10th Annual QCBio Retreat

Tuesday, September 24, 2024

9:00am – 7:00pm

Boyer Hall and Court of Sciences

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-richer 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 a new Systems Biology Home Area will launch next year 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 collabroation 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 10th year we hosted 69 students within the B.I.G. Summer Undergraduate Research Program this year with 49 laboratories participating! (https://qcb.ucla.edu/big-summer/big2024/). 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 biociences 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 partipate 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?

QCBio is here for you!
Special thanks to Caroline Baron for again organizing the Retreat this year! Alexander Hoffmann

Agenda

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

9:15 am **WELCOME**

9:30 am **STATUS REPORTS I**

- **Alexander Hoffmann**, Director, QCBio, BIG Summer
- **Dan Geschwind,** Director, Institute for Precision Health
- **Noah Zaitlen,** Computational Medicine, CGSI
- **Matteo Pellegrini**, Director, QCBio Collaboratory

9:50 am **KEYNOTE I – introduced by Jingyi Jessica Li**

- **Mark Handcock,** Distinguished Professor of Statistics, Interim Faculty Director of Fundamental Data Science, DataX, Department of Statistics and Data Science, UCLA

10:20 am **COFFEE & TEA BREAK**

10:45 am **SELECTED TALKS I chaired by Matteo Pellegrini**

- **Mohammadali Alidoost**, Bioengineering PhD student, Wilson Lab
- **Roni Haas**, Postdoc, Boutros Lab
- **Elaine Huang,** Bioinformatics student, Xiao Lab

11:40 am **STATUS REPORTS II**

- **Alex Bui**, Director, Medical Informatics Ph.D. Home Area
- **Roel Ophoff**, Bioinformatics Interdepartmental Ph.D. Program
- **Roy Wollman**, Interim Director, Systems Biology Ph.D. Home Area
- **Paivi Pajukanta**, Director, Genetic & Genomics, Ph.D. Program
- **Harold Pimentel**, Director, Biomathematics, Ph.D. Program

12:15 pm **LUNCH**

1:30 pm **KEYNOTE II – introduced by Wei Wang**

- **Nanyun (Violet) Peng**, Associate Professor in Computer Science, UCLA

2:00 pm **SELECTED TALKS II – chaired by Elisa Franco**

- **Francesco Musella**, Bioinformatics PhD student, Alber Lab
- **Eiji Nakamura**, Mechanical and Aerospace Engineering PhD student, Franco Lab
- **Richard Wolf,** Ecology and Evolutionary Biology PhD student, Garud lab

2:45 pm **COFFEE & TEA BREAK**

3:15 pm **SELECTED TALKS III – chaired by Eric Deeds**

- **Serena Hughes**, Bioinformatics PhD student, Deeds Lab
- **Bronson Jeong**, Computer Science PhD student, Sankararaman Lab
- **Lajoyce Mboning**, Chemistry and Biochemistry PhD student, Pellegrini Lab

4:00 pm **STATUS REPORTS III**

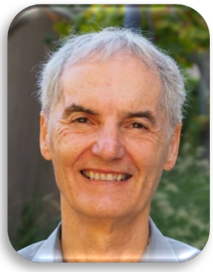
- **Matteo Pellegrini**, Director of the Computational and Systems Biology Major
- **XXXXXX**, Bioinformatics Minor
- **Eric Deeds**, Director of the Life Science Math Core
- **QBio-EDGE**

QCBio Research Excellence Award – TBN – awarded by Alexander Hoffmann

4:30 pm **RECEPTION & REFRESHMENTS**

POSTER SESSION

Keynote Speakers



Mark S. Handcock
Distinguished Professor of Statistics
Interim Faculty Director of Fundamental Data Science, DataX
Department of Statistics and Data Science
UCLA
https://faculty.stat.ucla.edu/handcock

DataX and some models for networks

DataX is an organized research unit with the objective of enhancing data-centric research and education on campus. It will work to enhance collaboration and communication between the data and domain sciences. I will provide an overview of DataX may help your work and its relationship to QCBio. I will give a brief overview of some statistical models for networks that may be of interest to those studying complex interacting systems.

Nanyun Peng
Assistant Professor
Computer Science
UCLA

Discourse-Level Natural Language Understanding For Biomedical Domains

Large language models (LLMs), such as ChatGPT and BioLM, have made significant strides in natural language understanding (NLU), excelling at tasks like summarization, question-answering, and entity extraction. In the biomedical field, the potential of LLMs is particularly exciting due to the sheer volume of complex, specialized text such as scientific papers, clinical notes, and drug data. Biomedical texts contain intricate relationships between entities like diseases, proteins, and drugs, making them challenging but critical for NLU. By fine-tuning LLMs for biomedical tasks, we can better extract valuable information—such as disease-drug interactions and clinical outcomes—from unstructured data, aiding research and clinical decision-making. However, while LLMs have demonstrated great success in general-domain tasks, applying them to biomedical NLU requires domain-specific adaptation and fine-tuning. Biomedical texts have unique characteristics: they contain highly specialized vocabulary, complex discourse structures, and critical contextual relationships that are not present in the open-domain datasets used to train most LLMs. In this talk, we explore advanced methods for discourse-level natural language understanding (NLU) tailored to the complexities of the biomedical domain, with applications to biomedical relation extraction for precision medicine, summarization for clinical notes, and clinical event extraction with LLMs.

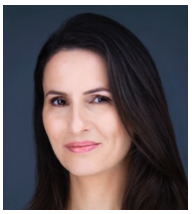
Selected Talks

- **Enhancing drug side effects detection through pathway phenotypes and drug-target interaction**

Mohammadali Alidoost[1], Jennifer L. Wilson[1]
[1] Department of Bioengineering, University of California Los Angeles, Los Angeles California, USA

Accurate prediction of drug side effects remains a significant challenge in pharmaceutical development because drug programs fail due to unforeseen adverse reactions. Protein-protein interaction (PPI) network models have the potential to predict adverse drug effects, but suffer from limited prediction performance, specifically over- and under- prediction. First, we refined pathway phenotypes by integrating key network genes and omics data into PathFX, our PPI model, and second, we incorporated new drug-binding targets from multiple databases. New pathway phenotypes and new drug targets enabled PathFX to predict previously unrecognized side effects (underprediction), and some pathway phenotypes eliminated PathFX false positive predictions (overprediction). In both scenarios, we observed a trade-off between specificity and sensitivity. Taken together, tuning new pathway definitions and drug target inputs suggests a path towards improving prediction performance and rationally using PPI models to anticipate drug-induced side effects.

- **Divergent Evolution in Bilateral Prostate Cancer: a Case Study**



Roni Haas[1,2,3,4,†], Yash Patel[1,2,3,4], Lydia Y. Liu[1,2,3,4], Rong Rong Huang[5], Adam Weiner[2,3,4], Takafumi N. Yamaguchi[1,2,3,4], Raag Agrawal[1,2,3] , Paul C. Boutros[1,2,3,4,†], Robert E. Reiter[2,3,4,†]
[1] Department of Human Genetics, University of California, Los Angeles, USA
[2] Department of Urology, University of California, Los Angeles, USA
[3] Jonsson Comprehensive Cancer Center, University of California, Los Angeles, USA
[4] Institute for Precision Health, University of California, Los Angeles, USA
[5] Department of Pathology, University of California, Los Angeles, USA
[†] Corresponding authors

Multifocal prostate cancer is a prevalent phenomenon, with most cases remaining uncharacterized from a genomic perspective. A patient presented with bilateral prostate cancer. On systematic biopsy, two indistinguishable clinicopathologic lesions were detected. Whole-genome sequencing displayed somatically unrelated tumours with distinct driver CNA regions, suggesting independent origins of the two tumors. We demonstrated that similar clinicopathologic multifocal tumours, which might be interpreted as clonal disease, can in fact represent independent cancers. Genetic prognostics can prevent mischaracterization of multifocal disease to enable optimal patient management.

- **Unveiling the hidden role of RNA stability as a link between genetic variation and disease**

Elaine Huang[1], Ting Fu[2], Guan'ao Yan[3], Ling Zhang[2], Kofi Amoah[1], Ryo Yamamoto[1], Sari Terrazas[4], Thuy Linh Nguyen[2], Carlos Gonzalez-Figueroa[2], Jae Hoon Bahn[2], Rajagopal Varada[2], Armen Khanbabaei[5], Jonatan Hervoso[1], Michelle T. Paulsen[6,7], Brian Magnuson[8], Mats Ljungman[6,7], Jingyi Jessica Li[1,3,9,10,11], Xinshu Xiao[1,2,4,5]
[1]Bioinformatics Interdepartmental Program, University of California, Los Angeles, CA [2]Department of Integrative Biology and Physiology, University of California, Los Angeles, CA [3]Department of Statistics, University of California, Los Angeles, CA [4]Molecular Biology Interdepartmental Program, University of California, Los Angeles, CA [5]Molecular, Cellular, and Integrative Physiology Interdepartmental Program, University of California, Los Angeles, CA [6]Center for RNA Biomedicine and Rogel Cancer Center, University of Michigan, Ann Arbor, MI [7]Department of Radiation Oncology and Environmental Health Sciences, University of Michigan, Ann Arbor, MI [8]Department of Pathology, University of Michigan, Ann Arbor, MI [9]Department of Biostatistics, University of California, Los Angeles, CA [10]Department of Computational Medicine, University of California, Los Angeles, CA [11]Department of Human Genetics, University of California, Los Angeles, CA

Gene expression is jointly modulated by transcriptional regulation and mRNA stability, yet the latter is often overlooked. Leveraging metabolic labeling data (Bru/BruChase-seq) and a new computational pipeline, RNAtracker, we categorize genes as allele-specific RNA stability (asRS) or allele-specific RNA transcription (asRT) events. We identify over 5,000 asRS variants among 665 genes across a panel of 11 ENCODE cell lines. These variants directly overlap conserved microRNA target regions and allele-specific RNA binding protein sites, illuminating mechanisms through which stability is mediated. Furthermore, we identified causal asRS variants using our massively parallel screen for variants that affect post-transcriptional mRNA abundance (MapUTR). Compared to asRT, asRS genes exhibit similar or higher prevalence in most cell lines analyzed.

Notably, asRS genes were enriched in immune-related pathways. This work highlights RNA stability as a critical, yet understudied mechanism linking genetic variation and disease.

- **DeepAtlas: A Tool to Explore the Manifold Hypothesis**

Serena J. Hughes [1-2], Timothy Hamilton [1-2], Ivy Xiong [1], Eric J. Deeds [1-3]
[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 manifold hypothesis states that high-dimensional datasets are sampled from low-dimensional latent manifolds, and thus can be studied in that low-dimensional space. Current standard "manifold learning" methods result in a global lower-dimensional embedding, but generally introduce high levels of topological distortion. Here we describe a new approach called the DeepAtlas that applies the mathematical definition of a manifold to generate low-dimensional models of data. The DeepAtlas generates local neighborhoods, embeds the neighborhoods in a lower dimension, and uses the embeddings to train continuous and invertible neural network models. Our tool includes a diagnostic step to directly evaluate whether the data is likely to be drawn from a manifold. We found that many popular datasets, including single-cell RNA sequencing data, actually do not have the expected manifold structure. The DeepAtlas can thus be used to directly test the manifold hypothesis and, if a manifold exists in the data, generate a robust mathematical model of that structure.

- **Scalable summary statistics-based heritability estimation method with individual genotype level accuracy**

Moonseong Jeong, Ali Pazokitoroudi, Zhengtong Liu, Sriram Sankararaman

SNP heritability, the proportion of phenotypic variation explained by genotyped SNPs, is an important parameter in understanding the genetic architecture underlying various diseases and traits. Methods that aim to estimate SNP heritability from individual genotype and phenotype data are limited by their ability to scale to Biobank-scale datasets and by the restrictions in access to individual-level data. These limitations have motivated the development of methods that only require summary statistics. While the availability of publicly accessible summary statistics makes them widely applicable, these methods lack the accuracy of methods that utilize individual genotypes. Here we present a SUMmary statistics-based Randomized Haseman-Elston regression (SUM-RHE), a method that can estimate the SNP heritability of complex phenotypes with accuracies comparable to approaches that require individual genotypes, while exclusively relying on summary statistics. SUM-RHE employs Genome-Wide Association Study (GWAS) summary statistics and statistics obtained on a reference population, which can be efficiently estimated and readily shared for public use. Our results demonstrate that SUM-RHE obtains estimates of SNP heritability that are substantially more accurate compared to other summary statistic methods and on par with methods that rely on individual-level data.

- **BayesAge 2.0: A Maximum Likelihood Algorithm to Predict Transcriptomic Age**

Lajoyce Mboning[1], Emma Katherine Costa[2, 3], Jingxun Chen[4], Anne Brunet[4], Tony-Wyss Coray[2], Louis-S Bouchard[1], Matteo Pellegrini [4]
[1] Department of Chemistry and Biochemistry, University of California, Los Angeles, United States
[2] Department of Neurology, and Neurological Sciences, Stanford University, United States
[3] Neurosciences Interdepartmental Program, Stanford University School of Medicine, United States
[4] Department of Genetics, Stanford University, United States
[5] Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, United States

Aging is a complex biological process influenced by various factors, including genetic and environmental influences. In this study, we present BayesAge 2.0, an improved version of our maximum likelihood algorithm designed for predicting transcriptomic age (tAge) from RNA-seq data. Building on the original BayesAge framework, which was developed for epigenetic age prediction, BayesAge 2.0 integrates a Poisson distribution to model count-based gene expression data and employs LOWESS smoothing to capture nonlinear gene-age relationships. BayesAge 2.0 provides significant improvements over traditional linear models, such as Elastic Net regression. Specifically, it addresses issues of age bias in predictions, with minimal age-associated bias observed in residuals. Its computational efficiency further distinguishes it from traditional models, as reference construction and cross-validation are completed more quickly compared to Elastic Net regression, which requires extensive hyperparameter tuning. Overall, BayesAge 2.0 represents a notable advance in transcriptomic age prediction, offering a robust, accurate, and efficient tool for aging research and biomarker development.

- **Infer Replication States from Multiplex FISH Imaging**

Francesco Musella[1,2,3], David Gilbert[4], Frank Alber[1,2,3]
[1] Bioinformatics IDP Program, University of California, Los Angeles, Los Angeles, CA, USA
[2] Department of Microbiology, Immunology and Molecular Genetics, Los Angeles, CA, USA
[3] Institute for Quantitative and Computational Biological Sciences, Los Angeles, CA, USA
[4] San Diego Biomedical Research Institute, San Diego, CA, USA

During replication, DNA is uniformly duplicated from initiation zones distributed across the genome. Although the genome 3D structure might regulate initiation firing, no experiment to date has jointly measured the single-cell DNA 3D structure and its replication states to investigate their relationship. Thus, we developed a statistical method to infer the replication states of genomic loci in single cells from DNA multiplex FISH imaging. Our method identifies each cell's cell cycle progression in G1, S, and G2 and estimates the replication state of each imaged locus in single cells. We applied our method to the mESC DNAseqFISH+ dataset and validated the results with FACS and RepliSeq. By providing accurate replication states to the imaged loci in multiplex FISH microscopy, our method allows us to investigate for the first time the relationship between genome structure and timing of replication initiation at the single cell level.s

- **Inversion properties of incoherent feedforward loop to process temporal information**

Eiji Nakamura[1,5], Franco Blanchini [2], Giulia Giordano[3], Alexander Hoffmann[4,5], Elisa Franco[1,5]
[1] Department of Mechanical and Aerospace Engineering, University of California Los Angeles, USA
[2] Department of Mathematics, Computer Science and Physics, University of Udine, Italy
[3] Department of Industrial Engineering, University of Trento, Italy
[4] Signaling Systems Laboratory, Department of Microbiology, Immunology, and Molecular Genetics, University of California Los Angeles, USA
[5] Institute for Quantitative and Computational Biosciences, University of California Los Angeles, USA

The incoherent feedforward loop (IFFL) motif is known for its versatile emergent properties, including pulse generation, adaptation, and fold-change detection. In this study, we describe a novel feature of the IFFL motif called the inversion property, where the input dose and either (or both) the duration and integral (area under the curve) of the output response are inversely related. The inversion property enables the IFFL to selectively respond to a pulsatile input. We first investigate the inversion property utilizing mathematical models of the IFFL to find parametrical/topological requirements for the inversion property. Additionally, we experimentally validate the inversion property of the IFFL using *in vitro* transcriptional networks. The inversion property has the potential to expand the uses of synthetic IFFL circuits and may contribute to our understanding of the mechanistic aspects of natural biological networks that process pulsatile signals.

- **Pervasive selective sweeps across human gut microbiomes**

Richard Wolff[1*], Nandita Garud[1,2*]
[1] Department of Ecology and Evolutionary Biology,
[2] Department of Human Genetics,
University of California Los Angeles, Los Angeles California, USA
* These authors contributed equally.

The human gut microbiome is composed of a highly diverse consortia of species which evolve within and across hosts. The ability to identify adaptations common to many host gut microbiomes would not only reveal shared pressures across hosts, but also key drivers of functional differentiation of the microbiome that may affect host traits. Here, we develop a novel selection scan statistic, named the integrated linkage disequilibrium score (iLDS), and apply this statistic to ~30 common commensal gut species. We find evidence of pervasive spread of positively selected alleles across human microbiomes, mediated by horizontal gene transfer. Moreover, we identify selective differences between the microbiomes of Westernized and non-Westernized hosts, including potential adaptations to metabolize specific components of ultra-processed food (such as the synthetic starch maltodextrin) in the Westernized cohort. In summary, we find that selective sweeps are a common feature of the human gut microbiome.

Collaboratory Fellows 2024-2025

Matteo Pellegrini, **Director**

Eloy Lopez, **Program Manager**

Fei-Man Hsu

Daniel Ha

Kelsey Jorgensen

Akos Rudas
New Fellow

Ivy Xiong

Giorgia Del Vecchio

Cameron Gill
New Fellow

Lukasz Salwinski

Seyoon Ko

Wenbin Guo

Montgomery Blencowe
New Fellow

Weihong Yan

Pavan S. Holur
New Fellow

Xianglong Tan
New Fellow

Shawn Cokus

Lingyu Zhan

Giovanni Quinones Valdez

Karolina Kaczor-Urbanowicz

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 our Incoming Medical Informatics Students!

Irsyad Adam
UCLA

Dominic Amara
UCSF

Chandler Beon
UC Berkeley

Dongwoo (Richard) Lee
Johns Hopkins

Steven Swee
UC San Diego

Welcome our Incoming Genetics & Genomics Students!

Jacob Argandona
UC Riverside

Mohammad Baig
York University

Timothy Derebenskiy
UC Santa Barbara

Joshua Hack
University of Arizona

Emily Hansen
Cal Poly San Luis Obispo

Emma Kumagai
University of Tsukuba

Suchita Lulla
University of Colorado Boulder

Noe' Reyna
University of Texas at Austin
B.I.G. SUMMER 2023 ALUMNUS

Annabel Sen
Brown University

Takafumi Yamaguchi
University of Manitoba

Welcome our Incoming Bioinformatics Masters Students!

Ryan Barney
UCLA
B.I.G. SUMMER 2024 ALUMNUS

Alison King
UCLA

Welcome our Incoming Bioinformatics Students!

Atef Ali
University of Minnesota

Hannah Faris
UCLA

Yiqian Gu
South China Normal University

Richard Hu
John Hopkins University

Aditya Iyer
University of North Carolina at Chapel Hill
B.I.G. SUMMER 2023 ALUMNUS

Nabil Mohammed
UC Santa Cruz University

Xinyue Qie
Yale University

Emma Wade
Mississippi State University
B.I.G. SUMMER 2021 ALUMNA

Yuxing Zhou
UCLA

Welcome our Incoming Biomathematics Students!

Vivek Agarwal
University of Maryland

Kai Akamatsu
UC San Diego

Mayuko Boffelli
UCLA

Karen Li
UC Santa Barbara & Yale University

Connor Morris
Brigham Young University & UCLA-Caltech MSTP

Joanne Qiu
UCLA

JJ Schirle
UCLA

Isha Tripathi
UCLA
B.I.G. SUMMER 2023 ALUMNA

Patrick Yuan
University of Melbourne

Poster Session

1. **Pathway Optimization Discovers Hidden Pathways and Mechanisms for Pulmonary Arterial Hypertension and its Potential Treatments**

Youzi Bi[1], Jennifer Wilson[1]
[1]Department of Bioengineering, Samueli School of Engineering, University of California, Los Angeles

Pulmonary arterial hypertension (PAH) is a disease associated with complex gene and cellular pathways. However, incorporating omics approaches often reveals incomplete pictures of disease changes. This study aims to use pathway optimization approaches to uncover the hidden key players in PAH. Using the pathway optimization algorithm, Prize Collecting Steiner Forest (PCSF), and Gene Ontology (GO) enrichment analysis, we analyzed high throughput RNA sequencing data in multiple tissue and cell types under various conditions. We uncovered hidden pathways with high confidence. Later, we assessed pathways in different tissues 15-HETE disease-induction and multiple treatments. While the differential genes and pathways were largely unique, they converged on similar GO functions. Our results indicate that treatments can achieve similar outcomes through distinct mechanisms but similar cellular processes. The work not only advances our understanding of the mechanisms of PAH and different treatments, but also opens opportunities for network methods to discover potential treatments.

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

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

Combination therapies are regarded as the future for treating cancer, yet predicting synergy between multiple drugs remains challenging. Even sophisticated machine learning models with ample single perturbation and cell line information report model-experiment correlations of ~0.24-0.48. Separately, protein-protein interaction network methods have successfully predicted drug effects, however, few have considered de novo synergy prediction solely from network topology. Thus, we measured the extent to which topological relationships between drug protein targets can predict experimental synergy. We quantified topology by exhaustively testing distance metrics and found that they were moderately aligned with synergy (correlations of 0.35-0.58), suggesting that network topology alone is as performant as experiment-informed models. We further used sensitivity analyses to understand performance differences between distance metrics encoded by varied network features. Guided by our reported optimal network features, models may be a powerful approach for de novo synergy prediction and finding effective drug combinations.

3. **Emerging zoonoses at the point of introduction: the effect of risk heterogeneities on dynamics and control**

Santiago Cardenas[1] and James Lloyd-Smith[1,2]
[1]Biomathematics, Department of Computational Medicine
[2]Depart ment of Ecology and Evolutionary Biology
University of California, Los Angeles

Controlling the incidence of newly introduced pathogens while they exhibit subcritical (i.e. $0<R_0<1$) transmission is crucial to reducing disease burden and minimizing the risk of outbreaks. For emerging zoonoses, cases arise from a mixture of introductions and onward transmission. Heterogeneities in the rates of spillover and transmission offer the potential for efficient targeted control, but the dynamics of subcritical pathogens in heterogeneous populations have not been studied. Here, we develop a two-group branching process model to characterize how risk heterogeneity can shape patterns of incidence for subcritical pathogens. We quantify disease burden using the distributions of the chain's total progeny and time to extinction. We then compare the impacts of applying prevention and mitigation control strategies to target subgroups and present recommendations under representative transmission models. Our study expands the theoretical foundation for understanding how heterogeneous risk behaviors shape zoonotic emergence and provides rational recommendations for control.

4. **Identifying Liver Transplantation Rejection Mechanisms Through Integration of Longitudinal Donor and Recipient Immunological Measurements**

Jackson L. Chin1, Cyrillus Zhixin Tan2, Aaron S. Meyer1, 2
1Department of Bioengineering, University of California, Los Angeles, CA
2Bioinformatics Interdepartmental Program, University of California, Los Angeles, CA

Up to 17% of liver transplants (LTs) result in chronic rejection. Liver ischemia-reperfusion injuries (LIRI) accrued during transplant are suspected to contribute to rejection, though the relationship between LIRI and LT outcome remains incompletely understood. As LIRI signatures manifest across time and donor/recipient tissues, we hypothesize that understanding LT rejection requires integration of longitudinal donor and recipient measurements. Here, we collected cytokine and liver function test measurements from LT donors and recipients at multiple timepoints along the LT process. We applied tensor partial least squares (tPLS)—a supervised, tensor-based decomposition method—to integrate these longitudinal measurements. We find that tPLS successfully captured determinants of LT outcome, predicting transplant rejection with an accuracy of 72% and highlighting that Th2-mediated immunoregulation and IFNγ-driven regeneration are critical for preventing LT rejection. Collectively, these efforts demonstrate the power of tPLS in integrating longitudinal measurements and highlight immunological targets for future therapeutic and screening efforts.

5. **Analysis of Intracellular Communication Reveals Consistent Gene Changes Associated with Early-Stage Acne Skin.**

Min Deng, Woodvine O. Odhiambo, Min Qin, Thao Tam To, Gregory M. Brewer, Alexander R. Kheshvadjian, Carol Cheng, & George W. Agak*
Division of Dermatology, David Geffen School of Medicine. University of California (UCLA), Los Angeles, CA-90095, USA
*Correspondence: Gagak@mednet.ucla.edu

A comprehensive understanding of the molecular changes driving cell interactions within acne lesions remains elusive. To address this, we analyzed early papules from six acne subjects using scRNA and spatial RNA-seq. Utilizing CellChat, we mapped an atlas of signaling pathways for healthy skin, and our comparative analysis revealed alterations in 49 pathways in lesional skin. Notably, we identified ten consistently dysregulated molecules across all donors, including GRN, IL-13RA1, and SDC1. We focused on GRN and IL-13RA1 due to their potential roles in inflammation and hyperkeratinization. Further investigation showed that GRN upregulated proinflammatory cytokines and chemokines, such as IL-18, CCL5, and CXCL2 in TREM2 macrophages. Additionally, IL-13 activation of IL-13RA1 in HaCaT cells led to the dysregulation of hyperkeratinization-associated genes. These findings highlight the GRN-SORT1 and IL-13-IL-13RA1 axes as critical players in acne pathogenesis, suggesting that targeting these pathways may offer promising new therapeutic strategies for acne.

6. **Systematic comparison of single-cell RNA-seq methods**

Alejandro Espinoza[1,2,3], Bruno Andrade[5*], Robert Modlin[4,5], Matteo Pellegrini[2,3]
[1] Department of Human Genetics,
[2] Institute for Quantitative and Computational Biosciences-The Collaboratory
[3]Department of Molecular Cell and Developmental Biology
[4]Division of Dermatology, Department of Medicine, David Geffen School of Medicine
[5]Department 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.

7. **A BrainSuite pipeline for quantifying neuroanatomical variation over time**

Siqi Fang[1], Yeun Kim[2], Anand Joshi[4], David Shattuck[3], Daniel Tward[1]

[1] Department of Computational Medicine, University of California, Los Angeles
[2] Department of Bioengineering, University of California, Los Angeles
[3] Department of Neurology, University of California, Los Angeles
[4] Department of Electrical and Computer Engineering, University of Southern California

Quantifying variability in anatomical form is critical to understanding growth, aging, or neurodegeneration. The latter recognized as a biomarker for Alzheimer's disease (AD) We designed a new computational pipeline to analyze neuroimaging data over time. We register sets of surfaces with an optimal diffeomorphism, which leads to a minimization problem over regularization and matching terms. From a given atlas template, we jointly estimate a sequence of diffeomorphic transformations, which provides an alignment between our atlas template and each observation. We can estimate changes to thickness and surface area without referring to sources of variability explained by residual deformations and differences in pose. We applied our pipeline to a set of T1-weighted MRI from OASIS-3. We computed the optimal deformation for patient samples and measured cortical thickness and surface area change. We used a mixed-effects model and showed statistically significant results that differentiated patients with AD from the control group.

8. **Identifying common disease trajectories of Alzheimer's disease with electronic health records**

Mingzhou Fu[1,2], Sriram Sankararaman[3], Bogdan Pasaniuc[4], Keith Vossel[1], Jason D. Hinman[1], Timothy S. Chang[1*]
[1] Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, 90095, United States
[2] Medical Informatics Home Area, Department of Bioinformatics, University of California, Los Angeles, Los Angeles, CA, 90024, United States
[3] Computational Medicine, University of California, Los Angeles, Los Angeles, CA, 90024, United States
[4] Genomics and Computational Biology, University of Pennsylvania, PA, 19104, United States

Alzheimer's disease (AD, poses a growing global public health challenge. While recent studies have identified AD risk factors, they often focus on specific comorbidities, neglecting the complex interrelations and temporal dynamics. We address this by analyzing AD progression through longitudinal trajectories, utilizing clinical diagnoses over time. Using machine learning and network analysis, we created a computational framework to identify common AD progression patterns. Our analysis included 24,473 eligible AD patients from UC Health Data Warehouse's Electronic Health Records. We identified four trajectory clusters: 1) a mental health cluster; 2) an encephalopathy cluster; 3) a neurodegenerative disease cluster; and 4) a vascular disease cluster. Significant differences were observed in demographics, symptoms, and AD features across clusters. Causal analysis indicated that 26.2% of the identified trajectory connections were causal. Our findings can significantly benefit patient care and medical research by moving toward earlier and more accurate diagnoses, along with personalized treatment, such as medical risk factors management and lifestyle modifications.

9. **Modeling heterogeneous signaling dynamics of macrophages reveals principles of information transmission in stimulus responses**

Xiaolu Guo[1,2], Adewunmi Adelaja[1,2,3], Apeksha Singh[1,2], Roy Wollman[1,4], Alexander Hoffmann*[,1,2]
[1] Department of Human Genetics,
[1] Institute for Quantitative and Computational Biosciences, University of California Los Angeles, Los Angeles, USA
[2] Department of Microbiology, Immunology, and Molecular Genetics, University of California Los Angeles, Los Angeles, USA
[3] Current address: Harvard combined Dermatology Residency Training Program, Boston, MA, USA
[4] Department of Integrative Biology and Physiology, University of California Los Angeles, Los Angeles, USA
* Corresponding author: ahoffmann@ucla.edu

Macrophages initiate pathogen-appropriate immune responses with the activation dynamics of transcription factor NFκB mediating specificity. Live-cell imaging NFκB dynamics enabled quantifying stimulus-response specificity (SRS) of populations of cells. To study SRS beyond what is experimentally accessible, we developed a mathematical model that captures the SRS performance of stimulus-specific heterogeneous NFκB dynamics. Complementing experimental data, extended-dose response simulations improved channel capacity estimates. By

collapsing parameter distributions, we located information loss to receptor modules, while the negative-feedback-containing core module showed remarkable signaling fidelity. Further, constructing single-cell network models enabled quantification of single-cell stimulus-response specificity (scSRS). We found that despite SRS limitations at the population level, the majority of single cells are capable of responding specifically to immune threats, and that the few instances of stimulus-pair confusion are highly uncorrelated. The diversity of "blindspots" enable small consortia of 2~6 macrophages to achieve perfect stimulus distinction.

10. **Minority Reporting: Leverage and its consequences for scRNA-seq.**

[1]Timothy Hamilton, [2]Juan V. Najar, [3]Elena Dry, [4]Eric J. Deeds
[1]Bioinformatics IDP,
[2]Department of Computational and Systems Biology,
[3]Department of Molecular, Cellular and Developmental Biology
[4]Department of Integrative Biology and Physiology, University of California, Los Angeles

Single-cell methods allow researchers to analyze the behavior of multicellular tissues at the level of individual cells. These methods have revealed surprising amounts of heterogeneity in well-studied classical systems, However, the current pipelines to analyze the data can be unstable and difficult to replicate. Here, we demonstrate that even removing a small number of cells randomly from the dataset (e.g. 5% or even less) can generate completely different clustering results from those generated by analyzing the entire dataset. We found that this instability is in part due to the fact that a small minority of cells are responsible for driving the covariance structure in scRNA-seq data, allowing a small number of cells to drive steps like PCA and clustering. Our findings suggest the need to develop more robust approaches to single-cell data analysis.

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

Mariana Harris[1], Bernard Kim[2], and Nandita Garud[3,4]
[1] Department of Computational Medicine Biomathematics Program, University of California Los Angeles, Los Angeles California, USA
[2]Department 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*.

12. **Temporally distinct 3D multi-omic dynamics in the developing human brain**

Matthew Heffel

The human hippocampus and prefrontal cortex play critical roles in learning and cognition [1,2], yet the dynamic molecular characteristics of their development remain enigmatic. Here we investigated the epigenomic and 3D chromatin conformational reorganization during the development of the hippocampus and prefrontal cortex, using more than 53,000 joint single-nucleus profiles of chromatin conformation and DNA methylation (snm3C-seq3) [3]. The remodeling of DNA methylation (mC) is temporally separated from chromatin conformation dynamics. Using single-cell profiling and multi-modal single-molecule imaging approaches, we have found that short-range chromatin interactions are enriched in neurons, whereas long-range interactions are enriched in glial cells and non-brain tissues. We reconstructed the regulatory programs of cell-type development and differentiation, finding putatively causal common variants for schizophrenia strongly overlapping with chromatin loop-connected, cell-type-specific

regulatory regions. Our data provides multimodal resources for studying gene regulatory dynamics in brain development and demonstrates that single-cell 3D multi-omics is a powerful approach for dissecting neuropsychiatric risk loci.

13. cfTools: an R/Bioconductor package for deconvolving cell-free DNA via methylation analysis

Ran Hu[1,2,3*], Shuo Li[2*], Mary L. Stackpole[2], Qingjiao Li[2], Xianghong Jasmine Zhou[2,3], Wenyuan Li[2]
[1] Bioinformatics Interdepartmental Program,
[2] Department of Pathology and Laboratory Medicine,
[3] Institute for Quantitative & Computational Biosciences,
University of California Los Angeles, Los Angeles California, USA
* These authors contributed equally.

Analyzing cell-free DNA (cfDNA) methylation profiles holds significant potential for cancer detection and monitoring. However, deconvolving cfDNA originating from tumors or specific tissues remains challenging due to a scarcity of specialized bioinformatics tools. To address this, we created cfTools, an R package encompassing three computational methods recently developed for deconvolving cfDNA based on its origin from tumor or normal tissues. For cancer detection and monitoring, cfTools offers a function that sensitively detects tumor derived cfDNA fragments and estimates tumor burden in cfDNA. For tracing tissue of origins and predicting the tissue composition of cfDNA, cfTools provides two functions: a state-of-the-art unsupervised probabilistic approach, and a novel supervised deep learning method. By identifying an abnormally elevated composition of specific tissues, cfTools can infer the presence of underlying pathological conditions, including but not limited to cancer.

14. DeepAtlas: A Tool to Explore the Manifold Hypothesis

Serena J. Hughes [1-2], Timothy Hamilton [1-2], Ivy Xiong [1], Eric J. Deeds [1-3]
[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 manifold hypothesis states that high-dimensional datasets are sampled from low-dimensional latent manifolds, and thus can be studied in that low-dimensional space. Current standard "manifold learning" methods result in a global lower-dimensional embedding, but generally introduce high levels of topological distortion. Here we describe a new approach called the DeepAtlas that applies the mathematical definition of a manifold to generate low-dimensional models of data. The DeepAtlas generates local neighborhoods, embeds the neighborhoods in a lower dimension, and uses the embeddings to train continuous and invertible neural network models. Our tool includes a diagnostic step to directly evaluate whether the data is likely to be drawn from a manifold. We found that many popular datasets, including single-cell RNA sequencing data, actually do not have the expected manifold structure. The DeepAtlas can thus be used to directly test the manifold hypothesis and, if a manifold exists in the data, generate a robust mathematical model of that structure.

15. Capturing ADAR1-associated transcriptomic changes during neuronal differentiation using combinatorial indexing scRNA-seq

Armen Khanbabaei[1-3*], Jae Hoon Bahn[2-3*], Cong Liu[2-3], Xinshu Xiao[1-3]
[1] Molecular, Cellular & Integrative Physiology (MCIP) IDP,
[2] Quantitative and Computational Biology (QCBio),
[3] Department of Integrative Biology and Physiology,
University of California Los Angeles, Los Angeles California, USA
[*] These authors contributed equally.

ADAR1 is the primary enzyme responsible for A-to-I RNA editing, a crucial process in post-transcriptional regulation. ADAR1 is known to play a critical role in brain development and neuronal function, with its dysregulation implicated in various neurodevelopmental and neurodegenerative disorders. In this study, we aim to investigate the role of ADAR1 during the differentiation of neural precursor cells, focusing on its impact on gene expression, splicing and RNA editing in single cells. To achieve full-length RNA coverage in scRNA-seq, we optimized dissociation protocols and redeveloped established combinatorial index-based scRNA-seq methods, including EasySci, SPLiT-seq, and Parse Bio Evercode. This technique allows full-length RNA interrogation, overcoming the limitation of other scRNA-seq techniques that focus only on the 3' or 5'

end of RNA. In addition, it does not require specialized equipment, confer flexibility in handling various cell sizes, and preserves whole-cell integrity. These enhancements allowed us to capture transcriptome-wide changes associated with ADAR1 in differentiating cells. This experimental technique will greatly facilitate investigations of transcript isoform and RNA editing patterns at single-cell resolution.

16. **A hierarchical Bayesian statistical model for identifying significant genes in high-throughput CRISPR screens**

Sandy Kim[1], Harold Pimentel[1,2]
[1] Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA
[2] Departments of Computational Medicine, Human Genetics, and Computer Science; Molecular Biology Institute; HHMI, UCLA, Los Angeles, CA

CRISPR-Cas9 has revolutionized genetic research by enabling large-scale, unbiased genetic perturbations through high-throughput CRISPR screens. Such screens introduce a guide RNA library into cells, which integrates into the DNA, allowing the interrogation of multiple phenotypes under selective pressure via high-throughput sequencing. Analysis of such experimental data poses itself as a challenge due to the high-dimensional and inherent relational nature of the data. Current methods for analyzing CRISPR screen data often ignore shared information among guide RNAs and rely on oversimplified assumptions, introducing biases. We introduce a Bayesian hierarchical model that leverages the hierarchical design of CRISPR screens, where multiple guide RNAs target a single gene. Our model infers effect sizes, variances, and gene significance while sharing information across gene, guide, and sample levels, providing a robust and flexible analysis tool. We show our method provides improved sensitivity in identifying significant genes and guide RNAs compared to existing methods and demonstrate its utility with experimental data.

17. **Modeling reveals the strength of weak interactions in stacked ring assembly**

Leonila Lagunes[1], Eric J. Deeds[1]
[1]Integrative Biology and physiology, University of California Los Angeles, Los Angeles, California, USA

While cells regulate vital processes with macromolecular machines, they synthesize these machines as individual components that assemble into functional complexes. A common motif is a stacked ring. Insights into stacked trimer assembly are crucial for understanding complex regulation. Here, we developed a mathematical model of stacked trimer assembly that accounts for different binding affinities between and within rings. Our main finding is that deadlock – a severe form of kinetic trapping– can be extremely long. Deadlock is worst when all the interfaces have high binding affinities. We predict that evolution avoids stacked trimers with uniformly strong affinities. Our findings reveal that most solved structures lack such strong interactions. To understand the origins of deadlock, our pathway analyses reveal that strong binding affinities lead to the use of multiple pathways, which consume subunits and intensifies deadlock. In sum, our work provides critical insight into the evolutionary pressures that have shaped stacked ring assembly.

18. **Detection of selective sweeps across human gut microbiomes**

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

Adaptation is widespread and pervasive in bacteria and underlies the evolution of important and well-studied traits, with antimicrobial resistance (AMR) as one example. In the human gut microbiome, AMR has been shown to evolve through mechanisms of horizontal gene transfer, where adaptations beneficial in multiple hosts spread via migration of a gene fragment rather than evolving *de novo* in individual hosts. Whether such adaptive mechanisms are common to gut commensals and facilitate the spread of non-AMR traits remains to be seen. Here, we conduct haplotype homozygosity scans in 25 gut commensal bacteria in the human gut to discover widely-shared genetic fragments driven by positive selection, and find abundant evidence of elevated haplotype homozygosity. Corresponding neutral diversity from recombination-and-diversity matched simulated bacterial populations are unable to recapitulate observed levels of homozygosity. We conclude that positive selection has been a common force shaping diversity across hosts, operating across multiple functional categories.

19. **L-Ornithine supplementation triggers neutral lipid accumulation in the microalgae *Chlamydomonas reinhardtii***

Leandro L. Lavandosque[1,2,3], Flavia V. Winck[1], Richard C. Law[3], Junyoung O. Park[3]
[1]Center for Nuclear Energy in Agriculture, University of São Paulo, São Paulo, São Paulo, Brazil
[2]Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, São Paulo, Brazil
[3]Department of Chemical and Biomolecular Engineering, University of California Los Angeles, Los Angeles California, USA

Chlamydomonas reinhardtii is known as a potential source of triacylglycerols for biotechnological applications, but feasible approaches to induce lipids accumulation still need to be pursued. Here, we investigated the accumulation of neutral lipids under culture supplementation with L-Ornithine. We used Nile Red staining to observe the lipid bodies' formation under confocal fluorescence microscopy and quantified relatively through fluorimetry. We further applied high-throughput phenotyping techniques such as shotgun label-free proteomics and time-resolved target metabolomics. We assessed candidate gene expression from the arginine catabolism pathway by qPCR. In addition to providing insight into unknown aspects of the ornithine metabolism in *Chlamydomonas*, these results can contribute to exploring the amino acid dynamics in microalgae and their implications for the lipid accumulation mechanism.

20. **Topological measures of antibody phylogenies reveal B-cell fate decisions**

Chengyuan Li[1], Haripriya Vaidehi Narayanan[2,3], Alexander Hoffmann[2,3]
1 BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
2 Institute for Quantitative and Computational Biosciences, UCLA
3 Department of Microbiology, Immunology, and Molecular Genetics, UCLA

The immune response to vaccination generates an antibody repertoire through Darwinian evolution involving selection, mutation, and expansion of B-cells. Impaired immune responses may be due to alterations in B-cell fate decisions but in vivo they cannot be directly observed. We asked whether cellular dynamics can be revealed from phylogenetic trees constructed from end-point antibody repertoire sequencing. We developed a mathematical model parameterizing survival, mutation, and expansion probabilities of B-cells under sequence-based selection, performed Monte Carlo simulations of phylogenetic trees in the repertoire, and analyzed distributions of graph-theoretic measures of topology and sequence abundance. We found that purely topological measures like root-to-tip depth are sensitive to mutation and death rates, while abundance-weighted measures reveal selection stringency and expansion rates. This novel approach yields quantitative insights across biological scales, inferring control of B-cell fate decisions from the resulting antibody repertoire. As a potential diagnostic measure it may inform strategies for personalizing vaccination.

21. **Reliable ligand discrimination in stochastic multistep kinetic proofreading: First passage time vs. product counting strategies**

Xiangting Li[1], Tom Chou[1,2]
[1] Department of Computational Medicine,
[2] Department of Mathematics,
University of California Los Angeles, Los Angeles California, USA

Cellular signaling, crucial for biological processes like immune response and homeostasis, relies on specificity and fidelity in signal transduction to accurately respond to stimuli amidst biological noise. Kinetic proofreading (KPR) is a key mechanism enhancing signaling specificity through time-delayed steps, although its effectiveness is debated due to intrinsic noise potentially reducing signal fidelity. In this study, we reformulate the theory of kinetic proofreading (KPR) by convolving multiple intermediate states into a single state and then define an overall “processing” time required to traverse these states. This simplification allows us to succinctly describe kinetic proofreading in terms of a single waiting time parameter, facilitating a more direct evaluation and comparison of KPR performance across different biological contexts such as DNA replication and T cell receptor (TCR) signaling. We find that loss of fidelity for longer proofreading steps relies on the specific strategy of information extraction and show that in the first-passage time (FPT) discrimination strategy, longer proofreading steps can exponentially improve the accuracy of KPR at the cost of speed. Thus, KPR can still be an effective discrimination mechanism in the high noise regime. However, in a product

concentration-based discrimination strategy, longer proofreading steps do not necessarily lead to an increase in performance. However, by introducing activation thresholds on product concentrations, can we decompose the product-based strategy into a series of FPT-based strategies to better resolve the subtleties of KPR-mediated product discrimination. Our findings underscore the importance of understanding KPR in the context of how information is extracted and processed in the cell.

22. **Comparative assessment of automatic segmentation methods for whole cell cryo X-ray tomography analysis**

Yuhui Li1,2, Kate White1, Raymond Stevens1, Frank Alber1,2
1Quantitative and Computational Biosciences, Department of Biological Science, University of California, 1050 Childs Way, Los Angeles, CA 90089, USA
2Institute of Quantitative and Computational Biosciences, Department of Microbiology, Immunology, and Molecular Genetics, University of California Los Angeles, 520 Boyer Hall, Los Angeles, CA 90095

Cryo-soft X-ray tomography is an imaging technique to study the cellular ultrastructure of cells at a resolution of 25-40 nm, which can be used to study the chromatin organization in nuclei of individual cells. Here, we investigate current state-of-the-art deep learning architectures for the automatic segmentation of soft X-ray tomograms of entire cells. Specifically, we focus on semantic segmentation to study the nuclear architecture in rat INS-1 pancreatic beta cells. We first focus on the automatic detection of the nuclear shape, followed by the distribution of heterochromatin and euchromatin as well as the detection of nucleoli and pericentromeric heterochromatin clusters. We perform our tests on 85 INS-1 cells for assessing the performance of gmm model. We use histogram equalization, label smoothing and image augmentation as preprocessing methods, model assembly and morphological operations as postprocessing methods. We examine trends of heterochromatin and euchromatin spatial distributions, which will be used with other data sources to study the genome architecture of pancreatic beta cells in future. Furthermore, we perform machine learning techniques and visualization of the features extracted by our model.

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

Michelle Loui[1], Scott Taylor[1], Crystal Xiao[1], Cyrillus Tan[2], Het Desai[1], Allison Brookhart[3], Sophia Ting[4], Gabriella DiBernardo[5], Laura B. James-Allan[5], Sanaz Memarzadeh[5], Aaron S. Meyer[1]
[1]Department of Bioengineering,
[2]Bioinformatics Interdepartmental Program,
[3]Institute for Quantitative and Computational Biosciences,
[4]Department of Microbiology, Immunology and Molecular Genetics,
[5]Department of Obstetrics and Gynecology, David Geffen School of Medicine,
University of California, Los Angeles, Los Angeles, CA , USA

In high-grade serous ovarian cancer (HGSOC) patients, malignant epithelial cells arise from the fallopian tube and ovarian surface epitheliums. Endogenous antibodies (anti-tumor antibodies; ATAbs) target these cells and should promote recognition by the immune system. Patient-derived tumors have been found to be frequently coated in IgG, and ATAbs are present both in the tumor mass and in the fluid that builds up in the peritoneum surrounding the tumor microenvironment. They are derived from B cells that have undergone somatic hypermutation, indicating an active immune response. However, despite their widespread abundance in HGSOC, ATAbs fail to eliminate the tumor cells. We hypothesized that ATAbs are unable to eliminate tumors due to the dysregulation of immune interactions via their Fc region. Therefore, we applied a quantitative, multiplexed assay for profiling the Fc properties and immune receptor interactions of ATAbs. Our data demonstrate that ATAbs interact poorly with FcγRIIIa—a potent activating receptor for antibody-dependent cellular cytotoxicity (ADCC) found on natural killer cells—due to fucosylation. Understanding the mechanisms of humoral immunity evasion will help with the prediction of therapeutic responses in cancer patients and uncover how immunotherapies might reactivate effective humoral immunity.

24. **Quantifying the shared genetic components of complex traits and Mendelian phenotypes**

Jonathan C. Mah[1], Kirk E. Lohmueller[2, 3,*], and Nandita Garud[2, 3,*]
1. Bioinformatics Interdepartmental Program, University of California, Los Angeles

2. Department of Ecology and Evolutionary Biology, University of California, Los Angeles
3. Department of Human Genetics, University of California, Los Angeles
* Co-correspondence to ngarud@ucla.edu and klohmueller@ucla.edu
Despite the importance of gut commensal microbiota to human health, there is little knowledge about their evolutionary histories, including their demographic histories and distributions of fitness effects (DFE) of mutations. Here, we infer the demographic histories and DFEs for amino-acid changing mutations of 27 of the most prevalent and abundant commensal gut microbial species in North Americans over timescales exceeding human generations. We find reductions in genetic variation in North American versus African rural microbiomes. Additionally, some species in North American microbiomes display contractions in population size and others expansions, potentially occurring at several key historical moments in human history. DFEs across species vary from highly to mildly deleterious, with accessory genes experiencing more drift compared to core genes. Within genera, DFEs tend to be more congruent, reflective of underlying phylogenetic relationships. Together, these findings suggest that gut microbes have distinct evolutionary histories, possibly reflecting their unique roles.

25. A probabilistic model of relapse in drug addiction

Sayun Mao[1], Tom Chou[1], Maria R. D'Orsogna[1, 2]
[1] Department of Computational Medicine, University of California Los Angeles, Los Angeles California, USA
[2] Department of Mathematics, California State University at Northridge, Los Angeles California, USA
More than 60% of individuals recovering from substance use disorder relapse within one year. Some will resume drug consumption even after decades of abstinence. The cognitive and psychological mechanisms that lead to relapse are not completely understood, but stressful life experiences and external stimuli that are associated with past drug-taking are known to play a primary role. Stressors and cues elicit memories of drug-induced euphoria and the expectation of relief from current anxiety, igniting an intense craving to use again; positive experiences and supportive environments may mitigate relapse. We present a mathematical model of relapse in drug addiction that draws on known psychiatric concepts such as the "positive activation; negative activation" paradigm and the "peak-end" rule to construct a relapse rate that depends on external factors (intensity and timing of life events) and individual traits (mental responses to these events). We analyze which combinations and ordering of stressors, cues, and positive events lead to the largest relapse probability and propose interventions to minimize the likelihood of relapse. We find that the best protective factor is exposure to a mild, yet continuous, source of contentment, rather than large, episodic jolts of happiness.

26. Detecting parallel adaptive changes in the gut microbiome

Aina Martinez i Zurita[1], Sharon Greenblum[2] and Nandita Garud[1,3]
[1] Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, USA
[2] Joint Genome Institute, Genome Technologies Group, Berkeley, CA
[3] Department of Ecology and Evolutionary Biology, University of California, Los Angeles, USA
The human gastrointestinal tract contains a diverse ecosystem of microorganisms involved in many aspects of human health and disease. Due to the large population size of bacteria in each host and their short generation time, it is estimated that an average human microbiome experiences billions of de-novo mutations every day. Hence, the same mutation may arise in multiple hosts, and if beneficial, could rise to high frequency in these hosts. Mutations changing in frequency in parallel across many hosts may be adaptive and thus have important functional relevance to the microbiome. We develop a statistical framework to detect cases of parallelism across many hosts. We demonstrate the effectiveness of our method using simulations. We further apply our framework to a large cohort of mother - baby dyads temporally sampled during the first year of life. Our work begins to uncover the dynamics of adaptive variants that were previously missed due to not causing severe allele fraction changes in the data.

27. Inversion properties of incoherent feedforward loop to process temporal information

Eiji Nakamura[1,5], Franco Blanchini [2], Giulia Giordano[3], Alexander Hoffmann[4,5], Elisa Franco[1,5]
[1] Department of Mechanical and Aerospace Engineering, University of California Los Angeles, USA
[2] Department of Mathematics, Computer Science and Physics, University of Udine, Italy
[3] Department of Industrial Engineering, University of Trento, Italy
[4] Signaling Systems Laboratory, Department of Microbiology, Immunology, and Molecular Genetics, University of California Los Angeles, USA
[5] Institute for Quantitative and Computational Biosciences, University of California Los Angeles, USA

The incoherent feedforward loop (IFFL) motif is known for its versatile emergent properties, including pulse generation, adaptation, and fold-change detection. In this study, we describe a novel feature of the IFFL motif called the inversion property, where the input dose and either (or both) the duration and integral (area under the curve) of the output response are inversely related. The inversion property enables the IFFL to selectively respond to a pulsatile input. We first investigate the inversion property utilizing mathematical models of the IFFL to find parametrical/topological requirements for the inversion property. Additionally, we experimentally validate the inversion property of the IFFL using *in vitro* transcriptional networks. The inversion property has the potential to expand the uses of synthetic IFFL circuits and may contribute to our understanding of the mechanistic aspects of natural biological networks that process pulsatile signals.

28. Functional Validation of Rare Pathogenic Variants in SETX Using a Disease Specific Transcriptional Biomarker

Kathie J. Ngo[1,2], Darice Y. Wong [1,3], Brent L. Fogel [1-4]
[1.] Department of Neurology, David Geffen School of Medicine, UCLA, Los Angeles, California
[2.] Bioinformatics Interdepartmental Program, UCLA, Los Angeles California
[3.] Clinical Neurogenomics Research Center, David Geffen School of Medicine, UCLA, Los Angeles California
[4.] Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles California

Genetic ataxias are clinically heterogenous neurodegenerative conditions and it is difficult to assign patphogenicity to rare gene variations solely based on DNA sequencing. *SETX* encodes a ubiquitous DNA/RNA helicase crucial for maintaining genome stability and loss-of-function mutations cause a recessive disorder, Ataxia with Oculomotor Apraxia Type 2 (AOA2). Here we utilize Weighted Gene Co-expression Network Analysis (WGCNA) to construct an AOA2-specific transcriptomic signature as a biomarker to functionally validate variants of uncertain significance in patients clinically suspected of AOA2. WGCNA from peripheral blood RNA of 20 AOA2 families identified five disease-specific modules, one of which was shown to effectively distinguish individuals with AOA2 from carriers (sensitivity 64, specificity 98%) and from patients with genetically distinct, yet phenotypically similar, neurological disorders. As further proof-of-concept, we utilized this transcriptomic biomarker to identify the first pathogenic mutation in a non-canonical *SETX* transcript, expanding the spectrum of mutations that contribute to AOA2.

29. Ensemble absolute metabolite quantitation in T cells reveals conserved features of immunometabolism

Samantha O'Keeffe[1], Elle McCue[2], Harin Sim[3], Junyoung O. Park[1*], [1]Department of Chemical and Biomolecular Engineering,
[2]Department of Mechanical and Aerospace Engineering,
[3]Department of Bioengineering,
University of California, Los Angeles, Los Angeles, CA, USA.

Metabolism plays a key role in immune cell proliferation, activation, and persistence. Quantitative understanding of immunometabolism underlies improving immune. Here we developed a simple technique for comprehensive absolute metabolite quantitation in T cells. We innovated an isotope-ratio-based approach that leverages the knowledge of absolute concentrations in the model systems, ^{13}C labeling, and liquid chromatography-mass spectrometry (LC-MS) to quantify T-cell metabolome *en masse*. Our approach involves simultaneous extraction of unlabeled T cells and ^{13}C-labeled reference cells to distinguish metabolite origins. We quantified absolute metabolite concentrations of ~80 metabolites in both Jurkat T cells and human primary CD4+ and CD8+ T cells. Absolute metabolite concentrations facilitate the

integration of metabolomics, proteomics and fluxomics using kinetic and thermodynamic laws. We obtained overall Gibbs free energy changes across glycolysis in both CD4+ (ΔG = –48 kJ/mol) and CD8+ (ΔG = –46 kJ/mol) T cells. Comprehensive absolute metabolite quantitation offers fundamental systems-level insights into immunometabolism.

30. Characterizing transcriptomic heterogeneity in pulmonary tuberculosis

Jonathan Perrie[1], Rosane Teles[2], Robert Modlin[2,5], Matteo Pellegrini[1,2,3,4,5]

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

[2] Division of Dermatology, David Geffen School of Medicine, University of California Los Angeles

[3] Molecular, Cell and Developmental Biology, University of California Los Angeles

[4] Human Genetics, University of California Los Angeles

[5] Microbiology, Immunology, and Molecular Genetics, University of California Los Angeles

TREM2+ macrophages play a critical role in active tuberculosis and have also been implicated in other diseases including acne and Alzheimer's. They are characterized by their plump, lipid-laden morphology, and in tuberculosis, have been found at both the site of initial infection within the alveoli and at the site of containment within the granuloma. In tuberculosis, there is a constant battle between containment and clearance of Mycobacterium tuberculosis. Notably, when phagocytized, the pathogen can evade host defenses by inhibiting phagosome-lysosome fusion and exploiting the lipids from TREM2+ macrophages as an energy source. Utilizing the NanoString CosMx platform, this study provides a spatially resolved transcriptomic analysis of lung tissue from patients with active tuberculosis. Every transcript has a set of coordinates and those that fall within a cell's segmented boundaries quantify its expression. Our results find specific gene programs characterizing the different cell types in the granuloma and alveoli including the TREM2+ macrophage. This analysis reveals a possible progression from initial infection to granuloma formation.

31. Integrative, high-resolution analysis of single cells across experimental conditions with PARAFAC2

Andrew Ramirez[1], Brian Orcutt-Jahns[1], Sean Pascoe[2], Armaan Abraham[1], Breanna Remigio[1], Nathaniel Thomas[1], Aaron Meyer[1]

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

[2]Department of Molecular Biosciences, Northwestern University, Evanston, IL, 60208, USA

Effective tools for exploration and analysis are needed to extract insights from large-scale single-cell measurement data. However, current techniques for handling single-cell studies performed across experimental conditions (e.g., samples, perturbations, or patients) require restrictive assumptions, lack flexibility, or do not adequately deconvolute condition-to-condition variation from cell-to-cell variation. Here, we report that the tensor decomposition method PARAFAC2 (Pf2) enables the dimensionality reduction of single-cell data across conditions. We demonstrate these benefits across two distinct contexts of single-cell RNA-sequencing (scRNA-seq) experiments of peripheral immune cells: pharmacologic drug perturbations and systemic lupus erythematosus (SLE) patient samples. By isolating relevant gene modules across cells and conditions, Pf2 enables straightforward associations of gene variation patterns across specific patients or perturbations while connecting each coordinated change to certain cells without pre-defining cell types. Thus, Pf2 provides an intuitive universal dimensionality reduction approach for multi-sample single-cell studies across diverse biological contexts.

32. The cis-regulatory logic of the type I interferon enhancer allows for tunable, stimulus-specific responses

Allison Schiffman, Zhang Cheng, Diana Ourthiague, and Alexander Hoffmann

Signaling Systems Laboratory, Department of Microbiology, Immunology, and Molecular Genetics, Institute for Quantitative and Computational Biosciences, and Molecular Biology Institute, UCLA

The expression of type I interferon (IFNβ), a critical determinant of the innate immune response, is controlled by the transcriptional activators NFκB and IRF. These activators bind to the enhancer region of IFNβ. Classic molecular biology studies of IFNβ expression suggest that NFκB and IRF function synergistically by forming an enhanceosome complex. However, this synergy is not a simple AND-gate, as knockout mouse studies have shown that dependence on NFκB is stimulus-specific. It was more recently discovered that an NFκB family member, p50:p50, competes with IRF at a Guanine-

rich binding site in the IFNβ enhancer to inhibit transcription in a stimulus-specific manner. We developed a quantitative thermodynamic state model that captures the stimulus-specific synergy between NFκB and IRF. We demonstrate that different binding equilibrium constants for the two IRF binding sites is essential for capturing the stimulus-specific dependence on NFκB. We show that the enhancer requires two neighboring proteins to be bound to have transcriptional activity and that a third bound protein does not always increase the amount of transcription from this enhancer. We also demonstrate how this regulation provides a mechanism for stimulus-specific responses to stimuli and repression by p50:p50.

33. Investigating multiple gene therapy approaches for DOCK8 Deficiency

Vrishti Sinha[1,2], Ashley Takenami[1], Joseph Long[1], Caroline Kuo[1]
[1]Department of Pediatrics
[2]GPB Genetics and Genomics

DOCK8 Deficiency is a rare, autosomal recessive primary immunodeficiency resulting from mutations in the *DOCK8* gene. *DOCK8* encodes for a Guanine Exchange Factor protein important for immune cell differentiation and actin cytoskeletal rearrangement. Roughly 1/1,000,000 individuals have this disease. Clinical manifestations include recurrent respiratory infections, eczema, and food allergies. While allogeneic hematopoietic stem cell transplantation (HSCT) is a viable therapy, there is a high chance for Graft vs Host Disease and an issue of finding a match donor. As such, our project aims to utilize autologous HSCT by correcting patient peripheral blood stem cells (PBSCs) through *ex vivo* stem cell therapy approaches. The large size of the *DOCK8* cDNA presents a challenge in donor delivery due to size limitations. Therefore, we are testing multiple donor methods, either splitting the cDNA construct which recombines at the protein level or using a vector with a higher carrying capacity for site-specific editing.

34. Cell packing frustration diversifies epigenetic alterations in epithelial monolayers

Iris Sloan[1], Jerry Chen[1], Alexandra Bermudez[1], Zoe Latham[1], and Neil Lin[1,2*]
[1]Bioengineering Department
[2]Mechanical and Aerospace Engineering Department, University of California, Los Angeles, California, USA

Intrapopulation heterogeneity invariably emerges as a single cell develops into a multicellular collective. In this study, we adopt a biophysical approach by combining thermodynamic theories and micropatterning techniques to elucidate the emergence of geometric frustration and respective histone modifications. We used the vertex model to simulate confined epithelial cells, in which the cell geometry is determined by an effective interfacial energy as a function of cell area and perimeter. As the system expands to three cells, geometric incompatibility emerges, leading to an energy gap between the ground state and the actual state (i.e., frustrated packing). We seeded Madin-Darby Canine Kidney (MDCK) cells on geometrically-confined micropatterned squares. To assess the histone modification, we stained the cells for the heterochromatin and euchromatin markers, H3K27me3 and H3K9ac, and quantified the corresponding intensities normalized to the DNA content. We demonstrated that cell heterogeneity arises from packing frustration, which is directly dictated by system size.

35. More precise heritability estimates and gene-environment interaction measurements are enabled by two novel, unbiased, sibling-based methods

Dylan Steinecke[1,2,*], Jonathan Jala[1], Daniel J. Benjamin[1,3,4], Patrick Turley[5,6]
1. Anderson School of Management, University of California Los Angeles, Los Angeles, CA 90095, USA.
2. Bioinformatics Interdepartmental Ph.D. Program, University of California, Los Angeles, CA, 90095, USA.
3. David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA.
4. National Bureau of Economic Research, Cambridge, MA 02138, USA.
5. Department of Economics, University of Southern California, Los Angeles, CA 90089, USA.
6. Center for Economic and Social Research, University of Southern California, Los Angeles, CA 90089, USA.
* Corresponding and presenting author (dylansteinecke@ucla.edu)

Insight into population health and causes of biological, medical, and behavioral phenotypes can be provided by measuring heritability—the fraction of variation in a phenotype arising from genetic variation in a population. However, heritability methods can be confounded by gene-environment

correlation, failing to measure causal effects. Here, we present two novel heritability estimators that are unbiased and substantially more precise than existing sibling-based methods. Indeed, our two methods respectively achieve standard errors up to two and twenty times smaller than the current state-of-the-art. Taken together, our new methods can also measure the role of gene-environment interactions (GxE) for a phenotype. We demonstrate this on simulated data and also apply our methods to 22,600 sibling pairs of diverse ancestries. By better estimating heritability and detecting GxE differences across social structures, these methods could bring us closer to understanding how genetics and the environment jointly give rise to an array of phenotypes.

36. Identifying 3D Genome Structure Features that Encode Gene Expression Information

Ye Wang[1,2,3], Lorenzo Boninsegna[1,2], Asli Yildirim[1,2], Xianghong Jasmine Zhou[1,3,*], Frank Alber[1,2,*]

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

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

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

This study explores how 3D genome architecture encodes gene expression information and represents a core project within the NIH 4D Nucleome consortium. By developing 3D chromatin models, the research reveals the relationship between nuclear organization and gene function, with a focus on how gene locations within the nucleus affect expression levels. Using Hi-C, Lamin B1 DamID, and SPRITE data, our team constructed 1,000 single-cell 3D genome structures for H1 human embryonic stem cells and HFFc6 human fibroblast cells. The study identifies 14 structural features that define a gene's 3D nuclear microenvironment. Results show that highly expressed genes are associated with nuclear speckles and interior nuclear regions, while low-expression genes are linked to the nuclear periphery. Changes in gene expression between cell types correlate with shifts in nuclear positioning, exemplified by the POU3F1 gene, which relocates to the nuclear periphery and is downregulated in HFFc6 cells. This research highlights the intrinsic link between gene expression and nuclear architecture.

37. Integrating adipose tissue bulk and single nucleus RNA-seq data to identify genes and cell-type level networks for insulin resistance

Zitian Wang[1,2], Seung Hyuk T. Lee[1], Asha Kar[1,2], Marcus Alvarez[1], Heather M. Stringham[3], Kyla Gelev[1], Markku Laakso[4,5], Laura Scott[3], Karen L. Mohlke[6], Päivi Pajukanta[1,2,7]

[1]Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, [2]Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA, [3]Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, MI, [4]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland, [5]Department of Medicine, Kuopio University Hospital, Kuopio, Finland, [6]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, [7]Institute for Precision Health, David Geffen School of Medicine at UCLA, Los Angeles, CA

Obesity predisposes to insulin resistance (IR) in subcutaneous adipose tissue (SAT). Through integrating SAT bulk RNA-seq data (n=335) to SAT single cell data using Scissor, we searched for insulin sensitive (IS) and IR cells across our SAT discovery (n=120) and replication (n=84) single nucleus RNA-seq (snRNA-seq) cohorts. Across both data sets, expression of the IR cells was significantly enriched for type 2 diabetes and C-reactive protein GWAS genes using scDRS. Next, we identified IS and IR - associated cell-type level SAT co-expression networks in the discovery cohort using hdWGCNA, all strongly preserved in the replication cohort. Consistently across both cohorts, the IS/IR cellular subpopulations showed significantly higher expression of the healthy/unhealthy network genes in each SAT cell-type. Together, our robustly replicated results identify genes and networks for IS and IR in SAT cell-types, thus providing new insight into system level transcriptome profiles of IS versus IR at the cell-type resolution.

38. Genetic homogeneity along the tract of the gut

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

[1]University of California, Los Angeles, Human Genetics, Los Angeles, CA, [2]University of California,

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

Environmental gradients exist throughout the digestive tract, driving spatial variation in the abundances of bacterial species along the gut. While spatial variation at the species level has been well-studied, less is known about the spatial distribution of within species genetic diversity. Within species genetic diversity in the gut exists from the co-colonization of multiple, genetically distinct strains and novel variants arising from within-host evolution. Here we ask whether the environmental gradients that drive shifts in species composition cause strains and their evolutionary modifications to be distributed heterogeneously along the gut as well. To characterize the spatial dynamics of strain colonization and within-host evolution, we analyzed five regions along the gut in eight mice inoculated with the same human stool sample. Surveying a broad panel of species, we found that strain abundances were relatively uniform along the gut in contrast to species abundances, showing that multiple, genetically diverged strains of the species can coexist within a host without spatially segregating. Similarly, evolutionary changes that arose within mice tended to sweep throughout the gut, showing little spatial specificity to particular gut segments. Together, our findings suggest that unlike at the species level, strains and their adaptations display spatial homogeneity along the gut.

39. Evolutionary dynamics of tumor development in FH-deficient renal cell carcinoma

Selina Wu[1,2], Helena K. Winata[2,3], Shinji Ohtake[4], Yash Patel[2], Randy Caliliw[4], Mao Tian[2], Sorel Fitz-Gibbon[2], Anhyo Jeong[4], Lydia Y. Liu[5], Mohammad F.E. Mootor[2], Yu Pan[6], Paul C. Boutros[2,4,7,*], Brian M. Shuch[2,4,*]

[1] Department of Medical Informatics, University of California, Los Angeles, California, USA

[2] Jonsson Comprehensive Cancer Center, University of California, Los Angeles, California, USA

[3] Department of Bioinformatics, University of California, Los Angeles, California, USA

[4] Department of Urology, University of California, Los Angeles, California, USA

[5] Department of Medical Biophysics, University of Toronto, Ontario, Canada

[6] Office of Health Informatics and Analytics, University of California, Los Angeles.

[7] Department of Human Genetics, University of California, Los Angeles, California, USA

* These authors contributed equally.

Hereditary leiomyomatosis and renal cell cancer (HLRCC) is a rare hereditary condition associated with risk of aggressive renal cell carcinoma (RCC), specifically fumarate hydratase (FH)- deficient RCC, characterized by loss of function in the *FH* gene. While RCC risk and cancer penetrance in HLRCC have been studied, the structural variant landscape of FH-deficient RCC remains largely unexplored. We performed whole genome sequencing on tumor/normal pairs from HLRCC patients to investigate the molecular architecture of FH-deficient RCC tumors. Through germline variant and somatic mutation calling, we have identified specific mutational drivers, including those at the structural variant level for the first time in HLRCC. Using tumor subclonal reconstruction, we inferred the clonality of RCC tumors and mapped the timing of key oncogenic events. These findings provide critical insights into the mutational processes and drivers underlying FH-deficient RCC, shedding light on mechanisms contributing to the lethality of this rare hereditary malignancy.

40. Stochasticity and epigenetic heritability of B-cell fate decisions are necessary for efficient affinity maturation of the antibody repertoire

Mark Xiang[1,2,3], Haripriya Vaidehi Narayanan[2,3], Alexander Hoffmann[2,3.]

[1] Bioinformatics Interdepartmental Program,

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

[3] Institute for Quantitative and Computational Biosciences,

University of California Los Angeles, Los Angeles California, USA

Antibody-mediated immunity relies on Darwinian evolution to produce a diverse repertoire of high-affinity antibodies. According to classical clonal selection theory, this involves the generation of B-cell receptor (BCR) diversity, followed by selection to determine cell fate of survival, proliferation, or differentiation into plasma cells. However, studies show substantial stochastic variability in B-cell fate

decisions, independent of BCR affinity. We developed a mathematical model to investigate how cell fate stochasticity among founder B-cells and their progeny impacts antibody affinity maturation. Our simulations reveal that while stochasticity is detrimental to a simple Darwinian process, it can enhance affinity maturation when plasma cell differentiation is included. This effect is amplified by heritability of cell fate decisions within clonal bursts, accelerating affinity maturation by leveraging the progeny at the highest BCR affinities for more affinity-enhancing mutations. These findings underscore the importance of cell fate stochasticity in antibody-mediated immunity, with potential implications for vaccine development.

41. Sexual dimorphism in renal metabolism, hemodynamics and diseases

Ivy Xiong[1,2,3], Kate Barrows[3], Lauryn Brown[3], Natalie Porat-Shliom[3], Eric J. Deeds[1,2], Alan Garfinkel[1,4]

1. Department of Integrative Biology and Physiology, 2. Institute for Quantitative and Computational Biosciences, University of California, Los Angeles, CA; 3. Cell Biology and Imaging Section, Thoracic and GI Malignancies Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD; 4. Department of Medicine (Cardiology), University of California, Los Angeles, CA

Age-related decline of renal function is faster in males than in age-matched females, manifesting in increased susceptibility to both chronic and acute kidney diseases among males, in humans and rodents. Glomerular filtration rate (GFR) is tightly regulated by tubuloglomerular feedback (TGF) within renal tubules to optimize filtration of plasma and reabsorption of essential molecules. *In vivo* measurements of GFR dynamics via intravital microscopy reveals sex-specific characteristics in TGF-mediated GFR oscillations, including frequency, amplitude and waveform. Mathematical modeling of TGF suggests that dynamic features of the system differ between male and female kidneys due to differential metabolic activities and feedback sensitivity in renal tubules, resulting in male kidneys being more prone to stress-induced damage, therefore providing an explanation for sexual dimorphism in renal aging and diseases. With this mechanistic model, our work also provides insight into how pharmacological interventions can be employed to confer renoprotection in chronic disease management.

42. scIsoSim: A Tool for Simulating Single-Cell RNA-seq Data with RNA Isoform Ground Truths

Guanao Yan[1], Weijian Wang[2], Zhiyin Liu[3], Jingyi Jessica Li[1,2]

[1] Department of Statistics and Data Science, University of California Los Angeles, Los Angeles California, USA

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

[3] Biochemistry and Cell Biology Program, Hong Kong University of Science and Technology, Hong Kong, China

Single-cell RNA sequencing (scRNA-seq) using next-generation sequencing (NGS) captures only short fragments of gene transcripts, making it challenging to accurately quantify isoform abundance and identify isoform structures. The lack of isoform-level ground truth in experimental data further necessitates simulators for benchmarking scRNA-seq computational tools. However, existing simulators cannot generate scRNA-seq reads at the isoform level. To address this, we introduce scIsoSim, the first simulator capable of producing realistic scRNA-seq reads that reflect the genetic splicing mechanism. scIsoSim supports various NGS technologies, including Smart-Seq2, 10x 3', and 5' protocols, and mimics real scRNA-seq data at both the sequencing and read count levels. It provides ground truths at two levels: isoform structure and transcript abundance at the isoform level, or alternative splicing events with corresponding transcript abundance at the exon level. Benchmarking with scIsoSim demonstrates that Kallisto achieves the highest accuracy in isoform quantification for Smart-Seq2, while BRIE2 excels in quantifying exon splicing events.

43. An Agentic System for Single Cell Bioinformatics Analysis

Yu Yan[12], Baradwaj Simha Sankar[2], Irsyad Adams[12], Wei Wang[3], Peipei Ping[12]

[1] Bioinformatics IDP, Medical Informatics HA,,

[2] Department of Physiology,

[3] Department of Computer Science,

University of California Los Angeles, Los Angeles California, USA

In recent years, single-cell RNA sequencing (scRNA-seq) has emerged as a transformative tool within the realm of cardiovascular research. However, the intricacies of scRNA-seq data analysis necessitates bioinformaticians with knowledge of advanced

computational techniques and best practices, considerable time and labor, and a sophisticated computational environment. Therefore, despite the wealth of public scRNA-seq data that holds answers to key questions about cellular states and temporospatial homogeneity and heterogeneity, the challenge of identifying and analyzing appropriate datasets to address a biomedical hypothesis remain significant. High-quality AI automation, including bioinformaticians in the loop, could greatly enhance this process. We introduce an innovative agentic system designed for single-cell bioinformatics analysis and demonstrate it with a use case in cardiovascular research. Our agentic system employs a suite of specialized agents, each dedicated to distinct roles within the bioinformatics workflow, all powered by Large Language Models (LLM). The architecture of the system includes agents responsible for: (1) Designing a bioinformatics analysis workflow that adheres to best practices, (2) retrieving relevant information from data to assist bioinformatics analysis, (3) writing, optimizing, and debugging scripts for the analysis, and (4) presenting a comprehensive report including code and statistical analysis for expert validation. This agentic system ensures a streamlined and efficient analysis pipeline, only requiring input of the biomedical question, the data, as well as metadata. To evaluate the performance and utility of our agentic system, we applied it to perform analyses on a publicly available single-cell cardiovascular dataset. The system successfully executed multiple bioinformatics analyses, demonstrating its capability to work as a senior bioinformatician with little human supervision. This innovative approach significantly expedites scRNA-seq analysis enabling cellular insights in cardiovascular biology to be efficiently and automatically extracted from public datasets.

A heartfelt thank you to our 36 faculty and 33 trainee mentors for making the 2024 B.I.G. Summer Program a resounding success!

- June 24 to August 16, 2024
- 10th Annual Program
- 430 applications received
- 67 admitted students
- 36 faculty mentors
- 33 direct research mentors





Precision Medicine (16)

Boutros Lab
Isabella Lamont
Elise Stagaman
Jacob Valenzuela
Adriana Wiggins

Meyer Lab
Salina Adhanom
Ayana Price
Ethan Hung

Chiang Lab
Alex Chen
Joy Cheng

Tward Lab
Josef Liem
Faria Tavacoli

Ophoff Lab
Alexander Everett
Beyza Duymayan

Hsu Lab
Leonard Garcia

Olde Loohuis Lab
Yael Beshaw
Lotem Efrat

Comp Genomics (18)

Geschwind Lab
Jesus Velazquez

Arboleda Lab
Samantha Scott

Ernst Lab
Gideon Shaked

DelLaTorreUbieta Lab
Jesus Mauricio
Martin Ibarra

Xiao Lab
Ryan Barney
Taylor Harris

Garud Lab
Evelyn Barajas
Alexandria Hunt

Ribas Lab
Cadence Chang

Luo Lab
Edward Chen
Thea Traw

O'Sullivan Lab
Jingtong Liang

Pellegrini Lab
Praveena Ratnavel
Lily Zello
Galen Heuer

Touma Lab
Charlotte Wolf

Yang Lab
Shunsuke Kikuchi

DOC (14)

Hu Lab
Troy Osborn
Keera Puett
Emily Nguyen

Wang Lab
Jarod Arroyo Ruiz
Anapaula Camou

Kim Lab
Ahana Goswani

Wei Lab
Louise Oh
Hassoon Sarwar

Wong Lab
Dev Trivedi
Aaron Zander

Pirih Lab
Angelina Savino
Gloria Wang

Lux Lab
Jennifer Rios-Rodriguez

Li Lab
Paul Chou

Bio Systems (19)

Bouhaddou Lab
Ricardo Roure
Matias Lee

Chou Lab
Henry Li

Hoffmann Lab
Chengyuan Li
David Mastro
Aryahi Deorukhkar

Knowles Lab
Naomi Barber-Choi
Sarah Bonver
Kyle Kalindjian
Jacob Kelman
Madelaine Leitman
Carly Rabun
Jacob Fisher

Wilson Lab
Lydia Longritz
Jiale Yang

DiPoppa Lab
Kate Jackson
Marcello Berger

Zhou Lab
Ethan Crofut

Venugopal Lab
Jenna Jabourian

2024

Talks, Posters and Abstracts : https://qcb.ucla.edu/big-summer/big2024/