

9th Annual QCBio Retreat

Tuesday, September 26, 2023

9:30am – 6:00pm

Boyer Hall and Court of Sciences



RETREAT
2023

PROGRAM

UCLA

Institute for Quantitative & Computational Biosciences

Welcome

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

UCLA provides numerous graduate training research opportunities in quantitative and computational biosciences (<https://qcb.ucla.edu/overview/>). Four graduate Programs in Bioinformatics, Medical Informatics, Biomathematics, and Genetics&Genomics are already coordinated and efforts are under way to add a Systems Biology Home Area to address student and faculty demand.

The QCB Collaboratory (<https://qcb.ucla.edu/collaboratory/>) is a postdoctoral training program that provides computational biology postdocs with opportunities to extend their teaching and collaboration skills, preparing them for maximum academic impact. In turn, Collaboratory Fellows have provided essential training in a broad range of skills and workflows through a 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. We maintain an Undergraduate Research Portal (<https://qcb.ucla.edu/research-portal/>) to connect potential mentors with eager undergraduate students – I encourage you to post your projects there.

Mentoring is a key component of our QCB community culture. Now in its 9th year we hosted 74 students within the B.I.G. Summer Undergraduate Research Program this year with 42 laboratories participating! (<https://qcb.ucla.edu/big-summer/big2023/>) The Program was a remarkable success and a model for UCLA. BIG Thank Yous to all faculty, postdoc, graduate student mentors!

For a vibrant academic culture, we recognize that diverse questions, approaches, viewpoints are rooted in diverse racial, ethnic, and cultural experiences. This is an important issue for quantitative and computational biosciences, where diversity lags behind other biosciences disciplines. With our webpage <https://qcb.ucla.edu/diversity-equity-inclusion/> we would like to provide resources and draw attention to our values and the activities that document our commitment. A shout-out to the student-led QBio-EDGE outreach program. Their insights have now appeared in print! <https://doi.org/10.1371/journal.pcbi.1011072>

The QCB Retreat marks the start the new academic year – I invite everyone to participate in and contribute to a thriving community. We will hold our weekly Research lunch; we can support your Affinity group meeting; we will host Career panels; and we’re eager to support your other initiatives and hear from you – so please add your ideas and comments to the prompts on the shared pages as they come up in the course of the day.

- [Ideas for our seminar series?](#)
- [Ideas for symposia or workshops?](#)
- [Ideas for supporting graduate students?](#)
- [Ideas for supporting postdoctoral fellows?](#)
- [Ideas for improving JEDI in our community?](#)
- [Ideas for anything else?](#)

QCB is here for you!

Special thanks to Caroline Baron for again organizing the Retreat this year! Alexander Hoffmann

Agenda

COFFEE, TEA, JUICE, BAGELS, PASTRIES

9:45 am **WELCOME**

10:00 am **STATUS REPORTS I**

- **Alexander Hoffmann**, Director, QCBio, BIG Summer
- **Eleazar Eskin**, Chair of Computational Medicine
- **Matteo Pellegrini**, Director, QCBio Collaboratory

10:20 am **SESSION I**

KEYNOTE I

- **Brunilda Balliu**, Assistant Professor in the Departments of Pathology & Laboratory Medicine and Computational Medicine at UCLA

10:50 am **SELECTED TALKS**

- **Boyang Fu**, Computer Science PhD student, Sankararaman Lab
- **Kangcheng Hou**, Bioinformatics PhD student, Pasaniuc Lab
- **Casey Barkan**, Physics PhD student, Shenshen Wang Lab

11:35 am **STATUS REPORTS II**

- **Alex Bui**, Director, Medical Informatics Ph.D. Program Home Area
- **Grace Xiao**, Director, Bioinformatics Interdepartmental Ph.D. Program
- **Paivi Pajukanta**, Director, Genetic & Genomics, Ph.D. Program
- **Eric Sobel**, Director, Biomathematics, Ph.D. Program

12:00 pm **LUNCH – POSTER SESSION I**

1:30 pm **SESSION II**

KEYNOTE II

- **Mehdi Bouhaddou**, Assistant Professor, Microbiology, Immunology, and Molecular Genetics (MIMG) at UCLA

2:00 pm **SELECTED TALKS**

- **Brian Orcutt-Jahns**, Bioengineering PhD student, Meyer Lab
- **Helena Winata**, Bioinformatics PhD student, Boutros Lab

2:30 pm **STATUS REPORTS III**

- **Matteo Pellegrini and Xia Yang**, Interim Directors of Computational and Systems Biology Major
- **Sriram Sankararaman**, Director of the Bioinformatics minor
- **Eric Deeds**, Director of the Life Science Math Core

2:50 pm **COFFEE & TEA BREAK**

3:15 pm **SESSION III**

SELECTED TALKS

- **Giovanni Quinones-Valdez**, Postdoc, Xiao lab
- **Breanne Sparta**, Postdoc, Deeds lab
- **Jonathan Mah**, Bioinformatics PhD student, Garud lab

4:00 pm **QCBIO-EDGE (Empowering Diversity and Growth in Education)**

CONCLUDING REMARKS

4:15 pm **RECEPTION & REFRESHMENTS - POSTER SESSION II**

Keynote Speakers



Brunilda Balliu

Assistant Professor

Departments of Pathology & Laboratory Medicine and Computational Medicine

Statistical Methods for Mapping Context-Dependent Regulatory Variants

A majority of the variants identified in GWAS fall in non-coding regions of the genome, indicating their mechanism of impact is mediated via gene expression. Leveraging this hypothesis, eQTL studies and TWAS across multiple contexts, e.g., tissues or cell types, have assisted in both the interpretation and discovery of additional genes associated with complex traits. However, existing methods for mapping eQTLs and conducting TWAS do not take full advantage of the intra-individual correlation inherently present in multi-context studies with repeated sampling, e.g., the GTEx project or single-cell RNA-Seq experiments. Moreover, they estimate effects that are mixtures of both context-specific and context-shared (pleiotropic) genetic effects and are often computationally inefficient. These oversights have significantly diminished the power to detect eQTLs and downstream interpretation of disease-associated variants. Here I will introduce three novel statistical methods that leverage the correlation structure of multi-context studies with repeated sampling to efficiently and powerfully map cis and trans-eQTLs as well as perform TWAS. We apply these methods to bulk multi-tissue and single-cell RNA-seq data sets and show that they provide a three-fold increase in precision to identify relevant contexts for GWAS variants and increase the number of locus-phenotype associations discovered by over 51%, relative to previous eQTL mapping and TWAS methods.



Mehdi Bouhaddou

Assistant Professor

Microbiology, Immunology, and Molecular Genetics (MIMG)

Systems to Mechanisms: Deciphering virus-host signaling networks

Viruses are obligate intracellular parasites that require their hosts to replicate. These organisms have long evolved alongside their hosts to hone their ability to remodel the host intracellular signaling environment to maximize their fitness, survival, and spread. One of the primary modalities of exchange between viruses and their hosts is protein-based, accomplished through direct virus-host protein-protein interactions. Such interactions can rewire protein complexes, regulate signaling cascades, and shift phenotypic outcomes. In this talk, I will discuss how we apply mass spectrometry proteomics and bioinformatics to big datasets (Systems) to extract testable mechanistic hypotheses for followup studies (Mechanisms). Our goal is to use this information to develop better antiviral therapies but, also, and importantly, to gain new understanding into how viruses, and by proxy, humans, function.

Selected Talks

▪ *Migration feedback yields novel critical transitions and emergent ecotypes in connected populations*



Casey O. Barkan¹, Shenshen Wang¹

¹Department of Physics and Astronomy,
University of California Los Angeles, Los Angeles, USA

Biological populations live and evolve in spatially extended and heterogeneous environments. From gut microbiota to antibiotic resistant bacteria, spatial heterogeneity of selection pressure can profoundly affect evolution. Yet, the effects of heterogeneity depend upon the migration patterns by which organisms explore their environment. We present a simple and general model of evolution on a network of interconnected habitats, showing that migration feedback generates non-local niches to which *emergent ecotypes* specialize. Varying migration rates induces continuous and discontinuous phase transitions at which ecotypes exchange stability. These transitions result in ecotype extinction or emergence, which may have significant health implications. We find that the discontinuous transitions arise via a novel mechanism of simultaneous transcritical bifurcations. Interestingly, these transitions show a 'fine structure', analogous to that of atomic spectra, when a symmetry of the competitive interactions is broken. The ecological nature of the transitions implies that feasible experiments could explore the predicted behaviors.

▪ *A Biobank-scale test of marginal epistasis reveals genome-wide signals of polygenic epistasis*



Boyang Fu^{1*}, Ali Pazokitoroudi^{1*}, Albert Xue², Aakarsh Anand¹, Prateek Anand¹, Noah Zaitlen^{3,4}, Sriram Sankararaman^{1,4,5}

¹ Department of Computer Science,

² Bioinformatics Interdepartmental Program,

³ Department of Neurology

⁴ Department of Computational Medicine, David Geffen School of Medicine

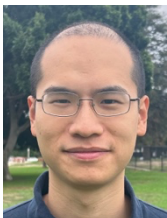
⁵ Department of Human Genetics, David Geffen School of Medicine

University of California Los Angeles, Los Angeles California, USA

* These authors contributed equally.

The contribution of epistasis (interactions among genes or genetic variants) to human complex trait variation remains poorly understood. Methods that aim to explicitly identify pairs of genetic variants, usually single nucleotide polymorphisms (SNPs), associated with a trait suffer from low power due to the large number of hypotheses tested while also having to deal with the computational problem of searching over a potentially large number of candidate pairs. An alternate approach involves testing whether a single SNP modulates variation in a trait against a polygenic background. While overcoming the limitation of low power, such tests of polygenic or marginal epistasis (ME) are infeasible on Biobank-scale data where hundreds of thousands of individuals are genotyped over millions of SNPs. We present a method to test for ME of a SNP on a trait that is applicable to biobank-scale data. We performed extensive simulations to show that our method provides calibrated tests of ME. We applied our method to test for ME at SNPs that are associated with 53 quantitative traits across ~300 K unrelated white British individuals in the UK Biobank. Testing 15,601 trait-loci associations, we identified 16 trait-loci pairs across 12 traits that demonstrate strong evidence of ME signals (p-value $p < 5e-8/53$). We further partitioned the significant ME signals across the genome to identify 6 trait-loci pairs with evidence of local (within-chromosome) ME while 15 show evidence of distal (cross-chromosome) ME. Across the 16 trait-loci pairs, we document that the proportion of trait variance explained by ME is about 12x as large as that explained by the GWAS effects on average (range: 0.59 to 43.89). Our results provide evidence for epistatic interactions modulating the effects of genetic variants on complex traits.

▪ *Calibrated prediction intervals for polygenic scores across diverse contexts*



Kangcheng Hou¹, Ziqi Xu², Yi Ding¹, Arbel Harpak^{3,4}, Bogdan Pasaniuc^{1,5,6,7}

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

² Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA

³ Department of Population Health, The University of Texas at Austin, Austin, TX, USA

⁴ Department of Integrative Biology, The University of Texas at Austin, Austin, TX, USA

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

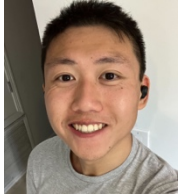
⁶ Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

⁷ Institute for Precision Health, University of California, Los Angeles, Los Angeles

Polygenic scores (PGS) have emerged as the tool of choice for genomic prediction in a wide range of fields. We analyze data from two large biobanks in the US and the UK to find widespread variability in PGS performance across contexts. Many contexts, including age, sex, and income, impact PGS accuracies with similar magnitudes as genetic ancestry. We introduce trait prediction intervals as a principled approach to account

for context-specific PGS. Our approach enables PGS-based trait predictions that are well-calibrated. We show that prediction intervals need to be adjusted for all considered traits. Adjustment of prediction intervals are dataset- and trait-specific; for example, prediction intervals for education years need to be adjusted by 90% in All of Us versus 8% in UK Biobank. Our results provide a path forward towards using PGS as a prediction tool across all individuals regardless of their contexts and highlight the importance of comprehensive profile of context.

▪ ***Inference of demographic histories and distributions of fitness effects of deleterious mutations from human gut microbiomes***



Jonathan C. Mah¹, Kirk E. Lohmueller^{2,3}, Nandita Garud^{2,3}

¹ Bioinformatics Interdepartmental Program,

² Department of Ecology and Evolutionary Biology, University of California, Los Angeles

³ Department of Human Genetics, University of California, Los Angeles

Human commensal gut microbes are crucial to host health, with functions including aiding digestion and metabolizing drugs. Despite their importance, we lack insight into the evolutionary histories of gut microbiota, including their population demographic histories and distributions of fitness effects (DFE). Here, we infer the demography and DFEs of 27 gut commensal microbiota prevalent in North Americans. We find overall reductions in genetic variation relative to microbes sampled from non-Western rural populations. Additionally, we infer varied demographic histories of North American gut microbiota, including contractions and expansions. For example, *Akkermansia muciniphila* displays a contraction ~10,000 years ago, coincident with the onset of agriculture. DFEs across species vary from highly to mildly deleterious, with accessory genes displaying less deleterious DFEs than core genes. Within genera, DFEs tend to be more congruent, reflective of underlying phylogenetic relationships. Taken together, these findings suggest that human commensal gut microbes have distinct evolutionary histories.

▪ ***Tensor factorization maps dysregulation of immune signaling in breast cancer patients***



Brian Orcutt-Jahns¹, Andrei Rodin², Joao Rodrigues Lima Junior², Peter Lee², Aaron Meyer^{1,3,4,5}

¹ Department of Bioengineering, University of California, Los Angeles

² City of Hope, Duarte, CA, United States of America

³ Department of Bioinformatics, University of California, Los Angeles, United States of America

⁴ Jonsson Comprehensive Cancer Center, University of California, Los Angeles, United States of America

⁵ Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, United States of America

Metastatic cancer correlates with dysregulation of immune signaling and function. For example, impaired T cell responsiveness to IL-6 has been shown to predict poor clinical outcome in ER⁺ breast cancer (BC). To systematically profile immune dysregulation in cases of BC, we stimulated human PBMCs from 10 healthy and 11 BC patients with a panel of 11 cytokines and growth factors. Cells were stained for 25 canonical cell type markers and 5 intracellular signaling proteins, allowing for the dissection of the cytokine responses of 30 immune cell subtypes. We arranged our cytokine response data into a 5D tensor and analyzed responses using tensor factorization techniques. Using this approach, we show that reduced responsiveness to IL-10 and increased STAT5 phosphorylation are predictive of BC disease state. Our approach lays the groundwork for analyzing the inherently tangled network of immune dysregulation common in BC.

▪ ***Long-read RNA-seq demarcates cis- and trans-directed alternative RNA splicing***



Giovanni Quinones-Valdez¹, Kofi Amoah², Xinshu Xiao^{1,2*}

¹ Department of Integrative Biology and Physiology

² Bioinformatics Interdepartmental Program

University of California, Los Angeles, Los Angeles, CA 90095, USA

RNA splicing results from the interplay of cis- and trans-acting elements; however, their respective contributions to specific splicing events are not clearly established. Here, we leveraged long-read RNA sequencing data to distinguish splicing events primarily directed by cis- or trans-regulatory mechanisms using our novel method, isoLASER. Analyzing human and mouse data, isoLASER identified 2,047 and 4,679 unique exonic regions, respectively, exhibiting cis-directed splicing. While most alternatively spliced exons are predominantly regulated by tissue-specific trans-acting factors, a subset of cis-regulated exons remains consistently spliced across tissues with shared genetic backgrounds. These cis-directed regions show low conservation across vertebrates and are enriched in immune-related loci. Furthermore, isoLASER enables cohort-level analysis, unveiling splicing-associated variants (SAVs). Notably, these SAVs exhibit significant alterations in RNA-protein binding and demonstrate high reproducibility in splicing quantitative trait loci (sQTLs) studies. In summary, we provide a framework for identifying genetically and non-genetically driven splicing and the associated variants in cohorts with limited sample sizes.

▪ **Graph homogeneity analysis of single-cell epigenetic states**



Breanne Sparta^{1,2}, Timothy Hamilton^{1,3}, and Eric J. Deeds^{1,2,3}

¹Institute for Quantitative and Computational Biosciences, University of California, Los Angeles,

²Department for Integrated Biology and Physiology, University of California, Los Angeles,

³Bioinformatics Interdepartmental Program, University of California, Los Angeles

The prevailing interpretation of Waddington's landscape is that attractors in gene expression space produce and stabilize distinct cell types. This notion is often applied in single-cell omics data, where groups of cells are clustered for analysis. We applied graph theory to characterize the distribution of cells in epigenetic space, using data from various tissues and organisms as well as various single-cell omics technologies. We found that cell types exist in the same regions of epigenetic space, with highly heterogeneous density distributions that are inconsistent with expected densities near an attractor. The lack of attractor structure could not be explained by technical noise, scale variance among genes, nor the subset of genes that were used; nor could it be rescued by any standard set of transformations. These findings pose a challenge for the robust analysis of single-cell data and open the possibility for alternative explanations of canalization during development. To address these practical issues, here, we develop a graph-homogeneity approach to characterize how tissue composition changes in health and disease.

▪ **EMulSI-Phy: Efficient Multi-Sample Inference of Cancer Phylogeny**



Helena K. Winata^{1,2}, Dan Knight^{1,2}, Pier Selenica³, Nicholas K. Wang^{1,2}, Caroline Kostrzewa⁴, Stefan Eng^{1,2}, Yingjie Zhu³, Jaron Arbet^{1,2}, Juber A. Patel³, Ronglai Shen⁴, Jorge-Reis Filho³, Pedram Razafi⁵, Paul C. Boutros^{1,2}

¹Department of Human Genetics, University of California, Los Angeles, CA, USA

²Jonsson Comprehensive Cancer Centre, University of California, Los Angeles, CA, USA

³Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

⁴Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

⁵Department of Medicine, Memorial Sloan Kettering Cancer Center and Weill Cornell Medical College, New York, NY, USA

Inferring tumor evolution from DNA sequencing data is crucial for understanding cancer progression and mutation patterns within clones. Decreasing sequencing costs enable routine sequencing of multiple tumor samples per patient, offering an opportunity to explore tumor evolution with unprecedented resolution across spatial regions and longitudinal timeframes. However, growing dataset complexity presents a challenge, necessitating innovative methods to transcend the computational constraints faced by current methods. Hence, we developed EMulSI-Phy, a rule-based algorithm that leverages the fundamental principles of cancer biology to enable more rapid and accurate subclonal reconstruction. Benchmarking on 576 simulated datasets demonstrated its computational efficiency and accuracy in resolving complex evolutionary structures. Applied to 12 breast cancer patients with 4-36 sequenced tumor regions, EMulSI-Phy reduced average runtime ten-fold, delineating and reconstructing the phylogeny of up to 88 subclones per patient, which is unfeasible with most methods.

Collaboratory Fellows 2023-2024

Matteo Pellegrini, **Director**



Eloy Lopez, **Program Manager**



Fei-Man Hsu
New Fellow



Daniel Ha



Kelsey Jorgensen



Giorgia Del Vecchio



Ivy Xiong
New Fellow



Shawn Cokus



Lukasz Salwinski



Seyoon Ko



Wenbin Guo



Don Vaughn



Weihong Yan



Karolina Kaczor-Urbanowicz



Giovanni Quinones Valdez



Lingyu Zhan

<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>

Welcome new Faculty!

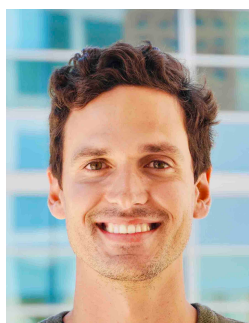


Brunilda Balliu, Ph.D.

Assistant Professor

Departments of Pathology & Laboratory Medicine and Computational Medicine

Dr. Balliu is an Assistant Professor in the Departments of Pathology and Laboratory Medicine, Computational Medicine, and Biostatistics at UCLA. Her lab develops novel statistical methods and computational tools for analyzing high-dimensional repeated measures and intensive longitudinal data arising from high-throughput genomic assays, mobile phone and wearable sensors, and electronic health records. She applies these methods to understand the genetic, molecular, cellular, and environmental mechanisms underlying complex human traits and diseases, with a focus on metabolic and psychiatric phenotypes. She is particularly interested in understanding the role of inherited genetic variation in disease through the diversity of changes observed in gene expression under different contexts, e.g. age, tissues & cell types, environmental perturbations, and developmental stages.



Mehdi Bouhaddou, Ph.D.

Assistant Professor

Microbiology, Immunology, and Molecular Genetics (MIMG)

Dr. Mehdi Bouhaddou performed his postdoctoral training with Dr. Nevan J. Krogan at UC San Francisco (UCSF) in virology, mass spectrometry proteomics, bioinformatics, and network modeling as a member of the Quantitative Biosciences Institute (QBI) Coronavirus Research Group (QCRG). During his postdoc, Dr. Bouhaddou received F32 (NCI) and K99 (NIAID) awards to study phosphorylation signaling and protein-protein interactions in the context of infectious disease and cancer, co-mentored by Danielle L. Swaney. He developed virus-host interaction networks for SARS-CoV-2 and other coronaviruses, and systematically compared the molecular response to emerging SARS-CoV-2 variants to pinpoint variant-specific mechanisms of pathogenesis. Prior to his postdoc, Dr. Bouhaddou worked at Roche with Drs. Li Yu and Antje-Christine Walz to develop pharmacokinetics and pharmacodynamics (PK/PD) mathematical models of epigenetic modifier drugs in cancer. He received his PhD in Biomedical Sciences advised by Dr. Marc Birtwistle at the Icahn School of Medicine at Mount Sinai in New York City, where he developed ordinary differential equation (ODE) models of cancer signaling to predict personalized therapeutic strategies tailored to specific cancer mutational contexts. Lastly, Dr. Bouhaddou received his Bachelor's degree from UC Berkeley in Cognitive Neuroscience.



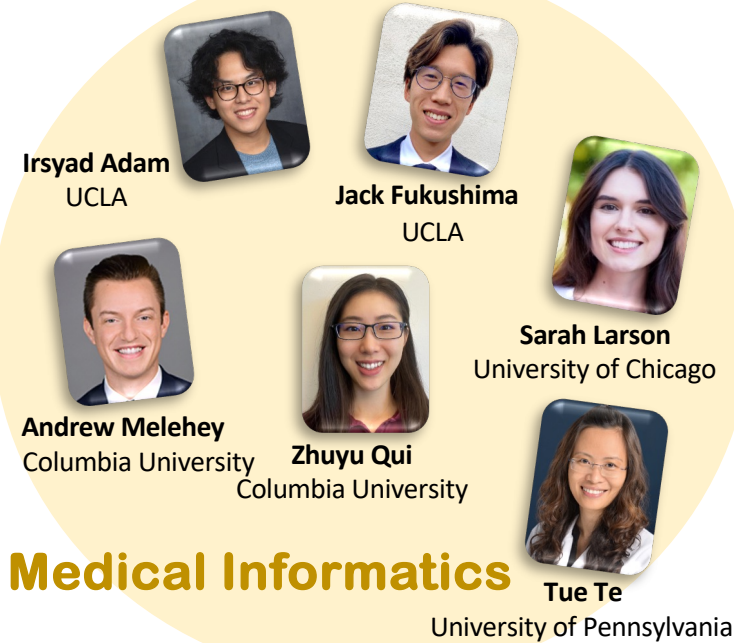
Danielle Schmitt, Ph.D.

Assistant Professor

Chemistry and Biochemistry

Dr. Danielle L. Schmitt is an Assistant Professor in the UCLA Department of Chemistry and Biochemistry. The Schmitt Lab develops genetically encoded tools for quantitative imaging to determine mechanisms for the compartmentalized regulation of cellular metabolism in health and disease. Prior to joining UCLA, Dr. Schmitt earned her B.S. in Chemistry at Ball State University, her PhD in Chemistry and Biochemistry at University of Maryland Baltimore County with Dr. Songon An. Most recently, Dr. Schmitt was a University of California President's Postdoctoral Fellow at University of California San Diego working with Dr. Jin Zhang.

Welcome Incoming Students!



Medical Informatics

Irsyad Adam
UCLA

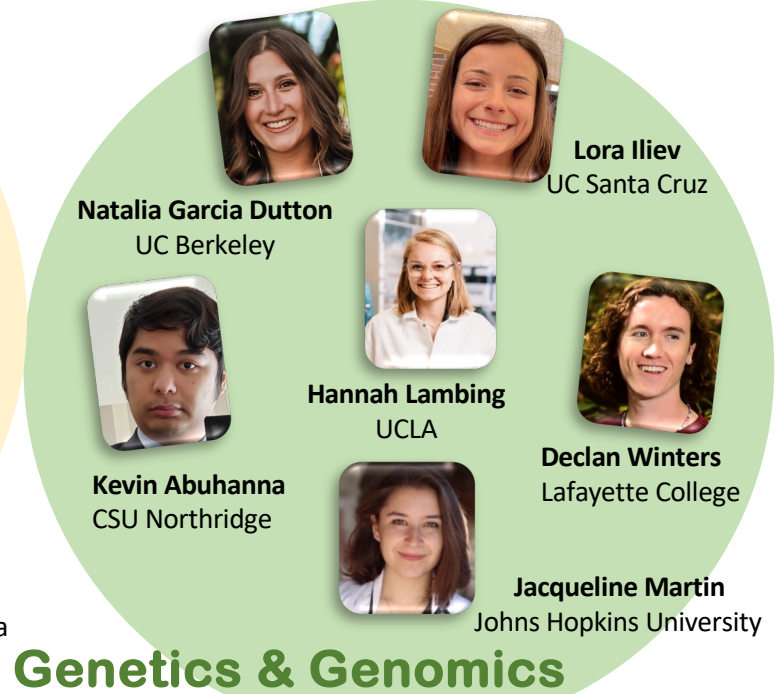
Jack Fukushima
UCLA

Andrew Melehey
Columbia University

Zhuyu Qui
Columbia University

Sarah Larson
University of Chicago

Tue Te
University of Pennsylvania



Genetics & Genomics

Natalia Garcia Dutton
UC Berkeley

Lora Iliev
UC Santa Cruz

Hannah Lambing
UCLA

Kevin Abuhanna
CSU Northridge

Declan Winters
Lafayette College

Jacqueline Martin
Johns Hopkins University



Biomathematics

Simon Lee
UC Santa Cruz
B.I.G. SUMMER 2022 ALUMNUS

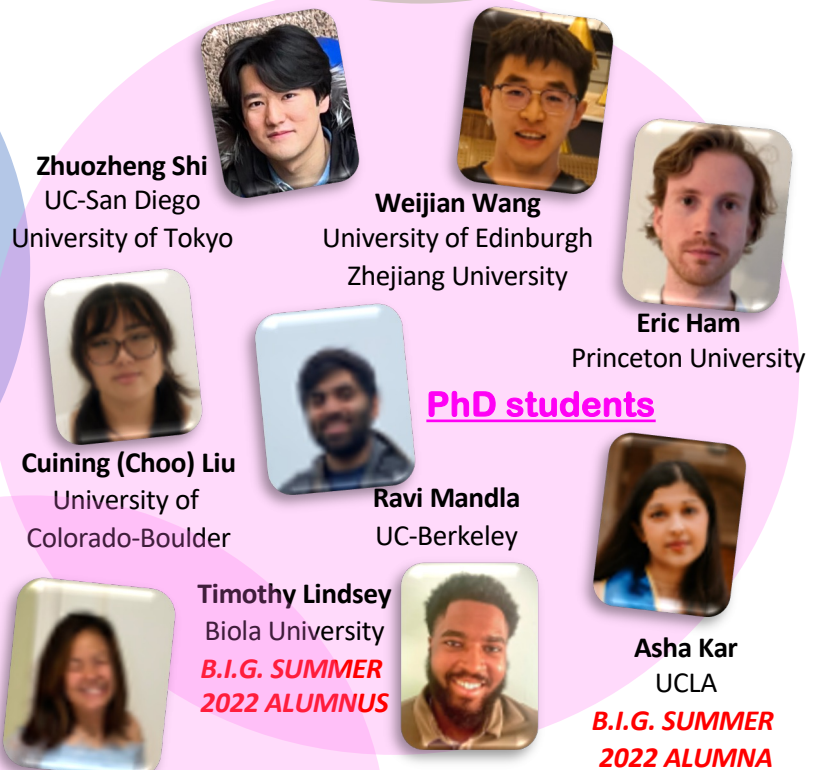
Yihui Cen
BS Liverpool University

Sayun Mao
Peking University

Arabdhia Biswas
UC Irvine

Jiahang (Hank) Sha
University of Pennsylvania

Zachary Schlamowitz
University of Arizona



PhD students

Zhuozheng Shi
UC-San Diego
University of Tokyo

Weijian Wang
University of Edinburgh
Zhejiang University

Eric Ham
Princeton University

Cuining (Choo) Liu
University of Colorado-Boulder

Ravi Mandla
UC-Berkeley

Timothy Lindsey
Biola University
B.I.G. SUMMER 2022 ALUMNUS

Asha Kar
UCLA
B.I.G. SUMMER 2022 ALUMNA



Masters Students

Calvin Lee
UC-San Diego

Melanie Tu
UCLA

Ronan Bennett
UCLA

Tadeo Spencer
UC-San Diego

Bioinformatics

Poster Session I – Lunch

1. The impact of sex and ancestry on the molecular hallmarks of pancreatic adenocarcinoma

Caroline Y. Chen^{1,2,3,4}, Roni Haas^{3,4,5}, Takafumi Yamaguchi^{3,4,5,8}, Constance Li^{7,8}, Paul C. Boutros^{3,4,5,6,8}

Department of Medicine Division of Hematology-Oncology, UCLA, USA¹

Bioinformatics Interdepartmental Program, UCLA, USA²

Jonsson Comprehensive Cancer Center, UCLA, USA³

Institute for Precision Health, UCLA, USA⁴

Department of Human Genetics, UCLA, USA⁵

Department of Urology, UCLA, USA⁶

National Cancer Centre Singapore, Singapore⁷

Department of Medical Biophysics, University of Toronto, Toronto, Canada⁸

Pancreatic adenocarcinoma (PDAC) is a highly lethal cancer of the exocrine pancreas and is projected to become the second leading cause of cancer mortality in the next decade. There is no robust screening method for the general population. PDAC is one of the most heritable cancers, with 1 in 10 cases associated with rare germline variants and GWAS has demonstrated that common variants play a role as well. Little is known about the interplay between germline and the development of somatic mutations in PDAC. In this work, we use publicly available cancer datasets to evaluate the impact of sex and ancestry on mutational load, top genes impacted by single nucleotide variants, copy number profile and RNA abundance in PDAC patients. Further investigation of germline-somatic interactions is warranted to understand the underlying mechanisms behind tumorigenesis and may result in more accurate screening practices, patient selection for various therapies and prognostication.

2. Human protein adaptation in response to viruses can happen through protein stability evolution

Chenlu Di^{1,2}, Jesus Murga-Moreno¹, David Enard¹

¹Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, USA

²Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles California, USA

Pathogens were important drivers of human protein adaptation. Viruses in particular drive a large amount of adaptation in the thousands of proteins that physically interact with viruses (VIPs for Virus-Interacting Proteins). However, numerous adaptations cannot be explained by a few known functioning amino acids. Here, we hypothesize that the evolution of protein stability, the balance between the folded and unfolded proteins, is a potential mechanism of virus-driven human protein adaptation in VIPs. We inferred human protein stability changes in high-throughputs based on AlphaFold 2 structures and found amino acid mutations that altered stability experienced highly elevated adaptive evolution in VIPs. We further find that stability in immune antiviral VIPs evolved under directional selection while VIPs needed by viruses evolved under compensatory evolution following viral epidemics. Together, these results suggest that stability evolution, and thus functional host protein abundance evolution, was a major mechanism of host protein adaptation during viral epidemics.

3. Improving genetic risk modeling of dementia from real-world data in underrepresented populations

Mingzhou Fu^{1,2}, Leopoldo Valiente-Banuet¹, Satpal S.

Wadhwa¹, UCLA Precision Health Data Discovery Repository Working Group, UCLA Precision Health ATLAS Working Group, Bogdan Pasaniuc³, Keith Vossell¹, Timothy S. Chang^{1*}

¹Mary S. Easton Center for Alzheimer's Research and Care, Department of Neurology,

²Medical Informatics Home Area, Department of Bioinformatics,

³Department of Computational Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

Early genetic risk modeling for broader dementia categories significantly benefits patients and healthcare providers. There has been a lack of studies developing dementia genetic risk models based on real-world data, particularly for underrepresented populations. Our study aimed to enhance dementia risk prediction models, specifically tailored for individuals from diverse genetic ancestries using electronic health record data. We proposed an Elastic Net model for individual dementia risk prediction using gene-annotated single nucleotide polymorphisms from various neurodegenerative disease genome-wide studies, which significantly improves prediction performance compared to *APOE* gene and polygenic risk score models across Amerindian, African, and East Asian genetic ancestry. We identified both shared and ancestry-specific risk genes and biological pathways contributing to dementia risks, which align with previous research while also adding new insights to the existing knowledge. Insights gained from our research present promising prospects for advancing precision medicine approaches in dementia diagnosis and risk assessment.

4. Modeling the heterogeneous NFkB dynamics of single immune cells

Xiaolu Guo, Adewunmi Adelaja, Supriya Sen, Alexander Hoffmann

Institute for Quantitative and Computational Biosciences, and Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, USA

Macrophages initiate pathogen-appropriate immune responses, with transcription factor NFkB dynamics mediating the specificity. Our laboratory generated extensive single-cell NFkB dynamic data using a fluorescent reporter mouse. Here, we developed a mathematical model of the signaling network that recapitulates the stimulus-specific but highly cell-to-cell heterogeneous NFkB dynamics. Biologically-varied parameters were fit with a non-linear mixed effect model. Visual inspection and quantitative evaluation revealed an excellent concordance between simulations and experimental data. The model allowed identifying biochemical reactions that cause heterogeneity in NFkB dynamics; investigating the dose response behavior of each ligand; exploring how different ligand responses combine in combinatorial ligands stimulation conditions. Each of these led to new insight that were experimentally tested and validated. Our work presents a new computational research tool for studying NFkB dynamics in single immune cells.

5. Machine learning to predict ceftriaxone resistance using single nucleotide polymorphisms within a global database of *Neisseria gonorrhoeae* genomes

Sung Min Ha¹, Eric Y. Lin², Jeffrey D. Klausner³, and Paul C. Adamson⁴

¹Department of Integrative Biology and Physiology, UCLA, Los Angeles, California, USA

²David Geffen School of Medicine at UCLA, Los Angeles, California, USA

³Departments of Population and Public Health Sciences and Medicine, Keck School of Medicine of University of Southern California, Los Angeles, California, USA

⁴Division of Infectious Diseases, David Geffen School of Medicine at UCLA, Los Angeles, California, USA

Gonorrhea, a sexually transmitted infection caused by the bacterium *Neisseria gonorrhoea*, is a pressing global health issue due to increasing antimicrobial resistance (AMR). This study employs a comprehensive dataset from the PathogenWatch database, featuring 12,936 *N. gonorrhoea* genomes, to train machine learning models aimed at predicting susceptibility to ceftriaxone, a cornerstone antibiotic in gonorrhea treatment. Utilizing a random forest algorithm, the model achieved a high level of predictive accuracy, with an AUC score of 0.965. Additionally, the study identified a select combination of single nucleotide polymorphisms (SNPs) that performed almost as well as using all 97 known SNPs related to ceftriaxone resistance. These groundbreaking findings not only demonstrate the efficacy of machine learning algorithms in predicting AMR but also offer practical applications for the development of advanced diagnostic tests, such as nucleic acid amplification tests (NAATs), which could revolutionize treatment protocols and curb the rise of AMR.

6. Hierarchical modeling and simulation of Cell Villages to improve experimental design

Chloe Hanson¹, Rachel Fox², Tim Derebenskiy¹, Michael Wells¹ and Harold Pimentel^{1,3}

¹ Department of Human Genetics,

² Department of Neuroscience,

³ Department of Computational Medicine,

University of California Los Angeles, Los Angeles California, USA

Cell Villages are a cutting edge tool that leverage pooled populations of 10-100 genetically distinct, human stem cell lines to study how genetic variation and environment influence a spectrum of molecular phenotypes. This platform allows researchers to apply selection and perturbation treatments to large, diverse cell populations, while simultaneously measuring phenotype differences between donors through molecular and omics assays. While Villages show a lot of promise, experimental methods need to be refined to ensure reproducibility and adequate statistical power. To address this limitation we developed a modeling framework of Village donor growth variation to assess how different experimental parameters affect reproducibility and the signals that can be detected in the data. Using this experimentally validated simulation framework I will build a user-friendly tool that allows researchers to input their own preliminary data to project how different design choices might impact experimental results.

7. The role of APOBEC3-induced mutations in the differential evolution of monkeypox virus

Xiangting Li^{1*}, Sara Habibipour^{2*}, Tom Chou^{1,3}, Otto O. Yang²

¹Department of Computational Medicine,

²Depts. of Medicine and Microbiology, Immunology, and Molecular Genetics,

³Department of Mathematics,

University of California Los Angeles, Los Angeles California, USA

These authors contributed equally.

Recent studies show that newly sampled monkeypox virus (MPXV) genomes exhibit mutations consistent with Apolipoprotein B mRNA Editing Catalytic Polypeptide-like3 (APOBEC3)-mediated editing, compared to MPXV genomes collected earlier. It is unclear whether these single nucleotide polymorphisms (SNPs) result from APOBEC3-induced editing or are a consequence of genetic drift within one or more MPXV animal reservoirs. We develop a simple method based on a generalization of the General-Time-Reversible (GTR) model to show that the observed SNPs are likely the result of APOBEC3-induced editing. The statistical features allow us to extract lineage information and estimate evolutionary events.

8. Graph-Based Analysis of Cell Lineages

Gunalan Natesan, Timothy Hamilton, Eric J. Deeds, Pavak K. Shah

¹Department Molecular, Cell, and Developmental Biology

²Bioinformatics Interdepartmental Program,

University of California Los Angeles, Los Angeles California, USA

Cell lineages in eutelic organisms, which possess a fixed number of somatic cells, have been a powerful tool in understanding development. *C. elegans*' eutelic characterization motivates the creation and benchmarking of tree distance metrics, noting the intuitive map of lineages to binary trees. We adapt the tree edit distance, which measures topological variation, and introduce the branch distance, an extension of the L2 norm, to analyze developing *C. elegans* lineages. We benchmark these metrics using a published database of wild type and RNAi-perturbed *C. elegans* embryos, revealing previously uncharacterized heterogeneity in effects of RNAi variability on developmental timing. These measurements identify an interesting role of Notch in the control of developmental timing. We analyze RNAi perturbations resulting in cell fate transformations where we find that, while developmental timing appears to be highly sensitive to genetic perturbation, RNAi against certain developmental regulators generate transformations that preserve lineage-specific developmental clocks.

9. Interpretable and scalable integration of single-cell measurements across conditions with PARAFAC2

Andrew Ramirez¹, Brian Orcutt-Jahns¹, Aaron Meyer¹

¹Department of Bioengineering, University of California, Los Angeles, CA 90024, USA

Analyzing how populations of single cells respond to multiple conditions or perturbations is a challenging task for current computational methods. To address this issue, we propose using PARAFAC2 (Pf2), a tensor decomposition method that can reduce the complexity of single-cell data in multi-condition and perturbational studies. By summarizing the data in a reduced space that explicitly separates alignment of cells from condition-specific effects, we can analyze how experimental conditions cause similar changes across single-cell experiments and how cells vary within experimental conditions. This version of irregular tensor decomposition (TD) is compatible with diverse single-cell methods and offers a significant improvement in the versatility of single-cell analyses. By harnessing the power of

irregular TD to track patterns of cellular response, our study represents a major step forward in single-cell applications.

10. Rosace: a Bayesian framework for analyzing growth-based deep mutational scanning data

Jingyou Rao¹, Christian Macdonald², Willow Coyote², Harold Pimentel^{1,3}

¹ Department of Computer Science, University of California Los Angeles, Los Angeles California, USA

² Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco California, USA

³ Department of Computational Medicine and Human Genetics, University of California Los Angeles, Los Angeles California, USA
Mutations provide information about how changes in gene sequences can impact function. Deep mutational scanning (DMS) enables functional insight into protein mutations with multiplexed measurements of thousands of genetic variants in a protein simultaneously. Here we present Rosace, a Bayesian framework for analyzing growth-based deep mutational scanning data. Rosace is the first method to leverage amino acid position information to increase power and provides reasonable shrinkage to control the false discovery rate. To benchmark Rosace against existing methods, we develop Rosette, a simulation framework that simulates the distributional properties of DMS. Importantly, Rosace and Rosette are not two views of the same model, and thus, we implicitly show the robustness of Rosace. Finally, we run Rosace on real datasets and it shows a much lower false discovery rate than existing methods while discovering almost all the validated variants.

11. Know2BIO: A Biomedical Knowledge Graph Benchmark for Machine Learning Methods

Dylan Steinecke², Yijia Xiao¹, Alexander Pelletier^{1,3}, Yushi Bai⁶, Peipei Ping^{3,4}, Wei Wang^{1,5}

¹ Department of Computer Science, UCLA, Los Angeles, California

² Medical Informatics Home Area, UCLA, Los Angeles, California

³ Department of Physiology, UCLA, Los Angeles, California

⁴ Department of Medicine, UCLA, Los Angeles, California

⁵ Department of Computational Medicine, UCLA, Los Angeles, California

⁶ Department of Computer Science, Tsinghua University, China
Knowledge graphs (KGs) have emerged as a powerful framework to represent and integrate complex biomedical information. However, there is a scarcity of biomedical KGs that assemble the most scientifically up-to-date data from diverse sources representing an abundance of biomedical data types from multiple modalities. To overcome these challenges, we propose Know2BIO, a general-purpose heterogeneous KG benchmark for the biomedical domain. Know2BIO integrates data from 30 diverse sources, capturing intricate relationships across 11 biomedical categories in ~219,000 nodes and ~6,200,000 edges. Know2BIO can be automatically updated to reflect the latest knowledge in biomedical science. Furthermore, Know2BIO can be enriched with multi-modal data: node features of natural language descriptions, molecular sequences, and molecular structures. We evaluate KG representation models on Know2BIO, demonstrating its effectiveness as a benchmark for KG representation learning in the biomedical field. Data and source code of Know2BIO are available at <https://github.com/Yijia-Xiao/Know2BIO>.

12. Mechanistic binding model quantifies specific antibody species from systems serology

Cyrrillus Tan¹, Armaan Abraham², Priyanka Shrestha³, Emily Bozich², Aaron S. Meyer^{1,2}

¹Bioinformatics Interdepartmental Program,

²Department of Bioengineering, University of California Los Angeles, Los Angeles, CA, USA

³Department of Computer Science, Stanford University, Palo Alto, CA, USA

Antibodies direct humoral immunity by binding to antigens specifically and signaling to immune effector cells via their Fc region. During infection or immunization, antibodies' Fc composition varies by subclass and glycan and directs the type of immune response. Systems serology assays profile serum antibodies by antigen specificity and Fc interactions, but these data are typically only used to identify correlative relationships to clinical outcomes. Here, to tie these measurements to specific molecular Fc features, we inferred antibody composition from measured fluorescence with a mechanistic binding model. We first validate the model's capability of quantifying different Fc domains in mixtures through simulations. Running it on COVID-19 datasets, we found that the model can impute unmeasured Fc interactions and discern patterns in antibody fucosylation. It also identified information-rich detection reagents, which can be used to optimize future assays. We expect this study to improve the application and granularity of systems serology assays.

13. Direct observation correlates NFκB cRel in B-cells with activating and terminating their proliferative program.

Haripriya Vaidehi Narayanan¹, Mark Y Xiang¹, Helen Huang¹, Sukanya Roy², Himani Makkar², Alexander Hoffmann¹, Koushik Roy²

¹Signaling Systems Laboratory, Department of Microbiology, Immunology, and Molecular Genetics, Institute for Quantitative and Computational Biosciences, University of California Los Angeles, Los Angeles, CA 90095.

²Division of Microbiology and Immunology, Department of Pathology, University of Utah, Salt Lake City, UT 84112.
B-cells are capable of substantial stimulus-responsive proliferation, but also show substantial cell-to-cell proliferative heterogeneity. Gene-knockout models established that NFκB cRel is essential for stimulus-driven B-cell proliferation, but whether its expression variability determines proliferative heterogeneity remains unknown. We developed a fluorescent reporter mTFP1-cRel mouse to directly observe natural cRel variation in B-cells and relate it to proliferation kinetics. We found that cRel abundance is heterogeneously distributed among naïve B-cells, enriched for high-expressors in a heavy-tailed distribution. High-cRel-expressors show fast activation of the B-cell proliferative program, but it isn't well sustained, with reduced overall population expansion. Using a molecular network model, we showed that cRel heterogeneity can arise from balancing positive feedback by auto-regulation and negative feedback by its inhibitor IκBe. By knocking out IκBe, we confirmed two-pronged control of the B-cell proliferative program by cRel. These findings emphasize the power of direct observation to understand how natural variations control biological function.

14. Sexual dimorphism in renal metabolism, hemodynamics and diseases

Ivy Xiong, Alan Garfinkel, Eric J. Deeds

Department of Integrative Biology and Physiology

University of California, Los Angeles, CA 90095, USA

Mammalian organs exhibit distinct physiology, disease susceptibility and injury responses between the sexes. Especially, males have increased susceptibility to both chronic and acute kidney diseases than in females. Recent work suggests that male proximal tubules (PTs) undergo excessive oxidative stress to meet the high energetic demand, while female PTs exhibit an anti-oxidation state. However, it remains elusive how the observed molecular differences relate to sex disparities in renal physiology. Widely known as an indicator of renal function, glomerular filtration rate (GFR) shows a sustained oscillatory pattern over time with a period of 30-45 seconds in rodents, and loss of GFR oscillations is associated with cessation of reabsorption activities and ischemia-reperfusion injuries in the kidney, as well as systemic hypertension. To study the relationship between intracellular events and organ physiology, we're developing a mathematical model of TGF linking metabolic regulation within PT cells to fluid handling and salt reabsorption in the nephron, using intra-vital imaging data from both sexes. We aim to perform bifurcation analysis and provide a mechanistic explanation for sexual dimorphism in kidney diseases.

1ng/mL were stratified into risk groups by the polygenic hazard score PHS290.

15. Evaluating biomarker potential of germline genomic factors for predicting clinical outcomes in prostate cancer.

Nicole Zeltser¹⁻³, Roni Haas¹⁻³, Kathleen Houlihan¹⁻³, Tyler Seibert⁴, Sarah Al-Hiyari¹⁻³, Stefan Eng¹⁻³, Yash Patel¹⁻³, Takafumi Yamaguchi¹⁻³, Shu Tao¹⁻³, Rong-Rong Huang⁵, Robert Reiter^{3,6}, Huihui Ye⁵, Adam Kinnaird⁷, Michael Liss⁸ and Paul Boutros^{1-3,6}

¹Department of Human Genetics, University of California, Los Angeles, CA, USA

²Institute for Precision Health, University of California, Los Angeles, CA, USA

³Jonsson Comprehensive Cancer Centre, University of California, Los Angeles, CA, USA

⁴University of California, San Diego, United States

⁵Department of Pathology and Laboratory Medicine, University of California, Los Angeles, USA

⁶Department of Urology, University of California, Los Angeles, CA, USA

⁷University of Alberta, Edmonton, Alberta, Canada

⁸University of Texas Health Science Center, San Antonio, TX, United States

Prostate cancer is the second-most diagnosed cancer and the second leading cause of cancer death in American men. Early detection is common, but is followed by the more challenging task of prognosing a variable clinical course. Current clinical risk-assessment strategies are highly imprecise. An improved method of risk stratification may lie in hereditary factors. Prostate cancer is highly heritable, with accumulating evidence associating rare and common variants and genetic ancestry to clinical outcomes. We developed a low-cost, prostate cancer-specific, targeted germline sequencing panel, that will pave the way for further exploration of inherited prostate cancer risk. In a pilot cohort of 48 prostate cancer patients, captured variants were successfully used to assess pathogenicity, calculate polygenic risk and infer genetic ancestry. In a second cohort of 310 patients in the SABOR study, men with baseline PSA $\geq$

Poster Session II - Afternoon

1. Landscape of shared genomic features in prostate cancer model systems

Raag Agrawal^{1,2,3,4}, Paul C. Boutros^{1,2,3,4,5}

¹ Department of Human Genetics, University of California, Los Angeles

² Department of Urology, University of California, Los Angeles

³ Jonsson Comprehensive Cancer Center, University of California, Los Angeles

⁴ Institute for Precision Health, University of California, Los Angeles

⁵ Department of Medical Biophysics, University of Toronto
Cancer cell lines are important tools as they allow us to study the behavior and genetics of cancer cells in a controlled environment. Despite their widespread use in research, the vast majority of prostate cancer model systems have not been DNA whole genome sequenced. Here, we sequence the most used prostate cancer model systems and, using a publicly available patient cohort, identify missing subtypes. We further quantify genomic features such as total mutational load, mutational signatures, driver mutations, and evolutionary features like mutational timing. These results fill a key gap in our understanding of prostate cancer model systems and creates a resource for the community to improve our understanding of model system representativeness, and ultimately lead to improved drug-screening studies.

2. Detection of Symptoms of Depression Using Data From the iPhone and Apple Watch

Samir Akre¹, Zachary D. Cohen², Amelia Welborn², Tomislav D. Zbozinek², Brunilda Balliu³, Jonathan Flint², Alex A.T. Bui¹, Michelle G. Craske²

¹ Medical Informatics Home Area,

² Department of Psychiatry and Biobehavioral Sciences

³ Department of Computational Medicine

University of California Los Angeles, Los Angeles California, USA
Digital health data from consumer wearable devices and smartphones have the potential to improve our understanding of mental illness. However, in conditions like depression, there is not yet a consistent uniform measurement tool whose result can be reliably used as a gold standard measure of depression severity. This work seeks to specify what symptoms and dimensions of depression can be detected using vitals, activity, and sleep monitored by consumer wearable devices. Machine learning models are fit to digital health data and used to detect responses to individual questions from self-reports as well as summary scores. Data is analyzed from 99 participants of an ongoing study with data from the Apple Watch, iPhone, and validated self-reports. The digital health data investigated was found to detect depression severity and specific symptoms like poor appetite, aspects of anhedonia, and sleep timings (ROC AUC 0.63 to 0.72).

3. Preclinical Side Effect Prediction through Pathway Engineering of Protein Interaction Network Models

Mohammadali Alidoost¹, Jennifer L. Wilson¹

¹ Department of Bioengineering, University of California Los Angeles, Los Angeles California, USA

Ensuring the safety of new drug development is crucial to avoid potential adverse effects and subsequent market failures. In this

study, we aimed to enhance the preclinical prediction of drug-induced safety events by leveraging protein interaction network models and pathway engineering strategies. We employed the PathFX protein-protein interaction tool to predict drug effects, using a drug toxicity dataset. To improve prediction accuracy and reduce over-prediction, we defined new pathways and implemented pathway engineering. By incorporating network proteins associated with true positive drug networks, we successfully eliminated over-prediction for side effects with sufficient true positive examples. Additionally, we leveraged an omics dataset to generate novel side effect pathways, improving side effects prediction. We ultimately evaluated our predictions by comparing them to the results of an animal testing study and demonstrated a trade-off between specificity and sensitivity values. Our study underscores the importance of side effect pathways in drug effect anticipation.

4. Long-read RNA-seq demarcates *cis*- and *trans*-directed alternative RNA splicing

Kofi Amoah², Giovanni Quinones-Valdez¹, Xinshu Xiao^{1,2}

¹ Department of Integrative Biology and Physiology

² Bioinformatics Interdepartmental Program

University of California, Los Angeles, Los Angeles, CA 90095, USA

RNA splicing results from the interplay of *cis*- and *trans*-acting elements, yet their contribution to specific splicing events is not clearly established. Here, we leveraged long-read RNA sequencing data to distinguish splicing events primarily directed by *cis*- or *trans*-regulatory mechanisms, using our novel method isoLASER. Analyzing human and mouse data, isoLASER identified 2,047 and 4,679 unique exonic regions, respectively, exhibiting *cis*-directed splicing. While most alternatively spliced exons are predominantly regulated by tissue-specific *trans*-acting factors, a subset of *cis*-regulated exons remains consistently spliced across tissues with shared genetic backgrounds. These *cis*-directed regions show low conservation across vertebrates and are enriched in immune-related loci. Furthermore, isoLASER enables cohort-level analysis, revealing splicing associated variants (SAVs). Notably, these SAVs exhibit significant alterations in RNA-protein binding and demonstrate high reproducibility in splicing quantitative trait loci (sQTLs) studies. Overall, we provide a framework for identifying genetically and non-genetically driven splicing and the associated variants in cohorts with limited sample sizes.

5. Migration feedback yields novel critical transitions and emergent ecotypes in connected populations

Casey O. Barkan¹, Shenshen Wang¹

¹ Department of Physics and Astronomy, University of California Los Angeles, Los Angeles, USA

Biological populations live and evolve in spatially extended and heterogeneous environments. From gut microbiota to antibiotic resistant bacteria, spatial heterogeneity of selection pressure can profoundly affect evolution. Yet, the effects of heterogeneity depend upon the migration patterns by which organisms explore their environment. We present a simple and general model of evolution on a network of interconnected habitats, showing that migration feedback generates non-local

niches to which *emergent ecotypes* specialize. Varying migration rates induces continuous and discontinuous phase transitions at which ecotypes exchange stability. These transitions result in ecotype extinction or emergence, which may have significant health implications. We find that the discontinuous transitions arise via a novel mechanism of simultaneous transcritical bifurcations. Interestingly, these transitions show a 'fine structure', analogous to that of atomic spectra, when a symmetry of the competitive interactions is broken. The ecological nature of the transitions implies that feasible experiments could explore the predicted behaviors.

6. Microenvironment-driven plasticity of tumor cellular states and lipid metabolism in gliomas

Nicholas A. Bayley^{1,2}, Christopher Tse¹, Henan Zhu¹, Jenna Minami¹, Weihong Yan³, Cassidy L. Andrasz^{1,4}, Aparna Bhaduri⁵, Thomas G. Graeber^{1,6*}, and David A. Nathanson^{1,6*}

¹Department of Molecular and Medical Pharmacology,

²Bioinformatics Interdepartmental Program,

³Department of Chemistry and Biochemistry,

⁴Department of Human Genetics,

⁵Department of Biological Chemistry,

⁶Jonsson Comprehensive Cancer Center, University of California Los Angeles, Los Angeles California, USA
* equal contribution from labs

Malignant gliomas are characterized by intratumoral heterogeneity and plasticity, with tumor cells transiently adopting gene expression profiles of different neurodevelopmental cell types. However, the combined cell-intrinsic and -extrinsic cues within the brain tumor microenvironment (TME) regulating these cellular states and their functional ramifications remain largely unknown. Here we performed bulk and single cell transcriptomic profiling of a clinically diverse set of patient gliomas and matched derivative models established in orthotopic mouse xenografts and gliomasphere cultures to study the impact of microenvironmental context on tumor heterogeneity. Comparative analyses of tumors grown in different TMEs revealed transcriptional signatures activated by specific microenvironmental factors, including an immune response lost in model TMEs and lipid metabolic changes necessary for growth *in vitro*. Single-cell profiling combined with bulk tumor deconvolution revealed divergent changes in tumor cellular state composition depending on model TME, with cellular state enrichment connected to distinct lipidomic profiles and model establishment phenotypes.

7. Reconstructing single-cell genome structures from multi-omics and imaging data

Lorenzo Boninsegna¹, Francesco Musella¹, Asli Yildirim¹, Frank Alber¹

¹ Institute for Quantitative and Computational Biosciences and Department of Microbiology, Immunology and Molecular Genetics,

University of California Los Angeles, Los Angeles CA 90095, USA
Data from state-of-the-art sequencing- and imaging-based technologies have revolutionized our view of the 4D genome and bolstered computational studies; in fact, experimental data are now widely used in simulations to uncover the mechanistic underpinnings of biological processes that regulate genome organization. We recently developed the Integrated Genome Modeling (IGM) platform to simulate highly predictive genome structures, by leveraging and complementing the information

content from orthogonal experiments (e.g., Hi-C, lamina DamID, SPRITE and 3D HiMap FISH data). Here, we discuss how the modeling process can be further improved by adding data from next-generation single-cell imaging experiments, including realistic shapes of nuclear bodies and spatial locations of selected chromatin regions. We simulated populations of single cell structures for selected cell lines from bulk Hi-C and imaging data (e.g. live-cell, DNA MERFISH, DNA seqFISH+). Simulated structures are validated with independent data and allow to predict single cell features relating to both local and global chromatin folding, on a gene-by-gene basis, which structurally characterize different types of functional chromatin. Our population-based approach can uniquely combine ensemble and single cell, imaging and multi-omics experiments to simulate the most realistic genome structures and thus unveil the functional role of chromatin organization.

8. Using the protein-protein interaction network to make drug synergy predictions

Emily R. Bozich¹, Jennifer L. Wilson¹

Department of Bioengineering,

University of California Los Angeles, Los Angeles California, USA
Complex diseases are caused by dysregulation of multiple molecular pathways and therefore can benefit from combination therapies that enhance treatment efficacy (i.e., synergy). However, due to the combinatorial explosion of potential drug combinations, it is difficult to identify synergistic combinations experimentally. Thus, we turned to network approaches applied to the protein interaction network to model the relationships between drug targets and dysregulated disease pathways in order to prioritize effective combinations. We modeled the cascading effects of target perturbations by defining compact and functionally coherent target neighborhoods. We then measured the topological relationships between a drug combination's target neighborhoods using various distance metrics. In correlating these distances to experimental gene interaction scores, we found that global diffusion-based distances are more variable with better experimental alignment compared to shortest paths distances. Overall, our computational pipeline can be a viable method for prioritizing synergistic combinations or more effectively designing resource-deprived experimental screens.

9. scGRNdb: A Cell Type Gene Regulatory Network Atlas for Human and Mouse

Michael Cheng^{1,2}, Julie Tran¹, Rainy Liu¹, Russell Littman^{1,2}, Xia Yang^{1,2,3,4,5}

¹Department of Integrative Biology & Physiology, UCLA, Los Angeles, CA, USA

²Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA, USA

³Institute for Quantitative and Computational Biosciences (QCBio), Los Angeles, CA, USA

⁴Molecular Biology Institute (MBI), Los Angeles, CA, USA

⁵Brain Research Institute (BRI), Los Angeles, CA, USA

Gene regulatory networks (GRNs) elucidate the complex regulatory landscape in cells and tissues, making them powerful tools for understanding mechanisms in disease pathophysiology and identifying therapeutic targets. The advent of single cell RNA-sequencing (scRNAseq) enables a more granular study of disease mechanisms using cell type-specific GRNs, but most existing GRN methods are not optimized for scRNAseq and robust network resources are scarce. We recently developed

SCING, which improves GRN performance on scRNAseq and spatial transcriptomics data compared to existing methods. Here we present scGRNdb: a GRN atlas of 1,000+ SCING GRNs for cell types across 12 human and mouse single cell data atlases. Functional annotation of these networks revealed subnetworks that recapitulate known cell type specific pathways and gene mechanisms for neurological and cardiovascular diseases. Furthermore, we will host scGRNdb and GRN analysis tools on a public web server to facilitate single cell research and biomedical discovery.

10. Single-Cell Analysis in Lung Adenocarcinoma Implicates RNA Editing in Cancer Innate Immunity and Patient Prognosis

Jack Dodson^{1,2,3}, Tracey Chan^{1,2,3}, Jaron Arbet^{2,4,5}, Paul Boutros^{1,2,4,5,6,7,8}, Xinshu Xiao^{1,2,3,6,7}

¹Bioinformatics Interdepartmental Program, University of California, Los Angeles, California.

²Jonsson Comprehensive Cancer Center, University of California, Los Angeles, Los Angeles, California.

³Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles, California.

⁴Department of Urology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California.

⁵Department of Human Genetics, University of California, Los Angeles, Los Angeles, California.

⁶Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California.

⁷Institute for Quantitative and Computational Sciences, University of California, Los Angeles, Los Angeles, California.

⁸Institute for Precision Health, University of California, Los Angeles, Los Angeles, California.

RNA editing modifies single nucleotides of RNAs, regulating primary protein structure and protein abundance. Recently, the complexity of gene regulation associated with RNA editing dysregulation has been increasingly appreciated in oncology. RNA editing in single cells within tumors has not been explored. By profiling editing in single cells from lung adenocarcinoma, we found that increased editing of bulk lung tumors was unique to cancer cells. Elevated editing levels were observed in treatment-resistant cancer cells, and editing sites associated with drug response were enriched. The editing level in cancer cells was correlated with somatic mutation burden. This observation motivated the novel definition of RNA editing load, reflecting the amount of RNA mutations created by editing. In lung cancer, RNA editing load was a stronger predictor of patient survival than TMB. This study provides the first single cell dissection of editing in cancer and highlights the importance of RNA editing in cancer.

11. Systematic comparison of single-cell RNA-seq methods

Alejandro Espinoza^{1,2,3}, Bruno Andrade^{5*}, Robert Modlin^{4,5}, Matteo Pellegrini^{2,3}

¹Department of Human Genetics,

²Institute for Quantitative and Computational Biosciences-The Collaboratory

³Department of Molecular Cell and Developmental Biology

⁴Division of Dermatology, Department of Medicine, David Geffen School of Medicine

⁵Department of Microbiology, Immunology and Molecular Genetics

University of California Los Angeles, Los Angeles California, USA
Recent single-cell RNA-sequencing methods have become more accessible at a lower cost and implementation. However,

specifically, Parse Biosciences' split-pool ligation sequencing has not been systematically compared. We compare two methods for single-cell sequencing in peripheral blood mononuclear cells (PBMCs) and PBMC-derived effector memory cells re-expressing CD45RA (TEMRA), generating 12 libraries. We evaluated sequenced samples by comparing performance in read alignments, multiplets produced, sensitivity, and transcriptomic information retrieved or biological information. Furthermore, we use our sequenced TEMRA cells to understand the effect of IL21 induction.

12. A Biobank-scale test of marginal epistasis reveals genome-wide signals of polygenic epistasis

Boyang Fu^{1*}, Ali Pazokitoroudi^{1*}, Albert Xue², Aakarsh Anand¹, Prateek Anand¹, Noah Zaitlen^{3,4}, Sriram Sankararaman^{1,4,5}

¹Department of Computer Science,

²Bioinformatics Interdepartmental Program,

³Department of Neurology

⁴Department of Computational Medicine, David Geffen School of Medicine

⁵Department of Human Genetics, David Geffen School of Medicine

University of California Los Angeles, Los Angeles California, USA
* These authors contributed equally.

The contribution of epistasis (interactions among genes or genetic variants) to human complex trait variation remains poorly understood. Methods that aim to explicitly identify pairs of genetic variants, usually single nucleotide polymorphisms (SNPs), associated with a trait suffer from low power due to the large number of hypotheses tested while also having to deal with the computational problem of searching over a potentially large number of candidate pairs. An alternate approach involves testing whether a single SNP modulates variation in a trait against a polygenic background. While overcoming the limitation of low power, such tests of polygenic or marginal epistasis (ME) are infeasible on Biobank-scale data where hundreds of thousands of individuals are genotyped over millions of SNPs. We present a method to test for ME of a SNP on a trait that is applicable to biobank-scale data. We performed extensive simulations to show that our method provides calibrated tests of ME. We applied our method to test for ME at SNPs that are associated with 53 quantitative traits across ~300 K unrelated white British individuals in the UK Biobank. Testing 15,601 trait-loci associations, we identified 16 trait-loci pairs across 12 traits that demonstrate strong evidence of ME signals (p-value $p < 5e-8/53$). We further partitioned the significant ME signals across the genome to identify 6 trait-loci pairs with evidence of local (within-chromosome) ME while 15 show evidence of distal (cross-chromosome) ME. Across the 16 trait-loci pairs, we document that the proportion of trait variance explained by ME is about 12x as large as that explained by the GWAS effects on average (range: 0.59 to 43.89). Our results provide evidence for epistatic interactions modulating the effects of genetic variants on complex traits.

13. Analyzing Exotic Density Distributions in Single-Cell Data

Timothy Hamilton^{1,2*}, Breanne Sparta^{2,3*}, Serena Hughes^{1,2}, Eric J. Deeds^{1,2,3},

¹Bioinformatics Interdepartmental Program, University of California, Los Angeles

²Institute of Quantitative and Computational Biology, University of California, Los Angeles

³ Department of Integrative Biology and Physiology, University of California, Los Angeles

* These authors contributed equally.

The proliferation of single cell sequencing has promised a revolution through multiple fields of biology. Methods to analyze this data are influenced by the “Waddington’s landscape” paradigm of the arrangement of cell types in gene expression space, which would produce well separated groups of cells that are arranged with a well understood density. Instead, our method has revealed that most cells are found in low density regions of the space with a few cells being found in areas that are 3-4 orders of magnitude denser than the lower density regions. This finding is inconsistent with the current interpretations and applications of Waddington’s Landscape commonly used in single cell analysis. Nevertheless, further work has allowed us to partially characterize the structure of single cell data. Ultimately, our findings have implications for how development must be conceptualized and reveal unseen diversity in how single cells behave that provides new opportunities for study.

14. Enrichment of hard sweeps on the X chromosome across six *Drosophila* species

Mariana Harris¹, Bernard Kim², and Nandita Garud^{3,4}

¹ Department of Computational Medicine Biomathematics Program, University of California Los Angeles, Los Angeles California, USA

²Department of Biology, Stanford University, Stanford, California, USA

³ Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles California, USA

⁴ Department of Human Genetics, University of California, Los Angeles, California, USA

The X chromosome is hemizygous in males, leaving it fully exposed one third of the time to the effects of natural selection and, thus, potentially subject to different evolutionary dynamics than autosomes. This is of particular interest given the importance of sex chromosomes in local adaptation, speciation, and sexual dimorphism. In our recent work, we found an enrichment of hard sweeps, expected when adaptation is gradual, on the X chromosome relative to the autosomes in a North American population of *D. melanogaster*. Now, we generalize these findings by analyzing diversity patterns across six *Drosophila* species, where we find consistently steeper reductions in diversity along with elevated haplotype homozygosity on the X chromosome compared to autosomes. To assess if these signatures are consistent with positive selection, we simulate a wide variety of evolutionary scenarios and find that the patterns observed on the X are most consistent with hard sweeps. Our findings highlight the importance of sex chromosomes in driving evolutionary processes and suggest that hard sweeps have played a significant role in shaping diversity patterns on the X chromosome across multiple *Drosophila* species.

15. Splicing-specific transcriptome-wide association uncovers novel genetic mechanisms for Schizophrenia

Jonatan L. Hervoso¹, Kofi Amoah¹, Jack Dodson¹, Mudra Choudhury¹, Arjun Bhattacharya², Giovanni Quinones-Valdez³, Bogdan Pasaniuc^{1,2,4,5*}, Xinshu Xiao^{1,3*}

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

²Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, California, USA

³Department of Integrative Biology and Physiology, University of California, Los Angeles, USA

⁴Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

⁵Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, USA
Recent studies have highlighted the essential role of RNA splicing in establishing connections between genetic variations and disease. Dysregulated RNA splicing has emerged as a major potential contributor to neurological and psychiatric disorders. Nevertheless, the direct association of genetically altered splicing with complex traits remains challenging. We introduce Spliced-Transcriptome-Wide Associations (SpliTAS), which integrates alternative splicing GWAS to identify genes associated with a trait through specific exon splicing events. We applied SpliTAS to two different schizophrenia (SCZ) RNA-seq datasets, identifying 137(84) trait-associated exons, and 88(67) exons (genes), respectively. SpliTAS variants impacted RNA-binding protein (RBP) binding sites, revealing potential disruption of RNA-protein interactions affecting splicing. Enriched biological functions in the associated gene sets converged on neuronal function and development, immune cell activation, cellular transport, pathways highly associated and related to SCZ. Overall, this study suggests that alternative splicing plays a significant role in the genetic basis of SCZ, providing valuable insights beyond conventional gene/exon expression analysis.

16. Calibrated prediction intervals for polygenic scores across diverse contexts

Kangcheng Hou¹, Ziqi Xu², Yi Ding¹, Arbel Harpak^{3,4}, Bogdan Pasaniuc^{1,5,6,7}

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

² Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA

³ Department of Population Health, The University of Texas at Austin, Austin, TX, USA

⁴ Department of Integrative Biology, The University of Texas at Austin, Austin, TX, USA

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

⁶ Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

⁷ Institute for Precision Health, University of California, Los Angeles, Los Angeles

Polygenic scores (PGS) have emerged as the tool of choice for genomic prediction in a wide range of fields. We analyze data from two large biobanks in the US and the UK to find widespread variability in PGS performance across contexts. Many contexts, including age, sex, and income, impact PGS accuracies with similar magnitudes as genetic ancestry. We introduce trait prediction intervals as a principled approach to account for context-specific PGS. Our approach enables PGS-based trait predictions that are well-calibrated. We show that prediction intervals need to be adjusted for all considered traits. Adjustment of prediction intervals are dataset- and trait-specific;

for example, prediction intervals for education years need to be adjusted by 90% in All of Us versus 8% in UK Biobank. Our results provide a path forward towards using PGS as a prediction tool across all individuals regardless of their contexts and highlight the importance of comprehensive profile of context.

17. Identifying epigenetic aging moderators with the Epigenetic Pacemaker

Chanyue Hu, Colin Farrell, Kalsuda Lapborisuth, Kyle Pu, Sagi Snir, and Matteo Pellegrini
Dept. of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90095, USA,
Dept. of Evolutionary Biology, University of Haifa, Israel
Epigenetic clocks are commonly employed to study age related biology. The difference between the predicted and observed age is often interpreted as a form of biological age acceleration. Here we compare penalized regression methods used to construct epigenetic clocks to an evolutionary framework of epigenetic aging, the epigenetic pacemaker (EPM) that directly models DNA methylation as a function of a time dependent epigenetic state. In simulations we show that the value of the epigenetic state is impacted by factors such as age, sex and cell type composition. In a dataset aggregated from previous studies, we show that the epigenetic state is also moderated by sex and cell type. We also found the epigenetic state is also moderated by toxins in a study of polybrominated biphenyl exposure. Thus, we find that the pacemaker provides a robust framework for the study of factors that impact epigenetic age acceleration.

18. Cell-free DNA methylation as a noninvasive prognostic indicator in hepatocellular carcinoma

Ran Hu^{1,2,3}, Benjamin Tran⁴, Mary L. Stackpole¹, Shuo Li¹, Xianghong Jasmine Zhou^{1,3,5}, Wenyan Li¹ & Vatche Agopian^{4,5}
¹Department of Pathology and Laboratory Medicine,
²Bioinformatics Interdepartmental Graduate Program,
³Institute for Quantitative & Computational Biosciences,
⁴Department of Surgery,
⁵Jonsson Comprehensive Cancer Center,
University of California Los Angeles, Los Angeles California, USA
The current noninvasive prognosis evaluation of hepatocellular carcinoma (HCC) using markers such as imaging features and serum biomarkers like alpha-fetoprotein is limited due to low sensitivity. Epigenetic changes in cell-free DNA (cfDNA) have shown promise in cancer diagnosis and prognosis. We aim to evaluate the potential of cfDNA methylation as a noninvasive predictor for pre-treatment prognostication in HCC. Using Illumina HumanMethylation450 array data of 377 HCC tumor and 50 adjacent normal tissues from TCGA, we identified 166 HCC-related DNA methylation markers associated with overall survival (OS). We validated the signature via the random survival forest model with 10-fold cross-validation in a Tissue Cohort of 30 patients and a Plasma Cohort of 52 patients. The cfDNA methylation-based OS risk score showed strong discriminatory power when evaluated as a single predictor and in combination with other clinical variables. Our study demonstrates that cfDNA methylation is a promising noninvasive prognostic predictor for HCC.

19. T-dependent B-cell selection and proliferation: a systems analysis of signal I and signal II integration

Helen Huang^{1,2,3}, HariPriya Vaidehi Narayanan^{2,3}, Alexander Hoffmann^{2,3}
¹ Bioinformatics Interdepartmental Program,

² Institute for Quantitative and Computational Biosciences (QCB),

³ Department of Microbiology, Immunology, and Molecular Genetics (MIMG),
University of California Los Angeles, Los Angeles, USA.

In response to vaccination or infection, a successful antibody response must enrich high-affinity antigen-reactive B-cells through positive selection, but eliminate autoreactive B-cells by negative selection. Reactive B-cells undergo proliferative bursts that are governed by signaling from the B-cell receptor (BCR) which binds the antigen and the CD40 signal provided by neighboring T-cells that recognize the antigen via its own MHC T cell receptor complex, when it is presented with the antigen. Both signals are a function of the BCR's affinity to the antigen. Little is known about the mechanism by which BCR and CD40 signaling are integrated, and thus jointly determine B-cell selection and proliferation. We quantitatively evaluated the population dynamics after stimulating B-cells in vitro through their BCR and CD40 receptors. We interpreted the data with a newly developed mathematical model of the BCR and CD40 signaling pathways and their control of B-cell fate decision machineries. Our results show that while CD40 and BCR costimulation induces more NFkB activity, no such potentiation is seen at the level of population expansion. Model simulations reveal that functional antagonism may be mediated by BCR-induced caspase activity triggering apoptosis in founder cells. We investigated in silico and in vitro the temporal relationship between these antagonistic signals and found that within a limited time window CD40 signaling may effectively rescue cell death triggered by BCR signaling. The window size depends on the strength of the BCR and CD40 signals, but a longer time gap does not allow for B-cell population expansion.

20. A Statistical Marker Gene Clustering Tool

Serena J. Hughes¹⁻², Timothy Hamilton¹⁻², Eric J. Deeds¹⁻³
¹ Institute for Quantitative and Computational Biosciences,
² Bioinformatics Interdepartmental Program,
³ Department of Integrative Biology and Physiology,
University of California Los Angeles, Los Angeles California, USA

The current standard practice for clustering analysis on single-cell RNA sequencing (scRNA-Seq) data often involves manual assignment of cell types to clusters based on marker gene expression, which can introduce bias and issues with reproducibility. To help mitigate this, we are developing a Statistical Marker Gene Clustering (SMGC) tool. This tool uniformly samples clustering parameters and quantitatively evaluates the output based on expected marker gene associations. In our exploration of data where the cell types are known, we observed that the pattern of marker gene expression lends itself to translation into barcodes that represent each cell type. The SMGC tool uses this concept of barcodes to compare clusters and individual cells to their assigned cell types. The SMGC tool contributes to the emerging set of algorithms for the clustering process that are shifting the field towards more rigorous and consistent analysis of scRNA-Seq data.

21. Stimulus-Response signaling dynamics characterize macrophage polarization states

Michael Iter^{1,2,3}, Apeksha Singh^{1,2,3}, Supriya Sen^{1,2,3}, Alexander Hoffmann^{1,2,3}

¹ Signaling Systems Laboratory,

² Department of Microbiology, Immunology, and Molecular Genetics,

³ Institute for Quantitative and Computational Biosciences, University of California Los Angeles, Los Angeles California, USA
Macrophages show remarkable functional pleiotropy that is further modulated by polarizing cytokines in their microenvironment. Given the importance of the transcription factor NFκB to macrophage responses, we hypothesized that its activation dynamics could characterize polarization states. Here, primary macrophages from a fluorescent reporter mouse were polarized into 6 states and single-cell trajectories of NFκB activation dynamics were collected following stimulation with different immune threats. Machine learning analyses identified that the TLR1/2 agonist, Pam3CSK, provided the greatest distinction between polarization states, and trajectory features informative for distinguishability were used to map the cell states. Next, utilizing an existing mechanistic model of NFκB signaling, we obtained single-cell parameter fits by minimizing the RMSD between the experimental trajectories and model simulations using a gradient-free optimization process with many initializations. Ultimately, we leveraged the fit parameter distributions to generate an alternative characterization of the single-cell states which recapitulates aspects of the feature-based map.

22. The theoretical ceiling of local ancestry inference

Moonseong Jeong¹, Sriram Sankararaman^{*1,3,4}, April Wei^{*2}

¹UCLA, Department of Computer Science, Los Angeles, CA, ²Cornell University, Department of Computational Biology, Ithaca, NY, ³UCLA, Department of Human Genetics, Los Angeles, CA, ⁴UCLA, Department of Computational Medicine, Los Angeles, CA

^{*}Contributed equally

Many statistical and machine learning inference methods have been developed to infer the local ancestry of admixed genomes from various demography. Despite their success, challenges remain for scenarios such as South Asians, as each method makes different assumptions on the population parameters and how populations interact. Thus, it is crucial to understand in which parameter regime a confident inference is possible. Here we present a coalescent framework with a single pulse of admixture to estimate the probability of correct local ancestry inference for a given demography. Through our model, we demonstrate how assumptions on the demography, including admix proportion, divergence/admixture times, and effective population sizes, impact the trustworthiness of inference based on references. In the Neanderthal introgression, we find that the effective population size has an outsized influence on the inference, whereas the admix proportion comparatively less so. Our method would be useful for determining whether a correct inference is possible for a given scenario.

23. Polygenic scores for major depressive disorder provide insights into medication use prediction in EHR-linked biobank.

Sandra Lapinska¹, Vidhya Venkateswaran^{2,5}, Kristin Boulter^{1,3}, Yi Ding¹, Aditya Pimplaskar¹, Loes M Olde Loohuis⁴, Bogdan Pasaniuc^{1,4,5}

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

²Department of Oral Biology, School of Dentistry, University of California, Los Angeles, Los Angeles, CA 90095, USA.,

³Department of Medicine, Division of Cardiology, University of California, Los Angeles, Los Angeles, CA 90095, USA.,

⁴Department of Human Genetics,

⁵Department of Pathology and Laboratory Medicine

Major depressive disorder (MDD) is the most common psychiatric illness. Studies have shown that MDD has a complex etiology with possible population differences in predisposing genetic factors. Investigating the relationship between ancestry and MDD could provide insight into how individuals may respond to certain treatments. We use a publicly available polygenic risk score (PGS) trained on European individuals from the UK Biobank data for major depressive disorder to investigate the clinical utility of PGS in medication use prediction within the diverse UCLA ATLAS biobank (N > 40,000). Our study suggests that individuals with higher genetic risk for major depressive disorder increase treatment resistance. This demonstrates a potential use of PGS to determine effective treatments for individuals diagnosed with major depressive disorder across ancestries.

24. Adaptation across host microbiomes using haplotype homozygosity

Peter Laurin^{1*}, Nandita Garud^{1,2}

¹Department of Ecology and Evolutionary Biology,

²Department of Human Genetics

University of California Los Angeles, Los Angeles California, USA
Adaptation is widespread and pervasive in bacteria and underlies the evolution of important and well-studied traits, with antibiotic resistance as one example. In the human gut microbiome, adaptation has thus far been uncharacterized due to the difficulty in distinguishing allele frequency changes in individual strains. However, adaptations may be beneficial in multiple hosts, spreading via migration rather than evolving multiple times *de novo*. Such scenarios leave behind signatures of elevated haplotype homozygosity in the vicinity of the selected site in across-host populations of bacteria. Here we examine haplotype homozygosity in the top ~40 gut commensal bacteria in the human gut and find evidence that positive selection has been a common force shaping diversity across hosts. Moreover, we find evidence that multiple origins of the same adaptive allele may have swept to high frequency in the broader population, indicative of the potential rapidity of adaptation and transmission in the gut microbiome.

25. Systems serology profiling of anti-tumor antibodies in high-grade serous ovarian cancer

Michelle Loui¹, Scott D. Taylor¹, Het Desai¹, Allison Brookhart¹, Meera Trisal¹, Gabriella DiBernardo², Sanaz Memarzadeh², Aaron Meyer¹

¹Department of Bioengineering,

²Department of Obstetrics and Gynecology, David Geffen School of Medicine,

University of California, Los Angeles, CA USA

In high-grade serous ovarian cancer (HGSOC) patients, malignant epithelial cells arise from the fallopian tube and ovarian surface epithelium. Endogenous antibodies (anti-tumor antibodies; ATAbs) target these cells and 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. ATAbs 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 tumor cells. We hypothesize that ATAbs are unable to eliminate tumors due to a dysregulation of their Fc region. Therefore, we applied a quantitative, multiplexed assay for profiling the Fc properties and immune receptor interactions of ATAbs. Understanding these mechanisms of tumor immune evasion will help uncover how immunotherapies might reactivate effective humoral immunity.

26. Cross-species and tissue imputation of species-level DNA methylation samples.

Emily Maciejewski^{1,2}, Steve Horvath³, Jason Ernst^{1,2}

¹ Department of Computer Science, University of California, Los Angeles

² Department of Biological Chemistry, University of California, Los Angeles

³ Altos Labs, San Diego, CA, USA

DNA methylation data is highly informative to study a variety of aspects of mammalian biology. The availability of such data for many mammals at conserved sites was recently vastly enhanced by the development and large-scale application of the mammalian methylation array. For instance, we consider here 13,245 samples profiled on this array representing 348 species and 59 tissues from 746 species-tissue combinations. While having some coverage of many different species and tissue types, this data only captures 3.6% of potential species-tissue combinations. We thus developed CMImpute (Cross-species Methylation Imputation) which uses a Conditional Variational Autoencoder to impute DNA methylation of non-profiled species-tissue combinations. In cross-validation, we show that CMImpute yields high correlation with held-out observed values, outperforming multiple baselines. We then train a model on all the data to impute 19,786 new species-tissue combinations. We expect CMImpute and our imputed data resource will be useful for DNA methylation analyses across mammalian species.

27. The impact of selection on synonymous mutations when inferring the distribution of fitness effects on nonsynonymous mutations

Aina Martinez i Zurita^{1,2}, Christopher C. Kyriazis², Kirk Lohmueller²

¹ Genetics and Genomics PhD Program,

² Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles California, USA
The DFE (Distribution of Fitness Effects) describes the distribution of selection coefficients of new mutations. Accurate measurements of the DFE are both important as a fundamental parameter in evolutionary genetics as well as because of its implications for our understanding of complex disease or inbreeding depression. Current computational methods to predict the DFE for nonsynonymous mutations from natural variation rely on the assumption that synonymous variants can be considered neutrally evolving. Synonymous variants are typically used to infer the effect of demography and background selection affecting the genome. However, some recent experimental evidence points toward synonymous mutations having measurable effects on fitness. To test whether selection on synonymous mutations affects inference of the DFE of nonsynonymous mutations, we simulated several possible synonyms selection models using SLiM and attempted to recover the DFE of nonsynonymous mutations using FitDaDi, a common method for demographic

and DFE inference. In our preliminary results, correct inference of demographic parameters is skewed by the presence of selection on synonymous sites. However, the effects on the DFE are more varied. Our work has implications for understanding how violations of population genetic models can affect downstream inferences.

28. Prediction in Electronic Health Records: Portability for 8 million patients across 5 health systems.

Rachel Mester¹, Alexis Stephens², Aaron Chin², Ruth Johnson³, Bogdan Pasaniuc^{1,4,6,7}, Manish Butte^{2,4}

¹Biomathematics Graduate Program, Department of Computational Medicine, UCLA

²Department of Pediatrics, Division of Immunology, Allergy and Rheumatology, UCLA

³Department of Biomedical Informatics, Harvard Medical School

⁴Department of Human Genetics, UCLA

⁵Department of Microbiology, Immunology, and Molecular Genetics, UCLA

⁶Department of Pathology and Laboratory Medicine, UCLA

⁷Department of Computational Medicine, UCLA

Over the past several years, the volume of research in electronic health record (EHR) data has exploded. However, large scale studies of best practices for combining EHR datasets to improve sample size and understand the portability of results is still very much an open question. We utilize the UC Health Data Warehouse, a combined dataset from five major University of California health systems which includes over eight million patients, to understand portability and reproducibility for disease prediction algorithms in the EHR. We conduct an evaluation of methods to combine datasets from these five health systems and assess the impact of sample size, choice of algorithm, disease heterogeneity, and phenotype uncertainty on the accuracy and portability of predictive algorithms. We find that this EHR dataset does not contain the type of structure that inhibits combining datasets, but that sample size differences and phenotype uncertainty do have an impact on portability.

29. 3D Whole-Genome Structure Modeling in Mouse Cerebellum Tissue by the Integrative Genome Modeling Platform

Francesco Musella^{3,4,5}, Ziyu Zhao^{1,2,#}, Omar Payán Parra^{1,2,#},

Nicolas Scrutton-Alvarado^{1,2}, Shinichiro Fujita¹, Frank Alber^{3,4,5}, Yue Yang¹, Tomoko Yamada¹

¹Department of Neurobiology, Northwestern University, Evanston, IL, USA

²Program in Interdisciplinary Biological Sciences, Northwestern University, Evanston, IL, USA

³Institute for Quantitative and Computational Biological Sciences and ⁴Department of Microbiology, Immunology and Molecular Genetics and ⁵Bioinformatics IDP Program, University of California, Los Angeles, Los Angeles, CA, USA

[#]These authors contributed equally

During mammalian brain development, chromatin structure rearranges to coordinate gene expression, but the understanding of DNA organization and its role in development is limited. We investigated 3D genome changes during postnatal brain development in mouse cerebellum at several times. Using Integrative Genome Modeling, we deconvoluted ensemble HiC data into single cell 3D genome structures. Analysis revealed an increase in long-range interactions between active chromatin during development, uncovering a new active compartment associated with enhancer-dense regions and long neuronal

genes. This compartment's nuclear topography is different from two other active compartments associated with speckles and gene-dense regions. Particularly, the enhancer-dense compartment is closer to the nuclear envelope and far from speckles, confirmed by FISH experiments. Our findings offer deeper insights into chromatin structure in brain development.

30. Characterizing Prostate Cancer Radiation Resistance and Local Recurrence

Beth K. Neilsen^{1,2,3}, Nicole Zeltser^{2,3,4}, Raag Agrawal^{2,3,4}, Selina Wu^{2,3,5}, Yash Patel^{2,3,4}, Takafumi N. Yamaguchi^{2,3,4}, Mohammed Faizal Eeman Mootor^{2,3,4}, Sorel T. Fitz-Gibbon^{2,3,4}, Rong Rong Huang⁶, Jian Yu Rao⁶, Anthony E. Sisk⁶, Cyriac Kandoth⁶, Liying Zhang⁶, Wolfgang Lilleby⁸, Paul C. Boutros^{2,3,4,7}, Amar U. Kishan^{1,2,3}

¹ Department of Radiation Oncology,

² Jonsson Comprehensive Cancer Center,

³ Institute of Precision Health,

⁴ Department of Human Genetics,

⁵ Department of Medical Informatics,

⁶ Department of Pathology and Laboratory Medicine,

⁷ Department of Urology,

University of California Los Angeles, Los Angeles California, USA

⁸ Department of Urologic Oncology,

Oslo University Hospital, Oslo, Norway

A large majority of men with prostate cancer are diagnosed with localized disease making them candidates for definitive treatments such as radical prostatectomy or radiation therapy. Prostate cancer recurs after definitive treatment in about one third of patients without reliable predictors for recurrence. Characterization of four patient-paired *de novo* and post-radiation, locally recurrent prostate cancer FFPE samples was completed using a targeted pan-cancer DNA sequencing panel and RNA sequencing. Samples were evaluated for treatment-induced changes (clonal selection *versus* clonal evolution) and compared to large, publicly available databases of *de novo* prostate cancer to identify targets and/or pathways enriched in locally recurrent prostate cancer. Evaluation of an expanded cohort is being undertaken to further characterize locally recurrent prostate cancer. These findings may be applied to generate prognostic and predictive biomarkers, provide for earlier detection of prostate cancer recurrence and supply candidate targets for the development of novel therapies and treatment regimens.

31. Characterizing the Genomic and Evolutionary Landscape of Oligometastatic Prostate Cancer

Jieun Oh^{1,2,3,4,5}, Tsumugi Gebo^{1,2,3,4}, Yash Patel^{1,2,3,4}, Sarah Al-Hiyari^{1,2,3,4}, Urooza Sarma, Paul Boutros^{1,2,3,4}, GAP6 Consortium

¹ Department of Human Genetics, UCLA, CA, USA

² Department of Urology, UCLA CA, USA

³ Jonsson Comprehensive Cancer Center, UCLA, CA, USA

⁴ Institute for Precision Health, UCLA, CA, USA

⁵ Bioinformatics Interdepartmental Program, UCLA, USA

⁶ Department of Anatomy and Developmental Biology, Development and Stem Cells Program, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia
While localized prostate cancer with low to intermediate risk has favorable prognosis, metastatic prostate cancer remains incurable. Oligometastatic prostate cancer (OMPC) is a controversial disease state characterized by limited metastatic spread. It represents an intermediate stage of metastasis that could potentially be cured before progressing to incurable

metastatic castration-resistant prostate cancer. Emerging evidence suggests a distinct biology in oligometastasis and several studies highlight the benefits of radiotherapy for prostate cancer with low metastatic burden. However, OMPC lacks genomic and biological definitions, with a clinical definition of less than three to five metastatic lesion. This definition is subject to change as imaging techniques become more sensitive. To address this, we propose the first comprehensive genomic and evolutionary characterization of OMPC, involving whole genome sequenced multi-region primary prostate regions and treatment-naive oligometastatic bone metastases. This study seeks to refine the definition of OMPC, differentiate it from localized and metastatic prostate cancer and identify the “molecular hallmarks of curability”.

32. VISTA: An integrated framework for structural variant discovery

Varuni Sarwal^{1*}, Seungmo Lee^{1*}, Jianzhi Yang², Sriram Sankararaman¹, Mark Chaisson², Eleazar Eskin^{1#}, Serghei Mangul^{2,3#}

¹ Department of Computer Science, UCLA

² Department of Quantitative and Computational Biology, USC

³ Department of Clinical Pharmacy, Alfred E. Mann School of Pharmacy, USC

With advances in whole genome sequencing (WGS) technologies, a plethora of Structural Variant (SV) detection methods have been developed. However, the majority of these methods are prone to a high false-positive rate, and unable to detect a full range of SV's present per sample. Here, we report an integrated SV calling framework, VISTA (Variant Identification and Structural Variant Analysis) that leverages the results of individual callers using a robust filtering and merging algorithm. VISTA executes various combinations of top-performing callers based on variant length and genomic coverage to generate highly accurate SV events. We benchmarked VISTA using the Genome-in-a-Bottle gold-standard SV set, haplotype-resolved *de novo* assemblies from the HPRC, and an in-house PCR-validated mouse set, where VISTA maintained the highest F1-score. VISTA represents a significant advancement in SV calling, offering a robust and accurate framework that outperforms existing tools and sets a new standard for SV detection in genomic research.

33. A mathematical model reveals the regulatory logic of interferon- β expression

Allison Schiffman^{1,2}, Zhang Cheng¹, Diana Ourthiague¹, Alexander Hoffmann¹

¹Department of Microbiology, Immunology, and Molecular Genetics, Institute for Quantitative and Computational Biology, and Molecular Biology Institute

²Molecular Biology Interdepartmental Doctoral Program University of California Los Angeles, Los Angeles California, USA
The expression of type I interferon (IFN β), a critical determinant of the innate immune response, is controlled by three key transcriptional activators, AP1, NF κ B and IRF, which bind to the enhancer region of IFN β . Classic molecular biology studies of IFN β expression suggest that NF κ B and IRF function synergistically by forming an enhanceosome complex. However, biochemical evidence shows that NF κ B and IRF do not have positive cooperativity in binding, and knockout mouse studies show that NF κ B is not always required for IFN β expression. We developed a quantitative model to account for both literature data and new measurements from NF κ B and IRF knockout cells.

We found that the Guanine-rich IRF binding site affinity may be tuned by competing repressor factors, which allows NFκB to boost IFNβ expression with low IRF.

34. PCR data accurately predict infectious virus: a meta-analysis of SARS-CoV-2 in non-human primates

Celine Snedden¹, James Lloyd-Smith^{1,2}

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

² Department of Computational Medicine, University of California Los Angeles, Los Angeles, California, USA

Researchers and clinicians often rely on molecular assays like PCR to identify and monitor viral infections instead of the resource-prohibitive gold standard of viral culture. However, it remains unclear when (if ever) PCR measurements of viral load are reliable indicators of replicating or infectious virus. Here, we compare total RNA, subgenomic RNA, and viral culture results from 24 studies of SARS-CoV-2 in non-human primates using bespoke statistical models. On out-of-sample data, our best models predict subgenomic RNA from total RNA with 91% accuracy, and they predict culture positivity with 85% accuracy. Total RNA and subgenomic RNA showed equivalent performance as predictors of culture positivity. Multiple cofactors, including exposure conditions and host traits, influence culture predictions for total RNA quantities spanning twelve orders of magnitude. Our model framework can be adapted to compare any assays, in any host species, and for any virus, to support laboratory analyses, medical decisions, and public health guidelines.

35. Liftover consequences: mapping artifacts and their impact on variant calling

Nicholas K. Wang^{1,2,3}, Paul C. Boutros^{2,3,4,5}

¹ Bioinformatics Interdepartmental Program, UCLA

² Department of Human Genetics, UCLA

³ Jonsson Comprehensive Cancer Centre, UCLA

⁴ Department of Urology, UCLA

⁵ Institute for Precision Health, UCLA

Released in 2013, the Genome Reference Consortium Human Build 38 (GRCh38) corrected thousands of small sequencing artifacts found in GRCh37 and expanded the use of alternate contigs. The UCSC LiftOver tool is used as part of standard practice to convert coordinates between genome builds. Despite widespread use, little is known about the reliability of such conversions and whether they introduce biases in variant calling that can affect downstream analyses. Here, we present a matched comparison of 50 whole-genome sequenced tumor-normal pairs aligned to both GRCh37 and GRCh38. We assessed variant concordance after LiftOver of single nucleotide variants and structural variation across both germline and tumor genomes and identified differences in sample-level metrics as well as biases among discordant variants. By characterizing the landscape of mapping artifacts between GRCh37 and GRCh38, we hope that our findings can guide future analyses and enable more robust comparisons across reference versions.

36. Benchmarking Brain Cell Type Label Transfer Methodologies on Code Ocean

Ning Wang^{1,2}, Changkyu Lee¹, Dong-Wook Kim¹, Zizhen Yao¹

¹ Allen Institute for Brain Science,

² Department of Integrative Biology and Physiology,

University of California, Los Angeles, Los Angeles, CA 90095, USA

The advent of next generation sequencing enables high-throughput, transcriptome-wide gene expression profiles at single cell resolution. Recently, the Allen Institute released the whole mouse brain taxonomy, with over 4 million cells and 5,200 cell types. Optimal usage of reference atlases requires benchmarking label transfer methodologies. Through reproducible pipelines on Code Ocean, we showed that our in-house mapping method was able to outperform the popular Seurat label transfer pipeline.

37. Unraveling the Role of 3D Genome Organization in Modulating DNA Damage Susceptibility and Gene Expression during Chemotherapy

Ye Wang^{1,2,3}, Asli Yildirim^{1,2}, Lorenzo Boninsegna^{1,2}, Charis Qi⁴, Ryan Cope⁴, Xianghong Jasmine Zhou^{1,3,*}, Frank Alber^{1,2,*}

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

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

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

⁴BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

Platinum-based chemotherapy drugs kill cancer cells by damaging DNA. Previous studies have found that the 3D genome structure of cells can influence the distribution of chemotherapy-induced DNA damage. However, the explicit role this regulation plays in drug resistance remains unexplained. Our research explores the relationship between alterations in 3D genome structure of cancer cells and their resistance to platinum-based chemotherapy. By comparing oxaliplatin-resistant and sensitive cells, we observed that changes in DNA damage levels across genomic regions correlate with their nuclear location. Genomic regions near the nuclear speckle displayed reduced damage levels in resistant cells, suggesting that 3D genome adaptations increase resistant cell survival. Additionally, we identified genes with high speckle association frequency that exhibited reduced damage and increased expression levels, implicating them in platinum-drug resistance. This study contributes to our understanding of the interplay between 3D genome structure and chemotherapy resistance, offering potential new avenues for cancer treatment strategies.

38. Microbiome diversity along the tract of the gut

Michael Wasney^{1,4,*}, Leah Briscoe^{2,5,*}, Ricky Wolff³, Hans Ghezzi⁶, Carolina Tropini^{6,7,8} and Nandita R. Garud^{1,2,3}

¹ Department of Human Genetics,

² Interdepartmental Program in Bioinformatics,

³ Department of Ecology and Evolutionary Biology,

⁴ Genetics and Genomics Program,

⁵ Bioinformatics Program,

University of California Los Angeles, Los Angeles California, USA

⁶ Department of Microbiology and Immunology,

⁷ School of Biomedical Engineering,

University of British Columbia, Vancouver British Columbia, Canada

⁸ Humans and the Microbiome Program, Canadian Institute for Advanced Research, Toronto Ontario, Canada

* These authors contributed equally.

The species and genetic diversity of the human gut microbiome has been extensively quantified from stool and associated with important host phenotypes. However, diversity measured from stool may not reflect the spatial heterogeneity of the microbiome along the tract of the gut. Here, we quantify spatial diversity along the gut at the species, strain, and gene level in the guts of eight mice. We find that species composition differs substantially between the upper and lower gut as well as between hosts on different diet regimes. Contrastingly, co-colonizing strains of the same species tend to coexist along the gut with little variation in strain abundance between gut regions, suggesting that little spatial heterogeneity occurs at the strain level. Finally, gene content was found to differ between the upper or lower gut. In sum, we show that host, tissue, and diet have different impacts on specific and subspecific levels of diversity, respectively.

39. EMulSI-Phy: Efficient Multi-Sample Inference of Cancer Phylogeny

Helena K. Winata^{1,2}, Dan Knight^{1,2}, Pier Selenica³, Nicholas K. Wang^{1,2}, Caroline Kostrzewa⁴, Stefan Eng^{1,2}, Yingjie Zhu³, Jaron Arbet^{1,2}, Juber A. Patel³, Ronglai Shen⁴, Jorge-Reis Filho³, Pedram Razafi⁵, Paul C. Boutros^{1,2}

¹Department of Human Genetics, University of California, Los Angeles, CA, USA

²Jonsson Comprehensive Cancer Centre, University of California, Los Angeles, CA, USA

³Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

⁴Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

⁵Department of Medicine, Memorial Sloan Kettering Cancer Center and Weill Cornell Medical College, New York, NY, USA

Inferring tumor evolution from DNA sequencing data is crucial for understanding cancer progression and mutation patterns within clones. Decreasing sequencing costs enable routine sequencing of multiple tumor samples per patient, offering an opportunity to explore tumor evolution with unprecedented resolution across spatial regions and longitudinal timeframes. However, growing dataset complexity presents a challenge, necessitating innovative methods to transcend the computational constraints faced by current methods. Hence, we developed EMulSI-Phy, a rule-based algorithm that leverages the fundamental principles of cancer biology to enable more rapid and accurate subclonal reconstruction. Benchmarking on 576 simulated datasets demonstrated its computational efficiency and accuracy in resolving complex evolutionary structures. Applied to 12 breast cancer patients with 4-36 sequenced tumor regions, EMulSI-Phy reduced average runtime ten-fold, delineating and reconstructing the phylogeny of up to 88 subclones per patient, which is unfeasible with most methods.

40. Pervasive selective sweeps in the human gut microbiome

Richard Wolff¹, Nandita Garud^{1,2}

1 : Department of Ecology and Evolutionary Biology

2 : Department of Human Genetics

The human gut microbiome is composed of hundreds of species that are finely adapted to live within human hosts. This adaptive process is continually ongoing, with new adaptive variants entering these populations daily. While the same variant may

arise in multiple hosts by recurrent *de novo* mutation, alleles may also spread across hosts through migration and horizontal gene transfer. Here, we develop a novel LD-based statistic (iLDS) to scan for positive selection across host microbiomes, and find that adaptive horizontal gene transfer is a pervasive force shaping patterns of genetic diversity across gut species. We find that genes related to the digestion of synthetic sugars and starches, which have increased dramatically in Western diets in recent years, are frequently and strongly selected for. Our findings highlight the importance of recombination in the ongoing process by which commensal bacteria adapt to their human hosts.

41. Unveiling the Genetic Landscape and Evolutionary Consequences of Tumor Initiation in Hereditary Leiomyomatosis and Renal Cell Cancer

Selina Wu^{1,2}, Brian Shuch³, Paul Boutros^{2,3,4}

¹Medical Informatics Interdepartmental Program,

²Jonsson Comprehensive Cancer Center,

³Department of Urology,

⁴Department of Human Genetics,

University of California Los Angeles, Los Angeles California, USA

Hereditary Leiomyomatosis and Renal Cell Cancer (HLRCC) is a rare hereditary condition affecting 1 in 200,000 individuals globally, with 15-20% developing aggressive papillary renal cell carcinoma. The genetic landscape and structural variants (SVs) of HLRCC tumors remain unexplored, but this study of three whole genome sequencing tumor-blood pairs aims to uncover genetic events triggering tumor initiation and potential early intervention targets. Mutational analysis reveals low single nucleotide variant density with no recurrent mutations or mutational signatures, suggesting alternative mechanisms drive HLRCC tumorigenesis such as SVs. Copy number alterations affect up to 45% of the genome in one tumor, with widespread gains on chromosomes 2 and 16 and extensive loss of heterogeneity across all samples. Subclonal reconstruction reveals the clonal and subclonal architecture, providing insight into tumor evolution events. Understanding HLRCC subclones can guide early detection biomarkers and personalized treatments for enhanced therapeutic efficacy.

42. From B-cell fate decisions to an antibody repertoire: the impact of non-genetic cell state heterogeneity and its heritability

Mark Xiang^{1,2,3}, Haripriya Vaidehi Narayanan^{2,3}, Alexander Hoffmann^{2,3}.

¹Bioinformatics Interdepartmental Program, ²Signaling Systems Laboratory, Department of Microbiology, Immunology and Molecular Genetics (MIMG), ³Institute for Quantitative and Computational Biosciences (QCB), University of California Los Angeles, CA.

Antibody-mediated immunity relies on a Darwinian selection process to produce a diverse repertoire of high-affinity antibodies. This process involves B-cell somatic hyper-mutation and affinity-based selection. However, B-cell fate decisions are affected by heterogeneous epigenetic states and molecular stochasticity. We present here a mathematical model to investigate how such heterogeneity affects the resulting antibody repertoire. Considering a simple Darwinian selection process, stochasticity in cell fate decisions renders antibody affinity maturation less effective. However, B-cell biology involves the differentiation of the highest affinity cells into antibody secreting plasma cells. This addition to the model caps

the maturation efficiency, but stochasticity in cell fate decisions actually recovers it. In this context, heritability of cell states within the proliferative burst further accelerates affinity maturation and diversifies the breadth of the antibody repertoire. These findings underscore the importance of considering non-genetic heterogeneity in order to understand antibody generation and design novel vaccine development strategies.

43. dsRID: Editing-free in silico identification of dsRNA region using long-read RNA-seq data

Ryo Yamamoto¹, Zhiheng Liu², Mudra Choudhury¹, Xinshu Xiao^{1,2,3*}

¹Bioinformatics Interdepartmental Program,

²Department of Integrative Biology and Physiology,

³Molecular Biology Institute,

University of California, Los Angeles, California, USA

Double-stranded RNAs (dsRNAs) are potent triggers of innate immune responses upon recognition by cytosolic dsRNA sensor proteins. Identification of endogenous dsRNAs helps to better understand the dsRNAome and its relevance to innate immunity related to human diseases. Here, we report dsRID (double-stranded RNA identifier), a machine learning-based method to predict dsRNA regions *in silico*, leveraging the power of long-read RNA-sequencing (RNA-seq) and molecular traits of dsRNAs. Using models trained with PacBio long-read RNA-seq data derived from Alzheimer's disease (AD) brain, we show that our approach is highly accurate in predicting dsRNA regions in multiple datasets. Applied to an AD cohort sequenced by the ENCODE consortium, we characterize the global dsRNA profile with potentially distinct expression patterns between AD and controls. Together, we show that dsRID provides an effective approach to capture global dsRNA profiles using long-read RNA-seq data.

44. Conformational Analysis of Chromosome Structures Reveals the Role of Chromosome Morphology in Gene Function

Yuxiang Zhan^{1,2}, Asli Yildirim¹, Lorenzo Boninsegni¹, Frank Alber^{1,2*}

¹Institute of Quantitative and Computational Biosciences, Department of Microbiology, Immunology, and Molecular Genetics, University of California Los Angeles, 520 Boyer Hall, Los Angeles, CA 90095

²Department of Quantitative and Computational Biology, University of Southern California, 1050 Childs Way, Los Angeles, CA 90089, USA

The 3D conformations of chromosomes are highly variant and stochastic between single cells. Recent progress in multiplexed 3D FISH imaging, single cell Hi-C and whole genome structure modeling allows a closer analysis of the structural variations of chromosomes between cells to infer the functional implications of structural heterogeneity. Here, we introduce a two-step dimensionality reduction method to classify a population of single cell 3D chromosome structures into dominant clusters with distinct chromosome morphologies. We found that almost half of all structures for each chromosome can be described by 5-12 dominant chromosome morphologies, which play a fundamental role in establishing global conformational variation of chromosomes. The same morphologies are observed in different cell types, but vary in their relative proportion of structures. Chromosome morphologies are distinguished by the presence or absence of characteristic chromosome territory domains, which cause some chromosomal regions to be exposed

to different nuclear environments in different morphologies. Our method provides an important approach to assess the variation of chromosome structures between cells and link differences in conformational states with distinct gene functions.

A B.I.G. Thank You

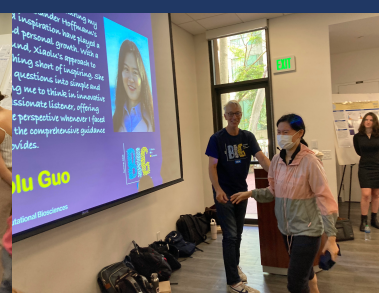
To all mentors for a successful Summer Program 2023!



- June 20 to August 11, 2023
- 9th Annual Program
- 440 applications received
- 74 admitted students
- 42 faculty mentors
- 43 direct research mentors



**Cheers to all the 74 accomplished students
and to our 42 dedicated faculty mentors
and 43 student mentors
who made the B.I.G. Summer program a success!**



Talks, Posters and Abstracts Posted on the QCBio Website at:
<https://qcb.ucla.edu/big-summer/big2023/>

UCLA Institute for Quantitative & Computational Biosciences