Principles of Systems Biology, No. 21

This month: relating single cells to populations (Cao/Packer, Wu/Altschuler, O'Brien, Friedman), an excess of ribosomes (Barkai), human pathology atlas (Uhlen), signatures of feedback (Rahi), and major genome redesign (Baumgart).

Single-Cell Transcriptional Profiling of a Multicellular Organism

Junyue Cao, Jonathan Packer, Robert Waterston, Cole Trapnell, and Jay Shendure, University of Washington

Principles

Single-cell RNA sequencing (sc-RNA-seq) has emerged as a powerful strategy for resolving cellular heterogeneity in multicellular organisms. One of the challenges in designing sc-RNA-seq experiments is that some cell types can be orders of magnitude less frequent than others, but costs per cell scale linearly. To address this challenge, we developed a combinatorial indexing strategy to profile the transcriptomes of large numbers of single cells at a sample preparation cost that grows sub-linearly with the number of cells processed (single-cell combinatorial indexing RNA-seq, or sci-RNA-seq) (Cao et al., Science 357, 661–667).

To demonstrate the potential of sci-RNA-seq, we dissociated whole *C. elegans* larvae and sequenced >50,000 cells from this population, effectively achieving over 50-fold "shotgun" coverage of the global cellular content of the organism. We identified 27 major cell types, as well as 40 neuron subtypes, including rare cells types corresponding to one or two cells in the *C. elegans*. All of our data and analysis scripts are available at http://atlas.gs.washington.edu.

> *". . .we developed a combinatorial indexing strategy to profile the transcriptomes of large numbers of single cells. . .effectively achieving over 50-fold "shotgun" coverage of the global cellular content of the organism (C. elegans* larvae)."

What's Next?

We anticipate that combinatorial indexing may be used to profile multiple aspects of cell state, along with lineage history, temporal, and spatial information, in the same single cells. We are also working toward profiling each cell throughout the life cycle of the worm.

Sampling to Capture Single-Cell Heterogeneity

Satwik Rajaram, University of Texas Southwestern; Lani F. Wu and Steven J. Altschuler, University of California, San Francisco

Principles

Large-scale experiments are increasingly performed to characterize cell-to-cell heterogeneity and identify components that have functional meaning. This often requires assessing heterogeneity across a large number of conditions. However, repeated samples may be difficult to obtain, particularly from scarce or nonrenewable patient tissue, and experiments can be costly. A critical question is: how many samples should be taken from a population to be confident that cellular heterogeneity has been captured well?

We develop a data-driven framework to analyze a user-provided, "reference" collection of samples and estimate the sampling depth required to reliably profile heterogeneity in prospective studies (Rajaram et al., Nat. Methods, published online September 4, 2017. http://dx.doi.org/10.1038/nmeth.4427). The method searches within the provided collection for a minimal number of subsamples that can confidently return a phenotypic distribution that is similar to that of the whole population. While applicable to any single-cell study, this approach is particularly important for image-based studies, where tissue heterogeneity may vary from region to region, and more subsamples would be required to characterize the true heterogeneity than for a well-mixed population.

> *"The method searches within the provided collection for a minimal number of subsamples that can confidently return a phenotypic distribution that is similar to that of the whole population."*

What's Next?

The current approach is designed to work with one cellular feature at a time. A natural extension will be to consider multivariate measures of phenotypic similarity between subsamples and whole populations. The approach depends on whether a reference collection of samples truly represents the underlying biological heterogeneity, and prospective studies may be useful feedback to update estimates of sampling depth.

Organ Size: Act Locally to Control Globally

Jackson Liang and Lucy Erin O'Brien, Stanford University

Principles

Most adult organs are in a state of continuous turnover, in which old differentiated cells are lost and replaced by new progeny of stem cells. For organs to maintain constant size, there must be strict equilibrium between cell loss and production. However, the mechanistic basis of this equilibrium has been poorly understood.

Examining the *Drosophila* adult midgut, we found that tissue-level equilibrium arises through local signals from dying enterocytes that release a block on stem cell division (Liang et al., Nature 548, 588–591). When an enterocyte undergoes its natural death, loss of E-cadherin triggers induction of *rhomboid* and consequent secretion of EGFs. These EGFs signal to nearby stem cells, which spurs a stem cell to divide and generate a new replacement. Only stem cells within a set radius of the dying cell receive the EGF signal. These zones of EGF activation, integrated across the expanse of the tissue, may set and maintain organ size over the organism's lifetime.

> *". . .our findings imply that the spatial density of stem cells is central to organ size control. . . ."*

What's Next?

First, our findings imply that the spatial density of stem cells is central to organ size control, but how is this density established and sustained? Second, we show that stem cells respond to enterocyte loss; conversely, do enterocytes sense stem cell divisions and adjust their lifespans accordingly?

Cell Systems Call

Cells Combine Cytokine Signals in a Hierarchical-Additive Manner

Inbal Eizenberg-Magar, Jacob Rimer, and Nir Friedman, Department of Immunology, Weizmann Institute of Science

Principles

Understanding the logic by which cells respond to complex signal combinations is of major interest. We studied signal integration by CD4 T cells, using flow cytometry to systematically map their response to 64 mixtures of cytokines that drive their differentiation (Eizenberg-Magar et al., Proceedings of the Nat. Academy of Sciences 114, E6447–E6456). We find that in response to varied input combinations, cell populations span a continuum of cell fates, as opposed to a limited number of discrete phenotypes. Furthermore, although the response of individual cells is highly heterogeneous, the mean response of cell populations is more predictable.

Mathematical modeling explains these results using hierarchical summation of cytokine inputs and correctly predicts responses to new input conditions. According to this model, the response to a combination of cytokines is the weighted sum of responses to individual cytokines, but some cytokines are more dominant and tune the weights by which the other cytokines contribute to the additive response.

". . .in response to varied input combinations, cell populations span a continuum of cell fates, as opposed to a limited number of discrete phenotypes. . . . Mathematical modeling explains these results using hierarchical summation of cytokine inputs. . . ."

What's Next?

Our findings suggest that complex cellular responses can be effectively described using simple hierarchical summation rules. This understanding can provide a general framework for prediction of responses to signal combinations also in other systems, with potential applications for generating therapeutic T cells with improved functionality.

Excess Ribosomes for Increased Translational Demands

Eyal Metzl-Raz, Moshe Kafri, Gilad Yaakov, Ilya Soifer, Yonat Gurvich, and Naama Barkai, Weizmann Institute of Science

Principles

To duplicate protein content at each division, growing cells must coordinate production of ribosomes with all other cellular proteins. It has long been appreciated that maximal growth rate is obtained when cells produce the highest amount of fully active ribosomes necessary for their immediate growth requirements. Indeed, if produced, inactive ribosomes will not contribute to growth but will still consume translational resources of other cellular proteins.

Do cells function at this limit of full ribosomal usage? Classical studies in bacteria suggested that this may indeed be the case. By contrast, we found that budding yeast, a model eukaryote, produces excess ribosomes, so that even during rapid logarithmic growth, a significant fraction (~8%) of its entire proteome consists of ribosomal proteins that are not actively translating at a given time (Metzl-Raz et al., eLife. http://dx.doi.org/10.7554/eLife.28034). This excess ribosome pool is employed when translation demands abruptly increase, e.g., during nutrient upshift. This suggests that yeast ribosome capacity is evolutionarily tuned not solely toward maximal growth rate but also toward changing environmental conditions.

". . .budding yeast. . .produces excess ribosomes, so that even during rapid logarithmic growth, a significant fraction (~8%) of its entire proteome consists of ribosomal proteins that are not actively translating at a given time."

What's Next?

What could be the selective advantage for maintaining excess inactive ribosomes at the expense of immediate growth? One possibility is that this design enables near-instantaneous adaptation to improved nutrient conditions. Future studies will probe potential evolutionary benefits of excess ribosomes and further address how different organisms resolve this trade-off between fast adaptation and immediate growth rate.

The Human Pathology Atlas

Adil Mardinoglu, Science for Life Laboratory, KTH–Royal Institute of Technology; Fredrik Ponten, Department of Immunology Genetics and Pathology, Uppsala University; Mathias Uhlen, Science for Life Laboratory, KTH–Royal Institute of Technology

Principles

It is necessary to unveil the underlying molecular mechanisms involved in the occurrence of individual tumors for the development of efficient treatment strategies. We recently generated a Human Pathology Atlas for protein coding genes in major human cancers (Uhlen et al., Science 357, eaan2507). We employed powerful systems biology tools to analyze the transcriptome of 17 main cancer types using data from 8,000 patients and generated more than 900,000 survival plots describing the consequence of RNA and protein levels on clinical survival. We also generated personalized genome-scale metabolic models for each cancer patient to identify key genes involved in tumor growth. Prognostic genes identified for lung and colorectal cancer were validated at the protein level in independent, prospective cancer cohorts using immunohistochemistry. The Pathology Atlas is available via an interactive open-access database (www.proteinatlas.org/pathology).

"We. . .generated more than 900,000 survival plots describing the consequence of RNA and protein levels on clinical survival."

What's Next?

The systems level approach used to create the Pathology Atlas demonstrated the potential use of omics technologies in transforming how medical research is performed. The Human Pathology Atlas and extension of similar efforts on more focused and clinically relevant cancer patient cohorts may accelerate the adoption of personalized cancer medicine. It may also foster the development of effective cancer drugs that target the identified driver genes associated with individual tumor growth.

Identifying Signatures of Feedback

Sahand Jamal Rahi and Frederick R. Cross, The Rockefeller University

Principles

Let us treat a biological system like a wrapped gift: shake it, listen to its contents rattling, and try to figure out what is inside. For biological systems, there are, in general, too many different possibilities that give similar responses to the same stimulus for us to make a reliable guess about the network of molecular interactions. So, we focused on systems that show adaptation, i.e., that desensitize to a continuous stimulus.

These systems, which are ubiquitous in biology, must contain incoherent feedforward loops or negative feedback loops. Can we discriminate these two fundamental circuit topologies? We settled on oscillatory on-off stimuli as the simplest patterns of perturbations that could do the task.

We found two identifying "response signatures": in contrast to incoherent feedforward loops, only negative feedback loops show period skipping (responding intermittently to periodic stimuli), and, generically, only negative feedback loops have a stable refractory period (the time it takes the system to resensitize after a stimulus ends) (Rahi et al., Nat. Methods, published online 28 August 2017. http://dx.doi.org/10.1038/nmeth.4408). We used these to identify a calcium negative feedback loop leading to adaptation in the *Caenorhabditis elegans* olfactory neuron AWA and the subsystem that stabilizes the cell cycle period in yeast.

". . .in contrast to incoherent feedforward loops, only negative feedback loops show period skipping. . .or have a stable refractory period. . . ."

What's Next?

For the numerous signaling pathways in biology that are poorly understood, this approach could bring insight by elucidating the underlying circuit motif. Also, both signatures (stable refractory periods and period skipping) are often observed when researchers stimulate cells without realizing that they indicate the presence of negative feedback loops.

Bacterial Chassis Strain C1*

Meike Baumgart and Stephan Noack, IBG-1: Biotechnology, Forschungszentrum Jülich

Principles

Targeted top-down genome reduction strategies are considered to have a high potential for providing robust basic strains for synthetic biology and industrial biotechnology. In contrast to other genome reduction projects aiming at a minimal cell, we want to construct chassis strains for research and industry with reduced complexity, while maintaining growth behavior and robustness of the wild-type. Recently, we created a library of 26 genome-reduced strains of *Corynebacterium glutamicum* carrying broad deletions in single gene clusters without impact on biological fitness. We proceeded by combining gene cluster deletions, and the final chassis strain C1* carries a genome reduction of 13.4% (412 deleted genes) (Baumgart et al., ACS Synth. Biol., published online August 14, 2017. http://dx.doi.org/10.1021/acssynbio.7b00261). C1* shows wild-type-like growth behavior in defined medium with D-glucose as carbon and energy source. Moreover, C1* proves to be robust against several stresses (including oxygen limitation) and shows long-term growth stability under defined and complex medium conditions. Alongside this study, a large strain library and a revised genome annotation list were generated, offering the opportunity to screen for irrelevant gene cluster under different growth conditions and to successively unravel the function of the still numerous uncharacterized genes in *C. glutamicum*.

"Major genome redesign. . . ."

What's Next?

As we have already targeted most of the large regions without essential genes, alternative strategies are required to accomplish significant further genome reduction. Major genome redesign, e.g., by joining all genes of a certain pathway in one synthetic gene cluster, is one promising option.