
# Poster Abstracts

***1. An immune complex binding model predicts effector cell response across diverse disease models***

**Z. Cyrillus Tan,** Anja Lux, Markus Biburger, Stephen Lees, Falk Nimmerjahn, Aaron S. Meyer
University of California Los Angeles
Friedrich-Alexander-University of Erlangen-Nürnberg

Antibodies are crucial and central regulators of the immune response. Those of the IgG isotype interact with Fc gamma receptors on effector cells. The effector response of IgGs encompasses multiple cell types (e.g., macrophages, monocytes, etc.) and multiple processes (e.g., antigen presentation, cytokine response, phagocytosis, etc.). An individual IgG can be either pro- or anti-inflammatory depending upon its Fc domain composition and the context. IgGs are particularly versatile agents for therapeutic treatment due to their immunotherapeutic effects as well as those of direct antigen binding and opsonization. Indeed, IgG molecules comprise a broad range of approved therapies, many of which are known to rely in part on effector cell response. At the same time, the multiplicity throughout—of constant region composition, Fc gamma receptors, cell populations, and antigen binding in combination—makes precisely understanding, measuring, and manipulating effector function a yet-elusive goal.

***2. Inference on the structure of gene regulatory networks***

**Yue Wang**[1,2] and Zikun Wang[3]
[1] Department of Computational Medicine, UCLA
[2] Institut des Hautes Études Scientifiques
[3] Laboratory of Genetics, The Rockefeller University

We conduct theoretical analyses on inferring the structure of gene regulatory networks. Depending on the experimental method and data type, the inference problem is classified into 20 different scenarios. For each scenario, we discuss the problem that with enough data, under what assumptions, what can be inferred about the structure. For scenarios that have been covered in the literature, we provide a brief review. For scenarios that have not been covered in literature, if the structure can be inferred, we propose new mathematical inference methods and evaluate them on simulated data. Otherwise, we prove that the structure cannot be inferred.

***3. Multivalency enhances the specificity of Fc-cytokine fusions***

**Brian Orcutt-Jahns**[1] Peter Emmel[1] Cori Posner[2] Scott Carlson[2] Aaron Meyer[1,3,4,5]
[1]Department of Bioengineering, University of California, Los Angeles
[2]Visterra, Inc., Waltham, MA
[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

Interleukin (IL)-2 has potential as a therapy in cancer and autoimmunity but is limited in effectiveness by its specificity toward desired immune populations. IL-2 muteins may have improved cell type selectivity through altered receptor-ligand binding kinetics. Here, we analyze the response of immune cells to a panel of IL-2 muteins in monomeric and dimeric Fc fusions using tensor factorization. We find that dimeric muteins have considerably altered selectivity profiles. We then dissect the mechanism of altered specificity in dimeric ligands using a multivalent binding model and show that the enhanced selectivity by dimeric muteins arises due to multivalent ligands' avidity for cells based on the abundance of their target receptors. In total, we show that multivalent cytokines have unique selectivity profiles, that modeling can help to aid their design, and that tensor factorization provides an effective approach to visualize ligand responses across diverse cell populations.

### *4. Optimal Design of Experiments for Simulation-Based Inference of Mechanistic Acyclic Biological Networks*

**Vincent Zaballa**[1], Elliot Hui[1]

[1] Department of Biomedical Engineering, University of California Irvine, Irvine, California, USA

Biological signaling pathways based upon proteins binding to one another to relay a signal for genetic expression, such as the Bone Morphogenetic Protein (BMP) signaling pathway, can be modeled by mass action kinetics and conservation laws that result in non-closed form polynomial equations. Accurately determining parameters of biological pathways that represent physically relevant features, such as binding affinity of proteins and their associated uncertainty, presents a challenge for biological models lacking an explicit likelihood function. Additionally, parameterizing non-closed form biological models requires copious amounts of data from expensive perturbation-response experiments to fit model parameters. We present an algorithm (SBIDOEMAN) for determining optimal experiments and parameters of systems biology models with implicit likelihoods. We evaluate our algorithm using simulations of held-out true parameter values and demonstrate an improvement in the rate of accurate parameter inference over random and equidistant experimental designs when evaluated on two simple models of the BMP signaling pathway with an implicit likelihood function.

### *5. Phenotypic variance of new mutations in a spin glass model of fitness landscape*

**Supriyo Bhattacharya**

Integrative Genomics Core, Beckman Research Institute, City of Hope National Medical Center, Duarte, CA 91010

Phenotypic variance is the basis for adaptation to changing environments. The capacity for variation in populations is measurable as distribution of fitness effects or DFE (frequencies of deleterious, neutral and advantageous mutations). A key contributor to phenotypic variance is the genotype itself. However, understanding the mechanisms of variation are complicated by epistasis, the interactions among multiple genes governing phenotypic traits. Using a genotype-phenotype map inspired by spin glass models, I derive analytical expressions for two key properties of the DFE, the mean and the variance, as function of the gene interaction network. These mathematical functions allow a deeper look into the role of epistasis on variation, including the contribution of individual network nodes. Monitoring these properties during simulations demonstrate dynamic modulation of the DFE during periods of environmental fluctuation, leading to efficient stabilization of population traits. The emerging insight is that, DFE actuated by epistatic interactions, finetunes adaptation.

### *6. Quantifying the phenotypic information in mRNA abundance*

**Evan Maltz**[1,2], Roy Wollman[1,2,3,*]

[1] Department of Chemistry and Biochemistry, UCLA

[2] Institute of Quantitative and Computational Bioscience, UCLA

[3] Department of Integrative Biology and Physiology, UCLA

* correspondence author (rwollman@ucla.edu)

Quantifying the dependency between mRNA abundance to downstream cellular phenotypic is a fundamental open problem in biology. New computational frameworks are needed to dissect the contribution of individual genes and gene combinations to a given phenotype. Using an information theory approach, we analyzed the expression of 83 genes in the $Ca^{2+}$ signaling network and the dynamic $Ca^{2+}$ response in the same cells. We found that the overall expression levels of these 83 genes explains approximately 60% of $Ca^{2+}$ signal entropy. The average contribution of each single gene was 16%, revealing a large degree of redundancy between genes. Using different heuristics, we estimated the dependency between size of gene sets and its information content, revealing that the average set of 53 genes contains 90% of $Ca^{2+}$ signaling information. Our results provide the first direct quantification of information content about complex cellular phenotype that exists in mRNA abundance measurements.

### *7. Multi-omics analysis reveals gut-brain interactions mediating the anti-obesity effect of Akkermansia muciniphila*

**Sung Min Ha,** In-Sook Ahn, Graciel Diamante, Ingrid Cely, Guanglin Zhang, Thomas Kowal-safron, Justin Yoon and Xia Yang. Department of Integrative Biology and Physiology, University of California Los Angeles. USA

Previous studies have implicated a causal role for the gut bacterium *Akkermansia muciniphila in counteracting* diet-induced obesity and metabolic dysfunctions. However, a systems level understanding of the molecular mechanisms underlying the anti-obesogenic effect of *A. muciniphila* is lacking. Using fructose-induced obesity as a model, we carried out multiomics studies to investigate the molecular cascades mediating the effect of *A. muciniphila*. We found that *A. muciniphila* colonization in a fructose-induced obesity model triggered significant shifts in gut microbiota composition as well as alterations in numerous gut and plasma metabolites and gene expression in the hypothalamus. Among these, we found that the metabolite oleoyl-ethanolamide in the gut and circulation and hypothalamic oxytocin are the key regulators of gut-brain interactions that underlie the *A. muciniphila* anti-obesity effect. Our multiomics investigation elucidates the molecular regulators and pathways involved in the communication between *A. muciniphilain* the gut and hypothalamic neurons that counter fructose-induced obesity.

### *8. A molecular taxonomy of tumors independent of tissue-of-origin*

**Peter T. Nguyen[1],** Simon G. Coetzee[1], Daniel L. Lakeland[2], Dennis J. Hazelett[1,3]

[1]The Center for Bioinformatics and Functional Genomics, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA

[2]Lakeland Applied Sciences LLC, Los Angeles, CA 91001, USA

[3]Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA

Cancer is an organism-level disease, impacting processes from cellular metabolism and the microenvironment to systemic immune response. Nevertheless, efforts to distinguish overarching mutational processes from interactions with the cell of origin for a tumor have seen limited success, presenting a barrier to individualized medicine. Here we present a pathway-centric approach, extracting somatic mutational profiles within and between tissues, largely orthogonal to cell of origin, mutational burden, or stage. Known predisposition variants are equally distributed among clusters, and largely independent of molecular subtype. Prognosis and risk of death vary jointly by cancer type and cluster. Analysis of metastatic tumors reveals that differences are largely cluster-specific and complementary, implicating convergent mechanisms that combine familiar driver genes with diverse low-frequency lesions in tumor-promoting pathways, ultimately producing distinct molecular phenotypes. The results shed new light on the interplay between organism-level dysfunction and tissue-specific lesions.

### *9. Development of a fluidic model system for studying cancer vasculogenic mimicry*

**Maya Rowell**[1], William Leineweber[1], Stephanie Fraley, Ph.D[1]

[1]University of California, San Diego, San Diego, CA

Vasculogenic mimicry (VM) is a process by which cancer cells establish an alternative perfusion pathway in an endothelial cell-free manner or via mosaic vessels that contain both endothelial and cancer cells. We set out to develop a microfluidic tissue engineered culture system capable of modeling the formation and 3D perfusion function of VM. In static conditions, our system revealed that culture within high density collagen induced 4T1 cells to upregulate several growth factors and secreted proteins associated with VM and began to form networks *in vitro*. To achieve hollow perfused networks, optimal medium and matrix conditions are implemented in a microfluidic device while also varying flow conditions to identify the range of pressure drop and/or VM factor gradients necessary to enable perfusion. Future work will expand the model to study the interactions of immune cells, thus providing a platform for identifying potential therapies targeting VM.

### *10. Synthetic mammalian signaling circuits for robust cell population control*

**Yitong Ma**[1, 6], Mark W. Budde[1, 2, 6], Michaëlle N. Mayalu[3], Junqin Zhu[4], Andrew C. Lu[1], Richard M. Murray[1, 3], and Michael B. Elowitz[1, 3, 5, 7]*

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA

[2]Primordium Labs, Arcadia, CA 91006, USA

[3]Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA

[4]Department of Biology, Stanford University, Stanford, CA 94305, USA

[5]Howard Hughes Medical Institute, California Institute of Technology, Pasadena, CA 91125

[6]These authors contributed equally

[7]Lead contact

In multicellular organisms, cells actively sense and control their own population density. Synthetic mammalian quorum sensing circuits could provide insight into principles of population control and extend cell therapies. However, a key challenge is reducing their inherent sensitivity to “cheater” mutations that evade control. Here, we repurposed the plant hormone auxin to enable orthogonal mammalian cell-cell communication and quorum sensing. We designed a paradoxical population control circuit, termed *Paradaux,* in which auxin stimulates and inhibits net cell growth at different concentrations. This circuit limited population size over extended timescales, of up to 42 days of continuous culture. By contrast, when operating in a non-paradoxical regime, the same cells limited population growth, but were more susceptible to mutational escape. These results establish auxin as a versatile “private” communication system, and demonstrate that paradoxical circuit architectures can provide robust population control.

### *11. Phasor Unmixing to Reveal Organelle Organization and Cellular Response*

**S. Zhu**, L.Scipioni, M.A. Digman

Department of Biomedical Engineering, The Henry Samuli School of Engineering

Organelles are a series of subcellular structures that perform various functions inside eukaryotic cells. The coordination between organelles plays an important role in understanding the intricate biochemical processes within eukaryotic cells as organelles must work in concert to maintain cell function. Here, we perform simultaneous excitation of seven fluorophores and spectral phasor unmixing to achieve robust emission separation of all seven stained subcellular compartments and organelles including the Golgi apparatus, tubulin, lipid droplets, lysosomes, mitochondria, nuclear and mitochondrial DNA. Our imaging approach uses a single wavelength, two-photon excitation at 780 nm coupled with multicolor organelle-directed staining to obtain multi-parametric physiological profiling of live MCF10A cells, a non-tumorigenic epithelial breast cancer cell line, under stress-related treatments. We created an analysis pipeline to spectrally unmix the contribution of each organelle and obtain phenotypic signatures under each type of treatment. Ultimately, we characterized significant differences in organelle phenotypes upon application

of a variety of treatments, including Antimycin A (mitochondrial respiration), hydrogen peroxide (oxidative stress), Nocodazole (microtubule depolarization), and serum starvation. The phasor approach provides a robust tool in separating up to seven simultaneous emission spectra and could be useful in separating even more components by complementing with lifetime imaging.

***12. The arteriovenous metabolic flux of kidneys during fast and fed states***

**Lauren A. Hoffner**[1,*], Alexis L. Anica[1,*], Cuauhtemoc B. Ramirez[1,2,*], Joohwan Kim[1], Cholsoon Jang[2,#], Gina Lee[1,#]

[1]Department of Microbiology and Molecular Genetics, Chao Family Comprehensive Cancer Center, University of California Irvine School of Medicine, Irvine, CA 92697-4025, USA

[2]Department of Biological Chemistry, Chao Family Comprehensive Cancer Center, University of California Irvine School of Medicine, Irvine, CA 92697-4025, USA

[*]These authors contributed equally to this work.

[#]Correspondence: Cholsoon Jang (choljang@uci.edu), Gina Lee (ginalee@uci.edu)

Kidney is a powerful chemical factory that removes wastes from the blood and reabsorbs nutrients before being secreted through urine. Despite its vital functions, in vivo kidney metabolism is not as widely studied compared to in vitro kidney metabolism, which lacks the physiological complexity of organs. Fasting and feeding are fundamental nutritional challenges in our daily life that can influence the various human pathophysiology. Thus, we aimed to investigate metabolic differences of kidneys in fast vs. refed mice. We analyzed renal metabolic flux by quantitatively measuring metabolites isolated from artery and renal blood using LC-MS. Such measurements allowed us to map kidney-specific metabolite release and uptake in fasting and feeding states. This study will provide a comprehensive understanding of kidney metabolism in vivo and demonstrate how important the kidney is in adapting to the availability of dietary nutrients and circulating metabolites in fasted and fed conditions.

***13. Hepatic lipogenesis mediates genetic vulnerability to fructose-induced steatosis***

**Cuauhtemoc B. Ramirez**, In-Sook Ahn, Gina Lee, Xia Yang, and Cholsoon Jang

Department of Microbiology and Molecular Genetics, University of California Irvine

Department of Biological Chemistry, University of California Irvine

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

Chao Family Comprehensive Cancer Center, University of California Irvine

Center for Complex Biological Systems, University of California Irvine

Center for Epigenetics and Metabolism, University of California Irvine

Excessive fructose intake is a risk factor for non-alcoholic fatty liver disease (NAFLD), yet how different individuals have distinct NAFLD vulnerabilities remains unknown. Using in vivo isotope tracing and metabolomics in three genetically divergent mouse strains (B6, DBA, and FVB), here we show that hepatic lipogenesis induction is the metabolic mal-adaptation that mediates NAFLD susceptibility. DBA strain is highly susceptible to fructose-induced NAFLD with increased hepatic lipid deposition and fibrosis. Quantitative analysis reveals that DBA strain shows distinctively strong induction of hepatic but not small intestinal or gut microbial fructose catabolism, which coincides with potent lipogenesis induction. Neither FVB strain showing constitutively active hepatic fructose catabolism nor B6 strain showing enhanced hepatic fructose catabolism without lipogenesis coupling develop NAFLD. Thus, individuals with fructose-sensitive lipogenesis induction are more likely susceptible to NAFLD, pinpointing lipogenesis assessment as a diagnostic tool for identifying such individuals and recommending dietary fructose restriction for NAFLD prevention.

***14. Modeling NFκB signaling and gene expression responses in heterogeneous single macrophages***

**Xiaolu Guo, Yijia Chen**, Xiaofei Lin, Adewunmi Adelaja, Alexander Hoffmann.

Signaling Systems Laboratory, 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 Empirical Bayes Estimates (EBSs) estimation. We then extended the signaling network model with an experimentally tuned model of gene expression based on a chromatin-associated mechanism of transcriptional initiation and mRNA decay. We used the model to predict the gene expression specificity and computed the expression profiles of 81 genes in single immune cells that were treated with 5 different stimuli of different concentrations. Overall, our models quantitatively predict the NFκB immune gene expression responses to pathogen threats. We discuss next steps to improve the fits to experimental data in our long-term effort to produce the 'virtual macrophage'.

***15. A mechanochemical instability drives vertebrate gastrulation***

**Mattia Serra**[1], Cornelis J. Weijer[2], L. Mahadevan[3]
[1]UCSD, Department of Physics, CA USA
[2]University of Dundee, Division of Cell and Developmental Biology, UK
[3]School of Engineering and Applied Sciences, Harvard University, MA USA

Gastrulation is a critical event in vertebrate morphogenesis, characterized by coordinated large-scale multi-cellular movements. One grand challenge in modern biology is understanding how spatio-temporal morphological structures emerge from cellular processes in a developing organism and vary across vertebrates. We derive a theoretical framework that couples tissue flows, stress-dependent myosin activity, and actomyosin cable orientation. Our model, consisting of a set of nonlinear coupled PDEs, predicts the onset and development of observed experimental patterns of wild-type and perturbations of chick gastrulation as a spontaneous instability of a uniform state. We use analysis and numerics to show how our model recapitulates the phase space of gastrulation morphologies seen across vertebrates, consistent with experiments. Altogether, this suggests that early embryonic self-organization follows from a minimal predictive theory of active mechano-sensitive flows.

***16. The Role of TFPI Inhibition in Coagulation Under Flow***

**Amandeep Kaur**[1], Dougald Monroe[3], Karin Leiderman[2], and Suzanne Sindi[1]
[1]Applied Mathematics Department, University of California, Merced.
[2]Applied Mathematics and Statistics, Colorado School of Mines, Colorado.
[3]Department of Medicine, University of North Carolina, Chapel Hill.

Blood coagulation is a complex network of biochemical reactions necessary to form a blood clot. The process occurs in three stages (initiation, amplification, and propagation) under flow with inhibition mechanisms happening at each stage to regulate clotting. Tissue factor pathway inhibitor (TFPI) is an important inhibitor of initiation phase, but its precise role in inhibition under flow remains unknown. Previous mathematical models of TFPI show that flow itself is a more important inhibitor of the system than TFPI. In this study, we re-investigate the role of TFPI with and without the presence of flow. First, we employ mathematical models and constrained optimization to fit the data given by previous static experimental studies of TFPI. Our work shows the data fits our model better when two mechanisms for TFPI inhibition are included. We then study our model under flow to understand the relationship between flow and both mechanisms of TFPI inhibition.

***17. MLR-OOD: a Markov chain based Likelihood Ratio method for Out-Of-Distribution detection of genomic sequences.***

**Xin Bai**[1], Jie Ren[2], Fengzhu Sun[1].

[1]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, 90089, USA.

[2]Google Research, Brain Team, USA.

Machine learning or deep learning models have been widely used for taxonomic classification of metagenomic sequences and many studies reported high classification accuracy. Such models are usually trained based on sequences in several training classes in hope of accurately classifying unknown sequences into these classes. However, when deploying the classification models on real testing data sets, sequences that do not belong to any of the training classes may be present and are falsely assigned to one of the training classes with high confidence. Such sequences are referred to as out-of-distribution (OOD) sequences and are ubiquitous in metagenomic studies. To address this problem, we develop a deep generative model-based method, MLR-OOD, that measures the probability of a testing sequencing belonging to OOD by the likelihood ratio of the maximum of the in-distribution (ID) class conditional likelihoods and the Markov chain likelihood of the testing sequence measuring the sequence complexity. We compose three different microbial data sets consisting of bacterial, viral, and plasmid sequences for comprehensively benchmarking OOD detection methods. We show that MLR-OOD achieves the state-of-the-art performance demonstrating the generality of MLR-OOD to various types of microbial data sets. It is also shown that MLR-OOD is robust to the GC content, which is a major confounding effect for OOD detection of genomic sequences. In conclusion, MLR-OOD will greatly reduce false positives caused by OOD sequences in metagenomic sequence classification.

***18. Quantitatively Comparing Tumor Images to Agent-Based Models for Parameter Fitting***

**Colin G. Cess,** Stacey D. Finley.

Biomedical Engineering, University of Southern California. USA.

Agent-based models (ABMs) of tumors have produced great insight into tumor behavior and interactions with other cells. However, ABMs are difficult to parameterize as there is no current way to quantitatively compare ABM simulations to tumor images, which is the data type required as it provides spatial information. Therefore, with this study, we describe a method for quantitative comparison by using a Siamese Neural Network to calculate the similarity between tumor images and ABM simulations as a single, continuous value. This value can then be used in a parameter fitting algorithm in order to estimate the parameters of the ABM. Here, we use model-generated data in place of actual tumor images in order to thoroughly test the algorithm's robustness and speed. We use two different ABMs at several parameter sets in order to produce simulations with very different spatial states.

***19. NFκB dynamics characterize Response Specificities of polarization-specific macrophage states***

**Apeksha Singh**, Supriya Sen, Adewunmi Adelaja, Alexander Hoffmann

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

Macrophages show remarkable functional pleiotropy that is dependent on microenvironmental context. Prior studies have characterized how polarizing cytokines alter the transcriptomic and epigenetic landscape. Here we characterized the immune-threat appropriate responses of polarized macrophages by measuring the single-cell signaling dynamics of transcription factor NFκB. Leveraging a fluorescent protein reporter mouse, primary macrophages were polarized into 6 states and stimulated with 8 different stimuli resulting in a vast dataset. Linear Discriminant Analysis revealed how NFκB signaling codons compose the immune threat level of stimuli, placing polarization states along a linear continuum between the M1/M2 dichotomy. Machine learning classification revealed losses of stimulus distinguishability with polarization,

which reflect a switch from sentinel to more canalized effector functions. However, the stimulus response dynamics and discrimination patterns did not fit the M1/M2 continuum. Instead, our analysis suggests macrophage functional niches within a multi-dimensional polarization landscape.

### *20. HiCBin: binning metagenomic contigs and recovering metagenome-assembled genomes using Hi-C contact maps*

**Yuxuan Du,** Fengzhu Sun.
Department of Quantitative and Computational Biology, University of Southern California. USA.

Recovering high-quality metagenome-assembled genomes (MAGs) from complex microbial ecosystems remains challenging. Conventional shotgun-based binning approaches may encounter barriers when multiple samples are scarce. Recently, high-throughput chromosome conformation capture (Hi-C) has been applied to simultaneously study multiple genomes in natural microbial communities. We develop HiCBin, a novel open-source pipeline, to resolve high-quality MAGs utilizing Hi-C contact maps. HiCBin employs the HiCzin normalization method and the Leiden community detection algorithm based on the Potts spin-glass model and includes the spurious contact detection into binning pipelines for the first time. We validate our method and compare the capability to recover high-quality MAGs of HiCBin against other state-of-the-art Hi-C-based binning tools including ProxiMeta, bin3C, and MetaTOR, and one popular shotgun-based binning software MetaBAT2 on a human gut sample and a wastewater sample. HiCBin provides the best performance and applicability in resolving MAGs.

### *21. 16S rRNA and metagenomic shotgun sequencing data revealed consistent patterns of gut microbiome signature in pediatric ulcerative colitis*

**Wenxuan Zuo**, Beibei Wang, Xin Bai, Yihui Luan, Yingying Fan, Sonia Michail and Fengzhu Sun
Quantitative and Computational Biology Department, University of Southern California. USA.

Dysbiosis of human gut microbiota has been reported in association with ulcerative colitis (UC) in both children and adults using either 16S rRNA gene or shotgun sequencing data. However, these studies used either 16S rRNA or metagenomic shotgun sequencing but not both. We sequenced feces samples from 19 pediatric UC and 23 healthy children ages between 7 to 21 years using both 16S rRNA and metagenomic shotgun sequencing. The samples were analyzed using three different types of data: 16S rRNA genus level abundance, microbial species and pathway abundance profiles. We demonstrated that a) the alpha diversity of pediatric UC cases is lower than that of healthy controls; b) the beta diversity within children with UC is more variable than within the healthy children; c) several microbial families including Akkermansiaceae, Clostridiaceae, Eggerthellaceae, Lachnospiraceae, and Oscillospiraceae, contain species that are depleted in pediatric UC compared to controls; d) a few associated species unique to pediatric UC, but not adult UC, were also identified, e.g. some species in the Christensenellaceae family were found to be depleted and some species in the Enterobacteriaceae family were found to be enriched in pediatric UC; and e) both 16S rRNA and shotgun sequencing data can predict pediatric UC status with area under the receiver operating characteristic curve (AUROC) of close to 0.90 based on cross validation. We showed that 16S rRNA data yielded similar results as shotgun data in terms of alpha diversity, beta diversity, and prediction accuracy. Our study demonstrated that pediatric UC subjects harbor a dysbiotic and less diverse gut microbial population with distinct differences from healthy children. We also showed that 16S rRNA data yielded accurate disease prediction results in comparison to shotgun data, which can be more expensive and laborious. These conclusions were confirmed in an independent data set of 7 pediatric UC cases and 8 controls.

### *22. DeepLINK: Deep Learning Inference Using Knockoffs with Applications to Genomics*

**Zifan Zhu**[1], Yingying Fan[2], Yinfei Kong[3], Jinchi Lv[2], and Fengzhu Sun[1]
[1]Quantitative and Computational Biology Department, University of Southern California, Los Angeles, CA 90089

[2]Data Sciences and Operations Department, Marshall School of Business, University of Southern California, Los Angeles, CA 90089
[3]Department of Information Systems and Decision Sciences, California State University Fullerton, Fullerton, CA 92831
We propose a deep learning based knockoffs inference framework, DeepLINK, that guarantees the False Discovery Rate (FDR) control in high-dimensional settings. DeepLINK is applicable to a broad class of covariate distributions described by the possibly nonlinear latent factor models. It consists of two major parts: an autoencoder network for the knockoff variable construction and a multilayer perceptron network for feature selection with the FDR control. The empirical performance of DeepLINK is investigated through extensive simulation studies, where it is shown to achieve FDR control in feature selection with both high selection power and high prediction accuracy. We also apply DeepLINK to three real data applications to demonstrate its practical utility.

### *23. Epigenomic and Chromosomal Architecture Reconfiguration in the Developing Human Prefrontal Cortex*

**Matthew G. Heffel[1,2]**, Yi Zhang[1], Terence Li[1], Kevin Abuhanna[1], Mercedes Parades[3], Tom Nowakowski[4], Eran Mukamel[5], Chongyuan Luo[2].
Bioinformatics Interdepartmental Program, University of California, Los Angeles (UCLA), Los Angeles, CA, USA[1]
Department of Human Genetics, University of California Los Angeles, Los Angeles, 90095, CA, USA[2]
Department of Neurology, University of California, San Francisco, San Francisco, CA, USA[3]
Department of Anatomy, University of California, San Francisco, San Francisco, CA, USA[4]
Department of Cognitive Science, University of California, 9500 Gilman Dr., La Jolla, San Diego, 92093, CA, USA[5]
The human prefrontal cortex is highly involved in cognitive control functions; influencing attention, impulse, and memory. The specification and maturation of human cortical cell types are associated with global epigenomic reconfiguration including the pronounced accumulation of neuronal non-CG DNA methylation, adjustments in CG DNA methylation, and 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 prefrontal cortex development using more than 29,000 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 cell differentiation and maturation and identified brain regional specific regulatory signatures.

### *24. Uncovering the Causal Network of GPCR:G protein interactions that Regulate G protein Coupling*

**Elizaveta Mukhaleva[1]**, Ning Ma[1], Sergio Branciamore[1], Andrei Rodin[1], Nagarajan Vaidehi[1]
[1]Department of Computational & Quantitative Medicine, Beckman Research Institute of the City of Hope
G-protein coupled receptors (GPCRs) and G-protein interactions are dynamic and transient, so it is challenging to probe dynamic properties such as the chemical nature, spatiotemporal persistence and co-operativity of GPCR:G protein intermolecular contacts modulate the G protein coupling strength. In this work, we used a combination of Molecular Dynamics (MD) simulations and Bayesian Network (BN) models from Systems Biology field to decipher the GPCR:G protein residue pairs in the interface that play a critical role in coupling. By analyzing the spatial and temporal sampling of GPCR:G$\alpha$ protein contacts from MD simulations of 6 GPCR:G protein complexes (consisting of 2Gs, 2Gi and 2Gq coupled complexes), we have derived causal BN models for each system. Using the node strength property of BN models we have identified the GPCR:G$\alpha$ protein residue pairs that show high level of co-operativity and are causal to the G protein coupling.

### *25. Ensemble-based Genome-scale Modeling Predicts Metabolic Differences between Macrophages Subtypes in Colorectal Cancer*

**Patrick E Gelbach**[1], Stacey D Finley[1,2,3]

[1]Department of Biomedical Engineering, USC, Los Angeles, CA.

[2]Mork Family Department of Chemical Engineering and Materials Science, USC, Los Angeles, CA.

[3]Department of Quantitative and Computational Biology, USC, Los Angeles, CA.

The tumor microenvironment contributes to the difficulty in properly treating Colorectal Cancer. Macrophages are among the most abundant immune cells in the TME, and are generally classified into two subtypes: the classically activated (M1) phenotype, known for its inflammatory and anti-cancer properties; or the alternatively activated (M2) subtype, also known as Tumor-Associated Macrophages for their role in encouraging tumor proliferation, mediating immunosuppression, and promoting angiogenesis. This subclassification scheme is increasingly viewed as dependent on variation in cellular metabolism; however, the metabolic differences between the subtypes are not fully elucidated. Constraint-based analyses of genome-scale metabolic network reconstructions (GENREs) have emerged as a possible approach to tackle this challenge. GENREs are computational representations of cellular metabolism that generate predictions about the cell's metabolic capabilities. In this work, we generate a suite of M1- and M2-specific macrophage metabolic models and assess differences in predicted cell composition and metabolic capability. We identify key and consistent metabolic differences between the subtypes' metabolic signatures, increasing our understanding of the role of immune cells in CRC. Our metabolic models can be used to identify novel targets to convert tumor- supporting macrophages towards the M1 state, thus "re-educating" the cells to combat the CRC tumor.

### *26. State-Transition Analysis of Transcriptome Dynamics in BCOR Mutated Acute Myeloid Leukemia*

**Cheyenne L. Anderson** [1], Denis O'Meally [1] , Lisa Uechi [1] , David E. Frankhouser [1] , Ryan Jackson [2], Vanina Tomasian [2], Dongqing Gu [2], Ya-Huei Kuo [3], Guido Marcucci [3], Russell C. Rockne [1, 3], Michelle Afkhami [2]

[1]Division of Mathematical Oncology, Department of Computational and Quantitative Medicine,
City of Hope. USA.

[2] Department of Pathology, City of Hope. USA

[3] Department of Hematologic Malignancies Translational Science, City of Hope. USA.

Acute myeloid leukemia (AML) is a heterogeneous, highly malignant cancer of the blood and bone marrow. We have previously shown that AML can be viewed as a system-wide state-transition of the transcriptome, where critical points can be derived and visualized in a state-space that determines the probability of leukemia development over time. Prior literature has associated *BCOR* as a gene of interest in AML disease progression. This abstract aims to analyze the temporal dynamics of the bulk transcriptome in *BCOR* mutated AML patients. A dimensionality reduction analysis using the singular value decomposition was conducted on time-series RNASeq data from City of Hope's CLIA-certified targeted RNASeq assay derived from AML patients at City of Hope. From these data we were able to create AML transcriptome state-spaces by plotting singular vectors that clearly differentiate control from the leukemic data points.

### *27. Developing an automated live cell microscopy tracking for B-cell lineages*

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

[1]Bioinformatics IDP, University of California Los Angeles. USA.

[2]Signaling Systems, Institute for Quantitative and Computational Biosciences, University of California Los Angeles. USA.

Immune cells respond to various stimuli with complex signaling dynamics and cell fate decisions to survive, die, divide, or differentiate. Both can be captured by live cell microcopy, but linking the two requires a robust cell tracking pipeline that can follow the same cell and its progeny over many microscopy frames

over multiple days. However, tracking B-cells over several days of culture is challenging because they are unattached, move rapidly, and are small and without prominent features. We describe progress towards the development of a new automated tracking algorithm to re-construct in B-cell lineages from microscopy movies. The first step, cell detection based on segmentation, is 97% accurate. For the next step, tracking lineages, we implemented hidden Markov Models with the Viterbi algorithm. To benchmark this workflow, we simulated data and characterized the parameter sensitivities. The preliminary results from these studies will be presented and discussed.

***28. Metabolic fingerprinting HUVEC phenotypic shifts using Fluorescence Lifetime Image Microscopy (FLIM)***

**Karen L. Lopez**, Giulia Tedeschi, Francesco Palomba, Michelle A. Digman.
Department of Biomedical Engineering, University of California Irvine, Irvine, CA, USA.

Adult human vascular endothelial cells (ECs) change their phenotype in response to changes in vascular tumor microenvironments, oxygen concentration and during wound healing. During angiogenesis, ECs undergo an 'angiogenic switch'; they change from a resting (quiescent) phenotype to a more elongated and motile (tip cells) or proliferative and mobile (stalk cells) phenotype. Currently, the most common approaches to differentiate between phenotypes include Seahorse assays, fluorescent labeling or transfections in an attempt to categorize the HUVEC phenotypes by their metabolic signature. Although these methods have shown the importance of metabolic profiling to separate the cells, they are invasive and may ultimately alter the cellular metabolism. We propose to classify HUVECS phenotypes by their metabolic signature by characterizing NADH, a metabolic molecule that is naturally autofluorescent and indicative of the metabolic state (glycolysis or oxidative phosphorylation). These results will provide an important insight on EC phenotype, metabolism, and migration patterns which will benefit studies regarding endothelial dysfunction and cancer metastasis.

***29. Phasor approach FLIM as an indicator for NADPH oxidase during exposure to secondary organic aerosols***

**Yu-kai Huang**, Ting Fang, Jinlai Wei, Jessica E. Monterrosa Mena, Pascale S. J. Lakey, Michael T. Kleinman, Shiraiwa Manabu, Michelle A. Digman

Macrophage exposure to air pollutants has been an interesting topic to study how particle matters affect the human lung. One of the well-known mechanisms to defend inhaled particles is nicotinamide adenine dinucleotide phosphate oxidase (NOX). During activation of NOX, macrophages produce significant amount of reactive oxygen species (ROS) which contributes to lethal impacts such as lipid peroxidation or cancer. In this study, we introduce phasor approach FLIM to monitor the production rate of ROS by measuring the dynamic changes of free to bound NAD(P)H lifetime. The fraction change matches with chemiluminescene readings but drops even more to a longer lifetime area on the phasor plot which is suspected to be oxidized lipids. Other than lifetime changes, FLIM enables localization of NOX activity on live cells which happen on cell membranes. This gives the opportunity to investigate membrane fluidity changes due to these particle matters. In conclusion, phasor approach FLIM of NAD(P)H could serve as monitoring ROS production and provide spatial information of NOX activation.

***30. In-silico and experimental studies show 14-3-3 can increase paradoxical response to RAF inhibitors***

**Gaurav Mendiratta[1]**, Kodye Abbott[1], Yao-Cheng Li[2], Jingting Yu[3], Jianfeng Huang[1], Maxim N. Shokhirev[3], Thomas McFall[4], Geoffrey M. Wahl[2] and Edward C. Stites[1]
[1]Integrative Biology Laboratory, Salk Institute for Biological Studies; La Jolla, CA 92037, USA.
[2]Gene Expression Laboratory, Salk Institute for Biological Studies; La Jolla, CA 92037, USA.
[3]The Razavi Newman Integrative Genomics and Bioinformatics Core (IGC), Salk Institute for Biological Studies; La Jolla, CA 92037, USA.
[4] Department of Biochemistry and MCW Cancer Center, Medical College of Wisconsin; Milwaukee, WI 53226, USA.

Mutations in *RAF* are found in almost 8% of cancer patients. FDA approved, kinase inhibitor therapies to treat active RAF mutant cancers have existed for over a decade. However, a phenomenon of drug-induced activation of RAF signaling mediates side-effects and limits wider use of these inhibitors. One mechanism which explains this phenomenon is based on conformational autoinhibition equilibrium of RAF. We create an in-silico representation of RAF activation which includes this mechanism. We analytically solve this model to identify conditions under which paradoxical activation may arise. We also show that a protein which could modulate the inactive state of RAF could generate additional paradoxical activation. Our experiments utilizing such a protein called 14-3-3 validate our predictions. Our work explains the details of how this puzzling phenomenon of inhibitor-induced-activation may arise in the case of RAF and identifies a potential resistance mechanism to next generation drugs that target RAF.

***31. Understanding bacterial proteasome assembly***

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

Tuberculosis affects over 8,000 individuals in the United States every year. *Mycobacterium tuberculosis* (Mtb) able to resist the immune system in part through the function of the proteasome, which is a large macromolecular structure responsible for the degradation of misfolded or short-lived proteins in cells. The enzymatically active part of this complex, known as the Core Particle (CP), is composed of four stacked heptameric rings in a barrel-like structure. The Mtb CP is assembled from two types of monomers, α and β, and is only active once fully assembled. Here, we focused on understanding the mechanisms that regulate proteasome assembly in the Mtb system using both mathematical modeling and experimental analyses. Studying the regulation of CP assembly should yield insights that are critical to the development of a new class of therapies aimed at targeting tuberculosis infection. I hypothesized that Mtb proteasome has evolved a set of mechanisms that maximize yield and thus bacterial immune resistance. We used a mathematical model to determine the role kinetic parameters play in ring formation and ultimately proteasome assembly. We showed that a mathematical model of stacked ring assembly has interesting dynamics and can produce deadlock, a type of kinetic trapping. Our results suggest that the details of the assembly pathway play a critical role in determining assembly kinetics. In particular, the dynamics in the assembly of stacked rings depends on the binding affinities between rings and between monomers, and we see significant deadlock depending on the binding strengths between and within rings. This suggests that evolution has likely generated a specific pattern of binding affinities in order to maximize yield in the *Mtb* CP and related structures.

***32. Difference in permeability can result in steady-state enantiomeric excess across an N-membrane protocell***

**Celia Blanco**, Julien Derr, Donna Blackmond, Jack Szostak, Irene Chen Department of Chemical and Biomolecular Engineering. UCLA Samueli School of Engineering.
University of California, Los Angeles

There is no clear answer as to why ribose is the sugar in the nucleic acid backbone. Similarly, we do not know why sugars are D and amino acids are L in life. We know that modern living organisms are organized into cells. Indeed, life can be regarded as a property that emerges when a replicating genome is encapsulated in a semi-permeable membrane that undergoes spontaneous growth and replication. Since compartmentalization seems to be a stepping-stone toward the origin of life, it is of special interest to study mechanisms that favor one substrate over another in protocell systems. It is known that ribose is more membrane permeable than the other aldopentose [1], and more recently, it was found that chiral membranes are more permeable for one enantiomer than the other for prebiotically plausible amino acids [2]. If a membrane is selectively permeable favoring one substrate over another, does a vesicle having multiple lamellae receive amplified enrichment? If so, under what conditions is enrichment amplified? We modeled the effect of multiple lamellae on permeability differences across a protocell membrane using partial differential equations and numerical simulations. Our model can be applied to

both the case of different types of substrate, and the case of two enantiomeric species of the same substrate. Our results indicate that in the absence of metabolism inside a vesicle, there can be a transient imbalance with multiple lamellae; ultimately, at the steady-state, the two types of substrate have equilibrated. However, if the substrates are consumed on the interior, either at the same or different rates, there can be a steady-state imbalance whose magnitude is proportional to the number of lamellae. We also simulate enantioselective polymerization inside the protocell [3] and find that homochiral polymerization is possible even if initial concentrations are zero at every layer, including the interior of the protocell. This is important because the enantioselective polymerization model itself can only lead to homochiral polymers if there is an initial enantiomeric excess in the system.

[1] Sacerdote, M.G. & Szostak, J.W. (2005) Proc. Natl. Acad. Sci. U.S.A. 102 (17) 6004-6008.
[2] Hu J., Cochrane W.G., Jones A.X., Blackmond D.G. & Paegel B.M. (2021) Nat Chem. 13 (8):786-791.
[3] Blanco C. & Hochberg D. (2011) Phys. Chem. Chem. Phys. 13, 839-849

***33. Fractal-like density distributions in single-cell data***

**Timothy Hamilton,** Breanne Sparta, Serena Hughes, Eric J. Deeds
Deeds Lab, Institute for Quantitative and Computational Biosciences, University of California Los Angeles. USA.

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.

***34. A lack of distinct cellular identities in single cell data: revisiting Waddington's landscape***

**Breanne Sparta**[1], Timothy Hamilton[1], Eric J. Deeds[1]
[1]Institute for Quantitative and Computational Biosciences, University of California Los Angeles

Our current understanding of the molecular basis of cellular differentiation rest upon concepts from dynamical systems theory. These concepts are schematized in Waddington's epigenetic landscape: cells move through a landscape of epigenetic constraints, descending through valleys that guide the production of terminally differentiated cell fates. In this representation, gene regulatory networks underly each cell fate, generating attractors in gene expression space and ensuring the production of distinct cell types. Here, cells of a particular type are expected to have similar molecular compositions, phenotypes, and functions. Until recently, transcriptome-wide and other global cellular measurement technologies lacked the single-cell resolution needed to evaluate these foundational assumptions. In this work, we directly characterize how cells from various tissues and organisms are distributed in gene expression space, using epigenetic

data from various single-cell omics technologies. We test the hypothesis that cell-type attractors should produce distinct groups of cells that are localized within dense clusters in gene expression space. To do this, we developed a graph theoretical approach that characterizes distances between cells over a range of thresholds. We found that none of the epigenetic data we analyzed is consistent with the structure predicted by Waddington's landscape. Rather than finding distinct clusters of cells that map cell types to specific regions of gene expression space, we find that cells of very distinct types and lineages occupy the same region of space. We were unable to find any evidence that cell types map to distinct epigenetic states regardless of numerous factors, including: the subset of genes used (whether using feature selection methods or supervised approaches), the measurement modality (whether scRNA-seq, scATAC-seq, MERFISH, or CITE-seq, etc), whether transformations were applied (including various normalization techniques and principal component analysis), and regardless of the type of biological sample measured (including post-mitotic brain tissues). The absence of cell-type clusters in epigenetic space presents both practical and theoretical challenges for the era of single-cell biology. Overall, these findings challenge the idea that cell fates are produced by attractor states of gene regulatory networks, and encourage us to reexamine our assumptions about the molecular basis of cellular differentiation.

### *35. Scalable Spatial Cell Type Mapping of the Mouse Brain with dredFISH*

**Zachary E Hemminger**[1], Gaby Sanchez Tam[1], Fangming Xie, Hanpeng Xu[2], Hong Wei Dong[2], Roy Wollman[1,3]

[1]University of California Los Angeles, Chemistry and Biochemistry, Los Angeles, CA,

[2]University of California Los Angeles, Neurobiology, Los Angeles, CA,

[3]University 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.

### *36. Deciphering the relationship between the DC and the centrosome using proximity- labeling proteomics*

**Surenna Pecchia,** Natasha Jones, Maxwell Z. Wilson

Department of Molecular, Cellular, and Developmental Biology, University of California Santa Barbara

The Wnt pathway converges at the $\beta$-catenin destruction complex (DC), a liquid-like structure that regulates cellular $\beta$-catenin levels and consequently, cell fate decisions. Truncatations in the DC scaffolding protein APC are highly prevalent in colorectal carcinomas. Our lab identified that the centrosome, a key cell cycle regulatory hub, nucleates DC formation into liquid-like biomolecular condensates. Here, we utilize biotin proximity labeling (BioID) to interrogate molecular coupling of the DC and centrosome, as well as how Wnt pathway activation and cancer-causing mutations may alter this structure. After fusing DC scaffolds Axin and APC to BirA*, a promiscuous biotin ligase, we verified a tight biotin labeling radius around

BirA* fusion proteins without perturbation of DC structure or localization. Streptavidin pull-down coupled with liquid chromatography mass spectrometry identified several centrosomal and cell cycle regulatory proteins. We have demonstrated that BioID is a valuable technique for interrogating dynamic protein-protein interactions within mesoscale structures.

### *37. Decoding the non-coding RNA Regulatory Map in Cellular Senescence*

**William M. Sparks**[1], Huiling Huang[2], Mithun Mitra[1,3], and Hilary A. Coller[1,3,4]

[1]Department of Molecular, Cell, and Developmental Biology, UCLA

[2]Bioinformatics Interdepartmental Program, UCLA

[3]Department of Biological Chemistry, David Geffen School of Medicine, UCLA

[4]Molecular Biology Institute, UCLA

Cellular senescence is a state of irreversible cell cycle exit and is associated with age-related chronic diseases and cancers. Identification of senescent cells is challenging due to the lack of universal biomarkers and known regulatory pathways. To address these gaps, we investigated long non-coding RNAs (lncRNAs) in senescent cells as lncRNAs have diverse gene-regulatory roles and are cell- and tissue-specific. Using RNA-seq comparative studies, we identified 67 core senescence lncRNAs for two independent senescence models. Sixty percent of these lncRNAs were upregulated with senescence and >80% were not differentially expressed (DE) with quiescence. 38 core lncRNAs were predicted to participate in a 3-way cellular RNA network comprising of lncRNAs, miRNAs, and mRNAs. Also, 24 core lncRNAs showed a strong propensity to form RNA:DNA triplexes. Taken together, our results indicate multiple layers of potential regulation of protein-coding gene expression and pathways by lncRNAs in cellular senescence.

### *38. From Molecular Networks to Antibody Repertoires: Multi-scale Modeling of B-cell fates*

**Helen Huang**, Haripriya Vaidehi Narayanan, Alexander Hoffmann

Signaling Systems Laboratory, Institute for Quantitative and Computational Biosciences (QCB), and Department of Microbiology, Immunology, and Molecular Genetics (MIMG), University of California Los Angeles. USA.

Vaccines and pathogens elicit an antibody repertoire as a result of stochastic mutation of antibody genes and selection of those with high affinity for the antigen. With a significant amount of stochasticity and complexity in the process, it is difficult to predict how SNPs or pre-existing conditions affect the vaccine-responsive antibody repertoire. We thus propose to develop a quantitative multi-scale model to bridge molecular mass action kinetics, stochastic cell-cell interactions, and B-cell evolution to predict the dynamics of antibody repertoire for rational vaccine design. We will use data generated from *in vitro* biochemical assays as well as flow cytometry assays to parametrize the model, and *in vivo* antibody sequencing data under different perturbations to further optimize the model. By simulating B-cell fate decisions and generating lineage trees, we can test if the underlying evolutionary driving force in *in silico* B-cell lineages matches the *in vivo* experimental B-cell phylogeny inferred from antibody sequencing.

### *39. N-ACT: Interpretable Deep Learning Model for Automatic Cell-Type Identification*

**A. Ali Heydari**[1,2], Oscar Davalos[2,3], Katrina Hoyer[2,3], Suzanne Sindi[1,2]

*[1]Department of Applied Mathematics, [2]Health Sciences Research Institute, [3]Department of Molecular and Cell Biology, University of California, Merced.*

A major limitation in most single-cell RNA sequencing (scRNAseq) analysis pipelines is the reliance on manual annotations to determine cell identities, which are time-consuming and laborious. Given the growth in scale of sequenced cells, supervised methods, especially Deep Learning (DL) models, are at the forefront of automatic cell-type identification (ACTI), achieving high accuracy and providing scalability. However, all existing DL models for ACTI lack interpretability and are used as "black-box" models. We present N-ACT (Neural-Attention for Cell Type identification): the first interpretable deep neural network for ACTI that utilizes a novel attention mechanism for identifying "attentive" genes (similar to marker

genes) used to identify cell-types. N-ACT can perform probabilistic annotations of scRNAseq datasets, or transfer labels from manual annotations to new cells (from similar populations). Our results demonstrate that N-ACT outperforms the current state-of-the-art ACTI on all tested datasets and identifies landmark genes that explain its decision-making process.

### *40. Longitudinal changes to perfusion and interstitial flow in a k-luc glioma mouse model from dynamic MRI*

**Ryan T. Woodall**[1], Weidong Hu[2], Tove Olafsen[2], Margarita Gutova[2], Christine E. Brown[3,4], Jennifer M. Munson[5], Russell C. Rockne[1]

[1]Division of Mathematical Oncology, Department of Computational and Quantitative Medicine, Beckman Research Institute, City of Hope, Duarte, California, USA

[2]Department of Stem Cell Biology and Regenerative Medicine, Beckman Research Institute, City of Hope, Duarte, California, USA

[3]Department of Hematology & Hematopoietic Cell Transplantation, Beckman Research Institute, City of Hope, Duarte, California, USA

[4]Department of Immuno-Oncology, Beckman Research Institute, City of Hope, Duarte, California, USA

[5]Department of Biomedical Engineering & Mechanics, Fralin Biomedical Research Institute, Virginia Tech, Roanoke, Virginia, USA

Angiogenesis and tumor invasion are two hallmarks of cancer which may be measured at tissue scale through medical imaging. Longitudinal dynamic contrast-enhanced MRI, coupled with mathematical inverse models of fluid transport, allow for the quantification of perfusion, vascularity, and interstitial fluid transport, and changes to these measurables in time. In this abstract, we track changes in perfusion ($K^{trans}$, $min^{-1}$), vascularity ($v_p$, volume fraction), and interstitial fluid velocity (IFV, mm $s^{-1}$), in a *k-luc* mouse glioma model. Tumor cells are implanted on day 0, and DCE-MRI is performed on days 14, 16, 21, 28, 35, and 42 post-engraftment. From day 14 to day 42, we observe 24.6%, 881%, 114%, 342% changes in mean $K^{trans}$, mean $v_p$, mean IFV, and tumor volume respectively. We aim to use this pre-clinical pilot study to guide our findings with larger scale pre-clinical and clinical studies of gliomas, with the ultimate goal of using these imaging biomarkers as a method of predicting tumor growth, invasion, response to therapy, and aiding surgeons in resection margin selection.

### *41. Multiscale Modeling of Prion Aggregate Dynamics in Yeast*

**Mikahl Banwarth-Kuhn,** Jordan Collignon, Suzanne Sindi

Applied Mathematics Department, University of California Merced, Merced, California, USA

Prion proteins are responsible for a variety of fatal neurodegenerative diseases in mammals but are harmless to Baker's yeast (*S. cerevisiae*) - making it an ideal system for investigating the protein dynamics associated with prion diseases. Most mathematical frameworks for modeling prion aggregate dynamics either focus on protein dynamics in isolation, absent from a changing cellular environment, or modeling prion aggregate dynamics in a population of cells by considering the "average" behavior. However, such models are unable to reproduce *in vivo* properties of distinct yeast prion strains. Here, I will present some recent results addressing how common experimentally observed outcomes depend on population heterogeneity across different biological scales.

### *42. Patterns of gene and microRNA expression identify dynamics that predict AML development*

**David E. Frankhouser**[1,2], Sergio Branchiamore[3], Lisa Uechi[2], Denis O'Meally[4], Lianjun Zhang[5,6], Ying-Chieh Chen[5,6], Man Li[5,6], Guido Marcucci[5,6], Ya-Huei Kuo[5,6], Russell C. Rockne[2]

[1]Department of Population Sciences, City of Hope National Medical Center, Duarte, CA 91010, USA.

[2]Department of Computational and Quantitative Medicine, Division of Mathematical Oncology, City of Hope National Medical Center, Duarte, CA 91010, USA.

[3]Department of Diabetes Complications and Metabolism, City of Hope National Medical Center, Duarte, CA 91010, USA.

[4]Center for Gene Therapy, City of Hope National Medical Center, Duarte, CA 91010, USA.
[5]Department of Hematological Malignancies Translational Science, Hematology & Hematopoietic Cell Transplantation, City of Hope National Medical Center, Duarte, CA 91010, USA.
[6]The Gehr Family Center for Leukemia Research, City of Hope National Medical Center, Duarte, CA 91010, USA.
Cancer dynamics during development are often overlooked by typical tumor-normal comparisons. We applied a state-transition model to explore the expression dynamics of acute myeloid leukemia (AML) using time-series blood draws from a mouse model of AML. We constructed AML state-spaces from both micro-RNA (miR) and messenger RNA (mRNA) transcriptomes then modeled the samples' movement in both state-spaces as a particle undergoing Brownian motion in a double-well potential. Our model proved more sensitive than the current standard of clinical disease detection and was able to predict each sample's disease course. We also identified nonlinear expression dynamics that associate with the stationary points in the state-space of AML development. Finally, we show that the state-space provides a mathematical foundation for integrating the dynamics of mRNA and miR expression with respect to the biological state-space. Our approach provides a novel analytical framework to investigate both cancer development and other biological state-transitions.

**43. ASXL1 mutations in Bohring-Opitz Syndrome dysregulate gene expression & chromatin accessibility**

**Angela Wei**[1,2,3,4], Isabella Lin[2,3,4,5], Meghna Singh[2,3,4], Bianca Russell[6], Valerie Arboleda[1,2,3,5]
[1]Interdepartmental Bioinformatics Program, 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 Computational Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA
[4]Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA
[5]Medical Scientist Training Program, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA
[6]Division of Medical Genetics, Department of Pediatrics, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA
*ASXL1* encodes a putative polycomb protein that regulates epigenetic mechanisms and transcription. *De novo* truncating mutations in *ASXL1* cause Bohring-Optiz Syndrome (BOS), characterized by intellectual disability, developmental delay, seizures, and heart defects. The pathogenesis of *ASXL1* mutations in BOS is not well characterized. To understand BOS, we performed RNA-sequencing on blood (13 BOS & 14 controls) and patient-derived fibroblasts (8 BOS & 16 controls), and ATAC-sequencing on fibroblasts. Both RNAseq datasets showed upregulation of gene expression in BOS patients compared to controls (fibroblasts: 129/177 DEGs, blood: 1166/2118 DEGs, $p_{adj} < 0.05$). Fibroblast ATAC-seq showed increased chromatin accessibility (3036/4336 differential peaks, $p_{adj} < 0.05$) and intersection with RNA-seq revealed strong positive correlation between changes in chromatin accessibility and corresponding gene expression. This suggests direct effects of *ASXL1* on gene expression through chromatin modification. Intersection of RNAseq datasets identified 43 common dysregulated genes, pointing to strong *ASXL1* effects shared between tissues.

***44. Computational analysis of 4-1BB-induced NFκB signaling suggests improvements to CAR cell design***

**Vardges Tserunyan**[1] and Stacey Finley[1,2,3]
[1]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA
[2]Department of Biomedical Engineering, University of Southern California, Los Angeles, CA
[3]Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, CA
Chimeric antigen receptor (CAR)-expressing cells are a modality of adoptive cell therapy against cancer. The dynamics of signaling events initiated by antigen binding depend on the costimulatory domain within the structure of the CAR. One such costimulatory domain is 4-1BB, which affects cellular response via the NFκB

pathway. We developed a mathematical model for describing canonical NFκB signaling activated by 4-1BB. Particularly, we determined the dose response of NFκB activation to a range of antigen concentrations and quantified the sensitivity of our model to the values of model parameters. Next, we used an information-theoretic analysis to propose manipulations to the NFκB signaling network that could improve the performance of CAR-4-1BB cells. Specifically, overexpressing
NEMO and disabling IKKβ deactivation are predicted to increase the mutual information between antigen levels and NFκB activation. Such alterations can fine tune the response of CAR cells to the antigen concentrations they are likely to encounter.

***45. Virtual interrogation of the ISR reveals a biphasic intensity encoding and stress memory landscape***

**Taivan Batjargal**, Ryan Grant, Francesca Zappa, Diego Acosta-Alvear, Max Wilson
Department of Molecular Cell and Developmental Biology/Biomolecular Science and Engineering, UC Santa Barbara, Santa Barbara, CA

The integrated stress response (ISR) detects cellular damage and computes homeostatic or terminal responses. Until now, current methods of interrogating the ISR employed toxic molecules which are inflexible in changes post initiation, and have pleiotropic effects, occluding interrogation of the pure ISR. To overcome these difficulties, we engineered an optogenetic tool to interrogate the pure ISR by creating a photo-switchable stress sensor kinase (opto-PKR). Virtual stress input to the pathway through opto-PKR mimics the canonical ISR signal cascade and under continuous activation exhibits a biphasic response in both transcriptome and proteome. Furthermore, by modulating virtual stress magnitude we find that input magnitude correlates with the peak pathway response of each ISR phase. Lastly, through pulsatile inputs, we show that the ISR remembers past input events and we mapped the sensitivity changes following activation. As such, our work begins to unravel the computation behind ISR's complex behavior.

***46. tauFisher: Circadian time prediction from genomic data.***

**Junyan Duan**[1*], Michelle Ngo[1*], Babak Shahbaba[1,2], John Lowengrub[1,3], Bogi Andersen[1,4,5]

[*]Equal contribution
[1]Center for Complex Biological Systems, University of California, Irvine, CA
[2]Department of Statistics, University of California, Irvine, CA
[3]Department of Mathematics, University of California Irvine, CA
[4]Department of Biological Chemistry, University of California, Irvine, CA
[5]Department of Medicine, Division of Endocrinology, School of Medicine, University of California, Irvine, CA

The circadian clock is a transcription-translation feedback loop that generates intrinsic rhythms with a period of around 24 hours. Since the circadian machinery is intertwined with fundamental biological processes such as metabolism, correlating disease management plans with one's circadian rhythm can maximize clinical outcomes. Given the robustness of the circadian machinery, people have proposed to predict one's internal circadian time using transcriptomic biomarkers. However, existing methods suffer from low accuracy, platform dependency, and/or failure to predict with just one sample. Here we present **tauFisher**, an algorithm that accurately predicts the circadian time from a single sample collected from various species and tissues as well as platforms and resolutions.

***47. Modeling sex differences and tubuloglomerular dynamics in the kidney***

**Lingyun (Ivy) Xiong**[1,2], Adam MacLean[1], Andrew McMahon[2]
[1]Dept. of Quantitative and Computational Biology,
[2]Dept. of Stem Cell Biology and Regenerative Medicine, University of Southern California

The kidney autoregulates its blood flow, to counteract natural fluctuations in arterial blood pressure and facilitate optimal reabsorption of water and solutes from the renal tubule back into the plasma. One essential mechanism in renal responses to blood pressure changes is the tubuloglomerular feedback (TGF) in the

nephron, resulting in sustained oscillations in glomerular filtration and tubular fluid flow, with a characteristic period of 30-45 seconds. Recent studies revealed prominent molecular differences between male and female kidneys in the rodents, especially among the proximal tubule cells, where over 70% of NaCl and water reabsorption occur. In this study, we investigate the impact of the observed sex differences on TGF dynamics and system physiology, to explain known sex differences in susceptibilities to kidney diseases and hypertension. Our tissue-level low-dimensional mathematical model of the TGF system recapitulates the oscillatory patterns of tubular fluid flow in the nephron. Bifurcation analysis reveals how the oscillatory states depends on the physiological parameters of the model, where sex differences operate. Extension of the mathematical model to the cellular and organ levels would allow for a multiscale understanding of the link between the observed molecular differences and renal physiology.

***48. Impact of age and sex differences on the transcriptome of bone marrow monocytes***

**Juliana E Shin**[1,2], Pijus K Barman[1,2], Di Wu[3], Yizhou Wang[3], Berenice A. Benayoun[4], Helen S Goodridge[1,2,5]

[1] Regenerative Medicine Institute, Cedars-Sinai Medical Center, Los Angeles, CA

[2] Research Division of Immunology, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA

[3] Applied Genomics, Computation & Translational Core, Cedars-Sinai Medical Center, Los Angeles, CA

[4] Department of Gerontology, University of Southern California, Los Angeles, CA

[5] David Geffen School of Medicine at UCLA, Los Angeles, CA

Monocytes, which are produced by hematopoietic stem cells in the bone marrow, can patrol the circulation or be recruited to tissues and differentiate into macrophages and monocyte-derived dendritic cells. Aging is associated with an overall decline in immune function, known as "immunosenescence", and monocytes correspondingly undergo gene expression changes with age. However, aging-associated changes in monocyte gene expression and their impact on monocyte function are not well defined. In this study, we performed single-cell RNA sequencing of classical bone marrow monocytes isolated from young and aged mice of both sexes, allowing us to assess the impact of age and sex on monocyte gene expression and heterogeneity. We found that sex is a stronger driver of variation in monocyte gene expression than age. We also found that several *H2* (MHC) genes were upregulated with age in both sexes, correlating with increased MHCII protein expression. Moreover, we observed elevated expression of a long non-coding RNA, *Aw112010*, which may be responsible for increased MHCII expression during aging in male and female mice. In ongoing studies, we are further investigating the sexual dimorphism of aging-related monocyte gene expression.

***49. The Destructosome: A Biophysical Mechanism for Coordinating* Wnt Signaling with the Cell Cycle**

**Hannah Lock, Alex Wang,** Maxwell Z. Wilson

Department of Molecular Cellular and Developmental Biology, University of California, Santa Barbara. USA

The construction and maintenance of biological tissues requires close coordination between cell proliferation and identity. These morphogenic processes cooperate adaptively to ensure the proper spatial development in proliferating and differentiating cells. The Wnt signaling system is an essential morphogenic pathway thought to direct both proliferation and differentiation, but how these distinct behaviors are coordinated by a single pathway simultaneously is unknown. The recent finding that the β-catenin Destruction Complex (DC)—the pathway's central signaling node—forms phase-separated droplets nucleated by the centrosome suggests that coordination between the cell cycle and Wnt activation states may occur within this newfound structure. Combining measurements of DC material state and signaling output in live cells, we find that cell cycle synchronization reduces the heterogeneity of partitioning of the DC scaffold Axin into centrosomal droplets and amplifies the transcriptional response to upstream Wnt input. These results support the idea that cell cycle phase alters Wnt signal processing through the DC via changes to its material state and provide a mechanism for integrating morphogenic inputs with proliferation.

***50. Single-Cell Parameter Inference of Calcium Pathway Models***

**Xiaojun Wu**[1, 3], Evan Maltz[2], Roy Wollman[2], Adam MacLean[1]
[1]Department of Quantitative and Computational Biology, University of Southern California
[2]Department of Chemistry and Biochemistry, University of California, Los Angeles
[3]Ph.D. in Computational Biology and Bioinformatics Program,
University of Southern California, Los Angeles, CA, USA

The Calcium signaling pathway is an important biological pathway in cells, which has been extensively studied using mathematical models built with systems of ordinary differential equations. Methods have been developed for learning parameters in those models from measured cellular Calcium response. The availability of single-cell gene expression data from live cells has opened up new possibilities for studying Calcium pathway models. We propose a new method that performs parameter inference for Calcium pathway models using single-cell Calcium response data together with single-cell gene expression data. Our method accelerates parameter inference and reveals roles of model parameters in Calcium dynamics and their relationship with Calcium signaling-related genes.

***51. Increasing prediction performance of colorectal cancer disease status using random forests classification based on metagenomic shotgun sequencing data***

**Yilin Gao**, Zifan Zhu, Fengzhu Sun
Department of Quantitative and Computational Biology, University of Southern California, 1050 Childs Way, RRI 201, Los Angeles, CA, United States

Dysfunction of microbial communities in various human body sites has been shown to be associated with a variety of diseases raising the possibility of predicting diseases based on metagenomic samples. Although many studies have investigated this problem, there are no consensus on the optimal approaches for predicting disease status based on metagenomic samples. Using six human gut metagenomic datasets consisting of large numbers of colorectal cancer patients and healthy controls from different countries, we investigated different software packages for extracting relative abundances of known microbial genomes and for integrating mapping and assembly approaches to obtain the relative abundance profiles of both known and novel genomes. The random forests (RF) classification algorithm was then used to predict colorectal cancer status based on the microbial relative abundance profiles. Based on within data cross-validation and cross-dataset prediction, we show that the RF prediction performance using the microbial relative abundance profiles estimated by Centrifuge is generally higher than that using the microbial relative abundance profiles estimated by MetaPhlAn2 and Bracken. We also develop a novel method to integrate the relative abundance profiles of both known and novel microbial organisms to further increase the prediction performance for colorectal cancer from metagenomes.

***52. Determining the Quantitative Relationship Between Erk Dynamics and Human Embryonic Patterning***

**Naomi Baxter**, Isobel Whitehead, Maani Bahador, Lauren Leung, Maxwell Wilson
University of California Santa Barbara, Department of Molecular Cell and Developmental Biology

It is becoming increasingly evident that one of the ways that cells interpret and encode information into multiple cell fates is by multiplexing information through dynamic encoding. Here we ask: does the MAPK/Erk pathway govern complex cell decisions through pathway dynamics? We combine 2D-micropatterning, cellular optogenetic control and live cell kinase activity reporters to probe the role of Erk signaling dynamics in human embryonic patterning. We utilize our computational pipeline to quantify the relationship between Erk activity and patterning of the three germ layers. Specifically, when Erk activity is increased by the addition of Erk activating growth factors, the width of the mesoderm band is increased. When Erk activity is decreased using Erk pathway inhibitors, the mesoderm band is either significantly reduced or eliminated depending on the inhibitor. Overall, this suggests that Erk plays a direct role in determining the width of the mesoderm band in the 2D gastruloid model.

### *53. Predicting phenotype to mechanotype relationship in cells based on intra-cellular signaling networks*

**Esra Tiftik**[1][2] Amy Turnlund[1] Stephanie I. Fraley, Ph.D.[2] Parag Katira,Ph.D.[1]
[1] Mechanical Engineering Department and Computational Science Research Center, San Diego State University, San Diego, California, USA
[2] Bioengineering Department, University of California, San Diego, San Diego, California, USA

Cells originating from the same tissue can respond differently to external signals depending on the genotypic and phenotypic state of the cell and its local environment. For example, normal breast epithelial cells do not respond to stress hormone signaling in discernable manner, whereas highly metastatic breast epithelial cancer cells respond by increasing migration and contractility. We have developed a semi-quantitative, computational model to analyze these intra-cellular signaling networks and their outcomes in the presence of multiple external signals including growth factors, hormones, and extracellular matrix. The model uses a hybrid boolean approach combined with compartmentalized kinetic modeling to predict how detailed cytoskeletal biochemical interactions influence mechanical response of cells in different phenotypic states. We apply this model to understand how small differences in environmental effectors can alter stress hormone driven biomechanical changes in breast cancer cells.

### *54. Destabilizing mutations in collagen systems promote tumor progression*

**Marissa L. Heintschel**[1], Fan Zheng[2], Marcus R. Kelly[2], Stephanie I. Fraley[1], Trey Ideker[2]
[1] Department of Bioengineering, University of California San Diego, La Jolla, CA
[2] Division of Genetics, Department of Medicine, University of California San Diego, La Jolla, CA

The development of cancer is a multistep process resulting in a vast array of genetic mutations. Due to the heterogeneous nature of cancer, interpreting the functional consequences of these mutation landscapes poses significant challenges. Previously, individual mutations in matrix genes were deemed insignificant. However, with a new statistical method called HiSig, we identified fibrillar collagen systems with significant mutational frequencies that were overlooked previously. To assess the functional impact of these mutations, we generated mutated and non-mutated collagen matrices using engineered fibroblasts to mimic the tumor microenvironment. Lung cancer cells were seeded on top of the decellularized matrices and assessed for proliferation and migration differences. Higher Ki67 signal was observed in cells seeded on the mutant collagen matrix, suggesting that collagen mutations aid in tumor progression. Ongoing work aims to characterize the matrices' biophysical and biochemical properties to establish how these mutations impact cellular behavior.

### *55. DeepDecon: A Deep-learning Method for Estimating Cell Fractions in Bulk RNA-seq Data with Applications to AML*

**Jiawei Huang**[1], Andres Stucky[2], Jiang F. Zhong[2], Fengzhu Sun[1]
[1]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089 USA
[2]Department of Basic Sciences, School of Medicine, Loma Linda University, Loma Linda, CA 92350 USA

Understanding the cellular composition of a disease-related tissue is important in disease diagnosis and downstream treatment. We present DeepDecon, a deep neural network model that uses gene expression information to predict the fraction of cancer cells in RNA-seq tissue data. DeepDecon was trained on single-cell RNA sequencing (scRNA-seq) data and was robust to bias and noise, making complex data preprocessing and profile feature selection unnecessary. When applied to bone marrow data, DeepDecon outperforms existing decomposition methods in both accuracy and robustness. We further show that cell number in bulk RNA-seq tissue is positively associated with decomposition precision. The stability and flexibility of DeepDecon make it an efficient tool to analyze the cellular composition of disease-related tissues.

### *56. Internalization of Aβ oligomers is highly active through a lipid-mediated mechanism is age- and disease-dependent*

**Balam Benítez-Mata**[1,2], Gregory Brewer[1,3], Michelle A. Digman[1,2]
Department of Biomedical Engineering, University of California Irvine[1]
Laboratory for Fluorescence Dynamics, University of California Irvine[2]
Institute for Memory Impairments and Neurological Disorders, University of California Irvine[3]
Keywords: spectral phasors, fluorescence microscopy, lipids, Alzheimer's Disease, Beta Amyloid
Alzheimer's disease is a neurodegenerative disease caused by the accumulation of amyloid beta (Aβ) oligomers and tau protein neurofibrillary tangles. Recent studies have focused on understanding the clearance mechanism for Aβ and disturbances at different processing stages, from the plasma membrane to the endo-lysosomal pathway. The endocytic mechanism by which Aβ is internalized is suggested to be lipid mediated, although there is a lack of evidence about the endocytic events after the anchoring to the plasma membrane and before the processing into late endosomes and lysosomes. Here, we investigate the lipid environment effects on the internalization of Aβ through a lipid-mediated mechanism and the changes in the amount and size of Aβ+ vesicles. We use the solvatochromic properties of the fluorescent environment-sensitive probe nile red to target lipids and their environment through the spectral phasor approach, to quantify vesicle distributions we use Phasor analysis of Local Image Correlation Spectroscopy (PLICS).

### *57. Modeling the global signaling state changes during bypass AXL activation in erlotinib-resistant lung cancer cells*

**Marc Creixell**[1], Scott Taylor[1], Jacqueline Gerritsen[2], Carmen Boixo[1], Forest M. White[2] and Aaron S. Meyer[1]
[1] Bioengineering Program, University of California, Los Angeles, Los Angeles, USA
[2]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, USA

The clinical benefit of receptor tyrosine kinase (RTK)-targeted therapies is invariably limited by drug resistance. A major mechanism of acquired resistance involves "bypass" switching to alternative pathways driven by non-targeted RTKs that reactivate proliferation. Overexpression of the RTK AXL is frequently observed in bypass resistant tumors which, in addition to cell survival, drives further malignant phenotypes such as EMT and migration. To explore these coordinated effects, we generated a panel of mutant PC9 cell lines in which each AXL intracellular tyrosine is mutated to phenylalanine, and measured phosphorylation signaling alongside cell viability, cell death, migration, and an erlotinib-induced cell clustering effect. We developed a tailored clustering method for mass spectrometry-based phosphorylation measurements that probabilistically clusters phosphosites based on both their phosphorylation responses and peptide sequences. A PLSR model using phospho-peptide clusters accurately predicted each phenotypic response. Our results indicate that ABL and Src-family kinases, upstream of two clusters, modulate cell proliferation via YAP whereas CK2 and its downstream cluster affect DNA damage response.

### *58. Phage-bacteria contig association prediction with a convolutional neural network*

**Tianqi Tang**[1], Shentwei Hou[2,3], Jed Fuhrman[3], Fengzhu Sun[1].
[1]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089, USA.
[2]Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China.
[3]Marine and Environmental Biology, Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089, USA
Motivation: Phage-host associations play important roles in microbial communities.
But in natural communities, as opposed to culture-based lab studies where phages are discovered and characterized metagenomically, their hosts are generally not known. Several programs have been developed for predicting which phage infects which host based on various sequence similarity measures or

machine learning approaches. These are often based on whole viral and host genomes, but in metagenomics-based studies we rarely have whole genomes but rather must rely on contigs that are sometimes as short as hundreds of bp long. Therefore, we need programs that predict hosts of phage contigs on the basis of these short contigs. Although most existing programs can be applied to metagenomic datasets for these predictions, their accuracies are generally low. Here we develop ContigNet, a convolutional neural network-based model capable of predicting phage-host matches based on relatively short contigs and compare to previously published VirHostMatcher and WIsH.
Results: On the validation set, ContigNet achieves 72-85% area under the receiver operating characteristic curve (AUROC) scores, compared to the maximum of 68\% by VirHostMatcher or WIsH for contigs of lengths between 200 bps to 50 kbps. We also apply the model to the Metagenomic Gut Virus (MGV) catalogue, a dataset containing a wide range of draft genomes from metagenomic samples and achieved 60-70% AUROC scores while VirHostMatcher and WIsH fulfilled 52%. Surprisingly, ContigNet can also be used to predict plasmid-host contig associations with high accuracy, indicating a similar genetic exchange between mobile genetic elements and their hosts.

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

**Farnaz Mohammadi**[1], Shakthi Visagan[1], Sean M. Gross[2], Luka Karginov[3], JC Lagarde[1], Laura M. Heiser[2], Aaron S. Meyer[1]
[1]Department of Bioengineering, University of California, Los Angeles
[2]Department of Biomedical Engineering and Knight Cancer Institute, Oregon Health and Science University, Portland
[3]Department of Bioengineering, University of Illinois, Urbana Champaign

The predominant methods of quantifying tumor-drug response operate on snapshot population-level measurements and therefore lack evolutionary dynamics. To preserve the relationship between the cells in a lineage structure, we used single cell tracking data of phenotypic measurements, and applied a tree-based hidden Markov model (tHMM) to learn the characteristic patterns of single cell heterogeneity and state transitions. The commonly used algorithms of hidden Markov models were manipulated to adapt with the tree-based structure of the cellular lineage data. We benchmarked our model using measurements of cell-cycle phase-specific cell fate and lifetime when treated with chemotherapy and growth factors. Our model identified subpopulations with distinct dynamical structure and quantified the effect of drugs in terms of cell cycle progression and arrest as well as cell death. 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.

### *60. Using a single cell transcriptomic approach to elucidate differences between volar and non-volar skin in humans*

**Julie Wiedemann**[1], Allison Billi[2], Johann Gudjonsson[2], Bogi Andersen[3,4,5]
[1]Mathematical, Computational, and Systems Biology; Departments of [3]Biological Chemistry and [4]Medicine, Division of Endocrinology, and [5]Institute for Genomics and Bioinformatics, University of California, Irvine, California; [2]Department of Dermatology, University of Michigan Medical School, Ann Arbor, MI

Volar skin, the thick skin located on the palms and soles, differs from the non-volar skin, which covers the rest of the body, in several key ways. These include increased epidermal thickness, presence of rete ridges, as well as specific patterns of keratinocyte expression, all of which contribute to the increased mechanical strength of volar skin. Additionally, volar skin lacks both hair follicles and sebaceous glands. These differences have been well characterized on a global level previously. However, transcriptomic analysis to understand cell type specific differences on the single cell level has yet to be undertaken and may be important to understanding the differences in disease susceptibility between volar and non-volar skin.

### *61. Core Regulatory Circuits in Cell Cycle Arrest*

**Sean Cheah**[1], Mithun Mitra[1,2], and Hilary A. Coller[1,2,3]
[1]Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles (UCLA), Los Angeles, CA, USA, 90095,
[2]Department of Biological Chemistry, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA, 90095,
[3]Molecular Biology Institute, UCLA, Los Angeles, CA, USA, 90095

Cellular quiescence and senescence, both states of cell cycle arrest, are known to be important across a wide range of processes ranging from embryogenesis to tumor suppression. Although their general characteristics are now well understood, a gene regulatory code for distinguishing between the two has yet to be identified. By mining public and in-house genomic data, we identified differences between their corresponding core regulatory circuits (CRCs), each of which comprise an interconnected group of self-regulated transcription factors (TF). Notably, ETV1 was present in the top scoring CRC for senescent cells only, consistent with its known role of indirectly activating the senescence-associated p53 tumor suppressor. Equally promising is the quiescence-specific inclusion of NFATc2, a known repressor of G0/G1 restriction point kinase CDK4. Given these and other differences between the putative CRCs, we believe further investigation into these TFs will help shed light on the regulatory differences between these two states.

### *62. Optimizing mRNA capture efficacy of highly multiplexed FISH experiments*

**Gabriela Sanchez**, Zachary Hemminger, Roy Wollman
Institute of Biochemistry, Molecular and Structural Biology, University of California Los Angeles. USA

Spatial transcriptomic technologies enable analysis of tissue architecture by analyzing how gene expression patterns are spatially organized. To do so, spatial transcriptomics method must capture the majority of mRNA is the tissue. Highly multiplexed FISH approaches, such as MERFISH provide a powerful tool in detecting hundreds to thousands of mRNAs with reported efficiencies ranging between 80-90%. However, these efficiencies are calculated only based on probability to detect a captured mRNA ignoring the fact that many mRNA are lost during sample preparations. Here we present an optimized pipeline for mRNA capture protocol used for multiplexed FISH experiments. Key changes to the protocol include anchoring of the mRNA to the gel, hybridization of the encoding probe to the mRNA, and efficient clearing of the sample. We demonstrate that these improvements are robust across multiple species, including mammalian cells, mouse, and zebra finch tissue section, and are sensitive enough to capture up to ~1000 transcripts per cell, with a false positive rate of less than 2%.

### *63. Computational methods for large-scale generation of CRISPR-guide RNA molecules*

**Siddharth Sabata**, Chongxu Qiu, Natasha Jones, Surenna Pecchia, Max Wilson Department of Molecular, Cellular, and Developmental Biology
University of California Santa Barbara, Santa Barbara California, USA
* These authors contributed equally.

Designing and testing CRISPR-guide RNA molecules can be done manually, but becomes difficult at a large scale. Roy et al. describes a CRISPR based method for accurate, high throughput genome editing with MAGESTIC (multiplexed accurate genome editing with short, trackable, integrated cellular barcodes). This technology requires hundreds of unique CRISPR-guide RNA molecules to work. Due to the number of guide RNAs required, bioinformatics methods are needed to overcome this issue. We use tools from BioPython and NCBI to enable the high throughput generation of viable CRISPR-guide RNA molecules.

References:

Roy, Kevin R et al. “Multiplexed precision genome editing with trackable genomic barcodes in yeast.” Nature biotechnology vol. 36,6 (2018): 512-520. doi:10.1038/nbt.4137

### *64. A High-throughput Deep Learning Pipeline to Analyze Sector-like Formations in Growing Yeast Colonies*

**Jordan Collignon**, Suzanne Sindi.
Department of Applied Mathematics, University of California, Merced, USA.

Prion proteins are most commonly associated with fatal neurodegenerative diseases in mammals, but are also responsible for a number of harmless heritable phenotypes in *S. cerevisiae*. While yeast colonies usually grow circular with uniform white or pink colors related to the fraction of normal (non-prion) protein, mild experimental manipulations can change protein aggregation dynamics causing the formation of red sector-like regions indicating disease loss within the colony. Images of sectored yeast colonies provide a rich data set for uncovering relationships between molecular processes and colony-level phenotypic transitions. I present ongoing work on an automated pipeline to segment and extract detailed information about sector-like regions in growing yeast colonies. My pipeline uses a convolutional neural network and circle Hough transform to quantify shape, size, and frequency of sectors. This approach aims to streamline structural quantification of yeast colonies from experimental images to offer additional insight into mechanisms driving colony-level phenotypic transitions.

### *65. Initial cell-matrix interactions determine collective migration and morphogenesis programs*

**Alyssa M. Walker**[‡1], Sural K. Ranamukhaarachchi[‡1], Man-Ho Tang[2], Wouter-Jan Rappel, Ph.D. [2], Stephanie I. Fraley, Ph.D. [1]
[1]Department of Bioengineering, University of California, San Diego, La Jolla, CA
[2]Department of Physics, University of California, San Diego, La Jolla, CA
‡ These authors contributed equally

Collective rotational and chain migration enable epithelial cells to form ducts and acini during morphogenesis. Understanding the rules guiding these behaviors could inform regenerative medicine and cancer treatment. Using a model system where breast cancer cells seeded sparsely as single cells develop clonally into either spheroids or invasive networks in collagen, we explore the underlying mechanisms. Spheroids arise from coherent rotational migration, while networks arise from non-coherent invasive migration. To identify the key cell-ECM interactions driving these distinct morphogenic outcomes, we conducted nine separate measurements characterizing the four primary processes underlying cell migration: cytoskeletal dynamics, cell-matrix adhesion, matrix remodeling, and contractility. The invasive phenotype is initiated by elongated cells with longer-lived protrusions, suggesting localized degradation that enabled cell spreading. Spheroid formation is initiated by rounded cells with shorter-lived protrusions. Modeling these interactions predicts that global inhibition of matrix degradation exclusively promotes rotational migration, which was experimentally validated through Marimastat treatment.

### *66. Data-driven modal analysis of time-series gene expression for discovery of transcriptional genetic sensors*

**Aqib Hasnain**, Enoch Yeung.
Department of Mechanical Engineering, University of California Santa Barbara. USA.

Accelerating the design of synthetic biological circuits requires broadening the currently available genetic toolkit, which currently limits the contributions of synthetic biology to sensing. Although whole-cell biosensors have been successfully engineered and deployed, particularly in applications such as environmental and medical diagnostics, novel sensing applications necessitate the discovery and optimization of novel biosensors. In this paper, we develop a data-driven approach that combines dynamic mode decomposition and observability analysis to extract salient endogenous genetic sensors for analytes of interest. To demonstrate our method, we discover and characterize 15 reporters for the organophosphate malathion in the host bacterium Pseudomonas fluorescens SBW25. Furthermore, we demonstrate how to enhance malathion sensing via virtual sensors, obfuscating the need to construct complex, multi-component genetic sensors. This library of living malathion sensors can be optimized for

use in environmental diagnostics while our machine learning tool can be applied to discover genetic sensors for many other environmentally and medically relevant analytes.

***67. Phasor Spectral and Lifetime imaging for the real time determination of cell state of living cells***

**Lorenzo Scipioni**[1], Atena Zahedi[2], Angel Balam Benitez Mata[1], Giulia Tedeschi[1], Michelle A. Digman[1], Aileen J. Anderson[2], Enrico Gratton[1]

[1] Department of Biomedical Engineering, University of California, Irvine (USA)

[2] Stem Cell Research Center, University of California, Irvine (USA)

Fluorescence Lifetime Imaging Microscopy (FLIM) and spectral imaging are two broadly applied methods for increasing dimensionality in microscopy. We developed Phasor S-FLIM, a novel technology capable of characterizing both spectral emission and fluorescence lifetime of environment-sensitive dyes (ESDs). ESDs are fluorophores that change their spectral emission and/or fluorescence lifetime according to changes in the nano-environment in which they localize, allowing us to measure important physiological parameters such as mitochondrial membrane potential, lipid droplets composition, chromatin compaction and membrane fluidity at the same time. We will present a series of applications of this framework, ranging from the study of the physiological profile of cells exposed to stress treatments, to the real time tracking of the differentiation state in PC12 cells, to the physiological fingerprinting of the differentiation human state of neural stem cells. Overall, our approach allows continuous monitoring of the differentiation state of living cells in a number of systems.

***68. Pathways' prediction of drug adverse events using PathFX***

**Mohammadali Alidoost**, Jennifer L Wilson.

Lab for the Understanding of Network Effects (LUNE), Department of Bioengineering, University of California Los Angeles. USA.

Preclinical prediction of drug-induced safety events is of the utmost importance. We aimed to benchmark PathFX [1], a phenotypic pathways method using protein-protein interactions, in predicting side-effects from drug labels. We used a drug toxicity dataset containing active ingredient-side-effect pairs extracted from drug labels [2]. We mapped active ingredients to DrugBank identifiers and developed an ensemble approach to find PathFX phenotypes relevant to labeled side-effects. To test PathFX predictions, we made an end-to-end approach by comparing our drug network model predictions to drug labels side-effects and calculated multiple evaluation metrics. Consequently, we directly mapped 32 of the 34 drug toxicity side-effects to a total of 65 PathFX phenotypes; 32 were direct matches and 33 were synonymous phenotypes. For the evaluation, we reported the confusion matrix for 869 mapped drugs. PathFX prediction accuracy varies by drug and side effect. Future work will consider additional data sources improving pathways-based prediction of side-effects.

[1] Wilson, Jennifer L., et al. "PathFX provides mechanistic insights into drug efficacy and safety for regulatory review and therapeutic development." PLoS computational biology 14.12 (2018): e1006614. [2] Wilson et al, in preparation.

***69. Characterizing the role of cell-cell communications in mouse melanoma tumorigenesis using single-cell transcriptomics***

**Hui Xiao**[1], Chi-fen Chen[2], Jessica Shiu[2], Rolando Ruiz-Vega[1], Arthur D Lander[1], Anand K Ganesan[2]

[1]Center for Complex Biological Systems, University of California, Irvine, Irvine, CA, USA, [2]Department of Dermatology, University of California, Irvine, Irvine, CA, USA.

Melanocytic nevi are benign neoplasms of melanocytes yet its malignant counterpart melanoma is the deadliest type of skin cancer. Though both lesions often acquire the BrafV600E oncogene mutation, melanocytic nevi undergo growth arrest whereas melanomas do not. Studies showed that tumorigenesis is tightly associated with the tumor microenvironment (TME). To identify the cell-cell communication networks in the TME, we performed single-cell RNA sequencing on nevi or melanoma bearing mouse skin. We identified populations of cancer-associated fibroblast (CAFs) and tumor-associated endothelial cells

(TECs) that were enriched in the tumor tissue. Looking into the melanocytic lineages, we characterized a population of melanocytic tumor, which we termed the "S" cells (Schwann-cell-like precursor cells), that were previously identified in Braf-inhibitor resistant patient-derived xenografts (PDXs). Collectively, these distinct populations may alter the TME and contribute to melanoma tumorigenesis. We currently aim to further elucidate the cell-cell communication networks among these cell types to identify novel pathways for rational drug design in melanoma treatment.

***70. Entropic Analysis of Antigen-Specific CDR3 Domains Identifies Essential Binding Motifs***

**Alexander M Xu***[1,2], William Chour[1,3,4], Diana C DeLucia[5], Yapeng Su[1], Ana Jimena Pavlovitch-Bedzyk[7], Rachel Ng[1], Mark Davis[8,9,10], John K Lee[5,6], James R Heath*[1]

[1]Institute for Systems Biology, Seattle, USA.
[2]Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, USA.
[3]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, USA.
[4]Keck School of Medicine, University of Southern California, Los Angeles, USA.
[5]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, USA.
[6]Division of Medical Oncology, Department of Medicine, University of Washington, Seattle, USA.
[7]Computational and Systems Immunology Program, Stanford University School of Medicine, Stanford, USA.
[8]Institute for Immunity, Transplantation and Infection, Stanford University School of Medicine, Stanford, USA.
[9]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, USA.
[10]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, USA.

T cell receptor (TCR) sequences are an essential pillar of adaptive immunity. TCR repertoires containing thousands of putative antigen-specific sequences are generated, but decoding the specificity of TCR recognition remains challenging. Bioinformatic approaches are often limited to analyzing TCRs with similar length CDR3 variable regions, and with specificities against similar antigens. We present SPAN-TCR as a tool that permits length-agnostic comparisons of TCR CDR3 sequences, and accommodates comparisons of amino acids and amino acid k-mers. When integrated with entropic analysis, SPAN-TCR identifies 2-mer motifs that strongly influence CDR3 diversity for T cell clonotypes specific to the same antigen. We find 'essential' motifs that limit diversity in the same CDR3 α or β chain containing the 2-mer, as well 'super-essential' motifs that limit diversity in both chains. We then expand this analysis, using two databases of viral antigen-specific TCRs, to identify striking similarities in TCR repertoires against different antigens.

***71. scAllele: a versatile tool for the detection and analysis of variants in scRNA-seq***

**Giovanni Quinones Valdez**[1], Ting Fu[2], Tracey Chan[3], Xinshu (Grace) Xiao[1,2,3,4,5,6]

[1]Department of Bioengineering, [2]Molecular, Cellular and Integrative Physiology Interdepartmental Program, [3]Bioinformatics Interdepartmental Program, [4]Department of Integrative Biology and Physiology, [5]Molecular Biology Institute, [6]Institute for Quantitative and Computational Biosciences, University of California, Los Angeles, CA 90095, USA.

Single-cell RNA sequencing (scRNA-seq) data contain rich information at the gene, transcript, and nucleotide levels. Most analyses of scRNA-seq have focused on gene expression profiles, and it remains challenging to extract nucleotide variants and isoform-specific information. Here, we present scAllele, an integrative approach that detects single nucleotide variants, insertions, deletions, and their allelic linkage with splicing patterns in scRNA-seq. We demonstrate that scAllele achieves better performance in identifying nucleotide variants than other commonly used tools. In addition, the read-specific variant calls by scAllele enables allele-specific splicing analysis, a unique feature not afforded by other methods. Applied to a lung cancer scRNA-seq data set, scAllele identified variants with strong allelic linkage to alternative splicing, some of which being cancer-specific. scAllele represents a versatile tool to uncover multi-layer information and novel biological insights from scRNA-seq data.

***72. Deep Learning on Protein Structures for Nucleic Acid Binding Prediction***

**Jared Sagendorf**, Raktim Mitra, Jiawei Huang, Remo Rohs

Protein binding to nucleic acids (NAs) is essential for biochemical processes involved in the regulation, maintenance, metabolism and structural organization of the genome and its transcripts. The ability of a protein to bind nucleic acids with high affinity and specificity is determined by the complex geometrical, chemical and electrostatic properties of its structure and the local spatial arrangement of solvent exposed residue side chains at the recognition site. Structural models of proteins bound to nucleic acids can provide great insight into the mechanisms that drive binding and recognition. However, structural models of protein-NA complexes are scarce relative to those of unbound proteins and recent advances in protein structure prediction indicate this imbalance is likely to further increase. We have a developed a novel deep learning-based approach for predicting protein-nucleic acid binding from the structure of an unbound protein, which we call GeoBind. Our method uses a graph representation of protein molecular surfaces and graph neural networks to predict nucleic acid binding based on the spatial distribution of chemical, geometrical and electrostatic properties of protein surfaces. We show that our models outperform other state of the art predictors on six benchmark datasets, generalize to structures of unbound proteins, can be applied to both experimental and predicted protein structures. As a case study, we applied our models to the HIV-1 restriction factor APOBEC3G and show our predictions are consistent with experimental RNA binding data. General applications of GeoBind include the discovery of new protein functions, identification of functional binding residues given a protein structure, better interpretation of biochemical data, and provide prior information about binding sites for modelling of protein-NA complexes (e.g. docking).

***73. RVAgene: generative modeling of gene expression time series data***

**Raktim Mitra**, Adam L MacLean

Methods to model dynamic changes in gene expression at a genome-wide level are not currently sufficient for large (temporally rich or single-cell) datasets. Variational autoencoders offer means to characterize large datasets and have been used effectively to characterize features of single-cell datasets. Here, we extend these methods for use with gene expression time series data. We present RVAgene: a recurrent variational autoencoder to model gene expression dynamics. RVAgene learns to accurately and efficiently reconstruct temporal gene profiles. It also learns a low dimensional representation of the data via a recurrent encoder network that can be used for biological feature discovery, and from which we can generate new gene expression data by sampling the latent space. We test RVAgene on simulated and real biological datasets, including embryonic stem cell differentiation and kidney injury response dynamics. In all cases, RVAgene accurately reconstructed complex gene expression temporal profiles. Via cross validation, we show that a low-error latent space representation can be learnt using only a fraction of the data. Through clustering and gene ontology term enrichment analysis on the latent space, we demonstrate the potential of RVAgene for unsupervised discovery. In particular, RVAgene identifies new programs of shared gene regulation of *Lox* family genes in response to kidney injury. All datasets analyzed in this manuscript are publicly available and have been published previously. RVAgene is available in Python, at GitHub.

**74. DNA Binding Specificity of Forkhead Transcription Factors Revealed by Multi-step Alignment of SELEX-seq Data**

Brendon H. Cooper[1], Ana Carolina Dantas Machado[1], Oscar M. Aparicio[2,3], and Remo Rohs[1,3,4]

[1] Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089, USA

[2] Molecular and Computational Biology, Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089, USA

[3] Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90033, USA

[4] Departments of Chemistry, Physics & Astronomy, and Computer Science, University of Southern California, Los Angeles, CA 90089, USA

Quantifying the preferences of DNA binding proteins is an essential step in determining how transcription factors (TFs) interact with their targets in the genome. The Forkhead box (FOX) family of TFs is known to play a crucial role in regulating a variety of key processes from proliferation and development, to tumor suppression and aging. By using high-sequencing depth SELEX-seq, we study the four known FOX homologs in *Saccharomyces cerevisiae*. We are able to precisely quantify the contribution of nucleotide positions flanking the core of the binding sites and measure small-scale differences between closely related homologs. Essential to this process is the alignment of our SELEX-seq reads to a set of candidate core sequences determined by two newly developed strategies of alignment and reprioritization applied to enriched 6-bp or 7-bp sequences.

### *75. Nucleosome SELEX-seq to identify binding affinity and pioneer activity of transcription factors*

**George Wang,** Brendon Cooper, Remo Rohs, Rohs Lab

The regulation of gene expression is centered on the binding of proteins known as transcription factors (TFs) to specific target sites in the genome. Recent advancements in sequencing technology has allowed for high-throughput analysis of transcription factor binding sites through methods such as SELEX-seq. The objective of this project is to adapt SELEX-seq methods for use with a randomized mononucleosome library to identify binding motifs and pioneer activity of forkhead transcription factors. This project will study four forkhead transcription factors – FKH1, FKH2, HCM1, and FHL1. Mononucleosomes will be reconstituted using a 147bp randomized library containing a 101bp randomized region. This large library and randomized region are necessary because of the size of DNA encapsulated by a mononucleosome, and to investigate the influence of nucleosomal position on TF binding affinity. TF-mononucleosome complexes will be formed and isolated using magnetic bead affinity immunoprecipitation, selecting for high affinity sequences. These sequences used as the library for the next successive step. The output at each step will be sequenced using next-generation sequencing and analyzed to identify binding motifs and possible pioneer activity of the forkhead TFs. Computational analysis of this sequencing data will focus on the challenges of analyzing SELEX-seq data from atypically large randomized regions and multiple binding of TFs for individual reads.

### *76. Quantitative Models of Protein–DNA Binding based on Functional Chemical Groups of Standard and Modified Base Pairs*

**Tsu-Pei Chiu**, Satyanarayan Rao, and Remo Rohs

DNA-binding proteins play important roles in various cellular processes, but the mechanisms by which proteins recognize target DNA sites in the genome remain not fully understood. DNA sequences contain functional groups on the edges of the base pairs exposed in DNA grooves that represent physicochemical signatures. These signatures enable proteins to form specific contacts between protein residues and DNA base pairs. Studies of these signatures in DNA grooves can therefore provide mechanistic insights into protein–DNA binding. Existing experimental methods, such as X-ray crystallography, reveal such mechanisms based on physicochemical interactions between proteins and their DNA target sites. However, the low throughput of such methods limits mechanistic insights for selection of many genomic sites. Several high-throughput binding assays enable prediction of potential target sites by determining relative binding affinities of a protein to massive numbers of DNA sequences. Available computational methods based on standard Watson–Crick base pairs assume that the contribution of overall binding affinity is independent for each base pair, or minimum unit in k-mer models such as dinucleotide models. These methods cannot directly determine physicochemical signatures and are not suitable to apply to DNA modifications or non-Watson–Crick DNA base pairs. These variations include DNA methylation, Hoogsteen or mismatched base pairs. The proposed method, DeepRec, is able to predict the relative binding affinities as function of physicochemical signatures and the effect of DNA methylation or other chemical modifications on binding. Current sequence-based modeling methods cannot achieve such mechanistic insights into recognition. Our chemistry-based modeling framework provides a new path towards understanding genome function at a mechanistic level.