


New single cell analysis tool

Researchers at Penn and colleagues have developed a tool to analyze single cells that assesses both the patterns of gene activation within a cell and which sibling cells shared a common progenitor.

 3D illustration of a cell being held by a pipette and a needle.

Arjun Raj of the School of Engineering and Applied Science and the Perelman School of Medicine, former postdoc Lee Richman, now of Brigham and Women's Hospital, and colleagues have developed a new analysis tool that combines a cell's unique gene expression data with information about the cell's origins. The method can be applied to identify new cell subsets throughout development and better understand drug resistance.

R

Recent advances in analyzing data at the single-cell level have helped biologists make great strides in uncovering new information about cells and their behaviors. One commonly used approach, known as clustering, allows scientists to group cells based on characteristics such as the unique patterns of active or inactive genes or by the progeny of duplicating cells, known as clones, over several generations.

Although single-cell clustering has led to many significant findings, for example, new cancer cell subsets or the way immature stem cells mature into “specialized” cells, researchers to this point had not been able to marry what they knew about gene-activation patterns with what they knew about clone lineages.

Now, research published in *Cell Genomics* (<https://www.sciencedirect.com/science/article/pii/S2666979X22002105?via%3Dihub>) led by [University of Pennsylvania](https://www.upenn.edu/) (<https://www.upenn.edu/>) professor of bioengineering [Arjun Raj](https://directory.seas.upenn.edu/arjun-raj/) (<https://directory.seas.upenn.edu/arjun-raj/>) has resulted in the development of [ClonoCluster](https://github.com/leerichman/ClonoCluster) (<https://github.com/leerichman/ClonoCluster>), an open-source tool that combines unique patterns of gene activation with clonal information. This produces hybrid cluster data that can quickly identify new cellular traits; that can then be used to better understand resistance to some cancer therapies.

“Before, these were independent modalities, where you would cluster the cells that express the same genes in one lot and cluster the others that share a common ancestor in another,” says Lee Richman, first paper author and a former postdoc in the Raj lab who is now at Brigham and Women's Hospital in Boston. “What's exciting is that this tool allows you to draw new lines around your clusters and explore their properties, which could help us identify new cell types, functions, and molecular pathways.”

Researchers in the Raj Lab use a technique known as barcoding to assign labels to cells they are interested in studying, particularly useful for tracking cells, clustering data based on cells' offspring, and following lineages over time. Believing they could parse more valuable information out of this data by incorporating the cell's unique patterns of gene activation, the researchers applied ClonoCluster to six experimental datasets that used barcoding to track dividing cells' offspring. Specifically, they looked at the development of chemotherapy resistance and of stem cells into specialized tissue types.

Much like how stereo systems allow a user to control the balance of sound between the left speaker relative to the right, the researchers added a tunable parameter called “alpha” to ClonoCluster. This allowed the team to fine-tune how much emphasis they were placing on clustering by clonal barcode relative to patterns of gene activation—with 0 corresponding to patterns of gene activation and 1 to clonal barcodes. On this 0 to 1 alpha scale, they found that the closer alpha was set to the clonal channel, the more their hybrid clusters expressed the unique gene levels associated with extracellular environment interactions and the translation of proteins from messenger RNA.

“Normally, when you consider how cell function is decided and regulated, you think it's likely intrinsic mechanisms that decide which genes will be expressed, or the cell's fate,” says Richman. “So, we believe that this hints at the fact that cells can also inherit traits through non-genetic means, possibly through environmental cues.”

The prevailing notion behind gene regulation suggests that proteins, known as transcription factors, bind to specific points on DNA sequences and activate or suppress certain genes, helping explain why most cells in an animal have the same genetic information but serve varying roles. As these cells bind to different sites, transcription factors signal to the protein-manufacturing apparatus which proteins the cell should produce to stay functionally relevant. Scientists can then use these biomarkers to monitor cellular properties and make predictions about what the cell might do next.

Enhancing scientific understanding of how cells behave and differentiate could improve what's known about cellular cloning. It could also offer insight into what goes awry when duplication runs rampant, such as in the case for cancers and certain conditions associated with genetic mutations characterized by overrepresentation of clone cells in a given cellular population.

To take this work a step further, the researchers developed a complementary tool to ClonoCluster that they called Warp Factor. This incorporates an existing machine-learning tool that simplifies dataset visualization to better explain the relationship between clonal data and their hybrid clusters. Using both tools, the team identified distinct expression patterns that previously seemed too dispersed to draw inferences on cell identity, information that could help researchers understand more about how cells differentiate and function.

“We think this work shows that tracking cellular identity over time can provide a new layer of information that researchers have been unaware of and hopefully can be exploited to deepen our understanding of cancer and other diseases,” Raj says.

He and colleagues will continue to update and refine ClonoCluster. They believe this is a significant advance in the field of computational cell biology with the potential to provide a more complete picture of cellular characteristics and function in the future.

Arjun Raj is a professor of bioengineering in [Department of Bioengineering \(https://be.seas.upenn.edu/\)](https://be.seas.upenn.edu/) in the School of Engineering and Applied Science and a professor of genetics in the [Department of Genetics \(https://genetics.med.upenn.edu/\)](https://genetics.med.upenn.edu/) in the Perelman School of Medicine at the University of Pennsylvania. He runs the [Raj Lab for Systems Biology \(https://rajlab.seas.upenn.edu/\)](https://rajlab.seas.upenn.edu/).

Lee Richman, a former postdoctoral researcher in the Raj Lab for Systems Biology in the Department of Bioengineering in the School of Engineering and Applied Science, is currently a medical resident at Brigham and Women's Hospital.

Other authors include Connie Jiang, an MD/PhD student in the Perelman School of Medicine, and Yogesh Goyal of Northwestern University.

This work was supported by the Burroughs Wellcome Fund Career Awards at the Scientific Interface, the Jane Coffin Childs Memorial Fund, the Schmidt Science Fellowship, and the National Institute of Health's Director's Transformative Research Award (grants R01 238237, R01 CA232256, P30 CA016520, SPORE P50 CA174523, and U01 CA227550.)

CREDITS **Nathi Magubane**
Writer

Nathi Magubane
Media Contact

DATE **January 19, 2023**

SUBTOPICS **Bioengineering, Biology, Medicine, Research**

SCHOOLS **School of Engineering & Applied Science,
Perelman School of Medicine**