11th Annual QCBio Retreat

Monday, September 22, 2025

9:00am – 6:00pm

Hershey Hall & Courtyard

QCBio

10100O11011010

PROGRAM

UCLA Institute for Quantitative & Computational Biosciences

Welcome

The Institute for Quantitative and Computational extends a warm Welcome to all. In its essence, QCBio is a community of scholars, researchers, trainees and educators: we are not only working in Computational Biology and Bioinformatics, but catalyze an ongoing revolution of biological and biomedical sciences towards data-richer capabilities and more quantitatively predictive insights.

UCLA provides five graduate research training opportunities in quantitative and computational biosciences (https://qcb.ucla.edu/overview/), namely in Bioinformatics, Medical Informatics, Systems Biology, Biomathematics, and Genetics&Genomics. While each has a distinct identity, coordination benefits all.

The QCBio Collaboratory (https://qcb.ucla.edu/collaboratory/) is a postdoctoral training program of computational biology postdocs extending their teaching and collabroation skills. In turn, Collaboratory Fellows have provided essential training through a rich workshop series. This has benefitted thousands of UCLA researchers, and is able to prepare an increasing population of undergraduates to be "research-ready".

Indeed, more and more undergraduates are flocking to the Computational Biology Major and Minors (https://casb.ucla.edu/) and the Bioinformatics Minor (https://bioinformatics.ucla.edu/undergraduate-bioinformatics/). A majority of these talented undergraduates are involved in research and contribute to projects. In fact UCLA has become the primary producer of applicants to top Comp Bio PhD programs. We maintain an Undergraduate Research Portal (https://qcb.ucla.edu/research-portal/) to connect potential mentors with eager undergraduate students – we encourage you to post your projects there.

Mentoring is indeed a key component of our QCBio community culture. Now in its 11th year we hosted 36 students within the B.I.G. Summer Undergraduate Research Program with 27 laboratories participating! (https://qcb.ucla.edu/big-summer/big2025/). BIG Thank Yous to all faculty, postdoc, graduate student mentors! Given federal funding constraints B.I.G. Summer has morphed from outreach to focus primarily UCLA undergraduates providing them the summer opportunity to beef up their research portfolio. So encourage your undergraduate researchers to apply and make a real contribution to your research project.

Over the past summer we have seen some leadership changes in our programs. We are grateful to the efforts and achievements of past leaders and are excited about the renewed energy that comes with fresh legs and ideas. Our status reports will detail these to inform and invite you to contribute in charting the next phase.

The QCBio Retreat marks the start the new academic year – we invite everyone to contribute to a thriving community. Based on your interests we will host affinity group meetings; career panels for postdocs, grad student or undergrads. In fact we're eager to support your other initiatives and hear from you.

- Nominations for our seminar series?
- Topic areas for symposia or workshops?
- Ideas for supporting graduate students?
- Ideas for supporting postdoctoral fellows?

The recent uncertainties about research funding it is clear that each of us must prioritize our public-facing communication and engagement more. Some of us are more practiced in developing narrative, but all of us can improve – QCBio would like to support these efforts, and we will start at the lunch breakout session !

Special thanks to Caroline Baron for again organizing the QCBio Retreat this year!
Alexander Hoffmann, QCBio Director Kirk Lohmueller, Bioinformatics IDP Director

Agenda

9:00 am **COFFEE, TEA, JUICE, FRUIT, BAGELS**

9:10 am **WELCOME**

9:15 am **STATUS REPORTS I**

- **Alexander Hoffmann**, Director, QCBio
- **Clara Lajonchere,** Deputy Director, Institute for Precision Health
- **Eleazar Eskin,** Executive Director, DataX
- **Matteo Pellegrini**, Director, QCBio Collaboratory

9:45 am **SELECTED TALKS I - chaired by Brunilda Balliu**

- **Timothy Hamilton**, Bioinformatics PhD student, Deeds Lab
- **Matthew Heffel**, Bioinformatics PhD student, Luo Lab
- **Emily Maciejewski**, Computer Science PhD student, Ernst Lab

10:30 am **COFFEE & TEA BREAK**

11:00 am **STATUS REPORTS II**

- **Kirk Lohmueller**, Bioinformatics Interdepartmental Ph.D. Program
- **Roy Wollman**, Director, Systems Biology Ph.D. Home Area
- **William Hsu**, Director, Medical Informatics Ph.D. Home Area
- **Paivi Pajukanta**, Director, Genetic & Genomics, Ph.D. Program
- **Brunilda Balliu**, Biomathematics, Ph.D. Program

11:30 am **DISTINGUISHED SPEAKER – introduced by Pavak Shah**

- **Arjun Raj**, Richard K. Lubin professor of Bioengineering, Professor of Genetics, University of Pennsylvania

12:20 pm **LUNCH + BREAKOUTS: Why it matters**

1:30 pm **QCBio SCIENCE SLAM – Why it matters**

2:00 pm **INVITED TALK – introduced by Alexander Hoffmann**

- **Jonathan Herman,** Assistant Professor-in-Residence, Division of Infectious Diseases UCLA David Geffen School of Medicine

2:30 pm **SELECTED TALKS II – chaired by Xia Yang**

- **Clara Frydman-Gani**, Bioinformatics PhD student, Olde Loohuis Lab
- **Fei-Man Hsu**, Postdoc, Pellegrini Lab

3:00 pm **COFFEE & TEA BREAK**

3:30 pm **SELECTED TALKS III – chaired by Grace Xiao**

- **Ye Wang,** Bioinformatics PhD student, Alber Lab
- **Michael Wasney,** Genetics & Genomics PhD student, Garud Lab
- **Haripriya Vaidehi Narayanan**, Postdoc, Hoffmann Lab

4:15 pm **STATUS REPORTS III**

- **Eric Deeds**, Director, Life Science Math Core
- **Noa Pinter-Wollman**, Director, Computational & Systems Biology Major
- **Sriram Sankararaman**, Director, Bioinformatics Minor
- **Alexander Hoffmann**, BIG Summer
- **QBio-EDGE**

Postdoctoral QCBio Research Excellence Awards

4:45 pm **RECEPTION + POSTER SESSION**

Distinguished Speaker

Thanks to the generous support of an anonymous donor, QCBio is excited to launch a new Distinguished Speaker Series**.** This series will bring visionary scientists from around the world to UCLA to share their work, deliver two seminars, and engage with our community, including a celebratory faculty dinner. A seminar committee guides the selection of QCBio Distinguished speakers

We are honored to welcome **Dr. Arjun Raj** as our inaugural Distinguished Speaker, joining us today at the QCBio Retreat. His second lecture will be on Tues Sept 23, 11am in Boyer 159

Arjun Raj
Richard K. Lubin Professor of Bioengineering
Professor of Genetics
University of Pennsylvania

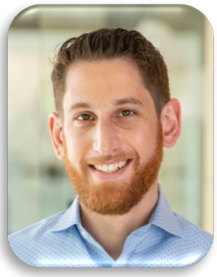
Link to biosketch

Arjun grew up in Ithaca, New York, then went to UC Berkeley for his undergraduate education, where he majored in math and physics. He then earned his PhD in mathematics from the Courant Institute at NYU, followed by postdoctoral training at MIT before joining the faculty at Penn in 2010. He is currently a Professor of Bioengineering and Professor of Genetics. His research focuses on the development and application of experimental techniques for making quantitative measurements in single cells and models for linking those measurements to cellular function. His ultimate goal is to achieve a quantitative understanding of the molecular underpinnings of cellular behavior.

Can a single cell learn?

Are cells nothing more than automatons, destined to follow their built-in programs without question? Or do they have the ability to encode past experiences and learn from them, allowing them to flexibly respond to new, unanticipated challenges? We present data documenting the ability of cells to learn, both cancer cells in the context of therapy resistance and macrophages in the context of trained immunity. We find the molecular mechanisms by which such regulatory flexibility is achieved. We believe cellular learning is a critical means by which cells are able to cope with the seemingly infinite complexity of their environments.

Invited Speaker



Jonathan D. Herman MD PhD
Assistant Professor-in-Residence
Division of Infectious Diseases
UCLA David Geffen School of Medicine

Jonathan D. Herman, M.D., Ph.D., is a physician-scientist who investigates how the human immune system protects itself against infectious diseases. His research focuses on understanding vaccine-induced and natural immunity to pathogens like malaria, with the goal of developing new vaccines and therapies to prevent and treat infections. Herman employs advanced systems biology techniques to study how the human humoral immune system — specifically antibodies and B cells — responds to infections. His research program centers around analyzing these immune responses to develop effective vaccines and treatments for infectious diseases such as malaria, COVID-19 and HIV. He utilizes and develops systems-level tools to understand the human humoral immune system and then engineer it to protect patients from infectious disease. His current methodologies include systems serology to study Fc- and Fab-driven antibody responses; phage-display immunoprecipitation sequencing (PhIP-Seq) to measure whole-proteome anti-pathogen antibody responses; and the development of new antibody and B-cell technologies to interrogate antigen-specific responses at a massive scale. By integrating these findings, Herman aims to inform the design of next-generation vaccines and immunotherapies that can provide robust protection against malaria and other infectious diseases, ultimately improving global health outcomes.

Systems Antibody Immunology to Inform Malaria Vaccine Design

How do you develop a vaccine? Past success in vaccine development have relied on iterative trial-and-error. When they have worked, humanity has benefited. When they haven't, we've been left without a clear understanding of why. In the Herman Antibodyomics Lab we use systems biology approaches to study the human antibody response and rationally design novel vaccines. In order to design a better vaccine for malaria, a protozoan pathogen that infects over 200 million people a year, we develop a malaria phage display immunoprecipitation assay (PhIP-Seq) tool that enables us to measure antibody responses to the entire malaria proteome. By applying this PhIP-Seq tool we call *Plasmoscan* to a live cell vaccine, we have identified a novel form of immune evasion we large implication for how we design future malaria vaccines.

Selected Talks

- **Large Language Models for Psychiatric Phenotype Extraction from Electronic Health Records**

Clara Frydman-Gani[1], Alejandro Arias[2], Maria Perez Vallejo[3], John Daniel Londoño Martínez[3], Johanna Valencia-Echeverry[3], Mauricio Castaño[2], Alex A. T. Bui[4], Nelson B. Freimer[1], Carlos Lopez-Jaramillo[3], Loes M. Olde Loohuis[1]
[1] Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, UCLA, LA, CA, USA
[2] Department of Mental Health and Human Behavior, University of Caldas, Manizales, Colombia
[3] Research Group in Psychiatry (GIPSI), Department of Psychiatry, School of Medicine, UdeA, Medellin, Colombia
[4] Department of Radiological Sciences, UCLA, LA, CA, USA

The accurate detection of clinical phenotypes from electronic health records (EHRs) is pivotal for advancing large-scale genetic and longitudinal studies in psychiatry. Clinical notes are an essential source of symptom-level information, particularly in psychiatry; however, the automated extraction of symptoms from text remains challenging. Here, we evaluated 11 large language models (LLMs) for their ability to detect 109 psychiatric phenotypes from Spanish-language clinical text, using EHR notes from a clinic in Colombia. LLMs were evaluated both "out-of-the-box" and after fine-tuning, and compared against a traditional natural language processing (tNLP) method developed from the same data. To generate a publishable fine-tuned LLM, we created a synthetic dataset free of PHI but containing original annotations. We fine-tuned an LLM on this data, creating "Mistral-small-psych", which can detect psychiatric phenotypes with performance comparable to that of LLMs trained on real EHR data (macro-F1=0.79). Our study underscores the value of domain-specific adaptation of LLMs and introduces a new model for accurate phenotyping in Spanish text.

- **Minority Reporting: Leverage and its effects on the analysis of scRNA-seq data**

Timothy Hamilton[1], Juan Vergara Najar[2], Eric J. Deeds[3,4]
[1] Bioinformatics Interdepartmental Program,
[2] Department of Computation and Systems Biology,
[3] Department of Integrative Biology,
[4] Institute of Quantitative and Computational Biology,
University of California Los Angeles, Los Angeles California, USA

The advent of single cell methods has revolutionized the fields of developmental and systems biology by enabling the investigation of complex tissues at the level of individual cells. Supporting these advancements is the fundamental idea that the gene expression of single cells can reveal distinct expression profiles that correspond to biologically distinct and meaningful groups. Unfortunately, recent work belies that notion; if a small percentage of cells is removed from the dataset and the typical analysis pipeline is repeated, the partitioning of remaining cells is significantly changed. We developed a new geometric method to determine whether the unique geometric structure of scRNA-seq data is the cause. Our findings imply that the analysis pipeline used to analyze scRNA-seq data are extremely sensitive to dataset composition, implying that the insights generated by the analysis of single cell data may not be generalizable beyond the original experiment.

Single-cell Epigenomic Atlasing of the Developing Human Cortex, Hippocampus, and Basal Ganglia

Matthew G. Heffel[1,2], Jingtian Zhou[3,4], Yi Zhang[3], Dong-Sung Lee[5], Kangcheng Hou[1,2], Oier Pastor-Alonso[6], Kevin D. Abuhanna[3], Joseph Galasso[7], Colin Kern[3], Chu-Yi Tai[7], Carlos Garcia-Padilla[3], Mahsa Nafisi[7], Yi Zhou[3], Anthony D. Schmitt[8], Terence Li[3], Maximilian Haeussler[9], Brittney Wick[3], Martin Jinye Zhang[3], Fangming Xie[3], Ryan S. Ziffra[10,11,12,13], Eran A. Mukamel[14], Eleazar Eskin[1,2,15], Tomasz J. Nowakowski[7,11,12,13], Jesse R. Dixon[5], Bogdan Pasaniuc[1,2,15], Joseph R. Ecker[5,16], Quan Zhu[3], Bogdan Bintu[17,18,19], Mercedes F. Paredes[7,13,20], Chongyuan Luo[1]

1. Department of Human Genetics, University of California Los Angeles, Los Angeles, CA 90095, USA
2. Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA 90095, USA
3. Center for Epigenomics, Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093, USA
4. Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, CA 92093, USA
5. Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA
6. Department of Neurosciences, University of California, San Diego, La Jolla, CA 92037, USA
7. Department of Neurology, University of California, San Francisco, San Francisco, CA 94158, USA
8. Arima Genomics, San Diego, CA 92121, USA
9. Genomics Institute, University of California Santa Cruz, Santa Cruz, CA 95064, USA
10. Department of Neurological Surgery, University of California San Francisco, San Francisco, CA 94158, USA
11. Department of Anatomy, University of California San Francisco, San Francisco, CA 94158, USA

12. Department of Psychiatry and Behavioral Sciences, University of California San Francisco, San Francisco, CA 94158, USA
13. Eli and Edythe Broad Center for Regeneration Medicine and Stem Cell Research, University of California San Francisco, San Francisco, CA 94158, USA
14. Department of Cognitive Science, University of California, La Jolla, CA 92037, USA
15. Department of Computational Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA
16. Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA
17. Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
18. Department of Developmental Biology, Stanford University, Stanford, CA 94305, USA
19. ChEM-H Institute, Stanford University, Stanford, CA 94305, USA
20. Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA 94158, USA

Large scale single cell atlasing projects of the developing human brain are becoming highly relevant in neuroscience research, yet largely exist only in transcriptomic space and have primarily focused on cortical regions. Here we investigated the epigenomic and three-dimensional chromatin conformational reorganization during the development of the prefrontal cortex and several basal ganglia regions including the hippocampus, striatum, globus pallidus, and more. Sequencing more than 200,000 joint single-nucleus profiles of chromatin conformation and DNA methylation generated by single-nucleus methyl-3C sequencing (snm3C-seq3) we are able to reconstruct several developmental cell lineages and explore the directionality of epigenomic changes across modalities. The inclusion of the ganglionic eminence allows for the complete trajectory dissection of inhibitory neurons before, during, and after regional migration. Using single-cell profiling and multimodal single-molecule imaging approaches, we have found that short-range chromatin interactions are enriched in neurons, whereas long-range interactions are enriched in glial cells and non-brain tissues. We reconstructed the regulatory programs of cell-type development and differentiation, finding putatively causal common variants for schizophrenia strongly overlapping with chromatin loop-connected, cell-type-specific regulatory regions. Our data provide multimodal resources for studying gene regulatory dynamics in brain development and demonstrate that single-cell three-dimensional multi-omics is a powerful approach for dissecting neuropsychiatric risk loci.

- **First-year post-transplant respiratory DNA methylation profile predicts chronic lung allograft dysfunction**

Fei-Man Hsu[1,2], John A. Belperio[3], S. Samuel Weigt[3], Vyacheslav Palchevskiy[3], Matteo Pellegrini[1,2,*], Joanna M. Schaenman[3,*]
[1] Department of Molecular, Cell and Developmental Biology,
[2] Institute for Quantitative and Computational Biosciences – The Collaboratory,
[3] Department of Medicine,
University of California Los Angeles, Los Angeles California, USA

Lung transplant (LTx) improves the quality and duration of life for patients with severe lung disease. Chronic lung allograft dysfunction (CLAD) occurs in approximately 50% of LTx recipients, leading to significant lung allograft morbidity and mortality. However, there is a lack of an early-detection method for CLAD. This study aims to characterize DNA methylation, and epigenetic modification, as a biomarker for the early detection of CLAD onset. We collected longitudinal bronchoalveolar lavage (BAL) samples from a multi-center cohort of 155 patients who underwent LTx over a 3-year follow-up period, and profiled DNA methylation using targeted bisulfite sequencing (TBS-seq). Our results showed that BAL DNA methylation could predict CLAD outcome accurately even within the first year post-LTx, and the derived epi-CLAD score is highly associated with time-to-CLAD, suggesting it could serve as a new modality to characterize patients at high risk of CLAD.

- **Computational approaches for expanding species and genomic coverage in DNA methylation array samples**

Emily Maciejewski[1,2,3], Jason Ernst[1,2,3]
[1]Department of Computer Science, University of California, Los Angeles, Los Angeles, CA 90095, USA
[2]Department of Computational Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA
[3]Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA

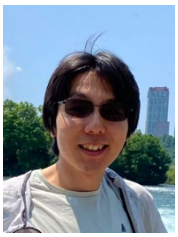
Emerging large compendia of DNA methylation (DNAm) data such as methylation array data for over 15,000 samples from over 300 species and about 60 tissue types from the Mammalian Methylation Consortium and whole genome bisulfite (WGBS) data for over 600 human reference epigenomes from the International Human Epigenome Consortium (IHEC) are presenting new analytical opportunities and challenges. The cross-species dataset while covering many species and tissues only covers a small percentage of potential species-tissue combinations. Thus, we develop CMImpute (Cross-species Methylation Imputation) to impute DNAm samples representing currently unavailable species and tissue combinations. The IHEC WGBS data provides the opportunity for imputation of unmeasured CpGs in samples profiled on methylation arrays, typically used in human epigenome-wide association studies. To address this, we develop KNN and regression-based reference based imputation methods. Both methods yield strong correlation with held-out data and we expect to be useful in epigenetic analyses.

- **Topological measures on phylogenetic trees of the antibody repertoire reveal the underlying kinetics of B-cell fate decisions**

Haripriya Vaidehi Narayanan[1,5 #], Chengyuan Li [1,2 *], Jianche Liu[3 *], Mark Y. Xiang[1,4], Helen Huang[1,4], Alexander Hoffmann[1,5 #]
1 Institute for Quantitative and Computational Biosciences, UCLA
2 Department of Molecular, Cell, and Developmental Biology, UCLA
3 Zhejiang University-University of Edinburgh Institute, Zhejiang University, China
4 Bioinformatics Interdepartmental Program, UCLA
5 Department of Microbiology, Immunology, and Molecular Genetics, UCLA
* These authors contributed equally
# Corresponding authors

Antibodies arise from a Darwinian process of mutation, affinity-dependent selection, and expansion. However, when the underlying B-cell fate decisions are altered by SNPs or chronic inflammation, the antibody repertoire may be impaired. Can we interpret the antibody repertoire to pinpoint what aspects if B-cell regulation were perturbed? We developed a mathematical model of antibody generation via B-cell survival, proliferation, and mutation under antigen-dependent selection and performed Monte Carlo simulations to produce phylogenetic trees. Our findings reveal that purely topological measures on the forest of trees are sensitive to mutation rates, while measures weighted by sequence frequency reflect positive selection. We showed that parameters are independently identified with accuracy 70-95% given perfect data, but this is drastically diminished by B-cell sampling losses above 50%. We suggest experimental guidelines, provide confidence intervals on inferred cell fate parameters, and a model-free approach for comparison across individuals. Our concept derives B-cell regulatory control from the end-point antibody repertoire, as a potential diagnostic measure informing personalized vaccination.

- **Studying the Role of Subnuclear Positioning in Gene Expression**



Ye Wang[1,2,3], Lorenzo Boninsegna[1,2], Yuxiang Zhan[1,2,4], Xianghong Jasmine Zhou[1,3,*], Frank Alber[1,2,*]
[1]Institute of Quantitative and Computational Biosciences (QCBio), University of California Los Angeles, Los Angeles, CA 90095, USA
[2]Department of Microbiology, Immunology, and Molecular Genetics, University of California Los Angeles, 520 Boyer Hall, Los Angeles, CA 90095, USA
[3]Department of Pathology, David Geffen School of Medicine, University of California Los Angeles, 10833 Le Conte Ave, Los Angeles, CA 90095, USA
[4]Department of Quantitative and Computational Biology, University of Southern California, 1050 Childs Way, Los Angeles, CA 90089, USA

The spatial organization of the genome is crucial to gene regulation. To assess how subnuclear context influences expression, we analyzed genome-wide relationships between nuclear positioning and transcription at single-cell resolution. Using 3D genome structures from the Integrative Genome Modeling (IGM) framework, we classified highly expressed genes by nuclear microenvironment (nuclear position, speckle distance, structural features). Two categories emerged: Class I genes (≈90%, including 73% of housekeeping genes) were speckle-associated, nuclear-interior, and positionally stable. Class II genes (≈10%) showed minimal speckle association, high positional variability, greater length, and favored long-range over nearby enhancers. Some housekeeping genes fell into Class II, challenging the conventional view of stable nuclear positioning. Expression changes between stem and fibroblast cells also correlated with microenvironment shifts. In summary, our findings reveal a complex relationship between subnuclear positioning and gene expression, influenced by the chromatin region's genomic properties.

- **Genetic homogeneity along the tract of the gut**

Michael W Wasney[1], Leah Briscoe[2], Ricky Wolff[3], Hans Ghezzi[4], Carolina Tropini[4,5,6], Nandita Garud[1,2,3]
[1]University of California, Los Angeles, Human Genetics, Los Angeles, CA, [2]University of California, Los Angeles, Interdepartmental Program in Bioinformatics, Los Angeles, CA, [3]University of California, Los Angeles, Ecology and Evolutionary Biology, Los Angeles, CA, [4] University of British Columbia, Microbiology and Immunology, Vancouver, Canada, [5]University of British Columbia, School of Biomedical Engineering, Vancouver, Canada, [6]Canadian Institute for Advanced Research, Humans and the Microbiome Program, Toronto, Canada

Environmental gradients in the gastrointestinal tract drive spatial variation in the abundance and membership of bacterial species along the gut. Less is known about the spatial distribution of within-species bacterial genetic variation, which confers important functions. We analyzed genetic diversity of approximately 30 species at different sites along the guts of humanized mice inoculated with the same human stool. Unlike species composition, genetic diversity was uniformly distributed across different gut sites within hosts, due to multiple genetically distinct intraspecific strains coexisting at similar frequencies along the gut. Additionally, roughly 60 evolutionary changes arising within mice showed no spatial specificity. Genetic variation detected in metagenomic samples collected along the human guts revealed similar spatial uniformity, indicating potential generalizability of these findings to the natural human microbiome. Our results imply that species presence-absence may play a larger role than genetic variation in responding to spatially varied environmental pressures in the gut microbiome.

Collaboratory Fellows 2025-2026

Matteo Pellegrini, **Director**

Eloy Lopez, **Program Manager**

Fei-Man Hsu

Daniel Ha

Xiaolu Guo
New Fellow

Amelia Palermo
New Fellow

Ivy Xiong

Giorgia Del Vecchio

Shreyas Rajesh
New Fellow

Lukasz Salwinski

Seyoon Ko

Xingbo Shang
New Fellow

Montgomery Blencowe

Weihong Yan

Maya Weissman
New Fellow

Xianglong Tan

Shawn Cokus

Haripriya Vaidehi Narayanan
New Fellow

Lingyu Zhan

Nick Wiltsie
New Fellow

Karolina
Kaczor-Urbanowicz

Fangming Xie
New Fellow

https://qcb.ucla.edu/collaboratory/people/

Please visit our website to learn more about the Collaboratory, our classes offered, and class schedule and of course to learn more about our Postdoctoral Fellows.

http://qcb.ucla.edu/collaboratory

Meet our new Bioinformatics students!

Yashas Appaji
Georgia Tech

Ronan Bennet
UCLA

Darin Boyes
Carnegie Mellon University

Claire Hsieh
University of California, Davis

Sakin Kirti
Case Western Reserve University

Lily Kuhn
Bernard College

Calvin Lee
UC San Diego

Zhiyin Liu
Hong Kong University of Science and Technology (HKUST)

Anna McDonald
Washington University

Shreya Nakhawa
Northeastern University

Elliot Outland
University of Texas Southwestern

Praveena Ratnavel
UCLA
B.I.G. SUMMER 2024 ALUMNA

Ziqi Rong
University of Michigan

Brian Schweitzer Wang
University of Michigan, Ann Arbor

Qingru Xu
Harvard University

Meet our new Medical Informatics students!

Rohil Ahuja
Washington University St. Louis

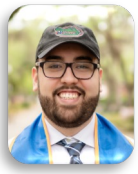
Joey Sungjoo Han
Korea Advanced Institute of Science and Technology (KAIST)

Jiayuan (Alice) Wang
University of California, Irvine

Meet our new Systems Biology students!

Ruoshui (Rainy) Liu
UCLA
B.I.G. Summer 2023 alumna

Nithish Narasimman
UC San Diego



Ricardo Roure
University of Florida
B.I.G. Summer 2024 Alumnus

Neel Tangella
USC

Miles Tran
New York University

Rohan Vanheusden
UC San Diego

Meet our new Genetics & Genomics students!

Adam Alnihmy
UC San Diego

Alfred Kibowen
Amherst College

Amman Klair
UC San Diego

Esaria Oliver
Eckerd College

Shivani Patel
UCLA

Varsha Rajesh
UC San Diego

Nicholas Shamlin
UCLA

Meet our new Biomathematics students!

Stephen Dorn
University of Wisconsin-Madison

Lotem Efrat
University of North Carolina at Chapel Hill
B.I.G. Summer 2024 Alumna

Sirui Li
Southern University of Science and Technology

Rilyn McKallip
University of Richmond

Eric Sun
Georgia Institute of Technology

QCBio Postdoctoral Excellence Award

QCBio proudly recognizes postdoctoral fellows nominated by their PIs for their outstanding research, creativity, and contributions to our community.

The QCBio Research Excellence Awards, created in 2024, are awarded to postdoctoral fellows selected from nominations by QCBio faculty. It is intended to complement and encourage nominations to the UCLA Chancellor's Postdoctoral Excellence Awards.

✨2025 Awardees✨

Pan Liu, PhD, since early 2024 postdoctoral scholar in the laboratory of Jessica Li, has already left her mark with two manuscripts in press, describing a principled method for grouping notoriously sparse single-cell RNAseq data into meta-cells for more quantitative downstream analysis. Appropriately called mcRigor, this method has been recognized at RECOMB2025 and an NHGRI TDCC Opportunity Fund grant. As a QCBio Collaboratory Fellow Dr. Liu's scRNAseq workshops have been particularly popular.

Chenghao (Trevor) Zhu, PhD, postdoctoral scholar in the laboratory of Paul Boutros 2020-2025, will begin his independent career at Sanford Burnham Prebys Medical Discovery Institute. He developed new methods in proteogenomics, critical for the identification of peptides and neo-epitopes that fuel novel cancer therapies. His landmark paper in Nature Biotechnology in 2025 has cracked opened the field and led to almost a dozen collaborative papers. He has excelled in team work and mentoring including BIG Summer.

Haripriya Vaidehi Narayanan, PhD, James S McDonnel Foundation postdoctoral fellow, and then Damon Runyon Quantitative Biology postdoctoral, has pioneered a whole new research area for the Hoffmann lab, to develop the tools that enable precision or personalized vaccination. She first recruited and mentored two Bioinformatics PhD students, and has included numerous undergraduates some of whom have gone to prestigious PhD programs. She has published first author work in PNAS and Cell Systems, and enabled studies in Molecular Systems Biology, Cell Reports, and Cell Death and Disease. She was instrumental in securing funds from the UCLA Immunology initiative and a fundable score in an R01 proposal. As a QCBio Collaboratory fellow, her image analysis workshops have provided valuable training to appreciative participants.

Poster Session

1. **Synonymous and Non-synonymous Variants Empower Deep Learning to Classify Selective Sweeps in Ancient Human DNA**

Brendan Aeria[1], Mariana Harris[2,3], Maya Weissman[3], Nandita Garud[3,4]
[1] BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
[2] Department of Computational Medicine, UCLA
[3] Department of Ecology and Evolutionary Biology, UCLA
[4] Department of Human Genetics, UCLA

Beneficial mutations arise and rapidly spread through a population to high frequency in a process known as selective sweeps. Sweeps fall into two categories: hard, if the beneficial mutation originates from a single source; soft, if the mutation was already present on multiple genetic backgrounds. Distinguishing between these scenarios in ancient human DNA (aDNA) can therefore reveal the primary mode of adaptation in human history. However, large amounts of missing data and complex demography hinder this classification. While existing methods, including deep learning, have shown promise, they often overlook a rich source of information: the distinct evolutionary patterns of synonymous versus non-synonymous mutations. By comparing a CNN with access to this information to one without, we aim to demonstrate the value of synonymous versus non-synonymous patterns and improve deep learning-based classification of selective sweeps. This could clarify our evolutionary history while offering a powerful new method for other species.

2. **Techniques for deriving cell type selectivity of multivalent ligands**

Samuel Amidon[1], Helen Kaidantzis[2], Emily Lin[1], Brian Orcutt-Jahns[1], Aaron Meyer[1]
[1] Department of Bioengineering,
[2] Department of Computational and Systems Biology,
University of California Los Angeles, Los Angeles California, USA

Efforts to enhance or thwart immune function by systemic administration of cytokines suffer widely from off target binding and adverse effects. Here, we explore an approach to combat these limitations: multivalent molecules containing unique ligands for both cell type-targeting and signal delivery. As a preliminary predictor of a molecule's ability to selectively engage a cell type, we quantify differences in receptor distributions of target and off target cells using Kullback-Liebler divergence and Earth Mover's Distance. We then apply a general multivalent binding model to obtain a direct quantification of selectivity. The model predicts how much receptor will be bound by a multivalent ligand, and when combined with single cell receptor data, it can predict target and off target binding and thus selectivity. We ultimately use these methods to guide the selection of ligands for multivalent formats that selectively deliver signaling molecules to any desired cell type.

3. **Learning prodromal Parkinson's disease risk from large-scale wearable accelerometry**

Aakarsh Anand[1], Ulzee An[1], Yuzhe Yang[1,2], Sriram Sankararaman[1,2,3,4]
[1] Department of Computer Science
[2] Department of Computational Medicine
[3] Department of Human Genetics
[4] Bioinformatics Interdepartmental Graduate Program
University of California Los Angeles, Los Angeles California, USA

Parkinson's disease (PD) is typically diagnosed years after neurodegeneration has begun, underscoring the need for scalable tools that detect early motor changes. We developed deep learning models using week-long wrist accelerometer recordings from ~100,000 UK

Biobank participants to predict future PD diagnoses. Our convolutional ResNet architecture, trained on diagnosed cases and controls, generalized effectively to an unseen cohort including prodromal individuals—those who were not yet diagnosed at the time of data collection but developed PD later. Compared to logistic regression on handcrafted activity features, the learned accelerometer embeddings significantly improved precision in identifying prodromal cases under realistic prevalence (<1%). Notably, predictive performance increased for subjects closer to diagnosis, achieving the highest accuracy within two years of disease onset. These results demonstrate that representation learning on raw sensor data can uncover subtle motor signatures of prodromal PD, highlighting the promise of wearable-based digital biomarkers for early detection.

4. **Probabilistic circuits enable practical and robust artificial genome generation while improving imputation accuracy in private populations**

Prateek Anand[1], Anji Liu[2], Meihua Dang[3], Boyang Fu[4], Xinzhu Wei[5], Guy Van den Broeck[1*], and Sriram Sankararaman[1,6,7*]
[1] Department of Computer Science, UCLA, Los Angeles, CA, USA
[2] Department of Computer Science, National University of Singapore, Singapore
[3] Department of Computer Science, Stanford University, Stanford, CA, USA
[4] Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
[5] Department of Computational Biology, Cornell University, Ithaca, NY, USA
[6] Department of Human Genetics, UCLA, Los Angeles, CA, USA
[7] Department of Computational Medicine, UCLA, Los Angeles, CA, USA
[*] These authors contributed equally.

Artificial genomes (AGs) are valuable for benchmarking pipelines and building reference panels without privacy risks, but generating realistic AGs efficiently remains challenging. We introduce a deep generative model based on hidden Chow-Liu trees (HCLTs) represented as probabilistic circuits (PCs). PCs capture long-range SNP dependencies, enable efficient inference, and require tuning only one hyperparameter. On the 1000 Genomes Project, PCs achieve higher held-out log-likelihood than HMMs, GANs, and RBMs while matching ability in reproducing allele frequencies, linkage disequilibrium, and population structure. As reference genomes, PC AGs improve imputation accuracy, especially for underrepresented populations: 7.6% over European-only panels and 9.7% over the next best method (<1% MAF: 83.9%). Combining PC AGs with European genomes further boosts accuracy (15.5% over real European genomes). Results replicate in African ancestry data and UK Biobank cohorts. PCs also preserve privacy, offering a practical, robust method for simulating AGs and inferring rare variants.

5. **Allele-specific alternative polyadenylation modulates RNA binding and explains disease-associated non-coding variants in the human brain**

Ryan M. Barney[1], Giovanni Quinonez-Valdez[2], Alison J. King[1], Kofi Amoah[1], Weijian Wang[1], and Xinshu Xiao[1,2]
[1]Bioinformatics Interdepartmental Program
[2]Department of Integrative Biology and Physiology
University of California Los Angeles, Los Angeles California, USA

Non-coding genetic variants are linked to human diseases, but their regulatory mechanisms remain unclear. One mechanism, alternative polyadenylation (APA), produces mRNA isoforms with distinct 3'UTRs and post-transcriptional fates. We analyzed allele-specific APA using RNA-seq from 1,047 postmortem brain samples across four regions in the Mount Sinai Brain Bank. Applying the ASARP method, we identified 9,195 asAPA events across 1,040 genes, many involved in vesicle trafficking and membrane pathways disrupted in Alzheimer's disease. A concordance-based framework

prioritized 732 high-confidence asAPA events, enriched for altered RNA-binding protein interactions, notably with FMRP. In FXS brains, FMRP dysregulation corresponded with widespread 3'UTR shortening. These SNPs were also enriched in aQTLs and GWAS loci for AD and autism, with several asAPA SNPs showing disease-specific allelic usage in AD. Our results highlight asAPA as a key post-transcriptional regulatory mechanism linking common genetic variants to RBP binding, 3'UTR isoform usage, and neurodegenerative disease risk.

6. **Lipid metabolic heterogeneity of glioma cellular states reveals environmental dependencies linked to cellular identity**

Nicholas A. Bayley[1,2], Jenna Minami[1], Aparna Bhaduri[3], Thomas G. Graeber[1,4*], and David A. Nathanson[1,4*]
[1]Department of Molecular and Medical Pharmacology,
[2]Bioinformatics Interdepartmental Program,
[3]Department of Biological Chemistry,
[4]Jonsson Comprehensive Cancer Center, University of California Los Angeles, Los Angeles California, USA
* equal contribution from labs

Gliomas are lethal malignancies composed of dynamic cellular states resembling normal neurodevelopmental cell types or responses to stressors within the tumor microenvironment (TME). To explore how metabolic requirements may regulate glioma state heterogeneity, we conducted multi-omic analysis across over 300 glioma specimens, including patient tumors, orthotopic xenografts, and gliomasphere cultures. Single-cell sequencing of a diverse subset of gliomas integrated with bulk transcriptome and lipidome profiling revealed cellular states associated with specific lipid metabolic profiles. Comparison of matched patients and models showed that non-native environments with new metabolic requirements drive adaptation towards distinct states. While tumors with glial lineage identity demonstrated metabolic plasticity, other states were tied to high or low *de novo* lipid synthesis capacity. Reduced *de novo* lipid synthesis in tumors with neuronal/oligodendroglial identity promoted reliance on exogenous lipid scavenging within the brain TME for survival. Together these results identify metabolic plasticity and liabilities linked to glioma state diversity.

7. **Balancing Stability and Complexity in Boolean Models of Biological Systems**

Sai Bavisetty[1,*], Matthew Wheeler[2], Reinhard Laubenbacher[2], Claus Kadelka[3]
[1]University of California, Los Angeles
[2]University of Florida, Gainesville
[3]Iowa State University
sbavisetty@ucla.edu

Boolean networks were first introduced by Kauffman as models for gene regulatory networks. These models have gained popularity because of their simplicity and ability to capture the complex behaviors of biological systems. Currently, the repository Biodivine Boolean Models contains more than 230 Boolean network models.

A systematic investigation of these biological models suggests that they are incredibly robust. In particular, they are resilient to perturbations and tend to reach the same phenotype despite small disturbances. An explanation of this phenomenon was first given by Kauffman, who showed empirically that a network's connectivity determines the stability of the Boolean network. This was further expanded on by Derrida, who provided a theoretical explanation for the effect of connectivity. This was succeeded by many works that looked at various features of biological networks to prove empirically the relation between a network parameter and stability. Building upon this foundational understanding of robustness in Boolean networks, our work delves into the intrinsic trade-off between phenotypic complexity and network stability. We extend a conjecture proposed by Willadsen, Triesch, and Wiles by proving that entropy (a proxy for phenotypic complexity) provides a tight asymptotic upper bound for coherence (a measure of stability). As a consequence, we

derive the Pareto frontier between complexity and stability, allowing us to determine exactly how much stability is achievable for any given level of complexity.

8. **Determining cytokine immunosuppressive potential with PARAFAC2-RISE**

Nicole Bedanova[1], Samuel Amidon[1], Aaron Meyer[1]
[1]Department of Bioengineering
University of California Los Angeles, Los Angeles, California, USA

Cytokines regulate immune responses in a cell type and environment-dependent manner, with individual cytokines capable of both suppressive and inflammatory signaling. Current immunosuppressive cytokine therapies for autoimmune diseases elicit a breadth of systemic effects, highlighting the need for more targeted approaches. Building on our lab's previous work in altering cytokine selectivity for targeted cell-type delivery, we sought to identify optimal candidates for mutein design. We assessed the immunosuppressive potential of 90 different cytokines using single-cell RNA sequencing (scRNA-seq) data from PBMCs individually stimulated with each cytokine. Analyzing the data across cells, genes, and stimulations poses a challenge as key associations are often obscured by conventional data flattening approaches. Here we use PARAFAC2-RISE tensor decomposition to preserve the multi-dimensional structure and find variation specific to cytokines, cell subpopulations, and gene modules. PARAFAC2-RISE identifies immunosuppressive signatures within each cytokine's broader response profile, providing a framework for comparing suppressive activity across all cytokines.

9. **Exploring the synergies of imaging-genomics data integration in producing accurate genome structures at higher resolution.**

Lorenzo Boninsegna[1], Francesco Musella[1,2], Frank Alber[1,2]
[1] Institute for Quantitative and Computational Biosciences and Department of Microbiology, Immunology and Molecular Genetics, University of California Los Angeles, CA 90095, USA
[2]Bioinformatics IDP Program, University of California Los Angeles, CA 90095, USA

Data from state-of-the-art sequencing- and imaging-based technologies have revolutionized our view of 4D genome and bolstered simulations to uncover the mechanistic underpinnings of biological processes. We recently developed the Integrated Genome Modeling (IGM) platform for data-driven simulation of single cell genome structures, jointly recapitulating multi-modal datasets. Here we discuss its expansion (IGM+) for incorporating in the modeling process single cell imaging data, such as chromatin tracing and volumetric data. This integration allows us to achieve accurate higher-resolution genome structures, while increasing the coverage of imaged genomic regions by a factor 100x. In addition, leveraging single cell spatial information allows us to fold chromatin into single-cell specific nuclear environments defined by variably shaped nuclei and nuclear bodies. As new and more accurate single-cell tracing technologies continue to emerge, their active integration into comprehensive computational models of genome organization will likely become a new gold standard in the field.

10. **Protein interaction network topology predicts in vitro synergy comparable to approaches trained with experimental data**

Emily R. Bozich[1], Jennifer L. Wilson[1]
Department of Bioengineering,
University of California Los Angeles, Los Angeles California, USA

Combination therapies are regarded as the future for treating cancer, yet predicting synergy between multiple drugs remains challenging. Even sophisticated machine learning models with ample single perturbation and cell line information report model-experiment correlations of ~0.24-0.48. Separately, protein-protein interaction network methods have successfully predicted drug effects, however,

few have considered de novo synergy prediction solely from network topology. Thus, we measured the extent to which topological relationships between drug protein targets can predict experimental synergy. We quantified topology by exhaustively testing distance metrics and found that they were moderately aligned with synergy (correlations of 0.35-0.58), suggesting that network topology alone is as performant as experiment-informed models. We further used sensitivity analyses to understand performance differences between distance metrics encoded by varied network features. Guided by our reported optimal network features, models may be a powerful approach for de novo synergy prediction and finding effective drug combinations.

11. Studying the long-term evolutionary dynamics of transposable elements

Sneha Chakraborty[1], Kirk Lohmueller[123]
[1]Department of Ecology and Evolutionary Biology,
[2]Department of Human Genetics, David Geffen School of Medicine,
[3]Bioinformatics Interdepartmental PhD Program,
University of California Los Angeles, Los Angeles California, USA

Transposable elements (TEs), once dismissed as "junk DNA," are now recognized as integral to genome architecture, yet their long-term evolutionary dynamics remain poorly understood. Here we use forward simulations in SLiM to model retrotransposon activity under diverse selective and demographic scenarios, analyzing the resulting site-frequency spectra to infer evolutionary patterns. Our framework incorporates a biologically realistic, copy-number–dependent (inverted U-shaped) transposition rate and examines its interplay with deletion and inactivation. We show that even under neutral conditions, where genetic drift is the sole evolutionary force, while both processes limit TE proliferation, deletion increases the persistence of high-frequency alleles compared to inactivation — highlighting the critical role of deletion and activity loss in shaping long-term TE dynamics. This survival-analysis perspective provides a novel validation of classic theory under more realistic assumptions and offers new insights into the maintenance of TE diversity across genomes.

12. GRNComm: global cell-cell communication inference from spatial transcriptomics and scRNAseq data

Michael Cheng[1,2], Xia Yang[1,2,3,4,5]
[1]Department of Integrative Biology & Physiology, UCLA, Los Angeles, CA, USA
[2]Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA, USA
[3]Institute for Quantitative and Computational Biosciences (QCBio), Los Angeles, CA, USA
[4]Molecular Biology Institute (MBI), Los Angeles, CA, USA
[5]Brain Research Institute (BRI), Los Angeles, CA, USA

The availability of single cell RNA sequencing (scRNAseq) and spatial transcriptomics provides a unique opportunity to study spatially dependent mechanisms in disease, such as cell-cell communication (CCC). Existing CCC methods, however, are limited by their reliance on known ligand-receptor pairs, often overlooking novel interactions. We developed GRNComm, a bootstrapped gradient boosting framework that integrates scRNAseq and spatial transcriptomics to address this gap. GRNComm performs an unbiased search for cross-cell-type interactions by analyzing global gene expression patterns, without requiring prior knowledge of communication genes. When comparing the enriched biological pathways in GRNComm's neuron-astrocyte cell communication networks with existing methods, we identified majority of shared pathways in axon guidance and synaptic transmission, as well as pathways specific to GRNComm in synapse development. These results suggest that GRNComm can retrieve biologically meaningful cell-cell interactions and signaling pathways that are missed when relying on predefined ligands and receptors.

13. **Sensitivity analysis of GRNmap and new feature dynamical systems modeling and viewing small- networks**

Kam D. Dahlquist[1], Ben G. Fitzpatrick[2], John David Nikki C. Chun[1,2], Cecilia J. Zaragoza[3], A'Kaia L. Phe Y. Zekarias[3]
[1]Department of Biology,
[2]Department of Mathematics,
[3]Department of Computer Science
Loyola Marymount University, Los Angeles, Califo

GRNmap uses ODEs to model dynamics of sr production rates, expression thresholds, and r transcription factor using gene expression c simulations. We previously used it to understan Now we examine its performance with smaller " parameter sensitivity. The least squares error bet weights was determined for all 21 possible 3-noc production rates in addition to weights and thres model was also sensitive to the direction anc parameters for networks with the same connecti displays a graph where edges are color-code repression relationships and nodes are colo expression data. Backend databases allow users t the nodes with expression data, and export an E in GRNmap.
http://kdahlquist.github.io/GRNmap/; https://do

14. **Mapping Neuronal Vulnerability to Amyloid-β and α-Synuclein in Neurodegenerative Disease Models**

Haley De Ocampo[1], Zachary Hemminger[1], Aihui Wang[1], Shujing Zhang[2], Chao Peng[2], Roy Wollman[3,4]
[1] Department of Chemistry and Biochemistry,
[2] Department of Neuroscience,
[3] Department of Bioinformatics,
[4] Department of Systems Biology,
University of California Los Angeles, Los Angeles California, USA

Pathological alpha-synuclein (α-syn) aggregates as Lewy bodies in Parkinson's disease with dementia (PDD), dementia with Lewy bodies (DLB), and some Alzheimer's disease (AD) cases. Although α-syn spreads along neuronal connectivity, distinct brain regions and cell types show variable vulnerability, suggesting molecular mechanisms that remain poorly understood. Amyloid-beta (Aβ) co-pathology further exacerbates susceptibility, underscoring the need for high-resolution approaches to study selective vulnerability. We propose using ATLAS (Atlas-scale Transcriptome Localization using Aggregate Signatures), a novel spatial transcriptomics method, to map neuronal subtypes sensitive to α-syn aggregation. Unlike single-molecule FISH, ATLAS approximates transcriptional states of millions of cells by leveraging gene expression patterns. Using a pre-formed fibril (PFF) injection model in 5xFAD mice, we will examine α-syn propagation in the context of Aβ pathology. Coupling ATLAS with α-syn antibody staining will generate a spatial atlas of α-syn transmission, enabling bioinformatic discovery of candidate genes that drive selective vulnerability and inform therapeutic strategies.

15. **The landscape of fitness effects of putatively functional noncoding mutations in humans**

Chenlu Di[1], Swetha Ramesh[1], Jason Ernst[2,3], Kirk E. Lohmueller[1,4]
1. Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA
2. Department of Computer Science, Samueli School of Engineering, University of California, Los Angeles, CA, USA
3. Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, CA, USA
4. Department of Human Genetics, University of California, Los Angeles, CA USA

While annotations of noncoding regions in the human genome are increasing, the fitness effects of mutations in these regions remain unclear. Here, we leverage these functional genomic annotations and human polymorphism data to infer the distributions of fitness effects of new noncoding mutations in humans. Our novel approach controls for mutation rate variation and linked selection along the genome. We find distinct patterns of selection in putative enhancers, promoters, and conserved noncoding regions. While mutations in enhancers are often neutral, approximately 30%

of mutations in promoters are deleterious. The most conserved noncoding regions, showing reduced divergence across mammals and primates, have the highest proportion of deleterious mutations. Notably, while we infer the most conserved sites across mammals and primates are enriched for deleterious mutations, such conserved sites only account for a minority of the deleterious mutations in noncoding regions. For example, the top 5% of conserved noncoding sites encompass fewer than 20% of deleterious mutations, indicating that functional noncoding regions vary widely in the distribution of their evolutionary constraint. Our findings highlight the dynamic evolution of gene regulation and shifting selection pressures over deep evolutionary timescales. Consistent with this finding, we infer mutations in ~7-9% of the noncoding genome are deleterious. These insights have broad implications for using comparative genomics to identify nonneutrally evolving sequences in the human genome.

16. Protein recruitment to dynamic DNA-RNA host condensates

Mahdi Dizani[1], Daniela Sorrentino[1], Siddharth Agarwal[1,2], Jaimie Marie Stewart[2], Elisa Franco[1,2,3]*

[1]Department of Mechanical & Aerospace Engineering, University of California at Los Angeles, CA 90095

[2] Department of Bioengineering, University of California at Los Angeles, CA 90095

[3]Molecular Biology Institute, University of California at Los Angeles, CA 90095

**Corresponding author: efranco@seas.ucla.edu*

We describe the design and characterization of artificial nucleic acid condensates that are engineered to recruit and locally concentrate proteins of interest *in vitro*. These condensates emerge from the programmed interactions of nanostructured motifs assembling from three DNA strands and one RNA strand that can include an aptamer domain for the recruitment of a target protein. Because condensates are designed to form regardless of the presence of target protein, they function as "host" compartments. As a model protein we consider Streptavidin (SA) due to its widespread use in binding assays. In addition to demonstrating protein recruitment, we describe two approaches to control the onset of condensation and protein recruitment. The first approach uses UV irradiation, a physical stimulus that bypasses the need for exchanging molecular inputs and is particularly convenient to control condensation in emulsion droplets. The second approach uses RNA transcription, a ubiquitous biochemical reaction that is central to the development of the next generation of living materials. We then show that the combination of RNA transcription and degradation leads to an autonomous dissipative system in which host condensates and protein recruitment occur transiently, and that the host condensate size as well as the timescale of the transient can be controlled by the level of RNA degrading enzyme. We conclude by demonstrating that biotinylated beads can be recruited to SA-host condensates, which may therefore find immediate use for the physical separation of a variety of biotin-tagged components.

17. Uncovering the role of converging antisense transcription in neuroinflammation through intermolecular dsRNA formation

Anna M.J. Fraser[1], Jae Hoon Bahn[2], Ting Fu[2], Elaine Huang[3], Ryo Yamamoto[3], Tadeo C. Spencer[3], Jonatan L. Hervoso[3], Xinshu Xiao[2]

[1] Molecular Biology Interdepartmental Program,

[2] Integrative Biology and Physiology,

[3] Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles California, USA

Antisense transcription is a pervasive feature of eukaryotic genomes, yet the mechanisms driving its occurrence and the functions of the resulting transcripts remain understudied. Intriguingly, converging bidirectional transcription produces complementary sense-antisense transcripts that may form intermolecular RNA duplexes. Endogenous

double-stranded RNAs (dsRNAs) are known to trigger immune response and have been linked to inflammatory pathologies, including Alzheimer's disease (AD). Here, we develop a bioinformatic pipeline leveraging stranded RNA-sequencing data from an AD patient cohort to systematically identify and characterize genomic regions harboring converging bidirectional transcription. Our analyses reveal a considerable subset of these regions associated with inflammatory and neurodegenerative disease phenotypes. Notably, many of these regions exhibit traits indicating the formation of intermolecular dsRNA structure *in vivo*, including A-to-I editing and low energetic favorability of intramolecular folding. This study implicates intermolecular dsRNA duplexes arising from antisense transcription as important contributors to the endogenous dsRNA species driving neuroinflammation in AD.

18. Minority Reporting: Leverage and its effects on the analysis of scRNA-seq data

Timothy Hamilton[1], Juan Vergara Najar[2], Eric J. Deeds[3,4]
[1] Bioinformatics Interdepartmental Program,
[2] Department of Computation and Systems Biology,
[3] Department of Integrative Biology,
[4] Institute of Quantitative and Computational Biology,
University of California Los Angeles, Los Angeles California, USA

The advent of single cell methods has revolutionized the fields of developmental and systems biology by enabling the investigation of complex tissues at the level of individual cells. Supporting these advancements is the fundamental idea that the gene expression of single cells can reveal distinct expression profiles that correspond to biologically distinct and meaningful groups. Unfortunately, recent work belies that notion; if a small percentage of cells is removed from the dataset and the typical analysis pipeline is repeated, the partitioning of remaining cells is significantly changed. We developed a new geometric method to determine whether the unique geometric structure of scRNA-seq data is the cause. Our findings imply that the analysis pipeline used to analyze scRNA-seq data are extremely sensitive to dataset composition, implying that the insights generated by the analysis of single cell data may not be generalizable beyond the original experiment.

19. First-year post-transplant respiratory DNA methylation profile predicts chronic lung allograft dysfunction

Fei-Man Hsu[1,2], John A. Belperio[3], S. Samuel Weigt[3], Vyacheslav Palchevskiy[3], Matteo Pellegrini[1,2,*], Joanna M. Schaenman[3,*]
[1] Department of Molecular, Cell and Developmental Biology,
[2] Institute for Quantitative and Computational Biosciences – The Collaboratory,
[3] Department of Medicine,
University of California Los Angeles, Los Angeles California, USA

Lung transplant (LTx) improves the quality and duration of life for patients with severe lung disease. Chronic lung allograft dysfunction (CLAD) occurs in approximately 50% of LTx recipients, leading to significant lung allograft morbidity and mortality. However, there is a lack of an early-detection method for CLAD. This study aims to characterize DNA methylation, and epigenetic modification, as a biomarker for the early detection of CLAD onset. We collected longitudinal bronchoalveolar lavage (BAL) samples from a multi-center cohort of 155 patients who underwent LTx over a 3-year follow-up period, and profiled DNA methylation using targeted bisulfite sequencing (TBS-seq). Our results showed that BAL DNA methylation could predict CLAD outcome accurately even within the first year post-LTx, and the derived epi-CLAD score is highly associated with time-to-CLAD, suggesting it could serve as a new modality to characterize patients at high risk of CLAD.

20. A Framework for Rigorous Cell Segmentation and Annotation for Xenium Spatial Transcriptomics

Helen Huang[1,2,3], Zhiyuan Zhu[1], Xiaolu Guo[2], Alexander Hoffmann[1,2]

[1] Institute for Quantitative and Computational Biosciences (QCB),

[2] Bioinformatics Interdepartmental Program,

[3] Signaling Systems Laboratory, Department of Microbiology, Immunology, and Molecular Genetics (MIMG),

University of California Los Angeles, Los Angeles, USA.

In imaging-based spatial transcriptomics like Xenium, cell segmentation and cell-type annotation (S&A) affect the reliability of biological conclusions. However, achieving accurate S&A is challenging due to tissue sectioning that leads to overlapping cells, nucleus-less cells, and lateral transcript diffusion. Although numerous S&A algorithms have been developed, evidence-based strategies for optimizing image analysis parameters and estimating accuracy are lacking. We evaluated widely used S&A tools by assessing single-cell expression profiles. Our results demonstrate that default parameters often lead to mis-segmentation and fail to classify cell types. To resolve these limitations, we developed a framework for optimizing segmentation parameters based on cellular morphological properties, combined with hierarchical annotation that is both knowledge- and data-driven. Our framework generated accurate cell maps, validated by marker purity. This work highlights the limitations of existing tools and provides a framework and criteria for parameter tuning to generate rigorous insights from complex spatial data.

21. Allele-specific splicing in the human brain reveals genetic variants linked to protein isoforms and Alzheimer's disease

Alison J. King[1], Kofi Amoah[1], Ryan M. Barney[1], Giovanni Quinonez Valdez[2], and Xinshu Xiao[1,2]

[1]Bioinformatics Interdepartmental Program, University of California, Los Angeles, USA

[2]Department of Integrative Biology and Physiology, University of California, Los Angeles, USA

Despite growing catalogs of genetic variation linked to human traits, their functional impact remains poorly understood. Alternative splicing, particularly in the human brain, represents a key layer of post-transcriptional regulation that may mediate genetic effects on gene expression and protein diversity. Here, we systematically map allele-specific alternative splicing (ASAS) events in postmortem brain tissues, identifying hundreds of genetically regulated splicing events across four brain regions. We nominate over 500 putative functional SNPs regulating ASAS, with many overlapping splicing QTLs, RBP binding sites, and GWAS loci for Alzheimer's disease (AD). ASAS events frequently occur in 5' UTRs, where they are associated with protein QTLs, alternative start codons, and isoform-specific domain changes. Importantly, we identify ASAS events exhibiting AD-specific splicing patterns, including within mitochondrial function- and neuronal signaling-related genes. Our results provide a brain-specific, splicing-resolved map of regulatory variation, uncovering novel mechanisms linking genetic variation to transcript and protein-level changes in AD.

22. FastGxC: A fast and powerful statistical method for mapping context-specific regulatory variants in bulk and scRNA-seq studies

Lena Krockenberger[1], Andrew Lu[2], Mike Thompson[3], Cuining Liu[1], M Grace Gordon[4], Andy Dahl[5], Chun Jimmie Ye[6], Noah Zaitlen[7,8,9], Brunilda Balliu[9,10,11]

[1]Bioinformatics Interdepartmental Graduate Program, University of California Los Angeles, Los Angeles, CA, USA

[2]UCLA-Caltech Medical Scientist Training Program, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

[3]Systems Biology Department, Center for Genomic Regulation (CRG), Barcelona, Spain

[4]Biological and Medical Informatics Graduate Program, University of California, San Francisco, San Francisco, CA, USA
[5]Section of Genetic Medicine, University of Chicago, Chicago, IL, USA
[6]Division of Rheumatology, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA
[7]Department of Computer Science, Samueli School of Engineering, UCLA
[8]Department of Neurology, David Geffen School of Medicine, UCLA
[9]Department of Computational Medicine, David Geffen School of Medicine, UCLA
[10]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA
[11]Department of Biostatistics, Fielding School of Public Health, UCLA

Context-specific cis expression quantitative trait loci (sp-eQTLs) underlie genetic risk factors for complex diseases. To limit cost and experimental heterogeneity, bulk and single cell eQTL studies often gather multiple samples across contexts (tissues and cell types) from the same donors. Linear mixed models (LMMs) are a natural analysis choice for such studies because they model intra-individual residual correlation and directly identify sp-eQTLs through a genotype-by-context (GxC) interaction term. However, LMM-GxC are computationally infeasible for eQTL studies. Hence, researchers rely on context-by-context eQTL mapping with post hoc examination of summary statistics to identify sp-eQTLs. While fast, these approaches are underpowered, as they do not account for shared noise within individuals, and rely on ad hoc definitions of sp-eQTLs, e.g. presence of eQTL in a single context. These shortcomings hinder characterization of sp-eQTLs and interpretation of disease-associated variants. Here, we introduce FastGxC, a fast and powerful method for sp-eQTL mapping in studies with repeated sampling. FastGxC removes intra-individual correlation by decomposing gene expression per individual into context-shared and context-specific components and estimates genetic effects on these components separately using ultra-fast implementations of linear regression. FastGxC eQTL effect sizes are a re-parameterization of LMM-GxC effect size estimates, enabling direct mapping of sp-eQTLs without post hoc analyses. FastGxC provides a global test to identify variants with heterogeneous effect sizes across contexts and a marginal test to identify the context(s) driving this heterogeneity. Its output integrates seamlessly with existing methods like mash and METASOFT, to improve the statistical power of eQTL mapping. In simulations, FastGxC is orders of magnitude faster and more powerful than previous approaches. We apply FastGxC to bulk multi-tissue RNA-seq data from GTEx (N=948) and PBMC single-cell RNA-Seq data from the CLUES (N=234) and OneK1K (N=982) cohorts, generating a comprehensive tissue- and PBMC cell type-specific cis-eQTL map. FastGxC identifies eQTLs for 70.2% and 36.1% of tested genes across tissues and PBMC cell types, 50.6% and 22.5% of which show specificity across tissues and cell types, respectively. FastGxC sp-eQTL are enriched in context-specific genomic annotations and provide a three-fold increase in precision to identify relevant contexts for GWAS variants across 138 complex traits. In summary, FastGxC sp-eQTLs can be used to understand context-specific gene regulatory mechanisms underlying complex human diseases.

23. Complex multisite adaptations underlie hard and soft sweeps across human gut microbiomes

Peter Laurin[1], Nandita Garud[1,2]
1. Department of Ecology and Evolutionary Biology,
2. Department of Human Genetics,
University of California Los Angeles, Los Angeles California, USA

In the human gut microbiome, adaptation is rapid, driven in part by bacterial populations' large sizes and associated immense mutational input. Recent work has also established adaptation proceeding across host microbiomes via horizontal gene transfer (HGT) of adaptive

gene fragments. We quantify how often host gut commensal species adapt via interhost-HGT vs. *de novo* mutation by conducting haplotype homozygosity scans for shared genetic fragments driven by positive selection, and characterize the number of *de novo* mutational origins from which adaptations arise. We find an abundance of across-host sweeps proceeding from surprisingly few genetic origins. Under a conservative model of interhost HGT, we recapitulate observed sweep frequencies under low mutation rates, consistent with complex multisite adaptation. In support of this hypothesis, genetic variation at peaks provides preliminary signs of introgression/interspecific origin. Overall, we propose that complex adaptation via across-host HGT proceeds through broadly different processes than rapid adaptation at single sites.

24. Epigenetic Footprints of Metabolic Memory in Type 2 Diabetes

Guan-Jun Lin [1, 2, 4], Ying-Chen Chen [3], Matteo Pellegrini [4], Feng-Chih Kuo [3], Pao-Yang Chen [1, 2, 4]

[1] Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

[2] Genome and Systems Biology Degree Program, Academia Sinica and National Taiwan University, Taipei, Taiwan

[3] Division of Endocrinology and Metabolism, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan

[4] UCLA-DOE Institute of Genomics and Proteomics, University of California, Los Angeles, Los Angeles, CA, United States

Metabolic memory, induced by hyperglycemia in diabetes, leads to various complications that often become apparent too late for treatment. This legacy effect has been observed in type-1 diabetes as DNA methylation changes at specific CpGs. To explore type-2 diabetes, we profiled genome-wide DNA methylation footprints, and tested a novel hypoglycemic drug. Peripheral blood mononuclear cells were collected from normoglycemic individuals, newly diagnosed patients, and long-duration patients with drug follow-up. We devised computational pipelines to track dynamic methylation changes across diabetes stages. Intriguingly we actually uncovered two distinct types of footprints; *new footprints* arising at diabetes onset and *long-term footprints* at separate loci, linked to persistent hyperglycemia and complications. While the methylation at *new footprints* could return to near-normal over years, the drug effectively erased them. Our work together with the bioinformatics pipeline advances current understanding of metabolic memory, and enables early prediction and personalized treatment of complications.

25. High-grade serous ovarian cancer autoantibodies interact poorly with cytotoxicity-inducing Fc receptors

Michelle Loui[1], Scott Taylor[1], Crystal Xiao[1], Cyrillus Tan[2], Het Desai[1], Allison Brookhart[3], Sophia Ting[4], Gabriella DiBernardo[5], Laura B. James-Allan[5], Sanaz Memarzadeh[5], Aaron S. Meyer[1]

[1]Department of Bioengineering,

[2]Bioinformatics Interdepartmental Program,

[3]Institute for Quantitative and Computational Biosciences,

[4]Department of Microbiology, Immunology and Molecular Genetics,

[5]Department of Obstetrics and Gynecology, David Geffen School of Medicine,

University of California, Los Angeles, Los Angeles, CA , USA

In high-grade serous ovarian cancer (HGSOC) patients, malignant epithelial cells arise from the fallopian tube and ovarian surface epitheliums. Endogenous antibodies (anti-tumor antibodies; ATAbs) target these cells and should promote recognition by the immune system. Patient-derived tumors have been found to be frequently coated in IgG, and ATAbs are present both in the tumor mass and in the fluid that builds up in the peritoneum surrounding the tumor microenvironment. They are derived from B cells that have undergone somatic hypermutation, indicating an active immune response. However, despite their widespread abundance in HGSOC, ATAbs fail to eliminate the tumor cells. We hypothesized that ATAbs are unable to eliminate tumors due to the dysregulation of immune interactions via their Fc region. Therefore, we applied a quantitative, multiplexed assay for profiling the Fc properties and immune receptor

interactions of ATAbs. Our data demonstrate that ATAbs interact poorly with FcγRIIIa—a potent activating receptor for antibody-dependent cellular cytotoxicity (ADCC) found on natural killer cells—due to fucosylation. Understanding the mechanisms of humoral immunity evasion will help with the prediction of therapeutic responses in cancer patients and uncover how immunotherapies might reactivate effective humoral immunity.

26. Inference of population demographic history captures differing evolutionary signals based on the number of individuals in the dataset

Jonathan C. Mah[1], Kirk E. Lohmueller[1, 2, 3]
[1]Bioinformatics Interdepartmental Program,
[2]Department of Ecology and Evolutionary Biology
[3]Department of Human Genetics,
University of California Los Angeles, Los Angeles California, USA

Accurate estimation of population demographic history is critical for population genetics inference but remains a challenging task. The site-frequency-spectrum (SFS) is a highly sensitive summary statistic of genetic variation which is commonly leveraged for inference of population demographic history; however, studies have shown that data analyzed at different sample sizes may yield qualitatively different demographic models. We analyzed two simulated datasets and one empirical dataset, all displaying an ancient population contraction and a recent population expansion. We fit a two-epoch demographic model to data subsampled to a variety of sample sizes, finding that the dominant inferred evolutionary signal changed from an ancient contraction at small sample sizes to a recent expansion at large sample sizes. Other summary statistics, such as Tajima's D and the proportion of rare variants, also change with sample size, suggesting that analysis of data at different sample sizes may shed insight into different evolutionary signals.

27. Detecting adaptation in the gut microbiome

Aina Martinez i Zurita[1], Sharon Greenblum[2], Nandita Garud[1,3]
[1] Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, USA
[2] Joint Genome Institute, Genome Technologies Group, Berkeley, CA
[3] Department of Ecology and Evolutionary Biology, University of California, Los Angeles, USA

The human gastrointestinal tract contains a diverse ecosystem of microorganisms that impact many aspects of human health. Due to the large population size of bacteria in each human host and their short generation time, it is estimated that an average human microbiome experiences billions of de-novo mutations every day. With such a large mutational input, the same mutation may arise in multiple hosts, and if beneficial, could rise to high frequency in these hosts. Mutations changing in frequency in parallel across many hosts in response to similar selection pressures may be adaptive and thus have important functional relevance to the microbiome. Statistical methods to detect such parallelism have proved insightful in metazoan populations, yet complementary methods that account for the unique features of complex bacterial populations have not yet been developed or rigorously tested. Here we develop a statistical framework to detect cases of parallelism in temporally sampled metagenomic datasets across many hosts. We test the effectiveness and sensitivity of our method using simulations. We demonstrate our method using a temporally sampled metagenomic dataset of 60 patients undergoing a 5-day ciprofloxacin antibiotic treatment. We successfully recover a signal in the main target for ciprofloxacin resistance, DNA gyrase subunit A (gyrA). We show elevated rates of parallelism during the antibiotic exposure and post-exposure in several bacterial species. Our work begins to uncover the dynamics of adaptive variants that may have been previously missed due to subtle yet consistent allele fraction changes.

28. Understanding accuracy, equity, and portability in cross-ancestry biobank phenotype imputation

Rachel Mester[1], Ulzee An[2], Aditya Gorla[3], Na Cai[4], Sriram Sankararaman[1,2,5]

[1]Department of Computational Medicine, University of California Los Angeles, Los Angeles, California

[2]Department of Computer Science, University of California Los Angeles, Los Angeles, California

[3]Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, California

[4]Helmholtz Pioneer Campus, Helmholtz Zentrum Munchen, Neuherberg, Germany

[5]Department of Human Genetics, University of California Los Angeles, Los Angeles, California

Phenotype imputation enables researchers to infer missing or unmeasured traits in biobank-scale datasets, expanding the scope of genetic and epidemiologic analyses. Recent developments in imputation algorithms have incorporated the complex missingness structure present in biobank data and improved the accuracy of biobank imputation in a cohort of white British individuals from the UK Biobank. However, the accuracy and portability of phenotype imputation models across diverse ancestral populations remains poorly understood, and improving phenotype imputation parity is essential to ensuring that the benefits of large-scale biobank research extend across global populations. In this study, we evaluate the cross-ancestry performance of phenotype imputation models. We quantify imputation accuracy across group, measuring the impact of sample size, missingness structure, and feature correlation on imputation. We investigate potential drivers of disparities in imputation accuracy between phenotypes and between ancestries and explore strategies to improve imputation accuracy in diverse populations.

29. Developing a disease-specific accessible transcriptional signature as a biomarker for ataxia with oculomotor apraxia type 2

Kathie J. Ngo[1], Darice Y. Wong [1,2], Alden Y. Huang[3], Hane lee[4], Stanley F. Nelson[1,5,6], Brent L. Fogel [1,2,6]

[1.] Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

[2.] Clinical Neurogenomics Research Center, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

[3.] Institute for Precision Health, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

[4.] 3billion, Inc., Seoul, Korea

[5.] Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

[6.] Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

Genetic ataxias are clinically heterogenous neurodegenerative conditions and it is difficult to assign pathogenicity to rare gene variants solely based on DNA sequencing. An effective functional assay from an easy-to-obtain biospecimen would aid this assessment and be of high clinical value. *SETX* encodes a ubiquitous DNA/RNA helicase crucial for maintaining genome stability and loss-of-function mutations cause a recessive disorder, Ataxia with Oculomotor Apraxia Type 2 (AOA2). We utilized Weighted Gene Co-expression Network Analysis (WGCNA) from patient blood to construct an AOA2-specific transcriptomic signature as a biomarker to evaluate *SETX* variants in patients clinically suspected of having AOA2. Overall, the transcriptional biomarker effectively distinguishes AOA2 subjects (n=32) from controls (n=35) with 72% sensitivity and 97% specificity and from other related neurological disorders. As proof-of-concept, we identified a previously undiagnosed ataxia patient with the first pathogenic mutation in a non-canonical

SETX transcript, expanding the spectrum of mutations that contribute to AOA2.

30. Generating forces in confinement via polymerization

Dino Osmanovic[1], Elisa Franco[1,2]
1, Mechanical and Aerospace Engineering, UCLA
2, Bioengineering, UCLA

Understanding how to produce forces using biomolecular building blocks is essential for the development of adaptive synthetic cells and living materials. Here we ask whether a dynamic polymer system can generate deformation forces in soft shells by pure self-assembly, motivated by the fact that biological polymer networks like the cytoskeleton can exert forces, move objects, and deform membranes by simply growing, even in the absence of molecular motors. We address this question by investigating polymer force generation by varying the release rate, the structure, and the interactions of self-assembling monomers. First, we develop a toy computational model of polymerization in a soft elastic shell that reveals the emergence of spontaneous bundling which enhances shell deformation. We then extend our model to account more explicitly for monomer binding dynamics. We find that the rate at which monomers are released into the interior of the shell is a crucial parameter for achieving deformation through polymer growth. Finally, we demonstrate that the introduction of multivalent particles that can join polymers can either improve or impede polymer performance, depending on the amount and on the structure of the multivalent particles. Our results provide guidance for the experimental realization of polymer systems that can perform work at the nanoscale, for example through rationally designed self-assembling proteins or nucleic acids.

31. Chromosome structure remodeling in innate immune training and gene regulation

Sohyeon Park[1,2], Ye Wang[2,3], Aleksandr Gorin[4], Frank Alber[1,2], Alexander Hoffmann[1,2]
[1.]Department of Microbiology, Immunology, and Molecular Genetics, University of California Los Angeles, Los Angeles, CA 90095
[2.]Institute for Quantitative and Computational Biosciences, University of California Los Angeles
[3.]Department of Pathology and Laboratory Medicine, University of California Los Angeles, Los Angeles, CA 90095
[4.]Department of Medicine, Division of Infectious Diseases, University of California Los Angeles, Los Angeles, CA 90095

Macrophage training has emerged as a promising target in cancer immunotherapy since macrophages are highly plastic and can adopt pro- or anti-tumor functions. Macrophages show the capacity for long-lasting immune memory, but the underlying mechanisms are not well understood. We hypothesize that immune training causes large-scale chromosome structural rearrangements that maintain the epigenetic enhancer landscape and regulate macrophage gene expression. Using high-depth Hi-C sequencing and differential compartment analysis, we reveal and map chromosome structural rearrangements in human macrophages and show that these changes are stimulus-specific. Integrating Hi-C with ATAC-seq and CUT&Tag (H3K4me1) data we found that chromosomal structural changes align with DNA accessibility and poised enhancer activity. By reconstructing chromosome structure at single-cell resolution via Integrative Genome Modeling, we identified genes whose distances to nuclear speckles changed and applied a simple quantitative model to relate these to RNA-seq data. Together, these findings support chromosome remodeling as a mechanism of innate immune memory in macrophages.

32. Identifying Common Host Immune States in MRSA and Candida Infections to Uncover Therapeutic Opportunities

James Popoli[1], Aaron Meyer[1]
[1]Department of Bioengineering, University of California, Los Angeles, Los Angeles California, USA

Methicillin-resistant *Staphylococcus aureus* (MRSA) and *Candida albicans* (CA) infections are significant causes of morbidity and mortality in hospital settings and often occur in similar, often immunocompromised, patient populations suggesting shared host vulnerabilities. Here, we test the hypothesis that MRSA bacteremia and candidemia elicit overlapping host transcriptional responses. We analyzed whole-blood RNA sequencing data from both infection types and performed principal component analysis (PCA) to identify common expression patterns. To place these findings in a broader infectious context, we applied PARAFAC2 tensor decomposition across our datasets and additional public transcriptomic data. Our results reveal a strong convergence in host gene expression between MRSA and CA infections, including shared PCA signatures predictive of patient outcomes. PARAFAC2 analysis further confirmed these similarities and uncovered shared immune responses across other infections. These findings highlight common host pathways as targets for broad-spectrum therapies or used to stratify patients across diverse infections.

33. ATLAS 3D: Large-scale, deep-tissue spatial transcriptomics through single cell mapping

Xingbo Shang[1], Kevin Suh[1], Ziwei Qiu[1], Aihui Wang[1], Thomas Underwood[1], Haley De Ocampo[1], Ricajoy Montero[1], Roy Wollman[1,2,3]
[1.] Department of Chemistry and Biochemistry, UCLA
[2.] Institute of Quantitative Biosciences, UCLA
[3.] Department of Integrative Biology and Physiology, UCLA

Spatial transcriptomics has greatly revolutionized the study of biological systems by offering a comprehensive overview of transcripts' levels and preserving the associated spatial information. Most spatial transcriptomic methodologies developed so far can only work on thin specimens with thickness less than 20µm in a low throughput manner. Consequently, it remains difficult to map the complete 3D spatial transcriptome of organ-scale objects (e.g. mouse brain). To solve this problem, we are currently developing ATLAS 3D, an imaging based technology capable of mapping spatial transcriptome in thick specimens (~400µm) in a high throughput manner. By utilizing light-sheet microscopy to measure a nonlinear dimensionality reduction learnt from a reference single-cell RNA-seq reference through a carefully designed neural network, ATLAS 3D is expected to process ~40 thick specimens per week, rendering it feasible to assemble complete 3D spatial transcriptomes for organ-scale objects.

34. Comparative Analysis of Drug Effects Across Species Using Gene and Pathway Signatures

Colette Shvager, Dr. Wilson, Department of Bioengineering, University of California Los Angeles, Los Angeles California, USA

Drugs are designed to target specific proteins, yet proteins function within complex networks, and even targeted drugs can produce adverse effects. Traditional animal models have yielded valuable insights, but physiological differences between species contribute to high translational failure rates. To address this challenge, new computational approaches are emerging. PathFX is a protein-network side effect prediction platform, and PharmOmics is a multi-species gene expression and pathway database. This study leverages PharmOmics to investigate conservation of drug-induced gene signatures across species. We focus on Vasodilator, analyzing both "genes_combined" (differentially expressed genes) and "enrichment_pathways" (associated biological pathways). Results reveal minimal overlap in activated genes across species, suggesting limited cross-species conservation. Thes findings underscore both the challenges and potential of integrating gene-based repositories with predictive tools like PathFX. By quantifying variation across species and networks, this work informs how large-scale datasets can be integrated into predictive algorithms for anticipating adverse drug effects.

35. Ants with ID tags help reveal nature's supply chains

Woncheol Song[1], Hyein Jo[1], Noa Pinter-Wollman[1]

[1] Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles California, USA

Social insects are valuable models to study information and resource flow through interaction networks. Argentine ants are important target species to study, because they are highly successful invasive species and exhibit a wide range of interaction behaviors. However, their sensitive anatomy and physiology make individual identification challenging. Here, we present novel methods for tagging and tracking the Argentine ants. To make low-cost, non-toxic, high-definition ID tags, we used cyanotype, an iron-based photographic printing technique, in conjunction with photolithography stencils used for integrated circuit manufacturing. Then the print was sealed in with veterinary adhesive before being attached to the ants. To detect the interactions in their crowded colony, we developed an image classification pipeline to augment the pose estimation and tag reading models. Our tools enable high-throughput analysis of resource flow in Argentine ants, which can reveal their ecology as well as inspire industrial logistics, communication infrastructure, and public health decisions.

36. Spatial cell type mapping of mouse lymph node to study B cell dynamics in germinal center.

Kevin Suh[1], Zachary Hemminger[1], Alexander Hoffman[2,3], Roy Wollman[1,3]

[1] Department of Chemistry and Biochemistry,

[2] Department of Microbiology, Immunology and Molecular Genetics,

[3] Institute for Quantitative and Computational Bioscience,

University of California Los Angeles, Los Angeles California, USA

The antibody repertoire maturation happens at germinal centers (GC) of lymph nodes. Migrating back and forth between light-zone and dark-zone of GC, B cell experiences Darwinian evolutional process to generate high affinity antibody for the antigen presented. Since this process includes dynamic cell-cell interaction between B cells and other types of cells such as antigen presenting cell and T cell in the GC microenvironment, thorough understanding of spatial structure of lymph node is necessary. To accomplish this goal, we will apply ATLAS, a hyper-high-throughput spatial transcriptomics platform, to map tens of millions of cells across entire mouse LNs, track change in the spatial cell type distribution during vaccination and compare the LN responses of different NF-κB mutant. Combined with spatially explicit computational modeling, this approach will reveal how cell interactions, migration, and spatial organization shape B-cell repertoire maturation, providing mechanistic insights into antibody evolution and identifying spatial features predictive of vaccine efficacy.

37. Quantifying the shared genetic components of complex traits and Mendelian phenotypes Mechanistic modeling of antibody-dependent cellular cytotoxicity

Meera Trisal[1], Victoria Ji Gong[1], Mana Bains[1], Aaron Meyer[1]

Department of Bioengineering[1]

University of California Los Angeles, Los Angeles California, USA

Efforts to alter Fc affinity, glycosylation, and antibody subclass have advanced our understanding of Fc-mediated effector responses in cancer, but these strategies remain incomplete because changing any single component perturbs a network of interconnected factors with multivariate outcomes. Here, we implement a mechanistic model of receptor-mediated cell adhesion to investigate how interactions within an immune-effector complex influence antibody-mediated cellular cytotoxicity. The model incorporates multiple variables, including affinity, receptor abundance, and antibody density, and reproduces preliminary findings of

affinity-dependent differences in cell killing that saturate at higher antibody density. These results establish a quantitative framework for examining the coordinated effects of multiple factors and generate predictions that can be experimentally validated to advance mechanistic understanding and guide the engineering of antibody-based therapies.

38. A framework for using drug pathway and genomic data to anticipate new pharmacogenomic associations

Elizabeth M. Trujillo[1] and Jennifer L. Wilson[1]
Department of Bioengineering,
University of California Los Angeles, Los Angeles California, USA
* These authors contributed equally.

Pharmacogenomics has advanced personalized treatment through genetic variation in drug targets, yet drug efficacy and safety often depend on downstream proteins that are rarely considered. To address this gap, we applied PathFX, a network-based algorithm that connects drugs and diseases through downstream pathway interactions. Our analysis of approved and predicted antidepressants revealed that disease-relevant variants are enriched not only in primary drug targets but also in downstream genes, which are functionally linked within shared biological pathways. To validate these findings, we leveraged the University of California Health Data Warehouse, a regional de-identified EHR containing drug exposure, treatment histories, side effects, and sequencing data. PathFX identified several antidepressants with downstream associations to GWAS variants, including the calcium-regulating gene CASR. Among ~8.6M patients, we identified 193 with CASR variants, with 130 representing coding mutations with potential functional effects. These findings highlight downstream proteins as underexplored determinants of pharmacogenomic response and side effects.

39. High-Plex HCR: A Scalable and Efficient Solution for Amplifying Multiplexed Spatial Transcriptomics

Thomas Underwood[1], Ziwei Qiu[1], Haley De Ocampo[1], Aihui Wang[1], Ricajoy Montero[1], Zachary Hemminger[1], Roy Wollman[1,2,3].
[1]Department of Chemistry and Biochemistry, UCLA
[2]Institute of Quantitative Biosciences, UCLA
[3]Department of Integrative Biology and Physiology, UCLA

Hybridization Chain Reaction (HCR) is a powerful signal amplification method for imaging-based spatial transcriptomics, but conventional protocols create major experimental bottlenecks. Achieving high multiplexing typically requires iterative cycles of encoding, amplification, and DNase-based stripping—a process that can extend experiments by up to 48 hours per imaging round and significantly increase the likelihood of failure. Here, we present high-plex HCR, a new HCR structure that employs strippable readout probes to eliminate the need for repeated amplification cycles. In parallel, we introduce a rolling circle amplification (RCA) based synthesis strategy that enables high-throughput production of HCR pools with improved purity. This approach supports the simultaneous screening of hundreds of HCRs using pooling and next-generation sequencing (NGS). Together, these advances provide a scalable and efficient framework for generating high-quality, multiplexable HCR amplifiers, streamlining spatial transcriptomics workflows.

40. Deep learning model uncovers genetic variants affecting mRNA translation and disease risk

Siqi Wang[1], Chuyun Chen[2], Xinshu Xiao[1, *], Zefeng Wang[2, 3, *]
[1] Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles CA, USA
[2] CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China

[3] School of Life Science, Southern University of Science and Technology, Shenzhen, China

Genetic variation can modulate protein abundance by altering mRNA translation, yet this regulatory layer is not well characterized. We built a deep learning model trained on ribosome profiling data from human lymphoblastoid cell lines to predict translation efficiency (TE) from full-length mRNA and quantify single-nucleotide variant (SNV) effects genome-wide. The model identified >90,000 TE-altering SNVs, mostly in untranslated regions, with position-, allele-, and context-dependent impacts. Missense variants, beyond their amino acid changes, also strongly affected TE, with proline substitutions reducing efficiency in proportion to poly-proline tract length. These variants were enriched in immune and cancer pathways, overrepresented in disease loci, and included GWAS-linked coding variants that decreased TE. Mechanistic analyses revealed disruptions to RNA secondary structure and RNA-binding protein interactions. Together, our results highlight translation as a pervasive target of genetic variation and provide a framework for prioritizing functional variants in complex traits and diseases.

41. Studying the Role of Subnuclear Positioning in Gene Expression

Ye Wang[1,2,3], Lorenzo Boninsegna[1,2], Yuxiang Zhan[1,2,4], Xianghong Jasmine Zhou[1,3,*], Frank Alber[1,2,*]

[1]Institute of Quantitative and Computational Biosciences (QCBio), University of California Los Angeles, Los Angeles, CA 90095, USA

[2]Department of Microbiology, Immunology, and Molecular Genetics, University of California Los Angeles, 520 Boyer Hall, Los Angeles, CA 90095, USA

[3]Department of Pathology, David Geffen School of Medicine, University of California Los Angeles, 10833 Le Conte Ave, Los Angeles, CA 90095, USA

[4]Department of Quantitative and Computational Biology, University of Southern California, 1050 Childs Way, Los Angeles, CA 90089, USA

The spatial organization of the genome is crucial to gene regulation. To assess how subnuclear context influences expression, we analyzed genome-wide relationships between nuclear positioning and transcription at single-cell resolution. Using 3D genome structures from the Integrative Genome Modeling (IGM) framework, we classified highly expressed genes by nuclear microenvironment (nuclear position, speckle distance, structural features). Two categories emerged: Class I genes (≈90%, including 73% of housekeeping genes) were speckle-associated, nuclear-interior, and positionally stable. Class II genes (≈10%) showed minimal speckle association, high positional variability, greater length, and favored long-range over nearby enhancers. Some housekeeping genes fell into Class II, challenging the conventional view of stable nuclear positioning. Expression changes between stem and fibroblast cells also correlated with microenvironment shifts. In summary, our findings reveal a complex relationship between subnuclear positioning and gene expression, influenced by the chromatin region's genomic properties.

42. Genetic homogeneity along the tract of the gut

Michael W Wasney[1], Leah Briscoe[2], Ricky Wolff[3], Hans Ghezzi[4], Carolina Tropini[4,5,6], Nandita Garud[1,2,3]

[1]University of California, Los Angeles, Human Genetics, Los Angeles, CA, [2]University of California, Los Angeles, Interdepartmental Program in Bioinformatics, Los Angeles, CA, [3]University of California, Los Angeles, Ecology and Evolutionary Biology, Los Angeles, CA, [4] University of British Columbia, Microbiology and Immunology, Vancouver, Canada, [5]University of British Columbia, School of Biomedical Engineering, Vancouver, Canada, [6]Canadian Institute for Advanced Research, Humans and the Microbiome Program, Toronto, Canada

Environmental gradients in the gastrointestinal tract drive spatial variation in the abundance and membership of bacterial species along the gut. Less is known about the spatial distribution of within-species bacterial genetic variation, which confers important functions. We analyzed genetic diversity of approximately 30 species at different sites along the guts of humanized mice inoculated with the same human stool. Unlike species composition, genetic diversity was uniformly distributed across different gut sites within hosts, due to multiple genetically distinct intraspecific strains coexisting at similar frequencies along the gut. Additionally, roughly 60 evolutionary changes arising within mice showed no spatial specificity. Genetic variation detected in metagenomic samples collected along the human guts revealed similar spatial uniformity, indicating potential generalizability of these findings to the natural human microbiome. Our results imply that species presence-absence may play a larger role than genetic variation in responding to spatially varied environmental pressures in the gut microbiome.

43. Push and pull: signatures of linkage disequilibrium and hitchhiking after a selective sweep

Maya Weissman[1,2], Nandita Garud[1,2,3,4]
1, UCLA Quantitative and Computational Biology; 2, UCLA Ecology And Evolutionary Biology; 3, Human Genetics; 4, UCLA Goodman-Luskin Microbiome Center

Adaptation is a process whereby a beneficial variant rises to high frequency in a population, a process known as a selective sweep. When this happens, nearby neutral and deleterious alleles often “hitchhike” with the sweeping variant, leading to locally elevated linkage disequilibrium (LD). Recent work shows that common nonsynonymous and synonymous sites may not equivalently be impacted: in a selective sweep, nonsynonymous variants should have higher LD with each other than synonymous variants. This is because deleterious variants private to the haplotype bearing the adaptive allele will be in tighter LD with one another in the vicinity of the sweep, while more common synonymous neutral variants will be in weaker LD due to drift. While others have used this signal to identify novel selective sweeps, this phenomenon is not fully understood. Here, we demonstrate that while elevated nonsynonymous LD is often a hallmark of sweeps, this is not true across all distances. Rather, excess nonsynonymous LD peaks at intermediate distances between variants, but is actually lower than synonymous LD at short distances. We demonstrate that this phenomenon is driven by the opposing forces of hitchhiking with beneficial alleles and repulsion between pairs of nearby deleterious alleles due to Hill-Robertson Interference. Finally, we show that this pattern is a key feature of numerous selective sweeps in real gut commensal species. Our research offers new insight into an important population genomic signal and how it applies to real world selective sweeps, and provides a useful metric for identifying novel sweeps across systems.

44. Lineage-Specific Developmental Variation in Wild *C. elegans* Isolates

Charlotte E Weymer[1], Ryan L Johnson[2,3], Erik C Andersen[4], Pavak K Shah[1,2,3,5]
[1] Bioinformatics Interdepartmental Program. University of California, Los Angeles. Los Angeles, CA 90095 USA
[2]Department of Molecular, Cell and Developmental Biology. University of California, Los Angeles. Los Angeles, CA 90095 USA
[3]Molecular Biology Institute. University of California, Los Angeles. University of California, Los Angeles. Los Angeles, CA 90095 USA
[4]Department of Biology. Johns Hopkins University. Baltimore, MD USA
[5]Institute for Quantitative and Computational Biosciences. University of California, Los Angeles. Los Angeles, CA 90095 USA

This study aimed to characterize geographically diverse wild isolate strains of *C. elegans*, particularly CB4856, MY23, DL238, ECA36, JU2001, and XZ1516 in conjunction with the lab strain N2 and the transgenic strain JIM113.

Label-free automated lineage tracing was achieved through the use of embGAN, a previously developed deep learning pipeline. The intersection branch distance was applied towards characterizing heterogeneity through cell cycle timing analysis. This metric is a measure of graph edit distance, modelling the invariant development as a weighted binary tree and assigning the weights with the cell cycle timing. Clustering analysis revealed four clusters of embryos with many of these strains(N2, JIM113, CB4856 DL238, JU2001, and MY23) exhibiting identifiable yet relatively minimal levels of intra-strain heterogeneity. Surprisingly, the standard deviation of the cell cycle timing scaled in a nonlinear manner with the mean cell cycle timing. Further, the variance of the cell cycle timing mean per cell demonstrated highest variability in the beginning and ending of early development, but not the middle, potentially indicating a developmental hourglass effect. On a lineage basis, the E lineage demonstrated the highest variability based on Principle Component Analysis as well as examining the mean absolute deviation of cell cycle timing on a per strain basis, which we hypothesize is due to gap phase introduction in the cell cycles in the E lineage. Future work will further characterize spatial heterogeneity throughout development through embryo alignment and application of the intersection branch distance.

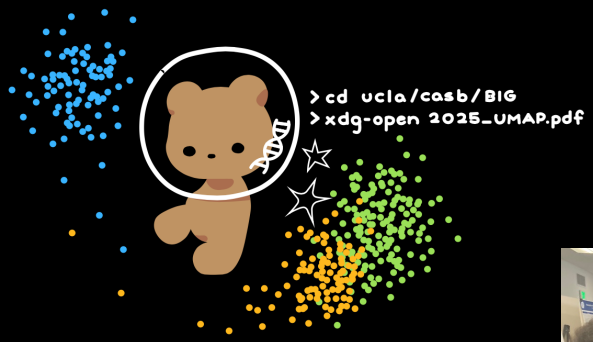
45. Functional coupling of cellular metabolism and renal hemodynamics underlies kidney sex differences

Lingyun (Ivy) Xiong[1], Alan Garfinkel[1], Kate Barrows[2], Lauryn Brown[2], Natalie Porat-Shliom[2], Eric J. Deeds[1]

1. Department of Integrative Biology and Physiology, University of California, Los Angeles, CA

2. National Cancer Institute, National Institutes of Health, Bethesda, MD

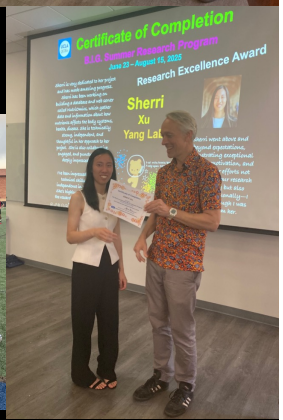
Age-related decline in renal function is faster in males than in age-matched females, manifesting in increased susceptibility to both chronic and acute kidney diseases among males, in humans and rodents. Sexually dimorphic gene activity maps predominantly to proximal tubule (PT) segments, where male-biased lipid metabolism implies exacerbated energy demands for reabsorption. Using intravital microscopy on mouse kidneys, we captured spontaneous metabolic oscillations in individual PT cells and real-time single-nephron glomerular filtration rate (GFR) oscillations, revealing functional coupling between cell-level metabolism and tissue-level fluid flow *in vivo*. GFR is tightly regulated by the tubuloglomerular feedback (TGF) mechanism. We found that TGF-mediated GFR oscillations also show sex-biased characteristics, with males exhibiting higher frequencies. Mathematical modeling of TGF indicates that differential metabolic states in renal tubules can account for sex differences in GFR dynamics. Male kidneys are more prone to loss of TGF-mediated autoregulation, thus explaining sexual dimorphism in renal aging and diseases.



UCLA Institute for Quantitative & Computational Biosciences

A sincere thank you to our 36 faculty and 33 trainee mentors whose dedication made the 2025 B.I.G. Summer Program such a success!

- June 23 to August 15, 2025
- 11th Annual Program
- 297 applications received
- 35 admitted students
- 28 faculty mentors
- 30 direct research mentors



Talks, Posters and Abstracts : https://qcb.ucla.edu/big-summer/big2025/