Apologies for any mistakes made in names, affiliations, or other information!

POSTER SESSION #1 (12:45 – 14:00)

Mohammadali Alidoost, UCLA

Engineering pathways improves prediction of side effects from protein-interaction network models

---

Preclinical prediction of drug-induced safety events is of the utmost importance. We discovered that downstream proteins, in addition to drug targets, are relevant for predicting drug side effects (SEs). We benchmarked PathFX, a protein-interaction network model, and discovered moderate sensitivity and specificity. In the PathFX mode, sensitivity and specificity are highly dependent on pathway definition and our goal was to engineer pathways by including identified key genes and other omics measurements and measure changes in specificity and sensitivity. We used a drug toxicity dataset containing active ingredient-SE pairs extracted from drug labels. We developed methods to evaluate PathFX performance per SE. We then defined new custom PathFX phenotype pathways using our new platform, PathFX_Gen, by including network proteins associated with true positive (TP) drug networks to evaluate TP and false positive (FP) SE predictions. To generalize our approach to cases with insufficient TP example networks, we generated novel SE pathways using an omics dataset. Baseline prediction performance was low and varied per SE. We modeled 890 drugs and matched drug toxicity SEs using an ensemble approach. For the prediction of 32 SEs, sensitivity and specificity were 0.18 and 0.83, respectively. Hypertension had a prediction sensitivity of 0.54 and a specificity of 0.63. However, despite the low performance, we found 224 and 143 genes associated with TP and FP pathways, respectively. Including key genes, we eliminated over-prediction for side effects with sufficient TP examples. Using the omics data, we found new pathways for Myocardial Infarction that had no signal previously. In conclusion, the discovery of TP hypertension pathways could emphasize the importance of finding core pathway genes. We demonstrated that pathway engineering using omics data can also improve the utility of protein interaction network models for preclinical SE prediction. Future work will consider additional sources for improving pathways-based prediction of SEs.

Lynne Cherchia, USC

Predictive modeling of signal transduction anchored by quantitative imaging

Signal transduction is challenging to investigate because it depends upon both spatial organization of signaling and receptor molecules and alterations of gene expression spanning multiple time and space scales. We argue that mathematical modeling provides the tools to build upon the results from quantitative imaging to analyze and interpret cell signaling across these scales. Quantitative imaging tools such as fluorescence fluctuation spectroscopy methods provide many of the key parameters needed for predictive modeling, including measurements of: protein mobility, oligomerization, and concentrations, as functions of time and space. Comprehensive predictive models can integrate such multi-dimensional and multi-scale data into frameworks that can be used to expand our understanding of biological signaling networks. Here, we present our efforts to combine quantitative imaging and computational modeling to explore JAK-STAT signal transduction, a key pathway in pancreatic beta cells. We demonstrate the ability of fluorescence correlation spectroscopy (FCS) to quantitatively capture protein dynamics of the fluorescently tagged human prolactin receptor, yielding data that can be integrated into a computational model. We use FCS measurements to extend an ordinary differential equation model, incorporating stochastic modeling via the Gillespie algorithm to account for the small numbers of transcription factors present in cells. Preliminary model simulations identify the rate of STAT5 transcription as an important factor in determining STAT5 accumulation in nuclei, which drives downstream pathway outputs such as cell proliferation. We explore this result in the stochastic modeling framework. We demonstrate the utility and robustness of FCS in quantitatively capturing the protein dynamics needed for integration into the computational model. This provides a powerful approach to quantitatively study pancreatic beta cell signaling, iterating between computational and experimental phases, to improve our quantitative understanding of pancreatic beta cells and provide the foundational knowledge needed for developing novel therapeutic approaches.

Tsu-Pei Chiu, USC

DNA-binding proteins play important roles in various cellular processes, but the mechanisms by which proteins recognize genomic target sites remain incompletely understood. Functional groups at the edges of the base pairs (bp) exposed in the DNA grooves represent physicochemical signatures. As these signatures enable proteins to form specific contacts between protein residues and bp, their study can provide mechanistic insights into protein-DNA binding. Existing experimental methods, such as X-ray crystallography, can reveal such mechanisms based on physicochemical interactions between proteins and their DNA target sites. However, the low throughput of structural biology methods limits mechanistic insights for selection of many genomic sites. High-throughput binding assays enable prediction of potential target sites by determining relative binding affinities of a protein to massive numbers of DNA sequences. Many currently available computational methods are based on the sequence of standard Watson-Crick bp. They assume that the contribution of overall binding affinity is independent for each base pair, or alternatively include dinucleotides or short k-mers. These methods cannot directly expand to physicochemical contacts, and they are not suitable to apply to DNA modifications or non-Watson-Crick bp. These variations include DNA methylation, and synthetic or mismatched bp. The proposed method, DeepRec, can predict relative binding affinities as function of

physicochemical signatures and the effect of DNA methylation or other chemical modifications on binding. Sequence-based modeling methods are in comparison a coarse-grain description and cannot achieve such insights. Our chemistry-based modeling framework provides a path towards understanding genome function at a mechanistic level.

Alice Decugis, Aarush Maddela, and Saebean Yi, UCLA

We would like to present our computational research project with the Zheng lab at the Stein Eye Institute of UCLA. As a group of three undergraduates, we are using computational biology to identify new potential genes that are part of the pathway causing Retinis Pigmentosa (RP) degenerative disease. More precisely, we are using a Python algorithm to run a set of 130 genes known to cause Retinis Pigmentosa through the STRING interface which finds protein-protein interactions. By looking at genes that interact with several RP-causing known genes, we can identify potential genes that are also part of the pathway of the disease. We are now working on mapping out what this pathway may look like, and categorizing which step of the process these genes all belong to.

Yi Fu, UCSD
Diego Comparison Theorems for Stochastic (Bio)Chemical Reaction Networks

---

Continuous-time Markov chains are frequently used as stochastic models for chemical reaction networks, especially in the growing field of systems biology. A fundamental problem for these Stochastic Chemical Reaction Networks (SCRNs) is to understand the dependence of the stochastic behavior of these systems on the chemical reaction rate parameters. Towards solving this problem, in this paper we develop theoretical tools called comparison theorems that provide stochastic ordering results for SCRNs. These theorems give sufficient conditions for monotonic dependence on parameters in these network models, which allow us to obtain, under suitable conditions, information about transient and steady state behavior. These theorems exploit structural properties of SCRNs, beyond those of general continuous-time Markov chains. Furthermore, we derive two theorems to compare stationary distributions and mean first passage times for SCRNs with different parameter values, or with the same parameters and different initial conditions. Our proof also yields a method for simultaneously simulating the sample paths of two comparable SCRNs. Our tools are developed for SCRNs taking values in a generic (finite or countably infinite) state space and can also be applied for non-mass-action kinetics models. We illustrate our results with applications to models of chromatin regulation and a network topology arising in Braess' paradox.

Belinda Garana, USC

Drug mechanism enrichment analysis improves prioritization of therapeutics for repurposing

BACKGROUND There is a pressing need for improved methods to identify effective therapeutics for diseases. Many computational approaches have been developed to repurpose existing drugs to meet this need. However, these tools often output long lists of candidate drugs that are difficult to interpret, and individual drug candidates may suffer from unknown off-target effects. We reasoned that an approach which aggregates information from multiple drugs that share a common mechanism of action (MOA) would increase on-target signal compared to evaluating drugs on an individual basis. In this study, we present Drug Mechanism Enrichment Analysis (DMEA), an adaptation of Gene Set Enrichment Analysis (GSEA), which groups drugs with shared MOAs to improve the prioritization of drug repurposing candidates.
RESULTS First, we tested DMEA on simulated data and showed that it can sensitively and robustly identify an enriched drug MOA. Next, we used DMEA on three types of rank-ordered drug lists: (1) perturbagen signatures based on gene expression data, (2) drug sensitivity scores based on high-throughput cancer cell line screening, and (3) molecular classification scores of intrinsic and acquired drug resistance. In each case, DMEA detected the expected MOA as well as other relevant MOAs. Furthermore, the rankings of MOAs generated by DMEA were better than the original single-drug rankings in all tested data sets. Finally, in a drug discovery experiment, we identified potential senescence-inducing and senolytic drug MOAs for primary human mammary epithelial cells and then experimentally validated that EGFR inhibitors can be senolytic.
CONCLUSIONS DMEA is a fast and versatile bioinformatic tool that can improve the prioritization of candidates for drug repurposing. By grouping drugs with a shared MOA, DMEA increases on-target signal and reduces off-target effects compared to analysis of individual drugs. DMEA is publicly available as both a web application and an R package at https://belindabgarana.github.io/DMEA.

Xiaolu Guo, UCLA

Modeling the heterogenous NFκB dynamics of single immune

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. The parameters that are subject to biological variation provide the potential to account for heterogeneity. We estimated parameter distributions using the Stochastic Approximation Expectation Maximization (SAEM) approach and then fit the individual cell data using Bayesian maximum a posteriori (MAP) estimation. Visual inspection revealed an excellent fit with the data. To quantitatively evaluate the fitting performance, we compared the experimental and predicted distributions of NFκB signaling codons. Further, we identified biochemical reactions that may account for the cellular heterogeneity in NFκB dynamics. Our results establish a mathematical modeling tool that may be used

to study the molecular determinants of response specificity and dynamical coding in immune sentinel cells at the single cell level.

Wayne Hayes
UCI

Topological Network Alignment aims to find functionally similar genes or gene products across species, independent of whether they possess sequence similarity. Recently I have published several papers demonstrating that network alignment driven by network topology alone is capable of discovering functional orthologs between species possessing ZERO sequence similarity and no known homology relationship. In this presentation, I will briefly outline the intuition behind my methods, how their functional predictions compare favorably against state-of-the-art sequence based methods, and outline the future of network-based functional prediction.

Holly Huber (USC)
Systematic Bayesian Posterior Analysis Facilitates Hypothesis Formation and Guides Investigation of Pancreatic Beta Cell Signaling

---

Computational models formalize biological hypotheses. This is particularly helpful as hypotheses increase in complexity. For example, a network of intracellular protein interactions can be characterized by a system of ordinary differential equations (ODEs), where variables represent protein concentrations and parameters represent reaction rate constants. When we evaluate such models, most commonly by comparing predictions to experimental data, we also evaluate the hypothesis that it characterizes. If the model fails to explain the data, an alternative biological hypothesis may exist. Bayesian inference characterizes what model parameters are plausible, given data. This is known as the posterior distribution. Evidence of alternative hypotheses can manifest in the posterior for one parameter - that is, the marginal posterior distribution. For example, it can manifest as a marginal that is inconsistent between different data subsets, or, as a marginal that is bimodal. Searching marginal posteriors offers an evaluation strategy that compliments comparing predictions to data; it not only indicates whether an alternative hypothesis may exist, but also associates this evidence with certain parameter(s). These parameter(s) can serve as starting points for model, and thereby hypothesis, refinement. For example, one rate constant of the aforementioned ODE model would implicate one interaction in the network. However, existing approaches to search marginal posteriors are largely ungeneralizable and unsystematic, limiting their applicability. Here, we show that ranking marginals by information gained from data provides a systematic and generalizable way to search for alternative hypothesis evidence; rather than searching at random, one can search per the ranking. We use an established measure, the Kullback-Leibler (KL) divergence, to quantify information gained. Our KL divergence ranking systematically uncovers different manifestations of alternative hypothesis evidence. Further, it outperforms ungeneralizable alternatives, like bimodality search. Overall, our approach facilitates model evaluation; by rigorously evaluating computational models, we better test the hypotheses they characterize.

Geena Ildefonso (USC)

A data-driven Boolean model explains memory subsets and evolution in CD8+ T cell exhaustion

T cells play a key role in a variety of immune responses, including infection and cancer. Upon stimulation, naïve CD8+ T cells proliferate and differentiate into a variety of memory and effector cell types; however, failure to clear antigens causes prolonged stimulation of CD8+ T cells, ultimately leading to T cell exhaustion (TCE). The functional and phenotypic changes that occur during CD8+ T cell differentiation are well characterized, but the underlying gene expression state changes are not completely understood. Here, we utilize a previously published data-driven Boolean model of gene regulatory interactions shown to mediate exhaustion. Our network analysis and modeling reveal the final gene expression states that correspond to TCE, along with the sequence of gene expression patterns that give rise to those final states. By examining the transcriptional gene pattern activation of pro-memory and exhausted-related genes, we found the population of in silico T cells are associated with eight distinct terminal exhausted gene networks, even when starting from identical underlying networks. We also evaluated how perturbing the network compared to cells with the baseline network wildtype. We encoded a PD1 checkpoint blockade in the model by repressing the ability for NFATC1 to activate PD1, a known activator of PD1. The activation of PD1 reveals an oscillatory pattern over time resulting in a shift to terminal pro-memory end state for all seven subsets, suggesting that the PD1 pathway may continue to function as a link of signals required for memory T cell maintenance. With a model that predicts the changes in gene expression that lead to TCE, we could evaluate strategies to inhibit the exhausted state. Overall, we demonstrate that a common pathway model of CD8+ T cell gene regulatory interactions can provide insights into the transcriptional changes underlying the evolution of cell states in TCE.

Francisco Lopes, Universidade Federal do Rio de Janeiro and UCLA

NFkB can induce a stochastic transition from the HER2 to triple negative breast cancer attractors in the Waddington's epigenetic landscape

NFkB induces a stochastic transition from the HER+ to triple negative breast cancer attractors in the Waddington's epigenetic landscape. Using the epigenetic landscape metaphor, Waddington proposes that the diversity of cell types in an organism is generated during development by a series of cell fate decisions like a ball rolling down along bifurcated valleys. Taking advantage of the dynamic systems theory, Kauffman proposed that all cellular types correspond to different attractors in a multidimensional space defined by the diversity of functional unities, its genes, in the organism genome. The Waddington bifurcated fate valleys and the Kauffman multidimensional space are generated by complex Gene Regulatory Network (GRN) established during the evolution. Here, we apply the above-described theoretical frameworks to characterize the breast cancer progression. We used qPCR data from two BC cell lines HCC-1954 (HER+) and MDA-MB-231 (TN), to build and calibrate a GRN model describing a positive regulation of NFkB over itself, TWIST1, SLUG and SNAIL. The epigenetic landscape associated to the model shows one attractor basins associated to each subtype besides an intermediate one with low residence time. Stochastic simulations show spontaneous irreversible transitions from HER+ to TN attractors. Different simulations exhibit the HER+ to TN transition at different times, indicating that stochasticity works in both levels, at the protein level

fluctuation and at the attractor basis transition. Specific variations in the GRN parameters, that can be caused by gene mutations, changes the epigenetic landscape resulting in increased probability in the HER+ to TN transition. The effect of drugs that reduces NFkB transcriptional activity can change the size basins attractors in the epigenetic landscape reducing the HER+ to TN transition probability or even induce inducing the reverse transition. Our results also indicates that a given genotype can produce different phenotype output, each one associated to a particular basis of attraction.

Jonathan Martinez
USC

Gene regulatory network inference with popInfer reveals dynamic regulation of hematopoietic stem cell quiescence upon diet restriction and aging

---

Inference of gene regulatory networks (GRNs) to reveal how cell state transitions are controlled is possible with single-cell genomics data. However, obstacles to temporal inference from snapshot data are difficult to overcome. Single-nuclei multiomics data offer means to bridge this gap and derive temporal information from snapshot data using joint measurements of gene expression and chromatin accessibility in the same single cells. We developed popInfer to infer networks that characterize lineage-specific dynamic cell state transitions from joint gene expression and chromatin accessibility data. Benchmarking against alternative methods for GRN inference, we showed that popInfer achieves higher accuracy in the GRNs inferred. popInfer was applied to study single-cell multiomics data characterizing hematopoietic stem cells (HSCs) and the transition from HSC to a multipotent progenitor cell state during murine hematopoiesis across age and dietary conditions. From networks predicted by popInfer, we discovered gene interactions controlling entry to/exit from HSC quiescence that are perturbed in response to diet or aging.

Gunalan Natesan, UCLA

Graph based Analysis of Embryonic Lineages

The study of cell lineages in eutelic organisms, which possess a fixed number of somatic cells, has been a powerful driving force in our understanding of fundamental developmental and biological processes. Cell lineages map intuitively to binary trees, and C. elegans' eutely property and extensive characterization make it an optimal candidate for the creation and benchmarking of distance metrics that compare tree structures. Here we adapt the tree edit distance, a metric that measures topological variations in trees, and introduce the branch distance, a generalization of the L2 norm, to the analysis of the C. elegans embryonic cell lineages. We benchmark these metrics using a published database of wild type and RNAi-perturbed C. elegans embryos, revealing previously uncharacterized heterogeneity in wild type lineages and in the phenotypic consequences of RNAi variability on developmental timing. Extending these measurements of lineage-specific timing to the detection of patterns of similarity in anterior cell lineages of the embryo and identifies a previously unappreciated role of Notch signaling in the control of developmental timing. Finally, we apply this approach to a systematic analysis of RNAi perturbations that result in cell fate transformations where we find that, while developmental timing

appears to be highly sensitive to genetic perturbation, RNAi against genes in a subset of important developmental regulators generate transformations that preserve lineage-specific developmental clocks.

Brian Orcutt-Jahns, UCLA

Immune evasion by tumors is characterized by dysregulation evident in the peripheral blood

---

Metastatic cancer has been shown to correlate with dysregulation of immune signaling and function. For example, impaired T cell responsiveness to IL-6, or reduced capacity of lymphocytes to response to IFNg has been shown to predict poor clinical outcome in HR+ breast cancer (BC). To systematically profile immune dysregulation in cases of BC, we stimulated human PBMCs from 18 healthy and 18 HR+ BC patients with a panel of 14 cytokines and growth factors. Cells were stained for 22 canonical cell type markers and 5 intracellular signaling proteins, allowing for the dissection of the cytokine responses of 23 immune cell subtypes. We arranged our cytokine response data into a 5D tensor and analyzed responses using tensor factorization techniques, which allowed us to visualize patterns of immune regulation, and thus identify variations in response specific to the BC cohort. We found that the components generated during tensor decomposition of basal and induced immune response were highly predictive of BC status, and that these components could be examined to gain insight as to the origins of cancer related signaling dysregulation. For example, we found that reduced responsiveness to IL-10, increased STAT5 phosphorylation, and increased basal levels of Smad1-2 phosphorylation are highly predictive of BC disease state. These novel findings demonstrate how that our approach represents a comprehensive and effective method for analyzing the inherently tangled network of immune dysregulation common in BC.

Nicolas Pelaez, Caltech and HHMI

During early human development, pluripotent stem cells differentiate and spatially self-organize into distinct lineages driven by secreted BMP, Wnt and Nodal ligands and inhibitors. Whether early human development also requires non-secreted ligands that mediate contact dependent cell-cell signaling through Notch remains largely unknown. Here, using human gastruloid colonies as synthetic models of the cell fate decisions occurring during early human development, we show that Notch signaling regulates differentiation, spatial patterning, and cell fate proportions of extra embryonic and embryonic lineages. Transcriptome measurements made using scRNAseq and spatial expression analysis by seqFISH of the Notch pathway show the Notch signaling components are present before and during gastruloid colony differentiation. Blocking Notch signaling chemically and genetically uncovered that Notch is necessary for cell differentiation and spatial pattern formation since Notch inhibition triggers absence of endodermal fates, reduction or absence of mesodermal and primordial germ cell progenitors, and expansion of extra-embryonic, ectodermal and epiblast-like states. Notch signaling regulates cell fate proportions through a proportion control mechanism that modulates the expression amplitude of key transcription factors that drive lineage-specific differentiation. Since

Notch signaling is necessary for human gastruloid colony differentiation, we suggest a model in which combinations of secreted morphogenes and non-secreted Notch signals pattern early cell fate choices during early human development.

Andrew Ramirez, UCLA

Interpretable and scalable single-cell transcriptomics integration across conditions with PARAFAC2

Single-cell measurements, such as flow cytometry, scRNA-seq, and ATAC-seq, have transformed our understanding of immune cell populations and their cell-to-cell heterogeneity. Many innovative clustering and data reduction tools identify and separate immune cells to study cell type differences in diverse biological contexts. However, current analysis tools, like tSNE and UMAP, cannot simultaneously distinguish cell subpopulations and analyze how cell populations respond across multiple conditions. Understanding how single cells respond across experimental conditions is vital for many biological questions; thus, a flexible technique that can analyze these complex single-cell datasets is imperative to maximize the potential of single-cell measurements. A solution to this problem includes using a data reduction technique PARAFAC2 (Pf2), a form of tensor decomposition that allows unaligned measurements along one dimension. Applying this technique allows us to analyze single-cell measurements across multi-condition and perturbational experiments, removing the constraint that the same single cells are measured for every experimental condition. By utilizing Pf2, we can investigate a myriad of single-cell modalities (proteomics, transcriptomics, etc.), track how experimental conditions results in similar or dissimilar variation across single-cell response, and determine cell-to-cell heterogeneity within experimental conditions. To demonstrate Pf2a's ability to analyze single-cell data with multiple conditions, we have analyzed a synthetic dataset and a perturbational experiment wherein human peripheral blood mononuclear cells were treated with several different drugs. We were able to successfully identify trends across the experimental conditions and highlight single cells with shared features. We believe Pf2 serves as a general method for single-cell experiments to find patterns across conditions.

Allison Schiffman, UCLA

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

---

The expression of type I interferon (IFNβ), a critical determinant of the innate immune response, is controlled by three key transcriptional activators, AP1, NFκB and IRF, which bind to the enhancer region of IFN β. Classic molecular biology studies of IFN β expression suggest that NFκB and IRF function synergistically by forming an enhanceosome complex. However, biochemical evidence shows that NFκB and IRF do not have positive cooperativity in binding, and knockout mouse studies show that NFκB is not always required for IFN β expression. We developed a quantitative model to account for both literature data and new measurements from NFκB and IRF knockout cells. We found that the second IRF site affinity may be tuned by competing repressor factors, which may control the degree to which IFN expression depends on NFκB in response to different stimuli.

Fakhar Singhera, USC

Single-Cell Cytokine Profiling of Bone-Marrow Derived Mesenchymal Stem Cells

Introduction: Charcot neuroarthropathy, also known as Charcot foot, is the progressive degeneration of the joints in the foot, potentially leading to permanent foot deformation is caused by peripheral neuropathy. Peripheral neuropathy, the loss of nerve function in the periphery of the body, results in an increase in pro-inflammatory cytokines which lead to the activation of the receptor activator of nuclear factor κB ligand (RANKL), an apoptosis regulation gene that has been identified to affect bone regeneration and remodeling. [1] In an attempt to better understand the differences between a healthy foot and a Charcot foot, pro-inflammatory cytokines that lead to the activation of the RANKL pathway were studied such as IL-6, IL-8, IL-10, EGF, FGF, VEGF, PDGF-AB, and PDGF-BB. Previous studies have shown that in the primary phase of fracture healing, IL-6, IL-8 and other cytokine factors are present within the first days of the tissue injury. [2]
Materials and Methods: We conducted single-cell cytokine profiling using bone-marrow derived mesenchymal stem cells (ATCC PCS-500-012) to represent non-diseased healthy humans. The culture media and the single-cell suspension was loaded onto the sub-nanoliter polydimethylsiloxane (PDMS) microchamber arrays and combined with the high-density antibody barcode chip to detect for the cytokines listed previously.
Results and Discussion: We analyzed the cytokine level in cultured media by testing for changes in cytokine expression over 3 different cell seeding densities and 4 different time points. Doing this, we were able to observe an increase in IL-8 and VEGF secretions. Cytokine expressions were verified using blot technologies.
Conclusions: The expression of cytokine changes in cultured media over time and that single cells show expression of multiple cytokines in healthy bone-marrow derived mesenchymal stem cells.

Breanne Sparta, UCLA

A prevailing interpretation of Waddington's landscape is that cells with similar physiologies exist within a shared basin of attraction and exhibit similar gene expression patterns. This notion is often applied in the analysis of single-cell omics data, where cells are clustered into groups that represent cells of similar identities, prior to the analysis of differential gene expression. Yet, until the advent of single-cell measurement technologies, it has been impossible to characterize the heterogeneity of cells in the neighborhoods of these attractor basins. In this work, we apply graph theory to characterize the distribution of cells in epigenetic space, using data from various tissues and organisms as well as various single-cell omics technologies. Rather than finding distinct clusters of cells that map cell types to specific regions of epigenetic space, we found that cells of very distinct types and lineages occupy the same region of space. Further, we found that the density distribution of cells is approximately power-law, which is not the density distribution we would expect to see in the neighborhood of an attractor. These two behaviors are universal in single-cell data on epigenetic state of multicellular organisms, regardless of the tissue, organism, measurement technique employed, or the approach used to select the subset of genes on which the analysis was performed. The fact that currently-available single-cell data is inconsistent with the predictions of Waddington's landscape poses a

challenge both for the robust analysis of these data and for our overall understanding of epigenesis in development.

Tianqi Tang, USC

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's host matches based on relatively short contigs, and compare it to previously published VirHostMatcher (VHM) and WIsH.

Niki Tavakoli, USC

Colorectal cancer is responsible for over 150,000 diagnoses and 50,000 deaths per year, characterizing itself as one of the fatal diseases in the United States. Cancer-associated fibroblasts (CAFs), a subtype of stromal cells, have been long known to impact and aid colorectal cancer cells (CRCs). Specifically, we focus on altered central carbon metabolism, which has become a crucial hallmark of cancer. However, the exact changes and mechanisms of these changes and the cells effects on one another remain unknown. Constraint-based modeling, with an integration of metabolomics profiles, is an important approach to better understand the metabolic crosstalk that occurs between CRCs and CAFs, and to apply model perturbations that mimic drug applications in a computational model. In this work, our group utilizes a systems biology approach by incorporating LC-MS metabolomics data (from cell culturing experiments) as constraints placed on our computational models. Additionally, the KRAS oncogene, whose mutation has been found to highly elevate colorectal cancer fatalities, was also looked at when comparing CRCs and CAFs. Our model predictions convey clear variations among key metabolic pathways in central carbon metabolism. Namely, there are clear shifts for mutant-KRAS CRCs co-cultured with CAFs versus CRCs cultured without them. Additionally, we performed single and pair-wise targeted enzymatic knockdowns in the metabolism network that provided us with knowledge of vulnerable reactions that are the best candidates for inhibiting certain areas of metabolic activity and growth. This research allows for efficient insight into a complex system whose full effects are difficult to understand through experimental data alone. The tumor-stromal metabolic crosstalk that is analyzed through our analyses can be used to not only improve the existing metabolic therapies of patients with colorectal cancer, but brings forth and validates novel methods that can be further explored upon and tested in the clinical setting.

Vardges Tserunyan, USC

Systems biology utilizes computational approaches to examine an array of biological processes, such as cell signaling, metabolomics and pharmacology. This includes mathematical modeling of CAR T cells, a modality of cancer therapy by which genetically engineered immune cells recognize and combat a cancerous target. While successful against hematologic malignancies, CAR T cells have shown limited success against other cancer types. Thus, more research is needed to understand their mechanism of action and leverage their full potential. In our work, we set out to apply information theory on a mathematical model of cell signaling to study CAR-mediated activation following antigen encounter. First, we estimated channel capacity for CAR-4-1BB-mediated NFκB signal transduction. Next, we evaluated the pathway's ability to distinguish contrasting "low" and "high" signals depending on the amount of intrinsic noise. Finally, we assessed the fidelity by which NFκB activation reflects encountered antigen concentration, depending on the prevalence of antigen-positive targets in tumor population. We found that in most scenarios, fold change in the nuclear concentration of NFκB carries a higher channel capacity for the pathway than NFκB's absolute response. Additionally, we found that most errors made by the pathway skew towards underestimating the concentration of encountered antigen. Finally, we found that disabling IKKα deactivation could increase signaling fidelity against targets with antigen-negative cells. Our information-theoretic analysis of signal transduction can provide novel perspectives on biological signaling, as well as enable a more informed path to cell engineering.

Haripriya Vaidehi Narayanan, UCLA

Antibody responses require selective proliferative expansion of B-cell clones in response to affinity-dependent signals from antigenic stimuli. Yet, the proliferative expansion of individual B-cells in response to an identical stimulus is itself highly heterogeneous, driven by non-genetic variations in the molecular network directing B-cell fate decisions. The NFkappaB subunit cRel is known to both transduce signals and promote proliferation in stimulated B-cells. Here, we developed a novel fluorescent reporter mTFP1-cRel to enable direct observation of cRel abundances in live intact B-cells and their relationship to proliferative outcomes. We found that cRel abundance is heterogeneously distributed across naïve B-cell populations, characterized by a heavy tail enriched for high expressors. We showed that while cRel abundance is predictive of the intrinsic proliferative kinetics of B-cells, the relationship is unexpectedly non-monotonic. High cRel expressing B-cells are primed to divide faster by reducing the time to first division, but at the cost of diminished proliferative capacity and population size. We computationally model how cRel heterogeneity arises from balancing positive feedback by autoinduction with negative feedback by its signaling inhibitor IkappaB-epsilon. Perturbing this balance in the IkappaB-epsilon knock-out recapitulates the proliferative kinetics of B-cells with high cRel abundance, indicating that negative feedback regulation is essential to keep cRel in check, ensuring appropriate timing and extent of B-cell proliferation. Thus, we present a quantitative study of the dynamical relationship between B-cell proliferation and regulatory factors in the molecular network, revealing non-intuitive and non-monotonic outcomes that may enable proliferative trade-offs, characterizing an effective population-level B-cell response with complex dynamics.

Mark Xiang, UCLA

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

Antibody-mediated immunity relies on the Darwinian selection process within the germinal center reaction to produce a diverse repertoire of high-affinity antibodies. This process involves B-cell somatic hyper-mutation and cell fate decisions, such as survival, proliferation, and differentiation into plasma cells. Although genetic variation in BCR sequences is considered the primary driver of affinity maturation, recent studies have shown that non-genetic factors also play a significant role. In this study, we developed a modeling approach to investigate how cell state heterogeneity and heritability impact affinity maturation outcomes. Our simulations demonstrate that heritability and heterogeneity lead to a more diverse repertoire of high-affinity plasma cells, thus improving the breadth and depth of the antibody response. These findings underscore the importance of considering non-genetic factors in understanding antibody-mediated immunity and may have implications for vaccine development and disease prevention strategies.

Ivy Xiong, USC

Direct androgen receptor regulation of sexually dimorphic gene expression in the mammalian kidney

Direct androgen receptor regulation of sexually dimorphic gene expression in the mammalian kidney. Previous studies have demonstrated sexual dimorphism in the mammalian kidney's physiology, susceptibility to disease and response to injury. In the mouse, proximal tubule (PT) cell types show distinct gene expression between the sexes. Here, we investigated the molecular mechanisms regulating dimorphic gene expression in the murine kidney. Gene- and isoform-level analysis highlighted differential functions related to peroxisomal lipid metabolism and nuclear receptor pathways, suggesting a sex-dependent role for metabolic regulation in renal function. Whole-kidney bulk RNA-seq from neonates to aged adult C57BL/6 mice demonstrated differential gene expression was established with the onset of sexual maturity from 4-8 weeks under gonadal control and stably maintained. Nephron-specific ablation of androgen receptor (AR) and estrogen receptor activity, and hormonal injection studies, showed testosterone-driven Ar activity in proximal tubule cells is the primary regulator of sexually distinct gene expression in the mouse kidney, distinguishing predominant sex-specific control mechanisms between the kidney and liver. Single-nuclear multiomic analysis of chromatin accessibility (snATAC-seq) and gene expression (snRNA-seq) demonstrated decreased chromatin accessibility for male-biased genes and increased accessibility for female-biased genes following Ar removal in male PT segments. Integrating published AR-ChIP-seq and these multiomic data identified putative androgen response elements near sex-biased genes. Examining human kidney datasets identified evidence of conserved dimorphic gene expression in the male and female human kidney. Collectively, these studies highlight organ-specific regulatory programs differentiating gene activity between organs in the mouse and provide a foundation for mechanistic analysis of sex difference in renal physiology and injury and disease modeling.

Wenxuan Zuo, USC

Consistent patterns of microbial diversity, prediction accuracy and associated bacterial organisms between pediatric ulcerative colitis and healthy children using 16S rRNA and metagenomic shotgun sequencing data

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. 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 demonstrate 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, the highest accuracy to date. We show that 16S rRNA data yield similar results as shotgun data in terms of alpha diversity, beta diversity, and prediction accuracy. Our study demonstrates that pediatric UC subjects harbor a dysbiotic and less diverse gut microbial population with distinct differences from healthy children. We also show that 16S rRNA data yield 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.

POSTER SESSION #2 (18:15 – 20:00)

Mika Caldwell, UCI

We often portray cancer as what happens when a renegade cell "escapes" control. To explain such events, we must first understand the strategies and goals of normal tissue growth control. Homeostasis (stability) is just one objective of many: tissues also attain precise sizes, regenerate rapidly and accurately from disturbances, minimize overshoots and oscillations, and compensate for stochasticity in the fates that daughter cells adopt after division. These challenges are usually (and, we argue, necessarily) met by systems of collective integral negative feedback control's where in a cell's probability of self-renewal depends on signals from neighboring cells. Mathematical models explain how these systems achieve superb control, yet, when such models are expanded to account for stochasticity in patterns of cell division, and the fact that tissues occupy space (so that collective signals decay over distance), curious things happen: Rarely, controlled proliferation spontaneously switches to unrestrained growth. We find these events result from implicit positive feedback loops that arise when stochasticity and space interact. Similar "escape" phenomena can also result from explicit positive feedback loops, without a need for spatial effects. These findings show that collective feedback control is structurally fragile: Superb control most of the time comes with a risk of total failure some of the time. Evolution likely adjusts parameters to reduce risk but cannot remove it entirely. These results suggest that some cancer-causing agents (oncogenic mutations, environmental influences, genetic predispositions) may work less by directing individual cells to grow faster (or die less readily) than by altering the parameters of a feedback control system, boosting the probability of rare, random escape. I will discuss how this idea reflects on current puzzles in cancer biology, such as oncogene expression in phenotypically normal cells, tumor dormancy, and the kinetics of relapse after cancer therapy

---

Ermioni Charalampopoulou, USC
Phospho-proteomic analysis of CAR-T cell signaling

Chimeric antigen receptors (CARs) are synthetic biomolecules comprised of an extracellular antigen recognition domain and intracellular signaling domains. When expressed in immune cells, CARs direct their host cells to kill diseased cells expressing the antigen recognized by the CAR. Although signaling pathways downstream of CAR activation control the cytotoxic function of CAR-expressing cells, phospho-proteomic studies of CAR signaling have been limited. Most approaches have used antibodies or soluble ligands, rather than cell-displayed antigens, to activate CAR signaling. Here, we demonstrate an efficient and cost-effective label- free phospho-proteomic approach to analyze CAR signaling in immune cells stimulated with antigen-presenting cancer cells. Following co-culture of CAR-T cells with cancer cells, we first preserved phospho-signaling by cross-linking proteins with formalin. Then, we used magnet-activated cell sorting (MACS) to isolate CAR-T cells from the co-culture. Validation experiments demonstrated that formalin fixation did not alter the phospho-proteome, that MACS achieved >90% CAR-T cell purity, and that activated CARs exhibited upregulation of phospho-tyrosine 142 on the CAR CD3 domain. Next, we compared the phospho-proteome in CAR-T cells stimulated with

either CD19-expressing or non-CD19-expressing SKOV3 ovarian cancer cells. This analysis revealed that CAR signaling activated known pathways including the mitogen- activated protein kinases (MAPKs) ERK1/2. Bioinformatic approaches further showed that CAR activation induced other signaling pathways including the MAPK p38, protein kinase A, and checkpoint kinase 1 (CHK1). Taken together, this work presents an easy and inexpensive method to better understand CAR immunotherapy by label-free phospho-proteomic analysis of CAR signaling in immune cells stimulated by antigen-presenting cancer cells.

Chuankai Cheng, USC

Functional Decomposition of Metabolism allows a system-level quantification of fluxes and protein allocation towards specific metabolic functions.
Quantifying the contribution of individual molecular components to complex cellular processes is a major challenge in systems biology. In this study, we introduce a theoretical framework called Functional Decomposition of Metabolism (FDM), which enables us to accurately quantify the role of metabolic reactions in various cellular processes, such as the synthesis of biomass building blocks and energy production through respiration and fermentation. Using FDM, we examined the energy and biosynthesis budget required for the growth of Escherichia coli in detail. Surprisingly, we found that the ATP generated during the biosynthesis of building blocks when metabolizing glucose almost balances the demand from protein synthesis, the largest energy expenditure in growing cells. This finding challenges the common notion that energy is a key limiting resource for growth since the bulk of energy generated by fermentation and respiration remains unaccounted for. Furthermore, FDM in combination with proteomics enabled us to quantify the contribution of individual enzymes to each metabolic function, leading to the development of a first-principles formulation of a coarse-grained model of global protein allocation based on the structure of the metabolic network.

Jackson Chin, UCLA

Liver ischemia-reperfusion injuries (LIRI) arise during liver transplantation and contribute to the high rate of chronic liver transplant rejection. LIRI occur across the liver transplant process, from pre-operation ischemic injuries accrued during the liver transport to post-operation reperfusion injuries accrued as the recipient's immune system responds to the donor's liver. While research has identified many types and mechanisms of LIRI, the relationship between LIRI and transplant outcome (donor liver acceptance or rejection) is incompletely understood, precluding development of preventative therapies. To better understand the mechanisms of LIRI and their relationship to transplant outcome, we applied tensor partial least squares (tPLS) - s supervised, tensor-based decomposition method that enables outcome-informed, high-dimensional data reduction - to regress transplant outcomes against longitudinal measurements derived from both transplant donor and recipient sources. Longitudinal measurements included peripheral blood cytokines and liver function test (LFT) measurements alongside cytokine measurements derived from the recipient portal vein shortly before and after liver transplant surgery. We find that tPLS successfully integrated pre- and post-transplant measurements

collected from the donor and recipient and captures mechanisms relevant to liver transplant rejection with just two components. These two components were highly predictive of transplant outcomes, predicting transplantation rejection with an accuracy of 69% and outperforming models trained solely on peripheral blood cytokine measurements (55% accuracy), LFT measurements (62% accuracy), or biopsy pathology score (51% accuracy). Biological interpretation of these components highlights that both Th1 polarization and Treg suppression are associated with liver transplant rejection, and that both processes combined increase the risk of liver transplant rejection. Ultimately, these efforts demonstrate the power of outcome-informed tensor decomposition techniques in capturing immunological signatures and highlighting potential therapeutic targets for improving liver transplant outcomes.

Yuxuan Du, USC

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

---

Recovering high-quality metagenome-assembled genomes (MAGs) from complex microbial ecosystems remains challenging. 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 clustering algorithm and includes the spurious contact detection into binning pipelines for the first time. HiCBin is validated on one synthetic and two real metagenomic samples and is shown to outperform the existing Hi-C-based binning methods.

Yilin Gao, USC

Heterogeneity in different genomic studies compromises the performance of machine learning models in cross-study phenotype predictions. Overcoming heterogeneity when incorporating different studies in terms of phenotype prediction is a challenging and critical step for developing machine learning algorithms with reproducible prediction performance on independent datasets. We investigated the best approaches to integrate different studies of the same type of omics data under a variety of different heterogeneities. We developed a comprehensive workflow to simulate a variety of different types of heterogeneity and evaluate the performances of different integration methods together with batch normalization by using ComBat. We also demonstrated the results through realistic applications on six colorectal cancer (CRC) metagenomic studies and six tuberculosis (TB) gene expression studies, respectively. We showed that heterogeneity in different genomic studies can markedly negatively impact the machine learning classifier's reproducibility. ComBat normalization improved the prediction performance of machine learning classifier when heterogeneous populations presented, and could successfully remove batch effects within the same population. We also showed that the machine learning classifier's prediction accuracy can be markedly decreased as the underlying disease model became more different in training and test populations. Comparing different merging and integration

methods, we found that merging and integration methods can outperform each other in different scenarios. In the realistic applications, we observed that the prediction accuracy improved when applying ComBat normalization with merging or integration methods in both CRC and TB studies. We illustrated that batch normalization is essential for mitigating both population differences of different studies and batch effects. We also showed that both merging strategy and integration methods can achieve good performances when combined with batch normalization. In addition, we explored the potential of boosting phenotype prediction performance by rank aggregation methods and showed that rank aggregation methods had similar performance as other ensemble learning approaches.

Patrick Gelbach, USC

Ensemble-based Genome-scale Modeling Predicts Metabolic Differences between Macrophage Subtypes in Colorectal Cancer

---

Colorectal cancer (CRC) shows high incidence and mortality, in part because of the role of the tumor microenvironment, which is often viewed as an active promoter of disease progression. Macrophages are among the most abundant cells in the tumor microenvironment. These cells are generally categorized into the classically activated (M1) phenotype with inflammatory and anti-cancer properties, or the alternatively activated (M2) subtype known as tumor-associated macrophages, which promote tumor proliferation and survival. It is thought that the M1/M2 subclassification scheme is strongly influenced by cellular metabolism; however, the metabolic divergence between the subtypes has not been fully elucidated. In particular, the metabolic states of M1 and M2 macrophages induced by cancer cells remains poorly understood. Computational modeling is needed to provide a systematic understanding of macrophage metabolism, due to the intricate and complex nature of their intracellular metabolic networks. In this work, we generated a suite of computational models that characterize the M1- and M2-specific metabolic states. The genome-scale metabolic models represent all the cell's known metabolic genes, reactions, and metabolites, and are simulated to estimate flow of material (flux) through the network, and thus capture the functional state of the cell. We apply the models to assess differences in the cells' predicted metabolic network and capability, and show key and consistent differences between the M1 and M2 macrophage metabolic signatures. Furthermore, we leverage the models to identify metabolic perturbations that cause the metabolic state of M2 macrophages to more closely resemble M1 cells, thereby identifying novel metabolic targets that may be of clinical relevance for cancer patients. Overall, this work increases understanding of the role of macrophages in CRC and elucidate strategies to promote the metabolic state of anti-tumor macrophages.

Timothy Hamilton, UCLA

Fractal-like Density Distributions in Single-Cell Data

The proliferation of single cell sequencing and other high-throughput methods to interrogate cell state has promised a revolution through multiple fields of biology. Methods to analyze this data are influenced by the "Waddington's landscape" paradigm of the arrangement of cell types in gene

expression space. 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. While this is a commonly held belief in the field, up to now there has not been any work to clearly demonstrate whether single-cell data can be interpreted in a manner consistent with “Waddington’s Landscape.” Here we demonstrate a novel method, called “epsilon networks” that uses local density to determine whether the underlying the structure of the data results from an attractor structure consistent with Waddington’s Landscape. If the data is consistent with Waddington’s Landscape, most of the cells would be found close to an underlying attractor, producing relatively dense regions, with fewer cells the further away from the center. Instead, our method has revealed that most cells are found in low density regions of the space with a few cells being found in areas that are 3-4 orders of magnitude denser than the lower density regions. This finding is inconsistent with the current interpretations and applications of Waddington’s Landscape. This “fractal-like” density distribution is not due to any form of technical noise or biological confounding and its presence may explain the heterogeneity shown in single cell data that makes it difficult to analyze and separate into distinct groups. Ultimately, our findings have implications for how development must be conceptualized and reveal unseen diversity in how single cells behave that provides new opportunities for exploration and discovery.

Jiawei Huang, USC

Understanding the cellular composition of a disease-related tissue is important in disease diagnosis, prognosis, and downstream treatment. The recent advances in single-cell RNA sequencing (scRNA-seq) technique and extensive high-quality datasets have allowed the measurement of gene expression profiles for individual cells and further made it possible to deconvolve cellular composition in bulk tissues. Here, we present DeepDecon, a deep neural network model leveraging single-cell gene expression information to accurately predict the fraction of cancer cells in bulk tissues. DeepDecon is trained based on single-cell RNA sequencing data and is robust to experimental bi- ases and noises. It will automatically select optimal models to recursively estimate malignant cell fractions and improve prediction accuracy. When applied to simulated and real data, it outperforms existing decomposition methods including Scaden, Bisque, MEAD, RNA-Sieve, CIBERSORTx, Mu- SiC, and NNLS in both accuracy and robustness. We further show that the DeepDecon is robust to the number of single cells that make up a bulk sample. DeepDecon is also able to estimate the confidence interval of the fraction of malignant cells in cancer tissue with coverage closes to the prespecified confidence than other existing methods.

Serena Hughes, UCLA
A Statistical Marker Gene Clustering Tool

The current standard practice for clustering analysis on single cell RNA sequencing (scRNA-Seq) data often involves manual assignment of cell types to clusters based on marker gene expression. During manual annotation, researchers may discover that their biological expectations do not align with the

clustering output and often address this by altering the algorithm parameters and rerunning it until a suitable clustering is achieved. While clusters that are consistent with biological expectations will likely lead to improved downstream analysis, this ad-hoc process introduces bias and issues with reproducibility. To help mitigate this, we are developing a Statistical Marker Gene Clustering (SMGC) tool. This tool uniformly samples clustering parameters and quantitatively evaluates the output based on expected marker gene associations. It takes as input the data to be clustered and a mapping of expected cell types to their associated marker genes. In our exploration of data where the cell types are known, we observed that the pattern of marker gene expression lends itself to translation into barcodes that represent each cell type. The SMGC tool uses this concept of barcodes to compare clusters and individual cells to their assigned cell types. It repeats clustering on sampled parameters and applies a scoring function to output the clustering that is most aligned with the researcher' expected cell types and marker genes. The SMGC tool contributes to the emerging set of algorithms for the clustering process that are shifting the field towards more rigorous and consistent analysis of scRNA-Seq data.

Matt Karikomi, UC Irvine

PCquery.jl: Automated Mining, Integration, and Annotation of Biological Signaling Pathways

---

Experimental knowledge of cell-signaling informs the teleological design of computational models. However, the curation of annotated pathways is both incomplete and highly-centralized, leading to inefficient use of existing knowledge and limited scalability in the face of increasingly rapid accumulation of new data. We present PCQuery: an automated pipeline that delivers decentralized, on-demand integration of7 biological signaling pathways. PCquery facilitates regularized and high-dimensional cell signaling8 models by exposing long-range signal-transduction cascades whose provenance is fully-annotated and9 verifiable. The modular package includes a collection of SPARQL clients to pull current annotations10 from UniProt, OntoDB, and NextProt. This structural, evolutionary, and clinical knowledge can be11 graphically displayed by a comprehensive set of pathway visualization tools.

Marcus Kelly, UCSD

A major goal of cancer research is to understand how mutations distributed across diverse genes affect common cellular systems, including multi-protein complexes and assemblies. Two challenges - how to comprehensively map such systems and how to identify which are under mutational selection - have hindered this understanding. Here, we create a comprehensive map of cancer protein systems integrating new and published multi-omic interaction data at multiple scales of analysis. We then develop a unified statistical model that pinpoints 395 specific systems under mutational selection across 13 cancer types. This map, called NeST (Nested Systems in Tumors), incorporates canonical processes and new discoveries, including a PIK3CA-actomyosin complex that inhibits PI3K signaling and recurrent mutations in collagen complexes that promote tumor proliferation. These systems can be

used as clinical biomarkers and implicate a total of 548 genes in cancer evolution and progression. This work shows how disparate tumor mutations converge on protein assemblies at different scales.

Stefanie Luecke, UCLA

Combinatorial and temporal coding by MAPK p38 and NFκB in the inflammatory response of macrophages

---

Macrophages are immune sentinel cells that orchestrate immune responses. They sense pathogens via numerous receptors but have only a limited set of signaling pathways at their disposal to produce stimulus-specific responses. Stimulus specificity is hypothesized to be achieved through combinatorial and temporal coding by these pathways. While NFκB is known to encode stimulus information temporally, temporal coding in other signaling pathways and how they function combinatorially remains unclear. Here, we investigate how temporal and combinatorial coding in the p38 MAPK and NFκB pathways interface in response to innate immune stimuli such as pathogen-associated molecular patterns (PAMPs) and the host cytokine TNF. We established a live-cell microscopy system to simultaneously analyze NFκB and p38 reporter activity over time in murine macrophages. Using information theoretic and machine learning approaches, we found that p38 activity dynamics distinguish TNF from PAMPs with high accuracy independently of NFκB, based on lower amplitude and duration of the TNF-induced p38 response. Classification of PAMPS (P3C4, CpG, and LPS) improved slightly when considering combined p38 and NFκB features, indicating that p38 and NFκB combinatorially contribute to stimulus-specificity. Although p38 did not contribute to dose distinction beyond that mediated by NFκB, p38 and NFκB showed stimulus-specific differential dose responses, suggesting that p38 reinforces the presence of high-dose stimulation and potentially broadens the distribution of this information to downstream effectors. The dynamic features of p38 and NFκB activities showed a surprisingly selective correlation pattern. Early dynamic features, determined receptor proximally (e.g. speed), showed some correlation, but later features (e.g. peak amplitude) showed little. Such weak correlations suggested p38 may contribute heterogeneity to macrophage gene expression responses. Analysis of single-cell gene expression revealed a higher variability for NFκB+p38-dependent vs. NFκB-dependent genes. The former were enriched for cytokine genes suggesting that variability in the p38 pathway prevents their aggregate overexpression in tissues.

Diamond Mangrum, USC

Using ODEs to Investigate the Impact of Intrinsic and Extrinsic Properties on Apoptotic Signaling

---

Malignant tumor growth is controlled by features of the tumor's microenvironment (i.e., ligand concentration) and properties of the tumor itself (i.e., receptor concentration, initial protein concentrations, and receptor-ligand binding affinity). Since these tumor cell-extrinsic and -intrinsic properties work together to optimize regulation of malignant tumor progression, many therapies have been ineffective in targeting both types of cell properties to inhibit tumor development and trigger apoptosis. Caspases are a common family of proteins found along the death-inducing signaling

complex that when activated, can lead to apoptosis signaling in individual cells. In this study, we use in-silico ordinary differential equation (ODE) modeling of the caspase mediated death-inducing signaling complex in tumor cells to examine how the combination of these properties can influence tumor cell apoptosis. By varying intrinsic and extrinsic properties of tumor cells, we have investigated conditions necessary for tumor cells to experience a phenotypic shift to an apoptotic state.

Raktim Mitra, USC

Deep dissection of protein-DNA interaction across families

Predicting specificity in Protein-DNA interactions is a challenging yet essential task for understanding gene regulation. Here, we present DeepPBS, a deep learning model capable of predicting binding specificity across protein families based on given Protein-DNA structures. Its modular architecture allows investigation of family-specific recognition patterns. DeepPBS employs a bipartite geometric convolutional architecture, enabling the calculation of biologically meaningful importance scores via a simple edge perturbation-based method. These importance scores, when aggregated at the protein residue level, conform well with alanine scanning mutagenesis experimental data. DeepPBS has potential utility in automated analysis of simulation trajectories, guiding experimental design and advancing our understanding of molecular interactions. Furthermore, the model opens up exciting future possibilities, such as specific DNA-targeted protein design and analysis of mismatched binding specificity. DeepPBS offers a promising foundation for future of machine aided molecular interaction studies.

Anh Nguyen, SDSU

The popularity of vaping has increased over the past decade, particularly among teenagers who adapt vaping to social activity. However, the information on potential health risks is limited. Although the carrier solutions associated with vaping (i.e. vape juice) are assumed to be non-toxic, there is growing evidence that they have a deleterious impact on the lung. Macrophages play a major role in the initiation, maintenance, and resolution of lung inflammation. The purpose of this study is to identify changes in lung macrophage gene expression associated with vaping. To accomplish this we are using spatial transcriptomics, a cutting edge technique that allows analysis of gene expression in the context of total lung architecture. We have analyzed macrophage gene expression in male and female mice exposed to vape juice for nine weeks. Gene expression in these animals was compared to control animals not exposed to vape juice. Both male and female mice were studied, allowing for analysis of sex-specific alterations in the macrophage response. We will use a combination of machine learning and statistical analysis to determine how macrophage activity is altered in lungs exposed to vape juice; potentially identifying biological pathways impacted by vaping.

Noa Popko, UCLA

RNAseq analysis of a mouse model of inflammation-associated aging in hematopoiesis

Aging and chronic inflammation are known to affect the hematopoietic process underlying blood production. However, it is not fully understood how the differentiation process is dysregulated through these perturbations. To study this question, we used a mouse model for chronic inflammation driven by NFκB ("IκBKO") and a rescue model ("IκBKO/TNFKO/IFNARKO/ cRelKO" or combined KO "cKO"). Based on the abundance of the CD45 cell surface marker, cells from the bone marrow were split into stromal (CD45-) and hematopoietic (CD45+) fractions. We then corrected cell counts in each fraction using information about the purity of the sorted samples by posing the correction as a non-negative least squares problem. To discover changes in the IκBKO, we searched for significant differences in gene expression between young and old WT mice, IκBKO mice, and cKO mice, with a specific focus on the stromal fraction which has the potential to influence the differentiation and development of hematopoietic cell lineages through cytokine signaling. We performed differential gene expression analysis on bulk RNA sequencing data between genotypes to identify important genes and pathways likely involved in phenotypic differences. Ultimately, we aim to understand how variable expression in genes in an inflamed and aged state contributes to altered hematopoietic differentiation.

Ali Risheh, CSULA

Novel Molecular Fingerprint using Rule based Graph Convolutional Layer (RGCL)

Molecular fingerprints are essential cheminformatics tools for machine learning with applications in drug discovery. Standard fingerprint software compute fixed-size feature vectors, and employ them as inputs to a deep neural network or other machine learning methods. Fixed-size fingerprint representation of molecules, however, requires extremely large vectors to encode all possible substructures. Limited accuracy and poor interpretation also occur due to the underlying neural network which emphasizes particular and exclusive aspects of the molecular structure. In this study, we develop a novel graph convolutional network (GCN) to predict the binding free energy of protein-ligand complexes. By adding a physics-based layer to the network architecture, the accuracy of the molecular fingerprint has been improved while potential overfitting has been avoided. In addition to standard information about the substuctures, our hyrbid physics-data model encodes atomic bonds features which captures structural features and enables further analysis. It has been shown that machine-optimized fingerprints, compared to fixed-sized fingerprints, can provide more accurate predictions, better performance, and more interpretable results. We show that the proposed GCN fingerprint outperforms the predictive performance of standard fingerprints on binding free energy of host-guest systems and PDBbind database.

Paulina Smaruj, USC

In mammals, interphase chromatin folding displays sequence dependence and evolutionary conservation. Two key factors underlie this genome folding: CTCF and cohesin, which segment the genome into a series of domains. Still, the sequence-dependent grammar of genome folding remains unknown. Despite much progress, experiments perturbing CTCF-binding sites in vivo remain limited in throughput, restricting our ability to probe this grammar. Recent improvements in computational models of genome folding from DNA sequence opened new possibilities to make high-throughput predictions in silico. Here we used an improved version of the Akita model to make rapid in silico predictions and better understand how sequence contributes to genome folding. We first varied the amount of flanking sequence around inserted CTCF binding sites. We found that ~15bp of flanking sequence makes a significant impact on predicted genome folding, with similar patterns across motif orientations. We then varied the spacing between inserted CTCF binding sites. We found substantial differences between orientations for inter-motif distances smaller than 400kb. In particular, two CTCF motifs in the divergent orientation show the highest boundary-creating potential. This demonstrates how a deep-learning model enables rapid and controlled in-silico experiments to obtain interpretable results. Importantly, our computational approach provides unexpected and experimentally testable hypotheses. Together, our analyses start to reveal the sequence-specific grammar governing genome folding at the megabase scale.

Cyrillus Tan, UCLA

Immunoglobulin (Ig)G antibodies coordinate immune effector responses by selectively binding to target antigens and then interacting with various effector cells via the Fc receptors. The Fc domain of IgG can promote or inhibit distinct effector responses across several different immune cell types through variation based on subclass and Fc domain glycosylation. Extensive characterization of these interactions has revealed how the inclusion of certain Fc subclasses or glycans results in distinct immune responses. During an immune response, however, IgG is produced with mixtures of Fc domain properties, so antigen-IgG immune complexes are likely to almost always be comprised of a combination of Fc forms. Whether and how this mixed composition influences immune effector responses has not been examined. Here, we measured Fc receptor binding to immune complexes of mixed Fc domain composition. We found that the binding properties of the mixed-composition immune complexes fell along a continuum between those of the corresponding pure cases. Binding quantitatively matched a mechanistic binding model, except for several low-affinity interactions mostly involving IgG2. We found that the affinities of these interactions are different than previously reported and that the binding model could be used to provide refined estimates of these affinities. Finally, we demonstrated that the binding model can predict effector-cell elicited platelet depletion in humanized mice, with the model inferring the relevant effector cell populations. Contrary to the previous view in which IgG2 poorly engages with effector populations, we observe appreciable binding through avidity, but insufficient amounts to observing immune effector responses. Overall, this work demonstrates a quantitative framework for reasoning about effector response regulation arising from IgG of mixed Fc composition.

Isaac Tate, UCR

To further understand how the diffusion of Arabidopsis transcription factors actually occurs, this model seeks to represent the tissue and transcription factor as the cell is growing. This is done by combining two models: one handling the chemical signaling of the transcription factors, and one handling the tissue's configuration by calculating cellular growth rates and intracellular forces.

Esra Tiftik Karabay
UCSD & SDSU

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. We have developed a semi-quantitative-computational model to analyze the intra-cellular signaling network and its outcome in the presence of multiple external signals including growth factors, hormones, and extracellular matrix. We use this model to analyze the cell response phase space to external stimuli and identify the key internal elements of the network that drive specific outcomes within this phase space. The model is

built upon Boolean approach to network modeling, where the state of any given node is determined using the state of the connecting nodes and boolean logic. This allows us to analyze the network behavior without the need to estimate all the various interaction rates between different cellular components. However, such an approach is limited in its ability to predict network dynamics and temporal evolution of the cell state. So, we introduce modularity in the model and incorporate dynamical aspects, mass-action kinetics, and chemo-mechanical effects on only certain transition rates within specific modules as required, creating a Boolean-Hybrid-Modular (BoHyM) signal transduction model. We present this model as a comprehensive, cell-type agnostic, user-modifiable tool to investigate how extra-and intra-cellular signaling can regular cellular cytoskeletal components and consequently cellular mechanical properties driving cell-substrate interactions, force generation and migration. Using this tool, we show how slight changes in signaling network architectures due to phenotypic changes can alter cellular response to stress hormone signaling in an environment dependent manner. The tool also allows isolating effector proteins driving specific cellular mechanical responses. Ultimately, we show the utility of the tool in analyzing transient chemo-mechanical dynamics of cells in response time-varying chemical stimuli.

Lisa Uechi, City of Hope

Acute myeloid leukemia (AML) is the most common type of leukemia in adults and can be found in 20% of pediatric leukemias. It has become clear that disease heterogeneity continues to present a major challenge in AML prognosis and treatment. Here we suggest a new interpretation of gene expression data in AML by using quantitative approaches from physics and chemistry. We propose to integrate two approaches, surprisal analysis and state-transition theory, for analyzing transcriptome profiles. Using publicly available AML RNA seq datasets, we calculate free energy deviations from the steady state for each normal and AML tissue and map those values into a space-state, representing the transition points of the disease. We find that the diseased samples are at a high free energy level and mapping free energy to critical points shows highest free energy levels in the unstable transition state. This suggests that the free energy characterization and mapping to state-space may not only provide a useful diagnostic feature for the heterogeneous AML tissues, but also indicate how AML states may be rationally manipulated therapeutically. We propose that once these processes are accurately resolved in each AML patient, the right set of individualized inhibitors can be rationally designed.

Daniel Velez-Ramirez, UCLA
cAMP-dependent spatial phosphoproteomic analysis of the eukaryotic flagellum

---

Trypanosoma brucei is a eukaryotic parasite and causative agent of sleeping sickness, a vector borne disease in Sub-Saharan Africa. Despite the recent efforts in eliminating this fatal neglected tropical disease, and the remarkable advances in developing safe and effective treatments, the eradication of it is still out of sight. A critical aspect of T. brucei biology is its ability to survive in two completely different environments: 1) the tsetse fly, which is T. brucei vector, and 2) the human body. Therefore, T. brucei requires constantly sensing its environment and responding to it, and there is cumulative

evidence that shows its flagellum is a cAMP signaling platform. However, the molecular mechanisms are almost completely unknown and such are potential drug targets. In our working model, adenylate cyclases localized at the tip of the flagellum produce cAMP, and a specific phosphodiesterase (PDEB1) along the flagellum, breaks down cAMP molecules diffusing away from the flagellum tip. The proper functioning of this cAMP microdomain, allows T. brucei cells to establish infection within the tsetse fly, making possible T. brucei transmission to humans. Considering that one of the main mechanisms of action of cAMP signaling is protein phosphorylation, we decided to obtain the flagellar cAMP-dependent phosphoproteome of T. brucei. Among the 153 proteins identified that undergo cAMP-dependent phosphorylation, we have identified two axonemal dynein heavy chains, and two subunits of the docking complex. We hypothesize that the phosphorylation of these axonemal proteins could have an effect on axonemal assembly and/or flagellum beating. Altogether, our results support that T. brucei flagellum is cAMP signaling platform, and it is subject of cAMP-dependent phosphorylation with scientifically interesting and medically relevant implications.

Sheng Wang, Caltech
Periodic spatial patterning with a single morphogen

Multicellular development employs periodic spatial patterning to generate repetitive structures such as digits, vertebrae, and teeth. Turing patterning has long provided a key paradigm for understanding such systems. The simplest Turing systems are believed to require at least two signals, or morphogens, that diffuse and react to spontaneously generate periodic patterns. Here, using mathematical modeling, we show that a minimal circuit comprising an intracellular positive feedback loop and a single diffusible morphogen is sufficient to generate stable, long-range spatially periodic cellular patterns. The model considers cells as discrete entities as a key feature, and incorporates transient boundary conditions. Linear stability analysis reveals that this single-morphogen patterning circuit can support a broad range of spatial wavelengths, including fine-grain patterns similar to those generated by classic lateral inhibition systems. Further, signals emanating from a boundary can initiate and stabilize propagating modes with a well-defined spatial wavelength. Once formed, patterns are self-sustaining and robust to noise. Finally, while noise can disrupt patterning in pre-patterned regions, its disruptive effect can be overcome by a bistable intracellular circuit loop, or by considering patterning in the context of growing tissue. Together, these results show that a single morphogen can be sufficient for robust spatial pattern formation, and should provide a foundation for engineering pattern formation in the emerging field of synthetic developmental biology.

Guanao Yan, UCLA
scReadSim: a single-cell RNA-seq and ATAC-seq read simulator

Benchmarking single-cell RNA-seq (scRNA-seq) and single-cell ATAC-seq (scATAC-seq) computational tools demands simulators to generate realistic sequencing reads. However, none of the few read simulators aim to mimic real data. To fill this gap, we introduce scReadSim, a single-cell RNA-seq and ATAC-seq read simulator that allows user-specified ground truths and generates synthetic sequencing

reads (in FASTQ and BAM formats) by mimicking real data. At both read-sequence and read-count levels, scReadSim mimics real scRNA-seq and scATAC-seq data. Moreover, scReadSim provides ground truths, including unique molecular identifier counts for scRNA-seq and open chromatin regions for scATAC-seq. In particular, scReadSim allows users to design cell-type-specific ground-truth open chromatin regions for scATAC-seq data generation. In benchmark applications of scReadSim, we show that cellranger is a preferred scRNA-seq UMI deduplication tool, and HMMRATAC and MACS3 achieve top performance in scATAC-seq peak calling.

Vincent Zaballa, UCI

Many models of biological systems can be simulated but we may not know the structure of the exact function that describes this forward process. This situation can be described as having an implicit likelihood function, one where the probability of the output depends on the inputs but in an unknown manner. Simulation-based inference (SBI) can overcome this task and provide an approximate likelihood function of observed data given conditional inputs to the function. Additionally, biological systems are expensive to perturb and study. We demonstrate a novel approach to both inferring the likelihood function of an implicit likelihood and the design of experiments for such a function in a unified optimization process. We show connections to mutual information estimation and an experimental validation on an implicit biological signaling pathway.