CellPress

Principles of Systems Biology, No. 13

This month: CRISPR flexes its massively parallel muscles (see also this month's editorial), two systems-level properties discovered in yeast, and a host of new tools and synthetic engineered systems.

Massively Parallel Single-Cell Writing

Andrew Garst, Tanya Lipscomb, and Ryan Gill, Muse Biotechnology, Ryan Gill, University of Colorado Boulder

Principles

CRISPR-based genome editing has transformed the biological sciences by enabling rapid genotype-phenotype mapping at unprecedented efficiencies. Further application of CRISPR to genome refactoring has been limited by the inability to multiplex at different genomic loci. We recently described CRISPR-enabled trackable genome engineering (CREATE), a method for multiplexing CRISPR editing at genome scale using trackable barcodes (Garst et al., Nat. Biotechnol., published online December 12, 2016. http://dx.doi.org/10.1038/nbt.3718). In CREATE, each guide RNA is associated with a homologous repair cassette that also functions as a barcode, allowing the phenotypes of individual cells to be linked to genetic modifications in pooled experiments.

> *"…a method for multiplexing CRISPR editing at genome scale using trackable barcodes."*

What's Next?

The CREATE technology is being commercialized by Muse Bio (www.musebio.com) in the form of an automated benchtop instrument for single-cell writing. Muse bio combines a powerful bioinformatics platform, novel genetic chemistries, and automation to bring massively parallel single-cell editing for designer proteins, pathways, or genomes to benchtops everywhere. CREATE should enable new approaches to studying genotype-phenotype associations by directly incorporating structural and biophysical hypotheses into the experimental design at genome scale.

Genome-Scale Screening for Functional lncRNAs

S. John Liu, Max A. Horlbeck, Jonathan S. Weissman, and Daniel A. Lim, University of California, San Francisco

Principles

High-throughput sequencing efforts have uncovered thousands of RNA transcripts in the human genome that do not code for proteins. A small number of these long non-coding RNAs (lncRNAs) have been found to play critical roles in biology, but the vast majority have not been functionally tested. Since it is not currently possible to predict which lncRNAs are functional, large-scale systematic approaches are required.

In Liu and Horlbeck et al. (Science, published online January 6, 2017. http://dx.doi.org/10.1126/science.aah7111), we developed a genome-scale screening platform based on CRISPR-mediated interference (CRISPRi), which uses a catalytically inactive (d)Cas9 fused to a KRAB repressor domain, to inhibit lncRNA transcription. We designed a CRISPRi Non-Coding Library (CRiNCL), consisting of ~170,000 sgRNAs, to target 16,401 lncRNA loci. Pooled screening across seven cancer and non-cancer cell types revealed 499 lncRNA loci that were required for robust cell growth and proliferation. Surprisingly, 89% of the functional lncRNAs modified growth in just one cell type tested, highlighting the importance of cellular context for lncRNA function.

> *"…89% of the functional lncRNAs modified growth in just one cell type tested…."*

What's Next?

Understanding the mechanisms underlying lncRNAs—and the dependence of function on cellular context—will require investigations of lncRNA subcellular localization, protein binding partners, and potential function of the lncRNA locus itself. In addition, the screening platform developed will also be used to study the roles of lncRNAs in higher-order cellular processes such as stem cell differentiation and responses to therapeutics.

Writing and Reading Cellular MEMOIRs with CRISPR and seqFISH

Long Cai and Michael Elowitz, California Institute of Technology

Principles

Mapping lineages and understanding cell fate decisions in developing organisms have been long standing goals in biology. Landmark work by Sulston and Horvitz elucidated the deterministic lineage maps of nematodes by direct observation. However, approaches that would scale to more complex, non-deterministic and non-transparent organisms, and that could access molecular event histories in individual lineages have been lacking.

We developed MEMOIR (memory by engineered mutagenesis with in situ readout) to address this challenge (Frieda et al., Nature *541*, 38–39). MEMOIR uses CRISPR/Cas9 to record lineage and gene expression information in the genome in a format that can be read out in situ by multiplexed, single-molecule, sequential fluorescence in situ hybridization (seqFISH). We demonstrated proof-of-principle that MEMOIR can record and recover lineage information in mouse embryonic stem cells. Beyond lineage, MEMOIR can in principle enable recording of molecular event histories, telling us when cells received specific signals or activated certain transcription factors in the past. Remarkably, this information can be inferred from end-point measurements of single cells within their spatial tissue context.

> *"MEMOIR uses CRISPR/Cas9 to record lineage and gene expression information in the genome in a format that can be read out in situ…by seqFISH."*

What's Next?

MEMOIR's capabilities can be dramatically expanded to increase the depth of lineage trees that can be tracked, the accuracy of reconstruction, and the number of distinct signals that can recorded at the same time. We envision MEMOIR being applied to study cell fate decision making in developing organisms and tumors, helping us understand how dynamic developmental signals and intrinsic events together specify cell fates and determine lineages.

Cell Detects Ratio not Quantity

Ertugrul Dalgic, Bulent Ecevit University School of Medicine

Principles

We know that the absolute concentration may not be important for a cell signaling event, but the changes in concentrations may be critical. There could be several mechanisms to achieve this, such as negative feed-back loops, receptor internalization, etc.

A new mechanism to produce this effect was shown in the pheromone response of *Saccharomyces cerevisiae* (Bush et al., Mol. Sys. Bio. *12*, 898). A ligand-dependent relay of signaling to opposite directions via an inhibitory regulator interaction with the receptor allowed the cell to compute the ratio of ligand-bound receptors. While the ligand-bound state of the system activated G protein switches, the unbound state inactivated them, causing an opposite effect. Therefore, the cell response is dependent on the ratio of number of the ligand-bound receptors to the free receptors. This seems to be a novel explanation of how biological systems achieve robustness against variations in the number of their components.

> *"…the cell response is dependent on the ratio of number of the ligand-bound receptors to the free receptors."*

What's Next?

G protein-coupled receptor signaling is involved in various cell signaling events. For example, the sensory cells mostly take information via these receptors. Therefore it will be critical to show the role of the identified mechanism for the robustness of other types of sensory information uptake.

Autonomous Metabolic Oscillations

Alexandros Papagiannakis, Bastian Niebel, Ernst C Wit, and Matthias Heinemann, University of Groningen, Netherlands

Principles

The eukaryotic cell cycle consists of strictly ordered molecular processes related to the replication and segregation of the genome. According to textbooks, the oscillatory activity of the cyclin/CDK machinery controls these processes. Recently, by means of dynamic metabolite measurements in single yeast cells, we found that yeast metabolism is an autonomous oscillator, meaning that it continues to oscillate, on the few-hour timescale, also in the absence of the cell cycle (Papagiannakis et al., Mol. Cell., published online December 10, 2016. http://dx.doi.org/10.1016/j.molcel.2016.11.018). Further, we discovered that the metabolic oscillator dynamically orchestrates the early and the late cell cycle, and determines cell cycle initiation in a metabolic frequency-dependent manner. The autonomous metabolic oscillator might constitute a primitive cell cycle regulator, on top of which later the CDK machinery might have been grafted.

> *"…we found that yeast metabolism is an autonomous oscillator…."*

What's Next?

We suggest that the metabolic oscillator and the cell cycle form a system of coupled oscillators, mutually influencing each other and together exerting the higher order function of cell cycle control. In order to understand the molecular functioning of this highly complex coupled oscillator system, next, it will be necessary (1) to identify the metabolic pathways whose temporal activity are responsible for the metabolite oscillations, (2) to unravel the mechanisms that drive these metabolic oscillations, and (3) to decipher the molecular interactions between the metabolic and cell cycle oscillators.

A Pathogen-Resistant Designer Microbiota

Bärbel Stecher, Ludwig-Maximilians-University of Munich; Thomas Clavel, Technical University of Munich; Alexander Loy and David Berry, Department of Microbiology and Ecosystem Science, University of Vienna

Principles

Our gut is home to hundreds of microbial species that play an essential role in preserving and promoting health. One of the primary functions of the microbiota is to provide colonization resistance, i.e. the ability to preclude pathogen infection. But what species and which of their capabilities are needed to create this protective shield? Partly due to the complexity of the native gut microbiota, little was known about the bacteria important for colonization resistance.

To address this, we designed a defined bacterial consortium that ensures protection against the human pathogen *Salmonella enterica* serovar Typhimurium in gnotobiotic mice (Brugiroux et al., Nat. Microbiol. *2*, 16215). We began with a consortium of 12 bacterial strains isolated from the mouse intestine that together provided only partial colonization resistance. Genome-based comparison of this minimal microbiota with natural gut communities and subsequent supplementation with further mouse gut strains (www.dsmz.de/miBC) identified facultative anaerobic commensals as crucial in mediating complete protection against *Salmonella*.

> *"…we designed a defined bacterial consortium that ensures protection against the human pathogen* Salmonella enterica *serovar Typhimurium in gnotobiotic mice."*

What's Next?

This work provides a foundation for future work in the field of microbe-host interactions, as the gnotobiotic mouse model can now be used for detailed mechanistic studies of colonization resistance and for the creation of standardized mouse models for other host phenotypes. Additionally, the continuous extension of the reference catalogue of well-characterized mouse gut strains along with analyses of their bacteriophages (i.e., the gut virome) might provide new means for targeted manipulation of individual defined-community members and thus for developing future microbiota-based therapies.

Cell Systems Call

Stealthy Control of Proteins and Cellular Networks in Live Cells

Nikolay V. Dokholyan, Department of Biochemistry & Biophysics, UNC-Chapel Hill; and Klaus M. Hahn, Department of Pharmacology, UNC-Chapel Hill

Principles

Can we establish control of diverse signaling cascades in living cells? This would open avenues for interrogating cellular networks and even producing designer cells that respond to disease environments in vivo. Unfortunately, modifying signaling proteins for optogenetic control usually results in perturbation of normal ligand interactions and/or protein function. Ideally, one would install a "handle" that responds to light or small ligands, but sits where it does not block endogenous interactions. Dagliyan et al. (Science *354*, 1441-1444) have accomplished this by modifying proteins at allosteric sites.

To modulate protein activity, we utilized the small domains LOV2 for light control and uniRapR for control by rapamycin, previously developed by Dagliyan et al. (Proc. Natl. Acad. Sci. U.S.A. *110*, 6800–6804). We harnessed the structural order of these domains: the presence of light or ligand determined whether the domains were in a disordered, unstructured state or more tightly folded. These domains could be easily inserted into protein surface loops identified using a straightforward computational approach. Through allostery the domains' structural order/disorder was coupled to the active site of the protein. We accessed the active site using "invisible" allosteric handles installed at a distance, thereby controlling live cell morphology.

> *"We accessed the active site using "invisible" allosteric handles installed at a distance, thereby controlling live cell morphology."*

What's Next?

Switching a single protein inside live cells enables interrogation of protein function and the role of spatio-temporal dynamics in vivo. The next step is to control multiple proteins simultaneously, to dissect and control interactions that generate cell behavior.

Thinking Fast and Slow: Synthetic Posttranslational Circuits for Dynamical Control of Cell Fate Decisions

Russell M. Gordley and Wendell A. Lim, University of California, San Francisco

Principles

Many long-term cellular decisions in development, synaptic plasticity, and immunity require that cells recognize patterns in how stimuli change over time, such as pulse duration or frequency. In these dynamically controlled cells, incoming stimuli are often filtered through a rapid-acting signaling layer, then passed to a downstream slow-acting layer that orchestrates the longer-term response. Directly testing the design principles of this dual-timescale network has been challenging, because most tools in synthetic biology allow rewiring of slow gene expression circuits but not of rapid signaling circuits.

We recently reported the development of synthetic phospho-regulons (Gordley et al., PNAS, *113*, 13528–13533). This method of phospho-engineering reroutes kinase signaling to rapidly modulate protein interaction, localization, and stability, and enables the assembly of dual-timescale regulatory networks. We used phospho-regulons to construct dynamical control circuits in yeast, and demonstrated that a simple combination of fast and slow regulation (sequential phosphorylation-based and transcription-based feedback loops) enables selective tuning of the cell's pulse duration sensitivity and the amplitude of its activated stable state.

> *"…phospho-engineering reroutes kinase signaling to rapidly modulate protein interaction, localization, and stability, and enables the assembly of dual-timescale regulatory networks"*

What's Next?

We anticipate that dual-timescale switches will be used to record the dynamics of cellular events. This tunable circuit can also be used to "read" messages encoded in the dynamics of a synthetic stimulus—empowering a single medium, such as light, to transmit multiple instructions to an engineered cell.

Comprehensive Benchmarking of RNA-Seq Aligners Indicates Large Variation in Performance

Giacomo Baruzzo, and Barbara Di Camillo, University of Padova; Katharina E Hayer, Eun Ji Kim, Garret A FitzGerald and Gregory R Grant, University of Pennsylvania

Principles

Read alignment is the first step in most RNA-seq analysis pipelines, and the accuracy of downstream analyses depends heavily on this step. There are at least 14 methods being used regularly in research; however there has not been an unbiased comprehensive benchmarking study to compare them. We performed such an analysis by simulating data from different organisms and at different levels of complexity (Baruzzo et al., Nat. Methods, published online on December 12, 2016. http://dx.doi.org/10.1038/nmeth.4106). We analyzed performance at the base, read, and junction levels using default and optimized parameters. Furthermore, we examined execution time and memory usage; the effect of untrimmed adapters; performance on indels, multi-mappers and canonical/non-canonical junctions. There are vast differences between the methods, in terms of accuracy and computational performance, highlighting the importance of choosing an aligner carefully.

> *"… the most reliable general-purpose aligners appear to be CLC Genomic Workbench, Novoalign, GSNAP and STAR."*

What's Next?

We found that accuracy and popularity were poorly correlated: the most widely used tool (TopHat2) underperforms for most metrics, sometimes dramatically, particularly when using default settings. Based on our analysis, the most reliable general-purpose aligners appear to be CLC Genomic Workbench, Novoalign, GSNAP and STAR.

Demystifying Complex Traits for Better Strain Engineering

Matthew J. Maurer, Jeffrey M. Skerker, and Adam P. Arkin, University of California, Berkeley

Principles

Engineering complex organismal phenotypes for real world applications is challenging because most traits of interest are multigenic and predicting the exact combination of optimal alleles is extremely difficult. Our goal is to understand the genetic basis of complex traits to make strain engineering easier and more rational. We studied two yeast strains of *Saccharomyces cerevisiae* with differing abilities to grow in a plant hydrolysate—a complex mixture of fermentable sugars and toxic compounds—that is a renewable feedstock for making biofuels, including lignocellulosic ethanol. To survive and produce ethanol in this harsh environment, the organism must deploy multiple pathways such as stress response, detoxification, and altered metabolism.

We developed an approach for discovering and optimizing host factors that impact fitness of *S. cerevisiae* during such fermentation (Maurer et al., ACS Synth. Bio., published online on December 12, 2016. http://dx.doi.org/10.1021/acssynbio.6b00264). We combined high-resolution Quantitative Trait Loci (QTL) mapping and Reciprocal Hemizygosity Analysis to discover loci that differentiate hydrolysate tolerance between an industrially derived and a laboratory yeast strain. We applied this "genetic blueprint", and a dual-guide CRISPR/Cas9-based locus replacement method, to engineer the weaker laboratory strain to have superior hydrolysate tolerance, demonstrating our overall approach.

> *"Our goal is to understand the genetic basis of complex traits to make strain engineering easier, faster, and more rational."*

What's Next?

We are applying this approach to dissect and engineer other complex traits, such as xylose utilization, and developing methods for higher-order trait stacking, identifying and controlling epistasis, and more efficient multiplex strain engineering.