8th Annual QCBio Retreat

Friday, September 16, 2022

9:30am – 6:30pm

Boyer Hall and Court of Sciences

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, more quantitative, more predictive insights and capabilities.

The Biosciences revolution is continuing full throttle. Our 2020 survey revealed that at UCLA 30% of the Biosciences research personnel is entirely dry lab, and also experimentalists are spending more and more of their time on computational data analysis, database queries and modeling.

UCLA provides numerus graduate training research opportunities in quantitative and computational biosciences (https://qcb.ucla.edu/overview/). Four graduate Programs in Bioinformatics, Medical Informatics, Biomathematics and Genetics Genomics are already coordinated and efforts are under way to add a Systems Biology Home Area, which applicants, faculty, and external reviewers are asking for. The UCLA Graduate Programs Navigator provides an overview and guides prospective applicants (https://qcb.ucla.edu/graduate-program-navigator/).

The QCB Collaboratory (https://qcb.ucla.edu/collaboratory/) is a postdoctoral training program that provides computational biology postdocs with opportunities for extending their teaching and collabroatiion skills and thus their academic impact. In turn, Collaboratory Fellows have provided essential training in a broad range of bioinformatics 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, undergraduates are increasingly flocking to the expanding Bioinformatics Minor (https://bioinformatics.ucla.edu/undergraduate-bioinformatics/) and the Computational and Systems Biology Major (https://casb.ucla.edu/). A majority of these talented undergraduates are involved in research. We have expanded the Undergraduate Research Portal (https://qcb.ucla.edu/research-portal/) to connect potential mentors with eager undergraduate students – an amazing human resource for driving research forward and future leaders.

Mentoring is therefore a key component of our QCB community culture. Now in its 8th year we hosted 73 students within the B.I.G. Summer Undergraduate Research Program this year with 34 laboratories participating! (https://qcb.ucla.edu/big-summer/big2022/) 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.

The 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. QCB is here for you!

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

Agenda

COFFEE, TEA & BAGELS

9:50 a.m. **WELCOME**

10:00 a.m. **STATUS REPORTS I**

- **Alexander Hoffmann**, Director, QCBio, BIG Summer
- **Daniel Geschwind,** Director, Institute of Precision Health
- **Matteo Pellegrini**, Director, QCBio Collaboratory
- **Eleazar Eskin**, Chair of Computational Medicine

10:25 a.m. **SESSION I**

KEYNOTE I

- **Wei Wang,** Leonard Kleinrock Professor in Computer Science; Director, Scalable Analytics Institute, UCLA

10:50 a.m. **SELECTED TALKS**

- **Kofi Amoah**, Bioinformatics PhD student, Xiao lab
- **Xiaolu Guo**, Postdoc, Hoffmann lab

11:20 a.m. **KEYNOTE II**

- **Quanquan Gu**, Associate Professor in Computer Science, UCLA

11:55 a.m. **STATUS REPORTS II**

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

12:20 p.m. **LUNCH - POSTER SESSION I**

1:45 p.m. **SESSION II**

KEYNOTE III

- **Safiya Noble**, Acting Director of Data X; Professor, Dept of Gender Studies and Dept. of African American Studies, UCLA

2:10 p.m. **SELECTED TALKS**

- **Christa Caggiano,** Bioinformatics PhD student, Zaitlen lab
- **Lorenzo Boninsegna**, Postdoc, Alber lab

2:40 p.m. **STATUS REPORTS III**

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

3:00 p.m. **COFFEE & TEA BREAK**

3:20 p.m. **SESSION III**

KEYNOTE IV

- **Deanna Needell,** Executive Director, Institute for Digital Research & Education; Dunn Family Endowed Chair in Data Theory; Professor, Department of Mathematics, UCLA

3:45 p.m. **SELECTED TALKS**

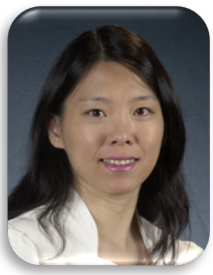
- **Russell Littman**, Bioinformatics PhD student, Yang lab
- **Yu Yan**, Medical Informatics PhD student, Ping lab

4:15 p.m. **QBIO-EDGE PRESENTATION (Empowering Diversity and Growth in Education)**

CONCLUDING REMARKS

4:30 p.m. **RECEPTION & REFRESHMENTS - POSTER SESSION II**

Keynote Speakers



Wei Wang
Leonard Kleinrock Professor in Computer Science
Director, Scalable Analytics Institute (ScAi)
University of California, Los Angeles

Wei Wang is the Leonard Kleinrock Chair Professor in Computer Science and Computational Medicine at University of California, Los Angeles and the director of the Scalable Analytics Institute (ScAi). She is also a member of the UCLA Jonsson Comprehensive Cancer Center, Institute for Quantitative and Computational Biology, and Bioinformatics Interdepartmental Graduate Program. She received her PhD degree in Computer Science from the University of California, Los Angeles in 1999. Dr. Wang's research interests include big data analytics, data mining, machine learning, natural language processing, bioinformatics and computational biology, and computational medicine. Dr. Wang received numerous awards in her career including an ACM fellow. She is the chair of ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD).

TITLE*: Knowledge Graph Representation Learning Enabling Biomedical Applications*

Quanquan Gu
Associate Professor, Department of Computer Science, UCLA

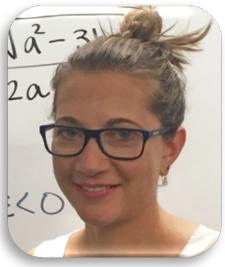
Quanquan Gu is an Associate Professor of Computer Science at UCLA. His research is in the area of artificial intelligence and machine learning, with a focus on developing and analyzing nonconvex optimization algorithms for machine learning to understand large-scale, dynamic, complex, and heterogeneous data and building the theoretical foundations of deep learning and reinforcement learning. He received his Ph.D. degree in Computer Science from the University of Illinois at Urbana-Champaign in 2014. He is a recipient of the Alfred P. Sloan Research Fellowship, NSF CAREER Award, Simons Berkeley Research Fellowship among other industrial research awards. He also serves as an associate/section editor for Journal of Artificial Intelligence and PLOS One, area chair for ICML, NeurIPS, ICLR, AISTATS and AAAI, and senior program committee member for IJCAI.

TITLE: *Self-supervised Hypergraph Convolutional Neural Networks for Biomedical Data Analytics*

Safiya Noble
Acting Director of Data X
Professor, Dept. of Gender Studies and Dept. of African American Studies, UCLA

Dr. Safiya U. Noble is an internet studies scholar and Professor of Gender Studies and African American Studies at the University of California, Los Angeles (UCLA). She is the Founder and Director of the newly launched National Center on Race and Digital Justice. She is the Interim Director of the DataX initiative at UCLA, Co-Founder and former Faculty Director of the UCLA Center for Critical Internet Inquiry (C2i2), and the Co-Director of the Minderoo Initiative on Tech & Power.

TITLE: *Taking on Big Tech: New Paradigms for New Possibilities*
The landscape of information is rapidly shifting as new demands are increasing investment in digital technologies. Yet, critical scholars continue to demonstrate how many technologies are shaped by and infused with values that are not impartial, disembodied, or lacking positionality. Technologies hold racial, gender, and class politics. In this talk, Dr. Safiya Noble from the UCLA Center for Critical Internet Inquiry will discuss new insights stemming from her recent book, Algorithms of Oppression, and posit emerging work that explores the impact of commercial technologies on the public



Deanna Needell
Executive Director, Institute for Digital Research & Education
Dunn Family Endowed Chair in Data Theory
Professor, Department of Mathematics, UCLA

Deanna Needell earned her PhD from UC Davis before working as a postdoctoral fellow at Stanford University. She is currently a full professor of mathematics at UCLA, the Dunn Family Endowed Chair in Data Theory, and the Executive Director for UCLA's Institute for Digital Research and Education. She has earned many awards including the Alfred P. Sloan fellowship, an NSF CAREER and other awards, the IMA prize in Applied Mathematics, and is a 2022 American Mathematical Society (AMS) Fellow. She has been a research professor fellow at several top research institutes including the Mathematical Sciences Research Institute and Simons Institute in Berkeley. She also serves as associate editor for IEEE Signal Processing Letters, Linear Algebra and its Applications, the SIAM Journal on Imaging Sciences, and Transactions in Mathematics and its Applications as well as on the organizing committee for SIAM sessions and the Association for Women in Mathematics.

TITLE*: Using Algebraic Factorizations for Interpretable Learning*
Non-negative Matrix Factorization (NMF) is a fundamental tool for dictionary learning problems, giving an approximate representation of complex data sets in terms of a reduced number of extracted features. In this talk, we will introduce the main concept of NMF, its implementation, and its online and streaming variations. We will showcase how mathematical tools like this can be used for interpretable learning tasks. These applications range from imaging and medicine to forecasting and collaborative filtering. Discussion and questions are welcome.

Selected Talks

- **Massively parallel screen uncovers many rare 3' UTR variants regulating mRNA abundance of cancer-relevant genes**

Kofi Amoah[2,3*], Ting Fu[1,3*], Tracey W. Chan[2,3], Jae-Hoon Bahn[3], Jae-Hyung Lee[3], Sari Terrazas[3,4], Rocky Cheung[5], Sriram Kosuri[5], Xinshu Xiao[1,2,3,4]
[1] Molecular, Cellular and Integrative Physiology Interdepartmental Program
[2] Bioinformatics Interdepartmental Program
[3] Department of Integrative Biology and Physiology
[4] Molecular Biology Interdepartmental Doctoral Program
[5] Department of Chemistry and Biochemistry
University of California, Los Angeles, Los Angeles, CA 90095, USA
*These authors contributed equally

Elucidating the functional impact of rare genetic variants, especially those in non-coding regions, represents a significant challenge. Here, we developed a massively parallel screen for rare 3' UTR variants (MapUTR) that affect mRNA abundance post-transcriptionally. Using two human cell lines, we assayed the function of 14,575 rare variants and found that 5,437 (37%) led to significant alterations of mRNA abundance in at least one cell line. These variants are enriched in miRNA target sites and binding sites of RNA-binding proteins, attesting to their functional relevance. Importantly, 71% of these variants are located in cancer-related genes. Further, 37 variants are associated with expression outliers in TCGA. Through prime editing, we characterized three variants in cancer-associated genes (*MFN2*, *FOSL2*, and *IRAK1*), confirming their impacts on mRNA stability. Importantly, all three variants significantly altered cellular proliferation, illustrating the effectiveness of MapUTR in pinpointing functional genetic variants.

- **Modeling the heterogenous NFκB dynamics of single immune cells**

Xiaolu Guo, Xiaofei Lin, Adewunmi Adelaja, Alexander Hoffmann
Institute for Quantitative and Computational Biosciences, and Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, USA

Macrophages function as immune sentinel cells, initiating appropriate and specialized immune responses to a great variety of pathogens. The transcription factor NFκB controls macrophage gene expression responses, and its temporal dynamic enables stimulus-specificity of these responses. Using a fluorescent reporter mouse our laboratory recently generated large amounts of single-cell NFκB dynamic data and identified dynamic features, termed 'signaling codons', that convey information to the nucleus about stimulus ligand and dose. Here, we aimed to recapitulate the stimulus-specific but highly cell-to-cell heterogeneous NFκB dynamics with a mathematical model of the signaling network. A parameter scan determined the potential parameters accounting for the heterogeneity. We estimated parameter distributions using the Stochastic Approximation Expectation Maximization (SAEM) approach and then fit the individual cell data using Bayesian maximum a posteriori (MAP) estimation. Visual inspection revealed an excellent fit with the data. To quantitatively evaluate the fitting performance, we compared the experimental and predicted distributions of NFκB signaling codons. Further, we identified biochemical reactions that may account for the cellular heterogeneity in NFκB dynamics. Our results establish a mathematical modeling tool that may be used to study the molecular determinants of information flow and dynamical coding in immune sentinel cells.
Key words: NFκB dynamic, cell heterogeneity, signaling network modeling, parameter estimation

- **A non-invasive biomarker candidate for amyotrophic lateral sclerosis**

Christa Caggiano[1,2], M. Morselli[3], B. Celona[4], R. Henderson[5], C. Lomen-Hoerth[6], F. Garton[7], M. Pellegrini[3], N. Zaitlen[2]
[1]Interdepartmental Program in Bioinformatics, UCLA
[2]Department of Neurology, UCLA
[3]Department of Molecular, Cell and Developmental Biology, UCLA
[4]Cardiovascular Research Institute, UCSF
[5]Department of Neurology, Royal Brisbane and Women's Hospital, Brisbane
[6]Department of Neurology, UCSF
[7]Institute for Molecular Biosciences, University of Queensland

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease with no clinical biomarker. In this work, we propose using cell-free DNA (cfDNA) as a non-invasive biomarker for ALS. CfDNA is enriched in ALS patients and can be used to infer the rate of cell death. We developed a DNA methylation capture panel technology, which captures disease and tissue informative CpG sites. Importantly, the panel has a projected cost of under $300 per sample, which can facilitate its use in large clinical cohorts. We tested 48 ALS cases and 48 controls and performed next-generation sequencing. A supervised machine learning model was developed to predict disease status. Through this approach, we found that the model could accurately predict ALS case-control status (cross-validated AUC: 0.95), which may reach the level of clinical utility. Together, these results suggest that cfDNA may be a valuable predictive biomarker candidate in ALS.

- **Reconstructing single cell whole genome structures from available experimental data**

Lorenzo Boninsegna[1], Asli Yildirim[1], Yuxiang Zhan[1], Frank Alber[1]
[1] Institute for Quantitative and Computational Biosciences (QCBio) and Department of Microbiology, Immunology and Molecular Genetics, University of California Los Angeles, Los Angeles (CA), USA

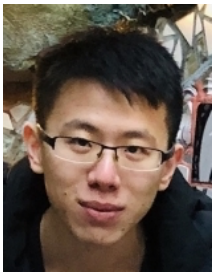
The availability of complementary data from different state-of-the-art technologies has revolutionized our view of 4D genomes and paved the way for data-driven approaches into investigating how genome architecture relates to its function. Here, we present a population-based modeling scheme that integrates heterogeneous information from genomics (e.g., ensemble Hi-C and lamina DamID) and imaging data modalities (e.g., high resolution FISH data) into simulating whole genome structures at a single-cell level. These structures are highly predictive for nuclear locations of genes and nuclear bodies and spatial segregation of functionally related chromatin, and they capture a similar heterogeneity to that observed in super-resolution imaging experiments. Our modeling platform is unique in that it can naturally accommodate single-cell data sources, such as shapes and locations of the nuclear bodies from segmented microscopy images and locations of selected chromatin regions from chromatin tracing experiments. We present a quantitative assessment of models from different combinations of data and show that multimodal data integration greatly increases model accuracy and coverage, and compensates for systematic biases in the input data sources. Our study also uncovers the key contributions of low-frequency inter-chromosomal contacts in accurately predicting the global nuclear architecture.

- **SCING: Single Cell INtegrative Gene regulatory network inference elucidates robust, interpretable gene regulatory networks**

Russell Littman[1,2], Ning Wang[1] & Xia Yang[1,2,3,4,5*]
[1]Department of Integrative Biology & Physiology, UCLA, Los Angeles, CA, USA
[2]Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA, USA
[3]Institute for Quantitative and Computational Biosciences (QCBio), Los Angeles, CA, USA
[4]Molecular Biology Institute (MBI), Los Angeles, CA, USA
[5]Brain Research Institute (BRI), Los Angeles, CA, USA
*Corresponding author: xyang123@ucla.edu

Gene regulatory network (GRN) inference is an integral part of understanding physiology and disease. Single cell/nuclei RNAseq (scRNAseq/snRNAseq) data has been used to elucidate cell-type GRNs; however, the accuracy and speed of current scRNAseq-based GRN approaches are suboptimal. Here, we present Single Cell INtegrative Gene regulatory network inference (SCING), a gradient boosting and mutual information based approach for identifying robust GRNs from scRNAseq, snRNAseq and spatial transcriptomics data. Performance evaluation using held-out data, Perturb-seq datasets, and the mouse cell atlas combined with the DisGeNET database demonstrates the improved accuracy and biological interpretability of SCING compared to existing methods. We applied SCING to the entire mouse single cell atlas, human Alzheimer's disease (AD), and mouse AD spatial transcriptomics. SCING GRNs reveal unique disease subnetwork modeling capabilities, have intrinsic capacity to correct for batch effects, retrieve disease relevant genes and pathways, and are informative on spatial specificity of disease pathogenesis.

- **MIND-S: a deep learning prediction model for elucidating protein PTMs in human diseases.**



Yu Yan[1,2], Jyun-Yu Jiang[3], Mingzhou Fu[1], Ding Wang[2], Alexander Pelletier[2,3], Dibakar Sigdel[2], Wei Wang[1,3], Peipei Ping[1,2,3]
[1] Medical Informatics, University of California at Los Angeles (UCLA), CA 90095, USA
[2] Department of Physiology, UCLA School of Medicine
[3] Scalable Analytics Institute (ScAi) at Department of Computer Science, UCLA School of Engineering,

Post-translational modification (PTM) of proteins plays a fundamental regulatory role in numerous biological processes; determining the specific amino acid residues modified is essential to understanding PTM-governed biological processes and disease pathogenesis. While significant efforts have been devoted to PTM site prediction, tools for multiple PTM predictions with accurate, efficient, and interpretable functionalities remain scarce. We report here a novel deep learning supported method, MIND-S, for PTM predictions. MIND-S utilizes the sequence and structure information of proteins, with techniques of multilabel and bootstrap approaches to achieve high performance with computational efficiency. MIND-S features interpretability through evaluating the saliency for the prediction of each residual. To demonstrate the practical usage of MIND-S, we show use cases of MIND-S in the cardiovascular research setting. Furthermore, with genetic information, MIND-S can annotate single nucleotide polymorphisms (SNPs) from a PTM perspective. We detail use cases where MIND is demonstrated as an accurate, interpretable, and efficient tool to understand biological processes in health and diseases.

Collaboratory Fellows 2022-2023

Haripriya Vaidehi Narayanan

Matteo Pellegrini, Director

Arjun Bhattacharya

Fangming Xie

Daniel Ha

Kelsey Jorgensen

Giorgia Del Vecchio

Karolina Kaczor-Urbanowicz

Shawn Cokus

Lukasz Salwinski

Seyoon Ko

Wenbin Guo

Don Vaughn

Weihong Yan

Eloy Lopez, Program Manager

Sergey Knyazev

Giovanni Quinones Valdez

Xinjun Zhang

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

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

http://qcb.ucla.edu/collaboratory

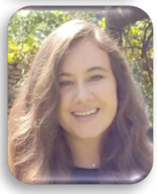
Welcome new Faculty!

Jeffrey Chiang, Ph.D.
Assistant Adjunct Professor
Dr. Jeff Chiang is interested in translating recent advances in big data and artificial intelligence to active clinical research, and his work addresses the computational challenges which arise when doing so. He works closely with clinical departments to identify risk factors and develop predictive models for negative outcomes such as age-related macular degeneration. On the way, Chiang and his team develop techniques which overcome limited data availability, combine and leverage health information from disparate sources, and are as free from bias as possible.
Dr. Chiang also leads the Computational Medicine Technology Core, which is involved in building the technical infrastructure for these computational models to be deployed in the clinic. Chiang obtained his B.S., M.A., and PhD in psychology (cognitive science) at UCLA. He then held research positions in industry and the Department of Computational Medicine prior to joining its faculty.

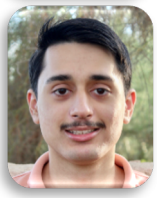
Welcome our Incoming Bioinformatics Students!

Akintunde (Tunde) Akinkuolie
MB & BS, University of Ibadan
MPH, Boston University

Jerome Freudenberg
BS, University of Chicago



Clara Frydman
BS, Technion-Israel Institute of Technology



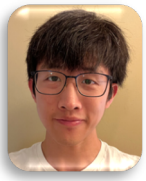
Joseph Galasso
BS, University of Dallas
B.I.G. SUMMER 2020 ALUMNUS

Greta Gerdes
BS, University of St. Thomas

Lena Krockenberger
BS, UC San Diego

Xinzhe Li
BS, UC Davis

Swetha Ramesh
BS, UC Bekerley



Jeremy Wang
BS, Brown University
B.I.G. SUMMER 2021 ALUMNUS

Zitian Wang
BS, UCLA

Yuxing Zhou
BS, Zhejiang University
BS, University of Edinburgh

Welcome our Incoming Medical Informatics Students!

Jeffrey Feng
BAS, University of Waterloo

Thai Tran
BS, UC Riverside

David Gibson
BS, UC Davis
M, UC San Francisco

Vedrana Ivezic
BA, Princeton University

Luoting (Lottie) Zhuang
BS, UCLA
MS, Harvard Medical School

Welcome our Incoming Genetics & Genomics Students!

Cassidy Andrasz
Cal Poly San Luis Obispo

Juliana Shin
Northwestern University

Kyla Gelev
USC

Maris Kamalu
Pomona College

Vrishti Sinha
UC Davis

Welcome our Incoming Biomathematics Students!

Santiago Cardenas
BS, The College of New Jersey

Yunbei Pan
BS, Liverpool University

Alexandra (Sasha) Schtein
BS, UCLA

Poster Session I – (Lunch/12:20pm)

1. **Dissecting environmental influences on transcriptional cellular state heterogeneity in glioma**

Nicholas A. Bayley[1,2], Christopher Tse[1], Henan Zhu[1], Jenna Minami[1], Weihong Yan[3], Aparna Bhaduri[4], Thomas G. Graeber[1,5*], and David A. Nathanson[1,5*]
[1]Department of Molecular and Medical Pharmacology,
[2]Bioinformatics Interdepartmental Program,
[3]Department of Chemistry and Biochemistry,
[4]Department of Biological Chemistry,
[5]Jonsson Comprehensive Cancer Center,
University of California Los Angeles, Los Angeles California, USA
* equal contribution from labs

Transcriptomic characterizations of cancer have revealed plastic cellular state heterogeneity. Glioma cells are said to reflect neurodevelopmental cell types, while aggressive glioblastomas include a Mesenchymal-like state associated with increased interactions with immune cells. Yet the gamut of environmental factors promoting state heterogeneity remains largely unexplored. Here we performed bulk and single-cell transcriptomic profiling of a diverse library of patient gliomas and derivative models established in immunodeficient orthotopic mouse xenografts and gliomasphere cultures. Data processing included estimation of normal cell contamination and removal of sequencing reads originating from mouse cells. Reciprocal PCA-based single-cell dataset integration identified nine cellular states with clinical and environmental specificity. CIBERSORTx-based bulk deconvolution was evaluated using matched bulk and single-cell sequencing. This highlighted limitations of deconvolution applied to admixed tumor samples and necessitated further modifications to our approach. Our results suggest that state heterogeneity is the combined product of varied microenvironmental stimuli and intrinsically constrained state plasticity.

2. **Reconstructing single cell whole genome structures from available experimental data**

Lorenzo Boninsegna[1], Asli Yildirim[1], Yuxiang Zhan[1], Frank Alber[1]
[1] Institute for Quantitative and Computational Biosciences (QCBio) and Department of Microbiology, Immunology and Molecular Genetics, University of California Los Angeles, Los Angeles (CA), USA

The availability of complementary data from different state-of-the-art technologies has revolutionized our view of 4D genomes and paved the way for data-driven approaches into investigating how genome architecture relates to its function. Here, we present a population-based modeling scheme that integrates heterogeneous information from genomics (e.g., ensemble Hi-C and lamina DamID) and imaging data modalities (e.g., high resolution FISH data) into simulating whole genome structures at a single-cell level. These structures are highly predictive for nuclear locations of genes and nuclear bodies and spatial segregation of functionally related chromatin, and they capture a similar heterogeneity to that observed in super-resolution imaging experiments. Our modeling platform is unique in that it can naturally accommodate single-cell data sources, such as shapes and locations of the nuclear bodies from segmented microscopy images and locations of selected chromatin regions from chromatin tracing experiments. We present a quantitative assessment of models from different combinations of data and show that multimodal data integration greatly increases model accuracy and coverage, and compensates for systematic biases in the input data sources. Our study also uncovers the key contributions of low-frequency inter-chromosomal contacts in accurately predicting the global nuclear architecture.

3. **Inferring HIV Transmission Patterns from Viral Deep-Sequence Data via Latent Spatial Poisson Processes**

Fan Bu[1], Oliver Ratmann[2], and Jason Xu[3]
[1] Department of Human Genetics, University of California – Los Angeles
[2] Department of Mathematics, Imperial College London
[3] Department of Statistical Science, Duke University

Viral deep-sequencing technologies play a crucial role toward understanding disease transmission network flows, because the higher resolution of these data compared to standard Sanger sequencing provide evidence into the direction of infectious disease transmission. To more fully utilize these rich data and account for the uncertainties in phylogenetic analysis outcomes, we propose a spatial Poisson process model to uncover HIV transmission flow patterns at the population level. We represent pairings of two individuals with viral sequence data as typed points, with coordinates representing covariates such as sex and age, and the point type representing the unobserved transmission statuses (linkage and direction). Points are associated with observed scores on the strength of evidence for each transmission status that are obtained through standard deep-sequenece phylogenetic analysis. Our method is able to jointly infer the latent transmission statues for all pairings and the transmission flow surface on the source-recipient covariate space. In contrast to existing methods, our framework does not require pre-classification of the transmission statuses of data points, instead learning them probabilistically through a fully Bayesian inference scheme. By directly modeling continuous spatial processes with smooth densities, our method enjoys significant computational advantages compared to previous methods that rely on discretization of the covariate space. We demonstrate that our framework can capture age structures in HIV transmission at high resolution, and bring valuable insights in a case study on viral deep-sequencing data from Rakai, Uganda.

4. **Asymmetric Branching Scale Factors as Features in Neuronal and Glial Cell-Type Classification**

Paheli Desai-Chowdhry[1], Alexander Brummer[2], Samhita Mallavarapu[3], Van Savage[1,3,4,5]
[1] Department of Computational Medicine, University of California Los Angeles (UCLA), Los Angeles, California (CA), United States of America (USA),
[2] Department of Mathematical Oncology, City of Hope, Duarte, CA, USA,
[3] Department of Computational and Systems Biology, UCLA, Los Angeles, CA, USA,
[4] Department of Ecology and Evolutionary Biology, UCLA, Los Angeles, CA, USA,
[5]Santa Fe Institute, Santa Fe, New Mexico, USA

Neurons are connected by complex branching processes - axons and dendrites - that process information for organisms to respond to their environment. Classifying neurons according to differences in structure or function is a fundamental piece of neuroscience. Here, we develop a novel unifying biophysical model that establishes a correspondence between neuron structure and function as mediated by the tradeoff of functional principles such as conduction time delay, energy efficiency, material costs, and space-filling. Moreover, from this model, we derive functionally relevant structural parameters that we use as features in machine learning classification methods to distinguish between different cell types. We find significant distinctions in the asymmetric scaling ratios between Purkinje cells and motoneurons and between axons and microglia, a specific class of electrically active non-neuronal brain cells. The performance of these classification methods gives us important insights into the correspondence between structure and function across different cell types.

5. **Genetic adaptations to potato starch digestion in the Peruvian Andes**

Kelsey Jorgensen[1], Obed A. Garcia[2], Melisa Kiyamu[3], Tom D. Brutsaert[4], Abigail W. Bigham[1]
[1]Department of Anthropology, University of California, Los Angeles,
[2]Department of Biomedical Data Science, Stanford University,
[3]Departamento de Ciencias Biológicas y Fisiológicas, Universidad Peruana Cayetano Heredia, [4]Department of Exercise Science, Syracuse University

Potatoes are an important staple crop cultivated as early as 10,000 years ago in the Andes. Ancient Andean populations that relied upon this high-

starch food to survive could possess genetic adaptation(s) to digest potato starch more efficiently. We applied several statistical tests to detect signatures of natural selection in genes associated with starch-digestion, *AMY1*, *AMY2*, *SI*, and *MGAM,* in order to identify whether this putative adaptation is still present in their modern-day Peruvian descendants. Results identified a regional haplotype in *MGAM* that is unique to Peruvians, and within a region of high transcriptional activity associated with the REST protein. The age of this haplotype is estimated to be around 9,938 years old, suggesting natural selection favored variants that led to more efficient digestion around the time of increased potato cultivation. This research demonstrates human dietary shifts can be a major driver of global variation in digestion ability and evolutionary change.

6. **Differentiating sex chromosomal effect from gonadal effect on sex-biased gene regulation in hypothalamic single cells and genetic association with metabolic diseases**

Gaoyan Li[1], In Sook Ahn[1], Xuqi Chen[1], Graciel Diamante[1], Douglas Arneson[1,2,3], Art Arnold[1], Xia Yang[1,2,3]
[1.] Department of Integrative Biology and Physiology, UCLA;
[2.] Bioinformatics Interdepartmental Program, UCLA;
[3.] Institute for Quantitative and Computational Biosciences (QCBio), UCLA

Hypothalamus, the central energy homeostasis regulator linking neuronal, endocrine, and metabolic systems, likely involved in determining sex differences in metabolic diseases. However, the contribution of gonadal sex effect (GSE) vs sex chromosomal effect (SCE) to hypothalamus remains poorly understood. Using single-cell-RNA-sequencing, we differentiate GSE from SCE in individual hypothalamic cell types and neuronal subtypes in the four-core-genotype model, which has XX mice with testes or ovaries as well as XY mice with testes or ovaries. We identified numerous sex-biased-genes affected by GSE or SCE or both in individual hypothalamic cell types. We further investigated the association of the sex-biased-genes with metabolic diseases based on enrichment patterns of hypothalamus-specific eQTL and sQTL associated with over 30 GWAS for metabolic diseases and phenotypes with known sex differences. Our results point to the role of specific hypothalamic cell types, their sex-biased genes and pathways, and the specific underlying sex factors in metabolic diseases.

7. **Stochastic dynamics and ribosome-RNAP interactions in Transcription-Translation Coupling**

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

Transcription-translation coupling (TTC) in prokaryotes is thought to control the timing of protein production relative to transcript formation. The marker for such coupling has typically been the measured time delay between the first completion of transcript and protein. We formulate a stochastic model for ribosome and RNAP elongation that also includes RNAP pausing and ribosome-RNAP binding. The model is able to predict how these processes control the distribution of delay times and the level of protection against premature termination. We find relative speed conditions under which ribosome-RNAP interactions can accelerate or decelerate transcription. Our analysis provides insight on the viability of potential TTC mechanisms under different conditions and suggests measurements that may be potentially informative.

8. **Flexible birth-death tree models with Markov random fields**

Andrew Magee[1], Sebastian Höhna[2], Adam Leaché[3], and Vladimir Minin[4]
[1]Department of Human Genetics, David Geffen School of Medicine at UCLA
[2]Division of Evolutionary Biology, LMU Munich
[3]University of Washington, Seattle
[4]Department of Statistics, UCI

Studying variation in rates of speciation enables researchers to examine the patterns and processes that shape the diversity of life on Earth. Similarly, studying variation in the rate of accumulation of new cases gives researchers the ability to understand the dynamics of the spread of infectious diseases. Birth-death process models give biologists a model-based framework in which both macroevolutionary and epidemiological questions can be addressed. We build two flexible birth-death process models that allow the birth rate to vary through time, with no need for an explicit choice of functional form. Through simulations, we show that one of these models, the Horseshoe Markov random field birth-death process, performs well in nonparametric estimation of time-varying birth rates.

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

Brian Orcutt-Jahns[1], Andrei Rodin[2], Joao Rodrigues Lima Junior[2], Peter Lee[2], Aaron Meyer[1,3,4,5]
[1]Department of Bioengineering, University of California, Los Angeles
[2]City of Hope, Duarte, CA, United Sates of America
[3]Department of Bioinformatics, University of California, Los Angeles, United States of America
[4]Jonsson Comprehensive Cancer Center, University of California, Los Angeles, United States of America
[5]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, United States of America

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

10. **Recapitulating Trends on IL-2 Mutein Responses Via GMMs & Tensor Decomposition**

Andrew Ramirez[1], Brian Orcutt-Jahns[1], Aaron Meyer[1]
[1]Department of Bioengineering, University of California, Los Angeles, CA 90024, USA

Cytokines are cell signaling proteins with crucial roles in immune system cellular communication. Interleukin-2 (IL-2) is one such cytokine of the common γ-chain receptor family that modulates immune activity and enacts pleiotropic regulation of both the adaptive and innate immune responses through proliferation of both regulatory and effector immune cells. Immune cells are extremely heterogeneous along with their responses, thus making it difficult to interpret singe cell analysis methods and uncover trends between cell types. Peripheral blood mononuclear cells (PBMCs) were profiled with IL-2 muteins at 12 doses, for single cells at 4 unique time points. In order to recapitulate known trends, variation in the response signals, and identify trends between markers, we propose a method with the basis of a combined form of Gaussian Mixture Model and Nonnegative Matrix Factorization. Our model serves as a general method for single cell experiments to find patterns across conditions..

11. **Utilizing PathFX to Analyze Drug-Gene Associations**

Anjali Sivanandan, Jennifer Wilson
Department of Bioengineering
University of California Los Angeles, Los Angeles California, USA

Protein-Protein Interaction network methods, like PathFX, are an increasingly popular way to predict drug downstream effects. However, these algorithms often predict more drug effects than evidence supports. These predictions can be validated through observational studies in the Electronic Health Record that test groups of drugs based on shared gene pathways. This study will focus on the disease areas of diabetes and lung cancer to illustrate how PathFX can be used to identify hypotheses for shared pathways. We analyzed PathFX networks for 44 and 34 drugs used to treat diabetes and lung cancer, respectively. We

used downstream proteins to cluster the drugs and found network clusters are distinct from ATC groups. We used GO enrichment to discover functions associated with network clusters and found distinct functional categories for each disease. We hypothesize that we will be able to distinguish clinical and non-clinical drugs by their downstream pathways and provide a means to reduce PathFX over-prediction.

12. Interactions between Valley Fever and genetic ancestry: the effect of genetic ancestry on risk of developing disseminated coccidioidomycosis

Sarah J. Spendlove[1,2], Samantha L. Jensen[2,3], Diego Orellana[2], Kangcheng Hou[1,2], Alexis V. Stephens[4], George R. Thompson[5,6,7], Royce H. Johnson[8,9], Arash Heidari[8,9], Rasha Kuran[8,9], Bogdan Pasaniuc[2,3,10], Manish J. Butte[3,4], and Valerie A. Arboleda[1,2,3,10]
[1] Interdepartmental Bioinformatics Program, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA
[2] Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA
[3] Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA
[4] Division of Immunology, Allergy, and Rheumatology, Department of Pediatrics, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA
[5] UC Davis Center for Valley Fever, UC Davis, Davis, CA, USA
[6] Department of Medical Microbiology and Immunology, UC Davis, Davis, CA, USA
[7] Department of Medicine and Division of Infectious Diseases, UC Davis, Davis, CA, USA
[8] Department of Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA
[9] Valley Fever Institute, Kern Medical, Bakersfield, CA, USA
[10] Department of Computational Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA

Coccidioidomycosis is a fungal infection endemic to the American Southwest. While most patients are asymptomatic after infection or develop only a respiratory infection, a small proportion of individuals develop severe disseminated coccidioidomycosis (DCM), which is associated with significant morbidity and mortality and requires prolonged antifungal treatment. Previous studies have shown that self-identified race and ethnicity (SIRE) is associated with risk of DCM, including a higher risk of DCM in individuals identified as African-American, Latino or Filipino. These used race and ethnicity, which is a social construct and not always equivalent to genetic ancestry, which is based on where individuals' genetic data cluster with reference populations. In this study we use a merged cohort of 479 individuals from the UC Davis Center for Valley Fever and 87 individuals from the Valley Fever Institute at Kern Medical (VFI) with 1000 genomes reference individuals. We calculated genetic ancestry through PCA analysis and also through global ancestry analysis. The patients have differing severity of coccidioidomycosis (either uncomplicated (UVF), disseminated (DCM), or other) and we show that having 20% or more of one's genetic ancestry clustering with the genetics of individuals from the African continent is associated with a 16.3 odds of DCM. Meanwhile having 20% or more genetic ancestry clustering with the genetics of individuals from Europe gives a decreased odds of 0.2. Next we used GWAS and admixture mapping methods to look for areas of the genome that may be transmitting this risk in some individuals. Finally, we imputed the 479 exomes using a 1000 genomes reference panel and then merged these 479 imputed genomes from UC Davis with the 87 genomes from VFI in order to run iLASH to detect identity-by-descent (IBD) segments that may be related to DCM risk. This project will allow us to leverage genetic biomarkers to identify individuals at the highest risk for DCM and who would benefit from early-treatment with immunomodulatory therapies.

13. A multivalent binding model predicts the binding and *in vivo* effector cell responses of antibody subclass mixtures

Cyrillus Tan[1], Anja Lux[2], Markus Biburger[2], Stephen Lees[3], Falk Nimmerjahn[2], Aaron S. Meyer[3]
[1]Bioinformatics Interdepartmental Program, University of California, Los Angeles, USA
[2]Friedrich-Alexander-University of Erlangen-Nürnberg, Germany
[3]Department of Bioengineering, University of California, Los Angeles, USA

Immunoglobulin (Ig)G antibodies are crucial regulators of immune responses that can direct immune effector responses by binding antigen specifically and interact with various effector cell via Fcγ receptors (FcγRs). While Fc domain of IgG presents great variations in subclasses and glycosylations, the contribution of having varied Fc forms in combination has not been investigated. Here, we use a computational model to predict binding to FcγRs and subsequent *in vivo* FcγR-dependent effects. This model could not only capture the trends of *in vitro* IgG mixture measurements, but also update the reported FcγR-IgG affinities and verify them in an independent dataset. We combined the model with generalized linear regression to analyze the role of each effector cell in murine antibody-mediated platelet depletion, and successfully identified that a subpopulation of Kupffer cells exerted most of the effector response. This work indicates that this binding model is an accurate tool for engineering IgG mixtures.

14. Determining the mechanisms by which exercise exerts its effect on cancer

Brandon Tsai[1], Lydia Liu[2], Stefan Eng[1], Whitney Underwood3, Courtenay Graham[3], Lee Jones[3] and Paul Boutros[1]
[1]University of California, Los Angeles, Los Angeles, CA, United States
[2]Ontario Institute for Cancer Research, Toronto, ON, Canada,
[3]Memorial Sloan Kettering Cancer Center, New York, NY, United States

While cancer incidence is a stochastic process, many factors influence the probability that an individual will develop cancer, including genetics, environmental exposures and lifestyle factors. Exercise is the strongest positive modifiable risk factor and has been linked to almost all cancer types and stages of disease progression. However, even within a specific cancer type, tumors appear to respond differently to exercise. Indeed, the molecular mechanisms by which exercise exerts its effect on cancer outcomes are almost entirely unclear. To fill this fundamental gap in our understanding of cancer etiology, I will employ three biological systems: (1) mouse xenografts of human breast cancer cell lines, (2) large cross-sectional retrospective patient cohorts and (3) a prospective longitudinal clinical trial. Taken together, these three experimental designs will allow me to test my hypothesis: exercise exerts tumor-dependent effects on the clinical and molecular behavior of cancers.

15. An age-structured Lotka-Volterra model and the emergence of overcompensation

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

There has been renewed interest in understanding the mathematical structure of ecological population models that lead to overcompensation, the process by which a population recovers to a higher level after suffering an increase in predation or harvesting. We construct an age-structured single-species population model that incorporates a Lotka-Volterra-type cannibalism interaction. Depending on the structure of the interaction, our model can exhibit overcompensation as well as oscillations of the total population; both phenomena have been observed in ecological systems. Analytic and numerical analysis of our model reveals sufficient conditions for overcompensation and oscillations. We also show how our structured population PDE model can be reduced to coupled ODE models representing piecewise constant parameter domains, providing additional mathematical insight into the emergence of overcompensation.

Poster Session II – (Reception/4:30pm)

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

Siddharth Agarwal[1,2], Dino Osmanovic[1], Melissa Klocke[1], Elisa Franco[1,2]
[1] Department of Mechanical and Aerospace Engineering
[2] Department of Bioengineering, University of California Los Angeles, Los Angeles California, USA

Phase separation of molecular condensates is emerging as a key mechanism in biology and biomaterials science. A major advantage of condensates is their capacity to form and reconfigure dynamically, generating responsive compartments that organize molecular targets and reactions in both space and time, in the absence of membranes. While condensation is known to depend on environmental conditions such as temperature and ionic strength, biological condensates in nature are likely influenced by fluctuating biochemical signals with high specificity. Here we ask whether the behavior of artificial condensates can be controlled via chemical reactions by design. Through theory and experiments we examine a model problem in which a phase separating component participates in chemical reactions that activate and deactivate its ability to self-attract. Our theoretical model indicates that such reactions have effects comparable to temperature, and illustrates the dependence of condensate kinetics on reaction parameters. We experimentally realize our model problem through a platform that combines DNA nanostar motifs to generate condensate droplets, and strand displacement reactions to kinetically control the nanostar valency. Our results show that DNA condensate dissolution and growth can be controlled reversibly via toehold-mediated strand displacement, and we characterize the influence of toehold and invasion domains, nanostar size, and nanostar valency. In some cases, the reduction of nanostar valency through invasion stabilizes the droplet size. Our results provide foundational methods for the development of dynamic nucleic acid condensates with potential applications in biomaterials science, nanofabrication, and drug delivery.

2. **Identifying Key Pathway Genes for Improving Network-Based Side-Effect Prediction**

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

Preclinical prediction of drug-induced safety events is of the utmost importance. We discovered that downstream proteins, in addition to drug targets, are relevant for predicting drug side effects (SEs). We benchmarked PathFX, a phenotypic pathways method using protein-protein interactions, in predicting SEs from drug labels. We used a drug toxicity dataset and mapped active ingredients to DrugBank identifiers and developed an ensemble approach to find PathFX phenotypes relevant to labeled SEs. We made two methods to evaluate PathFX performance per drug and per SE. In conclusion, baseline prediction performance was low and varied per SE or per drug. The discovery of pathway genes associated with true and false positive hypertension networks suggests that core hypertension pathway genes may be driving prediction and that under-predicted drugs may be missing links to these pathways. Future work will consider additional data sources for improving pathways-based prediction of SEs and emphasize discovering key genes.

3. **Unpacking the Columbian Exchange: viral persistence in trans-oceanic colonial shipping networks, c. 1492-1918**

Elizabeth Blackmore[1], James O. Lloyd-Smith[2]
[1] MS Biology, specializing in the Ecology and Evolution of Medicine, UCLA
[2] Department of Ecology and Evolutionary Biology, UCLA

From the fifteenth century onwards, expansions in colonial shipping opened new possibilities in global pathogen circulation. Yet, between lengthy journey times and small ship populations, fast-burning respiratory viruses such as smallpox, measles, and influenza could easily go extinct long before a ship reached port. We use a simple stochastic SEIR model to outline the conditions necessary for shipborne introduction of these pathogens. Our results indicate that pathogen-specific natural histories, journey times and frequencies, and on-board population characteristics all influenced the cumulative probability of introduction. Future work should investigate both technological advances and developments in the scale and purpose of transoceanic travel to understand how, where, and why global networks of acute infectious disease first arose.

4. **The genetics of gene expression at single-cell resolution in yeast**

James Boocock, Leonid Kruglyak
Department of Huang Genetics, University of California Los Angeles

Heritable differences in gene expression provide a molecular lens into the biology of traits. Gene expression levels are often measured from bulk populations, obscuring the differences between the effects of expression quantitative trait loci (eQTLs) on single cells due to factors such as their cell-cycle stage. We developed a one-pot approach to map eQTLs in *Saccharomyces cerevisiae* by single-cell RNA sequencing. We applied our approach to the transcriptomes of over 80,000 single cells of segregants from the extensively studied cross between the yeast strains RM and BY, as well as from two new crosses between strains YJM981 and CBS2888 and strains YPS163 and YJM454. We partitioned these cells into cell-cycle stages using unsupervised clustering in combination with marker genes and mapped tens of thousands of eQTLs, most of which acted independently of the cell-cycle stage. We found 12 new *trans*-acting eQTL hotspots, and we mapped 11 loci that influence the occupancy of different stages of the cell cycle. In segregants from the cross between YJM981 and CBS2888, we identified a trans-regulatory hotspot on chromosome X that influences the expression of thousands of genes. We fine-mapped the chromosome X hotspot to variants in the gene *CYR1*, which encodes adenylate cyclase, an enzyme that catalyzes the reaction that produces cyclic AMP (cAMP). This QTL also underlies fitness differences in a wide variety of conditions, and our single-cell data revealed that cells with the CBS2888 allele of *CYR1* more frequently occupy the G1 phase of the cell cycle and have increased fitness in 6 stressful conditions. Our results provide a more granular picture of how genetic variants can affect gene expression, cell-cycle progression, and condition-specific fitness.

5. **Using protein-protein interaction networks to predict drug synergy**

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

Combination therapies can improve efficacy and reduce adverse effects in treating complex diseases. However, due to the "combinatorial explosion" of potential drug combinations, it is difficult to identify synergistic combinations experimentally. Therefore, we have developed a network synergy pipeline to computationally identify synergistic combinations.The pipeline is derived from PathFX, a network tool which identifies proteins downstream of drug targets and measures associations of the target's network to multiple disease phenotypes. Here, we measure 43 single and 903 combination cancer-related drug network associations to 50 cancer phenotypes. Combination networks are then ranked based on a synergy score that measures if the combined networks have stronger associations to cancer phenotypes in comparison to their individual component networks. These scores are compared to experimental scores from the AstraZeneca DREAM challenge. Preliminary results reveal low-moderate correlation between predicted and experimental scores, dependence of predicted scores on phenotype pathway definitions, and distinct network topologies.

6. **Tensor Factorization for Interpreting the Mechanisms of Methicillin-Resistant Staphylococcus aureus Persistence**

Jackson L. Chin[1], Cyrillus Tan[2], Scott Taylor[1], Aaron Meyer[1,2]
[1]Department of Bioengineering

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

Methicillin-resistant Staphylococcus aureus (MRSA) bacteria is a common and life-threatening infection. While some antibiotics resolve MRSA infections *in vitro*, these antibiotics often fail to clear infections *in vivo*. Recent studies highlight the role of the host immune response in MRSA antibiotic resistance and have identified critical genetic and proteomic determinants of MRSA persistence, though the mechanisms that drive MRSA persistence are still poorly understood. Here, we implement tensor factorization to integrate transcriptomic and proteomic data collected from patients with persistent MRSA infections. Our factorization process identifies biological patterns across host factors and is able to explain 75% of the variance across transcriptomic and proteomic data with just 8 components. Further interpretation of these components highlights processes critical for MRSA persistence. Overall, these results suggest that tensor-based factorization can identify underlying mechanisms across host factors that both improve our ability to recognize and interpret the mechanisms that drive MRSA persistence.

7. Targeting Neuroendocrine Vulnerabilities in Platinum-Resistant Ovarian Cancer

Favour N. Esedebe[1,2], Ashvath Balgovind[2], Yi Jou (Ruby) Liao[2], Christopher Ochoa[3], Gabriella DiBernardo[3], Dr. Robert Damoiseaux[2,4,5], Dr. Sandra Orsulic[3,4], Dr. Tanya Singh[3,4,5], Dr. Sanaz Memarzadeh[2,3,4,5], Dr. Thomas Graeber[1,2,4,5]
[1] Bioinformatics Interdepartmental Program
[2] Department of Molecular and Medical Pharmacology, David Geffen School of Medicine
[3] Department of Obstetrics & Gynecology, David Geffen School of Medicine
[4] Jonsson Comprehensive Cancer Center
[5] Eli & Edythe Broad Center of Regenerative Medicine and Stem Cell Research
University of California Los Angeles, Los Angeles California, USA

Ovarian (OV) cancer is one of the deadliest gynecologic cancers in the U.S, with a low 5-year survival rate. While tumors initially shrink after platinum-based chemotherapy, they eventually relapse and become resistant to treatment. Patients with platinum-resistant OV cancer have few effective treatment options, hence the need to discover biomarkers and druggable targets. By analyzing bulk and single cell RNA-sequencing data collected from longitudinal matched patient and PDX tumor samples, we investigate the emergence of transdifferentiation as a means of resistance in OV cancer; here, we find previously unreported genetic and histologic evidence of a neuroendocrine phenotype in resistant recurrent OV tumors. Additionally, by performing multivariate integration of datasets from drug sensitivity and CRISPR knockout screens, we detected potential druggable gene and pathway targets associated with resistance and transdifferentiation in OV cancer cell lines. Overall, our findings provide us with unique vulnerabilities in platinum-resistant OV cancer that can be exploited therapeutically.

8. Simulation and Characterization of Randomly Generated Gene Networks

Hannah Michiko Faris, Timothy Hamilton, Breanne Sparta, Eric J. Deeds
[1]Institute for Quantitative and Computational Biosciences
[2]Department for Integrated Biology and Physiology, University of California Los Angeles

Gene expression levels are regarded to be the markers of cell types, with steady attractors yielding cell types. These steady attractors can best be identified by their dynamical behavior. However, the current approaches to studying gene regulatory networks eliminate key dynamical behaviors through over-simplification. They either ignore key interactions to focus on subsystems, or only allow for discrete expression levels. Our approach maintains both network size and complexity. Through this approach, we are able to randomly generate large networks so that we can study their inherent behavior. In these networks, two types of equilibrium occur: steady states and limit cycle attractors. However, these equilibria are not entirely stable. All equilibria are excitable, meaning that when they are perturbed they move away from the steady state equilibrium before returning. When characterizing these return paths and comparing to scRNA-seq data, we found similarities that may be consistent with similarities in attractor structure.

9. Genome-wide prediction of chromatin profiles from gene expression

Jingyuan Fu[1], Shan Sabri[2,3], Jason Ernst[1,2,3]
[1] Department of Computer Science,
[2] Department of Biological Chemistry,
[3] Interdepartmental Bioinformatics Program,
University of California Los Angeles, Los Angeles California, USA

High-throughput sequencing based methods for chromatin profiling have been developed to map epigenetic modifications such as histone methylation. Large compendia of these maps have been accumulated and used for biological questions including analyzing the contribution of non-coding genetic variants to disease. However, mapping such modifications at the single cell level remains challenging. In many cases it is easier to acquire gene expression data. This motivates the development of computational methods that impute chromatin marks by integrating compendia of chromatin and gene expression data into new cell types with only gene expression data available. Here we present a regression based method that makes predictions on histone modification based on gene expression featuers. We applied and evaluated the method against other state-of-the-art methods and demonstrate the potential application of this method in the deconvolution of bulk chromatin profiles into cell type specific chromatin profiles in the presence of single cell RNA-seq data.

10. Defining the distance between diseases using knowledge-graph embedding

Mingzhou Fu[1,2], Yu Yan[2], Loes Olde Loohuis[3,4], Timothy S Chang[1*]
[1] Movement Disorders Program, Department of Neurology, David Geffen School of Medicine,
[2] Medical Informatics Home Area, Department of Bioinformatics,
[3] Department of Psychiatry, David Geffen School of Medicine,
[4] Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine,
University of California Los Angeles, Los Angeles California, USA
* Correspondence author.

Characterizing disease relationships is essential to biomedical research to understand disease etiology and risk factors. Measurements of distance between disease pairs enable valuable research tasks. Distance metrics developed in prior work have focused on smaller, targeted disease sets, while measures to calculate distance between all disease pairs have not yet been defined, limiting the application to broader disease spectrum. Our current study defines disease distances for all diseases within the International Classification of Diseases (ICD), the diagnostic classification system universally used in electronic health records. Our proposed distance is based on a biomedical knowledge graph. We compared the knowledge graph-based metric to three other distance metrics based on the hierarchical structure of ICD, clinical comorbidity, and genetic correlation, to evaluate how each capture similar/unique aspects of disease relationships. We show that the knowledge graph-based distance metric captures known phenotypic, clinical, and molecular characteristics at a finer granularity than the other three distance metrics.

11. Many-core algorithms for high-dimensional gradients on phylogenetic trees

Karthik Gangavarapu[1], Xiang Ji[2], Guy Baele[3], Mathieu Fourment[4], Philippe Lemey[3], Frederick A. Matsen IV[5], Marc A. Suchard[1,6,7]
[1]Department of Biomathematics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, United States
[2]Department of Mathematics, Tulane University, New Orleans, United States
[3]Department of Microbiology and Immunology, Rega Institute, KU Leuven, Leuven, Belgium
[4]Three Institute, University of Technology Sydney, Ultimo, Australia
[5]Fred Hutchinson Cancer Research Center, Seattle, United States
[6]Department of Biostatistics, Jonathan and Karin Fielding School of Public Health, University of California, Los Angeles, United States
[7]Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, United States

Advancements in genomic sequencing have led to the generation of genomic data at an unprecedented rate spurring the need for statistical phylogenetic algorithms that can traverse the space of possible phylogenetic histories efficiently. A recent study proposed an algorithm to calculate the gradient of the log-likelihood in linear-time allowing Bayesian phylogenetic methods to take advantage of efficient gradient-based samplers. The CPU implementation of the algorithm makes the calculation of the gradient tractable for nucleotide-based models but falls short for models with larger state-size. Here, we describe novel algorithms to calculate the gradient utilizing the high parallelization afforded by graphics processing units. We report a >128-fold speedup of our algorithms over the CPU implementation for codon-based models. To demonstrate the utility of our algorithms, we date the first introduction of West Nile virus into the Unites States under a codon model and a relaxed molecular clock using a dataset of 501 full viral genomes.

12. Proteogenomic characterization of the molecular determinants of prostate cancer Radioresistance

Roni Haas[1,2,3,4], Shahbaz Khan[5], Gavin Frame[6], Wenyan Zhao[1,2,3,4], Benjamin Carlin[1,2,3,4], Takafumi N. Yamaguchi[2,3,4], Yuan Zhe Bugh[1,2,3,4], Julie Livingstone[1,2,3,4], Chenghao Zhu[1,2,3,4], Rupert Hugh-White[1,2,3,4], Shu Tao[1,2,3,4], Andrew Macklin[5], Vladimir Ignatchenko[5], Natalie Kurganovs[5], Geoff S. Higgins[7], Michelle R. Downes[8,9], Danny Vesprini[10], Andrew Loblaw[10], *Thomas Kislinger[5,6], *Paul C. Boutros[1,2,3,4,6], *Stanley. K Liu[6,10,11]

[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]Princess Margaret Cancer Centre, University Health Network, Toronto, Canada
[6]Department of Medical Biophysics, University of Toronto, Toronto, Canada
[7]Department of Oncology, Cancer Research UK/MRC Oxford Institute for Radiation Oncology, University of Oxford, Oxford, UK
[8]Division of Anatomic Pathology, Laboratory Medicine and Molecular Diagnostics, Sunnybrook
Health Sciences Centre, Toronto, Canada
[9]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada
[10]Department of Radiation Oncology, University of Toronto, Toronto, Canada
[11]Sunnybrook Research Institute, Sunnybrook Health Sciences Centre, Toronto, Canada
* co-senior authors

Prostate Cancer (PC) radioresistance is a major clinical concern. Hypo-Fractionated (HF) radiotherapy, with high radiation doses per treatment, has recently become clinically preferable over Conventional-Fractionation (CF). Radioresistance can emerge following both CF and HF, but it is unknown if the underlying resistance mechanisms are similar. We investigated the genomic, transcriptomic, and proteomic differences between HF and CF radioresistant PC cells. Compared to HF, CF radioresistant cells gained twice the number of mutations, demonstrated strong dysregulation of cancer related genes in RNA abundance profiles, and showed increased DNA-repair activity in cell nuclei on proteomic analysis. Collectively, we observed a far more aggressive phenotype in CF cells compared to HF. The clinical relevance of potential therapeutic targets was assessed using a cohort of 380 PC patients. We highlighted *POLQ* as a potential target and showed radiosensitization upon *POLQ* knockdown. Our study provides a platform for the development of therapies for radio-recurrent PC.

13. Fractal-like Density Distributions in Single-Cell Data

Timothy Hamilton[1,2*], Breanne Sparta[2,3*], Serena Hughes[1,2], Eric J. Deeds[1,2,3],
[1] Bioinformatics Interdepartmental Program, University of California, Los Angeles
[2] Institute of Quantitative and Computational Biology, University of California, Los Angeles
[3] Department of Integrative Biology and Physiology, University of California, Los Angeles
* These authors contributed equally.

With the advent of high-throughput sequencing and high-resolution single molecule microscopy, single-cell methods for analyzing various omics data have become increasingly prevalent. Many of the analyses that are applied to these data assume that the data itself should follow the classic "Waddington's landscape" picture of the arrangement of cell types in gene expression space. In this picture, cell types correspond to discrete "attractors" in the epigenetic landscape. They should thus form well-separated groups that each cluster around the center of the basin of attraction for a given cell type. Despite nearly a decade of analysis of single-cell data, however, the attractor structure of cell types in the gene expression space has never been directly characterized. Here, we used a novel analytical approach, which we term "epsilon networks," to characterize how cells are distributed in the underlying, high-dimensional gene expression space. If cell types corresponded to attractors, we would expect to see most points in the high-density regions in center of the basins of attraction, with points in lower-density regions being rarer. What we found instead, however, is that the density distribution is approximately power-law: most cells are in low-density regions, very far from other cells, while a small number of cells are found in extremely high-density regions. We found this behavior is universal in single-cell data on epigenetic state, regardless of the experimental technique employed. This "fractal-like" density distribution is inconsistent with the idea that cells are sampled smoothly from cell type attractors or developmental trajectories. Our findings have implications for the correspondence between bulk measurements and single cell measurements, as well as our overall picture of how stable cell types arise during development. Understanding how these fractal densities are generated and what they mean for the development and physiology of cell types represents a major challenge for the future of single-cell biology.

14. Comprehensive study of gene expression outliers and their regulation mechanisms in pan-cancer

Jee Yun Han[1,2,3,4], Paul C Boutros[1,2,3,4,5]
[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 Medical Biophysics, University of Toronto, Toronto, Canada

Cancer is a disease characterized by remarkable heterogeneity. Gene expression varies drastically between tumours and within cells of a single. This variability can generate extreme outliers: transcripts that show atypically high gene expression in a small percentage of cancers. These outliers increase the molecular and phenotypic diversity between individuals, contributing to tumour heterogeneity. However, the previous research has been limited to a single cancer type or a single gene. To fill the gap of our understanding, this study will identify the outliers from various cancer types using our novel statistical method and explore the biological functions of outliers. With integrated analysis with the clinical outcome, it will show how outliers affect the progression of cancer. This study is expected to deepen our understanding of the impact of outliers on different cancers by dissecting their dysregulation and will allow us to identify a novel cancer driver and potential drug target.

15. Contrasting the tempo and mode of adaptation on the X and autosomes in *Drosophila melanogaster*

Mariana Harris[1], Nandita Garud[2,3]
[1] Department of Computational Medicine Biomathematics Program, University of California Los Angeles, Los Angeles California, USA
[2] Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles California, USA

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

Adaptation on the X chromosome has attracted significant interest from evolutionary biologists because its dynamics seem to be distinct from that of autosomes. Here, we investigate the differences in the mode and tempo of adaptation in the X chromosome and autosomes. To do so, we quantify the incidence of hard versus soft sweeps, which are the signatures of selection when selection is gradual versus rapid, respectively. Using haplotype homozygosity statistics, we find an enrichment of hard sweeps on the X chromosome relative to the autosomes in North American *D. melanogaster* population genomic data, confirming predictions we make from simulations. Our results suggests that signatures of selection may differ between the X chromosome and the autosomes. Investigating these differences may enable a deeper understanding of how important phenotypes arise as well as the role of fundamental evolutionary parameters on adaptation, such as dominance, sex-specific selection, and sex-biased demography.

16. **Developmental Trajectory Analysis of Differentiating Mouse sensory Applying joint graph embedding to study Alzheimer's neurodegeneration patterns in volumetric data**

Rosemary He[1], Daniel Tward[2, 3]
[1] Department of Computer Science, UCLA, 90095
[2] Department of Computational Medicine, UCLA, 90095
[3] Department of Neurology, Ahmanson-Lovelace Brain Mapping Center, UCLA, 90095

Neurodegeneration measured through volumetry in timeseries MRI is recognized as a potential biomarker for Alzheimer's Disease (AD), but suffers from lack of specificity. Quantifying spatial patterns on a whole-brain scale, rather than some structures of interest, may help address this issue. Traditional approaches capture neurodegeneration over time, but may be too simplistic to capture more complex relations between structures and networks. In this work, we turn to network-based analyses and extend an existing joint graph embedding method to uncover networks that describe volume correlations during neurodegeneration. We identify significant networks that can distinguish mild from severe AD groups. We further develop a novel statistical testing procedure for identifying significant structure or pairs of structures within these networks. We find the most significant networks are dominated by structures associated to AD degeneration, including limbic and ventricular structures. The results show promise in studying AD and discovering neurodegeneration biomarkers with increased specificity.

17. **Epigenomic and chromosomal architectural reconfiguration in developing human frontal cortex and hippocampus**

Matthew G. Heffel[1,2], Yi Zhang[2], Kevin Abuhanna[2], Terence Li[1,2], Mercedes Parades[3], Tom Nowakowski[4], Eran Mukamel[5], Jesse Dixon[6], Chongyuan Luo[2]
[1] Bioinformatics Interdepartmental Program,
[2] Department of Human Genetics, UCLA, Los Angeles California, USA.
[3] Department Neurology,
[4] Department of Anatomy, UCSF, San Francisco California, USA.
[5] Department of Cognitive Science, UCSD, Los Angeles California, USA
[6] Gene Expression Laboratory, Salk Institute for Biological Sciences, La Jolla, USA

The specification and maturation of human cortical and hippocampal cell types are associated with global epigenomic reconfiguration including the pronounced accumulation of neuronal non-CG DNA methylation, as well as the remodeling of 3D-chromatin domains and chromatin contacts between enhancers and promoters. Here we investigated the epigenomic and 3D-chromatin conformational dynamics of human frontal cortex and hippocampus development using >53,000 joint single-nucleus profiles of chromatin conformation and DNA methylation (sn-m3C-seq) generated from mid-gestational, late-gestational, infant and adult human brains. We identified enriched dynamics of DNA methylation in late-gestational to early-post-natal development, preceded by the reconfiguration of chromatin conformation. We reconstructed 3D-connected regulatory hierarchies of cortical and hippocampal cell differentiation and maturation and identified brain regional-specific regulatory signatures. The single-cell 3D-regulatome approach further enabled the cell-type-specific dissection of neuropsychiatric risk loci during key neurodevelopmental stages as well as in adult human brains.

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

Ran Hu[1,3,4], Benjamin Tran[2], Mary Stackpole[1,3,4], Shuo Li[1,3,4], Vatche Agopian[2], Xianghong Jasmine Zhou[1,3], Wenyuan Li[1,3]
[1]Department of Pathology and Laboratory Medicine,
[2]Department of Surgery,
[3]Institute for Quantitative & Computational Biosciences,
[4]Bioinformatics Interdepartmental Program,
University of California Los Angeles, Los Angeles, California, USA

Current noninvasive prognosis evaluation of hepatocellular carcinoma (HCC) largely relies on imaging and alpha-fetoprotein (AFP) from blood. HCC prognosis is challenging when common prognostic factors are inconsistent and lack sensitivity. Here, we evaluate the use of methylation profiles of cell-free DNA (cfDNA) for the pre-treatment prognostication of overall survival (OS) in patients with HCC. We use 377 HCC tumor and 50 adjacent normal tissue Illumina 450K methylation array data to identify differentially methylated positions that are predictive of survival. We further develop random survival forest models based on selected markers to predict OS risk scores for two patient cohorts with reduced representation bisulfite sequencing (RRBS) data from 30 tumor tissue and 52 plasma samples. The risk score achieves higher overall time-dependent AUC compared with clinical features and stratifies patients into significantly different risk groups. Our results show the capability of cfDNA methylation as a promising noninvasive prognostic predictor in HCC.

19. **Global analysis of RNA editing in Alzheimer's disease across multiple brain regions**

Elaine Huang[1], Mudra Choudhury[1], Xinshu Xiao[1,2,3]
[1] Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA USA
[2] Molecular, Cellular, and Integrative Physiology Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA 90095, USA
[3] Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles, CA 90095, USA

RNA editing is a biological process that refers to the alteration of RNA transcripts through insertion, deletion, or substitution of nucleotides. Here, we aim to gain a comprehensive understanding of the relationship between RNA editing and Alzheimer's Disease (AD) pathology. We utilized RNA-seq data from the Mount Sinai Brain Bank and the ROS/MAP project. Each cohort encompass hundreds of samples from multiple brain regions. Using our *de novo* editing site detection pipeline, we found that thousands of sites (including protein-recoding sites in AD-related genes) have significantly different editing levels between AD and control samples. To better understand regulators of editing levels, we identified RNA binding proteins (RBPs) with expression levels that significantly correlate with editing. Finally, we performed ordinal regression analyses to identify sites that are related to the severity of various measures of AD pathology. Our work presents a systematic view of how RNA editing contributes to AD progression.

20. **Guilty until proven innocent: Antigen-BCR interactions prime B-cells for death until CD40 signaling comes to rescue**

Helen Huang[1,2,3], Haripriya Vaidehi Narayanan[2,3], Alexander Hoffmann[2,3]
[1] Bioinformatics Interdepartmental Program,
[2] Institute for Quantitative and Computational Biosciences (QCB),
[3] Department of Microbiology, Immunology, and Molecular Genetics (MIMG), University of California Los Angeles, Los Angeles, USA.

During an antibody response, B-cells undergo affinity maturation: the B-cell receptor (BCR) is randomly mutated and then subjected to selection based on interactions with antigen (through the BCR) and T-cells (through the CD40 receptor). A successful response must enrich high-affinity antigen-reactive B-cells through positive selection, but eliminate autoreactive B-cells by negative selection. Little is known about the mechanism by which BCR and CD40 signaling jointly determine B-cell selection. We quantitatively evaluated the population dynamics after

stimulating B-cells, to develop a mathematical model of the BCR and CD40 signaling pathways and their impact on B-cell fate decision machineries. Our results show that BCR signaling triggers apoptosis at short timescales, while CD40 signaling rescues B-cells from death, driving both proliferation and differentiation into antibody-secreting plasma cells at long timescales. The choice between death and division depends upon the strength of BCR activation, the strength and duration of CD40 stimulation, and the time delay between both. Thus, we propose a form of kinetic proofreading in B-cell affinity maturation, where BCR and CD40 signals work in opposition to determine B-cell fate decisions, discriminating B-cells with high selectivity.

21. Partitioning polygenic risk scores explains sex-specific differences in abdominal obesity

Huiling Huang[1,2], Asha Kar[1], Milena Deal[1], Marcus Alvarez[1], Karen L. Mohlke[5], Kirsi H. Pietiläinen[6,7], Markku Laakso[8], Janet S. Sinsheimer[1,3], Päivi Pajukanta[1,2,4]

[1]Department of Human Genetics,

[2]Bioinformatics Interdepartmental Program,

[3]Department of Computational Medicine,

[4]Institute for Precision Health,

University of California Los Angeles, Los Angeles California, USA

[5]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[6]Obesity Research Unit, Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[7]Obesity Center, Endocrinology, Abdominal Center, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland

[8]Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland

Partitioning polygenic risk scores (PRS) based on both genetic variants and fat tissue cell-type marker genes differentially expressed by sex can inform us about the key genomic elements driving sex differences in abdominal obesity. To this end, we used waist-hip ratio adjusted for body mass index (WHRadjBMI) as a proxy for abdominal obesity and compared sex-stratified WHRadjBMI PRSs on 3 sets of SNPs in the UK Biobank: SNPs in cis-regions of 31 sex-specific adipocyte (SSA) marker genes, sex-specific WHRadjBMI GWAS (SSG) SNPs, and full genome. When comparing SSA and SSG PRSs to the full genome that explains 9.53% of the WHRadjBMI variance in females and 2.77% in males, the female SSA PRS explains 2.77X more variance than expected by the number of covered loci, and the female SSG PRS explains almost half of the variance in full genome PRS. Overall, we discover that the 31 adipocyte marker gene regions and sex-stratified WHRadjBMI GWAS loci explain more abdominal obesity variance in females (4.84%) than males (0.51%), suggesting a larger role of environment and gene-environment interactions in male abdominal obesity.

22. A rigorous benchmarking of methods for SARS-CoV-2 lineage detection in wastewater.

Sergey Knyazev, Tianze Tao, Ke Wang, Bohdan Tyshchenko, Wenhao O. Ouyang, Alina Frolova, Pelin Icer Baykal, Bogdan Pasaniuc, Niko Beerenwinkel, Nicholas C. Wu, Alex Zelikovsky, Adam L. Smith, Serghei Mangul

The COVID-19 pandemic showed that an efficient real-time response to pandemics is required to minimize the economic, social, and public health burdens resulting from pandemics. As the SARS-CoV-2 virus continues to spread, evolve, and mutate, the need for a cost-effective and efficient way to detect the presence of lineages is beyond urgent. Using wastewater-based surveillance for COVID-19 showed its efficacy in numerous countries around the globe for monitoring viral prevalence in the population. Wastewater-based surveillance has numerous advantages including that it does not require interaction with patients and can simultaneously monitor entire communities including underserved and vulnerable populations as well as asymptomatic cases. The vast majority of current SARS-CoV-2 wastewater monitoring facilities are qPCR-based and can only quantify viral load without differentiating viral strains, which prevents monitoring strain prevalence and detecting novel strains. To quantify the novel and existing strains, the sequencing of wastewater samples coupled with advanced computational tools can promise to elucidate the relative abundances of known and novel strains. Some studies show that this approach allowed monitoring not only for the number of cases but also to quantify the prevalence of viral variants including detecting strains that were absent from clinical databases–suggesting that clinical databases may be delayed. However, scalable and effective methods are yet to be developed because wastewater genomic surveillance poses technical challenges including viral genome degradation in wastewater facilities resulting in poor sequencing quality and incomplete genome coverages. We propose benchmarking existing methods for lineage detection in wastewater samples containing SARS-CoV-2. We will get state-of-the-art methods and apply them on our inhouse benchmarks. The in-silico benchmarks are based on real wastewater samples and mimicking its lineage composure, genome degradation, and genomic bias identical to real samples. The genome-engineered in-vitro gold benchmarks provide positive and negative control for testing the tools. Total of 26 methods are going to be benchmarked in order to give a reference for wastewater genome based monitoring.

23. A machine-learning based pairwise functional similarity metric between genomic loci

Runjia (Luke) Li[1,2], Jason Ernst[2]

[1] Bioinformatics Interdepartmental Program,

[2] Department of Biological Chemistry,

University of California Los Angeles, Los Angeles California, USA

Genome-wide association studies (GWASs) is the primary tool for identifying trait-associated genetic variants. Since these associations typically reside in the non-coding genome, it remains a challenge to interpret their downstream consequences. Functional annotations such as histone modifications and ChromHMM state segmentations can be used to describe the epigenetic properties of variants. We reason that variants associated to the same trait should be sharing some specific epigenetic properties; therefore, we propose a model that learns a pairwise similarity score between two variants from their functional annotations. We annotated 100k pairs of variants taken from the EMBL-EBI GWAS catalog using 3000+ ChIP-seq and DNase-seq data from the Roadmap Epigenomics Projects and 24 different universal ChromHMM models, and trained Naive Bayes classifiers to predict whether the pairs are associated to the same trait. Preliminary results suggest ChromHMM state segmentations are predictive of the variants' phenotypic similarity while further model tuning is still required.

24. SCING: Single Cell INtegrative Gene regulatory network inference elucidates robust, interpretable gene regulatory networks

Russell Littman[1,2], Ning Wang[1] & Xia Yang[1,2,3,4,5*]

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

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

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

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

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

*Corresponding author: xyang123@ucla.edu

Gene regulatory network (GRN) inference is an integral part of understanding physiology and disease. Single cell/nuclei RNAseq (scRNAseq/snRNAseq) data has been used to elucidate cell-type GRNs; however, the accuracy and speed of current scRNAseq-based GRN approaches are suboptimal. Here, we present Single Cell INtegrative Gene regulatory network inference (SCING), a gradient boosting and mutual information based approach for identifying robust GRNs from scRNAseq, snRNAseq and spatial transcriptomics data. Performance evaluation using held-out data, Perturb-seq datasets, and the mouse cell atlas combined with the DisGeNET database demonstrates the improved accuracy and biological interpretability of SCING compared to existing methods. We applied SCING to the entire mouse single cell atlas, human Alzheimer's disease (AD), and mouse AD spatial transcriptomics. SCING GRNs reveal unique disease subnetwork modeling capabilities, have intrinsic capacity to correct for batch effects, retrieve disease relevant

genes and pathways, and are informative on spatial specificity of disease pathogenesis.

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

Emily Maciejewski[1,2], Steve Horvath[3,4], Jason Ernst[1,2]
1 Department of Computer Science
2 Department of Biological Chemistry,
3 Department of Human Genetics, University of California, Los Angeles, Los Angeles CA, USA
4 Altos Labs, San Diego CA, USA

DNA methylation is widely profiled in humans, but for many mammals there is more limited data available. Biological samples from certain tissues in some species can be difficult to obtain, motivating the development of methods to accurately impute DNA methylation samples representing species-tissue combinations that were not experimentally profiled. We designed a method that leverages cross-species DNA methylation and uses a Conditional Variational Autoencoder to generate DNA methylation value predictions for unobserved species-tissue combinations in its input compendium. We show that the method correlates well with held-out ground truth values and effectively preserves species and tissue relationships among samples. After performing cross-validation analyses, we use all available data to impute mean methylation samples for species-tissue combinations that have not previously been experimentally profiled. We expect the method and imputed data resource we have developed will be useful for DNA methylation analyses of species and tissue-level characteristics across mammalian species.

26. Inference of the demographic history of commensal gut microbes

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

Human commensal gut microbes play a crucial role in host health, including aiding with the digestion of foods that humans cannot digest themselves. Despite the importance of such microbes, there is little knowledge about the evolutionary history of commensal gut microbes. In this study, we infer the demographic history for 27 highly prevalent commensal gut microbial species in North Americans. Nine of these species show evidence of population contractions coincident with the onset of agricultural expansion approximately 10,000 years ago, in contrast with population expansions observed in a similar timeframe by the cavity-causing oral microbe, Streptococcus mutans. Additionally, we find reductions in diversity observed at the species and genetic level in commensal gut microbes sampled from Western populations relative to non-Western rural populations. Taken together, we infer a reduction in effective population size of gut microbiota coincident with the onset of agriculture in human history.

27. Impact of effect size and allele frequency heterogeneity by local ancestry on disease mapping in admixed populations.

Rachel Mester[1], Kangcheng Hou[2], Arjun Bhattacharya[3], Kathryn S. Burch[2], Yi Ding[2], Bogdan Pasaniuc[1,2,3,4]
[1]Department of Computational Medicine
[2]Bioinformatics Interdepartmental Program
[3]Department of Pathology and Laboratory Medicine
[4]Department of Human Genetics
University of California Los Angeles, Los Angeles California, USA

Admixed populations present both challenge and opportunity in disease mapping studies due to the need to balance the potential for capturing additional signal by conditioning variant effects on local ancestry with the anticipated loss of power due to correction for genetic structure within analyses. We show that methods that do not include local ancestry information are more powerful when allelic effects are similar across ancestries or allele frequencies differ decidedly by ancestry. Tests that condition on local ancestry perform optimally when a large enough level of heterogeneity in causal effects is present. We find the levels of effect size heterogeneity and allele frequency difference required in order to recommend each test. Using data from admixed individuals in the UK Biobank, we also investigate the extent to which differential linkage disequilibrium across local ancestries is expected to induce this amount of heterogeneity by local ancestry in causal effect sizes.

28. Deconvolving tissue spots

Jonathan Perrie[1], Charlotte Hu[1], Robert Modlin[2,3], Matteo Pellegrini[2,4,5,6]
[1] Bioinformatics Interdepartmental Program
[2] Department of Medicine
[3] Department of Microbiology, Immunology and Molecular Genetics
[4] Department of Molecular, Cell and Developmental Biology
[5] Department of Human Genetics
[6] California NanoSystems

Spatial transcriptomics enables the profiling of tissue samples with aggregate measures of gene expression coupled to a set of coordinates at a spot in the tissue. These spots are distributed across the tissue and can be composed of one or many cell types. Each spot can be thought as a linear combination of expression profiles for individual cell types, hence, treating this problem as a penalized least squares regression is appropriate given that the expression profiles of the spots and reference cell types are on the same scale. This mirrors an older problem: deconvolving bulk RNA-seq using a signature matrix, and we employ similar pre-processing techniques; however, spatial transcriptomics data generally has lower capture efficiency, hence, more noise. Here, we apply LASSO regression to samples from the spatial platforms, GeoMx and Visium, and compare their results.

29. A Machine Learning Approach to Predicting Thrombectomy Recanalization from Basic CT Imaging

Nidhi Ramesh[1], Jennifer Polson[1,3], Harry Zhang[1,3], and Corey W. Arnold[1-5] (Ph.D.)
[1]UCLA Computational Diagnostics Lab, University of California Los Angeles
[2]Department of Radiological Sciences, University of California Los Angeles
[3]Department of Bioengineering, University of California Los Angeles
[4]Department of Pathology & Laboratory Medicine, University of California Los Angeles
[5]Department of Electrical & Computer Engineering, University of California Los Angeles

When an acute ischemic stroke patient is treated with mechanical thrombectomy (MTB), the success of the treatment method is evaluated based on the modified treatment in cerebral infarction (mTICI) score. This project aims to create an automatic machine learning model that predicts mTICI scores, trained on a retrospective cohort of 177 patients who received MTB and had pretreatment computed tomography (CT) and CT angiography (CTA). A total of 215 quantitative radiomic features were extracted from both CT and CTA scans and used as features. The best performing model combination used Random Forest for both feature selection and classification, achieving 78.23% +/- 0.27% area under the curve (AUC) of receiver operating characteristic (ROC) curve, 78.23% +/- 0.69% sensitivity, and 78.27% +/- 0.71% specificity. These results suggest CT and CTA imaging in conjunction with a region-based mapping approach to predict successful recanalization in MTB patients show promising initial results.

30. Coordinated single-cell transcriptome dynamics enable the accuracy of macrophage responses to immune stimuli

Katherine M Sheu, Aditya Pimplaskar, Alexander Hoffmann
Institute for Quantitative and Computational Biosciences and
Department of Microbiology, Immunology, and Molecular Genetics

Macrophages respond in a context-dependent manner to appropriately neutralize immune threats, via the dynamic expression of hundreds of genes specific to the stimulus. However, each macrophage cell operates under noisy conditions that may interfere with the accuracy of transcriptomic responses. The dynamics of gene expression is thought to be critical for phasing immune response functions, but technological limitations have precluded measuring the dynamics of the transcriptomic response in single cells. Here, to overcome this limitation we first developed a method to reconstruct heterogeneous single-cell transcriptome time-courses contained within the population of macrophages, using time-point scRNAseq measurements. We then

calculated dynamical features across hundreds of expression trajectories and found which combinations of gene features were most important for distinguishing stimuli, across microenvironment contexts. Mathematical modeling of gene regulatory strategies explained the relationship between information contained in NFκB signaling versus transcriptome dynamics, elucidating how transcriptome dynamics maintain innate immune response accuracy.

31. Dissecting the dysregulation of hematopoietic progenitor cell fate specification caused by age-associated inflammation

Apeksha Singh[1,2], Jennifer Chia[2,3], Yu-Sheng Lin[2], Alexander Hoffmann[2]
[1] Biomathematics Graduate Program
[2] Signaling Systems Laboratory, MIMG Dept., & QCBio Institute
[3] Department of Pathology and Laboratory Medicine
University of California Los Angeles, Los Angeles, CA 90095

Aging, together with chronic low-grade inflammation ("inflammaging"), is associated with altered hematopoiesis resulting in increased myeloid and decreased lymphoid output, which increases infection susceptibility and reduces vaccine efficacy. It remains unclear, however, which hematopoietic stem and progenitor cells (HSPC) are affected and how their developmental dynamics are altered by chronic inflammation. To study hematopoietic dysregulation in inflammaging, we first formulate mathematical models of HSPC population dynamics to explore how perturbations to proliferation and differentiation parameters within the developmental pathway affect hematopoietic output. We then utilize mouse models of NFκB-driven inflammation, to identify cell fate decision changes that are compatible with the observed data of myeloid bias. Finally, we identify genes and pathways differentially expressed in HSPC subpopulations of these mouse models to elucidate molecular processes underlying dysregulated fate decisions. Our studies aim to provide mechanistic insights about myeloid-biased hematopoiesis in inflammaging, and hence suggest strategies for therapeutic intervention.

32. Harnessing the Power of Admixed Populations in GWAS
A lack of distinct cellular identities in single cell data: revisiting Waddington's landscape

Breanne Sparta[1,2], Timothy Hamilton[1,2,3], Eric J. Deeds[1,2]
[1]Institute for Quantitative and Computational Biosciences
[2]Department of Integrated Biology and Physiology
[3]Bioinformatics Interdepartmental Program
University of California Los Angeles, Los Angeles California, USA

A prevailing interpretation of Waddington's landscape is that cells with similar physiologies exist within a shared basin of attraction and exhibit similar gene expression patterns. To test this hypothesis, we used graph theory to characterize the distribution of cells in epigenetic space. Within a variety of single-cell omics datasets, we find that cell types exist in the same regions of epigenetic space, with a density distribution that is approximately power law. These findings are inconsistent with the idea that cell types are produced by attractors, and encourage us to consider alternative hypotheses for how epigenetic changes maintain physiological stability.

33. Text mining to quantify mitochondrial proteins, pathways, and function in cardiovascular disease

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

Mitochondrial function is essential for proper cellular function and health. Mitochondrial dysfunction has been implicated in numerous diseases, including cardiovascular diseases. A large body of work on mitochondrial proteins and cardiovascular diseases have been published, yet not fully utilized. By applying a text mining platform (CaseOLAP) to these publications, we can extract relevant information to provide us insight into mitochondrial proteins, their function, and the molecular pathways involved in multiple cardiovascular disease categories. Here, we present this approach and its findings: the mitochondrial proteins that ostensibly play the biggest role in cardiovascular diseases and the mitochondrial proteins most unique to each cardiovascular disease. Furthermore, we present the proteins' functions and pathways. By understanding the molecular underpinnings of these diseases, we can better understand how mitochondria are affected by and cause cardiovascular disease, paving the way for future disease treatments.

34. Germline structural variants shape prostate cancer clinical and molecular evolution

Nicholas K. Wang[2,3], Alexandre Rouette[1], Kathleen E. Houlahan[1,2,4,5,3,6], Takafumi N. Yamaguchi[2,3,5,6], Julie Livingstone[2,3,5,6], Chol-Hee Jung[8], Peter Georgeson[8], Michael Fraser[1], Yu-Jia Shiah[1], Cindy Q. Yao[1], Vincent Huang[1], Natalie S. Fox[1,2,3,4,5,6], Natalie Kurganovs[7,8], Katayoon Kasaian[1], Veronica Y. Sabelnykova[1], Jay Jayalath[1], Kenneth Weke[1], Helen Zhu[4,7], Theodorus van der Kwast[9], Tony Papenfuss[8], Housheng H. He[4,7], Niall M. Corcoran[10,11], Robert G. Bristow[4,7,12], Alexandre R. Zlotta[7], Christopher Hovens[8,10,11,§], Paul C. Boutros [2,3,4,5,6,13,§]
[1] Ontario Institute for Cancer Research
[2] Department of Human Genetics, UCLA
[3] Jonsson Comprehensive Cancer Centre, UCLA
[4] Department of Medical Biophysics, University of Toronto
[5] Department of Urology, UCLA
[6] Institute for Precision Health, UCLA
[7] Princess Margaret Cancer Center
[8] Australian Prostate Cancer Research Centre Epworth Richmond
[9] Department of Pathology and Laboratory Medicine, University of Toronto
[10] Department of Surgery, The University of Melbourne
[11] Department of Urology, Royal Melbourne Hospital
[12] Manchester Cancer Research Center
[13] Department of Pharmacology and Toxicology, University of Toronto
[§]Corresponding authors

Inherited genetic variation profoundly influences cancer risk and outcome. While the impact of germline single nucleotide polymorphisms has been well-studied in several cancer types, the effects of germline structural variants (gSVs) on cancer biology and clinical outcomes is largely unknown. From our cohort of 300 men with localized, intermediate risk prostate cancer, we identified 6,003 gSVs present in at least 3% of patients; 48 were associated with recurrent somatic alterations or clinical outcome. Of these, approximately 50% were associated with expression of nearby genes or intersected with exons or regulatory regions. Using external cohorts, we validated 3 gSVs that were strongly associated with poor clinical outcomes, including an inversion at chr14q24.1 present in ~20% of patients. Notably, a strong synergistic effect on outcome was observed in patients with somatic *TP53* alterations or high genomic instability, defining a new aggressive prostate cancer subtype with chr14^{INV} as a novel, recurrent biomarker.

35. Cell-cell Communication Gene Regulatory Network Inference Based on Single Cell Multiomics

Ning Wang[1*], Russell Littman[1,2,3*], and Xia Yang[1,2,3]
[1] Department of Integrative Biology and Physiology, UCLA
[2] Bioinformatics Interdepartmental PhD Program, UCLA
[3] Institute for Quantitative and Computational Biosciences, UCLA
* These authors contributed equally.

Gene regulatory networks (GRNs) depict which genes regulate one another through directed graphs, where nodes and edges represent genes and regulatory relationships between gene pairs. GRNs help illustrate the contribution of genes and pathways in physiology and disease. Single cell multiomics brings new opportunities to leverage multidimensional data to model GRNs across cell types. We designed and developed a new computational method, grnComm, to infer intercellular communication GRNs using gradient boosting decision trees based on collective information from single cell RNA sequencing, spatial transcriptomics, and receptor-ligand databases. To benchmark method performance, we will use spatial mutual information tests, pathway enrichment analysis, and differentially expressed genes predictions to compare grnComm against other network building methods. grnComm will enable integration of single cell multiomics data to derive accurate and comprehensive GRNs across cell types, which will inform on key cell types, genes, and pathways that regulate physiological homeostasis or disease etiology.

36. The 3D genome organization plays a key role in chemotherapy-induced DNA damage susceptibility

Ye Wang[1,2,3], Asli Yildirim[1,2], Lorenzo Boninsegna[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

Cisplatin is one of the most widely used chemotherapy drugs in cancer treatment. Although the cytotoxic and resistant mechanisms of cisplatin have been thoroughly explored, it remains elusive how the global 3D nuclear environment affects the susceptibility of DNA to cisplatin-induced damage. Here, we demonstrate that, the global nuclear location of genomic regions plays one of the key roles in modulating cisplatin DNA damage susceptibility in vivo. By integrating data from damage-seq experiments with information from 3D genome structures in human cells we show that the locations of genomic regions relative to specific nuclear bodies, such as nuclear speckles, and locations relative to the nuclear periphery can explain experimental patterns of DNA damage susceptibility. Thus, our results are consistent with recent observations of an enrichment of cisplatin in certain nuclear bodies. These results are relevant for a better understanding of cisplatin action and the development of cancer resistant cells.

37. Defining molecular dysregulation in Down Syndrome neocortex

Alexis Weber[1,2], Celine Vuong[1], Michael Margolis[1], Natalie Hawken[3], Michael Gandal[1], Daniel H. Geschwind[3], Luis de la Torre-Ubieta[1].

[1.] Department of Psychiatry and Biobehavioral Sciences, Intellectual and Developmental Disabilities Research Center, Semel Institute, David Geffen School of Medicine, University of California Los Angeles

[2.] Department of Human Genetics, Department of Biological Chemistry, and Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, United States.

[3.] Department of Neurology, University of California Los Angeles, Los Angeles, CA, USA.

Down syndrome (DS) is the most common form of genetic intellectual disability, in 1/700 newborns, presenting with cognitive, learning, memory, and language deficits. DS results from trisomy of chromosome 21(hsa21, T21) causing impaired cortical development and altered cell composition and brain architecture. However, the molecular mechanisms driving these pathologies remain unclear. We hypothesize that increased dosage of hsa21-encoded transcription factors (TFs) and epigenomic regulators changes global gene expression in neural progenitors, changing neural cell fate specification and differentiation. Here, we profiled 42k cells from 16 age/sex-matched DS and neurotypical (NT) neocortices via snMulti-ome technology. Using unbiased clustering, we identify 20 cell types in NT and DS samples, respectively. We observe an increase in the proportion of interneurons and upper layer neurons in DS, as compared to NT neocortex, consistent with DS adult cortex (Palmer et. al., 2022). We also identify cell-specific changes in transcriptomic profiles and differentially accessible regions of chromatin in developing DS brains. Together, these data have identified potential molecular mechanisms that are disrupted during DS development, which may provide a foundation to developing therapeutic targets in the future.

38. How does epigenetic heritability in B-cell fate decisions impact the antibody repertoire?

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

[1] Institute for Quantitative and Computational Biosciences,

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

[3] Bioinformatics Interdepartmental Program,

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

During immune responses, B-cells generate an antibody repertoire, which is characterized by its depth and breadth. Classical immunology describes the mechanism of repertoire generation as a Darwinian process of positive selection based solely on the genetically encoded B-cell receptor affinity for antigen. However, B-cells differ substantially in their epigenetic stimulus-response capacity, and this is remarkably heritable. How this epigenetic heritability impacts the antibody repertoire remains unknown. To address this question, we first develop a mathematical model of B-cell fate decisions within the cycles of the Darwinian positive selection process. Simulations show that the degree of epigenetic heritability vs stochastic change may determine the depth and breadth of the antibody repertoire. We then develop a workflow for measuring epigenetic heritability with live-cell imaging and automated image analysis. We anticipate that B-cells from old or sick mice show alterations in their epigenetic heritability, thereby leading to suboptimal antibody repertoires.

39. scReadSim: a single-cell RNA-seq and ATAC-seq read simulator

Guanao Yan[1], Jingyi Jessica Li[1,2,3,4]

[1] Department of Statistics,

[2] Department of Human Genetics,

[3] Department of Computational Medicin,

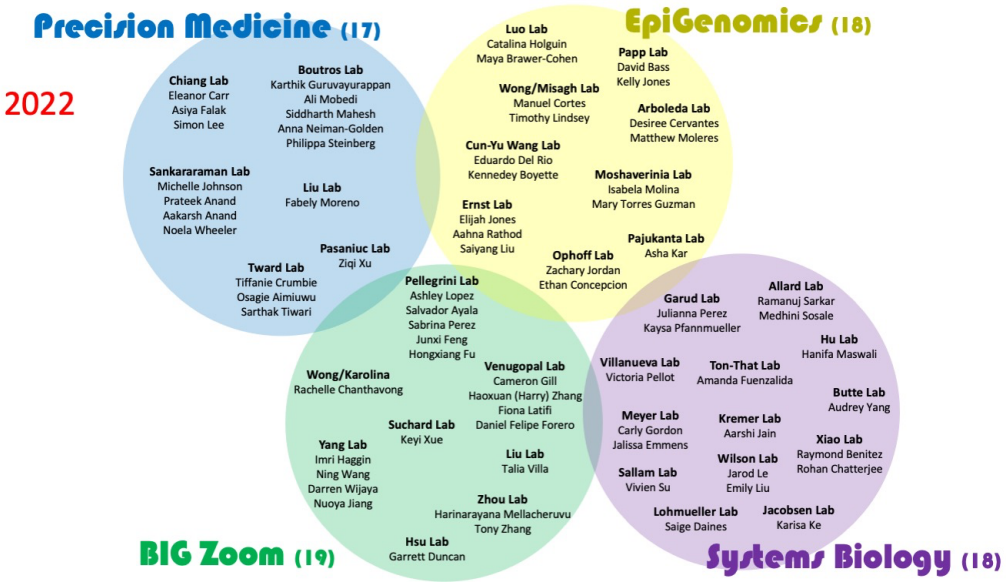
[4] Department of Biostatistics,

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

Rapid advances of single-cell RNA-seq and ATAC-seq technologies propelled the development of many computational tools, whose benchmarking demand realistic simulators. However, few simulators can generate sequencing reads, and none of the existing read simulators aim to mimic real cells, hurdling the benchmarking of low-level computational tools that process reads. To fill this gap, we propose scReadSim, a single-cell RNA-seq and ATAC-seq read simulator that generates synthetic cells to mimic real cells. Trained on real data, scReadSim can generate synthetic data in FASTQ and BAM formats. Deploying scReadSim on sci-ATAC-seq and 10x Multiome (ATAC+RNA) data, we show that the scReadSim synthetic data resemble the real data at both read and count levels. As a flexible simulator, scReadSim provides unique molecular identifier (UMI) counts for benchmarking scRNA-seq deduplication tools, and scReadSim can accommodate user-specified open chromatin regions ("ground truths'") to generate single-cell ATAC-seq data. Our benchmark applications of scReadSim show that UMItools is a preferred scRNA-seq deduplication tool, and MACS3 achieves top performance in scATAC-seq peak calling. Moreover, scReadSim can guide experimental design by allowing the cell number and sequencing depth to vary.

A B.I.G. Thank You

To all mentors for a successful 2022 Program!



- June 22 to August 12, 2022
- 8th Annual Program
- 357 applications received
- 73 admitted students
- 34 faculty mentors
- 53 direct research mentors

Talks, Posters and Abstracts Posted on the QCBio Website at: https://qcb.ucla.edu/big-summer/big2022/

B.I.G. SUMMER – Bruins-In-Genomics

Bruins-In-Genomics (B.I.G.) Summer Research Program is an 8-week full-time immersion program for undergraduates interested in learning how to read and analyze genes and genomes. Through this program students have the opportunity to experience graduate-level-cutting-edge research in UCLA laboratories and learn some of the latest research methods to solve real-world problems.
Please visit our website to learn more: https://qcb.ucla.edu/big-summer/