

UCLA

Institute for Quantitative & Computational Biosciences

7th ANNUAL RETREAT
QCBIO UCLA
SEPTEMBER 2021

Welcome

The QCB Community extends a warm welcome to all. Our annual Retreat is an opportunity to welcome new faculty, postdocs, and students, but also to take stock of where we are, to set goals, and define agendas.

I am really very proud that we as a community not only adapted to the challenges of the COVID pandemic, but excelled in several important ways: Our labs not only transitioned to remote research, meetings, and seminars, but many tackled important and diverse research questions to address the biology, the public health, and public policy challenges of the pandemic. Computational Medicine was in fact key to the development of SwabsSeq, UCLA's official COVID test. The Collaboratory stepped up to provide skills-focused workshops to a record number of workshop participants, and many of us engaged experimentalists who were shut out from their wet labs.

We also managed to hold the B.I.G. Summer Undergraduate Research Program this year with a record of 90 students, hosted by 43 laboratories! The Program was a remarkable success for the QCB community and a model for UCLA. BIG Thank Yous to our student, postdoc, staff, and faculty mentors, and Computational Medicine for being a great collaborator in this endeavor!

This year we also expanded our career development offerings to our postdocs. Thanks to Noa Pinter-Wollman who organized a well-appreciated workshop series to demystify the academic job search. We hope to offer that again in the coming year.

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 most experimentalists are spending more and more of their time on computational data analysis, database queries and modeling. The consequences of this revolution cannot be overstated. Leading research universities must transform their research training programs at postdoctoral, graduate and undergraduate levels, and UCLA is ahead many of them.

UCLA is in fact brimming with graduate training and undergraduate research opportunities in quantitative and computational biosciences across its multitude of departments. Last year we launched the UCLA Graduate Programs Navigator in Computational Biosciences, and expanded the Undergraduate Research Portal to connect potential mentors with eager undergraduate students, an amazing human resource for driving research forward and for training future leaders.

Those future leaders should be more diverse, as diverse questions, approaches, viewpoints are rooted in diverse racial, ethnic, and cultural experiences. This is critical for quantitative and computational biosciences, which – in this regard – lags behind other biosciences disciplines. With our new 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. Many thanks to Noa Pinter Wollman for serving as QCBio Equity Advisor, and I urge you to contact her if you want to get involved or have questions.

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 again; we can support your affinity group meeting; we will host career panels; and we're eager to support your workshop ideas, or other initiatives. QCB is here to support you!

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

Agenda

COFFEE, TEA & BAGELS

10:00 a.m. **WELCOME**

10:10 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

10:35 a.m. **SESSION I**

KEYNOTE I

- **Amjad Askary**, Assistant Professor, Department of Molecular, Cell and Developmental Biology

11:05 a.m. **SELECTED TALKS**

- **Leah Briscoe**, Bioinformatics PhD student, Garud lab
- **Karthik Gangavarapu**, Postdoc, Suchard lab

11:40 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

12:15 p.m. **LUNCH**

1:30 p.m. **SESSION II**

KEYNOTE II

- **Michael Wells**, Assistant Professor, Department of Human Genetics

2:00 p.m. **SELECTED TALKS**

- **Christa Caggiano**, Bioinformatics PhD student, Zaitlen lab
- **Sarah Spendlove**, Bioinformatics PhD student, Arboleda lab

2:35 p.m. **STATUS REPORTS III**

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

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

3:15 p.m. **SESSION III**

KEYNOTE III

- **Jennifer Wilson**, Assistant Professor, Department of Bioengineering

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

- **Farnaz Mohammadi**, Bioengineering PhD student, Meyer lab
- **Mingtao Xia**, Mathematics PhD student, Chou lab

4:20 p.m. **QBIO-EDGE PRESENTATION (Empowering Diversity and Growth in Education)**

CONCLUDING REMARKS

4:45 p.m. **RECEPTION & REFRESHMENTS**

POSTER SESSION

Keynote Speakers



Amjad Askary

Assistant Professor, Department of Molecular, Cell and Developmental Biology

Dr. Amjad Askary is an Assistant Professor at UCLA Department of Molecular, Cell and Developmental Biology. He studies cell fate specification in the mammalian retina using technologies that he has developed for synthetic recording and in situ readout of lineage and cellular history. Amjad received his undergraduate and Master's degree in Biotechnology and Bioprocess Engineering from University of Tehran in Iran and his PhD in Genetics, Molecular and Cellular Biology from University of Southern California. Most recently, he completed his postdoctoral fellowship, working with Dr. Michael Elowitz at Caltech.

Quantifying developmental variability by imaging lineage and gene expression

Systems-level understanding of cell fate decisions relies on our ability to quantitatively probe the molecular state and lineage history of individual cells. Progress in this area has been hampered by limitations of the existing methods. Traditionally, there has been a tradeoff between sequencing-based methods that provide a high dimensional profile of the cells but disrupt their spatial organization and imaging-based methods that capture native context of the cells but do not scale as well. Recent progress in spatial transcriptomics has started to bridge this gap. Here I present our approach to scalable imaging based genetic barcoding that enables decoding of high diversity combinatorial barcode libraries by microscopy. I show how we are applying this strategy to map the clonal composition of the mammalian retina. I will then discuss the ongoing directions in the lab to investigate cell fate specification in the retina by combining in situ gene expression profiling, lineage recording, and conditional perturbations.



Michael Wells

Assistant Professor, Department of Human Genetics

Michael F. Wells, PhD is an Assistant Professor in the UCLA Department of Human Genetics. He earned a B.S. in Biological Sciences from the University of Notre Dame in 2008, and a PhD in Neurobiology from Duke University in 2015 under the guidance of Dr. Guoping Feng. In 2021, he completed his postdoctoral training in the laboratory of Dr. Kevin Eggan at Harvard University and the Broad Institute. Michael's research focuses on discovering the disease mechanisms underlying neurodevelopmental disorders of genetic and viral origin using human stem cell-derived neural models and cerebral organoids. His work has been published in such high-impact journals as *Nature*, *Cell*, *Neuron*, and *Cell Stem Cell*, and has been funded by a F31 Predoctoral Fellowship, a K99/R00 Pathway to Independence award, and a Burroughs Wellcome Fund Postdoctoral Enrichment Program award. Outside of the laboratory, Michael serves as the creator and co-director of the COVID-19 National Scientist Volunteer Database (covid19sci.org), which is a resource for health officials

and decision-makers around the country looking to solve COVID-19-related problems. In addition, he is the Chair of the Society for Neuroscience (SfN) Trainee Advisory Committee, has previously participated in the SfN Early Career Policy Ambassador program. For more information, please visit MichaelFWellsPhD.com.

Exploration of human genetic and phenotypic diversity through cell villages

Our species is characterized by an immense diversity in neurological and psychological traits. Common and rare genetic variants have been linked to trait differences and disease risk in human populations, though the underlying biology is poorly understood and difficult to study at large scales. In this presentation, I will first describe a novel experimental platform that enables high-throughput investigations into the influence of human genetic variation on the earliest stages of brain development. This system, known as a "cell village" captures genetic, molecular, and phenotypic heterogeneity in a shared *in vitro* environment, thus facilitating the detection of relationships among human alleles, gene expression, and cellular behaviors. I will then describe how we used cell villages to identify a single nucleotide polymorphism in the *IFITM3* gene that could explain over half of the variance in neural progenitor cell susceptibility to the Zika virus, which is a pathogen that causes severe neurodevelopmental disorders. I will conclude with a brief discussion on the future of the village approach and how I plan to deploy this technology to investigate the role of progenitor cell competition in neurodevelopment and disease in my laboratory at UCLA.



Jennifer Wilson

Assistant Professor, Department of Bioengineering

Dr. Jennifer L. Wilson is an Assistant Professor at the UCLA Department of Bioengineering. In the Computational Systems Pharmacology Lab, Dr. Wilson studies how proteins downstream of drug targets affect drug-induced phenotypes – the ability to mitigate disease or cause side effects. The lab aims to develop engineering principles for rationally designing novel drug targets by accounting for downstream protein effects. Prior to coming to UCLA, Dr. Wilson earned a B.S. in Biomedical Engineering at the University of Virginia, she was an NSF Graduate Fellow with Doug Lauffenburger at M.I.T., and recently completed a CERSI fellowship in Regulatory Science (with Russ Altman) and SPARK fellowship (with Kevin Grimes) at Stanford University.

Modeling the multi-input, multi-output behavior of drug therapies using protein-protein interaction networks

A crucial problem in therapeutics is that drugs generally bind multiple proteins and affect multiple phenotypes, which are related to both drug safety and efficacy. However, my view is that this "bug" can be turned into a "feature" by embracing the multi-input, multi-output (MIMO) nature of drug therapies. Already drug therapies can affect multiple phenotype "outputs" - including disease symptoms or unintended side-effects - and several studies support that drugs bind several proteins ("inputs") within the cell. However, we don't yet rationally design drug therapies with these effects in mind. In the Lab for Understanding Network Effects (LUNE), we are using protein-protein interaction network models to better understand drug MIMO effects with the aim of eventually applying these methods predictively to identify new drug targets for untreated diseases. In this talk, I will highlight the extent to which protein-protein interaction networks describe drug effects and open questions about this modeling approach.

Selected Talks

* Source tracking using single nucleotide genetic signatures in the microbiome

Leah Briscoe^{1*}, Eran Halperin^{2,3,4,5,6}, Nandita R. Garud^{3,7}

¹Bioinformatics Interdepartmental Program, UCLA - ²Department of Computer Science, UCLA - ³Department of Human Genetics, UCLA - ⁴Department of Computational Medicine, UCLA - ⁵Department of Anesthesiology and Perioperative Medicine, UCLA - ⁶Institute of Precision Health, UCLA - ⁷Department of Ecology and Evolutionary Biology, UCLA

Microbiomes are communities of hundreds of species. Elucidating the colonization sources of these community members has been of great interest in the field. While emphasis has been placed on source-tracking using species composition, single nucleotide variants (SNVs) within species may be more precise for source-tracking due to their higher resolution. However, to date, SNV frequencies have not been leveraged for source-tracking. We assess the ability of SNVs versus species in a previously designed source-tracking algorithm FEAST (Shenhav et al. 2019) and find that SNVs can more accurately identify sources. Moreover, recapitulating previous findings, mother-infant species similarity increases with infant age, while SNV similarity decreases, indicating that sources other than mother contribute to the developing microbiome. Finally, with SNV FEAST we track migration of microbes across oceanic regions, including across the Suez and Panama canals. In sum, SNV FEAST reveals a more nuanced and complete picture of source-tracking than species FEAST.

* Identification of ancestry-specific health risks in a large cosmopolitan biobank

Christa Caggiano^{1,2}, Ruhollah Shemirani³, Alec Chiu¹, Ruth Johnson⁴, Valerie Arboleda^{5,6}, Gillian M. Belbin⁷, Noah Zaitlen^{2,8}

¹Interdepartmental Program in Bioinformatics, University of California, Los Angeles, CA - ²Department of Neurology, University of California, Los Angeles, CA - ³Information Science Institute, University of Southern California, Marina del Rey, California, USA - ⁴Department of Computer Science, University of California, Los Angeles, CA - ⁵Department of Pathology and Laboratory Medicine, University of California, Los Angeles, CA - ⁶Department of Human Genetics, University of California, Los Angeles, CA - ⁷Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA - ⁸Department of Computational Medicine, University of California, Los Angeles, CA

Genetic ancestry is a key factor that affecting disease risk. However, scientists often rely on race/ethnicity as noisy measures of ancestry, potentially leading to health disparities. To address this, we studied a large biobank of genotype and electronic health record (EHR) data. We called pairwise identity-by-descent (IBD) and then clustered individuals into genetic communities based on shared IBD. Many of these communities are understudied, such as the Persians or Armenians. We then used EHR data to estimate population-specific disease risks. We observed several well-known associations, including an elevated risk for Crohn's Disease in Ashkenazi Jews (LR p=3.82 x 10⁻⁸, OR 2.0 95% CI: 0.60-3.40), along with novel associations, like an increased risk for pruritus in Persians (LR p=8.43 x 10⁻⁴; OR: 2.5, 95% CI: 0.80-4.2). These results demonstrate the value of using genetic ancestry in precision medicine.

* Emergence of an early SARS-CoV-2 epidemic in the United States

Mark Zeller^{1*}, Karthik Gangavarapu^{1*}, Catelyn Anderson^{1*}, Allison R. Smither², John A. Vanchiere³, Rebecca Rose⁴, Daniel J. Snyder^{5,6}, Gytis Dudas⁷, Alexander Watts^{8,9}, Nathaniel L. Matteson¹, Refugio Robles-Sikisaka¹, Maximilian Marshall¹⁰, Amy K. Feehan¹¹, Gilberto Sabino-Santos Jr^{12,13}, Antoinette R. Bell-Kareem², Laura D. Hughes¹⁴, Manar Alkuzweny¹, Patricia Snarski^{15,16}, Julia Garcia-Diaz¹¹, Rona S. Scott¹⁷, Lilia I. Melnik², Raphaëlle Klitting¹, Michelle McGraw¹, Pedro Belda-Ferre^{18,19}, Peter DeHoff²⁰, Shashank Sathe^{21,22}, Clarisse Marotz^{18,23}, Nathan D. Grubaugh²⁴, David J. Nolan⁴, Arnaud C. Drouin^{2,16}, Kaylynn J. Genemaras^{2,25}, Karissa Chao^{2,25}, Sarah Topol²⁶, Emily Spencer²⁶, Laura Nicholson²⁶, Stefan Aigner^{21,22,27}, Gene W. Yeo^{21,22,27}, Lauge Farnaes²⁸, Charlotte A. Hobbs²⁸, Louise C. Laurent²⁰, Rob Knight^{18,19,29,30}, Emma B. Hodcroft³¹, Kamran Khan^{8,9,32}, Dahlene N. Fusco^{12,33}, Vaughn S. Cooper^{5,6}, Phillipe Lemey^{34,35}, Lauren Gardner¹⁰, Susanna L. Lamers⁴, Jeremy P. Kamil¹⁷, Robert F. Garry^{2,36}, Marc A. Suchard^{37,38}, Kristian G. Andersen^{1,26}

*These authors contributed equally to this work

[†]Senior authors

¹ Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037, USA - ² Department of Microbiology and Immunology, School of Medicine, Tulane University, New Orleans, LA 70112, USA - ³ Department of Pediatrics, Louisiana State University Health Sciences Center - Shreveport, Shreveport, LA 71130, USA - ⁴ BioInfoExperts LLC, Thibodaux, Louisiana, USA - ⁵ Department of Microbiology and Molecular Genetics, University of Pittsburgh, School of Medicine, Pittsburgh, PA, 15219, USA - ⁶ Center for Evolutionary Biology and Medicine, University of Pittsburgh, Pittsburgh, PA, 15219, USA - ⁷ Gothenburg Global Biodiversity Centre (GGBC), Gothenburg, Sweden - ⁸ Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada - ⁹ Bluedot, Toronto, Canada - ¹⁰ Department of Civil and Systems Engineering, Johns Hopkins University, Baltimore, MD, USA - ¹¹ Ochsner Clinic Foundation, New Orleans, Louisiana, USA - ¹² Department of Tropical Medicine, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA 70112, USA - ¹³ Centre for Virology Research, Ribeirão Preto Medical School, University of Sao Paulo, Ribeirão Preto, SP 14049900, Brazil - ¹⁴ Department of Integrative, Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA - ¹⁵ Heart and Vascular Institute, John W. Deming Department of Medicine, School of Medicine, Tulane University, New Orleans, LA 70112, USA - ¹⁶ Department of Physiology, Tulane University School of Medicine, New Orleans, LA 70112, USA - ¹⁷ Department of Microbiology and Immunology, Louisiana State University Health Science Center Shreveport, Shreveport, LA 71103, USA - ¹⁸ Department of Pediatrics, School of Medicine, University of California San Diego, La Jolla, California, USA - ¹⁹ Center for Microbiome Innovation, Jacobs School of Engineering, University of California San Diego, La Jolla, California, USA - ²⁰ Department of Obstetrics, Gynecology, and Reproductive Science, University of California, San Diego, La Jolla, CA 92037, USA - ²¹ Department of Cellular and Molecular Medicine, University of California at San Diego, La Jolla, California 92093, USA - ²² Stem Cell Program, University of California San Diego, La Jolla, CA 92093, USA - ²³ Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, USA - ²⁴ Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, 06510, USA - ²⁵ Bioinnovation Program, Tulane University, New Orleans, LA 70118, USA - ²⁶ Scripps Research Translational Institute, La Jolla, CA 92037, USA

²⁷ Institute for Genomic Medicine, University of California, San Diego, La Jolla, California, USA - ²⁸ Rady Children's Institute for Genomic Medicine, San Diego, CA 92123, USA; Rady Children's Hospital, San Diego, CA 92123, USA - ²⁹ Department of Computer Science and Engineering, Jacobs School of Engineering, University of California San Diego, La Jolla, California, USA - ³⁰ Department of Bioengineering, University of California San Diego, La Jolla, California, USA - ³¹ Biozentrum, University of Basel, Basel, Switzerland - ³² Department of Medicine, University of Toronto, Toronto, Canada - ³³ Department of Medicine, Tulane University School of Medicine, 1430 Tulane Avenue, New Orleans, LA, 70114, USA - ³⁴ Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Belgium - ³⁵ Global Virology Network - ³⁶ Zalgen Labs LLC, Germantown, MD, USA - ³⁷ Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA 90095, USA - ³⁸ Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA

The emergence of the COVID-19 epidemic in the United States (U.S.) went largely undetected due to inadequate testing. New Orleans experienced one of the earliest and fastest accelerating outbreaks, coinciding with Mardi Gras. To gain insight into the emergence of SARS-CoV-2 in the U.S. and how large-scale events accelerate transmission, we sequenced SARS-CoV-2 genomes during the first wave of the COVID-19 epidemic in Louisiana. We show that SARS-CoV-2 in Louisiana had limited diversity compared to other U.S. states, and that one introduction of SARS-CoV-2 led to almost all of the early transmission in Louisiana. By analyzing mobility and genomic data, we show that SARS-CoV-2 was already present in New Orleans before Mardi Gras, and the festival dramatically accelerated transmission. Our study provides an understanding of how superspreading during large-scale events played a key role during the early outbreak in the U.S. and can greatly accelerate epidemics.

* A lineage tree-based hidden Markov model to quantify cellular heterogeneity and plasticity

Farnaz Mohammadi¹, Shakthi Visagan¹, Sean M. Gross², Luka Karginov³, JC Lagarde¹, Laura M. Heiser², Aaron S. Meyer¹

¹Department of Bioengineering, University of California, Los Angeles - ²Department of Biomedical Engineering and Knight Cancer Institute, Oregon Health and Science University, Portland - ³Department of Bioengineering, University of Illinois, Urbana Champaign

Cell plasticity operates alongside other sources of cell-to-cell heterogeneity such as genetic mutations and variation in signaling which prevent most cancer therapies from being curative. The predominant methods of quantifying tumor-drug response operate on snapshot population-level measurements and therefore lack evolutionary dynamics. Here we apply a tree-based hidden Markov model (tHMM) to learn the characteristic patterns of single cell heterogeneity and state transitions. This model enables single cell classification based on the phenotype of individual cells and their relatives for improved specificity in pinpointing the dynamics of variability in drug response. We benchmarked our model using synthetic data of cell-cycle phase-specific cell fate and lifetime and demonstrated that the model successfully classifies cells within experimentally tractable dataset sizes. We analyzed experimental measurements of the same phenotypes in AU565 cells and identified 6 and 5 distinct subpopulations in lapatinib and gemcitabine-treated cells, and how much these drugs induce cell arrest and apoptosis in sub-populations.

* Polygenic risk score analysis of congenital heart disease phenotypes: elucidating the contribution of common variants to congenital heart defects.

Sarah Spendlove^{1,2}, Leroy Bondhus^{2,3}, Jae Hoon Sul^{1,4}, Valerie A. Arboleda^{1,2,3,5,6}

¹Interdepartmental Bioinformatics Program, David Geffen School of Medicine, UCLA - ²Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA - ³Department of Human Genetics, David Geffen School of Medicine, UCLA - ⁴Department of Psychiatry, David Geffen School of Medicine, UCLA - ⁵Division of Cardiology, Department of Medicine, David Geffen School of Medicine, UCLA - ⁶Department of Computational Medicine, David Geffen School of Medicine, UCLA

Congenital heart disease (CHD) occurs in ~1% of live births. CHD monogenic mutations are not completely penetrant and variably expressive. Thus polygenic inheritance may be important. Polygenic risk scores (PRS) aggregate effects of common variants across the genome to investigate whether they are associated with disease risk/severity. Using three CHD-related genome-wide association study (GWAS) summary statistics from the UK Biobank (UKBB) as our base study, and whole genome sequencing data from the CHD cohort ($n_1=711$ trios, $n_2=362$ European trios) of the Gabriella Miller Kids First dataset as our target study, we developed PRS for CHD. PRS estimated using the heart valve problem or heart murmur GWAS help explain variance in case-control status of CHD (all SNVs $p=0.00809919$, fetal cardiac $p=0.00819918$), and in severity of CHD (fetal cardiac $p=0.00769923$), showing that common genetic variants help explain the genetic basis of CHD of unknown genetic origin and differences in severity and penetrance.

* Controlling epidemics through optimal allocation of test kits and vaccine doses across networks

Mingtao Xia¹, Lucas Böttcher^{2,3}, Tom Chou^{1,2}

¹ Department of Mathematics, UCLA - ² Department of Computational Medicine, UCLA - ³ Frankfurt School of Finance and Management

Efficient testing and vaccination protocols are critical aspects of epidemic management. To study the optimal allocation of limited testing and vaccination resources in a heterogeneous contact network of interacting susceptible, recovered, and infected individuals, we present a degree-based testing and vaccination model for which we use control-theoretic methods to derive optimal testing and vaccination policies using control-theoretic methods. Within our framework, we find that optimal intervention policies first target high-degree nodes before shifting to lower-degree nodes in a time-dependent manner. Using such optimal policies, it is possible to delay outbreaks and reduce incidence rates to a greater extent than uniform and reinforcement-learning-based interventions, particularly on certain scale-free networks.

Welcome our Incoming Bioinformatics Students!



Kristin Boulier
BS, Molecular Biology
John Hopkins University
MD, Medicine
University of Virginia



Sandy Kim
BS, Computational Biology
MS, Bioinformatics
UCLA
B.I.G. SUMMER 2019 ALUMNA



Michael Cheng
BS, Molecular, Cell & Developmental
Biology
UCLA
B.I.G. SUMMER 2020 ALUMNUS



Sandra Lapinska
BS, Biometry and Statistics
MPS, Applied Statistics
Cornell University
B.I.G. SUMMER 2020 ALUMNA



Jack Dodson
BS, Genomics and Molecular Genetics
Michigan State University



Terence Li
BS, Quantitative Biology
MS, Quantitative and Computational
Biology
USC



Maria Flores
BS, Cell and Molecular Biology
San Francisco State University



Francesco Musella
BS, Physics
MS, Physics
University of Naples



Alexander Flynn-Carroll
BS, Ecological Science
BS, Computational Methods in Ecology
University of Edinburgh



Jieun "Grace" Oh
BS, Biochemistry
University of Hong Kong



Aditya Gorla
BS, Bioengineering
BA, Economics
UCLA



Aditya Pimplaskar
BS, Computational & System Biology
UCLA



Timothy Hamilton
BS, Molecular Cellular and Developmental
Biology
UCLA



Helena Winata
BS, Biophysics
University of British Columbia



Chanyue "Charlotte" Hu
BS, Computer Science
UC Irvine



Ryo Yamamoto
BA, Computer Science
UC Berkeley

Welcome our Incoming Medical Informatics Students!



Kaiyang "Victor" Cheng
BA, Computer Science
UC Berkeley



Saarang Panchavati
BS, Bioengineering and EECS
UC Berkeley



Elizabeth "Liz" Hutchins
BA, Neuroscience and Behavior
MS, Human Nutrition
MD, Medicine
Columbia University



Mara Pleasure
BA, Human Biology
BA, Media Arts & Practice
USC



Patrick "Patty" Liu
BA, Neuroscience and Economics
Pomona College
MPH, Global Health Metrics
University of Washington



Nathan Siu
BS, Bioinformatics
Davidson College
B.I.G. SUMMER 2020 ALUMNUS



Carlos Olivares
BA, Mathematics
University of Chicago



Selina Wu
BS, Biochemistry
Boston College

Welcome our Incoming Biomathematics Student!



Siqi Fang
BA (Mathematics) and MS (Systems Biology)
Cambridge

Welcome our Incoming Genetics & Genomics Students!



Jasmine Amerasekera
BA-Integrative Biology
UC Berkeley



Laila Sathe (will officially start in the Fall)



Chloe Hanson
BA-Environmental Science, MS-Integrative Biology
Oregon State University



Brandon Tsai (MSTP)
BS-Microbiology, Immunology and Molecular
Genetics
UCLA



Aina Martinez
BS-Physics
MIT



Michael Wasney
BA-Biological Sciences and English Language and
Literature
University of Chicago

Collaboratory Fellows 2021-2022



<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>

1. International COVID-19 Mortality Forecast Visualization: covidcompare.io

Samir Akre^{1,2}, Patrick Y. Liu¹, Joseph R. Friedman^{1,3}, Alex T. Bui^{1,2}

¹ Medical Informatics Home Area, University of California, Los Angeles, Los Angeles, USA - ² Department of Radiological Sciences, University of California, Los Angeles, Los Angeles, USA - ³ Center for Social Medicine and Humanities, University of California, Los Angeles, Los Angeles, USA

COVID-19 mortality forecasting models provide critical information about the trajectory of the pandemic, which is used by policymakers and public health officials to guide decision-making. However, thousands of published COVID-19 mortality forecasts now exist, many with their own unique methods, assumptions, format, and visualization. As a result, it is difficult to compare models and understand under which circumstances a model performs best. Here, we describe the construction and usability of covidcompare.io, a web tool built to compare numerous forecasts and offer insight into how each has performed over the course of the pandemic. From its launch in December 2020 to June 2021, we have seen 4,600 unique visitors from 85 countries. A study conducted with public health professionals showed high usability overall as formally assessed using a Post-Study System Usability Questionnaire (PSSUQ). We find that covidcompare.io is an impactful tool for the comparison of international COVID-19 mortality forecasting models.

2. Emergent Ecological Phenomena in Migrating and Evolving Populations

Casey Barkan¹ and Shenshen Wang¹

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

Spatial variations in an environment can have a major impact on the evolutionary dynamics of species that inhabit it. The degree to which spatial variations affect evolution depends on the rate at which organisms migrate through their heterogeneous environment. We study an evolutionary model of a species that can migrate through a network of habitats, each of which exerts unique selection pressure on the organisms. We find that distinct *quasi*-species evolve and compete with one another for resources. As migration rates are varied, some quasi-species go extinct, and others emerge. At each point of extinction or emergence a phase transition occurs and the system exhibits critical slowing down—a phenomenon known to often accompany ecological collapse such as extinction. Our work may be applicable to the evolution of human microbiomes, evolution of antibiotic-resistant bacteria, and selection of high-affinity antibodies during an immune response.

3. Intratumoral heterogeneity of gliomas is linked to heterogeneity of microenvironmental stimuli

Nicholas A. Bayley^{1,2}, Henan Zhu¹, Christopher Tse¹, Jenna Minami¹, Weihong Yan³, Timothy F. Cloughesy^{1,4}, Linda M. Liau^{1,5}, Thomas G. Graeber^{1*}, and David A. Nathanson^{1*}

¹ Department of Molecular and Medical Pharmacology, ² Bioinformatics Interdepartmental Program, ³ Department of Chemistry and Biochemistry, ⁴ Department of Neurology, David Geffen School of Medicine, ⁵ Department of Neurosurgery, David Geffen School of Medicine, University of California Los Angeles, Los Angeles California, USA

* equal contribution from labs

Gliomas are malignant brain tumors characterized by intratumoral heterogeneity of transcriptional cellular states mirroring neurodevelopmental cell types: oligodendrocyte and neural progenitors (OPC, NPC), astrocytes (AC), and mesenchymal (MES) cells. The MES state is linked to immune cell activity, however the full extent of microenvironmental influence and functional relevance of cell states remain to be elucidated. Here we perform bulk and single cell RNA sequencing and lipidomic profiling of patient gliomas and derived model systems established orthotopically in murine brains and as cell cultures.

We find that cells grown *in vitro* concurrently decrease their OPC fraction and gene expression related to neuroglial interactions while dynamically altering their lipid metabolism. Consequently, OPC-dominant tumors with specific lipidomic profiles frequently fail to establish *in vitro* while succeeding *in vivo*. This work broadens connections between environmental stimuli and intratumoral heterogeneity and unveils lipid metabolic heterogeneity connected to tumor growth, providing candidate targets for therapeutic intervention.

4. Integrative Genome Modeling Platform reveals essentiality of rare contact events in 3D genome organizations

Lorenzo Boninsegna^{1,2}, Asli Yildirim^{1,2}, Guido Polles^{1,2}, Sofia A. Quinodoz³, Elizabeth Finn⁴, Mitchell Guttman³, Xianghong Jasmine Zhou⁵, 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 Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA, ⁴ National Cancer Institute, NIH, Bethesda, MD 20892, USA, ⁵ Department of Pathology, David Geffen School of Medicine, University of California Los Angeles, 10833 Le Conte Ave, Los Angeles, CA 90095, USA

A multitude of sequencing-based and microscopy technologies provide the means to unravel the relationship between the three-dimensional (3D) organization of genomes and key regulatory processes of genome function. However, it remains a major challenge to systematically integrate all available data sources to characterize the nuclear organization of genomes across different spatial scales. Here, we develop a multi-modal data integration approach to produce genome structures that are highly predictive for nuclear locations of genes and nuclear bodies, local chromatin compaction, and spatial segregation of functionally related chromatin. We demonstrate that multimodal data integration can compensate for systematic errors and missing values in some of the data and thus, greatly increases accuracy and coverage of genome structure models. Our study reveals the key contributions of low-frequency inter-chromosomal contacts to accurately predicting the global nuclear architecture. Overall, our results highlight the benefits of multi-modal data integration for genome structure analysis.

5. PharmOmics: A Species- and Tissue-specific Drug Signature Database and Gene Network-based Drug Repositioning Tool

Yen-Wei Chen^{1,2,6}, Graciela Diamante^{1,2,6}, Jessica Ding^{1,3,6}, Thien Xuan Nghiem¹, Jessica Yang¹, Sung-min Ha¹, Peter Cohn¹, Douglas Arneson^{1,4}, Montgomery Blencowe^{1,3}, Jennifer Garcia², Nima Zaghari², Paul Patel², and Xia Yang^{1,2,3,4,5,*}

¹ Department of Integrative Biology and Physiology, UCLA,

² Interdepartmental Program of Molecular Toxicology, UCLA

³ Interdepartmental Program of Molecular, Cellular, & Integrative Physiology, UCLA,

⁴ Interdepartmental Program of Bioinformatics, UCLA

⁵ Institute for Quantitative and Computational Biosciences, UCLA

⁶ These authors contributed equally.

Drug development has been hampered by a high failure rate in clinical trials due to efficacy or safety issues not predicted by preclinical studies. A key contributor is incomplete understanding of drug functions across organ systems and species. Therefore, elucidating species- and tissue-specific drug actions can provide insights into therapeutic efficacy and potential adverse effects. Here, we present a drug knowledgebase and analytical tool, PharmOmics, comprised of footprints of drugs in individual tissues from human, mouse, and rat transcriptome data from GEO, ArrayExpress, TG-GATEs, and DrugMatrix. Using gene expression signatures as indicators of drug functions, we implemented gene network-based approach for drug repositioning. We demonstrate the potential of PharmOmics to retrieve known therapeutic drugs, identify

drugs with tissue toxicity, and predict FDA approved drugs for new disease indications which can be experimentally validated. PharmOmics serves as a complementary tool to support drug characterization and network-based medicine.

6. An Autoencoder for Extracting Latent Attributes from Acute Myeloid Leukemia Proteomic Data

Jackson Chin¹, Aaron Meyer¹

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

When paired with proteomic data, ordinary differential equations (ODE) can form powerful mechanistic models. Such models, however, are limited in their application to clinical proteomics due to patient-to-patient differences; variation between patients often necessitates patient-individualized models that require excessive individual data. Here, we present the development of an autoencoder to manage such patient-to-patient differences by extracting the latent attributes present in proteomic data derived from Acute Myeloid Leukemia tissues. We perform a grid-search to identify optimal autoencoder parameters then validate our encoding by decoding the latent attributes and comparing our decoded results to the original proteomic data. We demonstrate that we can reduce the proteomic data to 30 latent attributes that, following decoding, deviate from the original data with a mean-squared error of 0.823. We then compare latent attributes between patients and demonstrate that subsets of our latent attributes correlate to patient survival.

7. Motif-based phosphoproteome clustering improves modeling and interpretation

Marc Creixell¹, Aaron S. Meyer

¹UCLA, Department of Bioengineering

Mass spectrometry-based analysis of kinase signaling can provide a global view of kinase signaling regulation but making sense of these data is complicated by its stochastic coverage of the proteome, measurement of substrates rather than kinase signaling itself, and the scale of the data collected. Here, we propose a combined data and motif clustering strategy (DDMC) that simultaneously clusters substrate peptides into similarly regulated groups based on their variation within an experiment and their sequence profile. We show that this can help to identify putative upstream kinases and supply more robust clustering. We apply this clustering to large-scale clinical proteomic profiling of 110 treatment-naïve lung adenocarcinoma (LUAD) tumors and 101 paired normal adjacent tissues (NATs) from the NCI's CPTAC LUAD study and identify conserved proteomic signatures of tumorigenicity, genetic mutations, and tumor immune infiltration. We built a mixture model that probabilistically clusters phosphosites based on both their peptide sequence and abundance across samples. Clustering both the sequence and abundance measurements ensures that the resulting clusters are a function of both features, which we hypothesized would provide both more meaningful and robust clusters. The resulting clustering provides coordinated output that can be used in a few different ways. The cluster centers, by virtue of being a summary for the abundance changes of these peptides, can be regressed against phenotypic responses to establish associations between particular clusters and response. In parallel or independently, one can interrogate the resulting PSSMs to describe the overall sequence features of that cluster. This can be compared to other information such as experimentally generated specificity profiles of putative upstream kinases via PSPL, to infer upstream kinases. We first benchmarked and verified that DDMC clearly outperforms the use of standard methods. We then confirmed that DDMC correctly predicts AKT1 and ERK2 as upstream kinases of signaling clusters containing their experimentally validated substrates. Additionally, we found that incorporating the sequence information into the clustering criterion improves prediction of different phenotypes and indeed optimizes for the information content of both sequence motifs and phosphorylation behaviors. Finally, we comprehensively characterized those clusters that mostly contribute to explaining sample type, STK11/EGFR/ALK mutational status, and tumor immune infiltration.

8. Asymmetric Branching Scale Factors as Features in Neuronal and Glial Cell-Type Classification Using Machine Learning Methods

Paheli Desai-Chowdhry¹, Alexander Brummer^{1,2}, Van Savage^{1,2,3}

¹Department of Computational Medicine, University of California Los Angeles, Los Angeles, California, United States of America,

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

³Santa Fe Institute, Santa Fe, New Mexico, United States of America

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. In previous work, we constructed a biophysical theory that establishes a correspondence between neuron structure and function as mediated by principles such as time or power minimization, using undetermined Lagrange multipliers to predict scaling ratios for axon and dendrite sizes across branching levels. Here, we use scale factors related to asymmetric branching as features in machine learning classification 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.

9. Massively Parallel Screen for 3'UTR Variants Regulating mRNA Abundance

Ting Fu^{1,3*}, Kofi Amoah^{2,3*}, Tracey Chan^{2,3}, Jae-Hoon Bahn³, Jae-Hyung Lee³, Rocky Cheung⁴, Sriram Kosuri⁴, Xinshu Grace Xiao^{1,2,3}

¹Molecular, Cellular and Integrative Physiology Interdepartmental Program, ²Bioinformatics Interdepartmental Program, ³Department of Integrative Biology and Physiology, ⁴Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095

* These authors contributed equally

Genome-wide association studies have revealed that most disease-associated genetic variants are in non-coding regions of the genome. Using massively parallel reporter assays (MPRAs), previous studies have characterized numerous non-coding variants for transcriptional regulation. Yet, it remains elusive whether the non-coding variants could affect the fate of the mRNA transcript. Functional variants altering the binding sites of micro-RNAs and RNA-binding proteins could affect gene expression and downstream pathways by regulating RNA localization, translation and stability. However, a systematic study of functional 3'UTR variants is still lacking. To fill in this gap, we developed a massively parallel screen for 3'UTR variants that may affect RNA abundance. With this platform, we analyzed 30,000 rare 3'UTR variants and identified 2,788 functional candidates that regulate mRNA abundance in HEK293 cells. Our findings illustrate the usability of our method to understand disease-relevant variants, identify causal variants and uncover the associated functional mechanisms of post-transcriptional regulation.

10. Viral proteins stoichiometry critically impacts membrane fusion levels: illustration with henipaviruses

Amandine Gamble¹, David Buchholz², Yao Yu Yeo², James O. Lloyd-Smith¹ and Hector C. Aguilar²

¹Department of Ecology and Evolutionary Biology, University of California Los Angeles, California, USA, ²Department of Microbiology and Immunology, College of Veterinary Medicine, Cornell University, Ithaca, New York, USA

To enter host cells, most enveloped viruses first attach to the cells, then fuse with the cell membrane. For some viruses, this process requires the intervention of two viral proteins: an attachment protein (binding cell receptors) and a fusion protein (triggering fusion pore formation). This is notably the case for several paramyxoviruses, including henipaviruses, which draw lots of attention due to their broad host ranges, and the high case fatality rates associated with Nipah and Hendra viruses in humans.

While the roles of henipaviral attachment (G) and fusion (F) proteins in membrane fusion are relatively well understood, the quantitative aspects of their interactions remain elusive. However, data have shown that the F:G expression ratio can critically impact fusion levels, suggesting that protein stoichiometry is an important modulator of membrane fusion. In this study, we aim at characterizing the relationship between viral protein stoichiometry and membrane fusion levels. First, in order to accurately quantify the F:G expression ratio, we developed a protocol to obtain comparable measures of the expression levels of G and F from immunofluorescent labeling data. Then, we measured the effect of the F:G expression ratio on membrane fusion across henipaviruses. Our data revealed a bell-shaped relationship between F:G expression ratio and membrane fusion, with an optimum between 1:1 and 10:1 depending on the virus. Finally, we used these results to explore different hypotheses regarding the henipavirus-induced membrane fusion cascade, notably with respect to protein polymerization and recycling. The results we present have important implications for the optimization of *in vitro* studies of henipaviruses, and considerably enrich our understanding of virus-induced cell fusion.

11. Modeling the heterogeneous 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. 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 NFκB distributions of signaling codons. Five of the six codon distributions were well fitted, but the ‘oscillation’ codon was not. We discuss next steps to improve the modeling fits further and leverage the model for gaining new insights about NFκB dynamics.

12. Evaluating the ability of PCA to remove topological noise from sc-RNA-seq and sc-RNA-seq analogue data

Timothy Hamilton, Shamus Cooley, Samuel Aragonés, J. Christian J. Ray and Eric J. Deeds.

Due to its computational efficiency and high interpretability, Principal Component Analysis has become widely used within the field of single-cell RNA sequencing (sc-RNA-seq) analysis. Broadly speaking, PCA has two main uses within the field. The first is to reduce the dimensionality of the preprocessed sc-RNA-sequencing, while preserving the relevant information, thus reducing the computation and memory requirements for the downstream analyses. The second is to remove extraneous variation or “noise” by only selecting directions within the data that have the highest variance and thus removing variance that exists orthogonally to those selected directions. While previous work has shown that PCA cannot faithfully represent high dimensional data in low-dimensions, to date, it has not been shown whether PCA can effectively remove noise from data. To test this, we synthetically added noise to both real and simulated data sets, and then measured the ability of PCA to “de-noise” the data and recover the topology of the data set before the noise was added. We found that, in general, PCA failed to “de-noise” the data effectively. The only exception to this is when noise levels are very small and are completely orthogonal to the underlying biological or “true” variation in the data. Most real-world data sets, and particularly scRNA-

seq, likely do not have noise that follows this pattern. As such, using PCA to “denoise” data in this manner risks introducing significant distortion into the data, distortion that may introduce bias into subsequent unsupervised analyses.

13. Hard sweeps are the dominant mode of adaptation on the X chromosome in *Drosophila melanogaster*

Mariana Harris¹, Nandita Garud^{2,3}

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

²Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles California, USA, ³Department of Human Genetics, University of California, Los Angeles, California, United States of America

Sex chromosomes in *Drosophila* are found in a hemizygous state in males. Thus, deleterious mutations may be purged more rapidly on the X chromosome than on the autosomes. This purging can result in less available standing variation that can seed adaptation on the X chromosome. Here, we investigate how differences in the recessivity of alleles on the X chromosome versus autosomes impact the mode and tempo of adaptation. Specifically, we test the hypothesis that hard sweeps, in which a single adaptive mutation rises to high frequency, are more common on the X chromosome, while soft sweeps, in which multiple haplotypes rise to high frequency simultaneously, are more common on the autosomes. We analyze *D. melanogaster* genomic data from two populations, North Carolina and Zambia, and find that indeed hard sweeps are more prevalent on the X chromosome and soft sweeps are more common on the autosomes.

14. Developmental Trajectory Analysis of Differentiating Mouse sensory Interneurons to Further Improve Stem-cell Differentiation Protocols

Eric Heinrichs^{1,2,*}, Sandeep Gupta^{2*}, Riki Kawaguchi³, Bennett G.

Novitch², Samantha J. Butler²

¹ Department of Human Genetics, ² Department of Neurobiology

³ Department of Neurology, University of California Los Angeles, Los Angeles, CA

* These authors contributed equally

The developing spinal cord is populated with a variety of interneuron types responsible for transmitting sensory information, e.g. pain, itch and heat. These interneurons are required to repair damaged sensory circuits in injured spinal cord. However, current directed differentiation protocols are not capable of creating pure populations of single dorsal interneuron (dI) class but rather create heterogeneous populations of dIs. Here we utilize single cell transcriptomics to profile the differentiation of mouse dI4, dI5, and dI6 neuronal populations at day 9 in an *in vitro* differentiation. Using a combination of Seurat and Monocle3 packages in R, we plot the developmental trajectories of differentiating interneurons and perform pseudotemporal analyses to investigate the genetic drivers behind different developmental bifurcations. The goal of these studies is to identify novel factors which will provide the specificity in differentiation protocols to direct the pure dI cell populations needed for regenerative therapies and drug testing platforms.

15. Scalable Spatial Cell Type Mapping of the Mouse Brain with dredFISH

Zachary E Hemminger¹, Gaby Sanchez Tam¹, Igor Nikolski¹, Jonathan Perrie², Roy Wollman^{1,3}

¹University of California Los Angeles, Chemistry and Biochemistry, Los Angeles, CA, ²University of California Los Angeles, Bioinformatics, Los Angeles, CA, ³University of California Los Angeles, Integrated

Physiology, Los Angeles, CA

Understanding a biological system’s function often requires determining its structure. The connection between an organ’s anatomy (i.e., structure) and physiology (i.e., function) has been a core principle in investigating tissue biology as well as pathology. A key aspect of an organ’s anatomy is the spatial localization of cell types within that tissue. Currently, the gold standards for spatially profiling cell types within tissues are based on individual RNA counting using either spatially

barcoded sequencing approaches or single molecule optical approaches. These approaches are limited to relatively thin tissue sections and the total area profiled is limited by time (optical methods) or by cost (sequencing methods). One approach to circumvent these limitations is to optically profile whole cells rather than single molecules. Here we present dimensionality reduced Fluorescent In Situ Hybridization (dredFISH), a high throughput optical method for spatially profiling the cell types of individual cells and demonstrate this approach in mouse coronal brain sections. Using the Allen Brain scRNAseq Atlas, we designed an oligo probe set which encodes rich cell type information in a compact aggregate measurement of a cell's gene expression which is measured optically in low magnification. By circumventing the need for individual gene calls, dredFISH eliminates the stringent requirements of single-molecule detection providing much-needed increase in throughput that will enable the timely creation of 3D whole organ cellular atlases essential to fully understand tissue biology.

16. A novel approach for Spliced-Transcriptome-Wide Associations (SplitWAS)

Jonatan Hervoso^{1,2}, Mudra Choudhury^{1,2}, Arjun Bhattacharya^{4,5,6}, Giovanni Quinones-Valdez^{2,3}, Bogdan Pasaniuc^{1,4,5,6*}, Grace Xiao^{1,2,6*}
¹Bioinformatics Interdepartmental Program, ²Department of Integrative Biology and Physiology, ³Department of Bioengineering, ⁴Department of Human Genetics, ⁵Department of Pathology and Laboratory Medicine, ⁶Institute for Quantitative and Computational Biology University of California Los Angeles, Los Angeles California, USA
 * Corresponding authors

Transcriptome-wide association studies (TWAS) have been instrumental in identifying genomic risk regions and putative risk genes linked to complex traits and diseases. However, these approaches focused heavily on the genetic contribution to total gene expression, which cannot capture the transcriptome complexity enabled by alternative RNA processing. Recent studies have shown that RNA splicing, one of the main mechanisms of alternative RNA processing, is a primary link between genetic variations and disease. Genetic loci underlying RNA splicing variations contribute similarly or more substantially to complex traits, compared to those that affect total gene expression levels. Here, we establish a new approach, namely, Spliced-Transcriptome-Wide Associations (SplitWAS), to identify significant splicing-trait associations. We applied our approach to impute exon usage into GWAS summary statistics of Schizophrenia from the PGC consortium using two RNA-seq datasets as reference panels. This study provided new insights into the genetic regulation of splicing and the contribution of splicing to Schizophrenia. Our results showcase the power of SplitWAS to gain insights into the genetic basis of complex traits.

17. Analysis of Rare Non-coding Variants in Psychiatric Disorders through Integration of Functional Annotations

Runjia (Luke) Li^{1,2}, Jason Ernst^{1,2,3,4,5,6,7}
¹Bioinformatics Interdepartmental Program, UCLA, CA, USA
²Department of Biological Chemistry, UCLA, CA, USA, ³Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA, CA, USA, ⁴Computer Science Department, UCLA, CA, USA, ⁵Department of Computational Medicine, UCLA, CA, USA, ⁶Jonsson Comprehensive Cancer Center, UCLA, CA, USA, ⁷Molecular Biology Institute, UCLA, CA, USA

A large portion of common disease-associated variants identified by genome-wide association studies (GWASs) are located within the non-coding genome, while the contributions from rare, non-coding variants remain largely unexplored. The advance in whole-genome sequencing (WGS) has enabled systematic discovery of rare non-coding variants. However, it remains difficult to characterize the functional consequences of these variants. Here we propose a supervised machine learning framework that integrates various functional annotations and predicts an individual's disease risk from its rare non-coding variants, while correcting for covariates. Applying our framework to the WGS in the African-American cohort consisting of 1483 bipolar disorder cases, 3006 schizophrenia cases and 2374 controls, we show that the current selection of functional annotations are nominally predictive of

disease risk and highlight the challenges imposed by the potential presence of sequencing artefacts.

18. L-GIREMI uncovers RNA editing sites in long-read RNA-seq

Zhiheng Liu¹, Giovanni Quinones-Valdez², Ting Fu³, Xinshu Xiao^{1,2,3}
¹Department of Integrative Biology and Physiology, ²Department of Bioengineering, ³Molecular, Cellular, and Integrative Biology Program University of California, Los Angeles, California, USA

Using third-generation sequencers, long-read RNA-seq is increasingly applied in transcriptomic studies given its major advantage in characterizing full-length transcripts. A number of methods have been developed to analyze this new type of data for transcript isoforms and their abundance. Another application, which is significantly under-explored, is to identify and analyze single nucleotide variants (SNVs) in the RNA. Identification of SNVs, such as genetic mutations or RNA editing sites, is fundamental to many biomedical questions. In long-read RNA-seq, SNV analysis presents significant challenges, due to the well-known high error rates of the third-gen sequencers. Here, we present the first study to detect and analyze RNA editing sites in long-read RNA-seq. Our new method, L-GIREMI, effectively handles sequencing errors and biases in the reads, and uses a model-based approach to score RNA editing sites. Applied to PacBio Iso-Seq data, L-GIREMI affords a high accuracy in RNA editing identification. In addition, the unique advantage of long-read allowed us to uncover novel insights about RNA editing occurrence in single molecules and transcript-specific editing. L-GIREMI provides a valuable means to study RNA nucleotide variants in long-read RNA-seq.

19. Flexible birth-death tree models with Markov random fields

Andrew F. Magee^{1,2}, Sebastian Höhna^{3,4}, Tetyana I. Vasylyeva⁵, Adam D. Leaché¹, Vladimir N. Minin⁶
¹Department of Biology, University of Washington, Seattle, California, USA, ²Department of Human Genetics, University of California, Los Angeles, Washington, USA, ³GeoBio-Center, Ludwig-Maximilians-Universität München, 80333 Munich, Germany, ⁴Department of Earth and Environmental Sciences, Paleontology & Geobiology, Ludwig-Maximilians-Universität München, Munich, Germany, ⁵Department of Zoology, University of Oxford, Oxford, United Kingdom, ⁶Department of Statistics, University of California, Irvine, Washington, USA

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.

20. A pan-cancer analysis of global and functional genomics of tumor proliferation

Alexander Markowitz^{1,2,3,4}, Vinayak Bhandari^{5,6}, Chenghao Zhu^{1,2,3,4}, Eric Huang⁵, Robert G. Bristow^{7,8,9}, Paul C Boutros^{1,2,3,4}
¹Department of Human Genetics, University of California, Los Angeles, CA, USA, ²Department of Urology, University of California, Los Angeles, CA, USA, ³Jonsson Comprehensive Cancer Center, University of California, Los Angeles, CA, USA, ⁴Institute for Precision Health, University of California, Los Angeles, CA, USA, ⁵Ontario Institute for Cancer Research, Toronto, Canada, ⁶Department of Medical Biophysics, University of Toronto, Toronto, Canada, ⁷Division of Cancer Sciences, Faculty of Biology, Health and Medicine, University of Manchester, Manchester, UK, ⁸The Christie NHS Foundation Trust, Manchester, UK, ⁹CRUK Manchester Institute and Manchester Cancer Research Centre, Manchester, UK, ¹⁰Department of Pharmacology and Toxicology, University of Toronto, Toronto, Canada, ¹¹Vector Institute for Artificial Intelligence, Toronto, Canada

Rapid proliferation is a hallmark of cancer and predicts poor prognosis in many tumor types. Surprisingly, it remains unclear how proliferation is associated with mutational processes, molecular and evolutionary features across cancer types. We evaluated proliferation across primary cancers and cell lines using multi-omic data from 11,597 primary tumors and 1,804 cell lines across six major consortia. We hypothesize that mutations in specific genes underlies cellular proliferation in primary tumors and cancer cell lines. We test these hypotheses by integrating public data from primary cancers and cancer cell lines with rigorous statistical methods. Preliminary analysis suggests rapidly proliferating tumors display elevated burden of all types of mutations including *TP53* and *MYC*, and more mutations were attributed to specific mutational signatures, suggesting underlying mechanisms. By evaluating proliferation at these molecular levels, key insights will be yielded including expansive new resources on the variation of proliferation across and within cancer types.

21. Harnessing the Power of Admixed Populations in GWAS

Rachel Mester¹, Kangcheng Hou², Arjun Bhattacharya³, Kathryn S. Burch², Bogdan Pasaniuc^{1,2,3,4}

¹Department of Computational Medicine, ²Bioinformatics Interdepartmental Program, ³Department of Pathology and Laboratory Medicine, ⁴Department of Human Genetics, University of California Los Angeles, Los Angeles California, USA

Admixed individuals have traditionally been excluded from genome-wide association studies due to the difficulty in attributing SNPs to particular ancestry backgrounds and the anticipated loss of power due to correction for genetic structure within analyses. Additionally, it is unknown how common effect size heterogeneity between ancestries is and how much of it can be attributed to statistical factors such as linkage disequilibrium and SNP imputation. In this project, we use simulated admixed genotypes and phenotypes to assess different methods of GWAS in admixed populations to determine (1) which methods are best suited for traits with various levels of heterogeneity in allelic effect sizes and frequencies of causal variants, and (2) whether different types of genetic architectures induce heterogeneity of allelic effect sizes between ancestries. By providing guidance on the sources and detection of effect size heterogeneity, we hope to enable an increased volume of research in admixed populations.

22. The kinetics of breakthrough SARS-CoV-2 infections

Dylan H. Morris, Amandine Gamble, James O. Lloyd-Smith

As SARS-CoV-2 spreads globally and vaccinations are rolled out, an increasing fraction of symptomatic infections are immune "breakthroughs": infections of the vaccinated or previously-infected. Breakthrough infections can be facilitated by virus antigenic evolution, by waning of immune protection, by improved virus within-host replication, by individual susceptibility, and by high exposure levels. Disentangling these causes is crucial for policy; policymakers need to plan vaccination programs to ensure durable transmission reduction and protection against severe disease. We develop mechanistic, quantitative models of the infection process—the interaction between exposure to virions and the immune response—parameterized from real-world data. We show that the effectiveness of proposed policies for reducing breakthrough rates depends strongly on which mechanisms are driving breakthrough infection. For example, original virus booster shots are excellent for addressing breakthroughs caused by antibody waning, but less effective at addressing breakthroughs caused by antigenic evolution. Our results provide guidance for policymakers and identify key gaps in existing data on breakthrough infections.

23. A De Novo Pathogenic Variant in *SETX* causes a rapidly progressive neurodegenerative disorder of early childhood-onset with severe axonal polyneuropathy

Kathie J. Ngo^{1,2}, Aristides Hadjinicolaou³, Daniel Y. Conway¹, Craig L. Bennett⁴, Albert R. La Spada^{4,5}, Grace Yoon^{3,6}, Brent L. Fogel^{1,7,8}

¹ Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles California, USA, ² Bioinformatics Interdepartmental Program, University of California Los

Angeles, Los Angeles California, USA, ³ Division of Neurology, Department of Paediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada, ⁴ Department of Pathology & Laboratory Medicine, University of California, Irvine, USA, ⁵ Department of Neurology and Department of Biological Chemistry, UC Institute for Neurotherapeutics, University of California, Irvine, USA, ⁶ Division of Clinical and Metabolic Genetics, Department of Paediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada, ⁷ Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles California, USA, ⁸ Clinical Neurogenomics Research Center, David Geffen School of Medicine, University of California Los Angeles, Los Angeles California, USA

Pathogenic variants in *SETX* cause two distinct neurological diseases, a loss-of-function recessive disorder, ataxia with oculomotor apraxia type 2 (AOA2), and a dominant gain-of-function motor neuron disorder, amyotrophic lateral sclerosis type 4 (ALS4). We identified two unrelated patients with the same *de novo* c.23C>T (p.Thr8Met) variant in *SETX* presenting with an early-onset, severe polyneuropathy. We used weighted gene-correlation network analysis (WGCNA) to identify disease-associated modules from two different ALS4 mouse models that overlap with disease-associated modules in confirmed ALS4 patient data to derive an ALS4-specific transcriptional signature. WGCNA of whole blood RNA-sequencing data from a patient with the p.Thr8Met *SETX* variant was compared to ALS4 and control patients to determine if this signature could be used to identify affected patients. The expression profile of the patient carrying the c.23C>T (p.Thr8Met) variant was significantly associated with the human and mouse ALS4 signature, confirming the relationship between this *SETX* variant and disease.

24. Systems Engineering of IL-2 Identifies Multivalency as a Critical, Unexplored Axis of Regulatory T Cell-Specific Therapies

Brian Orcutt-Jahns¹ Peter Emmel¹ Aaron Meyer¹

¹Department of Bioengineering, University of California, Los Angeles

Interleukin (IL)-2 has potential as a therapy in autoimmunity but has limited effectiveness due to its limited specificity toward regulatory T cells (T_{reg} s). IL-2 muteins with altered receptor-ligand binding kinetics or valency can improve the cell type selectivity of the signaling response. Here, we analyze the response of immune cells to a panel of IL-2 muteins in both monomeric and dimeric formats and found that dimeric muteins demonstrate high selectivity for T_{reg} s. We then dissect the mechanism of enhanced T_{reg} specificity in dimeric ligands using a simple, two-step multivalent binding model. Our model showed that the enhanced activation of T_{reg} s by dimeric muteins arises due to multivalent ligands' enhanced avidity for cells based on the abundance of target receptors, and that ligands engineered in even higher valency formats may possess greater potential for cell type selective signaling, thus identifying a new pathway for the engineering of more selective cytokines.

25. Metabolic dependencies of ecDNA and HSR oncogene focal amplification modes and plasticity

Amelia Palermo¹, Kai Song¹, Johanna ten Hoeve-Scott^{1,2}, Gabriel Karin¹, Mark Goodman¹, Thomas G. Graeber^{1,2,3}

¹Department of Molecular and Medical Pharmacology, ²Crump Institute for Molecular Imaging, ³UCLA Center for Metabolomics, University of California Los Angeles, Los Angeles California, USA

The focal amplification (FA) of genes that support the uncontrolled growth and proliferation of cells (*i.e.*, oncogenes) on homogeneous chromosomal staining regions (HSRs) and extrachromosomal DNA (ecDNA) is a hallmark of resistant cancers. However, it remains an open question whether oncogene FA is affected by cellular metabolism and by the composition of the tumor microenvironment. This is important because cancers carrying oncogene FA, particularly on ecDNA, can evolve fast, and are highly malignant. We profiled an in-house *in vitro* model of resistant melanoma exhibiting high plasticity of ecDNA+/HSR+ BRAF FA upon combined BRAF plus MEK inhibitor escalation, by metabolomics and lipidomics analysis. Our data reveal that both FA- vs FA+, and ecDNA+ vs HSR+ cells are characterized by extensive

metabolism reprogramming and substantial changes in lipid composition. Ultimately, our observations support the hypothesis that oncogene focal amplification on ecDNA and HSRs can be modulated by leveraging unique metabolic vulnerabilities.

26. Differentiation status in uveal melanoma is predictive of lipid oxidation-ferroptosis sensitivity

Kim Paraiso¹, Jennifer Tsoi¹, Sal Alhani¹, Will Crosson¹, Mark Goodman¹, Nick Putney¹, Gerardo Javier¹, Silvio Gutkind², Tara McCannel³, Antoni Ribas⁴, Thomas Graeber^{1*}

¹Department of Molecular and Medical Pharmacology, UCLA, Los Angeles, CA; ²Department of Pharmacology, Moores Cancer Center, University of California, San Diego, California; ³Department of Ophthalmology, Stein Eye and Doheny Eye Institutes, University of California, Los Angeles, CA; ⁴Department of Medicine, University of California at Los Angeles Jonsson Comprehensive Cancer Center, Los Angeles, CA

While frontline therapies are effective at treating approximately 50% of uveal melanomas (UVM), the remaining half of patients eventually succumb to metastatic disease, primarily of the liver. UVM is of particular note in terms of clinically impactful molecular diagnostics in that biomarker tests of primary tumor samples are strongly predictive of patient prognosis. Namely, specific gene expression signatures, mutations, and DNA copy number events (e.g. monosomy 3) are associated with metastasis and poor patient outcome. We previously reported a new metabolic vulnerability in cutaneous melanomas where high oxidative stress leads to lipid oxidation and cell demise via the programmed death pathway of ferroptosis. Due to metabolic changes observed in UVM, we extended our ferroptosis studies to this subtype of melanoma. To determine if UVM are sensitive to ferroptosis, we treated our UVM cell line panel with the GPX4 inhibitor, RSL3. Similar to findings in cutaneous melanoma, we observed that a subset of UVM cell lines are sensitive to RSL3 mediated ferroptosis induction. We and others have reported that across multiple aggressive cancer types, dedifferentiation is a strong predictor for ferroptosis sensitivity. To determine if UVM differentiation states can predict for ferroptosis sensitivity, we performed linear regression and gene set enrichment analyses and correlated RSL3 sensitivities to differentiation-associated gene signatures. In line with our previous findings for cutaneous melanoma, UVM cell lines with a dedifferentiated gene signature were more sensitive to ferroptosis induction than cell lines with a differentiated gene signature. Taken together, our findings support a central role of heightened redox stress in aggressive tumor types, and point to ferroptosis as a potential therapeutic vulnerability in UVM.

27. PseudotimeDE: inference of differential gene expression along cell pseudotime with well-calibrated p-values from single-cell RNA sequencing data

Dongyuan Song¹, Jingyi Jessica Li^{2,3,4}

¹ Bioinformatics Interdepartmental Ph.D. Program, ² Department of Statistics, ³ Department of Human Genetics, ⁴ Department of Computational Medicine, University of California Los Angeles, Los Angeles California, USA

To investigate molecular mechanisms underlying cell state changes, a crucial analysis is to identify differentially expressed (DE) genes along the pseudotime inferred from single-cell RNA-sequencing data. However, existing methods do not account for pseudotime inference uncertainty, and they have either ill-posed *p*-values or restrictive models. Here we propose PseudotimeDE, a DE gene identification method that adapts to various pseudotime inference methods, accounts for pseudotime inference uncertainty, and outputs well-calibrated *p*-values. Comprehensive simulations and real-data applications verify that PseudotimeDE outperforms existing methods in false discovery rate control and power.

28. Tensor-structured decomposition improves systems serology analysis

Cyrillus Tan¹, Madeleine Murphy², Hakan Alpay³, Scott Taylor⁴, Aaron S. Meyer^{1,4}

¹ Bioinformatics Interdepartmental Program, ² Computational and Systems Biology, ³ Department of Computer Science, ⁴ Department of Bioengineering University of California, Los Angeles, Los Angeles, California, USA

Systems serology provides a broad view of humoral immunity by profiling both the antigen-binding and Fc properties of antibodies. These studies usually contain structured biophysical profiling across disease-relevant antigen targets, alongside additional in an antigen-generic manner. Identifying patterns in these measurements help guide vaccine and therapeutic antibody development and discover conserved regulatory mechanisms. Here, we report that coupled matrix-tensor factorization (CMTF), a form of tensor factorization, can reduce these data into consistently observed patterns. We use previous measurements of HIV- and SARS-CoV-2-infected subjects as examples. CMTF outperforms PCA in the extent of data reduction possible while making equivalently accurate but more replicable prediction of immune functional responses and subject classification. It also improves the interpretation through further data reduction, separation of Fc and antigen-binding effects, and recognition of consistent patterns across individuals. Therefore, we propose that CMTF is an effective general strategy for data exploration in systems serology.

29. Characterizing biological groups' representative chromatin state maps and detection of differential chromatin sites among groups

Ha Vu^{1,2}, Zane Koch¹, Petko Fiziev⁸, Jason Ernst^{1,2,3,4,5,6,7}

¹ Bioinformatics Interdepartmental Program, University of California, Los Angeles, CA, 90095, USA, ² Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA, ³ Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at University of California, Los Angeles, Los Angeles, CA 90095, USA, ⁴ Computer Science Department, University of California, Los Angeles, Los Angeles, CA 90095, USA, ⁵ Jonsson Comprehensive Cancer Center, University of California, Los Angeles, Los Angeles, CA 90095, USA, ⁶ Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA, ⁷ Computational Medicine Department, University of California, Los Angeles, Los Angeles, CA 90095, USA, ⁸ Illumina Inc.

Genome-wide maps of epigenetic modifications provide powerful resources for genome annotations. Maps of epigenetics marks have been integrated into widely-used cell-type-specific 'chromatin-state' annotations. In many cases, given a group of biologically similar samples, it is desirable to have a chromatin-state annotation that summarizes annotations of those samples. However, determining an effective summary annotation is challenging: there exists no explicit notion of states' similarities, while in practice some states are more biologically similar than others. Here, we developed CSREP-- method that accepts a set of chromatin-state annotations from a group of samples and probabilistically estimates the group's most representative annotation. CSREP trains a logistic regression classifier predicting the chromatin-state assignment of each sample, given the equivalent annotations from other samples, then averaging prediction probabilities. This enables implicitly learning a notion of states' distances. Additionally, the difference between two groups' representative chromatin-state maps helps identify differential chromatin regions. We designed a permutation-based test to statistically evaluate those differences. We applied CSREP to groups of reference epigenomes from Roadmap Epigenomics project. We demonstrate advantages of CSREP compared to a baseline method for this application. We also show CSREP can identify biologically relevant differences between groups with greater power than previous approaches.

30. Quantifying the ecological stability of strains in the human gut microbiome

Richard Wolff¹, William Shoemaker¹, Nandita Garud^{1,2}

¹Department of Ecology and Evolutionary Biology, ²Department of Human Genetics

The human gut microbiome is a complex community that harbors astounding taxonomic, phylogenetic, and functional diversity. This diversity extends below the species level, with nearly all species being made up of a handful of discrete strains present at intermediate frequencies. However, the ecological and evolutionary forces promoting strain coexistence remain obscure, in part because we lack a robust understanding of the dynamic behavior of strains. Here, we demonstrate that over periods of years, strains are overwhelmingly dynamically stable ecological units within hosts. First, we show that while levels of within-species genetic diversity experience short-term fluctuations, they are effectively stationary over longer periods. Next, we show that nearly all strains can be parsimoniously described by a stochastic logistic model, in which each population fluctuates around a fixed carrying capacity due to environmental noise. Lastly, we find that strains obey macroecological laws, including Taylor's Law – a power law scaling between the mean and variance in population abundance. Together, these findings indicate that ecological stability, one of the defining features of the human gut microbiome, emerges at the level of strains.

31. Widespread reduction in gene expression heritability drives increased heterogeneity through aging

Ryo Yamamoto^{1,2}, Ryan Chung³, Juan Manuel Vazquez¹, Huanjie Sheng¹, Nilah Ioannidis³, Peter H Sudmant¹

¹Department of Integrative Biology, University of California, Berkeley, ²Bioinformatics Interdepartmental Program, University of California, Los Angeles, ³Department of Electrical Engineering and Computer Science, University of California, Berkeley

Age is the primary risk factor for many common human diseases and determining how and why tissues age differently is key to understanding them. Here, we analyzed gene expression data collected across 27 tissues from 1,762 humans and identified two distinct signatures of age-associated dysregulation. We identify a subset of tissues in which, on average, gene expression patterns become more erratic with age. In these tissues eQTLs are less predictive of gene expression in older individuals compared to younger individuals. We show that this effect is largely driven by a decrease in the heritability of gene expression patterns in older individuals and the increased impact of the environment. These results suggest that changes in gene expression patterns associated with age are the result of a loss of genetic control concomitant.

32. Clinical Temporal Relation Extraction with Probabilistic Soft Logic Regularization and Global Inference

Yichao Zhou¹, Yu Yan^{2,4}, Rujun Han³, J. Harry Caufield², Kai-Wei Chang¹, Yizhou Sun¹, Peipei Ping², Wei Wang¹

¹Department of Computer Science, UCLA, ²Departments of Physiology, ³Department of Computer Science, USC, ⁴Medical Informatics, UCLA

There has been a steady need in the medical community to precisely extract the temporal relations between clinical events. In particular, temporal information can facilitate a variety of downstream applications such as case report retrieval and medical question answering. Existing methods either require expensive feature engineering or are incapable of modeling the global relational dependencies among the events. In this paper, we propose a novel method, Clinical Temporal Relation Extraction with Probabilistic Soft Logic Regularization and Global Inference (CTRL-PG), to tackle the problem at the document level. Extensive experiments on two benchmark datasets, I2B2-2012 and TB-Dense, demonstrate that CTRL-PG significantly outperforms baseline methods for temporal relation extraction.

33. Genetic control of deconvolved brain cell-types

Chloe X Yap^{1,2,3,4}, Chongyuan Luo⁵, Arjun Bhattacharya^{6,7}, Dorothea Seiler Vellame⁸, Alice Franklin⁸, Jonathan Mill⁸, Michael J Gandal^{1,2,5}

¹Department of Psychiatry, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA, ²Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA, ³Mater Research Institute, University of Queensland, Brisbane, Queensland, Australia, ⁴Institute for Molecular Biosciences, University of Queensland, Brisbane, Queensland, Australia, ⁵Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA, ⁶Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, USA, ⁷Institute for Quantitative and Computational Biosciences, David Geffen School of Medicine, University of California, Los Angeles, CA, USA, ⁸College of Medicine and Health, University of Exeter, Royal Devon and Exeter Hospital, Exeter, UK

Most neuropsychiatric diseases lack clear pathology in the brain. In neurodegenerative disorders, cell-type proportion shifts are intermediate traits that can reflect cellular vulnerability or activation states that relate to disease pathophysiology. Here, we deconvolved brain cell-type proportions from bulk prefrontal cortex DNA methylation, including participants with autism (N=51), schizophrenia (n=186), Alzheimer's Disease (N=300), and matched controls (N=680). We obtained estimates of NeuN +/- populations, and cell sub-type populations including excitatory neurons, inhibitory neurons, astrocytes, microglia, and oligodendrocytes. Preliminary analyses identified compositional shifts in inhibitory interneurons and oligodendrocytes for schizophrenia, and microglial changes in autism. Work is currently underway to investigate the heritability and genetic control of these cell-type proportions. Understanding the contribution of cell-types to neuropsychiatric disease may help to uncover novel mechanisms, and help to characterize relationships between genetics, epigenetics and neuropsychiatric traits.

34. Detecting A Genomic Basis for Physiological Adaptation to High Altitude in Peruvian Quechua

Kimberly T. Zhu¹, Abigail Bigham¹

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

Humans have inhabited the Andean Altiplano for over 11,000 years, where the partial pressure of oxygen is 35% lower than at sea level. Peruvian Quechua who thrive in this environment display a suite of adaptive phenotypes, including elevated hemoglobin (Hb) concentration. The genetic architecture underlying this adaptive phenotype is currently unknown, as this adaptation is extremely unique; other populations such as the Tibetan Monpa display an alternative suite of adaptive phenotypes to high altitude. We focus on the phenotype of elevated hemoglobin concentration and identify several genetic loci that have strong signatures of genetic selection through two selection scans. With these SNPs, we performed a genome-wide association study (GWAS) to detect significant association with hemoglobin concentration in high altitude Quechua participants. By investigating the genetic basis for this adaptive phenotype, this study contributes novel insights as to the occurrence of genetic adaptation in this high altitude population.