Advances in technology and computational analysis paving the way for translational studies

May 22 2015, by Lisa Girard, Broad Communications

The Human Genome Project gave us an incredible foundation from which to understand our potential genetic repertoire. In order to understand the actual roles of particular genes in disease, however, it is not only critical to identify genes, but also to know in which cells the genes are expressed and when.

"Genetics has gotten very good at finding genetic variation associated with disease. Going from genetic results to biological understanding of disease is the next important challenge," said Broad Institute member Steve McCarroll, also director of genetics in the Stanley Center for Psychiatric Research at the Broad and an assistant professor of genetics at Harvard Medical School.

Critical to these biological insights is the ability to analyze individual cells' roles in health and disease. New advances in single-cell technology and computational analysis by researchers at the Broad are enabling studies with a scale, ease, and affordability not possible before, opening the door for applications increasingly closer to the clinic.

The technology to sequence and analyze the genes expressed in a single cell—its transcriptome—has accelerated dramatically in just a few years. In 2012, Broad core institute member Aviv Regev and colleagues published a landmark analysis consisting of data from 18 cells. Last year, they expanded this to analyze hundreds of cells at once. Despite this

growth, access to single-cell analysis was limited by the cost and time involved in preparing hundreds of individual cells for sequencing.

McCarroll, Regev, and colleagues sought to change that. "We wanted to be able to chart the many cell types that are in the human body, human disease, and the many states and conditions that they may assume," said Regev. "That would mean assaying many millions of cells, requiring a radical change from the available technologies where every cell had to be physically isolated from each other on a plastic plate or device."

In an effort to do just that, Stanley Center research fellow Evan Macosko, from McCarroll's lab, and Klarman Cell Observatory postdoctoral fellow Anindita Basu, from Regev's lab and that of associate member David Weitz, worked together and with other Broad colleagues, including computational biologist James Nemesh, and MIT assistant professor and Broad associate member Alex Shalek, to develop an approach they call Drop-Seq, with generous funding from the Stanley Center for Psychiatric Research, the Klarman Cell Observatory, the Simons Center for the Social Brain, and, most recently, the BRAIN initiative. Featured in *Cell*, Drop-Seq takes thousands of single cells isolated from tissues and mixes each one with a tiny, DNA-barcoded microbead in a nanoliter-sized droplet. Once in the droplet, the cell is broken apart and its messenger RNA (mRNA) is captured on the microbead. The mRNA/microbead complex is called a STAMP, for Single-cell Transcriptomes Attached to MicroParticles. The mRNA is then retrieved and converted into DNA, which is amplified and sequenced. Thousands of STAMPs (corresponding to thousands of cells) can be sequenced in a single reaction. The STAMP barcode can then be used to determine from which cell each transcript came.

Tracking each cell in this new technique is critical. "We are replacing wells on a microtiter plate with drops," said Weitz, who has worked for years to develop single-cell droplets with his group at the Harvard School

of Engineering and Applied Sciences. "The problem is that the wells on a microtiter plate have spatial information. We lose that by going to drops and we have to somehow get that information back." Barcoding the microbeads allows cells to be tracked while running at great scale using very small reaction volumes.

Single-cell analysis is a powerful experimental approach, in part because information from rare cell types is not lost among the masses. Unlike approaches that analyze DNA from a population of cells at once, single-cell analysis captures the heterogeneity of cells in a tissue. Testing this ability with Drop-Seq, the researchers analyzed more than 45,000 cells from the mouse retina, a well-studied model system with a large number of cell types. From their data, the researchers were able to identify 39 retinal cell types, including known types to show that the method works and several potential new ones. Before Drop-Seq, such an achievement would have taken decades.

The high throughput capability of Drop-Seq helped researchers identify the new cell types. "When you are analyzing a single cell it's kind of like a snowflake," said Macosko. "It's going to be different from every other cell in your system, but even though all snowflakes are different, there are certain things that all snowflakes have in common. The only way you would be able to know that is if you saw thousands and thousands of snowflakes."

In addition to its high throughput capabilities, Drop-Seq addresses two additional single cell analysis challenges—high cost and risk of contamination. Drop-Seq reduces the cost of single-cell analysis to about six cents per cell—a 500-fold decrease from previous methods. While mechanically or physically separating cells into individual wells of a microtiter plate in other single cell approaches is expensive, Drop-Seq's nanoliter scale droplets are so tiny that 100 million of them fit in a test tube the size of a person's thumb. Since individual cells are tracked with

microbead barcodes, the whole sample can be pooled, shrinking the sequencing reaction volume and reducing costs significantly. Basu also noted that, "we don't have to worry about contamination as much since everything happens inside drops. The intact cell goes into a drop, it gets lysed, barcoded, and only then do you open the drop. At that point, contamination does not really affect it."

Drop-Seq, and other methods for single cell analysis, usually begins with a dissociated tissue, but inside the body a cell's location matters a great deal. New approaches in computational analysis are now tackling this challenge as well. A team including Rahul Satija, from Regev's lab, now at New York Genome Center, and Jeff Farrell, from associate member Alex Schier's group at Harvard University, developed a computational method known as Seurat to analyze gene expression patterns across complex tissues and derive cellular localization. Named after the French impressionist whose pointillist style is evocative of detailed spatial patterning information, Seurat uses machine learning approaches to map single cells by integrating single-cell RNA sequencing data with RNA in situ hybridization patterns from tissues. The team used Seurat to generate a map of spatial patterning in zebrafish embryos based on the transcriptome of over 800 cells. In *Nature Biotechnology*, they describe their work in the fish, but the approach is applicable to other systems including human tissue samples. For canonical tissues, Seurat may be unbounded in terms of scalability and, Regev said "Drop-Seq is its best friend," as parts of the Seurat package and work by Rahul Satija were integral for the analysis of cells and classification of the retina data obtained using Drop-Seq. "Seurat provides a framework in which to tackle the scale and type of data that experimental biologists pose when they use single cell genomics," Regev said.

With recent advances, single-cell analysis has become ripe for translational studies. Institute member Levi Garraway and his lab members at Dana Farber Cancer Institute, in collaboration with a team

from the Klarman Cell Observatory, including postdoctoral fellow Itay Tirosh, Shalek, and the Observatory's Associate Director, Orit Rozenblatt-Rosen, has established a workflow to study several different types of cancers. Early results indicate that it may reveal important information about tumor heterogeneity and the factors controlling whether tumors respond to different types of therapies. Since single-cell analysis can identify alterations in gene activity within a tumor that eludes bulk analysis, if differences are found, the data can provide clues about the molecular pathways involved in tumor response. Also, monitoring expression changes over time in response to a therapy could help predict when cells will fail to respond. "We think that single-cell analysis may help us discern heterogeneity—not simply for its own sake, but for what it tells us about tumor subpopulations, how well they respond to therapy, and how tumor cells evolve to drug resistance." said Garraway.

Another area in which single-cell analysis can have a tremendous impact is microbial genomics. "Being able to discover and understand microbial diversity is continuing to grow in importance," said Broad core institute member Paul Blainey, particularly as we learn more about the impact of the microbiome – the billions of microorganisms that live in and on the human body – on our health. Additionally, in response to the significant global threat posed by antibiotic resistance, researchers have stepped up efforts leveraging single cell analysis to find new antibiotics and identify antibiotic resistance genes.

With these advances, single-cell analyses at the Broad have expanded to the scale of tens of thousands of cells – up from 18 just three years ago. As Eric Lander, president and director of the Broad Institute noted, this new era of "quantitative biomedicine" will allow us to learn from two types of populations – populations of people and populations of cells. "We need to learn from the natural variation among cells," he said. "It's time to start thinking about a complete cell catalog." Such a "cell atlas"

will afford an unbiased way of identifying every cell type in the human body based on things like cell state, environment, lineage, and history. "A human cell atlas is now within reach," Regev said. "It will require some discipline, methods, and computational approaches that will be scalable, sensitive, and robust. Most significantly it will require a partnership across biology, technology, and medicine." Such a partnership would be a culmination of the interdisciplinary advances that have so successfully launched single cell genomics forward. It will cut across experimental, computational, and mathematical approaches and it has the potential to transform our understanding and treatment of human disease.

More information: "Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets." *Cell*. Online May 21, 2015. DOI: DOI: 10.1016/j.cell.2015.05.002

"Spatial reconstruction of single-cell gene expression data." *Nature Biotechnology*. Online April 13, 2015. [DOI: 10.1038/nbt.3192](DOI: 10.1038/nbt.3192)

Provided by Broad Institute of MIT and Harvard