CellPress

Principles of Systems Biology, No. 28

This month: new principles for engineering cells (Avalos/Toettcher, Li, Wang, Ellis/Stan), the daily rhythms of gene expression in a primate (Cooper/Panda), structure of the nuclear pore by an integrative approach (Rout/Akey/Sali), and scoring Mendelian disease risk using electronic medical records (Denny).

A Bright Future for Metabolic Engineering

José L. Avalos and Jared E. Toettcher, Princeton University.

Principles

Applying optogenetics to metabolic engineering is attractive because of the exquisite control that light could provide over engineered metabolic pathways. However, light penetration in fermentations is a real concern that has inhibited efforts to combine the two fields.

In a recent study (Zhao et.al., Nature, *555*, 683–687), we have taken an important step in solving this challenge. Using a highly light-sensitive transcription factor, EL222 from *E. litoralis*, we induced robust gene expression in high-cell-density yeast cultures (OD_{600} 50) in lab-scale reactors of up to 5 l. Furthermore, we developed optogenetic circuits that invert the transcriptional response to light, making it possible to strongly induce genes at any cell density by shifting from light to darkness.

These optogenetic tools—OptoEXP and OptoINVRT—can be combined to significantly improve isobutanol production in yeast. Under blue light, OptoEXP induces ethanol biosynthesis, an essential pathway that competes with many engineered metabolisms. Shifting to darkness induces a "production phase" where OptoINVRT drives isobutanol production. Remarkably, triggering additional bouts of ethanol production with brief light pulses during the production phase extends culture viability to further increase isobutanol yields. Such controlled sequences of light and darkness may provide completely new strategies to operate bioreactors. More broadly, our work begins to tear down the wall between optogenetics and metabolic engineering, letting some light shine through.

...controlled sequences of light and darkness may provide completely new strategies to operate bioreactors.

What's Next?

One could expand the multicolor optogenetic toolbox available for metabolic engineering or develop new strategies to optimize light and dark scheduling for further efficiency increases.

Enzyme Pathways Call for Precise Recipes

Jean-Benoît Lalanne and Gene-Wei Li, Massachusetts Institute of Technology

Principles

A fundamental challenge of systems biology is to combine individual, well-characterized, molecular players into a functional cell. Even a mundane-looking problem of mixing a few proteins to form a pathway can quickly become intractable due to the large combinatorial space of possible protein abundances. Is there a uniquely preferred enzyme stoichiometry inherent to each pathway? Or is there a plethora of solutions that are either degenerate or specific to organisms/conditions?

We answered these questions by combining high-precision expression-profiling techniques with large-scale analyses of enzyme pathways across divergent species (Lalanne et al., Cell, published online March 29, 2018, https://doi.org/10.1016/j.cell.2018.03.007). Using a new approach (Rend-seq) to reveal mRNA isoforms in bacteria, we found that many pathway-encoding operons have dramatically changed their transcript architectures in different lineages. However, at the level of protein synthesis, compensatory evolution between transcription and translation ensured that distant bacteria maintain conserved pathway-specific stoichiometry. Therefore, well-proportioned protein abundances are integral design features of biological pathways—a principle that transcends differences in evolutionary history, regulation, and protein sequence.

...well-proportioned protein abundances are integral design features of biological pathways—a principle that transcends differences in evolutionary history, regulation, and protein sequence.

What's Next?

Even for proteins with completely characterized molecular functions, we lack frameworks to predict what constitutes appropriate abundances and how susceptible cells are to expression perturbations. Our results broadly enable the discovery of optimization principles for biological functions, which will provide insights into engineering of cellular functions and elucidation of disease physiology due to proteome dysregulation.

Expanding the Regulatory Parts Toolbox in Diverse Microbes

Antonio L.C. Gomes, Nathan I. Johns, and Harris H. Wang, Department of Systems Biology, Columbia University

Principles

The ability to build genetic circuits with new and more sophisticated functionality requires well-characterized regulatory components, such as promoter and translation initiation sequences. However, the traditional synthetic biology toolbox has only a limited set of regulatory parts that are quantitatively characterized, especially for microbes beyond the model *Escherichia coli* bacteria.

In our recent work (Johns et al., Nature Methods, published online March 19, 2018. https://dx.doi.org/nmeth.4633), we describe a community resource of almost 30,000 regulatory sequences with transcriptomic and protein expression measurements in diverse microbes of biomedical and bio-industrial relevance. These sequences were mined from 184 microbial genomes and exhibited a wide range of gene expression activities and species-selective capacities in different bacterial hosts and growth conditions. This rich dataset enabled quantitative gene expression modeling. We further built gene circuits with unique expression profiles in different bacteria, thus paving a new path to engineer polymicrobial communities.

The use of well-characterized species-selective regulatory components enables programming of unique gene expression profiles across different microbes.

What's Next?

The use of well-characterized species-selective regulatory components enables programming of unique gene expression profiles across different microbes. Exploiting differences in microbial regulatory specificities will lead to genetic systems that propagate across microbial populations but only activate in designated target species in the community. This approach could lead to new applications to modify complex microbiomes.

Cell Systems Call

Improving Cell Engineering by Asking the Cell for Some Feedback

Tom Ellis and Guy-Bart Stan, Imperial College London

Principles

Whether you are engineering metabolic pathways or genetic circuits or producing proteins, expressing genes is a cost to the cell, using up its resources and machinery. This burden is a fundamental barrier for synthetic biology and forces those engineering cells to use design optimization to strike the right balance between synthetic expression and host-cell growth. Yet because burden is so fundamental, we reasoned that cells would naturally have evolved ways to sense it and that optimization could be made automatic if this cell's response could be fed to directly control the expression of synthetic gene constructs.

We performed RNA sequencing on strains of *E. coli* expressing different types of synthetic constructs and used the data to identify how the cell responds to burden. We found a set of chaperone promoters upregulated universally by burden (Ceroni et al., Nature Methods, published online March 26, 2018. https://dx.doi.org/10.1038/nmeth.4635). Using one of these, we built a general burden feedback controller to repress construct expression when it causes too much burden. This now automatically strikes a balance between expression and growth, and through dCas9 technology, it can be rationally tuned and quickly retargeted to control any construct.

…we built a general burden feedback controller…[that] automatically strikes a balance between expression and growth.

What's Next?

We're now exploring and improving the feedback dynamics to see how it acts in fluctuating environments. As this approach can be applied to automatically control the burden for any construct independent of the pathway/circuit being engineered, we are sharing it via http://www.addgene.org/ and looking forward to seeing what other constructs and cells it can help optimize.

Timing Is Everything (and Everywhere) in the Primate

Ludovic S. Mure and Satchidananda Panda, Salk Institute, La Jolla; Howard M. Cooper, INSERM, Lyon

Principles

Coordinated rhythms in gene expression and function are necessary for normal physiology, metabolism, and behavior. Disruption of these rhythms contributes to a range of chronic diseases. While our understanding of the genetic and genomic basis of circadian rhythms in mammals has progressed spectacularly in the last few decades, research has been based almost exclusively on a handful of tissues in nocturnal rodents.

In a recent study (Mure et al., Science, *359*, eaao0318), we undertook an analysis of the circadian transcriptome of major tissues of the diurnal primate using polyA+ RNA sequencing. Our dataset of 768 samples from 64 different tissues and organs sampled across the complete 24-hr day-night cycle is the most extensive transcriptional map and circadian transcriptome yet compiled for any terrestrial vertebrate.

This spatiotemporal atlas unveiled the unique features of the primate rhythmic gene expression. We found that the primate genome is highly rhythmic, with >80% of protein-coding genes showing daily rhythms in expression. In addition, tissue-specific gene expression, the repertoire, and phases of expression of cycling genes impart another layer of functional specialization. Ubiquitously expressed genes that participate in essential cellular functions are more likely to exhibit rhythmic expression in a tissue-specific manner.

...the primate genome is highly rhythmic, with >80% of protein-coding genes showing daily rhythms in expression.

What's Next?

This database will help identify novel potential drug-target pathways that display diurnal oscillation in target organs and may temporally optimize drug administration protocols. Such an atlas will also serve as a reference for future work on understanding how daily lifestyle factors shape human physiology, behavior, and metabolism.

A Hole New Picture

Javier Fernandez-Martinez and Michael P. Rout, The Rockefeller University; Christopher W. Akey, Boston University; Seung Joong Kim and Andrej Sali, University of California, San Francisco.

Principles

The nuclear pore complex (NPC) is a massive, evolutionarily conserved macromolecular machine that is the sole mediator of transport between the nucleus and the cytoplasm and provides a platform for the coordination of many essential cellular processes. Consequently, defects which affect NPC function are directly linked to human diseases.

To understand the function of the NPC, we determined its structure at subnanometer precision by an integrative approach (Kim et al., Nature *555*, 475–482), relying on data from quantitative proteomics, *in vivo* imaging, charge detection mass spectrometry, cryoelectron tomography, and extensive cross-linking with mass spectrometry. The NPC architecture resembles a suspension bridge in which rigid supporting columns are firmly anchored to a substrate while flexible suspension cables connect the columns and throughway, providing strength and resilience. We also describe the detailed organization of the transport machinery, rationalize how the spatial distribution of intrinsically disordered NPC components impacts their role in transport, and provide novel insights into the evolutionary origin of the NPC.

The NPC architecture resembles a suspension bridge in which rigid supporting columns are firmly anchored to a substrate while flexible suspension cables connect the columns and throughway, providing strength and resilience.

What's Next?

The published structure provides a roadmap to advance our understanding of NPC physiology and nuclear transport, while the integrative approach can be used to test how factors such as stress, viruses, drugs, or mutations modify the NPC structure to cause disease.

Cell Systems Call

Phenotype Risk Scores Find Mendelian Diseases Using Their Features

Lisa Bastarache and Joshua C. Denny, Vanderbilt University Medical Center

Principles

Mendelian diseases are generally caused by rare variants with large effect sizes and are often characterized by many phenotypes across different organ systems. Most genetic associations studies analyze individual phenotypes in isolation, potentially missing subpopulations with multiple phenotypes that share a common cause. We developed the phenotype risk score (PheRS), an approach to aggregate phenotypes based on clinical descriptions of Mendelian diseases (Bastarache et al., *Science*, 359, 1233–1239). The phenotypes are mined from electronic health records, making this a high-throughput approach. The method produces a score based on the aggregation of phenotypes and weighted by their rarity. In a validation study, PheRS effectively identified five known Mendelian diseases.

We used PheRS to search for genetic associations of phenotype patterns for 1,204 Mendelian diseases and 6,188 rare variants (frequency <1%) using a cohort of 21,701 individuals with exome array genotyping. We found 18 rare variant associations with Mendelian disease PheRS, 14 of which were novel. None of the 767 individuals with novel Mendelian variants were diagnosed with a genetic disease, despite some having severe manifestations of disease (e.g., organ transplants or cancer).

PheRS evaluates the pathogenicity of a variant and may also aid in identifying individuals with undiagnosed genetic diseases.

What's Next?

PheRS is applicable to any electronic health record linked to genetic information. PheRS evaluates the phenotypic impact of a variant and may also aid in identifying individuals with undiagnosed genetic diseases.