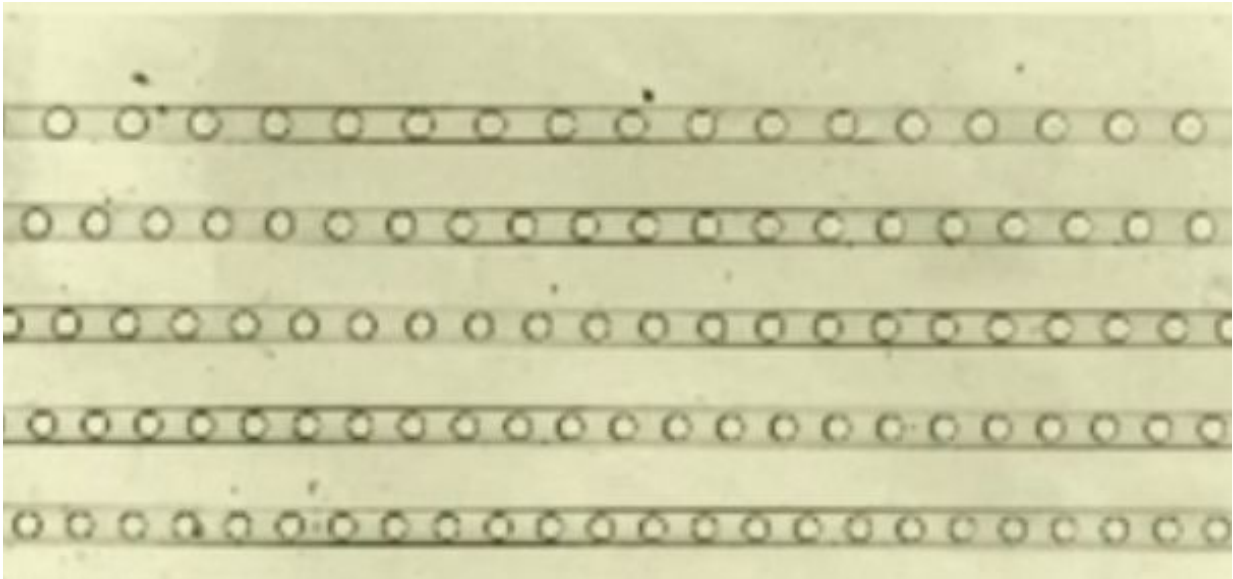# Beyond average: New platforms genetically barcode tens of thousands of cells at a time

May 21 2015

Imagine someone hands you a smoothie and asks you to identify everything that went into it.

You might be able to discern a hint of strawberry or the tang of yogurt. But overall it tastes like a blend of indiscernible ingredients.

Now imagine that the smoothie is made of 20,000 ground-up cells from, say, the brain.

You could run tests to determine what molecules are in the sample, which is what scientists do now. That would certainly give you useful information, but it wouldn't tell you which cells those molecules originally came from. It would provide only an average cell profile for the whole smoothie.

And when it comes to the tissues in our bodies, averages are almost always misleading. Just as you know there isn't an "average" food called strawbanaspinach-orangegurt, scientists know there isn't just one cell type in the brain.

"If you take a hunk of tissue and grind it up and analyze the RNA, you have no idea if it represents what every cell in that population is doing or what no cell in the population is doing," said Marc Kirschner, the John Franklin Enders University Professor of Systems Biology and chair of the Department of Systems Biology at Harvard Medical School. "Imagine if you had a population of men and women. If you assume everyone is an average of men and women, you [probably] wouldn't represent a single person in that population."

The trouble is, it's expensive, time-consuming and tricky to characterize tissues one cell, or cell type, at a time.

Kirschner and Steven McCarroll, assistant professor of genetics at HMS, reported this week in separate papers that their labs have developed high-throughput techniques to quickly, easily and inexpensively give every cell in a sample a unique genetic barcode before it goes into the blender.

As a result, scientists can analyze complex tissues by profiling each individual cell—no averaging required.

"Different cells in a tissue use the same genome in amazingly diverse ways: to engineer specialized cell shapes, accomplish diverse feats of

physiology, and mount distinct functional responses to the same stimulus. These techniques will finally let science understand how biological systems operate at that single-cell level," said McCarroll, who is also director of genetics for the Stanley Center for Psychiatric Research at the Broad Institute of Harvard and MIT. "We are so excited about the work ahead."

To make their tools, both teams collaborated with David Weitz, the Mallinckrodt Professor of Physics and Applied Physics at Harvard's School of Engineering and Applied Sciences and a pioneer in the field of microfluidics.

The teams expect that their techniques, published concurrently in the journal *Cell*, will equip biologists to discover and classify cell types in the body in much greater depth, map cell diversity in complex tissues such as the brain, better understand stem cell differentiation and gain more insights into the genetics of disease.

Harvard's Office of Technology Development has been working closely with the researchers to develop patent applications for various aspects of the technology, all with an eye toward commercialization.

'Two roads diverged in a yellow wood'

Evan Macosko and Allon Klein met in a microfluidics class a few years ago. Then they went their separate ways.

Unbeknownst to each other, they decided to develop methods to answer the same question: How could they obtain gene expression profiles for thousands of individual cells to better understand the complexity of gene expression within a tissue?

Gene expression—the pattern of gene activity in a particular

cell—underlies every process in biology, from cognition in the brain to development in the egg. Scientists have known for 50 years that gene expression varies from cell to cell like a fingerprint, making skin cells different from [liver cells](#) and making some liver cells different from others. But they haven't been able to measure it efficiently at the single-cell level in samples with many cell types.

Macosko, HMS instructor in psychiatry at Massachusetts General Hospital and a Stanley Neuroscience Fellow in the McCarroll lab, came up with a technique he called Drop-seq. Klein, assistant professor of [systems biology](#) at HMS, devised a method he called indexing droplets for sequencing, or inDrops.

Last fall, they learned about each other's work through the scientific conference circuit.

"It was kind of like meeting your doppelgänger," said Macosko. "He had been thinking about the same things I had for two years. Human beings have different ways of solving problems, and it was really cool to see how he did it."

How they work

The teams each developed ways of using tiny beads to deliver vast numbers of different DNA barcodes into hundreds of thousands of nanometer-sized water droplets simultaneously.

Thanks to Weitz's expertise, both methods were able to use microfluidic devices to co-encapsulate cells in these droplets along with the beads. The droplets get created in a tiny assembly line, streaming along a channel the width of a human hair.

The bead barcodes get attached to the genes in each cell, so that

scientists can sequence the genes all in one batch and still trace each gene back to the cell it came from.

Macosko and Klein make their beads in different ways. The droplets get broken up at different steps in the process. Other aspects of the chemistry diverge. But the result is the same.

After running a single batch of cells through Drop-seq or inDrops, scientists "can see which genes are expressed in the entire sample—and can sort by each individual cell," said Klein.

They can then use computer software to uncover patterns in the mix, including which cells have similar gene expression profiles. That provides a way to classify what cell types were in the original tissue—and to possibly discover new ones.

Current methods allow researchers to generate 96 single-cell expression profiles in a day for several thousand dollars. Drop-seq, by comparison, enables 10,000 profiles a day for 6.5 cents each.

"If you're a biologist with an interesting question in mind, this approach could shine a light on the problem without bankrupting you," said Macosko. "It finally makes gene expression profiling on a cell-by-cell level tractable and accessible. I think it's something biologists in a lot of fields will want to use."

Rather than competing with each other, the teams believe that having two options available in Drop-seq and inDrops will benefit the scientific community.

"Each method has unique elements that makes it better for different applications. Biologists will be able to choose which one is most appropriate for them," said Macosko.

Different goals

McCarroll, Macosko and their colleagues are excited to explore the brain with Drop-seq.

With luck, that will include discovering new cell types, constructing a global architecture of those cell types in the brain and understanding brain development and function as they relate to disease.

Among the questions they want to pursue are: What are all the cell types that make the brain work? How do these cell types vary in their functions and responses to stimuli? What cell populations are missing or malfunctioning in schizophrenia, autism and other disorders of the brain?

Classifying cell types may not sound exciting, said Joshua Sanes, the Jeff C. Tarr Professor of Molecular and Cellular Biology and the Paul J. Finnegan Family Director of the Center for Brain Science at Harvard University and a co-author of the Drop-seq paper, but it lays the foundation for mapping neuronal circuits and one day being able to probe the mystery of how the "wetware" of the brain gives rise to thoughts, emotions and behaviors.

In the shorter term, Sanes looks forward to completing a catalog of cell types in the mouse retina. Drop-seq has already revealed several new ones.

Kirschner, Klein and their colleagues, meanwhile, are keenly interested in other areas, including stem cell development.

"Does a population of cells that we initially think is uniform actually have some substructure?" Klein wants to know; he's trying to find out by studying immune cells and different kinds of adult stem cells. "What is

the nature of an early developing stem cell? What endows those cells with a pluripotent state? Is gene expression more plastic or does it have a well-defined state that's different from a more mature cell? How is its fate determined?"

Using inDrops, Klein and team have confirmed prior findings that suggest even embryonic stem cells are not uniform. They found previously undiscovered cell types in the population they studied, as well as cells in intermediate stages that they suspect are converting from one type to another.

Although both teams are excited by the massive amounts of data they and other researchers will obtain from Drop-seq and inDrops, they realize the sheer volume of information poses a problem as well.

"We have thousands of cells expressing tens of thousands of genes. We can't look in 20,000 directions to pick out interesting features," said Klein.

Machine learning is able to do some of that, and the teams have already employed new statistical techniques. Still, Kirschner has called on mathematicians and computer scientists to develop new ideas about how to analyze and extract useful information about our biology from the mountains of data that are on the horizon.

More information: A patent application has also been filed for the work described by Macosko et al.

Provided by Harvard Medical School

Citation: Beyond average: New platforms genetically barcode tens of thousands of cells at a time