute of Arts and Sciences established a teaching facility at Cold Spring Harbor. This Brooklyn initiative evolved over the years to be the global epicenter of modern biology and genetics. Through the DNA Learning Center program, the latest in life science is shared with students in real-life experiences and experiments. Today's event is a homecoming for Cold Spring Harbor Laboratory. We're thrilled to be back in Brooklyn and we hope to continue to make Brooklyn proud!"



Laurie Landeau speaks at the opening of *DNALC NYC*.

Many people helped us along the way to Brooklyn, including three who are no longer with us—Mary Jeanne Harris, Wendy van Der Poel Russell, and Arthur Spiro. In the 1970s, Jim Watson said that he hoped for an angel who would help him save CSHL from bankruptcy; the DNALC's angel and constant friend has been Laurie Landeau. In her remarks, Laurie recalled the many fits, starts, properties visited, and grants gained and lost on the road to Brooklyn. Her faith in our venture never waned, and she proposed that "patience" is a missing virtue in philanthropy.

In founding *Harlem DNA Lab* we were lucky to have the support of NYC Schools Chancellor Joel Klein. In opening the Brooklyn center, we have had the support of many within the CUNY System—most notably City Tech President Russ Hotzler, who has become a fast friend in the process. At the ceremony, CUNY Chancellor Félix Rodríguez noted this long-term collaborative and looked forward to engaging large numbers of undergraduate students in research experiences.

When we had the idea to open a center in NYC, we looked at more than a dozen properties in Manhattan in 2014–2015—just as the market rebounded and they became unaffordable. Considering the possibilities of the Borough of Brooklyn was a real eye-opener—because it was becoming clear to everyone that it was emblematic of the "new" New York. As Borough President, keynote speaker Eric Adams had been the primary architect of the tremendous revival of Brooklyn. In his remarks, he stressed the need for collaboration between government, education, and business to stimulate continued innovation in science and technology. True to this ideal, the Office of the Brooklyn Borough President provided $750,000 for state-of-the art scientific and exhibit equipment for *DNALC NYC* at City Tech.

The long road to Brooklyn began in Cold Spring Harbor in 1985, when we developed simple methods to allow high school students to analyze DNA molecules. Teachers on Long Island were enthusiastic to learn this technology, so we came up with the idea of mobile vans full of equipment that could outfit any classroom for a gene cloning experiment. For a decade we crisscrossed the country with two vans and trained thousands of teachers. In the meantime, in 1988, we started the DNALC at CSHL as the first science center devoted to public genetics education.

DNALC NYC Opening: (*Top left*) Brooklyn Borough President and mayoral candidate Eric Adams addressed attendees and (*top right*) helped cut the ribbon celebrating the opening of *DNALC NYC* on September 24. CSHL President Bruce Stillman, DNALC Committee Chairman Laurie J. Landeau, and CUNY Chancellor Félix V. Matos Rodríguez also spoke. (*Bottom left*) Daniel Master (second from *left*) discusses the Ashkelon burial with Marilyn Simons, Laurie Landeau, and Emily Master.

Over the next several decades we continued to develop laboratory and computer infrastructure that was used by hundreds of thousands of students per year. This supported the rapid growth of elective programs in biotechnology and student research. There are now convincing data showing that this approach works: Students who have a significant research experience are 20% more likely to stay interested in science. They are also 20% more likely to graduate on time, which shows that thinking like a researcher has broad impacts on a person's life. We call this *Research Ready*—in doing research, students are not only ready to compete in science, but they are also ready for further education and careers—ready for research, ready for life.

Wealthy schools realize this. Virtually all of the 124 school districts on Long Island and many in the New York metropolitan area have elective biotechnology and research courses—these schools accounted for 20% of 2020 finalists in the prestigious Regeneron Science Talent Search. We even started a Partnership program to help private schools in Manhattan keep up with the suburban schools. Our National Science Foundation (NSF)-supported study of 2,200

Russ Hotzler, Félix Rodriguez, Eric Adams, and Bruce Stillman.

biology teachers in 2018 showed that two-thirds of high schools with biotech electives were in zip codes with household incomes above the U.S. median. The best schools also infuse inquiry labs and project-based learning into regular science courses. Students from science-active schools are competitive for admission to the best science universities. They arrive at college ready to participate in early research opportunities that drive further success—including course-based undergraduate research experiences (CUREs) that reach many freshman students. They also learn skills and competencies that prepare them for the jobs of the future.

However, our own surveys of teachers showed that these great experiences were mostly only available in wealthy high schools. Students from resource-poor schools are not prepared to compete with their peers from schools with active biotech research programs for college admission and jobs—and in life. This is reflected in the persistent underrepresentation of Blacks, Hispanics, and Native Americans at all levels of science education and the workforce.

It was becoming obvious that the entire infrastructure we developed over several decades had failed to solve the hardest problem in science education: to involve a diverse population of students in real science. We realized we could never hope to solve this problem from Long Island; we needed to set up shop in NYC. So, in 2008, we teamed with the NYC Department of Education (DOE) to set up a single DNA lab in an NYC public school in Harlem. Our experiment in Harlem worked—there we found that we could readily reach a natural clientele that is about two-thirds disadvantaged and underrepresented minority (URM) students.

We then launched a decade-long search that led to *DNALC NYC*. It is the biggest physical infrastructure project we have tackled. Here, we will apply everything we have learned about science education over the last 35 years. Quite simply, we want to give to the students of NYC exactly the same practical and research experiences that Long Island students have enjoyed for decades.

DNALC NYC will bring *Research Ready* to level the biology playing field for URM and disadvantaged students through six major programs: (1) broadly available enrichment for precollege students: field trips, summer camps, and intensive research experiences; (2) virtual learning opportunities to reach more students in more situations; (3) comprehensive college biology and life preparation; (4) pathways to biology careers; (5) CUREs for CUNY students; and (6) public exhibition on shared inheritance and social justice. All these programs will help students to see themselves as people of science and to smoothly transition from high school to higher education and careers. This is especially important in NYC, where approximately two-thirds of public-school students are underrepresented minorities or from disadvantaged backgrounds and nearly 80% of freshmen in the CUNY system are graduates of NYC public schools. In parallel with the renovation of the facility, we established an $11 million endowment to provide support for core operations and programs to increase diversity, equity, and inclusion—including our commitment to provide scholarships to at least half of the public school students attending academic year programs.

For the last 30 years, the DNALC has worked intensively with selected Long Island school districts and New York City Schools to develop strong, lab-based programs in modern biology. This begins with substantial genetics units in middle school, progresses to molecular genetics lab modules in Regents and Advanced Placement (AP) Biology courses and biotechnology electives, and culminates in capstone research programs in DNA sequencing. Through *Research Ready,* we will offer these same opportunities for underresourced schools in NYC. First, we will work with school administrators and science teachers to develop a customized scope and sequence in biology, culminating in broadly available research experiences. Second, we will work with a selected group of motivated students to provide them a structured biology immersion over the last four years of high school—including summer and academic-year courses and research experiences. Research mentors will guide these students and prepare them for college. An alumni network will support students, especially as they adjust in their college freshman year. College alumni will be welcomed back as peer mentors for subsequent years of the program.

What DNA Says About Our Past and Future

During the year, we made substantial progress on the permanent exhibition for the Brooklyn center, *What DNA Says About Our Past and Future*. Funded by a grant from CSHL Trustee Paul Taubman, the exhibit is the culmination of several decades of work to popularize "personal genetics" for students and to present a hopeful message of genetics in our lives. "DNA Past" will take a side door into issues of social justice by emphasizing that "knowledge of our shared genetic ancestry is the best inoculation against racism." The DNA of any two people is 99.9 % identical; racial and ethnic differences acquired as we peopled the globe are, indeed, only "skin deep." By exploiting current interest in ancestry, we want to make *DNALC NYC* a destination for people who want to incorporate DNA into their family exploration. During the year we reached an agreement with scientists from 23 and Me to have exclusive access to their DNA data from hundreds of thousands of people. These data will be displayed on the centerpiece of the exhibit—a 9 × 16-foot "video wall" provided by a grant from the Borough of Brooklyn. A floor-to-ceiling world map will allow visitors to see personal DNA as data points in reconstructing migrations and diaspora that mixed people and their genes throughout history.

The weeks leading up to the September opening were a flurry of activity as the DNALC design team made ready the first component of our permanent exhibition, "Innovation and Gene Mixing in the Bronze Age." Two case studies—Ötzi the Iceman, from 3300 BC, and the early Philistines, from 1200 BC—examine how innovations in agriculture, animal husbandry, international trade, and metallurgy mixed genes in Eurasia. First, we retrieved a 3D recreation of Ötzi from his storage container in the CSHL warehouse in Syosset. He had languished there since his completion in 2017, with his twin currently on display at the DNALC in Cold Spring Harbor. The making of this replica from CT scan data was the subject of the PBS special, *Iceman Reborn*.

Ötzi lived in the Alps at the very beginning of the Bronze Age and illustrates the transition between hunting–gathering and farming. Although he had the tools of a hunter, his DNA type originated in the Middle East—showing that he descended from early agriculturalists who brought farming to Europe. In addition to presenting an ancient murder mystery, Ötzi illustrates the antiquity of supposedly modern maladies. DNA analysis revealed that he carried the genome of Borrelia, which causes Lyme disease. Although Ötzi ate a healthy "Neo(lithic)" diet, he carried DNA markers for heart disease and had atherosclerotic plaques in his major artery.

The second part of the exhibit revolves around events in Ashkelon, Israel, at the end of the Bronze Age and into the ensuing Iron Age. The Leon Levy Foundation, an endowment fund contributor, introduced us to Daniel Master, of Wheaton College, and Adam Aja, of the Harvard Museum of the Ancient Near East. These archaeologists concluded 20 years of excavations at Ashkelon and uncovered the first Philistine cemetery. DNA analysis from graveyard skeletons validated oral tradition that the original Philistines were migrants from Minoan Crete. However, it also showed that the Philistines were genetically assimilated into the local Canaanite population within several hundred years of their arrival! Using 3D scans provided by Dr. Master, we worked with Blue Rhino Studios to recreate a Philistine burial from 900 BC to 750 BC. The authenticity of the display is heightened by eight pieces of Philistine pottery loaned from the Israel Antiquities Authority—shipped and installed in time for the opening.

As the year ended, we began work with CSHL trustee Jeanne Moutoussamy-Ashe on a photo gallery that joins DNA past and future. "All the World in New York" will draw portraits from amateur photographers that highlight New York City as a modern melting pot—in the way the Middle East was during the Bronze Age. All people alive today are closely related by their common "DNA Past;" however, "DNA Future" points toward modern medicines that are tailored to the unique physical attributes of population groups and individuals. Among the DNA differences between any two people are mutations that make many diseases "personal." For example, similar cancers may be caused by different mutations in different people. Personalized, or precision, medicine tailors treatment to the specific mutations that drive disease in a particular person.

Adam Aja (*top left*) inspects artifacts from the Israel Antiquities Authority on arrival at *DNALC NYC*. Perfume juglet from Ashkelon (*top right*). Dan Master (*middle left*) consults with Adam during installation of the objects. Dave Micklos, Chun-hua Yang, and Dan Master place the vitrine over the artifact display (*middle right*). Staff unpack burial sculpture created by Blue Rhino Studio (*bottom left*). It was all hands on deck prepping the *DNA Past* section (*bottom right*) of the exhibition prior to the ribbon cutting.

"New York Nobels" will highlight 35 Nobel laureates who graduated from NYC high schools and show what students visiting the DNALC can attain. A life-sized model of CSHL Nobel Laureate Barbara McClintock at work with her microscope is a hyper-local story of success in science—she grew up in Brooklyn and graduated from Erasmus Hall High School!

High School DNA Barcoding Research Programs

The DNALC continued to support authentic biodiversity research with high school students using DNA barcoding. *Barcode Long Island* (*BLI*), funded by the National Institutes of Health (NIH),

involves students in "campaigns" 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 NYC. Science teachers are mentors for *BLI* and *UBP* students, while scientists from NYC institutions mentor *UBRP* students.

During the final year of funding through the NIH Science Education Partnership Award (SEPA), *BLI* supported 102 students in 34 teams. Despite the pandemic, 72 sequences were published in GenBank with student authors, including two new barcode records and 18 variants. Over seven years (see table), the program trained 234 Long Island teachers and supported nearly 1,500 high school students, resulting in 654 GenBank-published barcodes—providing important diversity and range information for more than 200 species. Although NIH funding ended, schools continue their involvement by covering costs or through scholarships.

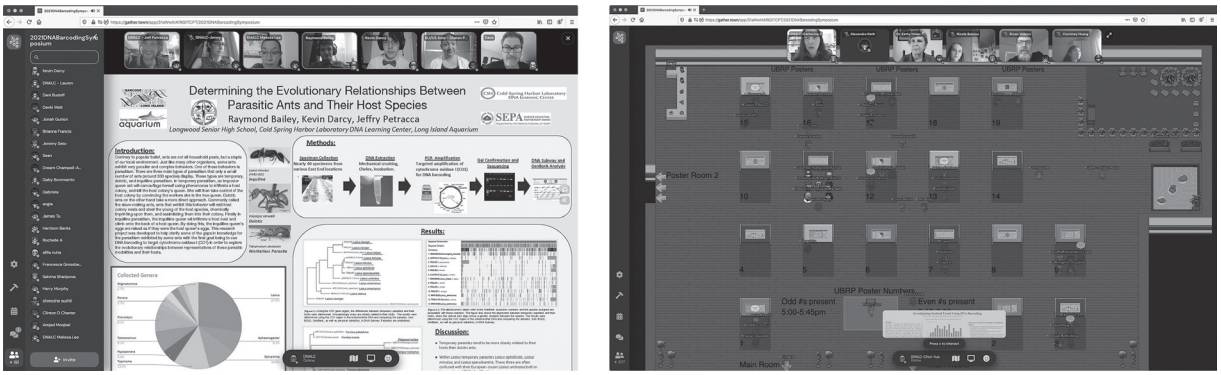
The SEPA project had important impacts on students who could be our next generation of scientists. Students ($n = 428$) reported increased knowledge of the principles and process of DNA barcoding (90%), indicated they were proud of their research (85%), learned a lot about conducting science (89%), and felt the problem-solving skills they learned would help in future courses (84%) and careers (75%). Many also reported increased interest in further pursuit of STEM (50%) and, despite its difficulty, bioinformatics (46%).

Objectives and outcomes from *Barcode Long Island* Cohorts, 2014–2021

		***Barcode Long Island* Cohort Results**							
Objective		**Cohort 1 2014–2015**	**Cohort 2 2015–2016**	**Cohort 3 2016–2017**	**Cohort 4 2017–2018**	**Cohort 5 2018–2019**	***Cohort 6 2019–2020**	***Cohort 7 2020–2021**	**Total**
Support 600 DNA barcoding research projects by 1,800 high school students	Teams supported	29	91	106	133	79	45	34	517
	Student participants grades 9–12	78	271	312	382	208	124	102	1477
	Open Labs hosted by DNALC and collaborators	6	19	13	18	7	4	5	72
	Students attending Open Labs	13	171	122	190	73	35	27	631
	Reagent and equipment footlockers distributed	6	36	27	39	16	14	13	151
	Students supported by footlockers	36	282	271	325	114	39	94	1161
	Samples collected	580	1274	1257	1200	800	283	372	5766
	DNA sequence reads obtained	467	1580	1402	1917	1218	366	548	7498
Train and support 240 Teacher Mentors	Teachers trained	71	82	28	30	17	6	–	234
	Teachers mentoring student teams	22	31	33	44	22	10	14	176
Publish DNA barcodes to GenBank database	Previously unpublished (novel) DNA sequences	1	1	1	22	11	3	2	41
	Variant DNA sequences	–	–	–	34	52	41	18	145
	Sequences that provide species GPS information	4	1	–	95	139	177	52	468

*Cohorts affected by the pandemic.

Ninety-six students working on 33 teams completed projects in the *UBP* and 34 students working on 19 teams completed projects in the *UBRP*. With many schools, DNALC facilities, and mentor lab spaces unavailable, 11 *UBP* and seven *UBRP* teams used at-home DNA extraction kits to complete their projects with assistance from DNALC staff or mentors. Four *UBRP* teams presented posters to peers and science professionals at the annual Science Research Mentoring Program (SRMP) Virtual Colloquium in June.



The barcoding research symposium featured a virtual poster session including presenter and attendee avatars that moved in the space and interacted via video in poster rooms.

Combined, during the year, 75 participants in the three high school barcoding programs completed project exit surveys. They were overwhelmingly proud of their research (90.9%) and valued the problem-solving approaches they learned (87.9%). More than three-quarters said they were more interested in continuing to study science (77.6%) and biology (73.0%)—impressive in this group of students who already chose to do science research.

The annual research symposium was held virtually for the second time in 2021. *UBRP*, *UBP*, and *BLI* students presented their research together with *Barcoding US Ants* participants on June 1 in a Gather Town space, allowing for both audio/visual and text-based interactions. Topics included wildlife across NYC and Long Island, biodiversity and trade, food fraud, and human health. Dr. Javier A. Izquierdo's (Hofstra University) keynote address on understanding, deconstructing, and rebuilding microbiomes to make a better world was streamed through YouTube.

Barcoding US Ants

The pilot cohort of *Barcoding US Ants*, our supplemental SEPA project that engages citizen scientists to identify and map ant species across the United States using DNA barcodes, wrapped up in January with a virtual mini-symposium. DNALC staff shared program results, including identification of 97 ant species from more than 300 collected samples. Program entomologists Dr. Shawn Dash and Jeffry Petracca discussed how these specimens illustrate the gaps in knowledge of even familiar species and the importance of citizen science in presenting new data that can advance science. Participants from 13 of 26 teams shared their results and experiences during the symposium.

During the year, 274 GenBank records were published with citizen scientists as authors, which included 70 variable DNA sequences and 18 previously unpublished DNA sequences. Another 15 specimens could not be identified using a combination of DNA barcode sequence and morphology and are under further investigation. One hundred twenty-two staff from science and nature centers, members of conservation organizations, educators, students, and nature enthusiasts completed projects.

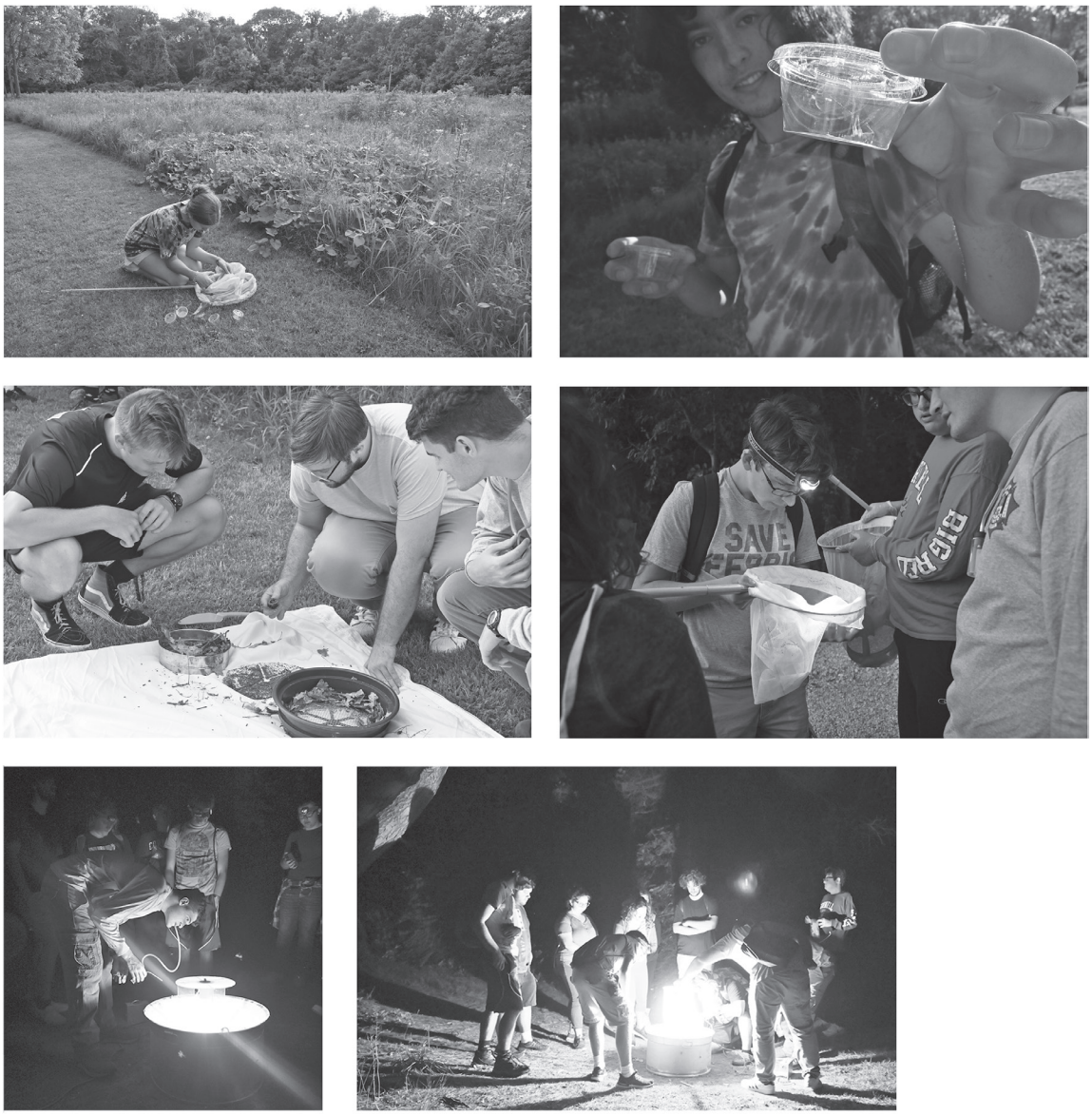
Following SEPA funding, 234 continuing or new participants investigate ant biodiversity through *Citizen DNA Barcode Network* (*CDBN*) or InnovATE*BIO*. Data analysis is ongoing.

Citizen DNA Barcode Network

This year, the *Citizen DNA Barcode Network*, our SEPA project following *BLI*, completed its first full year of programming. The project organizes local and national DNA barcoding campaigns for "citizen scientists" at science centers and conservation organizations. The aim is to engage

the public by having them learn about and contribute to range maps of diverse ants, beetles, and mosquitoes. These include vectors of human disease, invasive species, and economically important species whose ranges are shifting because of climate change.

Our collaborators at the New York Hall of Science (NYSCI) helped design, pilot, and refine retractable banners that introduce the public to DNA barcoding, orient participants in their roles as citizen scientists, and generate interest in upcoming events. NYSCI hosted *CDBN* events for the public and NYSCI interns. After training, collaborators at the Sweetbriar Nature Center (NY), the Long Island Aquarium (NY), the Cook Museum of Natural Science (AL), the HudsonAlpha Institute for Biotechnology (AL), and the Missoula Butterfly House and Insectarium (MT) also hosted public *CDBN* events. These activities included brief "tabletop" sessions introducing the use



We held a collection event for *CDBN* at Sweetbriar Nature Center in August where participants spent several hours learning sample collection techniques and gathering ant, beetle, and mosquito specimens.

In August, *CDBN* collaborators were introduced to DNA barcoding, laboratory techniques, bioinformatics, and program implementation during a five-day workshop.

of DNA to identify species; partial- or full-day, hands-on labs; DNA barcoding training during internships; specimen collection events; and a multiday DNA barcoding summer camp.

Despite limited facility access due to the pandemic and other natural disasters, the program supported more than 1,000 participants with a range of ages and skill levels. In total, 79 DNA barcodes were published to GenBank with citizen scientists as authors, including 22 previously unpublished barcode sequences and 24 new variants. Samples included a rare mosquito collected on Long Island (*Anopheles barberi*) whose larvae are generally sensitive to winter cold and a click beetle (*Diplostethus texanus)* collected outside of its known range in Louisiana.

The first set of *CDBN* banners was shipped out in December to collaborators at the Long Island Science Center (LISC). Image courtesy LISC.

DNA Barcoding in Undergraduate Classes

We continued to develop, disseminate, and assess DNA barcoding and metabarcoding as "formatted" solutions for post-secondary course-based research experiences (CUREs) through our $2 million NSF IUSE collaboration with James Madison University (JMU), CUNY City Tech, Bowie State University (BSU), and Austin Community College (ACC).

In June, we held a virtual five-day *DNA Barcoding for CUREs* workshop for 22 faculty from seven four-year universities, nine community colleges, and two high schools. Meeting our goal of reaching a diverse audience, 23% of participants were Black and 5% were Hispanic/Latino. The workshop covered the methods and logistics needed to teach DNA barcoding CUREs and included presentations by participants from the previous year's workshop describing their implementations. Supplemental videos demonstrating biochemical steps ensured all participants could master the protocols. Fifteen participants completed these steps using materials supplied by the DNALC and submitted samples for sequencing, allowing them to analyze their results while learning the bioinformatics steps during the workshop.

We also held a virtual metabarcoding CURE workshop during the summer. The workshop included 41 educators from 25 four-year and 11 two-year colleges and universities, including 8% Black and 8% Hispanic/Latino participants. In June, sample collection methods, experimental

design, and biochemistry were demonstrated, preparing participants to collect samples and then isolate and amplify DNA. JMU supported participant sequencing using their MiniSeq for 270 samples. In late June, participants learned to analyze their microbiome data and began plans for implementing CUREs in the classroom. During the workshop, seven participants from the previous year shared their efforts, providing insights into how to successfully implement metabarcoding CUREs. Two CURE institutes were also held at BSU, introducing barcoding and metabarcoding CUREs to 42 faculty.

During the year, DNA barcoding was taught in 32 classes reaching 896 students, and 25 classes included metabarcoding—taught to 431 students. As with the high school DNA barcoding programs, participation in these CUREs had positive impacts: 64.0% of survey respondents felt they were capable of going further in science after the experience, whereas more than half became more interested in technology, bioinformatics, and studying biology.

National Center for Biotechnology Education

The DNALC continued its work as a lead institution in the InnovATE*BIO* national biotechnology education center funded through NSF ATE, which supports training for America's workforce. Last year, we realized that Goal 4 of InnovATE*BIO* included a disparate set of activities that defied cohesion. This led to the concept of hubs to harness resident expertise and provide that expertise to the biotech community.

This year, we reorganized key InnovATE*BIO* activities into hubs as a means to enhance service to the community. The hubs support alumni and teachers; career options and entrepreneurship; high school to college pathways; workforce development; undergraduate research; supply chain models for training; and emerging technologies, including the hub we host, genomics, and a new immunotherapy hub.

In the fall, we added pages for each of the hubs to the center's website, aiming to position InnovATE*BIO* as a service organization that focuses on "What we can do for you." In this context, hubs function as national resources, as opposed to the traditional concept of regional resources. Structured interviews with community college biotech faculty members validated hubs as a useful way to organize InnovATE*BIO* activities. Given the positive feedback, design for an improved, web-search-optimized hubs site began at the end of the year.

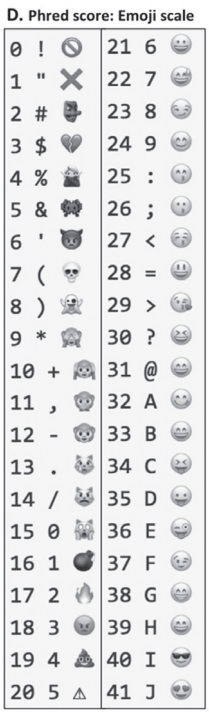
Through the Genomics hub, we continued to assist educators implementing genomics experiments and CUREs by supplying them with reagents, free sequencing, and technical or pedagogical support. We also supported development of a Sanger Sequencing Service at Liberal Arts and Science Academy (LASA) High School, Austin, funded by an ATE grant to Joseph Oleniczak (LASA and ACC) and Kissau Tchedre (ACC). Students from LASA's dual-credit biotechnology program perform DNA sequencing as a service, including for students at ACC, who visit the high school to learn Sanger sequencing.

Accessible Biomanufacturing to Teach Biotech Skills

In parallel to our work as part of InnovATE*BIO*, we started a new project to popularize the free isolation of *Taq* polymerase for classroom use in collaboration with Aron Kamajaya of Los Angeles Pierce College. Supporting the Center's supply chain hub, this ATE project is developing a biomanufacturing curriculum on the manufacture of *Taq* polymerase. We optimized simple isolation protocols to increase yield and constructed a new plasmid, p*SimpleTaq*, allowing constitutive expression of *Taq*, and removing the complexity and expense of induction. Expressing, purifying, and assaying *Taq* polymerase will familiarize students with all aspects of product manufacturing and quality control. Providing a free product to high school "consumers" will make polymerase chain reaction accessible to many more students.

In late fall, Dr. Kamajaya piloted *Taq* isolation with 12 Pierce College Biotech interns at the Pasadena Bio Collaborative Incubator. The interns prepared enough *Taq* extract to support 10,000 polymerase chain reactions (PCRs), and then developed PCR protocols for use by students in a microbiology course in the coming term.

NSF CyVerse



Although online training was the only option for faculty and students in 2020, constant disruptions and uncertainties made the pivot to virtual less than seamless. After plummeting more than 40% last year, *DNA Subway* usage rebounded in 2021 with all-time peaks of 48,035 registered users (12% increase), 1.18 million page views (29% increase from 2019), and 38,059 student projects (28% increase from 2020).

In addition to analyses on *DNA Subway* lines, we created new teaching resources for the command line. Working with long-time collaborator and DNALC workshop attendee Ray Enke from JMU, we published *A Fun Introductory Command Line Lesson: Next Generation Sequencing Quality Analysis with Emoji!*[*] in CourseSource, an online journal for biology and physics teaching. This lesson uses an application we developed on CyVerse to walk students through a quality control exercise analyzing next-generation DNA sequencing reads. Each DNA base sequenced is assigned a *Phred* score that indicates the probability it has been incorrectly called. A low score (e.g., 10) indicates a 1 in 10 chance of being wrong, and a high score (e.g., 40) indicates a 1 in 10,000 chance of being wrong—99.99% accurate. This fun lesson was developed as a *Jupyter* Notebook within the newly updated CyVerse *Discovery Environment*, allowing students to learn command line computing while working with real sequence data. Several of these notebooks and tools were also used by students in this year's new *Sequence a Genome* camp.

Finally, we continued our advanced training to cohorts of 130 graduate students, postdocs, and faculty in two 10-week online cohorts in our *Foundational Open Science Skills* course. We also reached another 363 faculty and students at various online training events and seminars including the International Society for Applied Microbiology Conference, North Carolina State Undergraduate Research Experience, James Madison University Center for Genome and Metagenome Studies workshop, and the NIH National Human Genomics Research Institute Short Course in Genomics.

Licensed Centers

With the continued closure of Regeneron's Sleepy Hollow campus, field trips and summer camps in the Westchester area remained virtual for the majority of the year. After an extensive fall e-mail campaign and several Open House events to reignite awareness of the new *Regeneron DNALC* facility, we hosted 260 students for in-person field trips. We also participated in a Virtual K–12 STEM Teacher Conference sponsored by Mercy College Center for STEM Education, where teachers were invited to learn more about DNALC programs in a virtual exhibition hall.

Campus restrictions prohibited the Notre Dame DNALC from conducting any in-person events through September. In summer, camps were offered virtually for 20 participants with materials kits shipped to students' homes, and processing of returned student samples handled onsite by staff. In-person programs resumed in the fall with 150 participants who attended an event at a Football Saturday, and a limited number of field trips.

[*]St. Jacques RM, Maza WM, Robertson SD, Lonsdale A, Murray CS, Williams JJ, Enke RA. 2021. A fun introductory command line lesson: next generation sequencing quality analysis with Emoji! *CourseSource* **8**. doi:10.24918/cs.2021.17

International Partnerships

China

Because of restricted international travel and leadership changes at Beijing No. 166 School in China, our collaboration was put on pause. Over the past three years, *Barcode Beijing* program students published 130 DNA sequences to GenBank, which included 122 species and 23 new variants. We think this noteworthy accomplishment—made by a single secondary school in China—may encourage more schools in China to engage in our brand of hands-on science.

In an effort to support science education for international students during the pandemic, we launched *Mentored On-Demand Camps*. These camps are a combination of our *Live Virtual* camps (taught live, but remote) and *On-Demand* camps (prerecorded instructional videos). In the hybrid camps, instruction is primarily through prerecorded videos, but several in-person sessions with DNALC educators are built into the schedule to provide time for Q&A, troubleshooting, and engagement with an instructor. This summer, 16 students from Keystone Academy, Shenzhen International School, and Guangdong International School participated in *Fun with DNA* and *DNA Barcoding Mentored On-Demand Camps*. Three teachers from Keystone Academy, a private school in Beijing, were trained to teach the camps using our prerecorded instructional videos and individually packed materials kits. Kit preparation and distribution was coordinated by *DNALC Asia* in Suzhou. One of the participating teachers reported, "The students were quite happy with the program and we were able to complete all experiments and activities."

We also began to advertise International Partner Membership modeled after our successful partnerships with independent schools in the New York metro area. Four international schools—Beijing No. 5 School, Shuangliu School, Keystone Academy, and Tsinglan School—have expressed interest. When travel restrictions are lifted, we will be able to send instructors to International Partner schools and host their students in New York for camps.

Michael Okoro and program participants at DNALC Nigeria.

DNA Learning Center Nigeria

Despite the challenges of the pandemic, *DNALC Nigeria* is now renovated and operating. Michael Okoro runs day-to-day operations and is supported by three teaching assistants and undergraduate interns who help deliver instruction.

More than 250 students and visitors attended workshops and research programs, including undergraduate courses and school visits. In January, the center hosted a metabarcoding workshop for students from Godfrey Okoye University, *DNALC Nigeria's* host institution. The group sampled microbial soil from lakes on the Ugwuomu campus. A second cohort in March included postgraduate students and faculty from other Nigerian institutions. Through a grant from the U.S. Consulate in Lagos, a series of DNA barcoding workshops kicked off in August. This program will ultimately reach 160 high school students and 40 teachers from Enugu State by its completion in March 2022. The center also participated in the DNALC *Barcoding US Ants* project and collected DNA from approximately 170 ant species in urban and rural parts of Southern Nigeria. Through a grant from the British Council, *DNALC Nigeria* participates in the *Innovation for African Universities* program, which supports the creation of STEM business and entrepreneurship through student training.

Dissemination at Professional Meetings

As in previous years, we continued to disseminate our programs at meetings. DNALC staff presented DNA barcoding and metabarcoding at the CSHL Genome Informatics Conference and NIH SEPA SciEd Conference. Our data science programs were also presented at the International Applied Microbiology Conference, while our analysis of American Science Education was presented at Cell Bio Virtual 2021, an American Society for Cell Biology/European Molecular Biology Organization meeting.

Lab Instruction and Outreach

Mid-pandemic, 2021 programs began with a complete shift to virtual instruction. As schools struggled to adjust to schedule changes, staggered cohort instruction, and restrictive health and safety guidelines, in-person field trips and school visits were not permitted. We provided remote field trip experiences to 10,179 students through demonstrations and kit-based experiments.

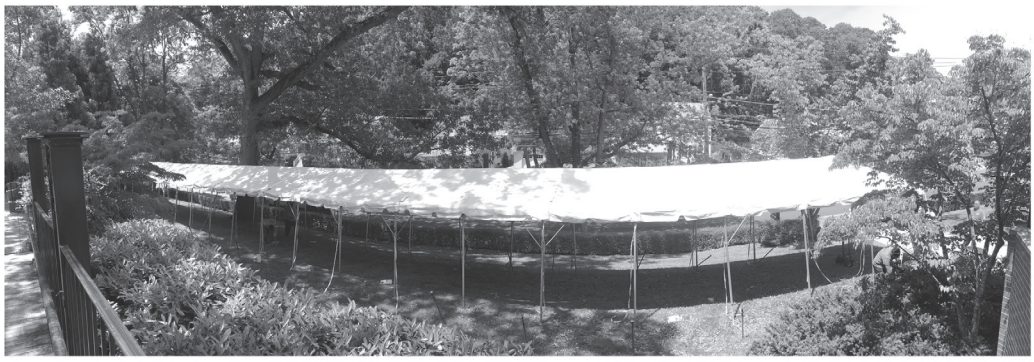
In lieu of school field trips, we were able to begin providing safe, on-site instruction in winter—a nice change of pace for students who were limited to remote instruction at school. Sixteen students attended AP Biology prep courses on Saturdays to complete a series of labs that either are required by the College Board or include techniques and concepts that are integral to the curriculum. Five other students attended a Saturday *Green Genes* series—one of the summer camps that could not be taught virtually. One hundred and ninety middle and high school students opted to forgo their winter and spring breaks to participate in a series of fun and engaging daily lab sessions on topics including: DNA structure and function, industrial enzymology, genetic engineering, forensics, and GMOs.

September brought less restrictive distancing requirements and an optimistic return to in-person instruction. A total of 3,135 students attended in-person lab field trips at our four facilities: Dolan DNA Learning Center, *Harlem DNA Lab*, *Regeneron DNALC,* and *DNALC NYC* at City Tech. In-school programs reached 1,789 students. Footlocker kits were used by 440 students, 64% of whom were conducting independent research through *BLI*, *UBP,* and *UBRP*.

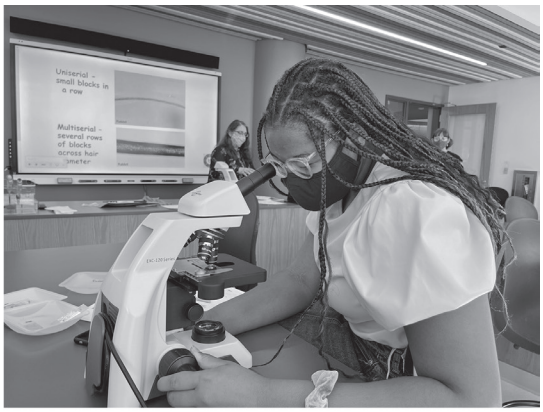
Field trip scholarships were provided for 174 students from Uniondale School District, Ossining School District, Danbury School District in Connecticut, Kearsarge School District in New Hampshire, Passaic Academy of Science and Engineering, and Jack and Jill of Central New Jersey. An additional 326 sixth grade students from Central Islip School District participated in a kit-based series of virtual labs supported by a grant from the National Grid Foundation. This year, 292 (12%) of the students who attended in-person field trips and virtual programs at the *Harlem DNA Lab* and *DNALC NYC* at City Tech came from Title I schools that qualified for tuition assistance supported by the William Townsend Porter Foundation.

In summer, we returned to in-person camps at the Dolan and NYC locations. With COVID-19 guidelines that included fully vaccinated instructors, distanced workstations, individual equipment, and daily health screening, we hosted 542 campers on Long Island, and 218 in NYC. We provided scholarships for 88 students. Another 26 campers participated in off-site workshops at Francis Lewis High School in Queens and Friends Academy in Locust Valley. One hundred and seventy-two students attended *Virtual Live* or *On-Demand* camps, including 48 middle schoolers who attended through City Tech's *Bridging the Gap,* a Science and Technology Entry Program (STEP) for pre-college students.

Now in its third year, the Science, Technology, and Research Scholars (STARS) program has become a community of young researchers from groups that have been traditionally underrepresented in STEM. This year we accepted 15 students from nine districts across Long Island and New York City for a two-week course in Cold Spring Harbor. Thanks to a donation from CSHL Trustee Laurie Landeau, half of our students benefitted from bussing provided by Suffolk Transportation. To kick things off, students and their families joined us for a Saturday afternoon orientation featuring STARS alumni joining via Zoom to present their summer research experiences.



Returning to in-person summer camps came with curb-side check-in and temperature checks (*bottom left*), an enormous tent to accommodate socially distanced lunches (*top*), and virtual parent days for middle school camps (*bottom right*).



DNALC NYC held its inaugural summer of camps, welcoming 183 students in 20 camps.

This year Harrison Banks (2020) had a summer internship in Robert Martienssen's Laboratory, and Diana Benedicto-Jimenez (2019) worked in Chris Vakoc's lab as part of the CSHL Partners for the Future Program. STARS students also made their own contributions to the rollout of the COVID-19 vaccination effort. Students created more than three dozen videos—posted to YouTube and the DNALC website—on the science behind COVID-19 and historical healthcare

Jason Williams prepares Duckweed samples during the new *Sequence a Genome* camp.

The STARS camp returned to in-person instruction for its third year; students perform lab experiments in the computer lab (*top row*). Dr. Lloyd Trotman spoke to the students about his research during the two-week program (*bottom left*). STARS participants and educators pose at the entrance to the DNALC (*bottom right*).

inequities in minority communities. We also completed an assessment of the impact of STARS on the first two student cohorts. One hundred percent agreed that STARS helped them prepare for a future research experience, 94% agreed that they felt mentored by STARS instructors, and 87% agreed that STARS helped them set goals for their careers. Importantly, 84% agreed that STARS helped them find role models with whom they identified, and 83% reported doing some activity that follows up on their interest in STEM—including high school research internships, undergraduate STEM majors and research internships.

A new summer camp, *Sequence a Genome*, brought students "into the genome age" by allowing them to fully sequence a genome using Oxford Nanopore DNA sequencing. In Nanopore sequencing, DNA is drawn through pores embedded in a membrane, generating an electrical signal that is reconstructed into DNA sequence. A single, handheld Nanopore device can generate billions of base pairs of DNA sequence in hours for about $1,000. We chose to sequence a strain of duckweed, *Spirodela polyrhiza*, which has a relatively small 150 million base pair genome. This plant, commonly seen growing on ponds, is being studied by the CSHL Martienssen laboratory as a candidate for biofuel production and carbon sequestration. Guest lectures included Sara Goodwin from the CSHL Genome Center and Alex Harkess from the HudsonAlpha Institute for Biotechnology, who has sequenced other duckweed strains. Students extracted DNA that was used to generate more than seven billion base pairs of sequence, and went on to do bioinformatics needed to analyze the data. Several continued to meet virtually in the fall to work towards completion of an entire assembly.

In partnership with CSHL Women in Science and Engineering (WiSE), we hosted the fifth annual WiSE *Fun with DNA* summer camp. Ten young female science enthusiasts had the

opportunity to meet and interact with engaging role models pursuing careers in the sciences. Each afternoon, the girls participated in WiSE activities on cancer research, neuroscience, and gene expression. On the final day of camp, parents and campers were invited for a guided tour of the main CSHL campus, and an outdoor lab activity.

The *Partner Member* program continued to provide custom instructional sequences and advanced electives for six independent schools in New York City, including kit-based virtual instruction in spring for a fully remote hands-on experience.

- Research teams from Fontbonne Hall Academy used DNA barcoding to produce novel GenBank entries for flowering plants and snails. AP Biology students studied viruses, vaccines, and protein structure and created 3D-printed models of the SARS-CoV-2 spike protein.
- At Marymount School of New York, genetics programs were incorporated as key parts of the biology curriculum at multiple grade levels. Students in *Independent Student Research* tested loop-mediated isothermal amplification (LAMP) PCR of the human PV92 *Alu* locus as a method for at-home PCR that does not require a thermal cycler.
- Sacred Heart Greenwich adopted a quarter membership with implementation of a DNA barcoding unit for its Science Research students.
- Lycée Français de New York implemented genetics programs in their AP Biology courses.
- The Chapin School coordinated genetics programs at several grade levels, including the advanced Molecular Genetics elective. Middle school students analyzed PCR product sizes from a mock locus to genotype a hypothetical family when studying a genetic disease.
- St. David's School integrated basic genetics with existing curricula in grade five. Grade eight used DNA barcoding to survey the ants of Central Park.

With the success of the *Partner Member* program in New York City we updated our school membership programs to include three different options. *Sustaining Members* receive field trips, advance registration, priority reservations for teacher workshops and special events, and opportunities to audit summer camps. *Associate Members* receive all the Sustaining benefits, with an additional 10 days of customized instructional time for project-based research, research mentorship, enrichment of existing curricula, or focused faculty training. Finally, *Partner Membership*—for local *and* international schools—includes 20 days of customized instructional time that can include developing and co-teaching advanced electives, access to emerging DNALC projects that incorporate cutting-edge techniques, and weeklong camps during school breaks. This year we were thrilled to add two new *Partner Members* from Long Island—Massapequa School District and Long Beach City School District—schools that have maintained sustaining membership for many years and want to strengthen their science brand through increased exposure and opportunity. Glen Cove City School District, a first-time DNALC member, joined as an *Associate Member* with the same goals.

As part of other ongoing local partnerships, the *Molecular and Genomic Biology* elective returned to Cold Spring Harbor High School after a pandemic hiatus. The 12 enrolled students spent the last two periods of every other school day at the DNALC, immersed in hands-on experiments in DNA barcoding, human and plant genomics, bacterial genetics and gene cloning, and bioinformatics. Similarly, 13 students from St. Dominic High School in Oyster Bay participated in a half-year *Molecular and Genomic Research* elective taught at the DNALC. As part of their course requirements, all students in both classes participated in *Barcode Long Island*.

Our Ötzi the Iceman exhibition had 100 in-person visitors, and 870 attended virtual museum tours guided by educators remotely, while the *DNALC Live* pre-recorded Ötzi museum tour on YouTube had a whopping 1,755,340 views in the year! Nineteen *Saturday DNA!* sessions,

both virtual and in-person, drew 200 participants. Virtually, participants used kits to extract DNA from strawberries and build their own gel electrophoresis chambers. They also explored how population size contributes to genetic drift and how our ancient ancestors contributed to modern human genetics. In-person participants learned about Mendel's laws of heredity, created works of art inspired by our beautiful Joe Rossano BOLD exhibition, and unraveled some forensic mysteries surrounding Ötzi the Iceman.

The "Meet a Scientist" remote lecture series continued to connect high school and public audiences with CSHL researchers, drawing 90 participants. Guest presenters shared their research, and stories of their journeys to becoming scientists. Dr. Lloyd Trotman presented his work on cancer metastasis and 3D analysis; Ph.D. student Miriam Ferrer Gonzalez talked to us about the whole-body response to cancer; graduate student Alexa Pagliaro delivered a talk on how maternal experience shapes brain activity; Dr. Hannah Mayer shared her work on how T cells distinguish friend from foe in our immune systems; Dr. Peter Koo introduced his research understanding gene regulation through deep learning; Dr. Doreen Ware presented her research on maize and how plants manage fluid genomes to adapt and evolve; and Dr. Christopher Vakoc talked about inventing new cancer medicines. All of the presentations are available on the DNALC website.

This year our collaboration with the CSHL School of Biological Sciences was put on hold. As part of their required curriculum, first-year graduate students usually work with DNALC instructors to develop skills needed to communicate science to a variety of audiences. Students complete 12 half-day sessions in which they progress from observation to co-instruction, and then independent teaching of lab classes. The 2021 cohort will participate in training in 2022.

BioMedia Visitation and Projects

In 2021, 7.68 million visitors accessed our suite of multimedia resources, a 144% increase from the previous year! This rise is attributed to another stunning increase in visitation to our YouTube channel, which received 4,633,125 views—274.7% over 2020! Watch time increased to 505,648 hours (627%) and we added 22,721 (170%) new subscribers. Two videos recorded early in the pandemic went viral:

- "Museum Tour: Ötzi the Iceman" (1.8+ million lifetime views) presented by Amanda McBrien,
- "What DNA Says About Our Human Family Episode II, Ancient Relatives: Neanderthals and Denisovans" (874+ thousand lifetime views) presented by Dave Micklos.

Google Analytics counted 2.6 million visits to DNALC websites, 84.6% of the prior year. With the end of support for Adobe Flash in all browsers at the end of 2020, we faced losing several of our content-based websites that were built using this technology. Fortunately, we were able to implement Ruffle, a Flash emulator that could be incorporated on some sites, including *Inside Cancer*, *Your Genes, Your Health*, and *DNA Interactive*. Some sites have been redirected to older non-Flash versions, such as the *Eugenics Archive* and *Weed to Wonder*. Unfortunately, *Genes to Cognition* and *Lab Center* are completely dependent on Flash and could not be revived. However, much of the content from these sites is available in alternative formats through our homepage's media search.

In 2021, 439,672 *3D Brain* and *Gene Screen* smartphone/tablet apps were downloaded. The 10-year-old *Weed to Wonder* E-book for iPads has been removed from the Apple App Store because the software used to develop it can no longer be used to update the app.

With the opening of *DNALC NYC* and the distance at-home instruction placed between teachers and students, the time was right to revamp how we reach parents and students with our summer camps. We decided to work with a digital marketing company, WebFX, to develop a new

online strategy. This included search engine optimization (SEO), which makes our website more "visible" to search engines by improving our ranking as an authoritative source and by optimizing for search terms people used to find us. WebFX has given us insight into how ranking factors used by Google (e.g., page authority, domain authority) can be improved to increase site traffic. In addition to SEO, WebFX has significantly enhanced our ability to reach parents directly through Google Ads and Facebook. Since 2016, we have tracked new summer camp participants by asking, "How did you learn about DNA Learning Center camps?" Traditionally, only about 3% of new summer camp parents find out about us through internet ads and search. Our marketing efforts increased that to 22% this year. We also benefitted from an ongoing nonprofit Google Ads grant that generated 117,291 impressions and 12,349 clicks—the equivalent of $19,546 in advertisement spending.

Staff and Interns

Congratulations are in order for DNALC staff! In April, DNALC Assistant Director Amanda McBrien was recognized by Crain's New York Business 2021 list of Notables in Nonprofits and Philanthropy. This year's list was "in celebration of 57 New York heroes who facilitated or led nonprofit work in the face of the pandemic."

Our Executive Director Dave Micklos was named the winner of the 2021 Bruce Alberts Award for Excellence in Science Education by the American Society for Cell Biology. This prestigious award recognizes career-spanning achievements in biology education.

The opening of *DNALC NYC* at City Tech brought a new staff to Brooklyn, and the Dolan DNALC also saw some staff changes.

Arden Feil started in June. She has both a Bachelor's and Master's degree in Biology from Wesleyan University. Her research focused on how chromosomes interact and exchange genetic information during meiotic cell division. She also majored in Science in Society, which taught her to consider the cultural, social, and political significance of her science training. This combination cemented her interest in making science accessible and relatable to everyone. She participated in science outreach efforts, from mentoring students to teaching after-school programs and summer camps in NYC. After graduating, she worked as a Research Specialist at the University of Chicago in a lab that studies cell signaling in tissue development. She realized that, although she loved the inquiry-driven, dynamic nature of lab work, she missed her favorite part of science—getting to share it with other people! Now, as an educator at *DNALC NYC*, she is excited to combine her research and teaching backgrounds to facilitate engaging learning experiences.

Andrea Mahee came aboard in June, as the *DNALC NYC* Administrative Manager. She brought to the table a myriad of skills including C-suite support; curating and managing special events; and conducting workshops, meetings, and trainings; and she is multilingual in French, Spanish, and Italian. She spent many years in the development arena; she was Chief of Staff at the Third Avenue Business Improvement District, executive assistant at the Bronx Council on the Arts, and development associate for the South Bronx Overall Economic Development Corporation.

In July, Kelsie Anson joined the Brooklyn team. Hailing from rural Wyoming, it was family trips to Yellowstone National Park that ignited her interest in science. She had a special interest in the thermophilic bacteria that lived in the hot springs in the park—where one of our important DNA engineering molecules, *Taq* polymerase, comes from! She earned a degree in molecular biology from Colgate University, where her senior research project explored how genetics might dictate social behavior in an ant colony. After graduation, she worked as an analytical chemist for an environmental testing company and then later took a research position at Rocky Mountain Laboratories in Montana, studying prion diseases—deadly brain diseases most common in elk herds in the Western

United States and Canada. In 2020, she was awarded her Ph.D. in biochemistry from the University of Colorado in Boulder. Her research used fluorescent sensors to measure fluctuations of zinc ions and determine the role these ions played in crucial cell signaling pathways. While at the University of Colorado, she also created an interactive teaching case study in which undergraduates learned how to genetically engineer bacteria to break down environmental toxins. Kelsie also tutored middle schoolers on optical physics and occasionally starred on the college podcast, *Buffs Talk Science*.

In July, the DNALC welcomed Donna Smith to the administrative staff. A graduate of Harborfields High School, Donna received her Bachelor's Degree in History from Moravian University, previously known as Moravian College, in Bethlehem, Pennsylvania. Donna worked for 12 years at Akorn Pharmaceuticals as a Lab Clerk in the Quality Control Raw Materials Lab, working on file management for new drug submissions to the FDA; sample intake for in-house testing; contract lab sample submissions for outside testing; and requisition creation for chemicals, standards, lab, and office supplies. Her duties at the DNALC include scheduling lab visits to the *Regeneron DNALC*, statistical record keeping, ordering office supplies, answering phones, and greeting visitors and students.

Anna Feitzinger started at *DNALC NYC* in August. After graduating with a degree in Chemistry from Hunter College, she worked as a research associate at the Skirball Institute at New York University in the developmental genetics department—gravitating to developmental and molecular biology. She never tired of imaging a glowing population of cells migrating across a live developing zebrafish embryo. It was this experience that prompted Anna to pursue a Ph.D. at the University of California, Davis, where she joined an evolution of development lab in the evolution and ecology department. Her interests in quantitative and developmental biology were united as she worked on characterizing the natural variation of gene expression of early-stage fruit fly embryos. While at UC Davis, she also worked as a teacher's assistant for a genetics and molecular biology lab course. It gave her great joy to communicate science and guide students during their first DNA extractions, gel electrophoresis, cloning, and fruit fly crosses.

Kelly Eames joined the Dolan staff as an educator in August. She has her parents to thank for encouraging her scientific endeavors and introducing her to genetics and microbiology—and more importantly to the DNALC. As an alum of our summer camps and *Saturday DNA!* programs, Kelly applied skills learned at the DNALC in her college courses. In 2013, she earned a bachelor's degree in biology and chemistry at Molloy College, then pursued a master's degree at Trinity College in Dublin, Ireland. The program focused on biodiversity and wildlife conservation; her thesis examined the different educational methods used to teach visitors to zoos and aquaria about conservation and community action. Engaging guests in interactive learning exercises was shown to better their understanding, especially when compared to passive learning (such as reading or videos). Now, 10 years after her last DNALC camp, Kelly works alongside some of the people who introduced her to these topics in high school. She is excited to offer the same opportunities to students today.

In November, Jack Kellogg joined the staff in preparation for reopening our Regeneron site. Growing up on farmland in a small rural town outside of Buffalo, New York, Jack loved exploring the woods and learning about the varied species of trees on their property. These explorations coupled with watching every episode of the nature docuseries, *Planet Earth*, sparked his fascination with the different behaviors, life histories, and morphologies of organisms. He attended Rutgers College, Newark, majoring in Biology. He had the privilege to join an insect evolution lab led by Dr. Jessica Ware. As an undergrad, he used computer vision to examine wing color patterns in tropical butterflies and investigated the effect major weather events have on dragonfly dispersal in New Jersey. He traveled to the Amazon as part of a field course focused on different methods in entomological research, including tree canopy surveys and bait traps. He and his group also did a small study related to butterfly mimicry. For his master's thesis, he used DNA barcoding to study the population genetics of aquatic insect communities.

We say goodbye to a number of our staff who have moved on in 2021: from our administrative department, Mary Lamont and Colette Riccardi; educators Melissa Lee, Louise Bodt, and Lina Ruiz-Grajales; and senior programmer Cornel Ghiban.

In 2003, Mary Lamont began as an Administrative Assistant for DNALC *West*, located within a Northwell Health facility in Lake Success. In an office no bigger than a closet, she quickly rose to Administrative Manager and Manager of Offsite Programs. DNALC *West* field trips included a tour of the core blood testing facility, often led by Mary. With proficient politeness, a keen "no-nonsense" attitude, and outstanding communication, she built excellent relationships with the schools and teachers who booked classes through her and did a wonderful job supporting instruction at *West*. When not at the DNALC, she is a traveling country singer with her husband. After 18 years of loyalty and dedication, Mary retired to continue singing and enjoy being a grandmother.

Colette Riccardi joined the administrative staff as our receptionist in March 2018. Her duties included answering phones and greeting guests at the DNALC, ordering all office supplies and arranging repairs for office equipment, and entering and organizing data. As an "Excel champion," she tackled complicated projects with ease. Her work ethic, dependability, and lovely personality also gave her an advantage when interacting with the public and forging relationships for the DNALC. Colette left the DNALC in February to begin preparing for full-time motherhood; her son Leo was born in May.

Harlem DNA Lab Manager Melissa Lee joined the DNALC family in 2011. She used her background in science education and lab management to deliver hands-on labs to thousands of students. Melissa managed the *Harlem DNA Lab*, taught high school and middle school classes, managed interns, and ran the footlocker equipment rental program. She also played a large role in the setup and management of the *Regeneron DNALC*. She directed NYC teacher training and worked with many of our partner schools. Melissa assisted in grant writing and in the execution of those grants, and she mentored students for symposiums and competitions. With a passion

2021 New staff: (*Top*) Arden Feil, Andrea Mahee, Kelsie Anson, and Donna Smith; (*bottom*) Ann Feitzinger, Jack Kellogg, and Kelly Eames.

for outreach and research, she became immersed in *UBP* and *UBRP*. Through her tenure at the DNALC, she ardently and continuously promoted our mission across NYC. In June, she decided to put her smart board markers away to become a full-time mom to her one-year-old son, Alex. We will miss her exuberant personality, her fine leadership skills, and her infectious enthusiasm.

Louise Bodt, a Brooklyn native, began her DNALC journey in August 2019 as an educator and *UBP* manager based out of our City Tech temporary lab. In addition to the instruction of high school and middle school students, she helped train and supervise interns in Brooklyn. She was the relationship manager for one of our partner schools, the Fontbonne Hall Academy for Girls in Brooklyn, as well as a *UBP* program manager. Louise also participated in grant writing for the *UBP*. At the onset of the pandemic, she helped to develop and promote virtual classes. Although she left the DNALC before the opening of *DNALC NYC*, she played a large role in promotion and outreach. She left in June to pursue her Ph.D. in Evolutionary Biology in the graduate program at the University of Chicago.

Lina Ruiz-Grajales joined the DNALC in November 2019 as an instructor for *Regeneron DNALC*, which had its grand opening that December. In addition to teaching high school and middle school, her strong research background allowed her to shoulder much of the lab preparation needed at our brand-new facility. When the facility closed temporarily in March of 2020, Lina shifted her focus to the development of virtual lessons and protocols. She also participated in developing protocols and laboratory techniques for *Barcoding US Ants*. Lina left in August to pursue her Ph.D. in Biological Sciences at Columbia University.

Cornel Ghiban, our computer programmer, began his odyssey at the DNALC in late 2006. Prior to arrival from his home country of Romania, he did contract work for us developing the "back-end" of our websites. He became the DNALC's "nervous system"; Cornel designed, updated, and maintained several tools used to engage students and researchers in biotechnology. He played a large role in creating and supporting new features for *DNA Subway* (maintaining the Blue Line databases and QIIME, adding new and updating old transcriptomes for the green line when necessary and making other changes as requested). As our enrollment increased, he developed online registration systems that could collect and update information from parents, teachers, students and researchers. He has also sustained our legacy websites, such as *DNAi*, *Inside Cancer*, and *YGYH*. When the pandemic hit, he facilitated our transition to virtual programming and helped us launch all of our virtual content—developing tools for the *On Demand* Program. Cornel is now working remotely as a software backend engineer for Vinli Inc. based in Dallas, Texas, a cloud-based data intelligence platform for mobility and transportation.

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. This year an amazing group of interns helped out, and we said farewell as others left for college:

High School Interns

Raymond Bailey, Longwood High School	Megan Jung, Jericho Senior High School
Raquel Belkin, Syosset High School	Sandhya LoGalbo, St. Dominic High School
Hayden Calabretta, Cold Spring Harbor High School	Brianna MacDonald, Commack High School
Kevin Darcy, Longwood High School	Rachel Morina, Huntington High School
Lauren Graziosi, Syosset High School	Maggie Wang, Northport High School
Min Hur, Jericho Senior High School	Matthew Warner, Saint Mary's High School

High School Interns Departing for College

Timothy Broadbent, Colby College	Neal Mehta, Boston College
Kaela Deriggi, The George Washington University	Aveline Roderick, Boston College
Ethan McGuinness, Northeastern University	

College Interns

Taehwan Cha, New York University	John O'Hara, University of Richmond
Christopher Cizmeciyan, Stony Brook University	Michael Stabile, Cornell University
Isabella Martino, Stony Brook University	Nicholas Stabile, University of Notre Dame
Jillian Maturo, Boston College	

Sites of Major Faculty Workshops

Program Key: *Middle School* High School **College**

Location/State	*Institution*	*Year(s)*
VIRTUAL	Host: **Atlanta University Center Consortium, Atlanta, Georgia**	**2021**
	Host: **Bowie State University, Bowie, Maryland**	**2020**
	Host: **DNA Learning Center, New York**	**2020**
	Host: ***Harlem DNA Lab* and *Regeneron DNALC*, Sleepy Hollow, New York**	**2020**
	Co-Host: **James Madison University, Harrisonburg, Virginia**	**2020, 2021 (3×)**
	Host: **North Carolina State University, Raleigh, North Carolina**	**2021**
	Host: **Quantitative Undergraduate Biology Education and Synthesis (QUBES) Project**	**2020**
	Co-hosts: **University of Arizona, Tucson, Arizona & DNA Learning Center, NY**	**2020**
ALABAMA	University of Alabama, Tuscaloosa	1987–1990
	HudsonAlpha 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, 2019–2020**
	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	**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	Bio-Link 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	2015–18
	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	**2014–15, 2018–19**
	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, 2021
	Environmental Science Center, Bergen Beach, Brooklyn	2015–2016
	Fostertown School, Newburgh	*1991*
	Harlem DNA Lab, East Harlem	2008–2009, 2011–1203, 2016–2019
	***Harlem DNA Lab*, East Harlem**	**2015–2016**
	Huntington High School	1986
	Irvington High School	1986
	K-12 Summer Institute, Kerrville	2019
	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 Brook	**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
	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
	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 System for Molecular Biology Conference, Dublin**	**2015**
	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	Godfrey 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, Fiskebackgkil	1995
	Uppsala University	2004
THE NETHERLANDS	**International Chromosome Conference, Amsterdam**	**2007**
	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**

Workshops and Visitors

January 13	*Saturday DNA!* "Virtual Lab: DNA Extraction from Strawberries," DNALC
	"Meet a Scientist: Dr. Lloyd Trotman," Virtual Event, DNALC
January 16	*Saturday DNA!* "Virtual Lab: DNA Extraction from Strawberries," DNALC
January 20	*Saturday DNA!* "Virtual Lab: As the Worm Turns," DNALC
January 23	*Saturday DNA!* "Virtual Lab: As the Worm Turns," DNALC
January 26	NIH *Barcoding U.S. Ants* Mini Virtual Symposium, DNALC
January 27	*Saturday DNA!* "Virtual Lab: Cracking the Code," DNALC
January 28	Agricultural Biotechnology: Emerging Technologies and Insights Mini Virtual Symposium 2021, "Agricultural Genomics: The Rise of Genomes," "Emerging Trends in Agricultural Diagnostics," "Clones, Carbon and Climate Change: The Epigenetics of Oil Production," The InnovATE*BIO* National Center for Biotechnology Education, Austin Community College, Austin, Texas
January 29	NIH *Citizen DNA Barcode Network* Collaborator Virtual Workshop, DNALC
January 30	*Saturday DNA!* "Virtual Lab: Cracking the Code," DNALC
	Green Genes Workshop, DNALC
	AP Bio/Restriction Analysis Part I Workshop, DNALC
February 1	NIH *Citizen DNA Barcode Network* Collaborator Virtual Workshop, DNALC
February 3	*Saturday DNA!* "Virtual Lab: DIY Electrophoresis Chamber," DNALC
	NIH *Citizen DNA Barcode Network* Collaborator Virtual Workshop, DNALC
February 5	NIH *Citizen DNA Barcode Network* Collaborator Virtual Workshop, DNALC
February 6	*Saturday DNA!* "Virtual Lab: DIY Electrophoresis Chamber," DNALC
	Green Genes Workshop, DNALC
	AP Bio/Restriction Analysis Part II Workshop, DNALC
February 13	*Green Genes* Workshop, DNALC
	AP Bio/Bacterial Transformation Workshop, DNALC
February 16	"DNA 101" Workshop, DNALC
	Bacterial Transformation Workshop, DNALC
	DNA Barcoding Virtual Workshop, DNALC
February 16–19	*Urban Barcode Research Program Conservation Genetics* Virtual Workshop, *DNALC NYC* at City Tech
February 17	"Got Lactase?" Workshop, DNALC
	"Forensics: Trace Evidence," Workshop, DNALC

February 18	NIH *Citizen DNA Barcode Network* Collaborator Virtual Workshop, DNALC
February 19	"Gene Therapy," Workshop, DNALC
	"Restriction Analysis," Workshop, DNALC
	DNA Barcoding Virtual Workshop, DNALC
	NIH *Citizen DNA Barcode Network* Collaborator Virtual Workshop, DNALC
February 20	*Green Genes* Workshop, DNALC
	AP Bio/Human DNA Fingerprinting, Workshop, DNALC
February 23	*DNA Barcoding* Virtual Workshop, DNALC
February 25	"DNA Fingerprinting," Workshops, DNALC
February 26	*DNA Barcoding* Virtual Workshop, DNALC
February 27	*Green Genes* Workshop, DNALC
	AP Bio/DNA Barcoding, Part I Workshop, DNALC
March 2	*DNA Barcoding* Virtual Workshop, DNALC
March 3	*Saturday DNA!* "Virtual Lab: Selection Detection," DNALC
March 5	*DNA Barcoding* Virtual Workshop, DNALC
March 6	*Saturday DNA!* "Virtual Lab: Selection Detection," DNALC
	Green Genes Workshop, DNALC
	AP Bio/DNA Barcoding, Part II Workshop, DNALC
March 9	*DNA Barcoding* Virtual Workshop, DNALC
March 11	"Meet a Scientist: Miriam Ferrer Gonzalez," Virtual Event, DNALC
March 12, 16	*DNA Barcoding* Virtual Workshop, DNALC
March 17	NIH *Barcode Long Island* Virtual Open Lab, DNALC
March 19, 23, 26	*DNA Barcoding* Virtual Workshop, DNALC
March 23	*DNA Barcoding* Virtual Workshop, DNALC
March 26	*DNA Barcoding* Virtual Workshop, DNALC
March 27	*AP Bio/Restriction Analysis,* Part I Workshop DNALC
March 29–April 2	*Urban Barcode Research Program* DNA Barcoding and Bioinformatics Virtual Workshop, *DNALC NYC* at City Tech
	Spring Break Workshops, *The Diversity of Life*, *Human Mitochondrial Sequencing: Part 1*, *Infectious Disease*, *Bacterial Transformation*, *Mendelian Genetics: Build a Creature*, *Restriction Analysis*, *Agar Art*, *Purification of Green Fluorescent Protein*, *Human Mitochondrial Sequencing*, DNALC
April 3	*AP Bio/Restriction Analysis,* Part II Workshop, DNALC
April 7	Informational Session, Virtual, *DNALC NYC* at City Tech
	DNA Barcoding Virtual Presentation for iNaturalist Discord Server Group, San Francisco, California, DNALC
April 8	"Meet a Scientist: Alexa Pagliaro," Virtual Webinar, DNALC
April 10	*Saturday DNA!* "Virtual Lab: Candy Bar Phylogenetics," DNALC
	AP Bio/Bacterial Transformation Workshop, DNALC
April 14–15	Informational Sessions, Virtual, *DNALC NYC* at City Tech
April 16	NSF Northeast Big Data Innovation Hub, "Student Data Corps Data Science Career Panel," Virtual Public Lecture, Data Science Institute, Columbia University, New York, New York
April 17	*Saturday DNA!* "Virtual Lab: History and Mystery of Cell Theory," DNALC
	AP Bio/Human DNA Fingerprinting Workshop, DNALC
	NIH *Barcode Long Island* Virtual Open Lab, DNALC
	DNA Extractions at Brooklyn Outdoor Earth Day Event, Brooklyn, New York
April 19	Earth Day/March for Science Virtual Science Expo, New York, New York
April 21	Informational Session, Virtual, *DNALC NYC* at City Tech
April 24	*Saturday DNA! "*Virtual Lab: Less is More: Population Size and Genetic Drift," DNALC
	AP Bio/DNA Barcoding, Part I Workshop, DNALC
April 28	NIH *Barcode Long Island* Virtual Open Lab, DNALC
May 1	*Saturday DNA!* "Virtual Lab: Microbe Mania!" DNALC
	AP Bio/DNA Barcoding, Part II Workshop, DNALC
May 5	Informational Session, Virtual, *DNALC NYC* at City Tech
May 8	*Saturday DNA!* "Virtual Lab: Ancient Ancestors & Me," DNALC
May 13	"Meet a Scientist: Dr. Hannah Meyer," Virtual Webinar, DNALC
May 14	*Mystery of Anastasia* Virtual Presentation for Siemens Healthineers, Malvern, Pennsylvania, DNALC
May 15	*Saturday DNA!* "Virtual Lab: EYE-volution!" DNALC
May 19–20	Informational Sessions, Virtual, *DNALC NYC* at City Tech
May 20–21 24–25	CyVerse Data Carpentry Genomics Virtual Workshop, Atlanta University Center Consortium, Atlanta, Georgia, DNALC
May 24–27	NIH SciEd 2021 Virtual Conference, "*Citizen DNA Barcode Network*: Support for Community-Based, Hands-On Science," Poster Session, Washington, D.C.

May 26–27	Informational Sessions, Virtual, *DNALC NYC* at City Tech
May 26–28	Introduction to Bioinformatics with CyVerse Virtual Workshop, North Carolina State University, Raleigh, North Carolina, DNALC
June 1	*Urban Barcode Project/Urban Barcode Research Program/Barcode Long Island/US Ants* DNA Barcoding Virtual Student Symposium, DNALC and *DNALC NYC* at City Tech
June 7–16	*DNA Barcoding* for CURES Virtual Workshop, James Madison University, Harrisonburg, Virginia, DNALC
June 8	International Applied Microbiology Virtual Conference 2021, "Introduction to RNA-Seq with the Kallisto and Sleuth Workflows," Faculty/Student Workshop, Society for Applied Microbiology, London, United Kingdom
June 9–10	NYC DOE STEM Virtual Conference, "The *Urban Barcode Project*—Hands-on Science at Home and in School," Teacher Workshop, Expo Hall, Long Island City, New York
June 10	"Meet a Scientist: Dr. Peter Koo," Virtual Webinar, DNALC
June 17	*DNA Barcoding* Training with Long Island Science Center, DNALC
June 21–25	*DNA Metabarcoding* for CURES Virtual Workshop, James Madison University, Harrisonburg, Virginia
June 28–July 2	*DNA Science* Workshop, DNALC
	World of Enzymes Workshop, DNALC
	Green Genes Workshop, DNALC
	DNA Science Workshop, DNALC
	Fun with DNA Workshop, DNALC
	DNA Science Workshop, *DNALC NYC* at City Tech
	Fun with DNA Workshop, *DNALC NYC* at City Tech
	Forensic Detectives Workshop, *DNALC NYC* at City Tech
	Urban Barcode Project Teacher Training Workshop, *DNALC NYC* at City Tech
June 29–July 1	CGEMS 2021 – Getting Started with R and CyVerse Virtual Workshop, James Madison University, Harrisonburg, Virginia, DNALC
July 5–9	*Genome Science* Workshop, DNALC
	Fun with DNA Workshop, DNALC
	Forensic Detectives Workshop, DNALC
	World of Enzymes Workshop, DNALC
	Fun with DNA Workshop, *DNALC NYC* at City Tech
	World of Enzymes Workshop, *DNALC NYC* at City Tech
	Bridging the Gap STEP: Fun with DNA Virtual Workshop, DNALC
July 12–16	*DNA Science* Workshop, DNALC
	Green Genes Workshop, DNALC
	Fun with DNA Workshop, DNALC
	Green Genes Workshop, DNALC
	BioCoding Workshop, DNALC
	DNA Science Workshop, *DNALC NYC* at City Tech
	Urban Barcode Research Program Conservation Genetics Workshop, *DNALC NYC* at City Tech
	Bridging the Gap Fun with DNA Virtual Workshop, *Regeneron DNALC*
	Bridging the Gap Forensic Detectives Virtual Workshop, *Regeneron DNALC*
	DNA Barcoding Workshop, Francis Lewis High School, Fresh Meadows, New York
July 19–23	*DNA Science* Workshop, DNALC
	Forensic Detectives Workshop, DNALC
	World of Enzymes Workshop, DNALC
	Fun with DNA Workshop, DNALC
	Forensic Detectives Workshop, *DNALC NYC* at City Tech
	Fun with DNA Workshop, *DNALC NYC* at City Tech
	DNA Science Workshop, *DNALC NYC* at City Tech
	Urban Barcode Research Program Barcoding Workshop, *DNALC NYC* at City Tech
	Urban Barcode Research Program Conservation Genetics Workshop, *Harlem DNA Lab*
	Bridging the Gap Forensic Detectives Virtual Workshop, *Regeneron DNALC*
July 26–30	*DNA Science* Workshop, DNALC
	Green Genes Workshop, DNALC
	World of Enzymes Workshop, DNALC
	DNA Barcoding Workshop, DNALC
	Genome Science Workshop, *DNALC NYC* at City Tech
	World of Enzymes Workshop, *DNALC NYC* at City Tech
	Urban Barcode Research Program Conservation Genetics Workshop, *DNALC NYC* at City Tech
	Urban Barcode Research Program Barcoding Workshop, *Harlem DNA Lab*
	Genome Science Virtual Workshop, DNALC

	Bridging the Gap STEP World of Enzymes Virtual Workshop, DNALC & *Regeneron DNALC*
	BioCoding Virtual Workshop, DNALC
July 26	"Extending the Possibilities of Biology with Open Science," Virtual Student Lecture, Prairie View A & M University College of Engineering, Prairie View, Texas, DNALC
July 30	NIH *Citizen DNA Barcode Network* Collection Event, Sweet Briar Nature Center, Smithtown, New York
August 2–6	*Green Genes* Workshop, DNALC
	Genome Science Workshop, DNALC
	Fun with DNA Workshop, DNALC
	Citizen DNA Barcode Network Barcoding Workshop, DNALC
	DNA Barcoding Workshop, *DNALC NYC* at City Tech
	Green Genes Workshop, *DNALC NYC* at City Tech
	Urban Barcode Research Program Barcoding Workshop, *DNALC NYC* at City Tech
	Urban Barcode Research Program Conservation Genetics Workshop, *Harlem DNA Lab*
	Genome Science Virtual Workshop, DNALC
	Fun with DNA Virtual Workshop, *Regeneron DNALC*
	Forensic Detectives Workshop, Friends Academy, Locust Valley, New York
August 3–5	*DNA Metabarcoding* for CURES Virtual Workshop, James Madison University Harrisonburg, Virginia, DNALC
August 9–13	*DNA Science* Workshop, DNALC
	STARS DNA Barcoding Workshop, DNALC
	Forensic Detectives Workshop, DNALC
	World of Enzymes Workshop, DNALC
	DNA Science Workshop, *DNALC NYC* at City Tech
	Forensic Detectives Workshop, *DNALC NYC* at City Tech
	Urban Barcode Research Program Conservation Genetics Workshop, DNALC NYC at City Tech
	Urban Barcode Research Program Barcoding Workshop, *Harlem DNA Lab*
	World of Enzymes Virtual Workshop, DNALC
	Forensic Detectives Virtual Workshop, *Regeneron DNALC*
August 16–20	*DNA Science* Workshop, DNALC
	Green Genes Workshop, DNALC
	World of Enzymes Workshop, DNALC
	Fun with DNA Workshop, DNALC
	STARS BioCoding Workshop, DNALC
	Fun with DNA Workshop, *DNALC NYC* at City Tech
	Genome Science Workshop, *DNALC NYC* at City Tech
	Urban Barcode Research Program Barcoding Workshop, *DNALC NYC* at City Tech
	DNA Barcoding Virtual Workshop, DNALC
August 23–27	*Sequence a Genome!* Workshop, DNALC
	DNA Barcoding Workshop, DNALC
	Forensic Detectives Workshop, DNALC
	WiSE Fun with DNA Workshop, DNALC
	DNA Science Workshop, *DNALC NYC* at City Tech
	World of Enzymes Workshop, *DNALC NYC* at City Tech
	Urban Barcode Research Program Bootcamp Workshop, *DNALC NYC* at City Tech
	Fun with DNA Virtual Workshop, DNALC
	DNA Barcoding Virtual Workshop, DNALC
August 25	Site Visit by Eric Adams, Brooklyn Borough & Russ Hotzler, City Tech, *DNALC NYC* at City Tech
August 30–	*DNA Science* Workshop, DNALC
September 3	*Fun with DNA* Workshop, DNALC
	Green Genes Workshop, DNALC
	Green Genes Workshop, *DNALC NYC* at City Tech
September 2	Chan Zuckerberg Initiative Training and Education for Open Science Virtual Workshop, Redwood City, California, DNALC
September 13, 15	NSF Virtual Panel, Alexandria, Virginia
September 20	Site Visit by Greg Borman, Department of STEM, Office of Curriculum Instruction & Professional Learning, NYC Department of Education, & Sheldon Young, ACCESS Citywide, NYC Department of Education, New York, New York, *DNALC NYC at City Tech*
September 21, 23	NSF Virtual Panel, Alexandria, Virginia
September 24	*DNALC NYC* at City Tech Ribbon Cutting Event

September 29	American Society of Cell Biology Public Engagement Webinar, "Relevance and Rapport; Promoting Inclusion through Public Engagement," Virtual Lecture Rockville, Maryland
September 30	Site Visit by Joseph Oleniczak and Kissaou Tchedre, Austin Community College, Austin, Texas, DNALC & *DNALC NYC* at City Tech
October 1	"A Day in the Life" Event, Massapequa High School, Massapequa, New York
October 5	*Urban Barcode Research Program* Student Orientation, Virtual and *DNALC NYC* at City Tech
October 8	CyVerse Webinar Series Virtual Teacher Workshop, "CyVerse Visual Interactive Computing Environment," DNALC
October 16	STEM Teachers NYC Training "Intro to Biotech, Part 1," *DNALC NYC* at City Tech
October 19	Site Visit by Nick Greiner and Maria Scoutas, Consul General's Office of Australia, New York, New York, with Bruce Stillman, DNALC
October 20	Regeneron Virtual Training Session "Day for Doing Good," DNALC
October 23	STEM Teachers NYC Training "Intro to Biotech, Part 2," *DNALC NYC* at City Tech *Saturday DNA!* "Tracking Traits," DNALC
October 26	Day for Doing Good Science Virtual Expo, *"DNA Extraction from Wheat Germ,"* Regeneron Pharmaceuticals, Tarrytown, New York
October 27	Massapequa School District Teacher Training Workshops, "Barcoding and Bioinformatics," and "DNA Extraction, Mendelian Inheritance, Bacteria & Antibiotics," & "Otzi the Iceman," DNALC Fall Open House, *DNALC NYC* at City Tech
October 28	NIH *Citizen DNA Barcode Network* Meeting at Jones Beach Energy and Nature Center, Jones Beach State Park, Wantagh, New York
October 29	Site Visit by Doug Torre, CSHL, DNALC Massapequa School District Teacher Training Workshops, "Restriction Analysis," DNALC
October 30	STEM Teachers NYC Training "Intro to Biotech, Part 3," *DNALC NYC* at City Tech
November 2	NIH *Barcode Long Island* Teacher Training Workshop, "DNA Barcoding," DNALC Teacher Training Workshop, "Human Mitochondrial Sequencing," DNALC *Urban Barcode Project* Training, Virtual Refresher Course, *DNALC NYC* at City Tech
November 3–5	CSHL Genome Informatics Virtual Conference, "Tools to Facilitate Student Metabarcoding Research," Poster Session, CSHL
November 4	Black Women in Computational Biology Seminar, "Opening Doors—Preparing the Next Generation of Computational Biologists," Virtual Lecture, Philadelphia, Pennsylvania Fall Open House, *DNALC NYC* at City Tech
November 8	Fall Open House, *DNALC NYC* at City Tech
November 9	Math for America Teacher Training "Human Mitochondrial Sequencing Part 1," *DNALC NYC* at City Tech Brooklyn North Virtual STEM Day Celebration, Brooklyn, New York
November 10	Fall Open House, *Regeneron DNALC*
November 11	*Ötzi the Iceman* Virtual Tour, DNALC
November 13	*Saturday DNA!* "A BOLD Connection," DNALC
November 15	Long Beach School District Teacher Training Workshop, "Bacterial Transformation," DNALC *Barcode Long Island* Virtual Teacher Training, "Bioinformatics," DNALC
November 16	Math for America Teacher Training "Human Mitochondrial Sequencing Part 2," *DNALC NYC* at City Tech
November 18	Fall Open House, *Regeneron DNALC* "Meet a Scientist: Dr. Doreen Ware," Virtual Webinar, DNALC
November 19	"Learning New Tricks—Career-spanning Learning in STEM," Virtual Student Lecture, Engaged STEM Scholars Program, Barry University, Miami Shores, Florida
November 23	Math for America Teacher Training "Human Mitochondrial Sequencing Part 3," *DNALC NYC* at City Tech
November 30	Site Visit & Tour by Justin Vázquez-Poritz, City Tech School of Arts and Sciences, Brooklyn, New York, *DNALC NYC* at City Tech
December 2	"Asking the Wrong Questions About American Science Education," Virtual Lecture, American Society for Cell Biology, Rockville, Maryland American Society for Cell Biology, Cell Bio Virtual 2021 Teacher Workshop, "Integrating Bioinformatics into Your Courses," Rockville, Maryland
December 2–3	Computing for Teaching and Learning with Jupyter Workshop, George Washington University, Washington D.C.
December 3	Shelter Island High School Science Fair, Shelter Island, New York
December 7, 9	Banbury Life Science Virtual Professional Development Conference, "Making Career-spanning Learning in the Life Sciences Inclusive and Effective for All," CSHL
December 8	Queensborough College Business Industry Leadership Team Kickoff Virtual Teacher Workshop, "Bringing Data into the Classroom," Queens, New York

December 9	CUNY TV Filming, *DNALC NYC* at City Tech
December 10	Site Visit by Brooklyn North e-STEM Team, Brooklyn, New York, *DNALC NYC* at City Tech
December 11	*Saturday DNA!* "A Day in the Life of the Iceman," DNALC
December 14	*Barcode Long Island* Open Lab, DNALC
December 16	Site Visit by Jeremy Seto (CUNY NYCCT) & City Tech Nursing Faculty, Brooklyn, New York, *DNALC NYC* at City Tech
	"Meet a Scientist: Dr. Christopher Vakoc," Virtual Webinar, DNALC

COLD SPRING HARBOR
LABORATORY PRESS

PRESS PUBLICATIONS

Serials

Genes & Development, Vol. 35 (www.genesdev.org)

Genome Research, Vol. 31 (www.genome.org)

Learning & Memory, Vol. 28 (www.learnmem.org)

RNA, Vol. 27 (www.rnajournal.org)

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*)

Influenza: The Cutting Edge, edited by Gabriele Neumann and Yoshihiro Kawaoka

T-Cell Memory, edited by David Masopust and Rafi Ahmed

Auxin Signaling: From Synthesis to Systems Biology, Second Edition, edited by Dolf Weijers, Karin Ljung, Mark Estelle, and Ottoline Leyser

The Nucleus, Second Edition, edited by Ana Pombo, Martin W. Hetzer, and Tom Misteli

Other

Ahead of the Curve: Women Scientists at the MRC Laboratory of Molecular Biology, by Kathleen Weston

CSHL Annual Report 2019, Yearbook Edition

E-books

T-Cell Memory, edited by David Masopust and Rafi Ahmed

Ahead of the Curve: Women Scientists at the MRC Laboratory of Molecular Biology, by Kathleen Weston

Auxin Signaling: From Synthesis to Systems Biology, Second Edition, edited by Dolf Weijers, Karin Ljung, Mark Estelle, and Ottoline Leyser

The Nucleus, Second Edition, edited by Ana Pombo, Martin W. Hetzer, and Tom Misteli

Combining Human Genetics and Causal Inference to Understand Human Disease and Development, edited by George Davey Smith, Rebecca Richmond, and Jean-Baptiste Pingault

Lung Cancer: Disease Biology and Its Potential for Clinical Translation, edited by Christine M. Lovly, David P. Carbone, and John D. Minna

Regeneration, edited by Kenneth D. Poss and Donald T. Fox

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

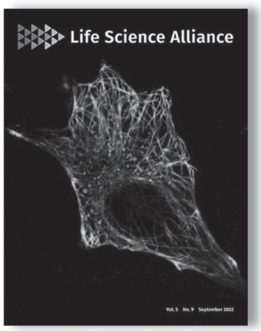
Cold Spring Harbor Laboratory Press provides scientists with resources to help advance their research and careers. These resources include books, manuals, and peer-reviewed journals.

The Press publishes nine journals and more than 270 books in print and electronic form. In 2021, its staff published more than 100 journal issues, five new print books, seven new e-books, and an audiobook, outputs that enabled the Press to make another substantial financial contribution to the Laboratory.

For more than a decade, the four established research journals, *Genes & Development*, *Genome Research*, *RNA*, and *Learning & Memory*, have combined subscription access for institutions with open access for individual papers paid for by authors or their funders. In 2020, prompted by publishing policies specified by science funders such as the Howard Hughes Medical Institute, a transition toward full open access began, aimed at the elimination of subscriptions for all four research journals by 2025. The transformational approach relies on an institutional offering known as "The Cold Spring Harbor Collection." Institutions that subscribe to one or any combination of our journals will be offered access to the entire Cold Spring Harbor Collection (all the research and review journals) in return for a single payment each year. In addition, the scientific staff at an institution with an agreement will be able to publish in the research journals without paying a per-article publication charge. The size of the institutional payment will depend on the current level of subscription expenditure at the institution concerned and the volume of articles expected to be published from that source, based on previous publication rates. In 2021, the number of institutional agreements rose to more than 50 and discussions with many more organizations took shape for completion in subsequent years.

For institutions not interested in access to our research journals, the review journals remain available on subscription, and subscriptions to all three increased during the year.

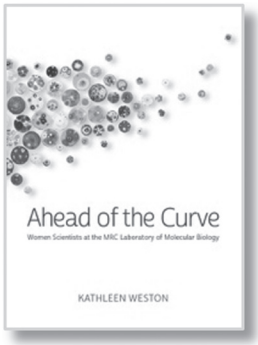
Anticipating this trend toward openness in research communication, our two newest journals, *Molecular Case Studies* and *Life Science Alliance*, were launched as open-access titles. *Molecular Case Studies* is a journal of precision medicine that enables sharing of insights from genomic and molecular analysis into the causes and potential treatment of cancer, neurological and metabolic disorders, and other diseases. In 2021, its readership continued to grow, although the pandemic affected new manuscript submissions because, as clinicians, a large percentage of this journal's potential authors were focused on patient concerns.



Life Science Alliance is owned and published jointly by Cold Spring Harbor Laboratory, the European Molecular Biology Organization (EMBO), and Rockefeller University. The journal's "cascade" model, publishing papers first declined by the partners' nine highly selective research journals, has become important in research communication in recent years. It benefits publishers, who gain revenue by retaining a greater proportion of submissions, and authors, whose papers find an appropriate vehicle for publication more efficiently. *Life Science Alliance* is a unique cross-publisher cascade and an alternative to the ecosystems of the dominant commercial publishers that compete for authors and journal content. Launched in April 2018, the journal's output has steadily increased, rising by 27% in 2021. It also received an encouraging first impact factor of 4.59.

Overall, online usage of Press journals remained strong, with a 6% increase to more than 21.1 million article downloads in the year.

COVID-19 hastened the shift to online delivery of books and e-books, and direct sales through the Press website accounted for 38% of all book sales during the year. The best sellers were the renowned reference works *Molecular Cloning* and *At the Helm*.



Other notable titles included *Ahead of the Curve: Women Scientists at the MRC Laboratory of Molecular Biology*, published as a paperback, an e-book, and as an audiobook narrated by the author, Kathleen Weston.

Faces of the Genome, edited by Ludmila Pollock, W. Richard McCombie, and Jan A. Witkowski, and *Conscience and Courage: How Visionary CEO Henri Termeer Built a Biotech Giant and Pioneered the Rare Disease Industry*, by John Hawkins, were published in Chinese. A Japanese edition of *The Digital Cell: Cell Biology as a Data Science*, by Stephen J. Royle, was released. Chinese and Russian translations of *Epigenetics*, Second Edition, edited by C. David Allis et al., are under way.

Staff

During the year, Linda Sussman retired after a 19-year career in which she held a variety of senior positions at the Press managing the technologies that support the journals and preprint servers. Sonali Bhattacharjee left the Press after a year in the Assistant Editor position of *Genes & Development*.

As a publishing organization, CSHL Press serves the research community by providing information of many types, in a variety of forms, that is reliable, accessible, and reasonably priced. In the second year of the COVID-19 pandemic, despite many personal and professional challenges, the Press staff continued to maintain the high quality of our books and journals and deliver them to the audience with efficiency and financial success. I thank all of them for their contributions and in particular 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, Linda Sussman, and Denise Weiss. And, as always, Mala Mazzullo was a warm, welcoming, and efficient presence at the heart of the organization.

John R. Inglis, Ph.D.
Executive Director and Publisher

PREPRINT SERVERS

PREPRINT SERVERS

A preprint is a research manuscript made freely available to the community by its authors without peer review. The Laboratory's preprint server (platform) for biology, bioRxiv, was founded eight years ago and the server for health sciences, medRxiv, co-managed with Yale University and the global health knowledge provider BMJ, is two years old. Both servers are free for readers and authors to use. Generous financial support is provided by the Laboratory and by grants from the Chan Zuckerberg Initiative.

bioRxiv and medRxiv are services that make possible the distribution of new research results two to four days after submission. Manuscripts are screened to ensure they conform to submission requirements, but there is no peer review as in the traditional publication model, so authors control the length, format, copyright, and content of their manuscript and the ability to update it. Once posted, the preprint can be shared to prompt feedback, cited in grant and job applications, and submitted to journals for publication.

In 2021, bioRxiv and medRxiv together posted 49,600 manuscripts. Managing this volume at speed required the exceptional dedication and expertise of a preprint team consisting of founders, content specialists, freelance manuscript screeners, clinical advisors, product specialists, developers, and user support specialists.

By year's end, bioRxiv contained 144,000 manuscripts, the work of 583,518 authors with 45,000 institutional addresses in 151 countries and territories. The United States, United Kingdom, Germany, China, and France accounted for most submissions. Stanford University, University of Oxford, University of Cambridge, University of Pennsylvania, and Harvard University were the most prolific institutions. The largest subject categories were neuroscience (17%), microbiology (10%), and bioinformatics (8%).

The entire repository of bioRxiv manuscripts had 87 million abstract page views and 38 million article downloads during 2021.

Thirty-six journals now give authors the opportunity to post a submitted manuscript simultaneously on bioRxiv. And 236 journals offer authors the opportunity to submit their preprints for editorial consideration directly from the platform. They are published by the American Chemical Society, European Molecular Biology Organization, Springer Nature, and many other prominent organizations. At least 68% of manuscripts posted to bioRxiv are published in a journal within two years, and nearly 81,000 papers first posted to bioRxiv have been published in more than 4,000 journals.

bioRxiv's momentum has aroused intense interest in biomedical preprints and, in turn, ways in which the work reported in preprints can be evaluated and the conclusions shared. This happens through on-site comments and Twitter but—increasingly through new, specific assessment projects initiated by organizations such as Review Commons, the *eLife* journal, and preLights, whose outputs are posted to the appropriate preprints. To assist readers in discovering these community evaluations, each bioRxiv preprint now has a dashboard that, through links and text, captures the diversity of reactions. It is launched from the preprint's abstract page.

In 2021, medRxiv posted 13,000 manuscripts originating in 156 countries and territories. The most prolific countries were the United States, United Kingdom, China, India, and Germany. The manuscripts had more than 92,000 authors. The University of Oxford, University College London, Imperial College London, and Stanford University were the most prolific sources. The largest subject categories are Infectious Diseases (25%), followed by Epidemiology (19%), and Public and Global Health (9%), which make up more than half of the total submissions.

medRxiv content had 71 million page views and 24 million article downloads in 2021.

The preprint team meeting on a weekly Zoom call.

medRxiv was 6 months old when the COVID-19 pandemic began in January 2020. A public health emergency on this scale required an exceptional response from the scientific and medical communities worldwide, and medRxiv was ideally positioned to provide the most rapid means of sharing new information as soon as it emerged. The deluge of pandemic-related medRxiv preprints in 2020 diminished in 2021 but, nevertheless, nearly 7,000 such preprints were posted during the year. According to National Library of Medicine data, medRxiv is the largest source of pandemic-related preprints.

John Inglis presented at The 4th Forum for World STM Journals, organized by CAST (China Association for Science and Technology), July 2021.

Usage of medRxiv remained high, reaching 5–9 million abstract page views per month. medRxiv preprints were covered in 1,500 news stories during the year and mentioned more than 75,000 times on Twitter.

One hundred and twenty-six journals offer authors the opportunity to submit their preprints for editorial consideration directly from medRxiv. These journals include those from the American Society of Clinical Oncology, BMJ, the JAMA Network, Oxford University Press, the Public Library of Science, and others. Approximately 10,500 papers first posted to medRxiv have been published in more than 2,300 journals.

In weekly medRxiv calls, the management and content teams from Yale, BMJ, and CSHL discuss policy and issues around submitted manuscripts. The CSHL Preprint Team also meet weekly to discuss content and production issues to facilitate smooth operations of both servers. Olaya Fernández Gayol was a welcome addition to the group of content specialists in September 2021.

During the year, co-founders John Inglis and Richard Sever and Content Team members Samantha Hindle, Sol Fereres Rapoport, and Sanchari Ghosh took part in virtual national and international conference talks, press interviews, panel presentations, webinars, and podcasts discussing the rapid adoption of preprints in biology and medicine and its implications.

bioRxiv and medRxiv are communication services that operate every day throughout the year, providing the international research community with the opportunity of sharing their work whenever they are ready to do so. This is made possible by the skill and diligence of the entire preprint team. Collectively they are empowering bioRxiv and medRxiv to transform the communication of biological and medical science, showing how it can be made faster, more equitable, and more interactive—"communication at the speed of science." The servers' continued progress adds to the Laboratory's already considerable reputation for innovation in the sharing of science worldwide.

John R. Inglis, Ph.D.
Co-Founder and Principal Investigator,
bioRxiv and medRxiv

FINANCE

FINANCIAL STATEMENTS

CONSOLIDATED BALANCE SHEETS
December 31, 2021
(with comparative financial information as of December 31, 2020)

	2021	2020
Assets:		
Cash and cash equivalents	$ 110,039,037	108,990,728
Grants receivable	11,023,970	8,898,818
Contributions receivable, net	109,223,435	112,118,879
Investments	785,156,520	730,782,048
Investment in employee residences	6,366,321	6,420,336
Restricted use assets	3,391,191	3,709,586
Other assets	7,585,315	6,831,860
Land, buildings, and equipment, net	286,004,083	266,300,564
Total assets	$ 1,318,789,872	1,244,052,819
Liabilities and net assets:		
Liabilities:		
Accounts payable and accrued expenses	$ 11,290,659	17,736,905
Deferred revenue	42,817,598	54,961,017
Interest rate swap	33,914,521	41,196,290
Bonds payable	96,006,502	95,940,233
Total liabilities	184,029,280	209,834,445
Commitments and contingencies		
Net assets:		
Without donor restrictions	695,295,250	598,160,188
With donor restrictions	439,465,342	436,058,186
Total net assets	1,134,760,592	1,034,218,374
Total liabilities and net assets	$ 1,318,789,872	1,244,052,819

CONSOLIDATED STATEMENT OF ACTIVITIES
Year ended December 31, 2021
(with summarized financial information for the year ended December 31, 2020)

	Without Donor Restrictions	With Donor Restrictions	2021 Total	2020 Total
Revenue and other support:				
Public support—contributions and nonfederal grant awards	$ 40,759,409	30,186,155	70,945,564	47,238,038
Federal grant awards	38,821,308	–	38,821,308	41,310,029
Indirect cost allowances	30,972,451	–	30,972,451	32,496,274
Investment return utilized	32,860,963	–	32,860,963	35,260,137
Royalty and license revenue	14,066,754	–	14,066,754	14,063,144
Program fees	6,177,475	–	6,177,475	6,027,763
Publications sales	9,193,214	–	9,193,214	9,665,668
Dining services	416,061	–	416,061	301,592
Rooms and apartments	1,036,749	–	1,036,749	1,162,678
Miscellaneous	996,292	–	996,292	945,231
Net assets released from restrictions	58,910,562	(58,910,562)	–	–
Total revenue and other support	234,211,238	(28,724,407)	205,486,831	188,470,554
Expenses:				
Research	110,705,498	–	110,705,498	107,061,748
Educational programs	13,781,654	–	13,781,654	13,907,463
Publications	9,555,000	–	9,555,000	9,611,376
Banbury Center conferences	1,327,685	–	1,327,685	1,418,130
DNA Learning Center programs	5,414,845	–	5,414,845	4,467,315
School of Biological Sciences programs	2,966,837	–	2,966,837	2,973,914
General and administrative	24,888,774	–	24,888,774	25,309,666
Total expenses	168,640,293	–	168,640,293	164,749,612
Excess (deficiency) of revenue and other support over expenses	65,570,945	(28,724,407)	36,846,538	23,720,942
Other changes in net assets:				
Investment return excluding amount utilized	24,282,348	32,131,563	56,413,911	45,901,739
Change in fair value of interest rate swap	7,281,769	–	7,281,769	(8,128,837)
Increase in net assets	97,135,062	3,407,156	100,542,218	61,493,844
Net assets at beginning of year	598,160,188	436,058,186	1,034,218,374	972,724,530
Net assets at end of year	$ 695,295,250	439,465,342	1,134,760,592	1,034,218,374

CONSOLIDATED STATEMENT OF CASH FLOWS
Year ended December 31, 2021
(with comparative financial information for the year ended December 31, 2020)

	2021	2020
Cash flows from operating activities:		
Increase in net assets	$ 100,542,218	61,493,844
Adjustments to reconcile increase in net assets to net cash provided by operating activities:		
Change in fair value of interest rate swap	(7,281,769)	8,128,837
Depreciation and amortization	16,065,677	14,740,266
Amortization of deferred bond costs	66,268	66,268
Net appreciation in fair value of investments	(73,881,062)	(75,871,001)
Contributions restricted for long-term investment	(3,810,763)	(3,911,559)
Changes in assets and liabilities:		
Grants receivable	(2,125,152)	(298,771)
Contributions receivable, net	13,930,309	20,319,211
Restricted use assets	318,395	(1,274,339)
Other assets	(753,454)	(971,593)
Accounts payable and accrued expenses	(6,446,246)	5,129,727
Deferred revenue	(12,143,419)	(13,170,825)
Net cash provided by operating activities	24,481,002	14,380,065
Cash flows from investment activities:		
Capital expenditures	(35,769,196)	(25,994,841)
Proceeds from sales and maturities of investments	123,699,817	256,049,061
Purchases of investments	(104,193,227)	(236,146,835)
Net change in investment in employee residences	54,015	546,411
Net cash used in investment activities	(16,208,591)	(5,546,204)
Cash flows from financing activities:		
Contributions restricted for long-term investment	123,296	154,116
Contributions restricted for investment in capital	3,687,467	3,757,443
(Increase) decrease in contributions receivable	(11,034,865)	3,632,367
Net cash (used in) provided by financing activities	(7,224,102)	7,543,926
Net increase in cash and cash equivalents	1,048,309	16,377,787
Cash and cash equivalents at beginning of year	108,990,728	92,612,941
Cash and cash equivalents at end of year	$ 110,039,037	108,990,728
Supplemental disclosure:		
Interest paid	$ 3,773,160	3,860,777
Purchases of capital expenditures in accounts payable	$ 1,167,068	1,974,654
Lease liability of right of use asset	$ 647,449	944,193

FINANCIAL SUPPORT OF THE LABORATORY

Cold Spring Harbor Laboratory, the DNA Learning Center, and the Banbury 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 2021.

GRANTS January 1–December 31, 2021

COLD SPRING HARBOR LABORATORY GRANTS

Grantor	*Program/Principal Investigator*	*Project Start Date*	*Project End Date*	*2021 Funding*[1]	
FEDERAL GRANTS					
NATIONAL INSTITUTES OF HEALTH					
Program Project and Center Support	Drs. Stillman/Egeblad/Krainer/ McCombie/Pappin/Spector/Vakoc	02/20/18	01/31/23	$ 4,484,444	
	Dr. Tuveson—Cancer Center Core	08/01/21	07/31/26	4,495,455	
Cooperative Research Agreement Support[2]	Drs. Huang/Gillis/Mitra/Osten	09/21/17	05/31/23	2,716,436	
	Drs. Osten/Albeanu/Mitra	09/20/17	05/31/22	1,733,826	
	Dr. Tuveson	03/06/18	02/28/23	425,196	
	Dr. Vakoc	08/01/19	07/31/24	439,200	
Research Support	Drs. Albeanu/Koulakov	09/11/18	08/31/23	1,089,335	
	Drs. Albeanu/Koulakov	06/15/19	03/31/24	904,662	
	Dr. Cheadle	03/14/19	02/28/23	243,850	
	Drs. Dobin/Gingeras	08/18/17	05/31/23	480,000	
	Dr. dos Santos	03/01/20	02/28/25	439,200	
	Drs. dos Santos/Siepel	09/30/20	05/31/25	357,683	
	Dr. Egeblad	03/01/20	02/28/25	530,642	
	Dr. Engel	09/01/21	08/31/24	488,137	*
	Dr. Furukawa	04/15/19	03/31/24	575,545	
	Dr. Furukawa	08/01/19	03/31/24	535,557	
	Drs. Gillis/Huang/Lee	07/13/17	05/31/22	480,000	
	Dr. Goodwin	09/11/19	08/31/24	122,709	
	Dr. M. Hammell	08/15/20	07/31/24	928,986	
	Dr. Kinney	09/01/19	08/31/24	480,000	
	Drs. Koulakov/Li	09/30/19	07/31/24	432,000	
	Dr. Krainer	07/01/17	06/30/22	806,400	
	Dr. Li	03/01/14	12/31/24	638,663	
	Dr. McCandlish	09/01/19	07/31/24	480,000	
	Dr. Mitra	09/14/20	09/13/23	672,312	
	Dr. Mitra	09/15/21	08/31/24	5,278,129	
	Dr. Moffitt	07/01/20	06/30/22	124,738	
	Dr. Park	09/01/21	08/31/26	248,045	*
	Dr. Pedmale	08/04/17	07/31/22	480,000	
	Dr. Schorn	08/17/20	06/30/25	403,200	
	Dr. Shea	06/01/19	03/31/24	480,000	
	Dr. Sheltzer	05/01/20	04/30/22	269,280	

[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 2021

Grantor	*Program/Principal Investigator*	*Project Start Date*	*Project End Date*	*2021 Funding*[1]	
	Dr. Siepel	03/01/18	02/28/23	479,215	
	Dr. D. Spector	08/20/19	07/31/24	768,000	
	Drs. Tollkuhn/Gillis	03/01/18	12/31/22	480,000	
	Dr. Tonks	05/01/21	06/30/26	768,908	*
	Dr. Tonks	03/17/20	01/31/24	480,000	
	Drs. Tuveson/Dobin/Preall	08/10/21	07/31/26	818,876	*
	Dr. Vakoc	12/18/18	11/30/23	496,969	
	Drs. Vakoc/Tuveson	07/02/19	06/30/24	413,053	
	Dr. Van Aelst	04/01/19	01/31/24	656,141	
	Dr. Van Aelst	04/01/20	01/31/25	586,028	
	Dr. Zador	08/01/20	07/31/23	1,101,008	
Active Contracts Continuing without Additional Support					
Leidos Biomedical Research, Inc.–NCI	Drs. Tuveson/D. Spector	01/09/17	09/30/21		
Active Research Awards Continuing without Additional Support	Dr. Joshua-Tor	06/10/16	03/31/21		
	Drs. Li/Huang	09/28/15	06/30/21		
	Dr. Tonks	05/14/15	04/30/21		
	Dr. Stillman	03/01/17	02/28/22		
	Drs. Tuveson/M. Hammell/Pappin	12/07/16	11/30/22		
	Dr. Vakoc	12/02/19	11/30/22		
Research Subcontracts					
NIH/Binghamton University Consortium Agreement	Dr. Navlakha	05/15/19	04/30/23	331,966	
	Dr. Zador	12/01/21	11/30/23	69,886	*
NIH/Columbia University Consortium Agreement	Dr. Zador	08/15/21	07/31/26	648,503	*
NIH/Cornell University Consortium Agreement	Dr. Siepel	05/01/19	02/28/23	313,407	
NIH/Duke University Consortium Agreement	Drs. Osten/Zador	09/25/20	09/22/23	802,207	*
NIH/Duke University Consortium Agreement	Drs. Gillis/Koo	05/01/21	04/30/24	252,180	
NIH/Harvard Medical School Consortium Agreement	Dr. Osten	07/01/17	10/31/21	42,000	
NIH/Memorial Sloan Kettering Cancer Center Consortium Agreement	Dr. Egeblad	03/15/20	12/31/24	96,000	
NIH/MIT Consortium Agreement	Dr. Mitra	09/01/21	06/30/24	422,983	*
NIH/New York University Consortium Agreement	Dr. Koulakov	09/01/19	05/31/24	431,374	
NIH/Oregon Health & Science University Consortium Agreement	Dr. Li	12/15/17	11/30/22	371,075	
NIH/The Salk Institute for Biological Studies Consortium Agreement	Dr. Zador	07/15/19	03/31/24	153,600	
NIH/The Research Foundation for the State of New York–Stony Brook Consortium Agreement	Dr. M. Hammell	09/15/17	03/31/22	37,837	
NIH/The Research Foundation for the State of New York–Stony Brook Consortium Agreement	Dr. Trotman	02/01/20	01/31/25	240,000	
NIH/University of California–San Diego Consortium Agreement	Dr. Gillis	09/18/19	08/31/22	188,060	

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2021

Grantor	*Program/Principal Investigator*	*Project Start Date*	*Project End Date*	*2021 Funding*[1]	
NIH/University of California–San Diego Consortium Agreement	Dr. Mitra	09/15/18	05/31/23	272,549	
NIH/University of Minnesota Consortium Agreement	Dr. dos Santos	12/01/19	11/30/24	108,236	
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–Austin Consortium Agreement	Dr. Osten	09/18/17	10/31/21	148,005	
Fellowship/Career Development Support	S. Balinth	07/16/20	07/15/23	30,036	
	J. Gegenhuber	09/01/20	08/31/23	30,036	
	Dr. Michalski	07/01/20	06/30/23	66,390	
	L. Tellez Perez	08/16/21	08/15/22	35,000	*
Institutional Training Program Support	Dr. Gann/CSHL School of Biological Sciences	09/01/21	08/31/22	289,234	
Course Support	Advanced Sequencing Technologies and Applications	04/10/12	06/30/26	80,019	
	Advanced Techniques in Molecular Neuroscience	04/15/20	02/28/25	128,639	
	Cell and Developmental Biology of *Xenopus*	06/01/20	05/31/25	106,959	
	Cellular Biology of Addiction	08/01/16	07/31/21	40,260	
	Computational Genomics	09/08/20	06/30/25	72,821	*
	Imaging Structure and Function in the Nervous System	04/15/20	02/28/25	158,630	
	Platform for Linking Circuits, Perception, and Behavior	04/01/19	03/31/24	29,000	
	Programming for Biology	09/01/20	07/31/25	96,492	
	Statistical Methods for Functional Genomics Course	09/07/21	06/30/25	92,149	
	X-Ray Methods in Structural Biology	09/01/17	08/31/23	71,720	
Meeting Support	Eukaryotic mRNA Processing	04/01/21	03/31/22	4,000	*
	Extramural Research Programs in the Neurological Disorders	02/01/21	01/31/22	5,000	*
	Extramural Research Programs in the Neurological Disorders	12/01/21	11/30/22	18,000	*
	Gene Expression and Signaling in the Immune System	12/01/21	11/30/22	16,000	*
	Genome Informatics	09/07/21	08/31/22	33,960	*
	Global Regulation of Gene Expression	09/10/19	07/31/24	29,774	*
	Mechanisms of Metabolic Signaling	06/11/21	03/31/22	15,000	*
	Microbial Pathogenesis and Host Response	07/01/21	06/30/22	16,000	*
	Microenvironment and Metastasis	07/05/21	06/30/22	12,503	*
	Network Biology Conference	02/01/21	01/31/26	32,465	*
	Neurobiology of *Drosophila* Conference	09/15/21	08/31/22	10,000	*

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2021

Grantor	*Program/Principal Investigator*	*Project Start Date*	*Project End Date*	*2021 Funding*[1]	
	Retroviruses	04/01/21	03/31/25	35,000	
	Systems Immunology	04/09/21	03/31/22	25,000	*
	Telomeres and Telomerase	04/01/21	03/31/22	29,935	*
	The Biology of Genomes	04/01/18	03/31/23	60,000	
NATIONAL SCIENCE FOUNDATION					
Multiple Project Award Support	Drs. McCombie/Martienssen	09/01/18	08/31/22	841,688	
Research Support	Dr. Jackson	08/01/19	07/31/22	241,603	
	Drs. Jackson/Lippman	09/01/21	08/31/25	956,695	*
	Dr. Navlakha	11/01/19	03/31/24	209,285	
	Dr. Ware	10/01/21	09/30/25	12,276	
Active Research Awards Continuing without Additional Support	Dr. Pedmale	09/01/18	08/31/22		
Research Subcontracts					
NSF/Cornell University Consortium Agreement	Dr. Siepel	09/01/18	08/31/22	42,499	*
NSF/New Jersey City University Consortium Agreement	Dr. Gingeras	06/01/21	05/31/24	53,760	
NSF/New York University Consortium Agreement	Drs. Jackson/Gillis	11/01/19	10/31/23	589,924	
Fellowship Support	L. Aguirre	06/01/19	05/31/22	46,000	
Institutional Training Program Support	Drs. Gillis/C. Hammell/Research Experiences for Undergraduates Program	04/01/20	03/31/23	133,840	
Course Support	Advanced Bacterial Genetics	06/15/17	05/31/22	90,000	
	Drosophila Neurobiology: Genes, Circuits and Behavior	05/01/20	04/30/24	101,869	
	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	118,880	
	Dr. Jackson	06/01/20	05/31/23	152,618	
	Dr. McCombie	09/10/19	09/09/22	1,356,006	
Active Research Awards Continuing without Additional Support	Dr. McCombie	09/15/17	09/14/22		
UNITED STATES DEPARTMENT OF THE ARMY					
Research Support	Dr. Egeblad	08/01/20	07/31/23	535,972	
	Dr. Lukey	01/15/21	01/14/24	269,441	*
	Dr. Sheltzer	08/15/20	08/14/21	287,927	
	Dr. Van Aelst	08/01/20	07/31/23	423,905	

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2021

Grantor	*Program/Principal Investigator*	*Project Start Date*	*Project End Date*	*2021 Funding*[1]	
UNITED STATES DEPARTMENT OF ENERGY					
Research Support					
Active Research Awards Continuing without Additional Support	Dr. Martienssen	09/15/17	09/14/22		
Active Research Subcontracts Continuing without Additional Support					
DOE/Brookhaven National Laboratory Consortium Agreement	Dr. Ware	09/01/19	05/31/22		
NONFEDERAL GRANTS					
MISCELLANEOUS SOURCES OF FUNDING					
Equipment Support					
F. M. Kirby Foundation, Inc.	Dr. Moses	06/30/20	05/31/22	115,000	
Pierre and Paula Gonthier Family Foundation	Dr. McCombie	06/01/21	05/31/22	66,000	*
The Sunshine Foundation	Dr. Moses	08/10/21	08/09/22	50,000	
Program Project Support					
The Simons Foundation/Autism	Dr. Wigler	01/01/21	12/31/24	2,095,952	
The Simons Foundation/Cancer	Dr. Wigler	01/01/21	12/31/24	3,127,278	
Research Support					
Allergan Sales, LLC	Dr. Furukawa	01/01/19	12/31/21	189,560	
American Association for Cancer Research	Dr. Borniger	12/01/20	11/30/23	150,038	
American Cancer Society	Dr. Sheltzer	01/01/20	08/31/21	142,893	
	Dr. Sheltzer	02/01/21	08/31/21	29,368	*
Anonymous	Dr. Tuveson	11/01/18	10/31/23	675,000	
Bessemer Trust Company	Dr. Janowitz	11/15/21	11/14/22	20,000	*
The Breast Cancer Research Foundation	Dr. Wigler	10/01/21	09/30/22	190,000	
The Joe W. and Dorothy Dorsett Brown Foundation	Dr. Sheltzer	01/01/21	08/31/21	100,000	*
Brown University	Dr. Tuveson	01/01/20	12/31/22	86,250	
Cedar Hill Foundation	Dr. Fearon	11/16/21	11/15/22	60,000	*
CSHL Translational Cancer Support	Dr. Beyaz	01/31/19	01/30/23	216,562	
	Dr. Beyaz	03/22/21	03/21/23	18,255	*
	Dr. Beyaz	03/22/21	03/21/23	51,654	*
	Dr. Beyaz	05/18/21	05/17/23	57,600	*
	Dr. Beyaz	11/01/21	10/31/22	160,134	*
	Dr. dos Santos	05/26/17	05/31/22	192,515	
	Dr. Gillis	03/01/18	02/28/23	238,314	
	Dr. Janowitz	08/22/18	08/21/23	237,168	
	Dr. Krasnitz	05/17/19	05/16/24	80,033	
	Dr. Krasnitz	11/15/21	11/14/23	66,264	*
	Dr. Krasnitz	03/01/18	02/28/23	258,460	
	Dr. Levy	05/17/19	05/16/24	63,460	
	Dr. Levy	03/01/18	02/28/23	21,238	
	Dr. Li	03/01/18	02/28/22	130,154	

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2021

Grantor	*Program/Principal Investigator*	*Project Start Date*	*Project End Date*	*2021 Funding*[1]	
	Dr. Li	03/01/18	02/28/22	128,508	
	Dr. Lyons	10/01/21	09/30/23	234,601	
	Dr. McCombie	11/15/21	11/14/23	300,405	*
	Dr. Meyer	05/01/21	04/30/24	220,838	*
	Dr. Meyer	05/01/21	04/30/24	27,763	*
	Dr. Mills	07/01/20	06/30/22	166,439	
	Dr. Mills	01/31/19	01/30/23	398,095	
	Dr. Mills	04/01/21	03/31/23	571,482	*
	Dr. Moses	04/01/21	03/31/23	178,796	*
	Dr. Nadella	08/10/20	08/19/23	129,285	
	Dr. Osten	07/21/17	01/10/22	162,641	
	Dr. Trotman	01/31/19	01/30/23	621,080	
	Dr. Trotman	08/26/19	08/25/22	523,520	
	Dr. Tuveson	02/23/21	02/22/23	132,065	*
	Dr. Tuveson	10/01/21	09/30/23	125,952	*
	Dr. Vakoc	05/17/19	05/16/22	711,420	
	Dr. Wigler	05/17/19	05/16/24	1,110,807	
	Dr. Wigler	05/17/19	05/16/24	30,056	
	Dr. Wigler	03/01/18	02/28/23	502,471	
	Dr. Yeh	06/16/20	06/15/23	535,496	
Duke University	Dr. Huang	08/17/20	12/31/21	196,108	
The Oliver S. and Jennie R. Donaldson Charitable Trust	Dr. Beyaz	12/07/21	12/06/24	294,500	*
Dr. Lee MacCormick Edwards Charitable Foundation	Dr. Janowitz	08/24/21	08/23/22	15,000	*
Dr. Lee MacCormick Edwards Charitable Foundation	Dr. Tollkuhn	12/28/21	12/27/22	15,000	*
Fortune Footwear	Dr. Janowitz	09/08/21	09/07/22	10,000	*
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	
Friends of TJ Foundation Inc	Dr. Vakoc	09/16/18	09/15/22	50,000	
Forbeck Foundation	Dr. Sheltzer	03/01/21	08/31/21	19,686	*
Bernard F. & Alva B. Gimbel Foundation	Dr. Krainer	02/01/20	01/31/23	5,000	*
Irving Hansen Foundation	Dr. Tonks	08/01/21	07/31/22	25,000	
The G. Harold and Leila Y. Mathers Foundation	Dr. Beyaz	06/01/21	05/31/24	251,986	*
Jo-Ellen and Ira Hazan	Dr. Tuveson	01/01/21	12/31/22	15,000	
Heartfelt Wings Foundation Inc.	Dr. Furukawa	09/05/18	09/04/22	250,000	
Lund University	Dr. M. Hammell	10/01/20	09/30/23	180,020	
The Lustgarten Foundation	Drs. Fearon/Lyons/Yeh	07/01/21	06/30/24	400,000	*
	Dr. Tuveson	09/01/17	12/31/22	1,000,000	
	Dr. Tuveson	07/01/21	06/30/22	70,000	
	Drs. Boyd/Tuveson	01/01/21	12/31/22	481,716	*
The Mark Foundation for Cancer Research Ltd.	Drs. Janowitz/Beyaz/Osten/Gillis/ Meyer	10/15/20	10/14/23	820,676	
The G. Harold and Leila Y. Mathers Charitable Foundation	Dr. Meyer	12/01/19	11/30/22	233,336	
Breast Cancer Awareness Day in Memory of Elizabeth McFarland	Dr. Wigler	01/01/21	12/31/22	48,795	
Mestag Therapeutics Limited	Drs. Chang/Dobin/Tuveson	07/01/20	03/31/22	378,023	
Maddie's Promise	Dr. Vakoc	09/16/14	09/15/23	90,000	*
The Marion I. & Henry J. Knott Foundation	Dr. Cheadle	12/07/21	12/06/22	100,000	*

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2021

Grantor	*Program/Principal Investigator*	*Project Start Date*	*Project End Date*	*2021 Funding*[1]	
The G. Harold and Leila Y. Mathers Foundation	Drs. Koulakov/Zador	10/01/21	09/30/24	225,423	*
Indian Institute of Technology–Madras	Dr. Mitra	01/01/15	06/30/23	28,994	*
Grace and Thor Larsen	Dr. Van Aelst	04/29/21	04/30/23	658,296	*
The McKnight Endowment Fund for Neuroscience	Dr. Cheadle	07/01/21	06/30/24	75,000	*
Mestag Therapeutics Ltd	Dr. Fearon	02/01/21	09/30/21	195,038	*
	Dr. Fearon	07/01/21	06/30/23	363,662	*
The Don Monti Memorial Research Foundation	Drs. Stillman/Tonks/Vakoc	01/01/21	12/31/22	100,000	
The Michelle Paternoster Foundation	Dr. Vakoc	09/16/14	09/15/23	25,000	
New York Genome Center	Dr. Beyaz	03/22/21	03/21/23	85,562	*
	Dr. Beyaz	05/18/21	05/17/23	70,000	*
	Drs. Krasnitz/Tuveson	02/23/21	02/22/23	140,431	*
New York State Department of Health	Dr. Trotman	06/01/21	05/31/23	173,333	*
Northwell Health, Inc.	Dr. Mills	08/13/21	08/12/23	50,000	*
Ono Pharmaceutical	Dr. Sheltzer	01/01/21	08/31/21	204,145	*
Robert J and Claire Pasarow Foundation	Dr. Fearon	10/25/21	10/24/22	50,000	*
The Pew Charitable Trusts	Dr. Navlakha	02/19/20	07/31/23	75,000	
Christina Renna Foundation Inc.	Dr. Vakoc	09/16/14	09/15/23	35,000	
Rita Allen Foundation	Dr. Cheadle	09/01/21	08/31/26	100,000	*
William J Riley Memorial Fund	Dr. Vakoc	09/16/14	09/15/23	30,000	
Charles and Marie Robertson Foundation	Dr. dos Santos	01/01/21	12/31/22	30,000	
	Dr. Shea	01/01/21	12/31/21	20,000	
The Mary Ruchalski Foundation, Inc.	Dr. Vakoc	09/16/14	09/15/23	85,000	
Sarah Edwards	Dr. Moses	08/01/21	07/31/22	6,066	*
The David and Lyn Silfen Foundation	Dr. Vakoc	10/12/21	10/11/22	5,000	*
Susan M. Schultz	Dr. Tuveson	08/25/21	08/24/22	25,000	*
The Eleanor Schwartz Charitable Foundation	Dr. Zador	06/18/21	06/17/22	150,000	*
The Simons Foundation	Dr. Engel	10/01/20	09/30/23	103,200	
	Dr. Zador	07/01/17	06/30/22	140,400	
	Dr. Iossifov	12/01/20	03/31/22	260,490	
The Simons Foundation/CSHL Innovative Center	Dr. Fearon	07/01/17	04/30/22	408,852	
	Dr. Vakoc	07/01/17	04/30/22	124,625	
Starr Cancer Consortium	Dr. Beyaz	01/01/20	06/30/22	202,111	
	Dr. Vakoc	01/01/20	12/31/22	213,500	
	Dr. Vakoc	01/01/21	12/31/22	96,000	*
Starr Companies	Dr. Moses	09/01/21	08/31/22	6,000	*
Swim Across America Nassau/Suffolk	Dr. Egeblad	12/21/21	12/20/22	50,000	
John Templeton Foundation	Dr. McCandlish	12/01/20	11/30/22	97,027	
The Toronto-Dominion Bank	Dr. Tuveson	11/19/19	12/18/22	157,400	
The University of Texas M.D. Anderson Cancer Center/Chan Zuckerberg Foundation	Dr. Lee	07/01/19	03/31/22	161,000	
Treeline Biosciences, Inc.	Dr. Vakoc	06/08/21	05/31/26	2,499,720	*
V Foundation	Dr. Krainer	11/01/21	11/01/24	200,000	*
Washington University in St. Louis	Dr. Kepecs	01/01/21	12/31/21	320,203	*
The Wasily Family Foundation, Inc.	Dr. Lukey	07/12/21	07/11/22	50,000	*
Joan & Sanford I. Weill Medical College	Dr. Fearon	01/01/21	06/30/22	187,205	
Women's Partnership in Science	Drs. Egeblad/M. Hammell/Joshua-Tor/Mills	01/01/21	12/31/21	202,602	

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2021

Grantor	*Program/Principal Investigator*	*Project Start Date*	*Project End Date*	*2021 Funding*[1]	
The Bradley Zankel Foundation, Inc.	Dr. Mills	11/15/21	11/14/22	20,000	
Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation	Dr. Beyaz	12/01/21	11/30/24	346,095	*
	Dr. M. Hammell	12/01/18	11/30/23	500,000	
Yale University	Dr. Sheltzer	10/25/21	10/31/22	159,230	*
Fellowship Support					
American Association for Cancer Research	X. He	07/01/21	06/30/23	60,000	*
American-Italian Cancer Foundation	G. Caligiuri	08/01/21	07/31/22	40,000	*
Brain & Behavior Research Foundation	Dr. Zhang	01/15/20	01/15/22	35,000	
	Dr. Dvorkin	01/15/20	01/14/22	35,000	
	Dr. Borniger	01/15/20	01/14/22	35,000	
	D. Fuerth	01/15/21	02/28/21	4,453	*
Cancer Research Institute	Dr. Adrover Montemayor	07/01/20	06/30/23	43,875	
Lola A. Goldring	Dr. Stillman	10/01/21	09/30/22	100,000	
Esther A. & Joseph Klingenstein Fund	Dr. Cheadle	07/01/21	06/30/24	75,000	*
Human Frontier Science Program	C. Soitu	07/01/21	06/30/24	66,660	*
Howard Hughes Medical Institute	D. Adams	01/01/21	08/30/22	59,677	
	B. Bibel	01/01/21	08/30/22	35,583	
	Dr. Mills	09/01/19	08/31/22	4,000	*
	Dr. Mills/D. Johnson	09/01/19	08/31/22	46,000	
	K. Meze	01/01/21	08/30/21	29,191	
Mary's Pence	D. Rupert	10/01/21	09/30/22	5,000	
The Meier and Linnartz Family Foundation	Dr. Cheadle	06/01/21	05/31/22	50,000	*
The New York Community Trust	J. Homer	07/01/21	06/30/23	75,000	*
John and Amy Phelan Foundation	CSHL School of Biological Sciences	09/01/17	12/31/21	100,000	
The Research Foundation for State University of New York–Stony Brook	J. Thalappillil	02/01/20	02/15/23	4,200	
	M. Shevik	07/01/20	06/30/23	4,200	
	K. Hills-Muckey	01/13/21	06/30/21	2,000	*
	C. Lopez-Cleary	03/16/21	03/15/22	4,200	*
	P. Moresco	07/16/21	07/15/22	4,200	*
Charles H. Revson Foundation, Inc.	Dr. Syrjanen	11/01/19	05/01/22	56,288	
	Dr. Furlan	10/01/19	04/01/22	56,288	
The Swartz Foundation	Dr. Koulakov	01/01/21	12/31/21	7,297	
	Drs. Engel/Langdon	01/01/21	12/31/21	65,116	
	Drs. Navlakha/Shen	01/01/21	12/31/21	62,587	
Course Support					
Howard Hughes Medical Institute	Course Program	08/01/20	07/31/25	500,000	
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	04/01/21	08/31/22	12,920	
Meeting Support					
Brains for Brain Onlus Foundation	Miscellaneous	04/01/21	12/31/21	1,500	*
JTech Foundation	Biological Data Science	12/01/20	11/30/22	1,500	
Merck Sharp & Dohme Corp.	HOST Meeting	09/01/21	08/31/22	5,000	*
Stemcell Technologies	3D Brain Meeting	07/01/21	12/31/22	5,000	*
ViiV Healthcare	Biohistory Series: 50 Years of Reverse Transcriptase	08/18/16	12/31/22	20,000	

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2021

Grantor	*Program/Principal Investigator*	*Project Start Date*	*Project End Date*	*2021 Funding*[1]	
Library Support					
Council on Library and Information Resources		05/01/21	04/30/22	25,940	*
The Ellen Brenner Memorial Fund		12/15/19	12/14/22	7,000	
Goelet LLC		01/01/21	12/31/22	12,600	*
The New York State Education Department		07/01/21	06/30/22	4,128	
The River Foundation/Celia and Wally Gilbert		10/27/21	10/26/22	30,000	*
Preprint Server for Biology					
Chan Zuckerberg Initiative	Dr. Inglis	06/02/20	08/31/22	1,052,469	
Anonymous	Dr. Inglis	05/01/17	07/30/22	1,867,384	

[1]Awarded, including direct and indirect costs
*New or competing renewals or supplements awarded in 2021

DNA LEARNING CENTER GRANTS

Grantor	*Program*	*Duration of Grant*	*2021 Funding*+
FEDERAL GRANTS			
National Institutes of Health	*Citizen DNA Barcode Network*	6/20–3/25	$252,292
National Science Foundation	*Implementing DNA Barcoding for Course- Based Undergraduate Research Experiences*	10/18–9/23	262,922
National Science Foundation (University of Arizona)	*CyVerse: Cyberinfrastructure for the Life Sciences*	8/18–7/23	164,532
National Science Foundation (The Washington University)	*RCN-UBE: Establishing a Genomics Education Alliance: Steps Towards Sustainability*	9/18–8/22	1,573
National Science Foundation (Austin CC)	InnovATE*BIO* National Biotechnology Education Center	10/19–9/24	175,444
National Science Foundation	*What Works in Workshops—Evolving Short Format Training to Serve Life Sciences STEM Professionals in the 21st Century*	3/21–9/22	14,051
NONFEDERAL GRANTS			
Beijing No. 166 High School	Chinese Collaboration Agreement	7/19–6/22	0
Breakthrough Prize Foundation	Laboratory Design and Teacher Training for Breakthrough Junior Challenge Prize Winners	12/15–12/22	18,224
Health Park	Health Park Agreement	12/15–12/22	864
Pinkerton Foundation	*Urban Barcode Research Program*	1/21–5/22	77,584
Richard Lounsbery Foundation	*Developing Independent Student Marine Biodiversity Research Using eDNA*	10/17–12//21	2,513
Paul Taubman	Paul Taubman support for *DNALC NYC* at City Tech Exhibit Development	6/21–6/22	113,396
The Simons Foundation	*Urban Barcode Research Company*	12/17–8/22	18,992
William Townsend Porter Foundation	*Harlem DNA Lab for Underprivileged Students*	1/20–1/22	2,983
Office of Brooklyn Borough President	*DNALC NYC* at City Tech Video Wall	10/21–10/22	87,010
Hudson River Park Trust	Environmental DNA Survey in Hudson River Park's Estuarine Sanctuary	1/20–12/21	54,926
NY Harbor Foundation	Billion Oyster Project	6/20–12/21	5,887
Department of Design and Construction	NYC—Department of Design and Construction DNALC-Brooklyn		438,336
National Grid Foundation	Genetics Education Program and Advanced Genetics	9/20–9/21	15,000

+ Includes direct and indirect costs.

The following schools and school districts participated in these membership programs of the DNALC:

Sustaining Memberships

Bellmore–Merrick Central High School District	$3,000
Elwood UFSD	$3,000
Herricks Union Free School District	$3,000
Huntington Union Free School District	$3,000
Island Trees Union Free School District	$3,000
Jericho High School	$3,000
Levittown Union Free School District	$3,000
North Shore Central School District	$1,750
Oceanside Union Free School District	$3,000
Oyster Bay–East Norwich Central School District	$3,000
Plainview–Old Bethpage Central School District	$3,000
Port Washington Union Free School District	$3,000
Portledge School	$3,000
Roslyn Union Free School District	$3,000
Syosset Central School District	$3,000
Yeshiva University High School for Girls	$3,000

Associate Memberships

Glen Cove Central School District	$16,000
St. Dominic High School	$16,500

Partner Memberships

Cold Spring Harbor Central School District	$33,000
Long Beach Central School District	$33,000
Massapequa Union Free School District	$33,000

The following schools participated in these membership programs of the *DNALC NYC* at City Tech:

Sustaining Membership	
Stuyvesant High School	$3,000
Associate Membership	
Portfolio School	$16,500
Partner Memberships	
The Chapin School	$33,000
Lycée Français de New York	$33,000
Marymount School of New York	$33,000
St. David's School	$26,150
Fontbonne Hall	$33,000

The following school participated in this membership program of the *Regeneron DNALC*

Sustaining Membership	
Archbishop Stepinac High School	$3,000

BANBURY CENTER GRANTS

Grantor	*Program*	*2021 Funding*
NONFEDERAL SUPPORT		
Cold Spring Harbor Laboratory Corporate Sponsor Program	Environmental Consequences of Deep-sea Mining, a Comparison with Land-Based Mining (virtual)	$37,321
Cold Spring Harbor Laboratory Corporate Sponsor Program	Cold Spring Harbor Laboratory Technology and Education Council (virtual)	2,000
Cold Spring Harbor Laboratory Corporate Sponsor Program	Making Career-spanning Learning in the Life Sciences Inclusive and Effective for All, Virtual Kickoff (virtual)	2,550

CORPORATE SPONSOR PROGRAM FOR MEETINGS SUPPORT

In 2021, the National Institutes of Health provided multiple grants for individual meetings, while Regeneron, Chan Zuckerberg Initiative, and the International Brain Research Organization provided targeted support to broaden access to certain meetings for a scientist at minority-serving institutions in the United States/Canada and from low- and middle-income countries.

Contributions from the following companies provide core support for the Cold Spring Harbor meetings program: **Corporate Benefactors:** Estee Lauder Companies, Regeneron, and Thermo Fisher Scientific; **Corporate Sponsors:** Agilent Technologies, Bayer, Biogen, Bristol-Myers Squibb, Calico Life Sciences LLC, Merck & Co., Inc., and New England BioLabs; **Corporate Partners:** Alexandria Real Estate, Enzo Biochem, Gilead, Novartis, and Pacific Biosciences.

We appreciate the ongoing major financial support for our courses from the following: Helmsley Charitable Trust, Howard Hughes Medical Institute, National Institutes of Health, and National Science Foundation. The course program is normally supported by major equipment and reagent companies and we appreciate their flexibility during the pandemic.

ADVANCEMENT

The year started with much optimism that the pandemic would end, and our lives would soon be back to normal. As we know, that did not happen, and yet there was still so much to be thankful for. We saw a return to in-person fundraising events, although some were at reduced capacity and mask wearing continued when social distancing was not possible. There seemed to be a renewed excitement from our donors and a strong awareness about the importance of supporting scientific research.

One of the year's highlights was the opening of the DNA Learning Center at City Tech. CSHL's President Bruce Stillman was joined by Trustee Laurie Landeau, CUNY Chancellor Félix Rodríguez, and then–New York City Mayoral candidate Eric Adams, along with other dignitaries to officially cut the ribbon and open the facility. At full capacity, DNALC NYC at City Tech will serve more than 30,000 New York City students annually. We are thankful to the donors who contributed to the $30 million raised toward this project.

The Double Helix Medals and its unique celebration of science never seems to get old, but this year an extra buzz surrounded the event. After nearly two years, our guests seemed especially happy to dust off their formal wear and gather together to celebrate our three honorees—Reggie Jackson, Leonard S. Schleifer, and George D. Yancopoulos. The evening ended with the surprise announcement of a $300,000 gift from Reggie's good friend Jim Crane, owner of the Houston Astros. Jim was inspired by stories of CSHL's efforts to engage minority students in STEM education.

Thank you to all our donors for making 2021 an extraordinary success. A record $8.5 million was raised in vital unrestricted funding toward our science and education programs.

Looking forward to 2022!

Charles V. Prizzi
Senior Vice President for Advancement and
Special Advisor to the President

Cold Spring Harbor Laboratory Corporate Advisory Board

Established in 1992, the Corporate Advisory Board (CAB) serves as a liaison to the corporate community and assists in securing unrestricted dollars for annual support of Cold Spring Harbor Laboratory. Comprised of influential business leaders from large and small companies on Long Island and Manhattan, the CAB is a necessary and vital component to the continued success and mission of Cold Spring Harbor Laboratory.

The goals of the CAB are to:

- act as an ambassador for Cold Spring Harbor Laboratory;
- offer a level of expertise in strategic planning and marketing that will ensure success;
- educate business leaders about the importance and values of science on Long Island.

The CAB meets two times per year. The annual financial commitment to CSHL by Board members is $5,000.

Corporate Advisory Board 2021

Edward A. Chernoff, MARS, *Chair*

Michael Aboff, Aboff's Inc.
David Altman, Brown & Altman, LLP
Paul Amoruso, Oxford & Simpson Realty
Todd Andrews, Centerbrook Architects and Planners
Stephen R. Barrese, Dilworth & Barrese
Eric Belfi, Labaton Sucharow LLP
Edward Blaskey, Sterling National Bank
Ryan J. Bohn, Tempus Financial Group
Jean Cacciabaudo, M.D.
Thomas J. Calabrese, Daniel Gale Sotheby's
Christopher Callaghan
John D. Catalano, Catalano Enterprises, LLC
Richard A. Catalano, KPMG, LLP
George Chacon, Bethpage Federal Credit Union
Jonathan Connors, Wells Fargo
Marian Conway, Ph.D., NYCB Foundation
Denis Cullen
Phil D'Avanzo, Cushman & Wakefield
Tracy Dellomo, UBS
Gregory DeRosa, Roanoke Holdings
Lauren Emr, Esq.
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
Thomas Gsell, R2DConsulting, LLC
Mary Jane Helenek
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, Enzo Life Sciences
Peter Klein, Aline Wealth
Amy Koreen, M.D.
Laurie J. Landeau, V.M.D.
Brian D. Lee, Newmark Grubb Knight Frank
David Lessing, Lessing's
Jeffrey L. Martin, M.D., Sight M.D.
Mark McAteer, The Laurel Group
Stephen F. Melore, Farrell Fritz, P.C.
Richard Nattis, M.D.
Robert Palatnick, DTTC
John G. Passarelli, M.D.
David Peikon
Nicole Prizzi, Harbor Digital Properties
Erin Rechler
Joseph Roberto, BankUnited, N.A.
Keith Rooney, National Grid
Stephen Ross, Nikon
Don Saladino, Drive 495
Raju Sarwal, M.D.
Carmella L. Stephens, Ph.D., Carter, DeLuca & Farrell
Edward Strohm, Three Strohm Sisters Family Foundation
John Topolovec, TD Bank
Deborah Messina Tsotsos, DO
Craig A. Weiss, T. Weiss Realty Corp.
Dave Zuklie, The Swoodle Society

Cold Spring Harbor Laboratory Association

Under the leadership of CSHL Association President Kristin Olson Smith, the CSHLA community raised $8.7 million through the annual support as well as fundraising and outreach events. Although the pandemic continued to pose event challenges, we were able to hold most of our events in person. The 20th annual Women's Partnership for Science event was Sunday, September 19. A record 300 attendees listened to CSHL Assistant Professor Jessica Tollkuhn speak about her work on estrogen's roles in brain development, mental health, and disease. Former director of the CSHL Association board Lori Bahnik was the honoree. The 17th annual Double Helix Medals event was on November 17, 2021, at the American Museum of Natural History, and raised a record $5 million to support biology research. After going virtual for the 2020 ceremony, guests of the 16th annual gala were excited to be back in person this year in New York City. Lesley Stahl, the American television journalist who emceed the dinner, said "there is still nothing like being together under the whale in the Hall of Ocean Life at the Museum of Natural History!" The event honored three people: baseball legend Reggie Jackson and Regeneron Pharmaceuticals, Inc., cofounders Leonard S. Schleifer and George D. Yancopoulos. New York City Mayor-elect Eric Adams took the stage to champion biological research and STEM (Science, Technology, Engineering, Math) education. The CSHLA Directors continue to do a stellar job representing CSHL as community ambassadors, and we are grateful for their service.

Officers

Kristin Olson Smith, *President*
James A. Woods, *Treasurer*
Ron Gottlieb, Virginia Knott, *Co-Secretaries*

Directors

Elizabeth Ainslie
Mary Auersperg
Hans E.R. Bosch
Lisa Broadbent
Barbara W. Callaghan
Eric Carlstrom
Edward A. Chernoff
Frank DellaFera
Debra Del Vecchio
Nelson DeMille
David Einbinder
David Goldring
Mark Hamer
Nancy Lippman Israeli, M.D.
Carissa Jordan
Ashley Jostrom
Joel Kassimir, M.D.
Jack Kelly
Terri Keogh
Errol Kitt
Nick Leopard
Madelyn Lombardi
Michael Maturo
Marcia Kramer Mayer
Michèle Bahnik Mercier
Nina Monell Morton
Eileen Otto
Patricia Petersen
Whitney F. Posillico
Rita Ranieri
Alicia Zarou Scanlon
Lonnie Shoff
Hope Geier Smith
David M. Stark
Mary Striano
Peter Tilles

Honorary Directors

Anne R. Meier
Cathy Cyphers Soref
Pearl F. Staller
Cynthia R. Stebbins

2021 CSHL Association Directors

Honor Roll of Donors

Lifetime Contributions of $5 million+

Dr. and Mrs. Lalit R. Bahl
The Arnold and Mabel Beckman Foundation
BGI
Jamie Nicholls and Fran Biondi
Mr. and Mrs. David Boies, Boies, Schiller & Flexner LLP
The Breast Cancer Research Foundation
Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation
The Dana Foundation
Henriette Darrell
DeMatteis Family Foundation
Mr. and Mrs. Charles Dolan, Dolan Family Foundation
The William Stamps Farish Fund
Charitable Lead Annuity Trust under the will of Louis Feil
Jacob Goldfield
Kate Medina Guthart and Leo A. Guthart
The Leona M. and Harry B. Helmsley Charitable Trust
Mr. and Mrs. Jeffrey E. Kelter
David H. Koch*
Laurie Landeau Foundation
Mr. and Mrs. Robert D. Lindsay and Family
Ivana Stolnik-Lourie and Robert W. Lourie
Lucille P. Markey Charitable Trust
The Lustgarten Foundation
Nancy Abeles Marks
The G. Harold and Leila Y. Mathers Charitable Foundation
Mercer Family Foundation
The Don Monti Memorial Research Foundation
Dr. and Mrs. Howard L. Morgan
New York Empire State Development Corporation
Pershing Square Foundation
Cynthia Hazen Polsky and Leon Polsky
The Quick Family
Charles and Marie Robertson Family
Edith Seligson
The Simons Foundation
Drs. Marilyn and James Simons
Alfred P. Sloan Foundation
St. Giles Foundation
The Starr Foundation
Dr. and Mrs. James M. Stone
The Swartz Foundation
Mr. and Mrs. Paul J. Taubman
The Thompson Family Foundation
Dr. and Mrs. James D. Watson
Mr. and Mrs. Henry Wendt III
The Roy J. Zuckerberg Family Foundation

Lifetime Contributions of $1 million+ or $100,000+ in 2021

Elizabeth and Lee Ainslie
Rita Allen Foundation, Inc.
The Bahnik Foundation
Carson Family Charitable Trust
The Mary K. Chapman Foundation
Mr. and Mrs. Edward A. Chernoff, MARS
Whitney and Jim Crane
Mr. and Mrs. Roderick H. Cushman
Davenport Family Foundation
Michel David-Weill
Pamela Hurst-Della Pietra and Stephen Della Pietra
Oliver S. and Jennie R. Donaldson Charitable Trust
Edward P. Evans Foundation
The Robert David Lion Gardiner Foundation
Robert A. Gay
Miriam and Alan E. Goldberg Foundation
Lola Goldring
The Oliver R. Grace Family
Michael J. Griffin
Guru Krupa Foundation, Inc.
Janet Strauss and Jeff Hawkins
Jo-Ellen and Ira Hazan
William Randolph Hearst Foundation
Heartfelt Wings Foundation
Elizabeth McCaul and Francis Ingrassia
F.M. Kirby Foundation, Inc.
Kissinger Family Foundation
Knott Family Foundation
Andre B. Lacy Revocable Trust
Betsy and Bryan H. Lawrence
The Lehrman Institute
Mr. and Mrs. Stephen M. Lessing
Jerome Levy Foundation
Mary D. Lindsay*
Louis Morin Charitable Trust
The Mark Foundation for Cancer Research
Dr. Marcia Kramer Mayer
Estates of Florence and Harold & Ethel McNeill
The Meier and Linnartz Family Foundation
Gillian and Eduardo Mestre
William R. Miller*
New York Community Trust
William C. and Joyce C. O'Neil Charitable Trust
The Pfizer Foundation
Amy and John Phelan
Emma Liu and Karl Qin
Mr. and Mrs. Thomas A. Saunders III
Alison Holtzschue and Douglas Schloss
Eleanor Schwartz Charitable Foundation
The Seraph Foundation
Pearl F. Staller
Cynthia R. Stebbins
Waclaw Szybalski, D.SC.*
Diana Taylor
TD Bank
Estate of Christianne Tsakis
Dr. and Mrs. Stuart T. Weisbrod
Anne Wojcicki Foundation
Drs. George Yancopoulos, Regeneron Pharmaceuticals

Contributions of $30,000+

Alnylam Pharmaceuticals, Inc.
Anonymous, Dr. and Mrs. Robert Biggar
Mary and Paul Auersperg
Mr. and Mrs. Hans E.R. Bosch
Cedar Hill Foundation
Casey and Ellen Cogut
Del Vecchio Family Foundation
Barbara Amonson and Vincent Della Pietra
Laura and John Desmarais
Ike, Molly and Steven Elias Foundation
Elizabeth Cogan Fascitelli and Michael D. Fascitelli
Douglas and Christine Fox
Friends of TJ Foundation
Celia and Wally Gilbert
Pierre & Paula Gonthier Family Foundation
Mr. and Mrs. Ronald A. Gottlieb
The Marc Haas Foundation
Mr. and Mrs. Mark W. Hamer
Annette Kade Fund
Mr. and Mrs. Errol Kitt
Peter Klein, Claire Friedlander Family Foundation
Vesna and Tomislav Kundic
Dr. Lee MacCormick Edwards Charitable Foundation
Maddie's Promise
Liz's Day For Breast Cancer Research
The Robert J. and Claire Pasarow Foundation
Lyon Polk, Morgan Stanley Private Wealth Management
Linda Johnson and Bruce Ratner
Christina Renna Foundation
William J. Riley Memorial Fund
Geoff Robertson, Robertson Family Fund
The Mary Ruchalski Foundation
Christine Anderson and Jake Siewert
Sight Growth Partners
Skanska USA Building, Inc.
Mr. and Mrs. Harry Slatkin
Kristin and Dustin Smith
Mr. and Mrs. Theodore Smith III, The Geier Foundation

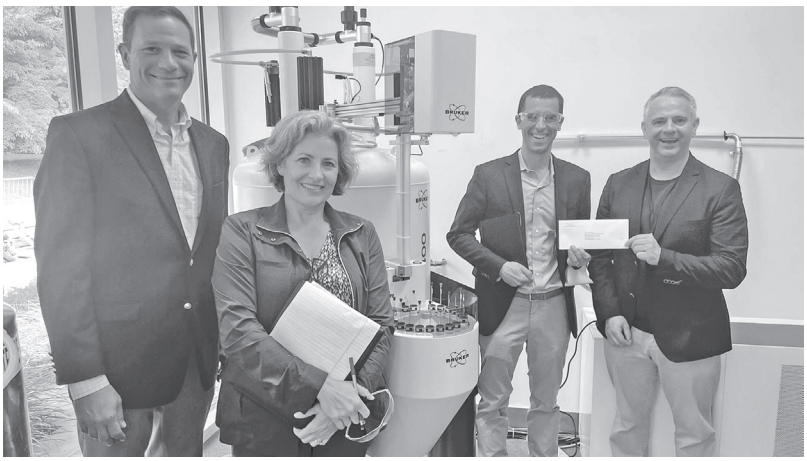
*Deceased

Darlene Carbone Brain Tumor Foundation presents a check to Dr. Alea Mills and her laboratory members.

Sunshine Foundation
Swim Across America Nassau-Suffolk
Three Strohm Sisters Family Foundation
The Wasily Family Foundation
Dr. and Mrs. Michael Wigler

Contributions of $10,000+

AKF Group
Mr. and Mrs. W. Dillaway Ayres, Jr.
Mr. and Mrs. Fred H. Baron
Ed Blaskey, Sterling National Bank
Botto Mechanical Corp.
Laura Louise Breyer
Broad Hollow Bioscience Park
Lisa and Tim Broadbent
Estate of G. Morgan Browne
The Caldwell Stair Charitable Fund
Barbara and Christopher Callaghan
Capital Group
Eric and Keri Carlstrom
Centerbrook Architects and Planners, LLC
Dr. Bayard D. Clarkson
Frederic R. Coudert Foundation
Maria and Geoff de Lesseps - Alpha Omega Charitable Foundation
Frank and Janet Della Fera Family Foundation
Nelson DeMille
Mr. and Mrs. Gregory DeRosa
Mr. and Mrs. Mark D. Eichorn
Dr. and Mrs. Philip Goelet
Jon & Mindy Gray Family Foundation
Randall and Mary Hack Foundation
Irving Hansen Foundation
Susan T. Harris
Nancy and Ron Israeli
Carissa and James Jordan



Diana Kostas and Justin Kiczek from the F.M. Kirby Foundation present a check to Dr. John Moses and Sr VP of Advancement Charlie Prizzi.

Ashley and Gabriel Jostrom
Dr. and Mrs. Joel J. Kassimir
Jack and Carmela Kelly
Terri and Peter Keogh
Andrea B. and Peter D. Klein
Gordon Lamb, Jefferson Family Charitable Foundation
Melissa and Nick Leopard
Rita and Art Levinson
Cynthia and David Lippe
Amanda and Tom Lister
Madelyn and Carl Lombardi
Michael Maturo, RXR Co Property Management LLC
The Ambrose Monell Foundation
Thomas Moore
Monique and Doug Morris
Ralph Nappi, Northwell Health
National Grid Foundation
Michael A. Nickols
Susanne Olin
O'Neill Family Charitable Trust
Jonathan and Eileen Otto
Lauryl and Robert Palatnick
Michelle Paternoster Foundation
David P. Pearson
Patricia Petersen, Daniel Gale Sotheby's International Realty
William Townsend Porter Foundation
John and Cheryl Pufahl
Thomas and Marina Purcell Family Foundation
Lewis and Rita Ranieri
Erin Rechler, The Morton and Beverly Rechler Family Foundation
John R. Reese
Pamela and Richard Rubinstein Foundation
Alicia Zarou Scanlon and Richard T. Scanlon
Susan M. Schultz*
John Sobolewski
David Stark, Teva Pharmaceutical Industries
Dr. and Mrs. Bruce W. Stillman
Mary and Vincent Striano, Striano Electric Co., Inc.
Anne Drackett Thomas
Dori and Peter Tilles
Joan P. Tilney
Douglas Torre
Leslie S. Farhangi and John P. Tuke
Ann Eden Woodward Foundation
The Bradley Zankel Foundation

Contributions of $5,000+

Aboff's, Inc.
David Altman, Brown & Altman
Mr. and Mrs. Paul F. Amoruso
Photeine Anagnostopoulos
Baranay Family Foundation

*Deceased

Double Helix honoree Reggie Jackson with emcee Lesley Stahl, Bruce Stillman, and Jim Crane.

Eric Belfi, Labaton Sucharow, LLP
Danielle Bellanger
Kathy and Gene Bernstein
Bethpage Federal Credit Union
Hadley and Ryan Bohn
Steven Brenner, M.D.
Lauren Emr and William Brunner
Jean Cacciabaudo, M.D. FACC
Mr. and Mrs. Thomas J. Calabrese, Jr.
Christiane Hurley Zoller and John T. Callaghan
Maddy Callahan, Helms Bros. Inc.
PJ Callahan Foundation
Lawrence and Teresa Cantwell
Mary and Richard Catalano
Tom and Beatriz Claugus, Foundation for a Better World
Rita M. Cleary
Mr. and Mrs. Jonathan Connors
Dr. Marian Conway, New York Community Bank Foundation
Cullen Family Fund
David Einbinder
Michael L. Focazio, Carissa Maringo Fund
Jim Ford, Eppendorf
Putney Cloos and Brian Fox
Amit Gandhi
Kelly Gaudreau
GEI Consulting Engineers & Scientists
Tom Giarraputo, Executive Cleaning Services
Gil-Bar Industries
Kevin Gillen, Current Fire Protections
Bernard F. and Alva B. Gimbel Foundation
Thomas and Kim Gsell
Robin Hadley
Reverend Linda Peyton Hancock
The Hastings Foundation, Inc.
Donald Hehir & Associates, LLC
Mary Jane Helenek
Robin and Jim Herrnstein
Paula and John Hornbostel
Yaoqing Huang Family Foundation
Richard Humann, H2M architects + engineers
Joan M. Hutchins
Bob Isaksen, Bank of America
Alan Jakimo
JSRM Foundation
Susan and John Kean, Kean Development Company, Inc.
Michael Keenan, Wells Fargo Private Bank
Sarah and Jim Kennedy
Shahnaz Batmanghelidj and Radford W. Klotz
Louise Hochberg and Brian Lee
Robert Levey, Certified Interiors
Hildegarde Mahoney, The Dana Foundation
Robert Marcus
Edward E. Matthews
Chris McIntosh, J.P. Morgan Private Bank
Stephen F. Melore, Farrell Fritz P.C.
Dr. and Mrs. Richard Nattis
James J. Norman
Hugh and Arianne O'Kane
Ashley and Frank O'Keefe
Tracy Dellomo and Clifford Packingham
Debbie Parmet Sondock
Dr. and Mrs. John Passarelli
Whitney Posillico
Spionkop Charitable Trust
Nicole and Charlie Prizzi
Drs. Christina Bucci-Rechtweg and Jay Rechtweg
Dr. Stephen Ross, Nikon Instruments, Inc.
Tara and Tim Ruocco
Dr. and Mrs. Raju Sarwal
The Silfen Family
Mr. and Mrs. Douglas S. Soref
Drs. David and Mona Spector
Lee Steinberg
Mr. and Mrs. George Tsunis
UBS
Diana and Roy Vagelos
John B. Vermylen
Mr. and Mrs. Paul A. Vermylen, Jr.
Vincent Vivolo
Thomas Vladeck
Craig Weiss, T. Weiss Realty Group
Mr. and Mrs. David Zuklie

Contributions of $1,000+

Mr. and Mrs. Todd Andrews
Debra and Scott Arenare
Erica and Henry Babcock
Robert and Elizabeth Bailenson
Dorothy T. Baldwin
BankUnited, N.A.
Caroline E. Bassett
Elizabeth Beach
Patrick Bencivenga
Joanne Berger-Sweeney, Trinity College
Elizabeth Bertani
Helen Bonebrake, The Corita Charitable Trust
Michael Botchan
Linda Riefler and Christopher Boyatt
Dr. William Braden, Porter Braden Fund at the Rhode Island Foundation
Vincent and Patricia Breitenbach
Gerard Brogan
Robert Brooks
Richard C. Cahn
Stephen and Carol Canter
Devon and Tommy Carroll
Louise Parent and John Casaly
Bernadette Castro
John Catalano
Mr. and Mrs. James C. Cook
Cosel-Pieper Family Foundation
Angela Creager
Lucy Cutting
Dr. James E. Darnell
Mr. and Mrs. Jack Dolce
Dr. Nancy Douzinas
Carol Durkin
Sarah Edwards
Frederick W. Engel
Diane and Bob Fagiola
Vicki and Joseph Ferrara
Floral Park Police Benevolent Association
E. Maxwell Geddes, Jr.

Double Helix Medals honorees Len Schleifer and George Yancopoulos with then NYC Mayor-elect Eric Adams.

Sander Gerber
Patricia L. Gondelman
Carolyn Gulotta
Mr. and Mrs. Frank Gundersen
Margaret M. Hargraves
Gwen Harris, M.D.
William A. Herzog
James B. Hoover
The Hotchkiss School
Mr. and Mrs. A. Joseph Iannucci
George Igel
J. McLaughlin
Walter B. James Fund No. 2
Drs. Clary Evans and Tobias Janowitz
Jean and Barclay Jones
Dr. Leemor Joshua-Tor
Mr. and Mrs. Clarke R. Keough
Drs. Kathleen and Victor Klein
The Kleinknecht Family Foundation
Patricia Klingenstein, John & Patricia Klingenstein Fund
Steven B. Klinsky
Laureen and Ragnar Knutsen
James Large
Suzanne and Richard Leeds
Nancy Metz and George N. Lindsay, Jr.
The Litwin Foundation
Kevin G. Mahony
Jane and Philip Mallinson
Masthead Cove Yacht Club
Ferrell and Chi McClean
Peggy McCormack
Neil and Amy McGoldrick
Nina Monell Morton
Jeanne Moutoussamy-Ashe
Mr. and Mrs. Stephen V. Murphy
Arthur J. Nagle
Nitorum Capital LP
Norwood Foundation, Inc.
Mr. and Mrs. Hugh O'Kane
Dr. and Mrs. James Paci
Carole and Nicholas Parks
Rhoda Parmet
Violetta and Douglas Partrick
Nicholas Paumgarten
Mr. and Mrs. David Peikon
Lou Ann Montana and J.W. Pflugrath
Mr. and Mrs. Robert A. Pilkington
Mr. and Mrs. Bennett Rechler
Joseph Roberto
Richard and Jean Roberts
Mr. and Mrs. William Roche
Linda Rodgers
Dr. and Mrs. Gerald M. Rubin
Brian and Ginny Ruder
Heather and Adam Savarese
Tami and Scott Schneider
Robin and Enrique Senior
Julie and Bryan Shepherd
Jane and James Spingarn
Alice and David Stamps
Marcia and Myron Stein
Joan Steitz
David Sterling
Jim and Pat Stewart
Stream Data Centers
Michele Miroff and Carl J. Studer
Yaakov Tambor
Pamela M. Thye
Louise Giffuni Tiernan Foundation
Patricia W. Timpson
Triumph Construction
Susie and Stanley Trotman
Debra Tuccillo-Kenney
Universe Kogaku (America), Inc.
Dr. Marjorie J. van de Stouwe
Richard John Walk
Douglas A. Warner III
Stacy Welsh
Gerald I. White
Sandy and Jennifer Williams
Steven and Georgene Winick

In-Kind

Mary Auersperg
Lori Bahnik
The Beauty Window
Benjamin Talks, Carissa Jordan, and Nikki Boulukos
Besito
Maddy Callahan
Eddie Chernoff
Suzanne Christenesen
Debra Del Vecchio
Daniel Gale
Nelson DeMille
Estee Lauder
Finnegan's Restaurant
Fiorello Dolce
Fuentes Cosmetic Surgery
Dagnia and Walter Goldschmidts
Kimberly Gorman Muto
Huntington Country Club
Il Posto di Joey
Jonathan's Ristorante
Natalie Kammer
Jeff Kelter
Ginny and Errol Kitt
Sarah Kitt
Kisner Plastic Surgery, SILK MD
Lessing's
Danny Miller
Nikon Instruments
Oheka Castle
Stephanie Orzel
Eileen Otto
Dr. John Passarelli
Corine Perrotta
Piccolos
Poll Restaurants
Annabel Romero Hernandez
Sandbar
Dr. Raju Sarwal
Alicia Scanlon
Alissa Shea
The Shed
Laura Slatkin
Dr. Bruce Stillman
Sweetie Pies
TD Bank
TERRA Cold Spring Harbor
Leslie Farhangi and John Tuke
Marjorie van de Stouwe
Andrew Viklund
Wilson
Windham House, Peggy King

Contributions in honor of

Edward A. Chernoff
Daniel Cordani
CSHL Legal & Business Development Teams

Golf Chair Eddie Chernoff, Bruce Stillman, and honoree John Passarelli.

Kristina P. Davison
Mary Epstein
Ron Gottlieb
Leo Guthart
Jessica and Thomas Heckel
Errol Kitt
Robert Lindsay
Dr. Marcia Kramer Mayer
Dr. Laurie Landeau
Dave Mirolli
Eleanor and Howard Morgan
Nina Morton
Timothy Motto
Nassau County Sheriff's COBA
Kate and Lilly O'Phelan
Port Authority PBA
Michael C. Schatz, Ph.D.
Douglas Schloss
Dr. Bruce W. Stillman
Suffolk County DSPBA
Suffolk County PBA
Suffolk County Probation Officers Association
Danielle and Paul Taubman
Kristine Thomas
Joan P. Tilney
Dr. Dave Tuveson
Sallie and Alex van Rensselaer
Dr. Evelyn M. Witkin

Contributions in memory of

Richard Ambrosio
Shirley R. Aprison
Josephine Auersperg
Giuseppe Bertani
Rev. Anthony Boateng-Mensah
Patricia Boudreau
Ellen Brenner
Ahmad Bukhari
Dr. Robert Cafferkey
Elfie Cairns
John P. Cleary
Andrew DeNatale
Carol Giacco
Christopher G. Goff, Ph.D.
Alan Goldstein
Teresa Haire
Harold "Hap" Hoffmann
Dr. Andy Kaplan
Billy and Jackie Kastner
Robert Jack King
Amar J.S. Klar
David Knott
Mary D. Lindsay
Carissa Maringo
Marie Lewis Matthews
Rusty McCormack
Marcia A. Meyer
Frank Morelli
Mother, Lulu
Kevin and Francis O'Connor
Robert C. Olby
Gertrude Ornstein
Carol Pfeifer
Preston Pumphrey
Andrew Rosenthal
Diane Emdin Sachs
Alan Seligson
Olga Smith
Francis Patiky Stein
Jason Stewart
Steven Strongin
Bob Suydam
Edward Travaglianti
Rajani Lakshmi Velivela
Evelyn and Frank Walkowski
Andrew Weiss
Dr. Norton D. Zinder
Dr. Mark Zoller

Women's Partnership for Science lunch—President Bruce Stillman with honoree Lori Bahnik and speaker Jessica Tollkuhn.

Helix Society

W. Dillaway Ayres, Jr.
Caroline E. Bassett
Elise Best
Mrs. William M. Blair, Jr.
Drs. Joan Brooks and James Garrels
John Broven
Mr. and Mrs. Thomas J. Calabrese, Jr.
Vicki Gruber Callahan
Kate C. Chapman
Edward A. Chernoff
Dr. Bayard D. Clarkson
Mr. and Mrs. Roderick H. Cushman
Henriette Darrell
Mrs. Donald L. Deming
Jane Duggan
Jan Eisenman
Mary Epstein
Michael L. Focazio
Mr. and Mrs. Douglas B. Fox
John H. Friedman
Jean G. Gardiner

Robert A. Gay
Eleanor J. Greenan
Michael J. Griffin
Michael Gurtowski
Robin Hadley
Mr. and Mrs. Mark W. Hamer
Margaret Hargraves
Mr. and Mrs. Herman M. Heinemann
Mrs. Valdemar F. Jacobsen
Mr. and Mrs. Peter D. Klein
Dr. Daniel F. Klessig
Laurie J. Landeau, V.M.D.
Mrs. Leslie S. Learned
Mrs. Henry Lewis III
Lisa M. Manche
Dr. Marcia Kramer Mayer
Susan Morgan
Richard S. Overton
Ann B. Parkinson
John S. Popeleski
Whitney F. Posillico
Dr. Gregory Prelich
Nicole and Charlie Prizzi
John R. Reese
Roger H. Samet
Edith Seligson
Jeffrey Shellan, Ph.D
Mr. and Mrs. Irwin Sicherman, O.D.
Suzanne Slocum
Mr. and Mrs. James L. Spingarn
Pearl F. Staller
Cynthia R. Stebbins
Lee Steinberg
Michele M. Miroff and Carl J. Studer
Dr. Robert Tjian
Mr. and Mrs. Henry Wendt III

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

Research Operations

Sydney Gary, Ph.D.
Director, Research Operations

Professors

Jeff Boyd, Ph.D.
Mikala Egeblad, Ph.D.
Douglas Fearon, M.D.
Hiroyasu Furukawa, Ph.D.
Thomas Gingeras, Ph.D.
Professor and Head, Functional Genomics
David Jackson, Ph.D.
Leemor Joshua-Tor*, Ph.D.
Chair, Cancer and Molecular Biology
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.
Chair, Genomics and Plant Biology
W. Richard McCombie, Ph.D.
Alea Mills, Ph.D.
Partha Mitra, Ph.D.
John Moses, Ph.D.
Adam Siepel, Ph.D.
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.
Chair, Neuroscience
Michael Wigler, Ph.D.
Anthony Zador, M.D., Ph.D.

Associate Professors

Dinu Albeanu, Ph.D.
Camila dos Santos, Ph.D.
Christopher Hammell, Ph.D.
Molly Hammell, Ph.D.
Ivan Iossifov, Ph.D.
Justin Kinney, Ph.D.
Dan Levy, Ph.D.
Saket Navlakha, Ph.D.
Stephen Shea, Ph.D.
Doreen Ware**, Ph.D.

Assistant Professors

Arkarup Bandyopadhyay, Ph.D.
Jeremy Borniger, Ph.D.
Lucas Cheadle, Ph.D.
Alexander Dobin, Ph.D.
Tatiana Engel, Ph.D.
Tobias Janowitz,M.D., Ph.D.
Peter Koo, Ph.D.
Je Hyuk Lee, M.D., Ph.D.
Michael Lukey, Ph.D.
David McCandlish, Ph.D.
Vincent Pedmale, Ph.D.
Andrea Schorn, Ph.D.
Jessica Tollkuhn, Ph.D.

Research Professors

Alexander Krasnitz, Ph.D.
Darryl Pappin, Ph.D.

Research Associate Professors

Kenneth Chang, Ph.D.
Jonathan Preall, Ph.D.
Research Assistant Professor and Head of Genomics Tech Development

Research Assistant Professors

Paolo Cifani, Ph.D.
Siran Li, Ph.D.
Scott Lyons, Ph.D.
Michael Ronemus, Ph.D.
Johannes Yeh, DDS, Ph.D.
Seungtai Yoon, Ph.D.

Senior CSH Fellow

Lingbo Zhang, Ph.D.

CSH Fellows

Semir Beyaz, Ph.D.
Hannah Meyer, Ph.D.

Clinical Fellows

Sandeep Nadella, M.D.
Kenneth Yu, M.D.

Adjunct Professors

Richard Barakat, M.D., M.B.A.
Anne Churchland, Ph.D.
Jesse Gillis, Ph.D.
Gregory Hannon, Ph.D.
Z. Josh Huang, Ph.D.
Adam Kepecs, Ph.D.
Scott Lowe, Ph.D.
Larry Norton, M.D.

Adjunct Associate Professors

Gurinder Atwal, Ph.D.
Pavel Osten, M.D., Ph.D.
Michael Schatz, Ph.D.

Adjunct Assistant Professor

Jason Sheltzer, 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 Investigators

Nicholas Gladman, Ph.D.
Youngkyu Park, Ph.D.
Damianos Skopelitis, Ph.D.
Zihua Wang, Ph.D.

Research Investigators

Joan Alexander, M.D.
Qing Gao, M.D., Ph.D.
Sara Goodwin, Ph.D.
Gilbert Henry Jr., Ph.D.
Manzar Hossain, Ph.D.
Jonathan Ipsaro*, Ph.D.
Pramod Kumar, Ph.D.
Yoon-Ha Lee, Ph.D.
Michael Regulski, Ph.D.
Benjamin Roche, Ph.D.
Yi-Jun Sheu, Ph.D.
Asya Stepansky, Ph.D.
Apurva Tandon, Ph.D.
Kai Yu, Ph.D.

Deputy Director of Administration for the Cancer Center

Lindsey Baker, Ph.D.

Visiting Scientists/ Collaborative Scientists

Arnon Arazi, Ph.D.
Taimour Baslan, Ph.D.
David Chitty, D.O.
Megan Crow, Ph.D.
Elad Elkayam*, Ph.D.
Takeo Fujii, M.D.
Kuan-Ting Lin, Ph.D.
Jung Park
Yongjun Qian, Ph.D.
James Roach, Ph.D.
Kenny Ye, Ph.D.

Visiting Professors

John Boockvar, M.D.
Matthew Ridley, Ph.D.
Stuart Weisbrod, Ph.D.

Visiting Clinical Researchers

Lyudmyla Demyan, M.D.
Oliver Standring, M.D.

Director, MAPseq Core Facility

Huiqing Zhan, Ph.D.

*Employee of the Howard Hughes Medical Institute
**Employee of the United States Department of Agriculture

Neuro AI Scholar

Nikhil Bhattasali

Senior Fellow

Andrea Moffitt, Ph.D.

Postdoctoral Fellows

Jose Adrover Montemayor, Ph.D.
Cina Aghamohammadi, Ph.D.
Aktan Alpsoy, Ph.D.
Armend Axhemi, Ph.D.
Balasooriya Balasooriya, Ph.D.
Walter Bast, Ph.D.
Pascal Belleau, Ph.D.
Kuhulika Bhalla, Ph.D.
Jalaj Bhandari, Ph.D.
Sonam Bhatia, Ph.D.
Debmalya Bhunia Ph.D.
Francesco Boato, Ph.D.
Janeen Braynen, M.D.
Jonathan Cahn*, Ph.D.
Giuseppina Caligiuri, Ph.D.
Mackenzie Callaway, Ph.D.
Hong Goo Chae, Ph.D.
Chen Chen, Ph.D.
Wei Chia Chen, Ph.D.
Yushu Chen, Ph.D.
Tsung Han Chou, Ph.D.
Taemoon Chung, Ph.D.
Samantha Cyrill, Ph.D.
Juliane Daßler Plenker, Ph.D.
Cristiane de Santis Alves Rosa*, Ph.D.
Mary Doherty, Ph.D.
Roman Dvorkin, Ph.D.
Max Epstein, Ph.D.
Austin Ferro, Ph.D.
Matthew Fisher, Ph.D.
Talitha Forcier, Ph.D.
Nikita Francis, Ph.D.
Daniel Fuerth, Ph.D.
Alessandro Furlan, Ph.D.
Minakshi Gandhi, Ph.D.
Yuan Gao, Ph.D.
Ankur Garg*, Ph.D.
Aybüke Garipcan, Ph.D.
Mikhail Genkin, Ph.D.
Shanu George, Ph.D.
Fabian Gerth, Ph.D.
Debarati Ghosh, Ph.D.
Bryan Gitschlag, Ph.D.
Ricardo Gomez Garcia, Ph.D.
Wuqiang Guan, Ph.D.
Priyanka Gupta, Ph.D.
Amber Habowski, Ph.D.
Clifford Harpole, Ph.D.
Rasmani Hazra, Ph.D.
Jia He*, Ph.D.
Xueyan He, Ph.D.
Anat Hendelman, Ph.D.
Diego Hernández Trejo, Ph.D.
Joshua Homer, Ph.D.
Yuma Ishigami, Ph.D.
Matt Jaremko*, Ph.D.
Risa Karakida, Ph.D.
Jonathan Kastan, Ph.D.
Brian Kinney, Ph.D.
Munenori Kitagawa, Ph.D.
Olaf Klingbeil, Ph.D.
Mahdi Kooshkbaghi, Ph.D.
Shan Kuang, Ph.D.
Wenjun Lan, Ph.D.
Amy Lanctot*, 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.
Penelope Lindsay, Ph.D.
Bodu Liu, Ph.D.
Ruchi Lohia, Ph.D.
Jason Lynn, Ph.D.
Andre Machado Xavier, Ph.D.
Antonio Majdandzic, Ph.D.
Vinay Mandati, Ph.D.
Carlos Marti Gomez Aldaravi, Ph.D.
Cristian Mateo Elizalde, Ph.D.
Olivia Mendivil Ramos, Ph.D.
Evdokia Michalopoulou, Ph.D.
Kevin Michalski, Ph.D.
Sushanta Mishra, Ph.D.
Mehreen Mughal, Ph.D.
Rodrigo Muñoz Castañeda, Ph.D.
Adriana Munoz Jimenez, Ph.D.
Sukalp Muzumdar, Ph.D.
Erik Eidy Nakagaki Silva, Ph.D.
Arun Narasimhan, Ph.D.
Karthick Natarajan, Ph.D.
Jeremy Nigri, Ph.D.
Kin On*, Ph.D.
Huiwu Ouyang, Ph.D.
Meng Ouyang, M.D., Ph.D.
Farhad Pashakhanloo, Ph.D.
Prasad Pofali, Ph.D.
Anna Posfai, Ph.D.
Yijian Qui, Ph.D.
Nissim Ranade, Ph.D.
Ashlan Reid, Ph.D.
Armin Scheben, Ph.D.
Mojdeh Shakiba, Ph.D.
Yang Shen, Ph.D.
Yanliang Shi, Ph.D.
Daniele Silva Pereira Rosado, Ph.D.
Cristian Soitu, Ph.D.
Lijuan Sun, Ph.D.
Qingtao Sun, Ph.D.
Shoujun Sun, Ph.D.
Simon Sun, Ph.D.
Xueqin Sun, Ph.D.
Kyle Swentowsky, Ph.D.
Johanna Syrjänen, Ph.D.
Hsiu-Chi Ting, Ph.D.
Pavel Tolmachev, Ph.D.
Claudia Tonelli, Ph.D.
Thu Tran, Ph.D.
Rohit Tripathy, Ph.D.
Nitinkumar Upadhyay, Ph.D.
Anand Vasudevan, Ph.D.
Prabhadevi Venkataramani, Ph.D.
Kaleb Vinehout, Ph.D.
Dharmendra Vishwakarma, Ph.D.
Ledong Wan, Ph.D.
Yue Wu, Ph.D.
Xiaosa Xu, Ph.D.
Alexander Xue, Ph.D.
Tao Yang, Ph.D.
Wen-Hsuan Yang, Ph.D.
Min Yao, Ph.D.
Toyoki Yoshimoto, Ph.D., M.D.
Li Yuan, Ph.D.
Sophia Zebell*, Ph.D.
Xian Zhang, Ph.D.
Xiang Zhao, Ph.D.
Yixin Zhao, Ph.D.
Imanol Zubiete Franco, Ph.D.

Graduate Students

Dexter Adams
Disha Aggarwal
Lyndsey Aguirre
Sergey Alekseev
Seamus Balinth
Batuhan Baserdem
Benjamin Berube*
Brianna Bibel*
Isobel Bolger
Sara Boyle
Leah Braviner
Paul Bunk
Salomé Carcy
Yuxin Cen
King Hei Cheng
Charlie Chung
Danielle Ciren
Cristian Cleary
Alberto Corona
Patrick Cunniff
Jed de Ruiter-Swain
Katherine Denney
Marie Dussauze
Ilgin Ergin
Santiago Espinosa
Miriam Ferrer Gonzalez
Connor Fitzpatrick
William Galbavy
Jordan Gegenhuber
Iacopo Gentile
Xiao Han
Amritha Varshin Hanasoge Somasundara
Samantha Henry
Kelly Hills-Muckey
Nikolas Holland
Yixin Hu
Yuzhao Hu
Sessen Daniel Iohannes
Emily Isko
Shruti Iyer
Hoda Jaber Ansari
David Johnson
Jessica Kahng
Vahag Kechejian
Seung Tea Kim
Sun Kim
Sam Kleeman
Alexander Kral
Christopher Krasniak
Asad Aziz Lakhani
Steven Lewis
Heng Liang
Lingjie Liu
Dennis Maharjan
Diogo Maia e Silva, M.D.
Craig Marshall
Ziyi Mo
Philip Moresco
Manojit Mosur Swamynathan
Emmanuella Nnuji-John
Kathryn O'Neill
Alexa Pagliaro
Michael Passalacqua
Matthew Peacey
Jordan Pearson
Rachel Polfer
Zhe Qian
Yihan Qin
Leonardo Jared Ramirez Sanchez
Ricardo Raudales
Jenelys Ruiz
Deborah Rupert
Khristina Samoilova

*Employee of the Howard Hughes Medical Institute

Danilo Segovia
Luqun Shen
Margaret Shevik
Padmina Shrestha
Nicole Sivetz
Dongyan Song
Martyna Sroka
Joshua Steinberg
Ziqi Tang
Lucia Tellez Perez
Jennifer Thalappillil
Shushan Toneyan
Gaurang Trivedi
Daniella van de Lisdonk
Dillon Voss
Julia Wang
Jonathan Werner
Xiaoli Wu
Cole Wunderlich
Wenbo Xu
Ran Yan
Lucia Yang
Zhezhen Yu
Chengxiang Yuan
Narges Zali
Qian Zhang
Xiaoyue Zheng
Xingyu Zheng

Computational Scientists

Saiteja Abbu
Samik Banerjee
Sarah Chapin
Kapeel Chougule
Astrid Deschenes
Corey Elowsky
Evan Ernst*
Brian He
Juliene Hinds
John Hover
Bingxing Huo, Ph.D.
Melissa Kramer
Sunita Kumari, Ph.D.
John Lee, Ph.D.
Nicholas Lee
Xu Li
Zhenyuan Lu
Steven Marks
Christopher Mezias, Ph.D.
Ethan Mulle
Andrew Olson
Jason Palmer
Xiaoli Qi, Ph.D.
Oliver Tam, Ph.D.
Marcela Tello-Ruiz, Ph.D.
Marygrace Trousdell
Raditya Utama, Ph.D.
Xuehong Wei
Fengyuanshan Xu

Biostatistician

Taehoon Ha

Research Associates

Kenneth Addison
Carmelita Bautista
Yon Chang
Hilary Cox
Jorg Drenkow
Adrian Gomez, Ph.D.
Sanjeev Kaushalya, Ph.D.
Hyun Soo Kim*, Ph.D.
Lisa Manche
Joseph Merrill
Joseph Milazzo
Suzanne Russo
Regina Shaw
Jianjun Sun
Wiktor Wadolowski
Caizhi Wu, M.D.
Melody Wu, Ph.D.
Lifang Zhang, Ph.D.

Scientific Project Managers

Katherine Brenner
Katherine Matho, Ph.D.

Laboratory Facility Managers

Sabrina Boettcher
Amy Brady
Julie Cheong
Robert Eifert
Elena Ghiban
Stephanie Goldsmith*
Kelly Ann Holfester
Pamela Moody
Timothy Mulligan
Michele Murrell
Eleni Noutsos, Ph.D.
Hardik Patel
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
Alexandra Boyle
Barbara Cascone
Melissa Daley
Becky Dong
Suzan Downey
Tricia Forys
Susan Fredericks
Lisa Kimoto
Caryn Koza
Idee Mallardi
Lindsay Monaco
Kylie Parker
Jessica Peluso
Catherine Scalise
Kara Spardy Sweeney
Madeline Wisnewski

Research Technicians

Yohan Auguste
Andalus Ayaz
Adrian Berisha
Artan Berisha
Lily Brine
Erica Bulzomi
Denise Cahn
Megan Chamberland
Jane Choe
Lisa Christensen
Michael Ciccone
Anthony Cocozzelli
Stefan Constantinou
Devon Cowan
Tania Cuevas Moreno
Martin Davis
Breanna Demestichas
Divyansha
Rhonda Drewes
Sintia Escobar Avelar
Audrey Fahey
Margaret Falkowski
Brian Farrell
Christy Felice
Jesse Filippazzo
Michael Fiore Jr.
Patrick Flannery
Sarah Ford
Corinne Freivaid
Victoria Gaeth
Libia Garcia
Thomas Genovese
Steven Gluf
Camila Gonzalez
Antoine Gruet
Inessa Hakker
Kaarina Hanington
Cheng Wen Hsu
Qingting Hu, M.D.
Asma Kaleem
Ece Kilic
Rebecca Koelin
Matthew Licht*
Dennis Lin
Ying Hsiu Liu
Jynelle Liverpool
Brianna Lodato
Senem Mavruk Eskipehlivan
Gregory McDermott
Colin McLaughlin
Julie Micko
Kristin Milicich
Francesca Minicozzi
Nicholas Mourikis
Stephanie Muller
Payal Naik
Benan Nalbant
Joseph O'Rourke
Kadir Ozler
Katherine Papciak
Umamaheswari Ramu
Danil Ratnikov
Claire Regan
Michael Riggs
Gina Robitaille*
Julie Rosenbaum
Ryan Santos*
Stephen Savoia
Brooke Seman
Liam Shanley
Radhashree Sharma
Sergey Shuvaev
Noriko Simorowski
Tara Skopelitis
Liam Spielmann
Logan Stecher
Luce St. Surin
Olivia Stamatatos
Danielle Stauder
Caitlin Tsang
Uma Vrudhula
Jing Wang
Robert Wappel
Yi-Chen Wu
Brian Yueh

Laboratory Facility Staff

Rachel Rubino, DVM
Director of Lab Animal Resources/Attending Veterinarian
Doris Abrego
Alexander Aguilera
Katherine Atehortua

*Employee of the Howard Hughes Medical Institute

Valerie Bagan
Carlos Ballon
Lisa Bianco
Amanda Bjertnes
Oral Bryan
Amy Burrows
Manuel Canas
Jodi Coblentz
Steven Collins
Kristin Cox
Charlene De Poto
Maria Demott
Eileen Earl
Sara Fredriksen
Pamela Guerra
Jill Habel
Amanda Jensen
Joseph Kelly
Michael Labarbera
Dylan Marcus
Daniel Mendoza
Maria Mosquera
Gustavo Munoz
Angel Oliveira
Jason Pellegrini
Bryan Rose
Laura Stonebridge
Walter Vigil Garcia

Media Makers/Farm Assistants

Nancy Bolanos
Terrance Chisum
Blaine Fitzgerald
Adriana Hincapie
Shouyun Qiao
Amy Qiu
Kyle Schlecht

EDUCATION

Banbury Center

Rebecca Leshan, Ph.D.
Executive Director
Alicia Franco
Hannah Stewart
Duncan Yates

CSHL Press

John R. Inglis, Ph.D.
Executive Director
Barbara Acosta
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
Sol Fereres Rapoport, Ph.D.
Olaya Fernández Gayol
Bibiane Garite
Sanchari Ghosh, Ph.D.
Michael Henigman
Samantha Hindle, Ph.D.
Nancy Hodson
Deborah Jarski
Brandon Kelly
Katherine Kelly
Tara Kulesa
Dana Macciola
Wayne Manos
Mala Mazzullo
Joanne McFadden
Mary Mulligan
Marisol Munoz
Stephanie Novara
Stephen Nussbaum
Dorothy Oddo
Robert Redmond
Theodore Roeder
Mary Sabala
Eric Sawey, Ph.D.
Diane Schubach
Richard Sever, Ph.D.
Inez Sialiano
Marcie Siconolfi
Michael Siragusa
Maria Smit, Ph.D.
Hillary Sussman, Ph.D.
Denise Weiss
Dinar Yunusov, Ph.D.

DNA Learning Center

David A. Micklos
Executive Director
Kelsie Anson, Ph.D.
Justin Burke
Megan Capobianco
Elna Carrasco
Taehwan Cha
Lauren Correri
Kelly Eames
Arden Feil
Anna Feitzinger
Cristina Fernandez Marco, Ph.D.
Lauren Graziosi
Jennifer Hackett, Ph.D.
Min Hur
Daniel Jacobs
Brittany Johnson
Jack Kellogg
Susan Lauter
Nicholas Liotta
Sandhya LoGalbo
Andrea Mahee
Allison Mayle, Ph.D.
Amanda McBrien
Valerie Meszaros
Eric Nash, Ph.D.
Sharon Pepenella, Ph.D.
Jeffry Petracca
Donna Smith
Maggie Wang
Jason Williams
Chun-hua Yang
Xiaoqun Zhang

Library and Archives

Ludmila T. Pollock
Executive Director
Paula Abisognio
Thomas Adams
Tara Bonet-Black
Jannette D'Esposito
Mary Longan
Sasha Luks-Morgan, Ph.D.
Elizabeth Pessala, JD
Stephanie Satalino
Alexander Sosnovski

Meetings and Courses

David J. Stewart, Ph.D.
Executive Director
Alfred Baman
Edward Campodonico
Joseph Carrieri
Meredith Cassuto
Betty Cavallaro
Reginald Cayo
Katiria Ceballos
Maoyen Chi, Ph.D.
Wendy Crowley
Matthew Denninger
William Dickerson
Bradley Frey
Constance Hallaran
Coleen Jackman
Patricia Klieger
Margaret Malone
Samantha Mastronardi
Dennis McGuire
Andrew Mendelsohn
Shannon Morris
Kenneth Orff
Valerie Pakaluk
Jonathan Parsons
Ira Russo
Mary Smith
Erica Stang
Kelly Tenbus
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.
SBS Professor

Cold Spring Harbor Laboratory 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
Senior Vice President, Advancement & Special Advisor to President
Diane Fagiola
Stephanie Garrett
Sarah Kitt
Karen Orzel
Jill Stone

Business Development & Technology Transfer

Andrew Whiteley
Vice President, Business Development/Technology Transfer
Vladimir Drozdoff, J.D., Ph.D.
Elizabeth Hand
Lindsey Karp, Ph.D.
Cynthia Miller
Fahema Nadira, M.D.
Radhakrishnan Narayanan, Ph.D.

Central Administration

Carol Gundersen
Carol Link
Elizabeth Panagot

Communications

Dagnia Goldschmidts
Vice President, Communications
Eliene Augenbraun, D.O., Ph.D.
Margot Bennett
Nicholas Fiore
Jasmine Lee
Philip Renna
Sara Roncero-Menendez
Susan Runkowski
Luis Sandoval, Ph.D.

Michael Skuthan
Benjamin Wigler

Culinary Services

James Hope
Director
Alan Aguirre
Jeannette Amato
Ingrid Amaya
Francis Bowdren
Maryam Brinkwart
Elmer Canales
William Carmona
Dilicia Castro Najera
Francis Cayamcela
Marvin Chavez
Melissa Davis
Erica Douglas
Saira Estrella
Bartola Garcia
Luis Garcia
Andres Gonzalez
Silvia Gonzalez
Bartholomew Isaac
Elizabeth Kirsch
Gabrielle Korwan
Juan Magana
Jose Maradiaga Gonzalez
Ann May
Frederik Munter
Adalia Pardo
Joel Pena Corvera
Jose Reyes
Susana Roman
Marlenis Romero
Carol Rupp
Lisa Spero
Jose Suarez
Selma Turk
Maria Vilorio
Alexandros Vogiatzis
Lidia Zuleta

Diversity, Equity, & Inclusion

Charla Lambert, Ph.D.
CSHL Diversity, Equity, & Inclusion Officer
Stephanie Franco

Facilities

Stephen Monez
Vice President, Chief Facilities Officer
Eliseo Amaya
Lucas Arevalo
Richard Bryant
Dessie Carter
Daniel Chapman
Rene Chavez Contreras
Anthony Ciccarella
Christopher Clark
Robert Collins
Juan Colocho
Drew Comer
Peter Dale
John Damico
Christina DeDora
Glen DiMaria
Wesley Dreusike
Paul Edwards
Elias Eifert
Rafael Estrella
Nicholas Favata
Vito Gassi
Robert Gensel
Leo Greene
Philip Grella, Jr.
Wilman Gutierrez
Jameel Hodges
Gerard Holler
Matthew Homburger
Joseph Houser
Matthew Hoyer
Christopher Hubert
Louis Hunter
Michael Hutchinson
Jeffrey Klaverweiden
Robert Kopetic, Jr.
Paulo Krizanovski
Oscar Lastra
James Lawshe
Joseph Leonard
Walter Leute
Zaharia Likoka
Jian Lin
Charles Mandell
Gregory Marotta
Patricia McAdams
Tahid McCoy
Christopher McEvoy
Kevin McHugh
Daniel Miller
Oscar Moreira
Guido Mosquera
Mary Muno
Michael Murphy
Kenneth Nastri
Herbert Neazer
Michael Nolasco
John O'Connor
Wayne Pav
Christopher Pearson II
Jose Pena Corvera
Sandro Peralta Zavala
Santos Perez Torres
John Pisciotta
Manuel Ramones
Wilson Ramones
Steven Randall
Jaqueline Reynolds
Jesus Rivera
Alfredo Rivera, Jr.
Richard Rutherford
Jeffrey Sanchez
Eric Sappell
Steven Sardone
John F. Schindler
John L. Schindler
Teri Schindler
Claudia Schmid
Lauren Schmidt
Ranjit Singh
Cesar Sisalima
Nicholas Smyth
Daniel Stahl
Raymond Stewart
Steven Tang
Tracey Anne Taylor
Hock Yew Teo
Kevin Titus
James Van Wie
Fredy Vasquez
Jose Velasquez
Trinidad Velasquez
Wilfredo Velasquez
Benjamin Veneable III
Hector Vigil Garcia
Graham Wildt
Daniel Zimmer

Finance and Accounting

Lari C. Russo
Chief Financial Officer
Elise Best
Karine Boyce
Caitlin Carroll
Damian Desiderio
Maryanne Detweiler
Colleen Eiermann
Karen Filasky
Mary Ellen Goldstein
Karen Heathwood
Winston Li
Monica Luna
Mary Mastropaolo
Linda Maurer
Daniel McClure
Alison McDermott
Denise Miller
Kristine Murphy
Nicole Povolny
Nicole Solomon
Patricia Urena

Human Resources

Katherine G. Raftery
Vice President, Chief Human Resources Officer
Andres Alarcon
Wendy Alexander
Camille Cava-Belluscio
Patricia Crawford
Sandra Dunne
Jessica Heckel
Lesley Inglis
Laura Magri
Laura Moran
Victoria Panebianco
Brian Ramsarran
Patricia Schoeffler
Bridget Shanley
Ashley Snyder

Information Technology

Douglas Torre
Vice President, Chief Information Officer
Alex Barton
Walter Benka III
Paul Braun
David Jonathan Castillo Torres
Frank Falco
Todd Heywood, Ph.D.
Thomas Karow
Thomas Keller
Leila Kelly
Sean Kelly
Matthew Lindsey
Michael Malave
Louis Malfi
Jon-Paul Maurin
Steven Northshield
Barbara Peters
Vincent Reale
Salvatore Serafino

Legal

Debra Arenare, J.D.
Vice President, General Counsel
Julie Block-Rosen, J.D.

Procurement

Michael Marchesiello
Vice President, Chief Procurement Officer
Robert Albinson
Peter Anderson
Deborah Aufiero
Christine Bernius
Victoria Bond
Kimberly Bronson
Timothy Cotone
Craig Darcy
Susan DeAngelo
Kevin Donohue

Jeffrey DuPree
Wayne Hamilton
Elizabeth Janow
Daniel Jusino
Anne Knoch
Christopher Oravitz
Jorge Ramirez
Maria Villatoro
Leslie Wenzel

Research Operations

Diane Esposito, Ph.D.
Director of Research Compliance/ Research Investigator
Priya Chembakasseril
Shanique Edwards, Ph.D.
Robert Gerdes
Susan Guterman
Grace McCrann
Joan O'Connor
Soma Prum
Denise Roberts, Ph.D.
Priya Sridevi, Ph.D.
Julie Sutherland
Michael Tramantano, Ph.D.

Sponsored Programs

Walter L. Goldschmidts, Ph.D.
Vice President, Executive Director of Sponsored Programs
Irina Armstrong
Renee Cercone-Rilo
Patryk Czyzo
Carol DuPree
Samantha Fagone
Daniel Fiorante
Donna Gavin
Jill Hemish, Ph.D.
Joanne Janelli
Josephine Marrali, J.D.
Andrea McCarthy
Cynthia McCormack
Dave Neale
Catherine Perdikoylis
Brett Roizman
Jaclyn Sammut
Gail Wagner

RESEARCH STAFF DEPARTURES DURING 2021

Clinical Fellow

Rintaro Koike, M.D.

Research Investigator and Metabolomics Core Manager

Ana Henriques da Costa, Ph.D.

Research Investigators

Carla Gauss, Ph.D.
Dennis Plenker, Ph.D.
Katharina Schmack, M.D., Ph.D.
Sze Wong, Ph.D.

Visiting Scientists/ Collaborative Scientists

Xu An, Ph.D.
Gary Goldberg, M.D.
Dhananjay Huilgol, Ph.D.
Angela Ianniciello, Ph.D.
Yi Li, Ph.D.
Hemanth Mohan, Ph.D.
Shreyas Mysore Suryanarayana, Ph.D.
Saikat Nandi, Ph.D.
Klaske Schukken, Ph.D.
Raffaella Sordella, Ph.D.
Shyam Srinivasan, Ph.D.
Sweta Srivas, Ph.D.
Julian Taranda, Ph.D.
Bor-Shuen Wang, Ph.D.
Shira Yomtoubian, Ph.D.
Na Zhou, Ph.D.

Visiting Clinicians

David Bonda, M.D.
Ariel Kredentser, M.D.

Research Compliance Specialist

Laura Lynn

Senior Fellows

Christian Bravo-Rivera, Ph.D.
Mohammad Rahman, Ph.D.

Postdoctoral Fellows

Oliver Artz, Ph.D.
Matthias Benoit, Ph.D.
Stephanie Yvonn Blockhuys, M.D., Ph.D.
Xiaoyin Chen, Ph.D.
Hanfei Deng, Ph.D.
Sarah Fumagalli, Ph.D.
Olivier Gschwend, Ph.D.
Longwen Huang, Ph.D.
Yuhan Huang, Ph.D.
Stellar Hur, Ph.D.
Alessandra Inguscio, Ph.D.
Gukhan Kim, Ph.D.
Lei Liu, Ph.D.
Jason Lynn, Ph.D.
Grinu Mathew, Ph.D.
Haiwei Mou, Ph.D.
Farhana Nasrin, Ph.D.
Sung Rye Park, Ph.D.
Nitin Pawar, Ph.D.
Donglai Shen, Ph.D.
Atsushi Shimada, Ph.D.
James Sturgill, Ph.D.
Bo Wang, Ph.D.
Minghui Wang, Ph.D.
Xingang Wang, Ph.D.
Xinyi Wang, Ph.D.
Zhikai Wang, Ph.D.
Xiao, Xiong, Ph.D.
Yunyao Xie, Ph.D.
Zhaolin Yang, Ph.D.
Juannan Zhou, Ph.D.

Computational Scientists

Daniel Blibaum
Nathan Fox
Ying Jin, Ph.D.
Benjamin Kaminow
Jude Kendall
Forrest Li
Fahimeh Mirhaj
Ritika Ramani
Ammar Tareen
Liya Wang, Ph.D.
Matthew Wroten, Ph.D.
Lilian Zhou

Research Associates

Habeeb Al Sudani
Ramesh Palaniswamy
Linda Rodgers