Inaugural Integrative Biology Symposium

Friday, September 13, 2019
Salk Institute for Biological Studies
La Jolla, California

***Important Note**: Abstracts published in this meeting book should not be cited in bibliographies. Material contained herein should be treated as personal communication and cited as such only with the consent of the author(s)*

Schedule

Thursday, September 12

Session 1

4:00pm	Keynote Presentation, Jay Shendure, University of Washington
5:00pm	Reception

Friday, September 13

8:30am	Registration Opens
9:00am	Breakfast & Coffee

Session 2

9:30am	Aleksandra Walczak, Ecole Normale Supérieure, France
10:00am	Rosalie Waller, University of Utah
10:15am	Tatyana Sharpee, Salk Institute for Biological Studies
10:35am	Coffee Break

Session 3

11:00am	Hannah Carter, University of California, San Diego
11:30am	Nehemiah Zewde, University of California, Riverside
11:45am	Ed Stites, Salk Institute for Biological Studies
12:05pm	Lunch
12:30pm	Poster Session

Session 4

1:30pm	Leonid Kruglyak, University of California, Los Angeles
2:00pm	Magnus Nordborg, Gregor Mendel Institute, Austria
2:30pm	Josep Vilarrasa Blasi, Stanford University
2:45pm	Wolfgang Busch, Salk Institute for Biological Studies
3:05pm	Coffee Break

Session 5

3:30pm	Jennifer Listgarten, University of California, Berkeley
4:00pm	Fangwei Si, University of California, San Diego
4:15pm	Saket Navlakha, Salk Institute for Biological Studies
4:35pm	Coffee Break

Session 6

5:00pm	Ben Lehner, Center for Genomic Regulation, Barcelona
5:30pm	Amelie Baud, University of California, San Diego
5:45pm	Graham McVicker, Salk Institute for Biological Studies
6:05pm	Banquet Dinner

Invited Speaker Bios

Inaugural Integrative Biology Symposium
Friday, September 13, 2019

Salk Institute for Biological Studies
La Jolla, California

Jay Shendure, M.D., Ph.D.

Director, Brotman-Baty Institute for Precision Medicine
Director, Allen Discovery Center for Lineaging Tracing
Howard Hughes Medical Institute Investigator
Department of Genome Sciences, University of Washington

Jay Shendure, M.D., Ph.D., is a Professor in the Department of Genome Sciences at the University of Washington and Director of the Brotman-Baty Institute for Precision Medicine as well as the Allen Discovery Center for Lineage Tracing. Since 2015 Dr. Shendure has been an Investigator of the Howard Hughes Medical Institute. Dr. Shendure completed his M.D. and Ph.D. at Harvard Medical School where he worked in the laboratory of George Church. It was at this time, that he invented one of the first next generation sequencing technologies that revolutionized genomics. Since starting his own laboratory in 2007 he has continuously pushed the boundaries of genomic technologies. Among his notable accomplishments are the first published use of exome sequencing and the application of exome sequencing to identify novel Mendelian disease genes. More recently his laboratory has developed a variety of methods for single-cell RNA-seq and ATAC-seq, high-throughput saturation mutagenesis, cell lineage tracking, and the sequencing of complete fetal genomes from non-invasive samples obtained from the parents. Dr. Shendure is a recipient of numerous awards, including the Curt Stern Award from the American Society of Human Genetics, an NIH Director's Pioneer Award, and a Fulbright Scholarship to study abroad in India.

Aleksandra Walczak, Ph.D.

Laboratoire de Physique Théorique
Ecole Normale Supérieure, Paris, France

Aleksandra Walczak, Ph.D., is a CNRS Directrice de Recherche based at the École Normale Supérieure, Paris. She received her M.Sc. in the Department of Physics, Warsaw University, Poland and her Ph.D. in physics at the University of California, San Diego. After a graduate fellowship at the Kavli Institute for Theoretical Physics (UCSB), she worked on applying information theory to signal processing in small gene regulatory networks at the Princeton Center for Theoretical Science. Her laboratory studies numerous topics in theoretical and computational biology including the effect of selection on population genealogies, the collective behavior of bird flocks, and statistical descriptions of the immune system. She has received numerous awards including the "Grand Prix Jacques Herbrand de l'Académie des Sciences" in 2014 and the bronze medal of CNRS in 2016.

Hannah Carter, Ph.D.

Department of Medicine, University of California, San Diego

Hannah Carter, Ph.D., is an Assistant Professor in the UCSD Department of Medicine, Division of Medical Genetics. She received her M.Eng. in Electrical and Computer Engineering from the University of Louisville and her Ph.D. in Biomedical Engineering from Johns Hopkins University where she studied cancer genomics with Rachel Karchin and Bert Vogelstein. Her laboratory uses computational tools to study the relationship between genotype and cancer-associated phenotypes at multiple scales. Recently her laboratory has performed pioneering work to identify critical germline-somatic interactions by which inherited genetic factors shape the evolutionary trajectory of tumors. Dr. Carter is an Azrieli Global Scholar, a Siebel Scholar, a recipient of a Mark Foundation for Cancer Research Emerging Leader Award, and a recipient of an NIH Director's Early Independence Award.

Leonid Kruglyak, Ph.D.

Department of Human Genetics and Department of Biological Chemistry,
University of California, Los Angeles
Howard Hughes Medical Institute Investigator

Leonid Kruglyak, Ph.D., is Professor of Human Genetics and Biological Chemistry at the David Geffen School of Medicine at UCLA. Dr. Kruglyak received his A.B. degree in physics from Princeton University and his M.S. and Ph.D. degrees, also in physics, from the University of California at Berkeley. After postdoctoral fellowships at the Institute for Advanced Study in Princeton and at Oxford University, he joined the Whitehead Institute as a research scientist. Subsequently, he held a faculty position at the Fred Hutchinson Cancer Research Center where he was also an Investigator of the Howard Hughes Medical Institute and an Affiliate Professor of Genome Sciences at the University of Washington. In 2005, Dr. Kruglyak returned to Princeton University as a Professor in the Department of Ecology and Evolutionary Biology and the Lewis-Sigler Institute for Integrative Genomics, before moving to UCLA in 2013. Dr. Kruglyak's laboratory studies the genetic basis of heritable traits by combining computational analyses with experiments in model organisms (currently, the yeast *Saccharomyces cerevisiae* and the nematode worm *Caenorhabditis elegans*). His laboratory has pioneered many quantitative methods in genetics and performed the first genome-wide analysis of gene expression quantitative trait loci. Dr. Kruglyak is a fellow of the American Association for the Advancement of Science, and the recipient of numerous awards including the Curt Stern Award from the American Society for Human Genetics and the Edward Novitski Prize from the Genetics Society of America.

Magnus Nordborg, Ph.D.

Director, Gregor Mendel Institute, Austrian Academy of Sciences

Magnus Nordborg, Ph.D., is the Scientific Director of the Gregor Mendel Institute in Vienna, Austria. Dr. Nordborg completed his Ph.D. at Stanford University where he worked with Marcus Feldman and he conducted postdoctoral research with Joy Bergelson, Brian Charlesworth, and Deborah Charlesworth at the University of Chicago. Dr. Nordborg was an assistant professor at Lund University in Sweden, and subsequently at the University of Southern California, before moving to his current position at the Gregor Mendel Institute in 2009. Dr. Nordborg is a population geneticist who combines fieldwork and benchwork with statistical and computational analysis of genomic data. Dr. Nordborg was a recipient of a Sloan Research Fellowship, is a fellow of the American Association for the Advancement of Science, a member of the European Molecular Biology Organization, and a corresponding member of the Austrian Academy of Sciences.

Jennifer Listgarten, Ph.D.

Department of Electrical Engineering and Computer Science
University of California, Berkeley

Jennifer Listgarten, Ph.D., is a Professor in the Department of Electrical Engineering and Computer Science, and Center for Computational Biology, at the University of California, Berkeley. She is also a member of the steering committee for the Berkeley AI Research (BAIR) Lab, and a Chan Zuckerberg investigator. From 2007 to 2017 she was at Microsoft Research, through Cambridge, MA (2014-2017), Los Angeles (2008-2014), and Redmond, WA (2007-2008). She completed her Ph.D. in the machine learning group in the Department of Computer Science at the University of Toronto, located in her home town. She has two undergraduate degrees, one in Physics and one in Computer Science, from Queen's University in Kingston, Ontario. Dr. Listgarten's research interests are broadly at the intersection of machine learning, applied statistics, molecular biology and science.

Ben Lehner, Ph.D.

ICREA Research Professor and Coordinator of CRG Systems Biology Program
Centre for Genomic Regulation (CRG),
Barcelona Institute of Science and Technology

Ben Lehner, Ph.D., is an ICREA Research Professor at the EMBL-CRG Systems Biology Program at the Centre for Genomic Regulation. He has a B.A. and a Ph.D. from the University of Cambridge and was a post-doctoral fellow at the Wellcome Trust Sanger Institute. The goal of his laboratory is to understand and predict phenotypic variation amongst individuals, including the distribution and effects of inherited genetic variation, somatic mutations, the environment and stochastic influences. His laboratory combines experimental and computational approaches, and, as model systems, they use yeast, worms and tumors in roughly equal proportions, Dr. Lehner is a member of the European Molecular Biology Organization, and has received numerous awards including the EMBO Gold Medal, Paper of the Year from the Catalan Society of Biology, the Bettencourt Prize for Life Sciences, The Genetics Society Balfour Lecture and the Eppendorf/Nature Award for a Young European Investigator.

Organizer Bios

Inaugural Integrative Biology Symposium
Friday, September 13, 2019

Salk Institute for Biological Studies
La Jolla, California

Tatyana Sharpee, Ph.D.

Integrative Biology Laboratory
Computational Neurobiology Laboratory
Salk Institute for Biological Studies

Tatyana Sharpee, Ph.D. is a Professor in the Computational Neurobiology Laboratory and Integrative Biology Laboratory at the Salk Institute for Biological Studies. Dr. Sharpee completed her M.S. in Theoretical Physics at Kiev National University in Ukraine, her Ph.D. in Theoretical Physics at Michigan State University, and a postdoc in Computational Neurobiology at the University of California, San Francisco. Her laboratory works to understand how the nervous system functions in its natural environment. This includes developing statistical tools to analyze responses of mid- and high-level sensory neurons to natural stimuli and building maximally informative behavioral strategies. They have also developed theories for how different neuronal types work together to achieve an accurate encoding of sensory stimuli and information contained in the responses of large neuronal populations. Dr. Sharpee is an elected fellow of the American Physical Society and has received numerous awards including the Alfred P. Sloan Research Fellowship, the Searle Scholar Award, the McKnight Scholar Award, and the NSF CAREER award.

Edward Stites, M.D., Ph.D.

Integrative Biology Laboratory
Salk Institute for Biological Studies

Edward Stites, M.D., Ph.D., is an Assistant Professor in the Integrative Biology Laboratory at the Salk Institute for Biological Studies. He worked in the laboratory of Kodi Ravichandran at the University of Virginia during the pursuit of his M.D. and Ph.D. degrees. A two-year Randy Pausch Scholar fellowship took him to the Translational Genomics Research Institute to work on early personalized medicine clinical trials, genomics, computational biology, and pancreatic cancer. He then completed a clinical residency in Clinical Pathology at Washington University in St. Louis, where he also completed postdoctoral research with Andrey Shaw. His research laboratory studies the RAS oncogene and cancer with the mathematical methods of dynamical systems and with traditional experimental biology.

Wolfgang Busch, Ph.D.

Plant Molecular and Cellular Biology Laboratory
Integrative Biology Laboratory
Salk Institute for Biological Studies

Wolfgang Busch, Ph.D., is an Associate Professor in the Plant Molecular and Cellular Biology Laboratory, as well as in the Integrative Biology Laboratory at the Salk Institute for Biological Studies in La Jolla, California. He is also an Associate Adjunct Professor in the Cell & Developmental Biology Division of Biological Sciences at the University of California, San Diego. His work focuses on understanding which genes, genetic networks, and molecular processes determine root phenotypes. For this, his laboratory exploits natural genetic variation in the model plant *Arabidopsis* and uses a systems genetics approach that combines large-scale phenotyping, genome wide association studies, genetics, and genomics to find and characterize genes, their alleles, and the genetic networks that ultimately determine root growth. Recently, Dr. Busch, together with the other four Salk plant biology faculty members, has founded the Salk Harnessing Plants Initiative in which the group is researching and developing plant varieties with enhanced carbon sequestration capabilities for removing CO_2 from the atmosphere and thereby counteracting climate change.

Dr. Busch did his undergraduate studies in Biology at the University of Tübingen, Germany, and obtained his PhD at the Max Planck Institute for Developmental Biology in Tübingen. He received his postdoctoral training at Duke University before he joined the Gregor Mendel Institute of Molecular Plant Biology in Vienna as a group leader in 2011. In 2017, he became an Associate Professor at the Salk Institute for Biological Studies in La Jolla. For his work, Dr. Busch received multiple international awards, including the Genome Web Young Investigator Award in 2014 and the SEB president's medal in 2015.

Saket Navlakha, Ph.D.

Pioneer Fund Developmental Chair
Integrative Biology Laboratory
Salk Institute for Biological Studies

Saket Navlakha, Ph.D., is an Associate Professor in the Integrative Biology Laboratory at the Salk Institute for Biological Studies. He received an AA from Simon's Rock College in 2002, a B.S. from Cornell University in 2005, and a Ph.D. in computer science from the University of Maryland College Park in 2010. He was then a post-doc in the Machine Learning Department at Carnegie Mellon University before starting his lab at the Salk Institute in 2014. He is interested in understanding the algorithms that biological systems have evolved to solve computational problems critical for survival. In 2018, he was named a Pew Biomedical Scholar and a NSF CAREER award recipient.

Graham McVicker, Ph.D.

Frederick B. Rentschler Developmental Chair
Laboratory of Genetics
Integrative Biology Laboratory
Salk Institute for Biological Studies

Graham McVicker, Ph.D., is an Assistant Professor in the Laboratory of Genetics and the Integrative Biology Laboratory at the Salk Institute for Biological Studies. Dr. McVicker received his B.Sc. in Computer Science from the University of British Columbia and subsequently worked on the Ensembl Genome Browser at the European Bioinformatics Institute. Dr. McVicker then did his Ph.D. with Phil Green at the University of Washington where he researched how selection shapes human genetic variation. During his post-doc with Jonathan Pritchard at the University of Chicago and Stanford University, Dr. McVicker studied how genetic variation affects gene regulation and chromatin. The McVicker Laboratory is currently developing new computational and statistical methods to discover somatic regulatory mutations in cancer genomes, and manipulating cell lines with high-throughput CRISPR technologies to discover new regulatory sequences.

Short Talk Abstracts

Inaugural Integrative Biology Symposium
Friday, September 13, 2019

Salk Institute for Biological Studies
La Jolla, California

Quantitative Tumor Expression Dimensions Deconstruct Myeloma Heterogeneity

Rosalie Waller[1], Michael Madsen[2], John Gardner[2], Douglas Sborov[3], and Nicola Camp[3]
[1]Biomedical Informatics, University of Utah
[2]Huntsman Cancer Institute, University of Utah
[3]Internal Medicine, University of Utah

Tumor heterogeneity hinders clinical management and research. Today approaches to reduce heterogeneity across patients focus on categorical subtype classification. However, these subtypes are mutually exclusive and may fail to capture important variation. We propose an alternative approach: quantitative tumor expression dimensions (QTEDs). Each dimension is an independent tumor characteristic capturing important global transcriptome variation across patients. QTEDs are unique linear combinations across key genes derived from principal component analysis (PCA) that alone or together can uncover associations with clinically-relevant endpoints.

A better understanding of tumor characterisitcs could improve prognosis in myeloma, the second most common hematologic malignancy. We derived QTEDs from RNA sequencing on treatment-naïve, sorted myeloma tumors from 768 patients. Multi-stage PCA was performed on the normalized gene-based expression counts. We investigated distinct dimensions and multi-dimension signatures for associations with demographic, clinical, and genetic characteristics using penalized linear regression modeling.

We identified 28 QTEDs of interest in the myeloma tumors and characterized multi-dimension signatures for each of the 768 patients. After correcting for multiple testing, 29 associations between QTEDs and clinical/demographic variables were significant. Dimensions 1-8 were associated with somatic alterations known to correlate with myeloma prognosis ($p = 8.2x10\text{-}5\text{-}1.6\ x10\text{-}23$). Dimensions 11 and 24 were significantly associated with African heritage ($p = 2.9x10\text{-}7$ and $8.0x10\text{-}7$), indicating the possibility that molecular variability may contribute to race disparities seen for myeloma which is twice as common in people of African descent. Ongoing work is investigating key genes driving the dimensions associated with clinical features.

Quantitative Model of the Complement System Links Susceptibility of Meningococcal Disease to Nasal Complement Levels

Nehemiah Zewde, Rohaine Hsu, and Dimitrios Morikis
Department of Bioengineering, University of California, Riverside, CA 92521

Neisseria meningitidis (*N. meningitidis*), an inhabitant of the human nasopharynx, is a pathogen responsible for meningococcal disease (MD). Studies have shown individuals with deficiencies in the complement system, notably the membrane attack complex (MAC), have a 7,000- to 10,000-fold higher risk of developing MD. Here, we developed a quantitative model of the complement system to assess dynamics of MAC production and predict why some individuals exposed to *N. meningitidis* develop MD while others remain healthy carriers with no signs of disease. Our model is comprised of the complete complement pathway (alternative, classical, and lectin) with 670 ordinary differential equations and 328 kinetic parameters. We then used fluid phase activation of the complement system to generate MAC concentration profiles on *N. meningitidis*. To account for complement activation in the bloodstream and nasopharynx, we used serum and nasal complement concentrations, respectively. Our model shows when *N. meningitidis* invades the bloodstream, MAC levels occupy <0.01% of the pathogen surface, whereas in the nasopharynx, MACs can occupy between 1.3% to 10% of the pathogen surface. Importantly, our model identified complement proteins C5, C6 and C9 as major components in MAC production against *N. meningitidis* (nasopharynx). Lastly, MAC production is dependent on a concentration barrier where complement activators should be at least three orders of magnitude higher to regulators. Altogether, our studies indicate nasal complement levels generate higher MACs on *N. meningitidis* if the concentration barrier is maintained. However, a compromised barrier leads to lower MACs and subsequently increases the potential of individuals developing meningococcal disease.

Systematic Characterization of Gene Functions in Photosynthetic Organisms

Josep Vilarrasa Blasi
Biology, Stanford University, Stanford, CA 94305

Photosynthetic organisms are essential for human life, yet most of their genes remain uncharacterized. Single-celled photosynthetic model systems have the potential to dramatically accelerate the connection of genotype to phenotype. Using a barcoded mutant library of the single-celled model eukaryotic alga *Chlamydomonas reinhardtii*, we determined the phenotypes of more than 60,000 mutants under more than 100 different environmental and chemical stress conditions. We place genes with unknown function into functional pathways based on similarity of mutant phenotypes. We illustrate the power of our approach for providing insights into conserved gene functions by confirming the same functional role in the land plant A*rabidopsis thaliana*. We anticipate that the phenotypic data presented in this study will guide the functional characterization of currently unknown genes across the green lineage of algae and plants.

Mechanistic Origin of Cell-size Control and Homeostasis in Bacteria

Fangwei Si[1], Guillaume Le Treut[1], John T. Sauls[1], Stephen Vadia[2],
Petra Anne Levin[2], and Suckjoon Jun[3]
[1]Department of Physics, University of California, San Diego, La Jolla, CA 92093
[2]Department of Biology, Washington University in St. Louis, MO 63130
[3]Department of Physics and Section of Molecular Biology,
University of California, San Diego, La Jolla, CA 92093

Evolutionarily divergent bacteria share a common phenomenological strategy for cell-size homeostasis under steady-state conditions. In the presence of inherent physiological stochasticity, cells following this "adder" principle gradually return to their steady-state size by adding a constant volume between birth and division regardless of their size at birth. However, the mechanism of the adder has been unknown despite intense efforts. In this work, we show that the adder is a direct consequence of two general processes in biology: (1) threshold -- accumulation of initiators and precursors required for cell division to a respective fixed number, and (2) balanced biosynthesis -- maintenance of their production proportional to volume growth. This mechanism is naturally robust to static growth inhibition, but also allows us to "reprogram" cell-size homeostasis in a quantitatively predictive manner in both Gram-negative *Escherichia coli* and Gram-positive *Bacillus subtilis*. By generating dynamic oscillations in the concentration of the division protein FtsZ, we were able to oscillate cell size at division and systematically break the adder. In contrast, periodic induction of replication initiator protein DnaA caused oscillations in cell size at initiation, but did not alter division size or the adder. Finally, we were able to restore the adder phenotype in slow-growing *E. coli*, the only known steady-state growth condition wherein *E. coli* significantly deviates from the adder, by repressing active degradation of division proteins. Together these results show that cell division and replication initiation are independently controlled at the gene-expression level, and that division processes exclusively drive cell-size homeostasis in bacteria.

Extra Legs in the Journey from Genotype to Phenotype

Amelie Baud[1], Francesco Paolo Casale[2], Abraham Palmer[1], and Oliver Stegle[2]
[1]Department of Psychiatry, University of California, San Diego, La Jolla, CA 92093
[2]EMBL - EBI, Hinxton, Cambridge, United Kingdom CB10 1SD

Genetic effects on an individual's phenotype can arise not only from the individual's own genotypes but also from genotypes of interacting partners. Such Social Genetic Effects (SGE) arise when heritable traits of social partners influence the phenotype of interest.
SGE have been largely ignored in biomedical genetics. To study SGE, we developed new statistical methods to jointly model 'traditional' (direct) genetic effects (DGE) and SGE in genetically heterogeneous populations. We analysed large datasets from laboratory mice and rats and quantified aggregate SGE from cage mates. We found that SGE significantly affected a broad range of behavioural, physiological and morphological phenotypes. SGE revealed unsuspected social effects: for example, wound healing was significantly associated with genetic makeup of cage mates in all three datasets examined, with SGE explaining up to 14% of phenotypic variation.
These results open new avenues for dissecting the mechanisms of social effects. For example, the genome-wide association study of SGE (sgeGWAS) can be used to identify genes of partners giving rise to SGE and, ultimately, traits of partners influencing the phenotype of interest. sgeGWAS in the aforementioned datasets provided proof of principle for this strategy and important insights into the comparative architectures of SGE and DGE. In particular, SGE and DGE acting on a given phenotype typically arise from different loci and SGE loci have lower effect sizes than DGE loci. Our results shed light on a new component of the genotype to phenotype path and opportunities to use SGE to better understand effects from the social environment.

Poster Abstracts

Inaugural Integrative Biology Symposium
Friday, September 13, 2019

Salk Institute for Biological Studies
La Jolla, California

Alternating Phases of AMPK and mTORC1: A Potential Mechanism of Mitochondrial Turnover Initiation during Electron Transport Chain Inhibition

Jennifer Baily
MCB, University of California, Davis, CA 95616

Maintaining adequate levels of energy is crucial to cellular survival. A key regulatory kinase involved in restoring energetic homeostasis is AMP-activated protein kinase (AMPK). AMPK monitors cellular energy charge and is activated by an increase in AMP:ATP ratios. In turn, AMPK activity can restore ATP pools through a variety of mechanisms, including: 1) activation of mitochondrial biogenesis and autophagic process, and 2) suppression of protein translation via inhibition of mTORC1. This balancing of anabolic and catabolic processes allows cells to recover during electron transport chain (ETC) inhibition, starvation, and other energetic assaults. Both AMPK activity and mitochondrial function decline with age, suggesting that AMPK acts as a feedback controller of mitochondrial biogenesis that can augment mitochondrial function to restore impaired ATP production. However, it is unclear how new mitochondria can be generated when AMPK suppresses protein translation. In single cells, AMPK undergoes oscillations in activity when cells are treated with ETC inhibitors, suggesting a potential mechanism by which the alternating phases of AMPK activity function to maintain energetic homeostasis. We hypothesize that the oscillatory kinetics enable AMPK to regulate the generation of mitochondria by initiating alternating periods of biogenesis and clearance during ATP depletion. Supporting this model, we have collected data that demonstrates that AMPK and mTORC1 are anticorrelated by monitoring the regulation and localization of the mTORC1 target, TFEB. The observed correspondence in AMPK and TFEB oscillations suggest a kinetic mechanism by which anabolic and catabolic processes are balanced at the single cell level during adaptation to energetic stress.

Integrating Mathematical Modeling with Machine Learning to Identify Cancer Driver Genes

Seo-Yeon Chung
Department of Mathematics, University of California, Irvine, CA
Northwood High School

The identification of cancer driver genes is a critical component of precision oncology. Given the large feature space of The Cancer Genome Atlas (TCGA), which catalogs millions of somatic mutations observed in human tumors, machine learning techniques are ideally suited to driver gene identification. In existing models, however, the objective assessment of such machine learners is complicated by unexplained errors in the mutational data used to train the algorithms and by the absence of a perfect drivers list. This study employs mathematical modeling in tandem with machine learning processes to construct an objective and accurate classifier that identifies cancer driver genes. A set of in-silico mutational data is generated by the stochastic simulation of a differential equations model of feedback-controlled cancer population dynamics. The synthetic dataset, validated through the assessment of mutational pattern distributions, trains a selected machine learning algorithm, producing a driver gene classifier. The gene classifier is shown to accurately prioritize high-impact driver genes in four cancer types. Additionally, the quality of the ranked list of putative driver genes is validated through enrichment analysis on a list of generally accepted driver genes and biological pathway analysis. Top colorectal cancer driver genes from the classifier hold key roles in the PI3K-AKT and Wnt pathways, which have well-documented implications in carcinogenesis. The interdisciplinary methodology developed here produces a more efficient and unbiased cancer driver gene classifier that can be utilized to identify henceforth unknown driver genes, providing insight essential for targeted cancer screening and treatment.

Overlap of Fetal-specific Cardiac Regulatory Variants and GWAS Lead Variants Supports Fetal Origins of Cardiovascular Disease

Matteo D'Antonio
Insitute for Genomic Medicine, UC San Diego
Matteo D'Antonio[1], Margaret Donovan[2], William Young Greenwald[2], David Jakubosky[3], Hiroko Matsui[1], Paola Benaglio[4], Erin Smith[4], Agni D'Antonio-Chronowska[1], and Kelly Frazer[5]
[1]Institute for Genomic Medicine,
[2]Bioinformatics and Systems Biology Graduate Program,
[3]Biomedical Sciences Graduate Program,
[4]Department of Pediatrics and Rady Children's Hospital, and
[5]Institute for Genomic Medicine and Department of Pediatrics and Rady Children's Hospital, University of California, San Diego, La Jolla, CA 92093

It has been hypothesized that many disease-causing variants exert their effects during development, rather than in adult cells. However, it is difficult to identify these variants and their effects as they could act in multiple different cell types, and there was a recent moratorium on research using fetal tissue. We recently established that iPSC-derived cardiovascular progenitor cells (CVPCs) are fetal-like, and can be utilized to identify cardiac regulatory variants. Here, we leveraged this system to identify fetal cell-type specific eQTLs that underlie GWAS signals for adult cardiac diseases.
We started by characterizing the differentiation of iPSCs into iPSC-CVPCs via scRNA-seq on eight samples, and found they were comprised of two cardiac cell types: cardiomyocytes (CMs) and epicardium derived cells (EPDCs). Next, we derived 180 iPSC-CVPCs, performed bulk RNA-seq, and used the scRNA-seq expression signatures to deconvolute and determine the relative proportions of CMs and EPDCs in each sample. We integrated these data with WGS and identified cell type-specific eQTLs (associated with only CMs or EPDCs). We next identified fetal-specific eQTLs by colocalizing our iPSC-CVPC eQTLs with all GTEx adult cardiac tissue eQTLs. To identify variants underlying the fetal origin of complex adult cardiac traits, we colocalized these fetal-specific eQTLs with cardiac traits GWAS summary statistics, and found 10 fetal-specific eGenes, including *CLPTM1*, which is associated with congenital malformations. Our findings provide genetic evidence supporting the fetal origin of cardiovascular disease and show that iPSC-derived tissues can be leveraged to study the fetal origins of diseases in relevant cell-types.

Empowering Disease Research Using Population-scale Genetic Datasets

Christopher DeBoever
Institute of Genomic Medicine, University of California, San Diego,
La Jolla, CA 92093

The advent of population-scale datasets that combine genetic and high-dimensional phenotype data for thousands of participants has revolutionized disease research. While biobanks contain a wealth of phenotypic information for each subject compared to traditional ascertainment-based studies, specific diseases are generally represented near their population prevalence which can hamper studies of individual diseases. New statistical methods that can jointly analyze multiple phenotypes are therefore needed to fully take advantage of biobank resources that link genetic data with health records and other phenotypic measurements. We have developed a suite of methods to study disease genetics in the context of multi-dimensional phenotyping that jointly leverage information from multiple phenotypes to estimate genetic parameters such as genetic correlations, model disease risk, and identify likely disease-associated variants and genes. These methods use GWAS summary statistics as input which allows meta-analyses across biobanks and with previous studies while protecting the privacy of participants. We have used these methods to evaluate digital phenotyping approaches that leverage unstructured questionnaire data for GWAS across more than 40 phenotypes in the UK Biobank. We have also applied these methods to identify novel gene drug targets by jointly analyzing summary statistics from diseases such as asthma and glaucoma along with related quantitative traits. This work demonstrates how combining the rich phenotypic data available for hundreds of thousands of subjects in biobank resources with multivariate analysis methods provides novel opportunities for understanding the genetic basis of disease and identifying therapeutic targets.

Cellular Deconvolution of GTEx Tissues Powers eQTL Studies to Discover Thousands of Novel Disease and Cell-type Associated Regulatory Variants

Margaret Donovan
Bioinformatics, University of California, San Diego, La Jolla, CA 92093

The Genotype-Tissue Expression (GTEx) resource has been used to study the regulatory impact of genetic variation on gene expression across tissues types, but has not yet been used to examine how variation acts at cell-type-specific resolutions. The cellular composition of heterogeneous tissues can be deconvoluted using expression signatures specific to each cell type identified in single-cell RNA-seq (scRNA-seq) generated from an analogous tissue. However, there are relatively few human scRNA-seq resources available, and thus only a small fraction of GTEx tissues can be currently be deconvoluted.

To address this gap, we used mouse scRNA-seq from the Tabula Muris resource to deconvolute bulk RNA-seq from over 6,000 GTEx samples representing 28 tissues. We detected distinct cell types within each of the GTEx tissues and the proportion of these cell types varied substantially across samples of the same tissue type. In two deconvoluted GTEx tissues (liver and skin), we performed eQTL analyses considering cellular compositions and identified thousands of cell-type-associated eQTLs. We examined the functional impact of the cell-type-associated eQTLs identified in skin, and showed that they colocalize with variants in GWAS loci for melanoma, malignant neoplasm, and infection signatures; including six eQTLs associated with both leukocytes and malignant skin neoplasms that regulate genes known to play a role in cancer progression or immune response (TCF19, ATAD3C, SERPINB9, NT5C2, and CD1E).

Our study provides a framework that can be implemented immediately to deconvolute the cellular composition of GTEx tissues for characterizing the functional impact of cell-type-associated genetic variation on human disease.

Identifying Novel Regulatory Elements using RELICS, a Statistical Framework for the Analysis of CRISPR Regulatory Screens

Patrick Fiaux[1], Hsiuyi Chen[1], Ishika Luthra[2], Carolyn O'Connor[3], and Graham McVicker[1]
[1]Integrative Biology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037
[2]Department of Engineering, Simon Fraser University, British Columbia, Canada V5A 1S6
[3]Flow Cytometry Core, Salk Institute for Biological Studies, La Jolla, CA 92037

High-throughput CRISPR/Cas9 screens are a powerful new tool for the systematic discovery of regulatory elements in the human genome. In these regulatory screens, thousands of guide RNAs (gRNAs) are delivered to cells to target potential regulatory sequences for mutation, activation or inhibition. The cells are then sorted into high- and low-expression pools based on the expression of a target gene. While, these screens have the potential to perform unbiased discovery of regulatory elements, they generate noisy data and the performance of analysis methods has not been rigorously assessed. Here we describe RELICS, a statistical framework for Regulatory Element Identification in CRISPR Screens. RELICS models the observed guide counts in different expression pools with a generalized linear mixed model. This approach is very flexible, can jointly model multiple expression pools (beyond just high and low), incorporate variability across guides, and accommodate over-dispersion. To assess the performance of RELICS we have developed a simulation framework for generating CRISPR regulatory screen data and simulated 1000s of data sets under a wide variety of experimental and biological conditions. RELICS outperforms existing analysis methods on the simulated data and we have applied it to identify regulatory elements in several published datasets. In addition, we have applied RELICS to data from a paired-guide regulatory screen that we performed for *GATA3* in Jurkat T cells. We identify a total of 23 putative regulatory elements within the 2MB targeted region surrounding *GATA3*. Notably 16 of the identified elements lie within the same topological associating domain as *GATA3*.

Predicting Age from the Transcriptome of Human Dermal Fibroblasts

Jason Fleischer[1], Roberta Schulte[2], Hsaio Tsai[2], Swati Tyagi[2], Maxim Shokirev[3], Ling Huang[3], Jerome Mertens[4], Fred Gage[4], Martin Hetzer[2], and Saket Navlakha[1]
[1]Integrative Biology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037
[2]Molecular and Cell Biology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037
[3]Bioinformatics Core Facility, Salk Institute for Biological Studies, La Jolla, CA 92037
[4]Laboratory of Genetics, Salk Institute for Biological Studies, La Jolla, CA 92037

Biomarkers of aging can be used to assess the health of individuals, and to study aging and age-related diseases. We generated a large dataset of genome-wide RNA-seq profiles of human dermal fibroblasts (133 individuals, aged 0-94 years old) to test whether signatures of aging are encoded within the transcriptome. We developed an ensemble machine learning method that predicted age to a median error of 4 years in leave one out cross validation, outperforming previous methods used to predict age. The ensemble and standard regression based methods were tested on 10 Progeria patients; the ensemble was the only method that correctly predicted accelerated aging in these patients. The ensemble was further validated on a second separate skin fibroblast RNA-seq dataset (51 samples, aged 0-89 years), where cross validation performance was again 4 years median error. However, cross dataset age prediction is currently not accurate. Solving this problem is the topic of ongoing research.

Genomic Properties of Structural Variants and Short Tandem Repeats that Impact Gene Expression and Complex Traits in Humans

David Jakubosky[1], Matteo D'Antonio[2], Margaret Donovan[1], Marc Bonder[3], Craig Smail[4], Agnieszka D'Antonio-Chronowska[2], Oliver Steagle[5], Stephen Montgomery[6], Christopher DeBoever[2], and Kelly Frazer[7]
[1]Biomedical Informatics, University of California, San Diego, La Jolla, CA 92093
[2]Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA 92093
[3]Genome Biology, European Molecular Biology Laboratory, Genome Biology Unit
[4]Biomedical Data Science, Stanford University, Stanford, CA 94305
[5]Genome Biology, European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, United Kingdom, CB10 1SD
[6]Genetics, Stanford University, Stanford, CA 94305
[7]Pediatrics, University of California, San Diego, La Jolla, CA 92093

Structural variants (SVs) and short tandem repeats (STRs) are important sources of genetic diversity but are not routinely analyzed in genetic studies because they are difficult to accurately identify and genotype. Here we assembled a set of 719 deep whole genome sequencing (WGS) samples (mean 42x) from 477 distinct individuals which we used to discover and genotype a wide spectrum of SV and STR variants using five algorithms. We used 177 unique pairs of genetic replicates to identify factors that affect variant call reproducibility and developed a systematic filtering strategy to create of one of the most complete and well characterized maps of SVs and STRs to date. Functional characterization showed that different SV classes and STRs vary in their impact on gene expression and complex traits. Among variants associated with gene expression, we found that those from different variant classes had unique properties such as their genomic locations relative to eGenes, likelihood of being associated with multiple eGenes, associated eGene types (e.g., coding, noncoding, level of evolutionary constraint), effect sizes, linkage disequilibrium with tagging single nucleotide variants used in GWAS, and likelihood of being associated with GWAS traits. We also identified a set of high-impact SVs/STRs associated with the expression of three or more eGenes via chromatin loops and showed that they are highly enriched for being associated with GWAS traits. Our study provides insights into the genomic properties of structural variant classes and short tandem repeats that impact gene expression and human traits.

Diabetic and Healthy Human Neutrophil Transcriptomics

Sarah Kleinstein[1], Jamison McCorrison[2], Alaa Ahmed[3], Hatice Hasturk[3], Thomas Van Dyke[3], and Marcelo Freire[4]
[1]Human Genomic Medicine, J. Craig Venter Institute, La Jolla, CA 92037
[2]Informatics, J. Craig Venter Institute, La Jolla, CA 92037
[3]The Forsyth Institute, Harvard School of Dental Medicine, Cambridge, MA 02142
[4]Genomic Medicine, J. Craig Venter Institute, La Jolla, CA 92037

Chronic inflammatory diseases, such as diabetes and cardiovascular disease, are heterogeneous and often co-morbid, with increasing global prevalence. Type 2 diabetes (T2D) can result in severe inflammatory complications. As neutrophils are essential to inflammation, we conducted RNA-seq transcriptomic analyses to investigate the association between neutrophil gene expression and T2D phenotype. Further, as specialized pro-resolving lipid mediators, such as resolvin E1 (RvE1), can act to resolve inflammation, we investigated the impact of RvE1 treatment on neutrophil gene expression using a perturbation study.
Cell isolation and RNA-seq analysis of neutrophils from N=11 T2D and N=7 healthy individuals with available clinical data was conducted. Additionally, cultured neutrophils (N=3 T2D, N=3 healthy) were perturbed with increasing RvE1 doses (0nM, 1nM, 10nM, or 100nM) prior to RNA-seq. Data was evaluated through a bioinformatics pipeline including pathway analysis and FDR-correction.
We observed significant differential expression of 50 genes ($p<0.05$) between T2D and healthy neutrophils, including decreased T2D gene expression in immune- and lipid-related genes *SLC9A4*, *NECTIN2,* and *PLPP3* ($p<0.003$). RvE1 treatment also induced differential gene expression (uncorrected $p<0.05$) across treatment groups, including 59 healthy and 216 T2D neutrophil genes. Comparing T2D to healthy neutrophils, 1097 genes were differentially expressed across treatment dosages, including two significant inflammatory genes: *LILRB5* and *AKR1C1* ($p<0.05$).
Immune and lipid genes were differentially expressed between T2D and healthy neutrophils. Additionally, RvE1 dose-response explained significant variation between T2D and healthy cells. Further validation of genomic differences between diabetic and healthy individuals could elucidate important inflammatory mechanisms and disease subsets.

Cell-to-cell Variability in AMPK Activation Reveals Autonomous Cycles in Cellular Energy Balance

Nont Kosaisawe
BMCDB, University of California, Davis, CA 95616

Cellular metabolism can be reconfigured to balance nutrient availability and cellular demands. However, the kinetics of reconfiguration within individual cells, and the extent of heterogeneity between cells, remain uncharacterized. Using live-cell imaging of ATP concentration, ADP/ATP ratio, and the energy stress sensor AMPK, we investigate the kinetics of bioenergetic adaptation in individual cells. In response to acute inhibition of oxidative phosphorylation, AMPK is activated bimodally, reflecting cell-to-cell variability in ADP/ATP ratio despite unchanging ATP concentrations. We use these responses to quantify the distribution of metabolic configurations within genetically identical cell populations, and identify compounds that shift this distribution, including inhibitors of glycolysis, insulin/Akt signaling, and protein synthesis. Through long-term analysis of cell lineages, we demonstrate that the configuration of cellular energy balance cycles over time within individual cells in a cell cycle-independent manner. Variability in AMPK activation is transmitted to inhibition of the ERK and mTOR cell growth signaling pathways. Our results establish kinetic details of cellular energy homeostasis, reveal dynamic fluctuations of energetic balance, and demonstrate how plasticity in metabolism influences the overall cell proliferation regulatory network.

Simultaneous Profiling of 3D Genome Structure And DNA Methylation in Single Human Brain Cells

Dongsung Lee
Clayton Foundation Laboratories for Peptide Biology Laboratory,
Salk Institute for Biological Studies, La Jolla, CA 92037

Dynamic 3D chromatin conformation is a critical mechanism for gene regulation during development and disease. Despite this, profiling of 3D genome structure from complex tissues with cell-type specific resolution remains challenging. Recent efforts have demonstrated that cell-type specific epigenomic features can be resolved in complex tissues using single-cell assays. However, it remains unclear whether single-cell Chromatin Conformation Capture (3C) or Hi-C profiles can effectively identify cell types and reconstruct cell-type specific chromatin conformation maps. To address these challenges, we have developed a multi-omic method, single-nucleus methyl-3C sequencing (sn-m3C-seq) to capture chromatin organization and DNA methylation information and robustly separate heterogeneous cell types. Applying this method to >4,200 single human brain prefrontal cortex cells, we reconstruct cell-type specific chromatin conformation maps from 14 cortical cell types. These datasets reveal the genome-wide association between cell-type specific chromatin conformation and differential DNA methylation, suggesting pervasive interactions between epigenetic processes regulating gene expression.

Integrated Analysis of NGS and Optical Mapping Resolves the Complex Structure of Highly Rearranged Focal Amplifications in Cancer

Jens Luebeck[1], Viraj Deshpande[2], Ceyda Coruh[3], Utkrisht Rajkumar[2], Siavash Dehkordi[2], Dave Pai[4], Kristen M. Turner[5], Julie A. Law[3], Paul S. Mischel[5], and Vineet Bafna[2]
[1]Bioinformatics & Systems Biology, University of California, San Diego, La Jolla, CA 92093
[2]Computer Science & Engineering, University of California, San Diego, La Jolla, CA 92093
[3]Plant Molecular and Cell Biology Laboratory, Salk Institute, La Jolla, CA 92037
[4]BioNano Genomics, Inc.
[5]Ludwig Institute for Cancer Research, University of California, San Diego, La Jolla, CA 92093

Copy number amplifications (CNA) are a hallmark of the cancer genome. Presence of such focal amplifications has been associated with genome instability, increased pathogenicity, as well as the presence of circular extrachromosomal DNA. Yet, the mechanisms causing focal CNAs and ecDNA are incompletely understood. New reports suggest circular extrachromosomal DNA (ecDNA) exists in up to 40% of cancer types, and are an important driver of CNA.

An earlier tool to analyze focal CNAs (Deshpande 2019), used next-generation sequencing (NGS) data to create a graph encoding rearrangement breakpoints, as a prelude to identifying the full structure. Paths extracted from breakpoint graphs provide signatures of rearrangement events, but are complex and rarely admit an unambiguous structure due to the complexity of focal CNAs. We present a method, Amplicon Reconstructor (AR), which integrates NGS data with optical mapping (OM) data or long-read data. OM data provides physical maps of DNA which can be assembled into ultra-long OM assemblies (N50 ~50 Mbp). AR employs a graph-based method to identify long paths and cycles in a breakpoint graph. Simulations validate AR-reconstructed amplicons, demonstrating high fidelity of our approach.

We applied AR to data from seven patient-derived and immortalized cell lines, using multiple cytogenetic approaches to validate our findings. AR reconstructed ecDNA structures, each larger than 1 Mbp, in three glioblastoma cell lines. In K562 cells, AR identified a complex rearrangement including the chr9-chr22 BCR-ABL fusion, and also genomic regions from chr13. In the HCC827 cell line, AR enabled complete reconstruction of a breakage fusion bridge. Our results provide reconstructions giving unique insight into focal CNA structure, unraveling the architecture of CNA in ecDNA-positive cancers.

In-Silico Models Suggest a Novel Mechanism for Paradoxical Activation by Kinase Inhibitors

Gaurav Mendiratta
Integrative Biology Laboratory, Salk Institute for Biological Studies,
La Jolla, CA 92037

An unexpected response to RAF oncogene inhibitor is observed with WT RAF where RAF inhibitors induce pathway activity. This 'Paradoxical Activation' occurs with drugs that successfully inhibit RAF kinase function but increase signaling in the pathway nonetheless. We use *in-silico* models of RAF kinase dimerization and drug binding to provide a novel explanation for the paradoxical activation. We generate hypotheses to test our model *in-vitro* and perform experiments to validate predictions based on our proposed mechanism. Our mechanism generalizes beyond the case of RAF kinase and suggests a potential for paradoxical activation with inhibitors for other dimer-active kinases such as PERK and RNaseL.

Growth and Stress Signaling Networks of Lung Epithelium in ALI

Devan Murphy
Molecular and Cellular Biology, University of California, Davis, CA 95616

The lung epithelium is a renewing tissue under continual stress from the environment. Lung disease, such as acute lung injury (ALI), results from an imbalance in the repair pathways that regulate lung healing. However, the current models used to ALI are inadequate to understand the complex signaling pathways and cellular behaviors that underlie the healing process. The core networks that dictate epithelial behavior include ERK, Akt, and mTOR, which modulate cell proliferation, and the stress response pathways of AMPK, NF-kB, and p38, which coordinate cytokine release and apoptosis. Our hypothesis is that alternating periods of pro-growth (ERK, Akt, and mTOR) and stress (AMPK, NF-kB, and p38) balance these cell fates to regulate epithelial maintenance, whereas uncoordinated signaling results in disease states seen during ALI. Our overall objective is to quantitatively decipher how proliferation, apoptosis, senescence, and cytokine secretion are governed by multi-kinase activity programs. A single-cell approach to this problem provides the temporal and spatial resolution essential to unmask the heterogeneity in cell fate decisions. Using live-cell imaging coupled with immunofluorescence, I will determine the underlying relationship between the signaling patterns (timing, intensity, and coordination between pathways) and cellular behavior. Our approach will develop the technology needed to detect and manipulate these temporal signaling programs to reestablish lung epithelial homeostasis in ALI. These findings will identify potential new routes of intervention for ALI, as well as have broader implications uncovering general concepts in lung epithelial homeostasis which may be utilized for the treatment of other lung malignancies.

Genomic Spatial Profiling of Archived Precancer Lesions Enabled by Improved Targeted Sequencing of Low Input FFPE DNA

Daniela Nachmanson[1], Joseph Steward[2], Huazhen Yao[3], Eliza Jeong[4], Thomas O'Keefe[2], Gillian L. Hirst[5], Farnaz Hasteh[6], Kristen Jepsen[3], Laura J. Esserman[7], and Olivier Harismendy[8]
[1]Bioinformatics and Systems Biology, University of California, San Diego, La Jolla, CA 92093
[2]Moores Cancer Center, University of California, San Diego, La Jolla, CA 92093
[3]Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA 92093
[4]Hellen Diller Family Comprehensive Cancer Center, University of California, San Francisco, CA 94143
[5]Surgery, University of California, San Francisco, CA 94143
[6]Pathology, University of California, San Diego, La Jolla, CA 92093
[7]Surgery, University of California, San Francisco, CA 94143
[8]Moores Cancer Center, Biomedical Informatics, University of California, San Diego, La Jolla, CA 92093

Treatment of precancer lesions presents an opportunity to prevent lethal cancers. In breast precancer, ductal carcinoma in situ (DCIS), most DCIS are treated though only a fraction will progress to breast cancer (10-30% at 10 years), with no reliable prognostic molecular biomarkers. Retrospective studies required to develop prognostic biomarkers require archival precancer biopsies which have historically presented a technical challenge. Archival precancer biopsies are often old, small, and damaged by formalin fixation (FFPE), thus their DNA unsuitable for standard targeted sequencing. Here we present an optimized workflow for the targeted sequencing and somatic variant calling from low input FFPE DNA. We tested the protocol on dilutions of 200, 50, 10, and 3ng of low quality FFPE DNA and evaluated variant calling accuracy against 200ng DNA from mirrored frozen tissue. Notable improvements include a blunt-end ligation increasing the fraction of targeted bases covered at 20X from ~15% to 99% from 3ng FFPE DNA as compared to standard protocols. Ensemble variant calling, and heuristic filters enabled us to call clonal somatic variants with 88% recall and 93% sensitivity and sensitively detect an *ERBB2* copy number amplification present in the samples. We evaluated the workflow on 10 laser-capture micro dissected regions from 3 DCIS blocks and demonstrate our ability to distinguish genomic clonal relationships of DCIS regions from different patients. Overall, this enhanced targeted sequencing workflow for low input FFPE DNA unlocks the ability to perform retrospective studies that relate genetic profiles of archived precancer lesions with patient outcome, enabling prognostic biomarker discovery.

Existence of Causation without Correlation in Transcriptional Networks

Gerald M. Pao, Ethan R. Deyle, Hao Ye, Junko Ogawa, Marisela Guaderrama, Manching Ku, Nina Tonnu, Nafeesa Khan, Eugene Ke, Curt Wittenberg,
Inder M. Verma, George Sugihara
Laboratory of Genetics, Salk Institute for Biological Studies,
La Jolla, CA 92037

It is commonly assumed that lack of correlation is evidence for lack of causal relationship. Here we show that in transcriptional networks, causal linkages can exist in the absence of correlation. We find that in well studied S. cerevisiae and in M. musculus cells, a substantial proportion of genes show evidence of nonlinear dynamics in their expression patterns (65-77%). By using a test that accommodates this fact, we uncover uncorrelated but strong causal relationships for the yeast transcriptional regulators *WHI5* and *YHP1*, and verify these relationships experimentally. These genes sit at important checkpoints in the cell cycle where multiple signals are integrated at single nodes, giving rise to causal but uncorrelated relationships. We show that although such links are invisible to correlation-based analyses, they can be uncovered with high accuracy (71-78%) with an empirical dynamic approach based on Takens theory based embeddings. Our analyses are consistent with the notion that the existence of Causation without Correlation is a consequence of multivariate state dependence where allowable states sit on the surface of low dimensional manifolds.

Exploring Gene Regulation and Sex Differences in Lung Adenocarcinoma Using Network Analysis

J. Garrett Parker[1], John Quackenbush[2], Camila Lopes-Ramos[2],
Steven Lawrence[3], and Katherine Cyr[2]
[1]Mathematics, University of California, San Diego, La Jolla, CA 92093
[2]Biostatistics, Harvard School of Public Health, Boston, MA 021115
[3]Biostatistics, Columbia University School of Public Health,
New York, NY 10032

The analysis of distinct, but related data sets can be accomplished using a network. In computational cancer research, there are many insights to be gained from investigating the relations across different data sets. In the case of lung adenocarcinoma, a network of differential expression levels, protein-protein interactions, and phenotypic data can be analyzed to identify highly regulated pathways across healthy and disease populations. There are several existing algorithms, such as PANDA, that can be utilized to construct this type of network. However, additional work is needed to identify useful information about the underlying molecular processes. Our research makes use of statistical methods, bioinformatics software and network analysis tools to investigate sexual dimorphism in lung adenocarcinoma patients. Also, we identify highly enriched pathways in female patients that raise interest for further investigation.

Quantifying ERK Dynamics Required for Oncogene Induced Senescence.

Abhineet Ram
Molecular and Cellular Biology, University of California, Davis, CA 95616

Extracellular signal-regulated kinase (ERK) controls cell proliferation and normal developmental programs in response to extracellular growth factors. Through pathway mutations or persistent extracellular signaling, hyper-activation of ERK leads to uncontrolled proliferation and oncogenesis. However, constitutive activation of the pathway also leads to permanent growth arrest, termed Oncogene Induced Senescence (OIS). Recent work at the single-cell level demonstrates that ERK activity is heterogeneous and has a large dynamic range. By virtue of these varied signaling characteristics, differential activation of ERK can lead to diverse cell fates; these include proliferation, differentiation, or senescence. The pleomorphism of ERK activation supports the idea that the decision between senescence and proliferation is dependent on its specific activation state, possibly dependent on an ERK activity threshold. Importantly, the amount and duration of active ERK needed to induce senescence in a single cell is not precisely defined. My project will quantify this threshold and identify the most important signaling characteristic of ERK activity at the single-cell level. To address these questions, I will use live, single-cell imaging to precisely quantify the dynamics of ERK in cells undergoing senescence.

Redundancy and Fragility in an Essential yet Evolutionarily Novel Character Identity Network in *C. elegans*

Scott Rifkin
Section of Ecology, Behavior and Evolution, Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92037

Although homologs often exhibit deep molecular conservation across huge evolutionary distances, the converse is also common - starkly different genes and development underlying the same character in different species. These findings challenge the concept of homology because they suggest that morphological characters can be historically continuous and homologous without the same being true of the genes and developmental processes that build them. Wagner and others have proposed that a core set of genes - a character identity network - links molecular and morphological homology and allows for downstream variation in genes giving a character its specific appearance and upstream variation in genes determining its specific position.

The intestinal specification network in *C. elegans* appears to be a canonical example of a character identity network underlying an extremely conserved character. The network consists of a short cascade of GATA-type transcription factors that interpret positional information from maternally deposited factors, regulate each other, and activate *elt-2*, which stays on through the worm's life, organizing and controlling intestinal development. This network exhibits redundancy in trans-regulation, but activation of *elt-2* relies on a single, essential GATA-factor binding site amidst an array of conserved but apparently dispensable GATA sites on the *elt-2* promoter. Most surprisingly, the network is an evolutionary novelty within the *Caenorhabditis* genus - all but one of the genes in the core network arose in a short burst of duplication within the genus itself. Even essential developmental processes can be dramatically rewired without any overt change at the phenotypic level.

Organizing Principles in the Transformation of Representations in Ventral Visual Pathway

Ryan Rowekamp and Tatuana Sharpee
Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037

The conversion of visual stimuli from a collection of signals representing the local intensities of light into a compressed set of signals representing the positions, qualities, and identities of objects in the environment. Previous work has shown that artificial convolutional networks are capable of simulating the responses of neurons in the early to middle visual cortex. We applied a two-layer quadratic convolutional model to neurons from macaque V1, V2, and V4. In all three areas, the models outperformed or matched our benchmark gradient boosted tree models. The first and second layers of the models tended to be dominated by either the linear or the quadratic term, so we divided neurons into four classes based on which term dominated which layer. For neurons that were quadratic in the first layer, we fit curved Gabors to characterize their feature selectivity and found that the excitatory and suppressive components tended to be locally orthogonal in V1, V2, and V4 as well as an increase in the relative curvature of these Gabors in V4 relative to V1 and V2. We also compared the dominant excitatory orientations between the first layer and second layer and found that they tended to be parallel with each other in V1, V2, and V4. Finally, we noted that excitation and suppression in neurons that were quadratic in the second layer tended to be aligned.

Systems Modeling Reveals Differences in Downstream Activation for Various GNAQ/GNA11 Pathway Mutations in Uveal Melanoma

Michael Trogdon and Edward Stites
Integrative Biology Laboratory, Salk Institute for Biological Studies,
La Jolla, CA 92037

Uveal melanoma (UM) is a rare but deadly cancer that arises in the melanocytes of the uveal tract in the eye. UM is biologically and genetically distinct from cutaneous melanoma. Mutually exclusive mutations in the GNAQ/11 signaling pathway occur in more than 95% of cases. Recently, progress has been made in characterizing the biochemical properties of mutant Gαq subunits and identifying the downstream targets of this signaling pathway that are critical for tumorigenesis. To elucidate the systems level consequences of oncogenic mutations, mathematical modeling has emerged as a powerful tool. In this work, we present a mechanistic mathematical model of oncogenic GNAQ/GNA11 signaling. Our mathematical model has provided several systems-level, biologically relevant insights into the potential mechanisms of oncogenic activation in UM. First, our model has helped to clarify the mechanisms by which different common point mutations can provide similar oncogenic signaling despite distinct biochemical properties. Second, our model was used to predict major differences in activation of downstream effectors between common GNAQ and CYSLTR2 mutations. While conventional wisdom would posit that mutually exclusive mutations in a particular pathway would provide similar downstream activation, our mechanistic model predicts otherwise. Lastly, we use our model to investigate the potential mechanisms of action for the recently developed inhibitors of Gαq subunits. Specifically, we show that despite being referred to as "constitutively active", the inhibition of mutant Gαq subunits is in fact more consistent with a dynamic equilibrium of active and inactive states, which could have implications for drug development going forward.

Quantitating the Epigenetic Transformation Contributing to Cholesterol Homeostasis Using Gaussian Process

Chao Wang and William Balch
Molecular Medicine, The Scripps Research Institute, La Jolla, CA 92037

To understand the impact of epigenetics on human misfolding disease, we applied Gaussian process regression based machine learning (GPR-ML) to generate population-based matrices describing the spatial covariance (SCV) relationships that link genetic diversity to the process of fitness in the individual in response to histone deacetylases inhibitors (HDACi). Niemann-Pick C1 (NPC1) is a Mendelian disorder caused by >300 variants in the NPC1 gene that disrupt cholesterol homeostasis leading to the onset and progression of neurodegenerative disease. Applying variation spatial profiling (VSP) that was recently developed by us https://www.cell.com/cell-reports/fulltext/S2211-1247(18)31162-8", Wang & Balch, Bridging Genomics to Phenomics at Atomic Resolution through Variation Spatial Profiling, https://www.cell.com/cell-reports/fulltext/S2211-1247(18)31162-8", 2018, Cell Reports), we determined the sequence-to-function-to-structure relationships of the NPC1 polypeptide fold required for membrane trafficking and generation of a tunnel that mediates cholesterol flow in late endosomal/lysosomal (LE/Ly) compartments. HDACi treatment reveals unanticipated epigenomic plasticity in SCV relationships that restore NPC1 functionality. VSP based matrices capture the epigenetic process impacting central dogma, providing an framework for quantifying the effect of the environment on the healthspan of the individual.

Hyperbolic Non-metric Multidimensional Scaling Reveals Intrinsic Geometric Structure in High-dimensional Data

Yuansheng Zhou and Tatyana Sharpee
Computative Neurobiology Laboratory, Salk Institute for Biological Studies,
La Jolla, CA 92037

Modern datasets characterize objects with respect to many variables and assign distances between objects based on a Euclidean metric. However, recent results suggest that for data produced by underlying hierarchical tree-like networks a hyperbolic metric might be more appropriate than a Euclidean one. We develop non-metric multidimensional scaling in hyperbolic space (HMDS) to perform hyperbolic embedding of points. Using simulations we find that the intrinsic Euclidean structure of objects can be fully represented by Euclidean data with the same or higher dimensions; however, the intrinsic hyperbolic structure can only be preserved by very high dimensional Euclidean representation. At the same time, it can be accurately represented using hyperbolic metric. Applying HMDS to human gene expression data, we show that the samples taken from local clusters show Euclidean structure, but samples taken broadly from the whole population show hyperbolic metric. The hyperbolic effects increase with the number of genes that are taken into account when describing differences between cells. We conclude that the human gene expression space is locally Euclidean but globally hyperbolic, and the hyperbolic structure requires high dimensional gene set to represent. Our method can be generalized to study the hidden geometric structures of diverse biological systems.