Day 2
Session 5

Capturing dynamics from multiple data sets to identify rules that govern signal processing from complex input

Michael J. Lee
Program in Systems Biology, University of Massachusetts Medical School Worcester, MA

Many complex and progressive diseases including cancers are generally treated by a combination of multiple drugs. Most of our best drugs function by modulating signaling pathways that control growth, survival, and cell death. These pathways are functionally integrated, very plastic, and incredibly sensitive to environmental context. Furthermore, a major shortcoming remains in our inability to predict how these pathways will respond to combinatorial stimuli, including complex drug cocktails. To begin to understand some of the "rules" governing how these signaling networks process complex input, we performed a screen of pairwise drugs in various *in vitro* models of breast cancer. Using genomic, proteomic, computational, and network-based tools, we focused primarily on the dynamics and adaptive properties that emerged from various drug combinations. This work has uncovered surprising levels of transcriptional plasticity that cancer cells use to evade therapies, a process that others and we now call adaptive resistance. This presentation will focus on our current understanding of this process and how the various data sets and types of data integration computation can help us decipher how signaling networks process complex inputs.

Data analyses and integration in a cell type selective manner

Olga Troyanoskaya
Department of Computer Science & Lewis-Sigler Institute of Integrative Genomics, Princeton University, Princeton, NJ, and Simons Center for Data Analysis, Simons Foundation, New York, NY

An immense molecular complexity forms the foundation of human disease. This complexity must be interpreted and distilled through accurate modeling of molecular networks and pathways whose malfunction promotes the emergence of complex human diseases. Although cell-lineage-specific gene expression and function underlie the development, function, and maintenance of diverse cell types within an organism and are critical to understanding molecular basis of disease, high-throughput data are rarely resolved with respect to specific cell lineages. In this talk, I will focus on our recent work developing integrative approaches that leverage functional genomics data collections to study how cellular pathways function in diverse cell types and identify genes and systems-level changes underlying complex human disease.

Data curation and database resources to enable quantitative systems biology

Kara Dolinski
Lewis-Sigler Institute of Integrative Genomics, Princeton University, Princeton, NJ

To produce simulations and predictions of complex physiological processes that determine human health and disease, a vast amount of data need to be generated, collected, and organized in a computable way. More specifically, a cellular parts list of proteins must be accurately and consistently annotated, along with each protein's biochemical and kinetic properties and participation in biochemical reactions. At a higher level, these proteins need to be organized into their larger, dynamic interaction networks and pathways, along with their context (cell-type and tissue specificity). While much of these data have yet to be generated, a substantial amount of it exists, but it lies in disparate databases, supplemental data files, or embedded in the text of the biomedical literature in forms not amenable to large-scale integration and calculation. These heterogeneous data types need to be systematically curated and collated from the literature and as they are continually produced via high-throughput methods and made publicly available in a robust way to enable the predictive modeling of human health and disease that will help lead to precision medicine. Here, I will present some of the existing resources that might be leveraged toward this end, and I will also present ideas and guide a discussion on what gaps need to be filled.

Session 6

Metabolomics to study enzymatic activity on a genome-wide scale

Theresa Fan
Department of Toxicology and Cancer Biology University of Kentucky College of Medicine, Lexington KY

Expertise in our laboratory integrates metabolomics with functional knockdown of key cellular components to understand quantitative relationships between cell and tissue metabolome and key reprogrammed enzymes in human cancer. This is illustrated with studies in lung cancer, which is a leading cause of cancer death in the United States and elsewhere. This disease is of particular concern in Kentucky, as Kentucky leads U.S. in both lung cancer incidence and fatality. The past decade of research in cancer metabolism reveals the value of exploring human metabolome for the discovery of novel therapeutic and diagnostic biomarkers for human cancers and other diseases. Using 13C6 glucose as tracer and NMR and MS-based stable isotope-resolved metabolomics (SIRM) analysis, we defined metabolic reprogramming in lung tumor tissues in 59 human patients with early stage NSCLC *in vivo*. In particular, we uncovered elevated anaplerotic pyruvate carboxylase (PC) activity in cancerous against paired benign tissues. Proliferating cancer cells require an active Krebs cycle for generating anabolic precursors, in addition to energy production. Diversion of the Krebs cycle intermediates to meet anabolic demands cannot be sustained without anaplerosis. Pyruvate carboxylation is one of the two major anaplerotic pathways that support cancer cell growth; the other involves glutaminolysis initiated by glutaminase (GLS). We then measured the expression of PC and GLS proteins in tumor and benign tissue pairs from 86 human NSCLC patients. PC protein was elevated (median 8-10 fold) in 94% of the tumor tissues, whereas GLS expression did not differ significantly between tumor and benign tissue pairs. We further examined the importance of PC in the growth and survival of NSCLC by employing the shRNA knockdown approach to demonstrate reduced lung cancer cell growth both *in vitro* and *in vivo*. PC knockdown not only attenuated PC-initiated Krebs cycle activity but also glutamine metabolism, as probed by 13C6-glucose and 13C5, 15N2-glutamine tracers coupled with SIRM analysis. These effects, in turn, led to reduced anabolic activity, such as the synthesis of fatty acyl chains of lipids, suggesting that both energy production and anabolic pathways were hindered, and blocking the PC pathway was not compensated by GLS activity. Applying the SIRM approach to *ex vivo* lung tissue slice cultures indicated that blocking PC anaplerosis with anti-cancer Se compounds can elicit massive necrosis with reduced mitotic index in tumor but not in paired benign lung tissues. Together, these results suggest that PC can be a promising therapeutic target for lung cancer. Such approaches can be very valuable in uncovering and deciphering the function of dysregulated cellular components in cells or directly in individual human tissues on a metabolome and genome-wide basis.

Quantitative profiling of lipid metabolities

Robert Murphy
Department of Pharmacology, University of Colorado School of Medicine, Aurora, CO

Lipids constitute a very large family of biomolecules with hydrophilic as well as hydrophobic properties that can be divided into six different groupings found in mammalian systems. These include fatty acyl compounds, glycerolipids, glycerophospholipids, sphingolipids, steroids, and prenol lipids. These molecules play critical roles in mammalian biochemistry, not only as sources of energy and ATP through fatty acid oxidation, but also physical-chemical roles as the major constituents of membranes that compartmentalize cells as well as subcellular vesicles. Also, some lipids are bioactive mediators, including prostaglandins, thromboxanes, leukotrienes, ceramides, phospholipids, and steroids. These diverse roles as well as the complexity with which lipids present themselves in the living cell requires demanding analytical techniques to both qualitatively and quantitatively assess their levels. Mass spectrometry has emerged as the predominate analytical technique largely because of the development of electrospray ionization, which makes analysis of all lipid substances possible. It is also possible to image the location of lipids within tissues and identify them with a high degree of structural specificity. The field of lipidomics has emerged as a powerful tool to understand the complexity of lipid biochemistry, the potential role of bioactive lipid mediators in physiology as well as pathophysiology. In this presentation, I will present approaches to comprehensively and quantitatively profile lipids in different human cell types

Integrative Informatics for cellular metabolites and macromolecules

Shankar Subramaniam
Department of Bioengineering, University of California, San Diego, San Diego, CA

My laboratory has extensive experience in developing and organizing large data sets and developing informatics tools. The Biology Workbench (the first web-based integrative infrastructure for protein and nucleic acid sequence analysis), the Signaling Gateway (the repository of functional states of signaling proteins), the Lipidomics Gateway (the most comprehensive resource for lipids), the Metabolomics Workbench (the NIH Common Fund's metabolomics resource for the scientific community), and more recently, the Alzheimer's Workbench (an integrative infrastructure for multiscale data and analysis for Alzheimer's disease) are examples of infrastructures developed in my laboratory. The metabolome can be regarded as end-point readout of cellular function with RNA and protein components interplaying to regulate, synthesize, and degrade metabolites. In essence, the metabolome by itself, or in conjunction with the transcriptome and proteome, can define the state of a physiological system and longitudinal measurements of these components provide insights into the dynamical changes in the state of the system. Additionally, the availability of absolute standards for thousands of metabolites provides a quantitative framework for measuring changes in metabolites, and this enables a kinetic analysis of a cellular phenotype. In this talk, I will address the challenges in developing infrastructures for cellular components focusing mainly on metabolites, discuss the challenges in quantitative measurements, and finally describe case studies where the integrative analysis of multi-omics data provides dynamical models of cellular function.

Session 7
Crowdsourcing approaches for parameter estimation and building whole cell models

Gustavo Stolovitzky
Functional Genomics and Systems Biology, IBM Research, Thomas J Watson Research Center, Yorktown Heights, NY

Once the gene regulatory network (GRN) of a biological process has been well characterized, predicting its dynamics depends on accurately estimating the parameters governing its biochemical kinetics. It is then essential to choose the most informative experiments and numerical approaches that will help estimate the parameter values. In the context of the Dialogue for Reverse Engineering Assessments and Methods (DREAM), my colleagues and I created an in silico test framework under which participants could apply their algorithms to determine the unknown parameters of different types of proposed GRNs. To perform their parameter estimation, participants could use a limited set of data chosen from in silico experimental assays and perturbations derived from the GRN ODE model. This successful approach was later applied to a much larger set of models, the so-called whole-cell models, that explicitly represent all cellular components at the molecular level and may enable genotype-to-phenotype predictions. Whole-cell models typically have thousands of parameters, many of which are poorly characterized or unknown. New algorithms are needed to estimate these parameters and enable researchers to build increasingly comprehensive models; hence, we asked Challenge participants to identify a subset of parameters of a modified Mycoplasma genitalium whole-cell model given the model's structure and in silico "experimental" data. The results from these DREAM challenges have been very informative regarding the ability to estimate the parameters of a whole range of GRN, and I will describe several lessons in the design and execution of crowdsourced challenges, and how such community efforts could be integrated with large-scale data gathering efforts.

Modeling approaches for whole cell dynamics: Challenges and opportunities

Chad Myers
Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN

One of the goals of systems biology is to build comprehensive, mechanistic models of the cell. After more than a decade of collecting genome-wide data capturing gene expression, protein levels, and physical or genetic interactions among proteins/genes, we have accumulated a wealth of data that should help us make progress towards this goal. As a community, we have made great strides in developing computational methods that leverage these genomic data to accomplish various prediction tasks, e.g. gene function or protein expression/localization prediction. While all of these are of great utility, I would argue that we are still quite far from systematic approaches for building mechanistic models that, for example, allow us to explain or simulate quantitative, dynamic behavior of biological systems. In my opinion, this highlights the fundamental difficulty of the mechanistic modeling problem, both in terms of the computational approaches required as well as in the focused, quantitative data necessary for the inference of such models. From the perspective of a few specific projects we have been involved in, I will highlight a few of the challenges we have encountered in making the transition between statistical and mechanistic models, and discuss opportunities for new research.

Modeling and analysis tools for whole cell modeling

Les Loew
Departments of Cell Biology, Computer Science and Engineering, University of Connecticut, School of Medicine, Farmington, CT

The shape of a cell, the size of subcellular compartments, and the spatial distribution of molecules within the cytoplasm can all control how molecules interact to produce a cellular behavior. This talk describes how these spatial features can be included in mechanistic mathematical models of cell biology. The Virtual Cell (VCell) computational modeling and simulation software is designed for spatial modeling of cellular reaction-diffusion systems. VCell facilitates choices between physical formulations that implicitly or explicitly account for cell geometry and between deterministic vs. stochastic formulations for biochemical reactions. It is also built on top of a database to facilitate sharing of models and reuse of model components. Recently, we have begun to address the special role of molecular aggregates and clusters in cell biology. New approaches for dealing with the combinatorial complexity associated with interactions of multivalent molecules will be described, including studies on signaling to the actin cytoskeleton.

Session 8

A *C. Elegans* pilot project

A.J. Marian Walhout
Program in Systems Biology, University of Massachusetts Medical School, Worcester, MA

The nematode *C. elegans* is a powerful model for the systems-level study of development and homeostasis. Its power comes from its short generation time, transparent body, outstanding genomic and genetic resources, and low cost of experimentation, among others. I will discuss how a pilot project of quantitative dynamics can be constructed for the worm, with a main focus on the integration of metabolic, transcriptomic and interactomic information, and how this relates to anabolic processes during animal growth and fecundity and catabolic processes to generate energy to support cellular and organismal activities. We propose that such a *C. elegans* project will demonstrate the power of a quantitative dynamics program, and that it will provide a blueprint to guide similar studies in more complex systems.

Multiscale modeling: Understanding adaptive immune responses in terms of dynamic molecular networks within each cell

Alexander Hoffmann
Institute for Quantitative and Computational Biosciences (QCBio), UCLA, Los Angeles, CA

Over the past decades, kinetic models have provided important quantitative insights in immunology at two scales: (1) kinetic models of lymphocyte populations have led to insights about the critical interplay between different lymphocyte subsets during chronic and acute infections; and (2) they helped reveal how a wealth of regulatory dynamics are generated in the molecular networks that determine lymphocyte function. In our recent work, we have begun to link these two scales through multiscale modeling that allow us to probe how the dynamics of molecular networks in each cell determine cellular cell fate decisions, and how these contribute to population-level functions of immune responses. I will outline how determination of kinetic parameters of key molecular networks would enable the development of a multiscale model of human immune system that will mechanistically link molecular and physiological scales, and thus, down the line, provide predictive power to interpret clinical presentations in the context of genetic variants and exposure histories, identify drug targets, and inform clinical decision making.

A human iPSC-derived cardiomyocytes cell project: All it takes to make & keep a beat

Ravi Iyengar
Department of Pharmacology and Systems Therapeutics and Systems Biology Center, New York, Icahn School of Medicine at Mount Sinai, New York, NY

Cardiomyocytes in culture can form functional sarcomeres and spontaneously beat, thus representing a cell type where there is a clearly observable phenotype to which cellular components and interactions can be correlated to a physiological function. Induced pluripotent stem cells (iPSCs) from human subjects can be readily differentiated into cardiomyocytes that beat. As the iPSCs can be derived from individual human subjects, the technology currently exists to obtain functionally active cardiomyocytes with well-characterized genomes. As part of the LINCS program, our DToxS Center has created a library of iPSC lines from 20 clinically-characterized healthy human subjects. These cell lines can be grown up in a large-scale manner, and they could potentially serve as the source material for a HQD pilot project. The iPSC-derived cardiomyocytes are primary cells, that once differentiated, do not divide; hence, they could provide a useful contrast to cell lines that divide in culture and thus may have an underlying proliferative program. Among the considerations in using iPSC-derived cardiomyocytes is the level of their differentiation: whether they adequately represent adult cells or more like neonatal or fetal cardiomyocytes. Both the pros and cons of considering iPSC-derived cardiomyocytes for a pilot project will be discussed.