Session 1

Initial overview of Human Quantitative Dynamics: A project to quantify the components of and interactions in a human cell

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

Understanding the relationship between genotype and phenotype is a multistep process that requires deciphering the relationship between the genome and the transcriptome, control of mRNA levels by genomic determinants as well as epigenomic regulators and expression patterns of mRNA in the different cell types. Each of these relationships has been or is being mapped in projects supported by the NIH, such as ENCODE and GTeX. A necessary next step is to identify and quantify the levels and interactions of cellular components including proteins and key small molecules and metabolites that give rise to whole cell behaviors that are characteristic of different cell types. The Human Quantitative Dynamics Project will focus on identifying the components of human cells and mapping their interactions in a precise and quantitative manner. Such a goal will require a complex integrated enterprise that uses multiple types of technologies, some of which currently exist and some that need to be developed, and the engagement of different biomedical and engineering communities. In this talk, I will present a listing of both the opportunities and challenges designing such a project poses, and provide suggestions for potential approaches including mechanisms that can enable collaborations between researchers in high-throughput and focused research communities. I will also present an aspirational view of how the output of such a project would look and how it could drive biomedical research in the coming years.

What can HQD do for basic research focused on understanding complex multiscale processes, such as development and homeostasis?

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

Systems biology studies of development and homeostasis have heavily focused on measurements of the transcriptome, proteome and interactome in model organisms such as yeast, worms, and flies, and to a lesser extent in mammalian tissues and cells. However, quantitative models that describe how the transcriptome and interactome function and interact are lacking. We will discuss the types of data that are required for this next important challenge: to obtain predictive models that capture large-scale systems-level data to obtain deep mechanistic insights.

What can HQD do for precision medicine

Garret FitzGerald
Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Clinical research has traditionally rested on the ability of randomized, double blind, placebo-controlled trials to detect large, average effects of drug efficacy and safety, and, to a lesser extent, data from observational studies claimed to reflect, albeit less precisely, “real life” situations. Presently, it is undergoing a fundamental shift. The rapid spread of electronic health records permit one to seek at scale disease associations which might never have been thought of or detected in the context of an individual’s patient experience. Linked biobanks might be exploited to seek biomarker evidence consistent with mechanistic linkage. If this thesis is supported, attention might then turn to deep phenotyping in small numbers of individuals – human phenomic science – to seek proof of concept before performing more efficiently than hitherto, targeted, refined (e.g., adaptive) prospective trials in pursuit of information more precisely linked to the individual. Such precision medicine strategies require the integration over time of diverse data sets: omics, imaging, physiological outputs like blood pressure and other remote sensing outputs, such as what and when foodstuffs are consumed, activity, and communication patterns, for example. The objective may be to develop predictive paradigms of disease evolution across the lifespan, or of efficacy or hazard of therapeutic interventions. Either way, repeated sampling over time in ways that permit iterative refinement of the models are necessary. In the case of therapeutics, the documentation of actual drug consumption and then the accurate measurement of systemic, and ideally local, drug exposure must be integrated with outputs reflective of drug response. Dynamical modelling can be directly integrated with hypothesis testing in human induced pluripotent stem cells *ex vivo*, and indeed, with data obtained in model systems, further to refine predictive paradigms in clinical research.

Session 2

One-hour proteomes and deep sequencing

Joshua Coon
Department of Chemistry, University of Wisconsin – Madison, Madison, WI

Modern biological and medicinal researchers increasingly rely on technologies for rapid and large-scale the comparative analysis of proteomes. Here, we describe the first identification of ~4,000 yeast proteins, a comprehensive yeast proteome, from just over an hour of analysis time. Next, we discuss the application of this technology to profile over one thousand proteomes and metabolomes for functional annotation of orphan genes. Prospects for extending this technology for ultrafast whole proteome quantification in human systems will also be presented. Finally, we describe technology for near complete mapping of the human proteome. With these methods we have achieved detection over 17,000 proteins in a single cell line and have detected peptides covering over 70% of each protein. Specifically, we recorded mass spectral evidence for 7 of every 10 amino acids in the human proteome. To our knowledge these data represent the deepest and most comprehensive proteomic analysis of the human proteome acquired to date.

The Human Proteome Atlas

Mathias Uhlen
KTH Biotechnology, AlbaNova University Center, Royal Institute of Technology (KTH), Stockholm, Sweden

The Human Protein Atlas (HPA) is a large-scale multidisciplinary project to generate and use antibodies to characterize the protein complement of the human genome. The project, which is part of the international efforts to map the human proteome, has engaged researchers from many different disciplines, and up to date, 13 million images have been annotated, showing the localization of proteins in the human body on a cellular and subcellular level. In the current version of the atlas, more than 24,000 antibodies have been used to analyze the protein-coding genes in a genome-wide manner. The proteomes of tissues representing all major tissues and organs in the human body are presented, containing spatial information down to the single cell level in the various compartments of these tissues. All images and data are publicly available as an open resource at the web portal www.proteinatlas.org. This data has been used to perform global analyses of the human proteome and sub-proteomes. Our current approaches can be used to characterize proteins in individual human cell types, and I will discuss these approaches in the context of the HQD project.

Analysis of network perturbations in human disease

Marc Vidal
Center for Cancer Systems Biology (CCSB), and Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School, Boston, MA

Our goal is to contextualize and functionalize human genetic variants for a wide range of disorders, from simple Mendelian diseases to complex traits, through direct interrogation of macromolecular networks. Along with network analysis of wild type and mutant proteins, we are developing concepts, technologies, and systematic datasets to functionally characterize large numbers of genotypes in terms of the effects they have on the molecular functions and physical and biochemical interactions mediated by the corresponding gene products, with implications for elucidating human gene-gene interactions and understanding heritability and improving genomic medicine. I will discuss how these approaches can be readily adapted to comprehensively map interactions within different human cell types.

The protein database landscape: Achievements and limitations

Rolf Apweiler
EMBL-EBI Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK

It is important to address informatics and database challenges if we want to move towards a detailed and comprehensive quantitative understanding of the components of various cell types and tissues in terms of their concentrations and kinetic interactions. In my talk, I will give my view on the current database landscape, from capturing of raw experimental data of a comprehensive quantitative project rooted in biochemistry and cell physiology to analysis and organization of the data into functionally annotated proteins, complexes and pathways. Finally, I will try to point out limitations in our current database landscape, and in our ability to integrate this knowledge with genomic and epigenomic information, and how to overcome these limitations.

Session 3

Overview

Sarah Dunsmore
NIGMS, National Institute of General Medical Sciences, Bethesda, MD

An overview of trans-NIH funding opportunities for large scale scientific projects will be discussed. Topics will include the NIH Common Fund, NIGMS Glue Grants, and NIGMS Systems Biology Centers. A brief introduction to the other speakers in Session 3 will be provided.

Multiomics projects for precision medicine: Dynamic analysis of healthy and disease states using personal omics profiling

Michael Snyder
Department of Genetics, Stanford University Stanford CA

Understanding health and disease requires a detailed analysis of both our DNA and the molecular events that determine human physiology. We performed an integrated Personal Omics Profiling (iPOP) on 100 healthy and prediabetic human subjects over periods of viral infection as well as during controlled weight gain and loss. Our iPOP integrates multiomics information from the host (genomics, epigenomics, transcriptomics, proteomics and metabolomics) and from the gut microbiome. Longitudinal multiomics profiling reveals extensive dynamic biomolecular changes occur during times of perturbation, and the different perturbations have distinct effects on different biomolecules in terms of the levels and duration of changes that occur. Overall, our results demonstrate a global and system-wide level of biochemical and cellular changes occur during environment exposures. I will discuss both the advantages and challenges of multiomics projects and how our experiences could be relevant for the proposed Human Quantitative Dynamics project.

Studying genetic interaction networks

Brenda J. Andrews
The Donnelly Centre for Cellular and Biomolecular Research, and Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

We generated a comprehensive genetic interaction network for the budding yeast, *Saccharomyces cerevisiae*, constructing 34 million double mutants to map ~450,000 negative and ~250,000 positive genetic interactions. The genetic interaction profile for each gene enabled quantification of functional similarity and the assembly of a hierarchical model of cellular function. Negative genetic interactions connected functionally related genes, highlighted essential genes as network hubs that provide a network scaffold for a set of core bioprocesses, and defined pleiotropic genes that mediate connections between numerous different bioprocesses. Positive interactions, especially among essential genes, appeared to reflect more general regulatory mechanisms. The organizing principles of the yeast genetic network, whereby coherent sets of positive or negative interactions occur within and between genes encoding complexes and pathways, can be exploited to decipher genetic interactions in humans.

The design and challenge the BRAIN Initiative

Andrea Beckel-Mitchener
NIMH, National Institute of Mental Health, Bethesda, MD

Brain Research through Advancing Innovative Neurotechnologies (BRAIN) is a relatively new initiative aimed at revolutionizing our understanding of the human brain. By accelerating the development and application of innovative technologies, researchers will be able to produce a revolutionary new dynamic picture of the brain that, for the first time, shows how individual cells and complex neural circuits interact in both time and space. The NIH is working in close collaboration with other government agencies, the scientific community, as well as private partners to ensure success through investment and cooperation. One component of BRAIN is an effort to generate a census of all cell types in the brain. Significant effort is devoted to understanding what cellular properties define a given cell type, how are these properties measured and, finally, what are the challenges associated with scaling up and harmonizing large data generation efforts

GTEx: Successes and challenges

Kristin Ardlie
Broad Institute, Cambridge, MA

Correlations between genotype and tissue-specific gene expression levels will help identify regions of the genome that influence whether and how much a gene is expressed. The Genotype-Tissue Expression (GTEx) project aims to provide a resource to the scientific community with which to study human gene expression and regulation and its relationship to genetic variation. This project collects and analyzes multiple human tissues from donors who are also densely genotyped, to assess genetic variation within their genomes. By analyzing global RNA expression within individual tissues and treating the expression levels of genes as quantitative traits, variations in gene expression that are highly correlated with genetic variation can be identified as expression quantitative trait loci, or eQTLs. The successes and challenges of this large-scale project will be discussed

Defining scope of big projects to enable success

Ajay Pillai
NHGRI, National Human Genome Research Institute, Bethesda, MD

In my presentation, I will provide an overview of considerations that go into success or failure of large-scale projects based on my experience working on multiple NIH Common Fund projects and other programs within NHGRI. Clearly articulating the goals of the project is crucial: is the project meant to answer specific questions or be a community resource? A sample of the types of questions that I will address include the following: How do you incentivize the research groups to keep overall project goals in mind throughout the period of the project? What specific incentives exist for junior scientists in each group to work collaboratively? Is the project suitable to include R01 scientists in some manner? How do you identify a community and include them in your journey during the project? What happens to the resources after the project is done?

Session 4

Proteomic approaches to identify protein-protein interactions and protein complexes

Anne-Claude Gingras
Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada

We develop experimental technologies and software tools for proteomics, primarily for the analysis of protein-protein interactions. Sample quality is key to interaction proteomics success, and the choice of the expression system, epitope tag, and purification protocol can greatly influence outcome. Our lab has developed robust protocols for the purification of interactors using epitope-tagged approaches. We also implemented the BioID *in vivo* biotinylation approach first introduced by Kyle Roux, which enables us to probe interactions that were previously very difficult to detect (e.g. with membrane-associated proteins). Another important factor in the success of interaction proteomics experiments is the ability to discriminate true interaction partners from background contaminants. Many excellent methods exist that utilize high stringency purification procedures to limit background, or employ isotope-based quantitative proteomics approaches to assist in the identification of true interactors. However, such approaches may not always be feasible or desirable, and we perform single affinity purification coupled to mass spectrometry from unlabeled cells. Since background contaminants are generally more abundant in these cases, we are developing computational tools to allow us to identify true interaction partners. We have developed, with Alexey Nesvizhskii and Hyungwon Choi, the SAINT (Significance Analysis of INTeractome) series of algorithms that provide confidence scores for each detected interaction. The relevance of these approaches to the Human Quantitative Dynamics project will be discussed.

Towards affinity quantification of protein-protein interactomes?

Gilles Travé
UMR 7242 CNRS, École Supérieure de Biotechnologie de Strasbourg, Illkirch, France

Most proteins establish multiple interactions with pools of alternative partners, which compete with each other based on individual binding affinities. Comparing the various affinities of a given protein to the full panel of its competing cellular partners, allows one to assess the binding specificity of this protein to each partner. Until now, high-throughput protein interactomics studies have mainly delivered binary data (“interact” or “not interact”) without quantitative information on affinities and specificities. The development of approaches for systematic determination of protein-protein interaction affinities thus represents a key challenge in systems biology. Unfortunately, most full-length proteins are difficult to produce and purify for binding assays, especially at high-throughput. Yet many protein-protein interactions boil down to interactions between globular domains and short linear sequence motifs. Most globular domains and peptides can be produced at high-throughput for interaction assays. Therefore, one way to tackle the interactome affinity determination problem is to measure domain-motif interactions at high-throughput. My talk will focus on this issue. I will in particular describe the "holdup" method for high-throughput domain-motif affinity determination, which we recently developed and applied to the PDZ domain family involved in various signaling processes (Vincentelli, Luck *et* al., Nature Methods 2015).

Measuring protein–ligand interactions

Anne-Claude Gavin
Structural and Computational Biology Unit, EMBL-Heidelberg, Heidelberg, Germany

Eukaryotic cells use membrane-bound organelles with unique lipid and protein compositions to regulate and spatially organize cellular functions and signaling. As part of this tight control, many proteins are regulated by lipids. In humans, the importance of these regulatory circuits is evident from the variety of disorders arising from altered protein–lipid interactions, which constitute attractive targets for pharmaceutical drug development. However, the full repertoire of interactions remains poorly explored and exploited because their detection is still difficult to achieve on a large and systematic scale. I will present pioneering technologies such as affinity-purification (AP)-lipidomics and the liposome microarray-based assay (LiMA) that will enable protein-lipid interactions to be deciphered on a system level. I will also illustrate the importance of these methods with two pilot studies conducted in yeast and human cell lines, which revealed surprising insights, such as the discovery of a new, conserved family of oxysterol-binding protein (OSBP) with unexpected specificities for phosphatidylserine, an important signaling lipid. The datasets also reveal cooperativity as a key mechanism for membrane recruitment of pleckstrin homology (PH) domains.

Quantitative profiling of protein-protein and drug-target interactions in human cells

Mikko Taipale
The Donnelly Center, and Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

We are interested in systematically and quantitatively profiling protein-protein and drug-target interactions in human cells. Our focus is on the cellular quality control machinery, including major cellular chaperones, such as Hsp90 and Hsp70, and E3 ligases and deubiquitinases that collectively regulate thousands of target proteins in the cell. We are particularly interested in how these quality control pathways impinge on cellular pathways underlying rare Mendelian disorders and tumorigenesis.

New technologies to study cellular quantitative dynamics

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

The cellular and biochemical systems that dictate human phenotypes are large, complex and incompletely understood -- both qualitatively and more so quantitatively. Predicting human quantitative dynamics will require new experimental technologies that provide data on a much more comprehensive scale to sufficiently constrain models of such large and complex systems. This talk will identify the latest techniques that are making strides in this direction. This will include brief descriptions of the different systems-level technologies that can be scaled up to obtain comprehensive characterization of different classes of cellular components in different cell types. I will also highlight areas that need the most technology development focus.

Systematizing the analysis and assembly of executable models

Peter Sorger
Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA

I will discuss recent approaches to the development of executable models of signal transduction (focusing on oncogenesis and drug mechanism of action) in which specialized computer languages are used to establish and maintain a hierarchy of abstractions that effectively capture both mechanistic and network-level representation of biological function. Our programs incorporate existing knowledge and are trained and tested against experimental data. Where uncertainty exists, we create families of alternative models and compare the ensemble to empirical data in a Bayesian framework that accounts for underlying uncertainties in data and assumptions; we also use executable models to optimize experimental design. The resulting reasoning system makes clear the assumptions inherent in different explanations and of necessity liberates the underlying data for re-analysis by others. Because the process from original hypothesis to experimental test to revised hypothesis is codified in open-source code and public data, findings becomes reusable, quantitative, and reproducible.