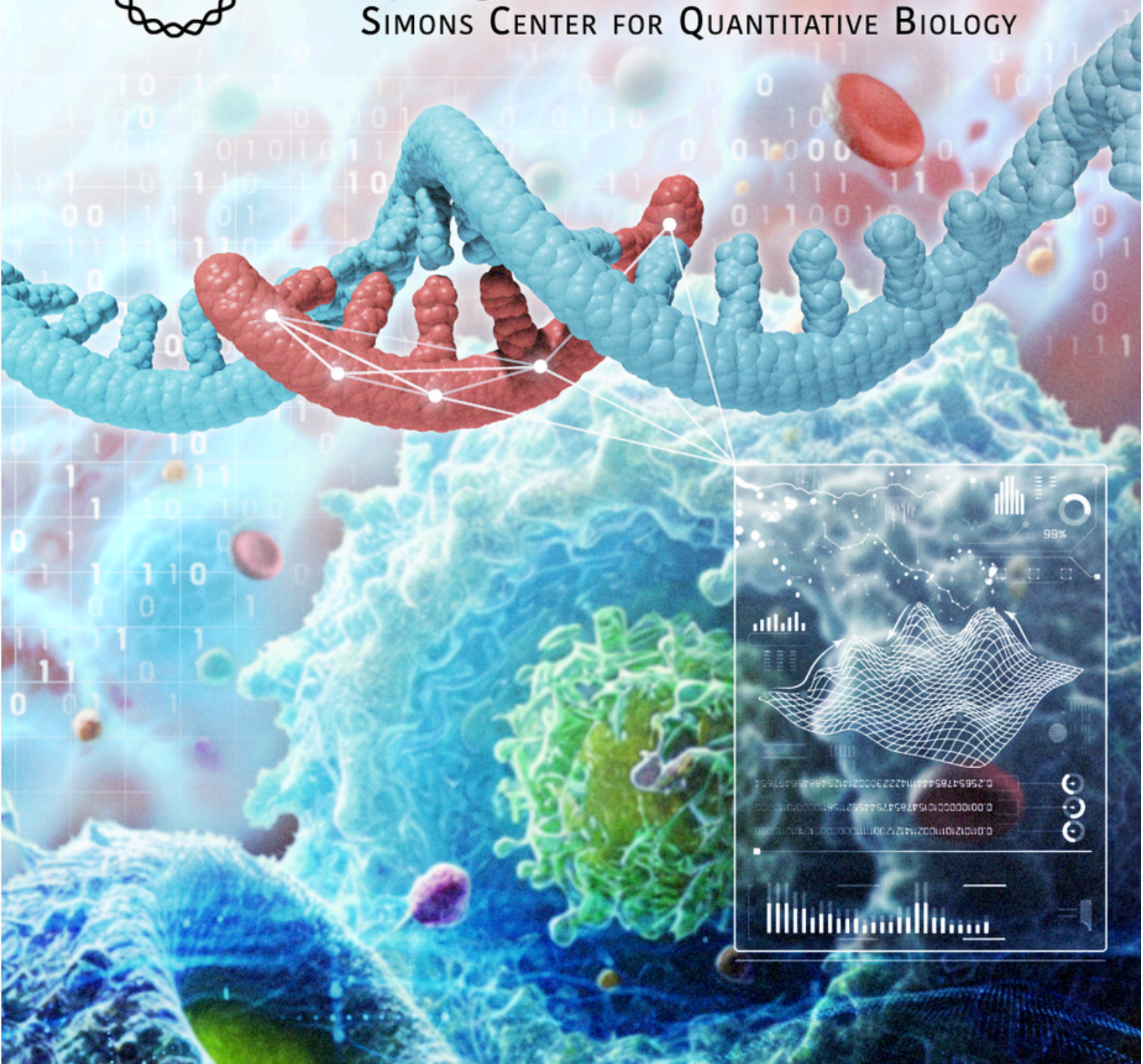
2024
ANNUAL REPORT

Cold Spring Harbor Laboratory
SIMONS CENTER FOR QUANTITATIVE BIOLOGY



Contents

Cover art by Evan Seitz, Postdoctoral Fellow, Kinney Laboratory

Letter from the Chair

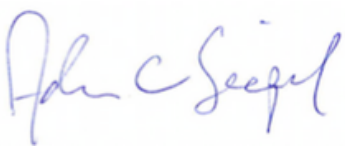
The story of 2024 at the SCQB can be summed up in two words: artificial intelligence. AI is everywhere these days, of course, from our phones to our homes and our cars, and it often drives the daily news cycle and the stock market. Over the last decade, the importance of AI in biological research has greatly accelerated. After several years of intensive investment in AI at CSHL, our efforts are finally beginning to pay off.

This year, investigators at the SCQB published papers on new AI-based methods for study of gene regulation (pp. 5-6), for predicting T-cell recognition (p. 7), for understanding patterns of metastasis in prostate cancer (pp. 8-9), and for revealing how splice-modifying drugs work (p. 11), among other applications. We obtained two NIH grants with major AI components totaling over $2.9M, as well as an AI-related "Pivot Fellowship" from the Simons Foundation (p. 4) and an AI-focused grant from CSHL-Northwell Health. AI guru Peter Koo, who was recruited in 2019, was promoted to associate professor (p. 4) and five our Ph.D. students graduated, four of them moving on to AI-related positions in industry (p. 12). In addition, Associate Professor Saket Navlakha is in the process of launching a new company, called knose.ai, for the use of artificial intelligence in odor detection (p. 8).

Until recently, efforts in AI at CSHL were split between quantitative biology and neuroscience, but we are in the process of unifying these efforts under the umbrella of the SCQB. We recently combined the Neuro AI and QB seminar series, to facilitate interaction and information exchange, and new research collaborations are already beginning to emerge (p. 14). Finally, quantitative biologists and computational neuroscientists worked together this year with CSHL's IT staff to establish a major upgrade to our high-performance computing resources, including new GPUs to support biological AI applications (p. 15).

As I step down as Chair of the SCQB in a few days, I plan to refocus my own research efforts in AI-based applications in tumor phylogenetics, transcriptional regulation, and population genetics. AI will not be the only thing we do at the SCQB, but it is clear that it will be a foundational technology that pervades research across our center and all of CSHL. I anticipate it will lead to many exciting biological discoveries in the years to come.

Best Wishes for the New Year,



Adam Siepel
December 27, 2024

External Advisory Committee

The Simons Center for Quantitative Biology (SCQB) at Cold Spring Harbor Laboratory (CSHL) benefits from the guidance of distinguished leaders in computational biology, mathematics, and related fields. We are grateful for their continued support and strategic counsel.

Andrew Clark, Ph.D.
Professor of Molecular Biology and Genetics
Cornell Univeristy

David Donoho, Ph.D.
Anne T. and Robert M. Bass Professor of Humanities and Sciences
Professor of Statistics
Stanford University

Molly Prezworksi, Ph.D.
Professor of Biological Sciences and Systems Biology
Columbia University

Eric D. Siggia, Ph.D.
Viola Ward Brinning and Elbert Calhoun Brinning Professor
Head of Laboratory of Theoretical Condensed Matter Physics
The Rockefeller University

Eero P. Simoncelli, Ph.D.
Silver Professor of Neural Science, Mathematics, Data Science, and Psychology | New York University
Scientific Director | Center for Computational Neuroscience, Flatiron Institute, Simons Foundation

Steven Salzberg, Ph.D. (Chair)
Bloomberg Distinguished Professor of Biomedical Engineering, Computer Science, and Biostatistics
Director, Center for Computational Biology
Johns Hopkins University

Committee's Role

- Provide strategic guidance on research directions and priorities
- Evaluate program effectiveness and impact
- Advise on recruitment and resource allocation
- Foster connections with external institutions and partners

Executive Summary

2024

In 2024, the SCQB achieved several important discoveries in AI-driven genomics and computational biology, marked by the development of novel interpretable AI tools like CREME and SQUID for decoding gene regulation, and BATMAN for predicting T-cell recognition patterns. The Center secured over $3.6M in new funding, expanded its computational infrastructure with state-of-the-art GPU clusters, and celebrated five exceptional PhD graduates transitioning into roles at leading institutions, including Meta, BioMarin, and Roche. Our interdisciplinary research yielded advances in cancer metastasis tracking through EvoTraceR, splicing drug mechanisms through innovative thermodynamic modeling, and autism genetics through the GPF platform. These advances continue to position the SCQB at the forefront of quantitative biology research.

Key Achievements in 2024

AI & Genomics Tools

Advanced genomics research through three innovative platforms for gene regulation, genomic interpretation, and T-cell predictions.

Disease Research

Pioneered new platforms for cancer metastasis tracking and autism genetics, with breakthroughs in drug mechanism modeling.

Neuro AI Integration

Launched new QB/Neuro AI Joint Seminar Series and expanded course offerings with Neuro AI faculty, strengthening the bridge between quantitative biology and computational neuroscience approaches.

Educational Excellence

Five PhD graduates advancing to Meta, BioMarin, Roche, and other leading institutions. Shushan Toneyan received the Birnstiel Award.

Research Funding

Secured $3.6M in new grants, including $2.9M in federal and $634K in private funding.

Next-Gen Computing

Deployed 20 NVIDIA H100 GPUs bringing the total to 60 GPUs with 12 PB storage system, enabling 30x faster processing for advanced computational research.

Faculty & Leadership

Our nine faculty members bring together expertise in mathematical biology, computational methods, and genomic technologies to understand genome function, evolution, and disease mechanisms.

Adam Siepel
Professor & Chair
Computational genomics and evolution

Justin Kinney
Associate Professor
Machine learning and biophysics in gene regulation

Ivan Iossifov
Professor
Autism genetics and computational biology

Peter Koo
Associate Professor
Innovation in deep learning for genomic analysis

Alexander Krasnitz
Research Professor
Cancer genomics and genetics ancestry research

Dan Levy
Associate Professor
Statistical genetics and computational methods

David McCandlish
Associate Professor
Protein evolution and quantitative modeling

Hannah Meyer
Assistant Professor
Immunological systems and computational biology

Saket Navlakha
Associate Professor
Algorithms in nature and neuro-inspired AI

Advances & Awards

★ **Peter Koo: Promotion and Cancer Center Appointment**

Peter Koo has been promoted to Associate Professor and appointed co-Faculty Head of the CSHL Cancer Center's Sequencing Technologies and Analysis Core

Saket Navlakha: Simons Foundation Pivot Fellowship

Associate Professor Saket Navlakha has received a Simons Foundation Pivot Fellowship to study immunology under the mentorship of Assistant Professor Hannah Meyer. Their collaboration will explore connections between immune system adaptation and machine learning.

Research Highlights

AI and Machine Learning in Genomics

In 2024, several important new AI interpretation tools emerged from the SCQB, including CREME and SQUID, which enhanced scientists' ability to understand how deep learning models analyze gene regulation. These platforms make AI systems more transparent and biologically interpretable, helping researchers connect computational predictions with biological mechanisms. The development of BATMAN demonstrated similar advances in immunology, achieving 20% improved accuracy in predicting T-cell recognition patterns and cross-reactivity. Meanwhile, Knose.ai showed how computational biology approaches could extend beyond genomics, applying brain-inspired algorithms to digital olfaction with potential healthcare applications at major medical centers.

New AI Tool Decodes Gene Expression Patterns

During her graduate studies at the SCQB, Shushan Toneyan, worked with Associate Professor **Peter Koo** to develop a powerful new computational tool that helps scientists peek inside the "black box" of artificial intelligence systems studying gene regulation.

Modern AI systems, particularly deep neural networks (DNNs), have become remarkably accurate at predicting gene expression from DNA sequences. However, determining how these AI models identify and interpret the regulatory DNA elements that control gene activity remains challenging. Understanding how DNNs make these decisions is crucial since it could reveal the fundamental rules of gene regulation in cells.

Toneyan and Koo's solution, called CREME (cis-regulatory element model explanations), is a sophisticated diagnostic tool for AI models. It systematically probes how the AI responds to carefully designed changes in DNA sequences, revealing the rules the system has learned about gene regulation. Unlike previous methods for genomic AI interpretability, which focused mainly on short DNA patterns (motifs), CREME is able to decipher longer sequences and complex interactions.

CREME therefore addresses a critical need in the group field of AI-powered genomics: translating computational predictions into biological insights that scientists can use to better understand how our genes are regulated.

Dr. Shushan Toneyan (left) and Assistant Professor Peter Koo (right)

Toneyan S, Koo PK. Interpreting cis-regulatory interactions from large-scale deep neural networks. Nat Genet. 2024 Sep 16.

Notable Advances

- CREME can analyze how different regions of DNA work together to control gene activity
- The tool works across multiple scales, from broad regulatory regions down to specific DNA sequences
- It can identify both enhancer and silencer elements that control gene expression

Research Highlights

Making Deep Learning Interpretable for Genomics

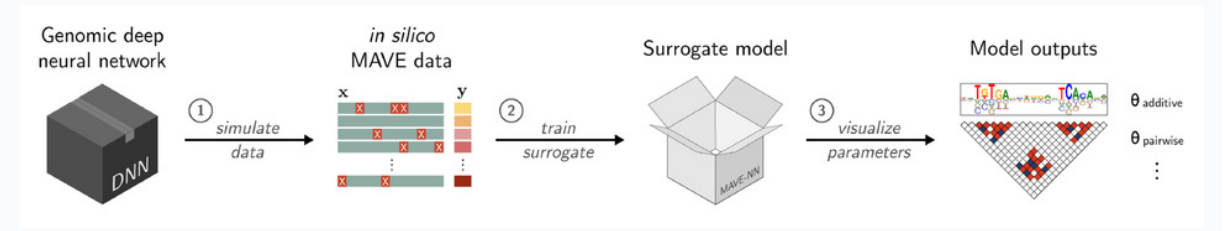
SQUID (Surrogate Quantitative Interpretability for Deepnets) represents another significant advance in genomic AI interpretation. It was developed through collaboration at the SCQB by Associate Professors **David McCandlish, Justin Kinney**, and **Peter Koo**, and postdoctoral researcher Evan Seitz. Their work, published in *Nature Machine Intelligence*, brings together the complementary research expertise of McCandlish in evolutionary modeling, Kinney in regulatory genomics, and Koo in deep learning to address a fundamental challenge in genomic deep learning.

While deep neural networks (DNNs) can predict how DNA sequences control genome function, current AI analysis tools struggle to provide consistent biological insights because they analyze DNA sequences like they would analyze images or text. These interpretation methods often miss how DNA sequences can have effects based on their genomic context and fail to capture the intricate regulatory interactions between distant DNA elements that jointly control gene expression.

SQUID takes a biology-centered approach using surrogate models—simplified, inherently interpretable mathematical representations—to approximate the neural network's behavior in specific regions of "sequence space." The method incorporates domain-specific knowledge about genomic data and accounts for biological complexities like nonlinear effects and context-dependent noise that can confound other methods.

The collaborative team's method has demonstrated remarkable effectiveness, showing more consistent identification of DNA binding motifs and improved predictions for how DNA mutations affect gene regulation. SQUID can quantify complex phenomena, such as interactions between regulatory elements in DNA, and offers flexibility to test various biological hypotheses. SQUID is particularly well-suited for in-depth study of specific genomic regions of interest, such as disease-associated DNA sequences.

Seitz EE, McCandlish DM, Kinney JB, Koo PK. Interpreting cis-regulatory mechanisms from genomic deep neural networks using surrogate models. Nat Mach Intell. 2024 Jun 21.



An illustration outlining the SQUID computational pipeline

Notable Advances

- SQUID interprets genomic AI by using simplified models to analyze DNA regions
- Integration of biological principles to analyze DNA regulation and interactions
- Enhanced prediction of DNA sequence patterns and mutation effects

Research Highlights

Interdisciplinary Scholars in Experimental & Quantitative Biology

Amitava Banerjee

The Interdisciplinary Scholars in Experimental Quantitative Biology (ISEQB) program at the SCQB awarded funding to Assistant Professor **Hannah Meyer** and Associate Professor **Saket Navlakha** in 2023. Through their collaboration, ISEQB Scholar Amitava Banerjee has advanced our understanding of T cell protein recognition, bridging immunology and computational biology. Their findings are available as a preprint on *bioRxiv*.

AI Predicts T-Cell Recognition Patterns

Hannah Meyer and Saket Navlakha together with their shared postdoc and ISEQB Scholar Amitava Banerjee have developed a new computational tool called Bayesian Inference of Activation of TCR by Mutant Antigens (BATMAN) that significantly improves our ability to predict how T lymphocytes, crucial components of our adaptive immune system, interact with different protein fragments (peptides). This advance addresses a major challenge in immunology: predicting which peptides will activate specific T cells.

T cells use specialized receptors (T cell receptors, TCRs) to recognize potentially dangerous peptides bound to major histocompatibility complex (MHC) molecules. A single TCR can recognize multiple different peptides — a property called cross-reactivity. Understanding this cross-reactivity is essential for developing better cancer immunotherapies and predicting viral mutations that might escape immune detection. However, existing computational methods have struggled to predict how single amino acid substitutions in peptides affect TCR activation.

To solve this problem, the team first created an extensive database of over 10,000 TCR-peptide-MHC interactions, systematically testing how single amino acid changes in peptides affect T cell activation. This comprehensive dataset provided crucial positive and negative examples for training their computational model.

BATMAN uses a novel approach that combines position-specific weights with learned amino-acid substitution patterns to predict TCR activation. The model outperformed existing methods by 20% and, importantly, captured known biochemical principles of TCR-peptide-MHC interactions. For example, it confirmed that mutations in the central residues of peptides have greater effects on TCR activation than changes at terminal positions and identified specific amino acid substitutions that consistently impact TCR recognition. This work represents a significant step forward in our ability to predict immune responses computationally, with potential applications in vaccine design, cancer immunotherapy, and understanding autoimmune diseases.

Banerjee A, Pattinson DJ, Wincek CL, Bunk P, Chapin SR, Navlakha S, Meyer HV. BATMAN: Improved T cell receptor cross-reactivity prediction benchmarked on a comprehensive mutational scan database. bioRxiv [Preprint]. 2024 Feb 8.

Notable Advances

- Built AI-powered immune recognition model (BATMAN) achieving 20% improved prediction accuracy, enabling precise prediction of immune cell responses.

Research Highlights

SCQB Research Advances Digital Olfaction

knose.ai
AN AI PLATFORM FOR DETECTING SMELLS

SCQB Associate Professor **Saket Navlakha** and Salk Institute Professor Sreekanth Chalasani have developed a pioneering AI platform that combines specialized gas sensors with brain-inspired algorithms to detect and analyze odors. This technology forms the foundation of their startup company, Knose.ai, demonstrating how computational biology research can lead to practical innovations.

Initial healthcare applications are being explored through pilot studies with Memorial Sloan Kettering and Northwell Health, with potential extensions to environmental monitoring and agricultural safety.

Technology Foundations

- Brain-inspired algorithms
- 98% detection accuracy

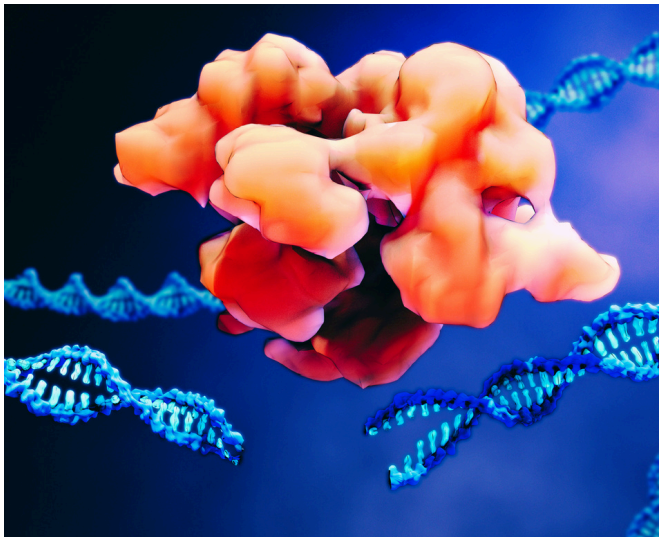
A demonstration of how the SCQB's computational biology research drives practical innovation

Genomics and Disease Research

Recent breakthroughs at the SCQB showcase innovative computational approaches to understanding complex medical challenges. From mapping cancer metastasis pathways using CRISPR technology to developing sophisticated models for analyzing drug mechanisms and creating platforms for autism genetic research, these advances demonstrate how computational biology and interdisciplinary collaboration are accelerating our understanding of human diseases.

Cancer Research Uncovers New Metastasis Pathways

Recent research from Professor **Adam Siepel**'s lab, in collaboration with Dawid Nowak's team at Weill Cornell Medicine, provides new insights into prostate cancer metastasis. By combining bioluminescence imaging, CRISPR/Cas9-based barcoding, and computational methods, the team has developed a methodology for tracing how cancer spreads from its initial site through the body.



The study employed EvoCaP (Evolution in Cancer of the Prostate), a mouse model developed by Nowak's lab that mimics metastatic cancer spreading to bone, liver, lungs, and lymph nodes.

Research Highlights

Using CRISPR-Cas9 lineage tracing technology, the researchers "barcoded" cancer cells with a 260 base-pair sequence, allowing them to track cell movement through the body. This tracking was analyzed using EvoTraceR, a software package developed together by the two research groups. Their findings, published in *Cancer Discovery*, show that only a small number of cancer cell lineages drive most of the disease's spread, primarily through migrations from the prostate to bone or liver tissue.

Understanding these preferences could inform treatments for different metastatic sites. The team is exploring applications beyond prostate cancer, including bladder cancer and studying how treatments like androgen deprivation therapy affect the disease's evolution. This interdisciplinary collaboration between Siepel's computational team and Nowak's cancer researchers shows how cancer operates as an evolutionary process, with cells competing and faster-growing variants becoming dominant.

Serio RN, Scheben A, Lu B, Gargiulo DV, Patruno L, Buckholtz CL, Chaffee RJ, Jibilian MC, Persaud SG, Staklinski SJ, Hassett R, Brault LM, Ramazzotti D, Barbieri CE, Siepel AC, Nowak DG. Clonal Lineage Tracing with Somatic Delivery of Recordable Barcodes Reveals Migration Histories of Metastatic Prostate Cancer. Cancer Discov. 2024 Oct 4

Notable Advances

- **EvoCaP** mouse model authentically replicates human prostate cancer progression and enables precise tumor tracking via CRISPR/Cas9 barcoding
- **EvoTraceR** analytical pipeline processes complex clonal evolution data and maps metastatic patterns
- Integration of both systems quantifies cancer cell migration kinetics and reveals pathways for potential therapeutic targeting

Krasnitz Awarded NCI Grant to Study Cancer's Genetic Ancestry

In 2024, Research Professor **Alexander Krasnitz** (left) was awarded a collaborative grant (U01CA289357) from the National Cancer Institute's Informatics Technology for Cancer Research (ITCR) program to develop innovative software tools for inferring genetic ancestry from cancer-derived molecular data.

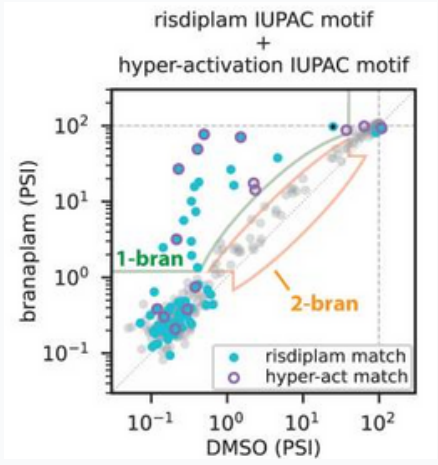
This project will unlock cancer genomic data repositories for ancestry studies without requiring patient genotypes or demographic data. The tools will analyze various molecular data types (whole-genome, RNA-seq, ATAC-seq) to infer continental-level ancestry with quantifiable accuracy. Through the ITCR Training Network, these FAIR-compliant tools will be freely available via GitHub and Galaxy, enabling large-scale studies of ancestry-specific patterns in cancer.

Research Highlights

Quantitative Models Decode Splicing Drug Mechanisms

In a new *Nature Communications* paper, Associate Professors Justin Kinney and David McCandlish, together with CSHL Professor Adrian Krainer, present novel quantitative models to understand how splice-modifying drugs work. The Kinney lab led experimental work and computational analyses to study two drugs—risdiplam (an FDA-approved drug for treating spinal muscular atrophy) and branaplam (an experimental drug being developed for Huntington's disease). Their experimental work included massively parallel reporter assays (a technique co-invented by Kinney) and RNA sequencing experiments, which provided rich datasets for sophisticated statistical analysis and thermodynamic modeling.

The research team developed a computational approach centered on two key innovations: a thermodynamic model for drug-dependent exon inclusion and a Bayesian inference framework for determining sequence-specific drug effects. This allowed them to separate the inherent effects of splice site sequences from drug-specific effects-something previous analyses could not achieve.



Branaplam's dual binding modes revealed through parallel splcing analysis of variant sequences.

Through their combined experimental and computational analyses, the team's results contradicted the prevailing "two-site hypothesis" for risidiplam's mechanism of action and instead revealed that branaplam interacts with RNA through two distinct binding modes. The team also used empirical Bayesian modeling to analyze drug dose-response curves. This analysis uncovered widespread cooperative effects for single drugs and widespread synergistic effects between pairs of drugs.

The integrated experimental and computational methods developed by the team provide a new framework for studying splice-modifying therapeutics, offering predictive power that could aid in developing new treatments. The work demonstrates how combining experimental approaches with sophisticated mathematical modeling can extract mechanistic insights from complex biological data.

Ishigami Y, Wong MS, Martí-Gómez C, Ayaz A, Kooshkbaghi M, Hanson SM, McCandlish DM, Krainer AR, Kinney JB. Specificity, synergy, and mechanisms of splice-modifying drugs. Nat Commun. 2024 Feb 29.

Notable Advances

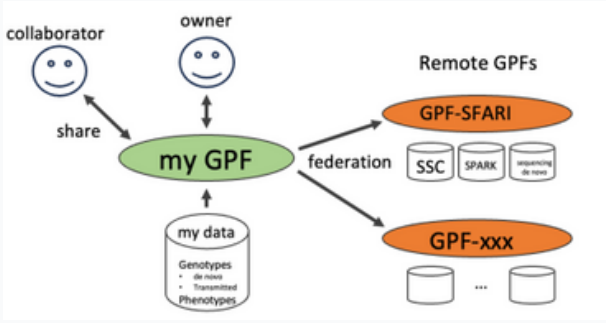
- New models separate drug-specific from sequence effects in splicing, improving drug design
- Revealed branaplam's two distinct binding modes, contradicting prior models and enabling targeted therapeutics
- Uncovered widespread drug synergies through statistical dose response analysis, suggesting combination therapies

Research Highlights

Mapping Autism's Genetic Landscape

Professor **Ivan Iossifov**'s lab at the SCQB has developed GPF (Genotypes and Phenotypes in Families), an advanced computational platform for managing and analyzing genetic and phenotypic data from large family collections. The GPF platform was initially created to handle data from the Simons Simplex Collection (SSC), comprising ~2,800 families with autism, and later expanded to accommodate the larger SPARK (Simons Foundation Powering Autism Research) collection of ~100,000 individuals with autism and their families.

The platform can process genetic variant types, from single nucleotide changes to large copy number variations, while handling different inheritance patterns and family structures. What makes GPF particularly valuable is its ability to help researchers ask sophisticated questions about how specific genetic changes might relate to autism characteristics. Scientists can use the platform to identify patterns in genetic mutations, explore how these mutations might affect behavior and development, and compare findings across different groups of individuals.



GFP: A personal research database that connects and shares data across research networks.

The platform also helps protect patient privacy while advancing research. While some sensitive data requires careful authorization to access, GPF makes certain aggregate information freely available to all researchers. This includes the "Sequencing *de novo*" dataset, which contains published genetic mutations found in individuals with autism and other developmental conditions, as well as summary statistics about genes that might play a role in autism.

Iossifov and his team designed GPF to be highly flexible and open-source, meaning other researchers can adapt it for their own genetic studies beyond autism research. The system can handle data from millions of individuals and support various types of family structures and genetic analysis.

Chorbadjiev L, Cokol M, Weinstein Z, Shi K, Fleisch C, Dimitrov N, Mladenov S, Xu S, Hall J, Ford S, Lee YH, Yamrom B, Marks S, Munoz A, Lash A, Volfovsky N, Iossifov I. The Genotype and Phenotypes in Families (GPF) platform manages the large and complex data at SFARI. bioRxiv [Preprint]. 2024 Feb 11.

Notable Advances

- Built scalable genomics platform processing millions of individuals' variant/phenotype data, enabling precise autism variant analysis
- Deployed at SFARI (~300K individuals), balancing secure access for sensitive data with public availability of key findings
- Created open-source system allowing researchers worldwide to analyze genetic data collaboratively while maintaining privacy.

Education and Training

Graduate Program Success at the SCQB

The SCQB Graduate Program, in collaboration with Cold Spring Harbor Laboratory's School of Biological Sciences and Stony Brook University, continues to demonstrate excellence in training the next generation of quantitative biologists. SCQB labs are a vibrant training ground for graduate students from both CSHL's School of Biological Sciences and Stony Brook University, currently hosting 12 graduate students with 14 PhDs completed since 2018. Our comprehensive training includes a QB Bootcamp, core quantitative biology coursework, and mentored studies. This approach combines rigorous computational training with cutting-edge biological research, preparing students for impactful careers in academia and industry.

2024 Doctoral Recipients

Salomé Carcy

Advisor
Hannah Meyer
Current Role
Medical student at Université Paris Cité

Lingjie Lu

Advisor
Adam Siepel
Current Role
Scientist I in Computational Biology at BioMarin Pharmaceuticals

Ziyi Mo

Advisor
Adam Siepel
Current Role
Research Scientist, Machine Learning at Meta

Ziqi (Amber)Tang

Advisor
Peter Koo
Current Role
Research Scientist, InstaDeep

Shushan Toneyan

Advisor
Peter Koo
Current Role
Postdoctoral Fellow, Roche

Key Achievements & Impact

Outstanding Recognition

- Shushan Toneyan received the prestigious 2024 International Birnstiel Award for Doctoral Research in Molecular Life Sciences, one of only six recipients worldwide.

Class of 2024 Research Impact

- The five graduating students have collectively produced 16 papers (13 publications, 3 preprints), with 8 as first authors.
- Ziyi Mo's *PLOS Genetics* paper on using domain adaptation to improve population genetic inference has been widely read and cited.
- Ziqi (Amber) Tang's comprehensive evaluation of DNA language models for regulatory genomics, published in *bioRxiv*, directly led to her role at InstaDeep, where she applies these insights to develop foundation models for biology.

CSHL's Postbaccalaureate Research Program and the SCQB

The SCQB participates in CSHL's Postbaccalaureate Research Education Program (PREP), which began in 2023. The program provides research experiences across multiple disciplines, including cancer, neuroscience, genomics, quantitative biology, and plant biology, combined with mentoring, coursework, and professional development for students interested in pursuing PhD programs.

The QB Program has seen increased participation in PREP, growing from one student in 2023 to two students in 2024.

PREP Students in QB Labs

2023 Cohort

Pretty Garcia worked in Peter Koo's lab, where she learned deep learning approaches for studying RNA-protein interactions.

She was accepted to multiple PhD programs and enrolled in the **Tri-Institutional PhD Program in Computational Biology and Medicine**, a collaborative program between Weill Cornell Medicine, Memorial Sloan Kettering Cancer Center, and The Rockefeller University in New York City.

2024 Cohort

Two PREP scholars joined QB labs:

Nadia Pressad joined Hannah Meyer's lab, studying thymocyte development.

Nirali Soma joined Peter Koo's lab, working on diffusion models for regulatory DNA sequence generation.

Coursework

PREP scholars can participate in CSHL's graduate-level courses, including the QB course. The curriculum includes:

- Statistics and programming
- Machine learning
- Algorithms
- Evolution
- Genomics
- Biophysics

Faculty members Ivan Iossifov, Justin Kinney, Peter Koo, Adam Siepel, David McCandlish, and Hannah Meyer lead the courses.

The program combines research experience, mentorship, and coursework to prepare students for graduate studies in quantitative biology and related fields.

2024 PREP Scholars

Community and Collaboration

SCQB labs continue to foster collaborations across CSHL's research areas. These partnerships demonstration the Center's commitment to integrating quantitative approaches throughout the institutions scientific programs.

Plant Biology

The SCQB collaborates with CSHL's Plant Biology Program, applying quantitative and computational methods to understand complex developmental patterns and genetic interactions in plant systems.

Active Collaborations

- David McCandlish & Zachary Lippman — *Plant epistasis*
- Saket Navlakha & Ullas Pedmale, David Jackson — *Voronoi patterning in plant ventation*

Neuroscience

The SCQB and CSHL's Neuroscience Program launched a joint AI seminar series bridging genomics and neural computation. This partnership leverages our shared strengths in machine learning, where Peter Koo and Saket Navlakha's work in interpretable AI complements Anthony Zador and Ben Cowley's Neuro AI advances. The collaboration extends to education, with Neuroscience faculty Ben Cowley and Helen Hou contributing to the QB course.

Active Collaborations

- Saket Navlakha & Florin Albeanu, Arka Banerjee, Alexei Koulakov — *Computational studies of olfactory system*

Cancer

The SCQB works closely with CSHL's NCI-designated Cancer Center, with seven of our nine Core Faculty serving as members. Our faculty contribute key leadership, with Adam Siepel co-leading Cancer Genetics & Genomics, Justin Kinney co-leading Gene Regulation & Inheritance, and Peter Koo serving as co-Faculty Head of Sequencing Technologies & Analysis.

Active Collaborations

- Adam Siepel & Lloyd Trotman — *Prostate cancer metastasis*
- Justin Kinney & Adrian Krainer — *MAVES for splicing*
- Justin Kinney & Bruce Stillman — *MAVES for origin replication*
- Peter Koo & Camila dos Santos — *Gene regulation in breast cancer*
- Hannah Meyer & Tobias Janowitz — *Cancer systems biology*

Infrastructure and Resources

Next-Generation AI Computing at the SCQB

The SCQB and CSHL's Neuro AI group are spearheading a transformative leap in the Laboratory's artificial intelligence capabilities. A $2 million federal grant, matched with institutional funding has equipped CSHL with computing power that redefines what is possible in computational biology.

Why It Matters

Analyzing a single gene regulatory network involves processing millions of data points. Tasks that once took a month can now be completed in a day, dramatically accelerating the research cycle. For QB researchers studying complex biological systems, this means testing more hypotheses and iterating models faster than ever before.

A New Era of AI Leadership: SCQB + Neuro AI

The SCQB drives CSHL's machine learning innovations, with Associate Professor **Peter Koo** and Associate Professor **Saket Navlakha** developing cutting-edge approaches in interpretable AI and computational genomics. This expertise is amplified through collaboration with CSHL's Neuro AI group, where Professor Anthony Zador and Assistant Professor Ben Cowley apply advanced machine learning to computational neuroscience. The synergy between QB's foundational work in biological AI and Neuro AI's computational approaches positions CSHL at the forefront of machine learning in biology.

Computing Power at Scale

- **Speed:** 30x faster AI processing enables real-time analysis of massive genomic datasets
- **Storage:** 12 petabyte system transfers data at 180 GB/second
- **Power:** 20 NVIDIA H100 GPUs specifically optimized for biological AI applications, bringing the total to 60 GPUs
- **Scale:** Enables analysis of entire genomes and spatial transcriptomics at unprecedented resolution

Enabling New Science

This integrated system allows SCQB researchers to:

- Build comprehensive models of gene regulation across entire genomes
- Develop interpretable deep learning archictures
- Process spatial transcriptomics data at cellular resolution
- Model complex biological networks at scale

Scientific Publications

Peer-Reviewed Articles

Berube B, Ernst E, Cahn J, Roche B, de Santis Alves C, Lynn J, Scheben A, Grimanelli D, **Siepel A**, Ross-Ibarra J, Kermicle J, Martienssen RA. Teosinte Pollen Drive guides maize diversification and domestication by RNAi. *Nature*. 2024 Sep;633(8029):380-388. doi: 10.1038/s41586-024-07788-0. Epub 2024 Aug 7. PMID: 39112710; PMCID: PMC11390486.

Chen WC, Zhou J, **McCandlish DM.** Density estimation for ordinal biological sequences and its applications. Phys. Rev. 2024 Oct 30; E110, 044408.

Dasgupta S, Meirovitch Y, Zheng X, Bush I, Lichtman JW, **Navlakha S**. A neural algorithm for computing bipartite matchings. *Proc Natl Acad Sci U S A*. 2024 Sep 10;121(37):e2321032121. doi: 10.1073/pnas.2321032121. Epub 2024 Sep 3. PMID: 39226341; PMCID: PMC11406297.

Gitschlag BL, Pereira CV, Held JP, **McCandlish DM**, Patel MR. Multiple distinct evolutionary mechanisms govern the dynamics of selfish mitochondrial genomes in Caenorhabditis elegans. *Nat Commun*. 2024 Sep 19;15(1):8237. doi: 10.1038/s41467-024-52596-9. PMID: 39300074; PMCID: PMC11413162.

Ishigami Y, Wong MS, Martí-Gómez C, Ayaz A, Kooshkbaghi M, Hanson SM, **McCandlish DM**, Krainer AR, **Kinney JB.** Specificity, synergy, and mechanisms of splice-modifying drugs. *Nat Commun*. 2024 Feb 29;15(1):1880. doi: 10.1038/s41467-024-46090-5. PMID: 38424098; PMCID: PMC10904865.

Kuderna LFK, Ulirsch JC, Rashid S, Ameen M, Sundaram L, Hickey G, Cox AJ, Gao H, Kumar A, Aguet F, Christmas MJ, Clawson H, Haeussler M, Janiak MC, Kuhlwilm M, Orkin JD, Bataillon T, Manu S, Valenzuela A, Bergman J, Rouselle M, Silva FE, Agueda L, Blanc J, Gut M, de Vries D, Goodhead I, Harris RA, Raveendran M, Jensen A, Chuma IS, Horvath JE, Hvilsom C, Juan D, Frandsen P, Schraiber JG, de Melo FR, Bertuol F, Byrne H, Sampaio I, Farias I, Valsecchi J, Messias M, da Silva MNF, Trivedi M, Rossi R, Hrbek T, Andriaholinirina N, Rabarivola CJ, Zaramody A, Jolly CJ, Phillips-Conroy J, Wilkerson G, Abee C, Simmons JH, Fernandez-Duque E, Kanthaswamy S, Shiferaw F, Wu D, Zhou L, Shao Y, Zhang G, Keyyu JD, Knauf S, Le MD, Lizano E, Merker S, Navarro A, Nadler T, Khor CC, Lee J, Tan P, Lim WK, Kitchener AC, Zinner D, Gut I, Melin AD, Guschanski K, Schierup MH, Beck RMD, Karakikes I, Wang KC, Umapathy G, Roos C, Boubli JP, **Siepel A**, Kundaje A, Paten B, Lindblad-Toh K, Rogers J, Marques Bonet T, Farh KK. Identification of constrained sequence elements across 239 primate genomes. *Nature*. 2024 Jan;625(7996):735-742. doi: 10.1038/s41586-023-06798-8. Epub 2023 Nov 29. PMID: 38030727; PMCID: PMC10808062.

Lammertse E, Li S, Kendall J, Kim C, Morris P, Ranade N, **Levy D,** Wigler M, Brouzes E. Magnetically functionalized hydrogels for high-throughput genomic applications. *Adv Mater Technol*. 2024 Jan 22;9(2):2301155. doi: 10.1002/admt.202301155. Epub 2023 Dec 14. PMID: 38645306; PMCID: PMC11029686.

Lingjie Liu, Yixin Zhao, Rebecca Hassett, Shushan Toneyan, Peter K. Koo, and **Adam Siepel**. Probabilistic and machine-learning methods for predicting local rates of transcription elongation from nascent RNA sequencing data. *Nucleic Acids Research*. In press.

Scientific Publications

Loh L, Carcy S, Krovi HS, Domenico J, Spengler A, Lin Y, Torres J, Prabakar RK, Palmer W, Norman PJ, Stone M, Brunetti T, **Meyer HV**, Gapin L. Unraveling the phenotypic states of human innate-like T cells: Comparative insights with conventional T cells and mouse models. *Cell Rep.* 2024 Sep 24;43(9):114705. doi: 10.1016/j.celrep.2024.114705. Epub 2024 Sep 10. PMID: 39264810.

Ortiz JR, Lewis SM, Ciccone M, Chatterjee D, Henry S, **Siepel A,** Dos Santos CO. Single-Cell Transcription Mapping of Murine and Human Mammary Organoids Responses to Female Hormones. *J Mammary Gland Biol Neoplasia*. 2024 Jan 30;29(1):3. doi: 10.1007/s10911-023-09553-x. PMID: 38289401; PMCID: PMC10827859.

Rozhoňová H, Martí-Gómez C, **McCandlish DM**, Payne JL. Robust genetic codes enhance protein evolvability. *PLoS Biol.* 2024 May 16;22(5):e3002594. doi: 10.1371/journal.pbio.3002594. PMID: 38754362; PMCID: PMC11098591.

Seitz, E.E., **McCandlish, D.M.**, **Kinney, J.B.**, **Koo, P.K.** Interpreting cis-regulatory mechanisms from genomic deep neural networks using surrogate models. *Nat Mach Intell* 6, 701–713 (2024). https://doi.org/10.1038/s42256-024-00851-5.

Serio RN, Scheben A, Lu B, Gargiulo DV, Patruno L, Buckholtz CL, Chaffee RJ, Jibilian MC, Persaud SG, Staklinski SJ, Hassett R, Brault LM, Ramazzotti D, Barbieri CE, **Siepel AC**, Nowak DG. Clonal Lineage Tracing with Somatic Delivery of Recordable Barcodes Reveals Migration Histories of Metastatic Prostate Cancer. *Cancer Discov*. 2024 Oct 4;14(10):1990-2009. doi: 10.1158/2159-8290.CD-23-1332. PMID: 38969342.

Toneyan S, **Koo PK.** Interpreting cis-regulatory interactions from large-scale deep neural networks. *Nat Genet.* 2024 Sep 16. doi: 10.1038/s41588-024-01923-3. Epub ahead of print. PMID: 39284975.

Yoko Tajima, César D. M. Vargas, Keiichi Ito, Wei Wang, Ji-Dung Luo, Jiawei Xing, Nurdan Kuru, Luiz Carlos Machado, **Adam Siepel**, Thomas S. Carroll, Erich D. Jarvis, Robert B. Darnell. A humanized NOVA1 splicing factor alters mouse vocal communications. *Nature Communications*. In press.

Yu Y, Muthukumar S, **Koo PK.** EvoAug-TF: extending evolution-inspired data augmentations for genomic deep learning to TensorFlow. *Bioinformatics.* 2024 Mar 4;40(3):btae092. doi: 10.1093/bioinformatics/btae092. PMID: 38366935; PMCID: PMC10918628.

Preprints

Banerjee A, Pattinson DJ, Wincek CL, Bunk P, Chapin SR, **Navlakha S**, **Meyer HV**. BATMAN: Improved T cell receptor cross-reactivity prediction benchmarked on a comprehensive mutational scan database. *bioRxiv* [Preprint]. 2024 Feb 8:2024.01.22.576714. doi: 10.1101/2024.01.22.576714. PMID: 38370810; PMCID: PMC10871174.

Scientific Publications

Chorbadjiev L, Cokol M, Weinstein Z, Shi K, Fleisch C, Dimitrov N, Mladenov S, Xu S, Hall J, Ford S, Lee YH, Yamrom B, Marks S, Munoz A, Lash A, Volfovsky N, **Iossifov I**. The Genotype and Phenotypes in Families (GPF) platform manages the large and complex data at SFARI. *bioRxiv* [Preprint]. 2024 Feb 11:2024.02.08.579330. doi: 10.1101/2024.02.08.579330. PMID: 38370639; PMCID: PMC10871337.

Kovaleva VA, Pattinson DJ, Barton C, Chapin SR, Minervina AA, Richards KA, Sant AJ, Thomas PG, Pogorelyy MV, **Meyer HV**. copepodTCR: Identification of Antigen-Specific T Cell Receptors with combinatorial peptide pooling. *bioRxiv* [Preprint]. 2024 Feb 8:2023.11.28.569052. doi: 10.1101/2023.11.28.569052. PMID: 38077028; PMCID: PMC10705480.

Livesey BJ, Badonyi M, Dias M, Frazer J, Kumar S, Lindorff-Larsen K, **McCandlish DM**, Orenbuch R, Shearer CA, Muffley L, Foreman J, Glazer AM, Lehner B, Marks DS, Roth FP, Rubin AF, Starita LM, Marsh JA. Guidelines for releasing a variant effect predictor. ArXiv [Preprint]. 2024 Apr 16:*arXiv:*2404.10807v1. PMID: 38699161; PMCID: PMC11065047.

Posfai A, **McCandlish DM**, **Kinney JB**. Symmetry, gauge freedoms, and the interpretability of sequence-function relationships. *bioRxiv* [Preprint]. 2024 Jun 24:2024.05.12.593774. doi: 10.1101/2024.05.12.593774. PMID: 38798625; PMCID: PMC11118426.

Sokirniy I, Inam H, Tomaszkiewicz M, Reynolds J, **McCandlish D**, Pritchard J. A side-by-side comparison of variant function measurements using deep mutational scanning and base editing. *bioRxiv* [Preprint]. 2024 Sep 26:2024.06.30.601444. doi: 10.1101/2024.06.30.601444. PMID: 39005366; PMCID: PMC11244880.

Staklinski SJ, Scheben A, **Siepel A**, Kilberg MS. Utility of AlphaMissense predictions in Asparagine Synthetase deficiency variant classification. *bioRxiv* [Preprint]. 2023 Nov 2:2023.10.30.564808. doi: 10.1101/2023.10.30.564808. PMID: 37961642; PMCID: PMC10634951.

Sun M, Stoltzfus A, **McCandlish DM**. A fitness distribution law for amino-acid replacements. *bioRxiv* [Preprint]. 2024 Oct 15:2024.10.11.617952. doi: 10.1101/2024.10.11.617952. PMID: 39464166; PMCID: PMC11507765.

Tang Z, Somia N, Yu Y, **Koo PK**. Evaluating the representational power of pre-trained DNA language models for regulatory genomics. *bioRxiv* [Preprint]. 2024 Sep 25:2024.02.29.582810. doi: 10.1101/2024.02.29.582810. PMID: 38464101; PMCID: PMC10925287.

Thompson M, Martín M, Olmo TS, Rajesh C, **Koo PK**, Bolognesi B, Lehner B. Massive experimental quantification of amyloid nucleation allows interpretable deep learning of protein aggregation. *bioRxiv* [Preprint]. 2024 Oct 1:2024.07.13.603366. doi: 10.1101/2024.07.13.603366. PMID: 39071305; PMCID: PMC11275847.

Zheng X, Venezia M, Blum E, Pedmale UV, Jackson D, Prusinkiewicz P, **Navlakha S**. Reticulate leaf venation in Pilea peperomioides is a Voronoi diagram. *bioRxiv* [Preprint]. 2024 Jul 1. doi: 10.1101/2024.07.01.601217.

2024 New External Funding Awards

Total New Funding	Federal Funding	Private Funding
$3,568,748	$2,935,094	$633,654

PI	Total Award*	Years	Funding Source	Project Tital
Koo, Peter	$235,112	2024-2025	CSHL-Northwell Health	Vision predicting suspicious pancreatic findings using computer vision
Krasnitz, Alexander	$1,225,889	2024-2027	NIH	Computational tools for accurate inference of genetic ancestry from cancer-derived molecular data
Navlakha, Saket Meyer Hannah	$326,542 $72,000	2025	Simons Foundation International	ImmunoAI: Exploring how the immune system solves fundamental machine learning problems
Siepel, Adam Peter Koo†	$1,309,521 $399, 684	2024-2028	NIH	A unified probabilistic model and software implementation for analysis of nascent RNA sequencing data

*Total award including direct costs
† Faculty member is co-investigator

2024 Laboratory Membership

CORE FACULTY

Iossifov Laboratory

[Autism genetics and computational biology]

Ivan Iossifov, PhD - Professor, Principal Investigator	Since 2008
Yoon-Ha Lee,PhD - Research Investigator	Since 2005
Steve Marks, PhD - Computational Science Developer I	Since 2010

Kinney Laboratory

[Machine learning and biophysics in gene regulation]

Justin Kinney, PhD - Associate Professor, Principal Investigator	Since 2009
Andalus Ayaz, BS - Research Technician II	Since 2016
John Desmarais, PhD - Postdoc Computational	Since 2023
Taehoon Ha, MS - Biostatistician	Since 2020
Zhihan Liu, PhD - Postdoc Computation	Since 2024
Kaiser Loell, PhD - Postdoc Computational	Since 2023
Evan Seitz, PhD - Postdoc Computational	Since 2022
Deborah Tenenbaum, PhD. - Postdoc Computational	Since 2022

Koo Laboratory

[Innovation in deep learning for genomic analysis]

Peter Koo, PhD - Assistant Professor, Principal Investigator	Since 2019
Alessandro Crnjar, PhD - Postdoc Computational	Since 2023
Jakub Kacsmarzyk, BS - Graduate Student, Visiting	Since 2021
Yijie Kang, MS - Graduate Student, SBU	Since 2023
Masayuki Nagai, BA - Graduate Student, SBU	Since 2023
Chandana Rajesh, BS - Graduate Student, SBU	Since 2022
Kaeli Rizzo, BS - Graduate Student, CSHL	Since 2022

2024 Laboratory Membership

James Rouse, MS - Computational Science Analyst I	Since 2019
Anirban Sarkar, PhD - Postdoc Computational	Since 2023
Brian Schilder, PhD - Postdoc Computational	Since 2024
Raditya Utama, PhD - Manager, Bioinformatics Core Facility	Since 2016
Jessica Zhou, PhD - Postdoc Computational	Since 2023

Krasnitz Laboratory

[Cancer genomics and genetics ancestry research]

Alexander Krasnitz, PhD - Research Professor, Principal Investigator	Since 2005
Pascal Belleau, PhD - Research Investigator	Since 2017
Xintong Li, BS - Graduate Student, SBU	Since 2024

Levy Laboratory

[Statistical genetics and computational biology]

Dan Levy, PhD - Associate Professor, Principal Investigator	Since 2007
Michael Gleyzer, MS - Computational Science Developer II	Since 2024
Matthew Moss, PhD - Postdoc Computational	Since 2024

McCandlish Laboratory

[Protein evolution and quantitative modeling]

David McCandlish, PhD - Associate Professor, Principal Investigator	Since 2017
Bryan Gitschlag, PhD - Postdoc Computational	Since 2021
Carlos Martí Gomez Aldaravi, PhD - Postdoc Computational	Since 2021
Mengyi Sun, PhD - Postdoc Computational	Since 2023

2024 Laboratory Membership

Meyer Laboratory

[Immunological systems and computational biology]

Hannah Meyer, PhD - Assistant Professor, Principal Investigator	Since 2019
Sarah Chapin, AB - Computational Science Developer I	Since 2020
Todor Cvetanovic, MS - Graduate Student, CSHL	Since 2023
Alexander Grosh, BS - Volunteer	Since 2024
Rishvanth Kaliappan Prabakar, PhD - Postdoc Computational	Since 2023
Alexandra Kiedrowski - Visiting Undergraduate Student	Since 2024
Madison Lapine, BS - Research Technician I	Since 2024
Yong Lin, PhD - Postdoc Computational	Since 2022
Deena Netz, BS - Research Technician I	Since 2024
Lijuan Sun, PhD - Visiting Scientist	Since 2023
Maha Syed, BS - Graduate Student, CSHL	Since 2022
Yunxin Xie, BS - Graduate Student, CSHL	Since 2023

Navlakha Laboratory

[Algorithms in nature and neuro-inspired AI]

Saket Navlakha, PhD - Associate Professor, Principal Investigator	Since 2019
Amitava Banerjee, PhD - Postdoc Computational (shared with Meyer)	Since 2023
Jonathan Zhang, BS - Volunteer	Since 2024
Xingyu Zheng, BS - Graduate Student, CSHL	Since 2020

Siepel Laboratory

[Computational genomics and evolution]

Adam Siepel, PhD - Professor, Principal Investigator	Since 2014
Rebecca Hassett, MS - Computational Science Developer II	Since 2022
Nurdan Kuru, PhD - Postdoc Computational	Since 2024

2024 Laboratory Membership

Luiz Machado, MS – Graduate Student, SBH	Since 2022
Stephen Stalinski, BS – Graduate Student, CSHL	Since 2022
Jiawei Xing, PhD – Postdoc Computational	Since 2024
Alexander Xue, PhD – Postdoc Computational	Since 2018
Xin Zeng, PhD – Postdoc Computational	Since 2024

ADMINISTRATIVE SUPPORT

Katherine Brenner, BS – Assistant Director of Administration	Since 2018
Susan Fredericks, BA – Sr. Scientific Administrator & Assistant to the Chair	Since 2013
Antonia Little, BA – Sr. Scientific Administrator	Since 2023

Cold Spring Harbor Laboratory
Simons Center for Quantitative Biology

One Bungtown Road
Cold Spring Harbor, NY 11724

Cold Spring Harbor Laboratory
Simons Center for Quantitative Biology

cshl.edu/research/quantitative-biology/

SCQB@cshl.edu
facebook.com/cshlnews/
linkedin.com/schoool/cold-spring-harbor-laboratory
@CSHL