2023
ANNUAL REPORT

CSH | Cold Spring Harbor Laboratory
SIMONS CENTER FOR QUANTITATIVE BIOLOGY

2023 SCQB ANNUAL REPORT

ADVISORY BOARD

ANDREW G. CLARK, PH.D.
Professor of Molecular Biology and Genetics
Cornell University

DAVID L. DONOHO, PH.D.
Anne T. and Robert M. Bass Professor of
Humanities
Professor of Statistics
Stanford University

MOLLY PRZEWORSKI, PH.D.
Professor of Biological Sciences and Systems
Biology
Columbia University

ERIC D. SIGGIA, PH.D.
Viola Ward Brinning and Elbert Calhoun Professor
Head of Laboratory of Theoretical Condensed
Matter Physics
The Rockefeller University

EERO P. SIMONCELLI, PH.D.
Silver Professor of Neuroscience, Mathematics,
Data Science, and Psychology
New York University
Scientific Director, Center for Neuroscience
Flatiron Institute, Simons Foundation

STEVEN L. SALZBERG, PH.D. (CHAIR)
Bloomberg Distinguished Professor of Biomedical
Engineering, Computer Science, and Biostatistics
Director, Center for Computational Biology
Johns Hopkins University

Cover Image: Image generated using DALL-E3, a new AI text-to-image system powered by Open AI.

TABLE OF CONTENTS

CORE FACULTY

Ivan Iossifov

Justin Kinney

Peter Koo

Alexander Krasnitz

Dan Levy

David McCandlish

Hannah Meyer

Saket Navlakha

Adam Siepel

FROM THE CHAIR

As I look back on my 9 1/2 years at CSHL, it seems that 2023 was the year when the SCQB finally hit its stride. It wasn't until the end of last year that we completely shook free of constraints of the pandemic and began to live and breathe normally again. And the years prior to the pandemic had been marked by intensive recruitment and growth. 2023 was the year that we finally settled into more quietly productive scientific work. Our faculty, now nine in number, has matured, with Ivan Iossifov promoted this year to Professor and former Fellow Hannah Meyer to Assistant Professor (p. 3). Several research groups have grown considerably—the Koo lab now lists twelve members and the Meyer lab lists nine— and a slew of new postdocs and students have recently arrived (p. 3). We are now in a steady rhythm with regular activities such as teaching our QB course and bootcamp (p. 8) and organizing the ProbGen meeting (p. 10).

The usual measures of research productivity are grants and scientific papers, and 2023 was a solid year by these metrics. Our grant productivity was down somewhat from its high of over $10M last year, but we still brought in a respectable $6.5 M in new grants, including four competitive awards from the National Institutes of Health (p. 12). We published 32 papers, exactly the same as last year, and down only slightly from our high during the pandemic. Among them were several exciting results in deep learning (p. 4), autism research (pp. 5-6), plant genetics (p. 6), and comparative genomics (p. 7). Our graduate students were particularly productive this year, playing leading roles in several major papers (p. 9).

To me, however, these numbers are less important than the priceless sense of community and support the group here has provided. I personally had a year unlike any other, coming uncomfortably close in May to dying on the pavement in Huntington when I was struck by a pickup truck while bicycling. My injuries were severe, and even after an emergency surgery at Huntington Hospital saved my life, the doctors could not say whether or not I would walk again. But after months in hospitals, rehab, and physical therapy, I am happy to report that I am now walking almost normally, and steadily improving in strength and balance. The SCQB community was wonderfully supportive throughout this ordeal, through hospital visits, calls and emails, and several members stepped forward to take up the slack created by my absence. I would especially like to thank the members of my laboratory and my assistant Susan Fredericks for their warmth, dedication, and support.

Next year will be my last as Chair of the SCQB. I will step down at the end of 2024, and will place leadership of the Center in the capable hands of my colleague Justin Kinney. Justin has been at the SCQB for most of his career, starting as a postdoc in 2009, and no one has shown more dedication to our efforts than he has. In addition to excelling as a researcher, Justin has been exceptionally active in teaching and support of graduate students, and he has collaborated broadly across the CSHL campus. After a decade under my leadership, the Center will surely benefit from his new ideas, energy, enthusiasm, and talent.

Best Wishes for the New Year,

Adam Siepel. Ph.D., Chair
December 22, 2023

OUR MEMBERS

The **Simons Center for Quantitative Biology (SCQB)** at Cold Spring Harbor Laboratory (CSHL) now includes 60 members, ranging from its faculty to postdocs, graduate students, and research staff. This year marked a season of recognition and transition within the community. **Ivan Iossifov** was promoted to Professor, while **Hannah Meyer,** formerly a CSHL Fellow, was invited to join the faculty as an Assistant Professor. Meanwhile, the community bid farewell to former Associate Professor Molly Gale-Hammell, who has embarked on an exciting new journey at NYU's Grossman School of Medicine.

**Professor Ivan Iossifov and Assistant Professor Hannah Meyer.**

60

Total Members

The year also witnessed a major influx of new talent at the Center, as a large group of young scientists arrived to begin postdoctoral work in SCQB laboratories. Among these additions were John Desmarais, Kaisor Loell, Jessica Zhou, Anirban Sakar, Alessandro Crnjar, Mengyi Sun, Yong Lin, Rish Prabaker, Amitava Banerjee, and Bahruz Jabiyevi, each contributing unique perspectives and expertise to enrich the community.

Additionally, three graduate students—Kaeli Rizzo (Koo lab), Yijie Kang (Koo lab), and Stephen Stalinski (Siepel lab)—found homes for their thesis work within SCQB labs this year.
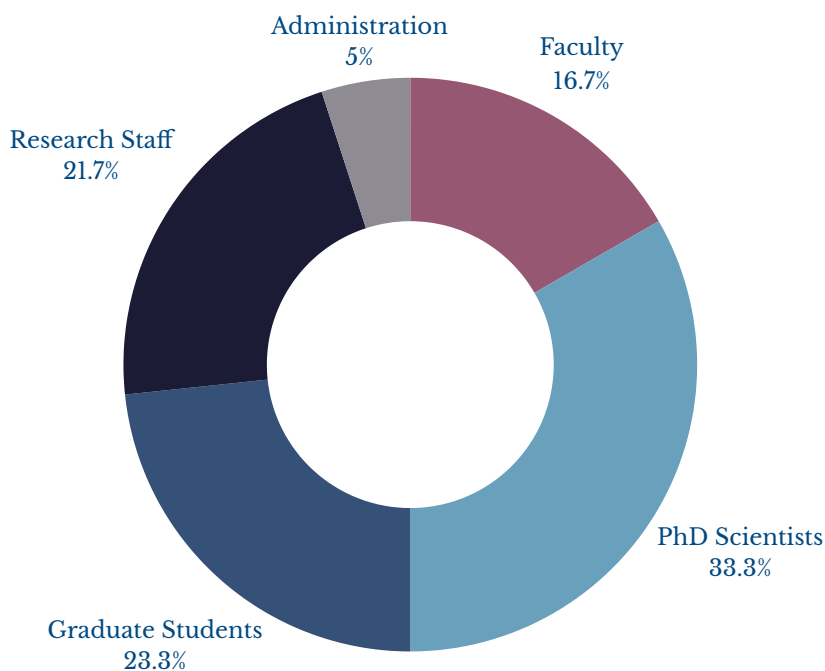
Postdocs Jessica Zhou, Mengyi Sun, and Amitava Banerjee

Graduate Students Kaeli Rizzo, Yijie Kang, and Stephen Stalinski

Researchers at the **SCQB** have continued their pioneering work at the intersection of mathematics, computation, and biology. The Center's major research themes—**Gene Regulation**, **Evolutionary Genomics**, **Genomic Disease Research**, and **Genomic Technology**—have not changed for several years, but within these themes there has been increased emphasis on artificial intelligence, big data, and evolutionary theory. This year SCQB researchers made significant strides in uncovering hidden genomic regulatory interactions, exploring genetic evolution across diverse contexts, and delving into the genetic factors underlying complex diseases like autism and cancer. A few projects are highlighted here.

Unlocking the mysteries of deep learning in genomics

In the field of genomics, deep neural networks (DNNs) have shown promise for deciphering DNA sequences and predicting regulatory functions. Understanding exactly what these powerful networks do, however, remains extremely challenging. In a new study, Assistant Professor **Peter Koo**'s lab has unveiled a new statistical technique to enhance the interpretability of DNNs in genomics. DNNs excel in genomic prediction tasks, where they are used to analyze DNA sequences and uncover regulatory insights. Special attribution methods assign significance scores to nucleotides to illuminate how deep neural networks (DNNs) function internally. These scores help reveal important patterns in DNA, such as binding motifs for transcription factors, and measure the impact of single-nucleotide mutations. But these methods often produce noisy maps, making it difficult to detecting crucial patterns. In their study, Assistant Professor Peter Koo's lab identified a previously overlooked source of noise in attribution maps, which arises from the particular way in which deep neural networks encode DNA sequences.

Koo's team found a clever but straightforward solution for this noise: a precise statistical correction involving just one line of computer code. Despite its simplicity, their correction results in dramatic improvements in the gradient-based attribution analysis used to make sense of DNNs. Its benefits apply not only to DNA sequences but also to protein and RNA sequences, making it a highly versatile tool for computational biology. Work continues on the critical problem of interpreting DNNs, by the Koo lab and many others around the world, but this new innovation is one important step forward, something that consistently improves the gradient-based attribution maps that are essential components of the toolkits of genomics researchers.

Majdandzic A, Rajesh C, Koo PK. Correcting gradient-based interpretations of deep neural networks for genomics. *Genome Biol*. 2023;24: 109.
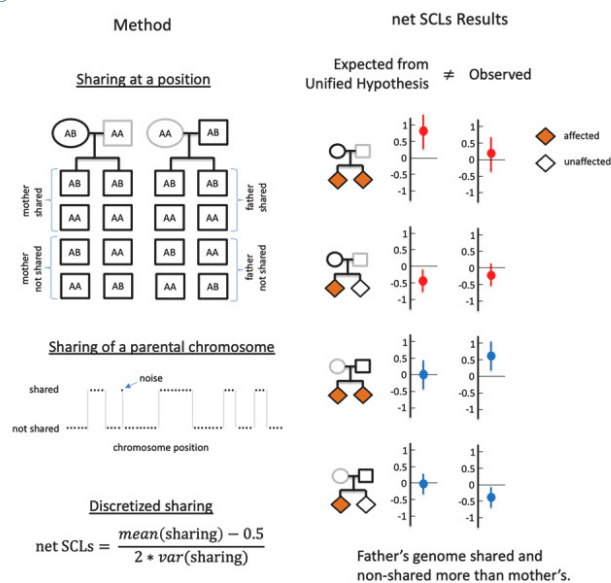
Unveiling autism genetics: Siblings inherit more from fathers than mothers

A recent study conducted by Professor **Ivan Iossifov** and Associate Professor **Dan Levy,** published in Cell Genomics, has unveiled intriguing revelations about the genetic basis of Autism Spectrum Disorder (ASD). In this extensive study, the researchers explored the genetic factors contributing to ASD incidence by investigating thousands of families, ultimately challenging established models and shedding light on the roles of parental genomes. ASD has long been recognized as a neurodevelopmental disorder with a strong genetic component. Factors such as high concordance rates in twins, known causal loci, and increased instances of deleterious de novo mutations in affected individuals all pointed to a genetic basis for the condition.

However, the study conducted by Iossifov and Levy sought to delve deeper into the specifics of this genetic contribution.

The study showed that siblings concordant for ASD—meaning they both had the condition—shared more of their parental genomes than would be expected by chance. Consistently, siblings discordant for autism—where one had the condition and the other did not—shared less of their parental genomes than expected. The increased sharing in concordant siblings was particularly pronounced from the father's side, with a highly significant p-value of 0.0014. However, after the necessary adjustments to account for differences in paternal and maternal meiotic recombination rates, a direct comparison of excess sharing of the maternal and paternal genomes in concordant siblings showed no statistically significant difference. Yet, this finding contradicts earlier models that predicted a higher sharing of the maternal genome, a function of female resistance to ASD. Iossifov and Levy's study builds upon the unified hypothesis, which previously explained various aspects of ASD incidence but struggled to account for the increased sharing of the father's genome. The research suggests that a more complex and flexible genetic model may be necessary to explain the observed patterns fully. As research in this field continues, we can anticipate further insights into the complex genetic landscape of ASD, potentially paving the way for enhanced diagnostic and therapeutic approaches.

Method — net SCLs Results

Sharing at a position

mother shared / mother not shared / father shared / father not shared

Sharing of a parental chromosome

shared / not shared / noise / chromosome position

Discretized sharing

$$net\ SCLs = \frac{mean(sharing) - 0.5}{2 * var(sharing)}$$

Expected from Unified Hypothesis ≠ Observed

◆ affected   ◇ unaffected

Father's genome shared and non-shared more than mother's.

Graphical Abstract

Deciphering gene epistasis in tomato fruit size regulation

A recent study co-led by Associate Professor **David McCandlish** in collaboration with HHMI Investigator Zachary Lippman has delved into the intricate world of gene interactions, known as epistasis, shedding new light on its role in controlling tomato fruit size. While epistasis analysis has traditionally focused on mutations with notable effects on protein function to understand gene relationships, the majority of natural genetic variation is due to differences in the expression of genes. Despite this, little is known about the role of epistatic interactions in the regulation of gene expression. This study takes a novel approach by examining the impact of natural and engineered cis-regulatory alleles within a plant stem-cell circuit, revealing an under-explored dimension of gene interaction that challenges our existing understanding. The research explores epistasis in regulatory mutations, which often have weaker effects on gene function by altering gene expression levels or timing.
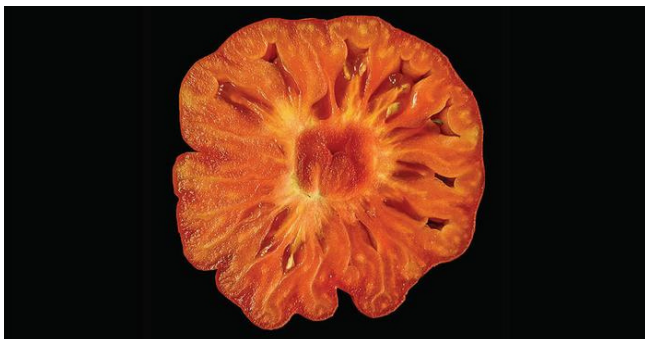
The study focuses on the *CLAVATA-WUSCHEL* (*CLV-WUS*) gene regulatory circuit in tomato plants, specifically the *CLAVATA3* (*CLV3*) gene responsible for encoding a signaling peptide that regulates stem-cell proliferation. This peptide represses the WUS gene, creating an intricate feedback loop. By leveraging CRISPR technology, the researchers engineered mutations in the *SlCLV3* gene, known to increase fruit size. They combined these mutations with others in genes that interact with *SlCLV3*, resulting in 46 tomato strains with different combinations of mutations. While some mutations produced predictable changes in tomato size, others yielded random outcomes. Surprisingly, the study revealed that two mutations central to tomato domestication, which arose millennia ago, had the most beneficial effect on fruit size when combined. This research challenges the predictability of gene interactions, especially within cis-regulatory alleles, and emphasizes the need to consider cis-regulatory variation in epistasis analysis, particularly in complex biological systems. These findings provide valuable insights into the dynamic role of epistasis in shaping traits during evolution and open new avenues for manipulating gene regulatory networks in agriculture and human health.

Aguirre L, Hendelman A, Hutton SF, McCandlish DM, Lippman ZB. Idiosyncratic and dose-dependent epistasis drives variation in tomato fruit size. *Science*. 2023 Oct 20;382(6668):315-320.

Revealing the genomic secrets of bats: insights into immunity, evolution, and health

In a recent publication in *Genome Biology*, members of Professor Adam Siepel's lab collaborated with the McCombie lab at CSHL to investigate the genomics of bats, providing valuable insights into their extraordinary features, such as their powered flight, extended lifespans and resilient immune systems. The study revealed a unique feature in the bat immune system – a significant reduction in interferon (IFN) genes, challenging previous assumptions about bat immune responses. Contrary to the belief that bats constitutively express IFN-α, three bat species were found to have completely lost these genes, shifting their reliance to IFN-ω. This shift may contribute to bats' ability to tolerate viral infections, making them common virus reservoirs with potential implications for human health.

Furthermore, the study identified rapid evolution in antiviral genes stimulated by type I IFNs, including duplications and positive selection in IFN-induced transmembrane genes. Additionally, several tumor suppressor and DNA-repair genes showed signs of positive selection, potentially contributing to bats' longevity and lower cancer rates. The research highlights the intricate interplay between bat-wide and lineage-specific evolution in the immune gene repertoire, suggesting diverse strategies for battling diseases. These findings underscore the importance of bats as a model for studying genotypic-phenotypic relationships, with implications for understanding zoonotic outbreaks, cancer, and immunity. The study's release of high-quality genomic resources for bats is expected to catalyze further research into addressing pressing challenges in human health, including infectious diseases and cancer.

Scheben A, Mendivil Ramos O, Kramer M, Goodwin S, Oppenheim S, Becker DJ, Schatz MC, Simmons NB, Siepel A, McCombie WR. Long-Read Sequencing Reveals Rapid Evolution of Immunity- and Cancer-Related Genes in Bats. *Genome Biol Evol.* 2023 Sep 4;15(9):evad148.
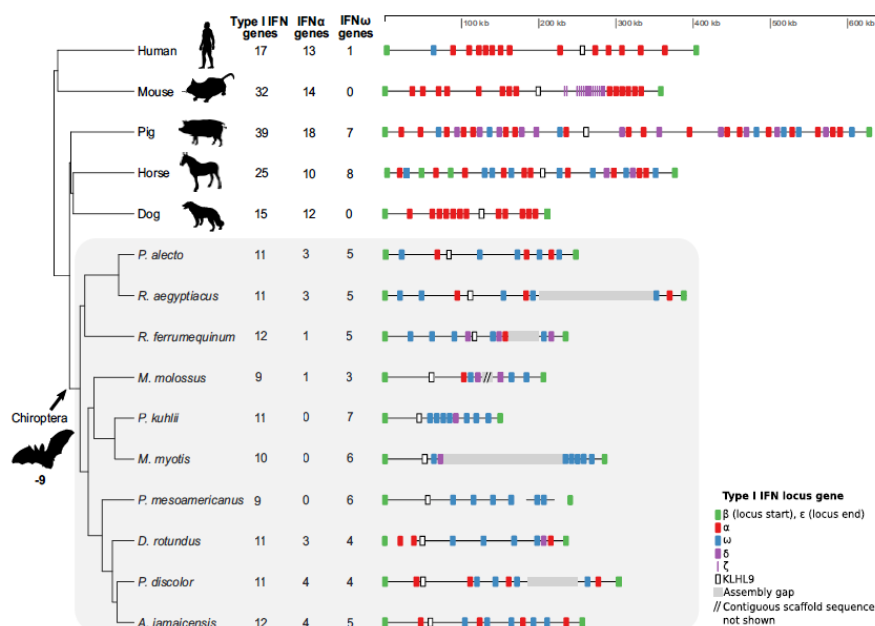
Comparison of the type I interferon (IFN) gene locus in bats with other mammals.

Ph.D. Training Program in Quantitative Biology

Together with their focus on research, the SCQB faculty maintain a strong commitment to education. Our Ph.D. Training Program in Quantitative Biology is a cornerstone of this commitment, aiming to foster a new generation of scientists who can bridge experimental and quantitative biology.

Program Success:

The community of graduate students in the SCQB continues to grow and thrive. There are now **13** dedicated graduate students actively enrolled and engaged in QB labs, helping to drive cutting-edge research and innovation. Since 2018, **10** Ph.D. students have graduated from QB labs, taking the skills and knowledge they acquired at the SCQB to many corners of the world.

Cold Spring Harbor School of Biological Science 2023 Doctoral Recipient

Program Highlights:

- **QB Bootcamp:** A hands-on 2.5-day QB Bootcamp, where incoming students are introduced to Python programming and high-performance computing. It serves as a solid foundation for their quantitative exploration.

- **QB Course:** A team of experienced QB faculty leads our 22-lecture QB course, covering a spectrum of vital topics in quantitative biology, from statistics to machine learning, algorithms, evolution, genomics, and biophysics.

- **Advanced Coursework:** Students are encouraged to participate in the SCQB Mentored Independent Study Program. Here, under the guidance of CSHL faculty, students take part in personalized curricula of directed reading in quantitative material, gaining expertise in contemporary quantitative methods.

COLE WUNDERLICH

Dr. Wunderlich conducted his research in Dr. Molly Gale-Hammell's laboratory. His thesis was entitled "Quantifying transposable element expression in single-cell RNA sequencing data."

Cutting-Edge Research by SCQB Graduate Students in 2023

Several graduate students from SCQB labs showcased their contributions to genomics and cancer research this year. **Shushan Toneyan**, working under Peter Koo's guidance, uncovered the critical ETV6 dependency in Ewing sarcoma, offering a promising strategy to combat this pediatric cancer. Additionally, Toneyan was a key team member who introduced EvoAug, a new method to enhance the performance and interpretability of genomic deep neural networks.

Shushan Toneyan and Peter Koo.

Member of the Siepel Lab from left to right: Xander Xue, Armin Scheben, Adam Siepel, Yixin Zhao, Lingjie Liu, and Ziyi Mo.

In Adam Siepel's lab, **Ziyi Mo** tackled the challenge of simulation mis-specification in population genetics, employing domain-adaptive neural networks to improve evolutionary analyses. Another student in the Siepel lab, **Lingjie Liu**, contributed to a mathematical model characterizing transcription initiation and promoter-proximal pausing dynamics, shedding light on critical gene expression processes. Together, these accomplishments underscore the impressive research endeavors of SCQB graduate students in 2023.

Citations:

Gao Y, He XY, Wu XS, Huang YH, **Toneyan S**, Ha T, Ipsaro JJ, Koo PK, Joshua-Tor L, Bailey KM, Egeblad M, Vakoc CR. ETV6 dependency in Ewing sarcoma by antagonism of EWS-FLI1-mediated enhancer activation. *Nat Cell Biol.* 2023;25: 298–308.

Lee NK, Tang Z, **Toneyan S**, Koo PK. EvoAug: improving generalization and interpretability of genomic deep neural networks with evolution-inspired data augmentations. *Genome Biol.* 2023;24: 105.

Mo Z, Siepel A. Domain-adaptive neural networks improve supervised machine learning based on simulated population genetic data. *PLoS Genet.* 2023 Nov 7;19(11):e1011032.

Zhao Y, **Liu L**, Hassett R, Siepel A. Model-based characterization of the equilibrium dynamics of transcription initiation and promoter-proximal pausing in human cells. *Nucleic Acids Res.* 2023.

Every year, the SCQB sponsors or organizes one or more scientific meetings to encourage collaboration and communication of ideas. This year we took the lead in organizing the international Probabilistic Modeling in Genomics, or **ProbGen**, meeting, held at CSHL in March. During this event, CSHL graduate student **Ziyi Mo** delivered a well-received talk entitled "Domain adaptive neural networks improve robustness of supervised machine learning models based on simulated population genetic data."

Members of the SCQB also continue to be involved in the **New York Genome Center (NYGC).** Professor Ivan Iossifov maintains a joint appointment at NYGC, focusing on autism sequencing projects. Professor Siepel co-leads the Population Genomics Working Group at the NYGC.

In addition, SCQB members increasingly contribute to the **CSHL Cancer Center.** This year, Associate Professor Justin Kinney became co-Program Leader for Gene Regulation and Inheritance in the Cancer Center.

Cover Art by Postdoc Armin Scheben: Probabilistic Modeling in Genomics Meeting, March 2-11, 2023.

Major Invited Talks

- **Associate Professor Justin Kinney** gave a talk entitled "Specificity, synergy, and mechanisms of splice-modifying drugs " at the ReviR Therapeutics Seminar Series, Brisbane, CA, January 2023.

- **Associate Professor Saket Navlakha** gave a talk entitled "Algorithms and data structures in the fruit fly brain" at the IIT Jodhpur, Next-Gen AI: Inspiration from Brain Science Conference. Jodhpur, India [Virtual], January 2023.

- **Associate Professor Justin Kinney** gave a talk entitled "Massively parallel reporter assays, quantitative sequence-function relationships, and splice modifying drugs" at the Department of Systems Biology Seminar Series, University of Massachusetts Chan Medical School, Worcester, MA, March 2023

- **Associate Professor Justin Kinney** gave a talk entitled "Learning biophysical mechanisms of gene regulation from massively parallel reporter assays" at the 14th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, Houston, TX, September 2023.

- **Assistant Professor Peter Koo** gave a talk entitled "Interpreting cis-regulatory mechanisms from genomic deep learning with surrogate models" at the Computational Regulatory Genomics, KIPOI Summit, Germany, September 2023.

- **Assistant Professor Peter Koo** gave a talk entitled "Interpreting cis-regulatory mechanisms learned by genomic deep learning" at the Intelligence: Natural, Artificial, and Synthetic, Annual Symposium of Barcelona Collaboratorium, CRG Barcelona, Spain, October 2023.

Remembering Yifei Huang
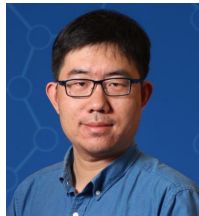
Yifei Huang, a brilliant computational biologist, left an indelible mark on the scientific community. Diagnosed with lung cancer in March 2022 at age 39, he sadly passed away in October 2023. Yifei earned his Ph.D. under Brian Golding at McMaster University in Canada and conducted postdoctoral research in **Adam Siepel**'s lab. He was a kind, supportive, and engaging colleague, always present at seminars, journal clubs, and social events, and a popular teaching assistant for the QB course. When he was diagnosed with cancer, Yifei had recently begun as an Assistant Professor in Biology at Penn State University, and had already been awarded a prestigious NIH R35 grant.

Yifei had a profound and lasting impact on research in the Siepel lab. He was a fountain of creative ideas, and either led or made key contributions to the development of several widely used tools, including *LINSIGHT, FitCons2,* and *LASSIE.* Several years after he left CSHL, work continues on ideas he suggested. One project he began, on patterns of "ultraselection" in the human genome, was only recently published. His work also led to new collaborations in which his methods have been applied to agriculturally important plants. Yifei will be remembered fondly for both his professional and his personal influence on all of us.

Event Highlights

Obeying the adage about "all work and no play," members of the SCQB do sometimes manage to leave their computer screens and enjoy life on the CSHL campus. This year, **Peter Koo** and **Adam Siepel**'s laboratories both participated in the eighth annual **CSHL Raft Race**, where an impressive 14 teams from across the campus demonstrated their ingenuity by crafting vessels from an array of materials, including yoga mats, pool noodles, and water jugs. The winner would secure a coveted spot on the champions' list, commemorated on the oar displayed in Blackford Hall. While victory eluded both teams, the Koo lab earned an honorable mention for their imaginative team name, "Koo's Booze Cruise: 2 Boozed 2 Lose," paying tribute to last year's champions. Meanwhile, at the **CSHL Volleyball Tournament**, **Hannah Meyer** led her team to triumph, securing their second Tiernan Cup in three years.
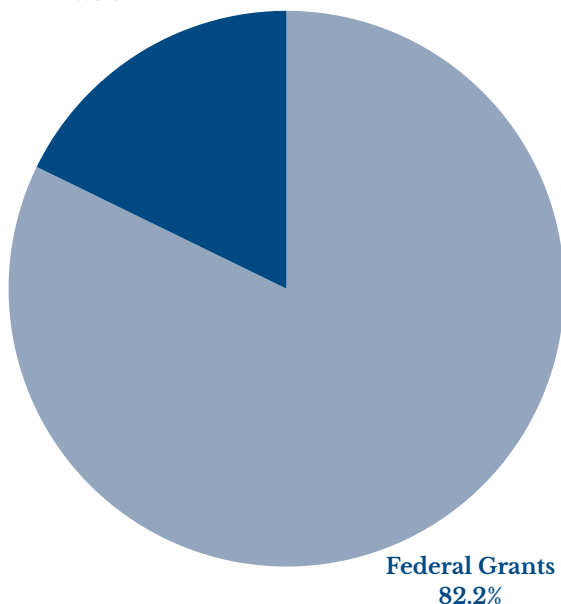
Members of the SCQB secured four new federal grants totaling **$5,304,903** in 2023. as well as five new private and foundation grants, providing an additional **$1,147,040** in support.

9
GRANTS

$6,451,943*
AWARDED

*Direct + Indirect Costs

Private/Foundation Grants
17.8%

Federal Grants
82.2%

2023 FUNDING BREAKDOWN

Top Funded Grants

Grant Title: Evolutionary human genomics: demography, natural selection, and transcriptional regulation

- **PI: Adam Siepel**
- **Funding: $2,951,205**

Grant Title: Reliable post hoc interpretations of deep learning in genomics

- **PI: Peter Koo**
- **Funding: $1,535,000**

APPENDIX

1. Aguirre L, Hendelman A, Hutton SF, **McCandlish DM**, Lippman ZB. Idiosyncratic and dose-dependent epistasis drives variation in tomato fruit size. *Science*. 2023;382: 315–320.

2. Avizemer Z, Martí-Gómez C, Hoch SY, **McCandlish DM**, Fleishman SJ. Evolutionary paths that link orthogonal pairs of binding proteins. *Res Sq.* 2023. doi:10.21203/rs.3.rs-2836905/v1

3. Belleau P, Deschênes A, Tuveson DA, **Krasnitz A**. RAIDS. *Bioconductor*; 2023. doi:10.18129/B9.BIOC.RAIDS

4. Berube B, Ernst E, Cahn J, Roche B, de Santis Alves C, Lynn J, Scheben A, **Siepel A**, Ross-Ibarra J, Kermicle J, Martienssen RA. Teosinte Pollen Drive guides maize diversification and domestication by RNAi. bioRxiv. 2023. doi:10.1101/2023.07.12.548689. In revision for *Nature*.

5. Cano AV, Gitschlag BL, Rozhoňová H, Stoltzfus A, **McCandlish DM**, Payne JL. Mutation bias and the predictability of evolution. *Philos Trans R Soc Lond B Biol Sci.* 2023;378: 20220055.

6. Einson J, Glinos D, Boerwinkle E, Castaldi P, Darbar D, de Andrade M, Ellinor P, Fornage M, Gabriel S, Germer S, Gibbs R, Hersh CP, Johnsen J, Kaplan R, Konkle BA, Kooperberg C, Nassir R, Loos RJF, Meyers DA, Mitchell BD, Psaty B, Vasan RS, Rich SS, Rienstra M, Rotter JI, Saferali A, Shoemaker MB, Silverman E, Smith AV; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium; Mohammadi P, Castel SE, **Iossifov I,** Lappalainen T. Genetic control of mRNA splicing as a potential mechanism for incomplete penetrance of rare coding variants. *Genetics*. 2023;224.

7. Ferrer M, Anthony TG, Ayres JS, Biffi G, Brown JC, Caan BJ, Cespedes Feliciano EM, Coll AP, Dunne RF, Goncalves MD, Grethlein J, Heymsfield SB, Hui S, Jamal-Hanjani M, Lam JM, Lewis DY, **McCandlish D**, Mustian KM, O'Rahilly S, Perrimon N, White EP, Janowitz T. Cachexia: A systemic consequence of progressive, unresolved disease. *Cell*. 2023;186: 1824–1845.

8. Gao Y, He XY, Wu XS, Huang YH, Toneyan S, Ha T, Ipsaro JJ, **Koo PK**, Joshua-Tor L, Bailey KM, Egeblad M, Vakoc CR. ETV6 dependency in Ewing sarcoma by antagonism of EWS-FLI1-mediated enhancer activation. *Nat Cell Biol*. 2023;25: 298–308.

9. Gitschlag BL, Cano AV, Payne JL, **McCandlish DM**, Stoltzfus A. Mutation and Selection Induce Correlations between Selection Coefficients and Mutation Rates. *Am Nat*. 2023;202: 534–557.

10	Ishigami Y, Wong MS, Martí-Gómez C, Ayaz A, Kooshkbaghi M, Hanson S, **McCandlish DM**, Krainer AR, **Kinney JB**. Specificity, synergy, and mechanisms of splice-modifying drugs. *bioRxiv*. 2022. doi:10.1101/2022.12.30.522303
11	Kaczmarzyk JR, Gupta R, Kurc TM, Abousamra S, Saltz JH, **Koo PK**. ChampKit: A framework for rapid evaluation of deep neural networks for patch-based histopathology classification. *Comput Methods Programs Biomed*. 2023;239: 107631.
12	Kleeman SO, Thakir TM, Demestichas B, Mourikis N, Loiero D, Ferrer M, Bankier S, Riazat-Kesh YJRA, Lee H, Chantzichristos D, Regan C, Preall J, Sinha S, Rosin N, Yipp B, de Almeida LGN, Biernaskie J, Dufour A, Tober-Lau P, Ruusalepp A, Bjorkegren JLM, Ralser M, Kurth F, Demichev V, Heywood T, Gao Q, Johannsson G, Koelzer VH, Walker BR, **Meyer HV,** Janowitz T. Cystatin C is glucocorticoid responsive, directs recruitment of Trem2+ macrophages, and predicts failure of cancer immunotherapy. *Cell Genom*. 2023;3: 100347.
13	**Koo PK**, Ploenzke M, Anand P, Paul S, Majdandzic A. ResidualBind: Uncovering Sequence-Structure Preferences of RNA-Binding Proteins with Deep Neural Networks. *Methods Mol Biol*. 2023;2586: 197–215.
14	Lee NK, Tang Z, Toneyan S, **Koo PK**. EvoAug: improving generalization and interpretability of genomic deep neural networks with evolution-inspired data augmentations. *Genome Biol*. 2023;24: 105.
15	Li S, Alexander J, Kendall J, Andrews P, Rose E, Orjuela H, Park S, Podszus C, Shanely L, Ma R, Ranade N, Ronemus M, Rishi A, Donoho DL, Goldberg GL, **Levy D**, Wigler M. High-throughput single-nucleus hybrid sequencing reveals genome-transcriptome correlations in cancer. bioRxiv. 2023. p. 2023.10.04.560973. doi:10.1101/2023.10.04.560973
16	Liu L, Zhao Y, **Siepel A.** DNA-sequence and epigenomic determinants of local rates of transcription elongation. *bioRxiv* 2023. doi: 10.1101/2023.12.21.572932
17	Majdandzic A, Rajesh C, **Koo PK**. Correcting gradient-based interpretations of deep neural networks for genomics. *Genome Biol*. 2023;24: 109.
18	Moffitt AB, Alexander J, Stepansky A, Famulare C, Lopez CA, Hakim N, Wu V, Wang Z, Kolitz JE, Chiorazzi N, Allen SL, Levine RL, **Levy D**, Wigler M. Whole Genome Sequencing Comparison of Acute Myeloid Leukemia at Presentation and Remission Predicts Patient Outcome. *Blood*. 2022;140: 9167–9169.

19	Mo Z, **Siepel A**. Domain-adaptive neural networks improve supervised machine learning based on simulated population genetic data. *PLoS Genet*. 2023 Nov 7;19(11):e1011032.
20	Mou H, Eskiocak O, Özler KA, Gorman M, Yue J, Jin Y, Wang Z, Gao Y, Janowitz T, **Meyer HV,** Yu T, Wilkinson JE, Kucukural A, Ozata DM, Beyaz S. CRISPR-induced exon skipping of β-catenin reveals tumorigenic mutants driving distinct subtypes of liver cancer. *J Pathol*. 2023;259: 415–427.
21	Ortiz JR, Lewis SM, Ciccone MF, Chatterjee D, Henry S, **Siepel A**, et al. Single-cell transcription mapping of murine and human mammary organoids responses to female hormones. *bioRxiv*. 2023. doi:10.1101/2023.09.28.559971
22	Petti S, Bhattacharya N, Rao R, Dauparas J, Thomas N, Zhou J, Rush AM, **Koo P,** Ovchinnikov S. End-to-end learning of multiple sequence alignments with differentiable Smith-Waterman. *Bioinformatics*. 2023;39.
23	Scheben A, Mendivil Ramos O, Kramer M, Goodwin S, Oppenheim S, Becker DJ, Schatz MC, Simmons NB, **Siepel A**, McCombie WR. Long-Read Sequencing Reveals Rapid Evolution of Immunity- and Cancer-Related Genes in Bats. *Genome Biol Evol*. 2023;15.
24	Seitz EE, **McCandlish DM**, **Kinney JB**, **Koo PK**. Interpreting cis-regulatory mechanisms from genomic deep neural networks using surrogate models. *bioRxiv*. 2023. doi:10.1101/2023.11.14.567120
25	Shen Y, Dasgupta S, **Navlakha S**. Reducing Catastrophic Forgetting With Associative Learning: A Lesson From Fruit Flies. *Neural Comput*. 2023;35: 1797–1819.
26	Shi X, Lu C, Corman A, Nikish A, Zhou Y, Platt RJ, **Iossifov I,** Zhang F, Pan JQ, Sanjana NE. Heterozygous deletion of the autism-associated gene CHD8 impairs synaptic function through widespread changes in gene expression and chromatin compaction. *Am J Hum Genet*. 2023;110: 1750–1768.
27	Srinivasan S, Daste S, Modi MN, Turner GC, Fleischmann A, **Navlakha S.** Effects of stochastic coding on olfactory discrimination in flies and mice. *PLoS Biol*. 2023;21: e3002206.
28	Tang Z, Toneyan S, **Koo PK**. Current approaches to genomic deep learning struggle to fully capture human genetic variation. *Nat. Genet., News & Views*, 2023.

29	Toneyan S, **Koo PK**. Interpreting Cis-Regulatory Interactions from Large-Scale Deep Neural Networks for Genomics. *bioRxiv*. 2023. doi:10.1101/2023.07.03.547592
30	Weinstein JY, Martí-Gómez C, Lipsh-Sokolik R, Hoch SY, Liebermann D, Nevo R, Weissman H, Petrovich-Kopitman E, Margulies D, Ivankov D, **McCandlish DM**, Fleishman SJ. Designed active-site library reveals thousands of functional GFP variants. *Nat Commun*. 2023;14: 2890.
31	Wroten M, Yoon S, Andrews P, Yamrom B, Ronemus M, Buja A, Krieger AM, Levy D, Ye K, Wigler M, **Iossifov I**. Sharing parental genomes by siblings concordant or discordant for autism. *Cell Genom*. 2023 May 8;3(6):100319.
32	Zhao Y, Liu L, Hassett R, **Siepel A**. Model-based characterization of the equilibrium dynamics of transcription initiation and promoter-proximal pausing in human cells. *Nucleic Acids Res*. 2023.

NEW EXTERNAL FUNDING IN 2023

PI	TOTAL AWARD[1]	YEARS	FUNDING SOURCE	PROJECT TITLE
Koo, P.	$1,536,000	2023-2027	National Institutes of Health	Reliable post hoc interpretations of deep learning in genomics
dos Santos, C. **Koo, P.**[2]	$352,324	2023-2028	National Institutes of Health	Blockade of cMYC oncogenic function by pregnancy-induced alterations and remodeling of the mammary gland
Koo, P.	$219,180	2023-2024	CSHL-Northwell Health	Vision predicting suspicious pancreatic findings using computer vision
Wigler, M. **Krasnitz, A.**[2]	$447,351	2023-2027	CSHL-Northwell Health	Endometrial cancer sequence-based detection and outcome prediction
Tuveson, D. **Krasnitz, A.**[2]	$124,873	2023-2026	The New York Genome Center	Investigation of the dark genome influence on targeted and chemotherapy sensitivity in pancreatic cancer
Wigler, M. **Levy, D.**[2]	$205,636	2023-2027	CSHL-Northwell Health	Endometrial cancer sequence-based detection and outcome prediction
Nowak, D. **Siepel, A.**[2]	$465,374	2023-2028	National Institutes of Health	Novel mouse models for quantitative understanding of baseline and therapy-driven evolution of prostate cancer metastasis

NEW EXTERNAL FUNDING IN 2023

PI	TOTAL AWARD[1]	YEARS	FUNDING SOURCE	PROJECT TITLE
Siepel, A.	$2,951,205	2023-2028	National Institutes of Health	Evolutionary human genomics, demography, natural selection, and transcription regulation
Siepel, A.	$150,000	2023-2024	Starr Consortium	Determining and targeting evolutionary trajectories driving bladder cancer

[1]Total award, including indirect costs

[2]PI designated this amount to the individual QB investigator

2023 LABORATORY MEMBERSHIP

CORE FACULTY

FACULTY	LAB MEMBERS	POSITION	YEAR STARTED
Molly Gale-Hammell Associate Professor	Cole Wunderlich	Graduate Student, CSHL	2017
Ivan Iossifov Professor	Yoon-Ha Lee Steve Marks	Research Investigator Computational Science Developer III	2005 2010
Justin Kinney Associate Professor	John Desmarais Taehoon Ha Kaiser Loell Evan Seitz Debora Tenenbaum	Postdoc Computational Biostatistician Postdoc Computational Postdoc Computational Postdoc Computational	2023 2020 2023 2022 2022
Peter Koo Assistant Professor	Nika Chuzhoy Alessandro Crnjar Pretty Garcia Jakub Kaczmarzyk Yijie Kang Shivani Muthukumar Chandana Rajesh Kaeli Rizzo Anirban Sakar Ziqi Tang Shushan Toneyan Jessica Zhou	Volunteer Postdoc Computational PREP Scholar Graduate Student, Visiting Graduate Student, Rotating Volunteer Graduate Student, SBU Graduate Students, CSHL Postdoc Computational Graduate Student, CSHL Graduate Student, CSHL Postdoc Computational	2023 2023 2023 2021 2023 2023 2022 2022 2023 2019 2019 2023
Alexander Krasnitz Research Professor	Pascal Belleau	Research Investigator	2017
Dan Levy Associate Professor	Swati Negi	Computational Science Developer I	2023

CORE FACULTY

FACULTY	LAB MEMBERS	POSITION	YEAR STARTED
David McCandlish Associate Professor	Bryan Gitschlag Carlos Martí Gomez Aldaravi Megyi Sun	Postdoc Computational Postdoc Computational Postdoc Computational	2021 2021 2023
Hannah Meyer Assistant Professor	Salome Carcy Sarah Chapin Rishvanth Kaliappan Prabakar Vasilisa Kovaleva Yong Lin Lijuan Sun Maha Syed Johsua Torres Caroline Walter	Graduate Student, CSHL Computational Science Developer I Postdoc Computational Computational Science Developer I Postdoctoral Fellow Visiting Scientist Graduate Student, CSHL Research Technician II Volunteer	2020 2020 2023 2023 2022 2023 2022 2022 2022
Saket Navlakha Associate Professor	Amitava Banerjee* Bahruz Jabiyev Yang Shen Matthew Venezia Xingyu Zheng	Postdoc Computational Postdoc Computational Postdoc Computational Intern Lab, College Graduate Student, CSHL	2023 2023 2019 2022 2020
Adam Siepel Professor	Rebecca Hassett Lingjie Liu Luiz Carlos Machado de Oliveira Ziyi Mo Armin Scheben Stephen Staklinski Alexander Xue Yixin Zhao	Computational Science Developer I Graduate Student, SBU Graduate Student, SBU Graduate Student, CSHL Postdoc Computational Graduate Students, CSHL Postdoc Computational Research Investigator	2022 2020 2022 2018 2019 2022 2018 2018

*Interdisciplinary Scholar in Quantitative and Experimental Biology (ISEQB) co-mentored with Hannah Meyer.

ASSOCIATED MEMBERS

FACULTY	POSITION
Alexander Dobin	Assistant Professor
Alexei Koulakov	Professor
W. Richard McCombie	Professor
Partha Mitra	Professor
Doreen Ware	Professor

ADMINISTRATIVE SUPPORT

NAME	POSITION	YEAR STARTED
Katherine Brenner	Assistant Director of Administration	2018
Susan Fredericks	Assistant to the Chair and Sr. Scientific Administrator	2022
Antonia Little	Scientific Administrator	2023

Cold Spring Harbor Laboratory
SIMONS CENTER FOR QUANTITATIVE BIOLOGY

One Bungtown Road
Cold Spring Harbor, NY 11724