1.15pm Session I – chaired by Alexander Hoffmann

Integrating Signaling and Polygenic Risk Scores to Predict Immune Dysregulation in Common Variable Immunodeficiency

HUMZA A. KHAN[1,2], Timothy J. Thauland[3], and Manish J. Butte[2,3]
[1] BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
[2] Department of Microbiology, Immunology, and Molecular Genetics, UCLA
[3] Division of Immunology, Allergy, and Rheumatology, Department of Pediatrics, UCLA

Common Variable Immunodeficiency (CVID) is a collection of monogenic disorders that are characterized by defective antibody production. Clinical presentation of these patients varies widely–from susceptibility to infection to autoimmunity to cancer. Patients with immune dysregulation require much more aggressive treatment, but we have no predictors of patient disease course. We employed mass cytometry (CyTOF), exome sequencing, and extensive clinical phenotyping to identify patients with immune dysregulation. By CyTOF, we found a number of aberrant signaling pathways in immune dysregulated patients, including a newly described defective T cell STAT3 signaling module. Other defects in the STAT3 and AKT signaling axes were found as well. We also utilized polygenic risk scores (PRS) to segregate CVID patient by clinical phenotype. A previously published PRS of absolute lymphocyte counts successfully distinguished autoimmune and non-autoimmune patients. Using this approach, we aim to combine genomics, phenomics, and phospho-CyTOF to classify patients and inform therapeutic interventions.

Accounting for genetic relationship in rare variant statistical testing

THOMAS FELT[1], Timothy Chang[2], Daniel Geschwind[2]
[1] BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
[2] Department of Neurology, David Geffen School of Medicine, UCLA

Progressive Supranuclear Palsy (PSP) is a rare neurodegenerative condition that has parkinsonian features and dementia. Association studies have identified common variants contributing to PSP risk, but common variation only accounts for part of disease heritability, indicating that rare variation likely contributes to the unexplained genetic heritability. However, traditional rare variant analyses have decreased power compared to common variant analyses due to the overall lower frequency of rare variants. Recent tests incorporate genetic relationship among samples to better estimate variance of rare variants. Here, we used whole genome sequencing data from 1668 PSP and 3272 control subjects, jointly-called, where we accounted for genetic relationships by removing 3rd degree or closer relatives. We then compared rare variant gene burden results in PSP accounting for (SKATO+GRM) and not accounting for (SKATO) the genetic relationship among samples. Although the correlations between SKATO+GRM and SKATO p-values were highly significant (Pearson cor=0.43, $p<2.2E-16$), they were lower than might be expected from the same cohort. This suggests that accounting for genetic relationship can have an impact on rare variant burden tests due to latent relatedness.

Cell type specific changes in transcriptional networks underlying ASD

SHRUTI GUPTA[1], Brie Wamsley[2], Daniel H Geschwind[2,3]
[1] BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
[2] Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, UCLA
[3] Program in Neurobehavioral Genetics and Center for Autism Research and Treatment Semel Institute and Department of Human Genetics, David Geffen School of Medicine, UCLA

Advances in genomic technologies have played a major role in understanding the underlying cause of autism spectrum disorder (ASD). The cerebral cortex is built from a highly heterogenous group of cell types whose cooperative function underlies high-order cognitive functions commonly disrupted in ASD. Initial scRNAseq analysis has confirmed cell types that are most disrupted, but we lack an understanding of altered transcriptional networks across ASD cortical cells. We use SCENIC (single-cell regulatory network inference and clustering) on a large single-cell dataset (200,000 cells) composed of pre-frontal cortex from 10 unaffected individuals and 10 individuals diagnosed with ASD. SCENIC builds networks based on gene co-expression with transcription factors and their cis-regulatory elements found within each cell type. This analysis holds promise to extend our understanding of the molecular changes underlying ASD by unbiased linking of distinct transcriptional alterations to their genetic basis within specific cell types of the human cortex.

Genome-scale CRISPR-Cas9 Knockout Screen Identifies Genes Driving Chromosomal Instability in Cancer

KYLE M. SHEU[1,2], Kai Song[2], Nikolas G. Balanis[2], Daniel S. Yong[2], Xiangyu Ge[2], Thomas G. Graeber[2]
[1] BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
[2] Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, UCLA

Chromosomal instability (CIN) is a hallmark of cancer and represents a state of high mutational frequency within the cell genome. This stochastic variance provides a Darwinian landscape through which cancer cell populations adopt characteristics favorable to drug-resistance, immune evasion, and metastasis; clinically, high CIN correlates with poor patient prognosis. However, the genes that drive CIN remain imprecisely determined. Here, we employ a forward genetic screen using genome-scale CRISPR-Cas9 knockout lentiviral barcoded libraries to identify these genetic determinants of CIN, with intent to identify potentially novel targets for molecular therapeutics. To assess gene contribution to CIN, we analyze sequencing data from CIN-high and CIN-low cell populations with STARS and correct false-positives from copy-number-amplified genomic regions with CERES. We then compare data from our screen to publicly available gene dependency screens to assess genetic perturbations and resultant CIN—as identified in our screen—and their potential mechanistic role in cancer cell fate.

Mapping a Melanoma Drug Resistance Program by Fitting a Data-Driven Dynamical Model

MADISON WAHLSTEN[1], ZHAN ZHANG[2], Farnaz Mohammadi[3], Aaron Meyer[3]
[1] BIG Summer Program, Institute of Quantitative and Computational Biosciences,
[2] Cross-disciplinary Scholars in Science and Technology, UCLA
[3] Department of Bioengineering, UCLA

Though effective therapies exist for melanoma, resistance to these drugs inevitably develops. Previous studies have shown that resistance arises from rare cancer cells that are reprogrammed from a pre-resistant state. Several genes, including EGFR, NGFR, and AXL, are disproportionately expressed in pre-resistant cells and have been comprehensively profiled through knockouts models and gene expression measurement. However, the broader regulatory events by which a cell enters this rare state are unclear. A unified model for how these components interact would help uncover drivers of this process. We built an ordinary differential equation model of the concentrations of mRNA corresponding to pre-resistant genes. We used this as a data-driven framework to identify gene-gene interactions by allowing all possible interactions, then comparing to gene expression measurements from each knockout using optimization implemented in Julia. The interaction parameters inferred by the model can be used to identify key regulators driving melanoma drug resistance development.

DISCUSSION

1.45pm Session II – chaired by Eric Deeds

Predicting LPA-induced gene expression dynamics in M1 and M2 macrophages with a multiple regression model of histone modifications

KEVIN JIANG[1], Katherine Sheu[2], Alexander Hoffmann[2]
[1] BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
[2] Department of Microbiology Immunology, and Molecular Genetics, UCLA

Macrophages encountering immune stimuli upregulate specific sets of genes, which may differ when macrophages are reprogrammed by polarizing cytokines into M1 and M2 states. Which genes are induced may be determined by histone modifications. However, the relationship between histone marks and stimulus-induced gene expression remains unclear. Here we aimed to study whether differences in histone modifications among macrophage states are predictive of peak fold-induction of LPA-induced genes. We processed ChIP-Seq data of four histone marks at baseline and RNA-Seq data for LPA-stimulated naive, M1, and M2 macrophages. We used a multiple regression model to correlate histone marks to either gene expression levels or peak fold-induction. We found that while histone modifications are predictive of baseline gene expression, they are poor predictors of peak gene induction upon stimulation. Our results emphasize the important role of signal-dependent transcription factors in the stimulus-response and suggest that their interaction with chromatin requires further study.

The transcription factor GAF induces gene expression in a cell-type-specific manner

CHIDERA OKEKE[1], MINH NGUYEN[1], Quen Cheng[2], Alexander Hoffmann[3]
[1] BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
[2] Department of Medicine, Infectious Diseases, UCLA
[3] Department of Microbiology, Immunology, and Molecular Genetics, UCLA

Interferon (IFN) cytokines are key signaling molecules of the immune system. IFNs activate two transcription factors (TFs), ISGF3 and GAF, in a coordinated manner to regulate interferon stimulated genes (ISGs). Previous work in epithelial cells demonstrated that GAF collaborates with ISGF3 to enhance ISG expression, but GAF binding alone is insufficient to induce expression of nearby genes. As GAF has been more extensively characterized in macrophages, we asked whether a similar phenomenon exists in this cell type. We examined ChIP-seq and RNA-seq data of IFN-stimulated macrophages and identified 2173 binding events, of which 1281 and 552 were classified as ISGF3 and GAF binding, respectively. ISGF3 and GAF binding events correlated with the induction of 19% and 22% of nearby genes, respectively. We conclude that GAF behaves differently in the two cell types. This specificity may be driven by collaborating TFs that are present in macrophages but not epithelial cells.

Prediction of Subnuclear Compartmentalization of Genomes Using 3D Structural Modeling and Machine Learning

SERAFINA TURNER[1], Asli Yildirim[2], Jitin Singla[2,3], Frank Alber[2]
[1] B.I.G Summer Program, Institute for Computational and Quantitative Biosciences, UCLA
[2] Department of Microbiology, Immunology, and Molecular Genetics, UCLA
[3] Department of Quantitative and Computational Biology, USC

The 3D structure of the genome plays an important role in various functions, including gene expression and replication. The genome is organized at different structural scales, with one layer of organization being its subnuclear compartmentalization: the composition of transcriptionally active A and inactive B subcompartments varies over cell types and governs chromatin co-segregation into functional microenvironments. Subcompartment detection requires high sequencing depths not available for all cell types. Here we combine 3D structural modeling and machine learning to identify subcompartments in different cell types. We use structural features extracted from 3D genome models generated from HiC data and compare the performance of unsupervised and supervised machine learning algorithms, such as k-means clustering, logistic regression and neural networks. We observed that logistic regression and neural networks achieved ~80% prediction accuracy. We aim to create a robust method to accurately predict subcompartments across cell types by including additional graph based structural features.

An accessible method for dynamical behavior analysis of large gene regulatory networks

SANDY KIM[1], Shamus Cooley[2,4], and Eric J. Deeds[3,4]
[1] BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
[2] Bioinformatics Interdepartmental Program, UCLA
[3] Department of Integrative Biology and Physiology, UCLA
[4] Institute for Quantitative and Computational Biosciences, UCLA

The temporal and spatial patterns of gene expression are fundamentally integral in all organisms. These patterns are governed by a set of genes and their interactions are known as a gene regulatory network (GRN) that underlie and influence many critical processes in the cell such as development, differentiation, and responses to environmental changes. Thus, dysregulation of GRNs is extremely detrimental and often leads cells to disease states or even senescence. The ubiquity and necessary role of GRNs across all organisms make them of great interest to study. However, due to their often large and highly complex nature, little is currently known about the dynamic properties of many GRNs. Although advances in high-throughput sequencing methods and their applications to temporal studies have produced time course data for gene expression, there remain many challenges in analyzing such high-dimensional data. Using principal component analysis, a dimensionality reduction technique, along with other mathematical methods, we developed an easily-accessible method to analyze the dynamics of high-dimensional time-course gene-expression data to infer the behavior and robustness of GRNs. We demonstrate its ability to uncover the dynamics underlying a wide variety of gene expression data by applying our tool to the analysis of simulations of very large GRNs.

Comparing the Levels of Distortion Introduced by Dimensionality Reduction Techniques

ANNA SPIRO[1], Shamus Cooley[2], Serena Hughes[2], Eric J Deeds[3]
[1] BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
[2] Bioinformatics Interdepartmental PhD Program, UCLA
[3] Department of Integrative Biology and Physiology, UCLA

Existing techniques for reducing the dimensionality of high-dimensional datasets include linear approaches like Principal Component Analysis (PCA) and nonlinear approaches like t-distributed Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP). These techniques all introduce distortion, which we quantify using the Average Jaccard Distance (AJD), a measure of how much a lower-dimension embedding retains information about the local structure of the data. After developing the Deep Embedder (DE), a deep neural network approach to nonlinear dimensionality reduction, we compared the AJDs of the embeddings created by this technique to those created by PCA, t-SNE, and UMAP for nine machine learning datasets. We found that for sufficiently high embedding dimensions, the DE generally produces embeddings that are less distorted than t-SNE or UMAP embeddings and more distorted than PCA embeddings. We predict that adjustments to the DE algorithm will allow it to better approximate nonlinear manifolds than existing techniques.

DISCUSSION

2.15pm Session III – chaired by Hilary Coller

A Quantitative Approach to Study the Rates of Autophagosome Formation and Degradation for Fibroblasts in Cellular Quiescence

LAUREL EZE[1], IVAN A. PEREZ[1], Eric Deeds [2,3], Hilary Colller [2,3]
[1] BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
[2] Department of Computational Biology, UCLA
[3] Department of Molecular, Cell and Developmental Biology, UCLA

The quiescent cell state is often overlooked due to its lack of genome replication and cell division. However, the transition to and from quiescence is a highly regulated process and is essential for cellular and pathophysiology. Previous studies suggest quiescence may be maintained by autophagy which recycles macromolecules into metabolites. Additionally, our lab has observed that transition from proliferation to quiescence is accompanied by an increase in autophagosomes. In order to understand the rates of autophagic flux in proliferating and quiescent cells, we used experimental data, statistical and dynamical modeling to solve an equation for proliferating, contact inhibited and serum starved fibroblasts. We determined that autophagosome formation is similar in both proliferative and quiescent cells but the rate of degradation decreases for cells in the quiescent state. These results suggest that decreased autophagosome degradation may play an important role in the viability or reversibility of quiescence.

Programmatic Identification of Information-Rich Tracers for Metabolic Flux Analysis

SEAN DICKSON[1], Jacob Prohroff[2], Keunseok Park[3], Aliya Lakhani[3], Junyoung Park[3]
[1] BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
[2] CARE Program, Undergraduate Research Center (Sciences), UCLA
[3] Department of Chemical and Biomolecular Engineering, UCLA

Metabolic research is essential for understanding cell functions, identifying therapeutics for metabolic diseases, and engineering metabolism for biotechnological applications. Metabolic Flux Analysis (MFA) is currently the favored method for studying metabolism, relying on stable-isotope tracers and MS/NMR to gather information on metabolic networks. A major pitfall for MFA, however, is the exceedingly large number of tracers to choose from when designing MFA experiments. As a result, tracer selection to this point has been largely heuristic, making it nearly impossible to identify the most information rich tracers available for a given network. Here, we designed a tool that allows researchers to identify the most information rich tracers available for study of any metabolic network. We used the Elementary Metabolic Unit (EMU) model to simulate isotopologue distributions for all intermediates in a given network. Through these simulations, we determined information content of all possible EMU tracers and identified ones which gave us the most information about the network. Using this tool, metabolic researchers can avoid tedious trial-and-error tactics for finding information-rich tracers, effectively making MFA more efficient and fast-tracking metabolic research at large.

Deep Learning Predicts Early Apoptotic Commitment from Caspase 8 Activity

CLAIBORNE TYDINGS[1], Evan Maltz[2], Alon Oyler-Yaniv[2], Jennifer Oyler-Yaniv[2], Roy Wollman[2, 3]
[1] BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
[2] Institute for Quantitative and Computational Biosciences, UCLA
[3] Departments of Integrative Biology and Physiology and Chemistry and Biochemistry, UCLA

When cells are exposed to TNF, they build a death-inducing signaling complex, which includes caspase 8. Caspase 8 activation is a known commitment step for apoptosis. Is the apoptotic decision instantaneous or does it take into account accumulated information? A deep learning neural network, including an LSTM and an attention layer, was trained on caspase 8 activity collected from live-cell FRET reporter imaging, and used to predict cell death. Our neural network shows that early, pre-apoptotic caspase 8 information impacts the final decision of apoptotic commitment. As the point of apoptotic commitment is approached, the cumulative information in the caspase 8 activity better predicts apoptotic commitment. This work indicates that cellular decision making in the case of apoptosis is not an instantaneous decision, but depends upon accumulated cellular information.

Computational analysis and comparison of two recombinase-based oscillator designs with molecular sequestration

JUDITH LANDAU[1], Christian Cuba Samaniego[2], Elisa Franco[2,3]
[1] B.I.G. Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
[2] Department of Mechanical and Aerospace Engineering, UCLA
[3] Department of Bioengineering, UCLA

Synthetic gene oscillators are canonical examples of dynamic biological circuits which allow autonomous cycling of cellular processes. Since serine recombinases can rearrange and then reverse-rearrange DNA when bound to their RDFs (recombination directionality factors), scientists have proposed their use in gene oscillators to invert promoters, an unconventionally dynamic application of recombinases. However, a recombinase-based oscillator has yet to be built so the optimal design is unknown. We used MATLAB to compare the dynamic models of two recombinase-based oscillator designs, each with an inverting promoter. Design 1 has one constitutive and one inverting promoter while design 2 is a novel oscillator with a single promoter. Our results unexpectedly showed that the more novel design 2 will likely perform better. This warrants thorough experimental testing of both designs. We are characterizing the conditions for oscillations and assessing tunability of period and amplitude to direct the selection of circuit components in experiments.

Allometric Scaling of Antibiotic Efficacy

SHAILI MATHUR[1], Portia M. Mira[2], Pamela J. Yeh[2,4], Christopher P. Kempes[4], Van M. Savage[2,3, 4]
[1] BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
[2] Department of Ecology and Evolutionary Biology, UCLA
[3] Department of Biomathematics, David Geffen School of Medicine, UCLA
[4] Santa Fe Institute, Santa Fe, NM 87501, USA

How antibiotic efficacy varies with bacterial species is of basic and applied importance, including understanding of microbial dynamics in clinical and ecological contexts with possible consequences for the community structure of the microbiome. The scaling of cellular components in bacteria and their impact on metabolic, cellular, and evolutionary processes will help illuminate this question and possibly reveal an important role for cell size across bacterial species. Cellular components that antibiotics target—DNA, proteins, mRNA, tRNA, cellular envelope, and ribosomes—all scale non-linearly with cell volume. We model optimal strategies for cells to respond to antibiotics based on energetic constraints. We develop theory that shows how antibiotic efficacy may depend on cell size based on the specific cellular components targeted by the antibiotics and the nonlinearities between those components and cell size. Here, we present a general framework and detailed model for ribosome targeting antibiotics.

DISCUSSION