Principles of Systems Biology, No. 18

This month: imaging the organelle interactome (Lippincott-Schwartz), big data immunology (Pulendran, Ginhoux), protein interactomes expand (Barna, Harper, Marcotte), and computation/engineering insights (Milinkovitch, Silver, and Sastry).

Imaging Informatics Reveals the Organelle Interactome

Alex M. Valm and Sarah Cohen, NIH; and Jennifer Lippincott-Schwartz, HHMI Janelia Research Campus

Principles

It is increasingly understood that membrane contact sites (MCSs) function in the direct transfer of Ca2+ and lipids between subcellular organelles and in organelle biogenesis. Live-cell fluorescence imaging is a powerful technique to answer systems-level questions about the spatial and temporal dynamics of MCSs. However, the inability to distinguish fluorophores with highly overlapping emission spectra has limited its use to the labeling of a few different organelles in live cells.

Our laboratory has recently developed a cell labeling, imaging and computational analysis method to identify the systems-level dynamic organization of organelle interactions in cells (Valm et al., Nature *546*, 162–167). We targeted spectrally variant fluorescent fusion proteins to endoplasmic reticulum, mitochondria, Golgi, lysosomes, and peroxisomes, and used a vital dye to label lipid droplets. Surprisingly, we observed that although individual organelles are highly dynamic, the sum of all organelle contacts in fibroblast cells is stable over time, forming a consistent pattern that we termed the "organelle interactome."

> *"…although individual organelles are highly dynamic, the sum of all organelle contacts in fibroblast cells is stable over time, forming a consistent pattern that we termed the 'organelle interactome.'"*

What's Next?

This approach will be useful to investigate organelle organization in response to various environmental and developmental cues, including changes in the availability of nutrients, exposure to drugs, and infection with pathogens. The application of systems biology approaches to the study of MCSs may reveal new paradigms in organelle dynamics and cellular organization.

Orthogonal Data Integration to Define Immunometabolic-Phenotypes

Shuzhao Li, Bali Pulendran, Emory University, and Stanford University

Principles

Human immunology is witnessing unprecedented progress, driven largely by the advent of omics technologies and single-cell analysis. Experimental perturbation of the immune system by vaccination in controlled longitudinal studies provides a valuable window to understand human immunity, in a field now coined as "systems vaccinology". However, the character of this new "big data immunology" is far removed from the intracellular models from the early days of systems biology, with the latter relying much more on systemic measurements, which necessitate new computational approaches.

Using the live attenuated herpes zoster vaccine as a probe, we measured a broad array of parameters before and after vaccination in a human cohort (Li et al., Cell *169*, 862–877). We used a "multiscale, multifactorial response network" to integrate plasma metabolomics, blood transcriptomics, cytokines, and populations of immune cells. This revealed a striking biological convergence between the pathways identified by metabolomics and transcriptomics. Metabolic phenotypes, such as sterol metabolism and inositol phosphate metabolism, were major predictors of the immunological outcome.

> *"This revealed a striking biological convergence between the pathways identified by metabolomics and transcriptomics."*

What's Next?

Integrative models that combine orthogonal data types will have widespread applications in vaccinology and immunotherapy of cancer, infectious diseases, autoimmune diseases, and inflammation. Such an approach provides new opportunities to modulate the immune responses by redirecting metabolic pathways at both cellular and systemic levels.

Mapping the Human DC Lineage

Peter See and Florent Ginhoux, Singapore Immunology Network (SIgN), A*STAR

Principles

Dendritic cells (DC) are pathogen sensing and professional antigen-presenting cells that control immune responses. They comprise plasmacytoid DC (pDC) and two functionally specialized lineages of conventional DC (cDC1 and cDC2), whose origins and differentiation pathways remain incompletely defined. Recently, we employed an unbiased high dimensional approach to map the human DC lineage (See et al., Science *356*, eaag3009).

We used single-cell mRNA sequencing to map the continuum of DC differentiation from the bone marrow to the blood, thereby allowing the identification of novel populations. Combining single-cell mRNA sequencing with mass cytometry (CyTOF) analyses, we identified a discrete population of dendritic cell precursor (pre-DC) that was previously contained in the pDC gate due to their phenotypic overlap. We also showed that the pre-DC compartment contains functionally and phenotypically distinct lineage-committed subpopulations, including one early uncommitted $CD123^{+}$ pre-DC subset, and two $CD45RA^{+}CD123^{lo}$ lineage-committed subsets. Our study has also revealed a continuous process of differentiation that starts in the bone marrow with common DC progenitors (CDP), which diverges at the point of emergence of pre-DC and pDC potential, and culminates in maturation of both lineages in the blood and tissues.

> *"…we identified a discrete population of dendritic cell precursor (pre-DC) that was previously contained in the pDC gate due to their phenotypic overlap."*

What's Next?

The discovery of pre-DC subsets opens new avenues for therapeutic strategies of cDC subset-specific targeting, as the mechanisms that direct the differentiation of uncommitted pre-DC into cDC subsets or maintain these cells in their initial uncommitted state in health and disease could be better delineated.

Cell Systems Call

A Ribosome RAPsody

Deniz Simsek and Maria Barna, Stanford University

Principles

Ribosomes are emerging as platforms where multiple regulatory pathways intersect to dynamically control gene expression. Are the ~80 ribosomal proteins (RPs) that constitute the mammalian ribosome adequate to orchestrate this regulation? Limited case-by-case examples have suggested that additional proteins can functionally diversify ribosome function, although the comprehensive repertoire of such proteins is lacking.

We established a new ribosome affinity purification methodology in mammalian embryonic stem cells that surprisingly uncovered an additional ~400 ribosome-associated proteins (RAPs) that directly interact with ribosomes independently of mRNAs or nascent peptides (Simsek et al., Cell *169*, 1051–1065). RAPs belong to unanticipated functional categories including critical RNA and protein modifying enzymes as well as cell cycle, cell redox regulation, and energy metabolism proteins. Our studies further highlight that a metazoan-specific post-translational modification, ufmylation, can demarcate and diversify distinct pools of ribosomes. Moreover, we identify a key energy metabolism enzyme that interacts with endoplasmic reticulum (ER)-associated ribosomes and regulates the translation of ER destined mRNAs revealing that RAPs functionally diversify ribosomes within subcellular space.

> *"…uncovered an additional ~400 ribosome-associated proteins (RAPs) that directly interact with ribosomes independently of mRNAs or nascent peptides."*

What's Next?

Different sets of RAPs may exist in different cell types that extend ribosome functional diversity and heterogeneity. Studying RAPs during cellular differentiation and/or upon specific stimuli will reveal how the proteome is shaped in normal and diseased states. Future studies will lead to technologies to map the entire repertoire of RAPs at specific subcellular compartments such as the mitochondria, cell membrane, and ER to understand the mechanisms leading to control of localized translation and ribosome-mediated cellular decisions.

Toward a Proteome-Scale Social Network

Edward L. Huttlin, Steven P. Gygi, and J. Wade Harper, Department of Cell Biology, Harvard Medical School

Principles

Cellular physiology arises from thousands of proteins assembled into complexes, signaling pathways and organelles. Since even minor disruptions can cause dysfunction and disease, a comprehensive model of proteome architecture is required.

Previously, we coupled mass spectrometry with immunoaffinity enrichment of tagged human proteins expressed in HEK293T cells for large-scale interaction mapping (Huttlin et al. Cell *162*, 425–440). Thus, we target interactions formed in intact cells and can enrich even low-abundance interactions. Our latest network, BioPlex 2.0, surveys ~6,000 baits to generate the largest experimentally derived interaction network to date: 56,553 interactions (87% unknown) among 10,961 proteins (Huttlin et al. Nature *545*, 505–509). We continue to profile all ~10,500 bait clones available to us.

BioPlex affords many insights: first, we infer function and localization for uncharacterized proteins through guilt-by-association; second, statistical enrichment suggests associations among structural domains; finally, BioPlex subdivides into ~1,300 interconnected protein communities reflecting biological function, cellular fitness, and disease susceptibility.

> *"BioPlex subdivides into … interconnected protein communities reflecting biological function, cellular fitness, and disease susceptibility."*

What's Next?

While unparalleled in scope, BioPlex remains an incomplete model of the human interactome. In part, this reflects technical limitations inherent to purification of certain protein classes and recovery of low-affinity interactions. Proteome dynamics are a more fundamental challenge. In response, we are also profiling interactions in additional cell lines. Future low- and high-throughput studies of perturbed systems will further probe interaction dynamics. Finally, numerous BioPlex interactions have the potential to drive targeted, hypothesis-driven research.

A Working Draft of Human Protein Complexes

Kevin Drew, Chanjae Lee, Ryan Huizar, Fan Tu, Blake Borgeson, Claire McWhite, Yun Ma, John Wallingford, and Edward Marcotte; University of Texas at Austin

Principles

In order to fully understand molecular causes of disease, we must better understand how proteins act in concert to carry out their functions, but we lack a comprehensive description of human protein complexes. We therefore asked three questions: Can we systematically discover new complexes? Can we discover new components of known complexes? With a better complex map, can we discover new disease genes? Using machine learning, we integrated >9,000 mass spectrometry experiments from three large-scale protein interactome studies and identified >4,600 human protein complexes, many strongly linked to disease (Drew et al., Mol Syst Biol. *13*, 932). We highlight novel complexes functioning in cilia, and thus associated with ciliopathies. We also identify a new member of the intraflagellar transport machinery, ANKRD55, whose perturbation phenocopies known ciliopathy genes, establishing it as a disease candidate.

> *"…identified >4,600 human protein complexes, many strongly linked to disease."*

What's Next?

Although we greatly enhanced the set of protein complexes, our map (hu.MAP; http://proteincomplexes.org) covers only an estimated 10%–40% of predicted human protein interactions. There is great interest in the field to increase coverage of the interactome using high-throughput methods. An ongoing effort to integrate these datasets into more comprehensive complex maps will allow us to better understand the normal functions of human proteins and give mechanistic insights into their failure in disease.

Cell Systems Call

The Lizard that Computed Its Colour Pattern

Michel C. Milinkovitch, University of Geneva

Principles

Skin colour patterns in animals arise from microscopic interactions among colored cells (Kondo and Miura, Science *329*, 1616–1620; Singh, and Nusslein-Volhard, Curr. Biol. *25*, R81–R92) that obey non-linear reaction–diffusion equations discovered decades ago by the great mathematician Alan Turing. We recently showed that the skin color patterning in the ocellated lizard is at odds with that time- and space-continuous mechanism (Manukyan et al., Nature *544*, 173–179). We demonstrated that the skin scales form spatially discrete mesoscopic units that are either green or black and switch color depending on the colors of their neighbors. This corresponds to a cellular automaton, a computational system invented in 1948 by John von Neumann. Using histology, numerical simulations, and mathematical derivation, we identified that skin thickness variation generated by 3D morphogenesis restricts diffusion at the borders of scales and causes the Turing mechanism to transform into the von Neumann cellular automaton, allowing biology-driven research to link, for the first time, the work of these two mathematical giants.

"Our study indicates that cellular automaton are not merely abstract computational systems, but can directly correspond to processes generated … by biological evolution."

What's Next?

Our study indicates that cellular automaton are not merely abstract computational systems, but can directly correspond to processes generated in reptilio (Edelstein-Keshet, Nature *544*,170–171) by biological evolution. I tend to believe that transdisciplinarity stimulates imagination and fosters discovery. Non-classical model species will continue to inform us on yet unknown exciting biological and physical processes generating this complex and diverse living world.

Gut Feelings: Engineered Bacteria as Diagnostics

David T Riglar and Pamela A Silver, Department of Systems Biology, Harvard Medical School and Wyss Institute for Biologically Inspired Engineering

Principles

Engineered bacteria show promise as next-generation clinical diagnostics and therapeutics. For example, colonization by disease-responsive bacteria could achieve non-invasive symptom monitoring for chronic conditions. However, stability of synthetic circuits during extended growth in non-laboratory environments such as the mammalian gut has not been studied.

We built on previous technology from our lab to engineer commensal mouse *E. coli* to remember exposure to a transient inflammatory marker, tetrathionate, which is produced in the gut downstream of reactive oxygen species (Riglar et al., Nat Biotechnol, published online on May 29, 2017. http://doi.org/10.1038/nbt.3879). By recording environmental information from the gut this permitted sensing of inflammatory status in disease models via fecal testing.

Remarkably, the synthetic circuits were compatible with long-term colonization and monitoring; the engineered bacteria demonstrated functional and genetic stability over 6 months (~1,600 generations) of continuous colonization of the mouse gut.

"...the engineered bacteria demonstrated functional and genetic stability over 6 months (~1,600 generations) of continuous colonization of the mouse gut."

What's Next?

Better understanding of tolerable burden from a range of synthetic circuits in bacteria will be critical for future design of safe and effective engineered clinical bacteria. Our study suggests that further investigation of tetrathionate and similar molecules as inflammatory biomarkers across a range of disease models and in humans is warranted.

Machine Learning to Accelerate High-Throughput Protein Expression

Anand Sastry and Liz Brunk, University of California, San Diego

Principles

The Human Protein Atlas maps the cellular localization of thousands of proteins across the human body. This is achieved through high-throughput immunohistochemistry, where >40,000 unique antigen-tagged human protein fragments have been expressed and quantified in *E. coli*, but requires a large number of trial-and-error experiments to express soluble tagged protein.

To reduce the total number of experiments required to build a successful antigen tag, we applied machine learning techniques to probe the influence of biochemical properties, from the mRNA level to structural predictions, on the expression and solubility of the protein fragments (Sastry et al., Bioinformatics, published online on April 7, 2017. http://doi.org/10.1093/bioinformatics/btx207). Four machine learning algorithms were applied to these features, resulting in an ensemble model that outperformed each individual algorithm, with accuracies of 70% and 80% for expression and solubility, respectively. Using these predictions, we could reduce the number of required experiments by 39% to achieve the same coverage of successful protein tags.

"…we applied machine learning techniques to probe the influence of biochemical properties, from the mRNA level to structural predictions, on the final expression and solubility of the protein fragments."

What's Next?

The full pipeline, from calculating biomolecular features to economization of experiments, is available online at http://github.com/SBRG/Protein_ML as a series of user-friendly, scalable IPython notebooks that can be tailored to new datasets and applications. Further features, such as those based on 3D structures for whole proteins, can easily be introduced into the pipeline.