

## Principles of Systems Biology, No. 19

This month: building bacterial organelles (Kerfeld, Tullman-Ercek), the widening scope of single cell analysis (Takebe, Glass, Rabadan), insights from yeast (Swain, Dunham), and more (Bloom, Romesberg, Gibson).

### Structure of the Protein Membrane of a Bacterial Organelle

Cheryl A. Kerfeld and Markus Sutter, Michigan State University and Lawrence Berkeley National Laboratory

#### Principles

Bacterial microcompartments (BMCs) are organelles composed entirely of protein. BMCs are defined by a selectively permeable protein shell that allows the diffusion of metabolites required by the encapsulated enzymes. While the shell is highly conserved, BMCs are functionally diverse, catalyzing a range of reactions. The widespread distribution of BMCs across bacterial phyla highlights the adaptive value of metabolic compartmentalization. In contrast to their eukaryotic counterparts, which are delimited by lipids with integrated proteins for selective transport, the proteins composing BMC shells simultaneously fulfill both the role of barrier and of conduit as the interface with the rest of cytosolic metabolism.

We determined the structure of an intact 6.5 MDa BMC shell composed of five different types of shell proteins with a diameter of 42 nm (Sutter et al., *Science* 356, 1293–1297). Besides the overall arrangement of subunits in the intact organelle and their relative sidedness, we identify the specific amino acid interactions between the protein subunits mediating shell self-assembly.

*Our structure also provides a blueprint for engineering synthetic organelles...*

#### What's Next?

Given the conservation of the basic shell architecture across the bacterial kingdom, these principles apply to all functionally diverse BMCs and are essential to modeling flux across their shells. Our structure also provides a blueprint for engineering synthetic organelles as well as for the design of therapeutics that can selectively disrupt shell assembly and function in pathogenic microbes.

### Mix and Match Proteins for Custom Organelles in Bacteria

Marilyn Slininger Lee, University of California, Berkeley; and Danielle Tullman-Ercek, Northwestern University

#### Principles

Bacterial microcompartments (MCPs) are protein cages that natively surround metabolic enzymes. Recently, researchers have sought to engineer MCPs to form custom nano-bioreactors, in which non-native pathways may be encapsulated to enhance function. Toward this goal, we tested methods to alter the shell of the 1,2-propanediol utilization (Pdu) MCP by substituting shell proteins from a different MCP system (Slininger Lee et al., *ACS Synth. Biol.*, published online June 6, 2017. <http://dx.doi.org/10.1021/acssynbio.7b00042>).

We found that complementing the deletion of the shell protein PduA with the shell protein EutM from the ethanolamine utilization (Eut) MCP allows the formation of a complete MCP shell in *S. enterica*. Further, we found that the function of the natively encapsulated pathway is affected by this substitution. We hypothesized that pores formed by the shell proteins were responsible for regulating diffusion of small molecule metabolites across the shell. A library of mutations to a pore-forming residue of PduA resulted in a range of pathway behaviors, revealing a handle for tuning metabolite diffusion in bacteria.

*...mutations to a pore-forming residue of PduA resulted in a range of pathway behaviors, revealing a handle for tuning metabolite diffusion in bacteria.*

#### What's Next?

We are now interested in expanding this method to substitute other shell proteins from diverse MCP systems and testing the effect of altered shell pores on non-native enzyme cargos. This work will aid design strategies for custom organelles.

### Human Liver Bud Organoid Models Development

Takanori Takebe, Cincinnati Children's Hospital Medical Center; Keisuke Sekine, Yokohama City University; and Gray Camp and Barbara Treutlein, Max Planck Institute for Evolutionary Anthropology

#### Principles

Despite recent progress in understanding and use of organoids, it remains unclear how the presence of various cells affects the identity and differentiation of each cell lineage and how the maturation status of the organoid compares to developing human tissue.

To address these questions, we used single-cell RNA sequencing to reconstruct hepatic differentiation and analyze how cells change during 3D liver bud organoid development (Camp et al., *Nature* 546, 533–538). When compared to homogeneous 2D cultures, we find that the cells within the 3D organoid have molecular signatures that very closely resemble those found in naturally developing human liver. Among various signaling interactions uniquely activated in the organoids, we identify molecular crosstalk between VEGF (vascular endothelial growth factor) and KDR (kinase insert domain receptor) that instructs endothelial network formation and hepatic maturation within the organoid similar to natural hepatogenesis. Our molecular dissection reveals interlineage communication regulating organoid development and illuminates previously inaccessible aspects of human liver development.

*Our molecular dissection [of 3D liver bud organoids]... illuminates previously inaccessible aspects of human liver development.*

#### What's Next?

Hepatic cells within the liver bud did not fully mature into metabolically active adult hepatocytes. Thus, the differentiation of hepatic and other component cell lineages requires additional fine-tuning before liver bud organoids can be tested in clinical trials.

### **Epigenomic Recipe of Human Microglia**

David Gosselin and Christopher Glass, University of California, San Diego

#### **Principles**

Microglia are the macrophages of the brain and participate in its development, homeostasis, and defense against pathogens and injuries. Genetic evidence suggests that they are critically involved in neurodegenerative diseases, including Alzheimer's disease. However, most knowledge of human microglia is extrapolated from studies using animal models, and this severely limits understanding of human brain functions and pathologies.

To address this, we characterized the transcriptional profile of human microglia, using RNA-seq, and its underlying epigenomic regulatory network, using ChIP-seq and ATAC-seq (Gosselin et al., *Science* 356, eaal3222). Results showed that human microglia express robustly hundreds of genes that are either modulated in various brain pathologies or are linked to DNA variants associated with brain pathologies. Comparative analysis also indicated that the transcriptional profile of human microglia correlated favorably with that of mouse microglia. Notably, similar transcription factors specify the microglia transcriptional profiles in both species. Finally, transition of microglia to a tissue culture environment demonstrated that the maintenance of the cellular identity of human microglia requires constant exposure to factors present in the human brain.

*...human microglia express robustly hundreds of genes that are either modulated in various brain pathologies or are linked to DNA variants associated with brain pathologies.*

#### **What's Next?**

We are now investigating how DNA variants affect the epigenomic architecture, transcription factor binding, and transcription in human microglia. We are also working to define environment-specific factors in the brain that determine microglia identity.

### **Tracking Continuous Transcriptional Responses in Single Cells Using Topological Data Analysis**

Abbas Rizvi, Pablo Camara, Elena Kandror, Tom Maniatis, and Raul Rabadan, Columbia University

#### **Principles**

The advent of single-cell transcriptomics has resulted in the discovery of new cell types and their relationship to progenitor states. The recovery of developmental trajectories requires reconstructing transcriptional states associated with cellular fate. Absent developmental synchrony, homogeneity, linearity, and discretion, dimensional embedding schemes often lack the ability to capture cellular transition states, while providing statistical rigor. We have effectively addressed this problem by implementing topological data analysis (scTDA) to analyze large-scale single-cell transcription datasets (Rizvi et al., *Nature Biotechnology* 35, 551–560). This mathematical approach provides the means to examine cellular development with unprecedented resolution. We have uncovered an unexpected expansion of key transcriptional events and pivotal coding and noncoding regulators associated with the *in vitro* differentiation of embryonic stem cells into motor neurons. We have concluded that topological data analysis is an effective strategy to reconstruct developmental trajectories resulting from a continuum of differentiating single cells.

*...dimensional embedding schemes often lack the ability to capture cellular transition states, while providing statistical rigor. We have effectively addressed this problem by implementing topological data analysis to analyze large-scale single-cell transcription datasets.*

#### **What's Next?**

Single-cell transcriptomics and topological data analysis are not limited to developmental systems biology. Tracking heterogeneous cellular responses occurring within fully differentiated cellular systems has applicability to the study of mammalian diseases such as cancer and neurodegeneration. Our work may provide a direct path to a systems-level dissection of such complex biological processes.

### **Distributing Intracellular Tasks to Survive Stress**

Alejandro A. Granados, Imperial College London and University of Edinburgh; and Peter S. Swain, University of Edinburgh

#### **Principles**

An organism's response to environmental change has evolved to balance different trade-offs. For example, a higher accuracy of detecting change can compromise the speed of the response. Such trade-offs become heightened when organisms are under stress, and it is unknown how single cells contend with these constraints and the extent to which a failure to do so impacts fitness.

We used dynamic inputs, microfluidic assays, and mathematical modeling to show that a speed-accuracy trade-off is mitigated in the hyperosmotic response of budding yeast (Granados et al., *eLife* 6, e21415). Defining speed by the network's response time and accuracy by the correlation between the "perceived" stress (the network's response) and the actual stress (the change in cellular volume), we show that cells assign the requirements of high speed and high accuracy to two distinct input pathways in the corresponding signaling network and by doing so increase their probability of survival in stress. Indeed, mutants with only the less accurate, reflex-like pathway are fitter in sudden stress, and mutants with only the slow, more accurate pathway are fitter in gradually increasing but fluctuating stress.

*...cells assign the requirements of high speed and high accuracy to two distinct input pathways...and...so increase...survival in stress.*

#### **What's Next?**

Our results demonstrate that trade-offs faced by organisms can be key to understanding the biochemical structure of signal transduction and the importance of dynamic inputs for revealing fundamental principles of cellular behavior.

**Somatic Homolog Pairing in Yeast**

Seungsoo Kim, Jay Shendure, and Maitreya J. Dunham, University of Washington

**Principles**

In many eukaryotic organisms, each cell contains two copies of its genome, which must pair together during meiosis. In *Drosophila*, homologous chromosomes preferentially associate even in somatic cells, allowing *trans* gene regulation, or transvection. However, it remains unclear whether somatic homolog pairing occurs in other organisms, even in the well-studied model, budding yeast.

We investigated the three-dimensional conformation of the diploid yeast genome using a combination of high-throughput chromosome conformation capture on diverged hybrid yeasts and live-cell imaging (Kim et al., eLife 6, e23623). Overall, homologous loci interact more frequently than expected by chance, but the strength of homolog pairing varies widely across the genome. The strongest pairing occurs between the chromosome XII homologs, primarily due to their rDNA arrays, which form the nucleolus. In addition, the homologous copies of the gene *TDA1* move to the nuclear periphery and pair upon activation in a condition-specific manner, partly through interactions with nuclear pore complexes. We conclude that somatic homolog pairing is dynamic and occurs through multiple mechanisms.

*We conclude that somatic homolog pairing is dynamic and occurs through multiple mechanisms.*

**What's Next?**

It remains unknown how prevalent somatic homolog pairing is in other eukaryotic species and how substantial of a role it plays in genome functions like transcription and DNA repair. A more detailed understanding of the underlying mechanisms may guide future studies to perturb somatic homolog pairing and thus directly assess its functional consequences.

**Looking Deep inside Influenza Infections**

Katherine S. Xue and Jesse D. Bloom, University of Washington and Fred Hutchinson Cancer Research Center

**Principles**

Influenza virus's rapid global evolution begins with de novo mutations that arise within individual patient infections, but few variants transmit between hosts and eventually spread on a global scale. Selection, hitchhiking, and transmission bottlenecks all affect how within-host genetic diversity is transformed into global variation.

We tracked influenza's evolution within humans by deep-sequencing longitudinal viral samples from four immunocompromised patients with lengthy infections (Xue et al., eLife 6, e26875). New mutations often arose and fluctuated in frequency, rarely fixing. Multiple mutant lineages competed with one another throughout the infection, providing a vivid clinical example of clonal interference.

Surprisingly, certain mutations arose independently in multiple patient infections. Most of these recurrent mutations were in hemagglutinin, the viral surface protein that is the main target of immune selection. Many within-host mutations in hemagglutinin also arose at sites of global viral variation. These similarities may reflect concordant antigenic selection at the within-host and global scales.

*...within-host diversity may act as a noisy early measurement of global viral evolution.*

**What's Next?**

Many influenza samples are deep-sequenced each year as part of routine surveillance. Viral variation in these samples may provide a glimpse of future evolutionary trends. Our results suggest that within-host diversity may act as a noisy early measurement of global viral evolution.

**Synthetic Parts for Synthetic Life**

Sydney E. Morris and Floyd E. Romesberg, The Scripps Research Institute

**Principles**

Synthetic biology aims to develop optimized and generalized "parts" that may be deployed in cells to impart them with novel attributes. Toward this goal, we have developed two synthetic nucleotides that form an unnatural base pair (UBP) that can be propagated in the DNA of *E. coli*. However, the extent of "optimization" (efficient polymerase recognition) and "standardization" (sequence-independent recognition) of the nucleoside triphosphate parts remained to be determined. Thus, we characterized the efficiency with which the Klenow fragment of *E. coli* DNA polymerase I synthesizes the UBP and its mispairs in a variety of sequence contexts (Morris et al., ACS Synth. Biol., published online on June 27, 2017. <http://dx.doi.org/10.1021/acssynbio.7b00115>).

Apart from one sequence, under steady-state conditions, UBP synthesis was faster than polymerase conformational changes, while the synthesis of every mispair was slower. This demonstrates that the synthetic triphosphate parts are recognized by the polymerase as natural-like, at least under the experimental conditions employed, and thus they are well optimized and standardized for this step of UBP replication.

*...synthetic triphosphate parts are recognized by the polymerase as natural-like...*

**What's Next?**

UBP replication involves synthesis, explored in the current work, as well as extension, which corresponds to the addition of the next nucleotide. A similar characterization of UBP extension is underway.

### **Printing Biologics on Demand**

Krishna Kannan, Kent S. Boles, and Daniel G. Gibson, Synthetic Genomics, Inc.

#### **Principles**

Distributed manufacturing of biologics could enable precision medicine-based therapies to scale and allow a rapid response to disease outbreaks. With this in mind, we developed a prototypical digital-to-biological converter (DBC), an essential first step to bringing on-demand and distributed manufacturing in biology to fruition (Boles et al., Nature Biotechnology, published online on May 29, 2017. <http://dx.doi.org/10.1038/nbt.3859>).

The DBC constitutes fully integrated instruments and underlying software that are coordinated to perform a series of tasks without human intervention. The DBC is preloaded with reagents required to make a designated biopolymer from digitally transmitted DNA sequences. Overlapping oligonucleotide sequences are chemically synthesized then assembled and processed to build a high-quality synthetic DNA amplicon, which may be further converted into RNA, proteins, or viral particles by user-designated instructions. We generated influenza virus genes, antibody polypeptides, an RNA vaccine, and  $\Phi$ X174 viral particles on the DBC.

*...we developed a prototypical digital-to-biological converter (DBC), an essential first step to bringing on-demand and distributed manufacturing in biology to fruition.*

#### **What's Next?**

By integrating novel DNA-synthesis technology and quality-control methodologies, the DBC prototype will be reduced to a portable machine with greatly diminished footprint that can reliably manufacture biological materials. Increased portability and reliability will support successful deployment of the instrument in laboratory, manufacturing, and clinical settings, thus accelerating the pace of research and development of new and custom therapeutics on demand, at a designated location.