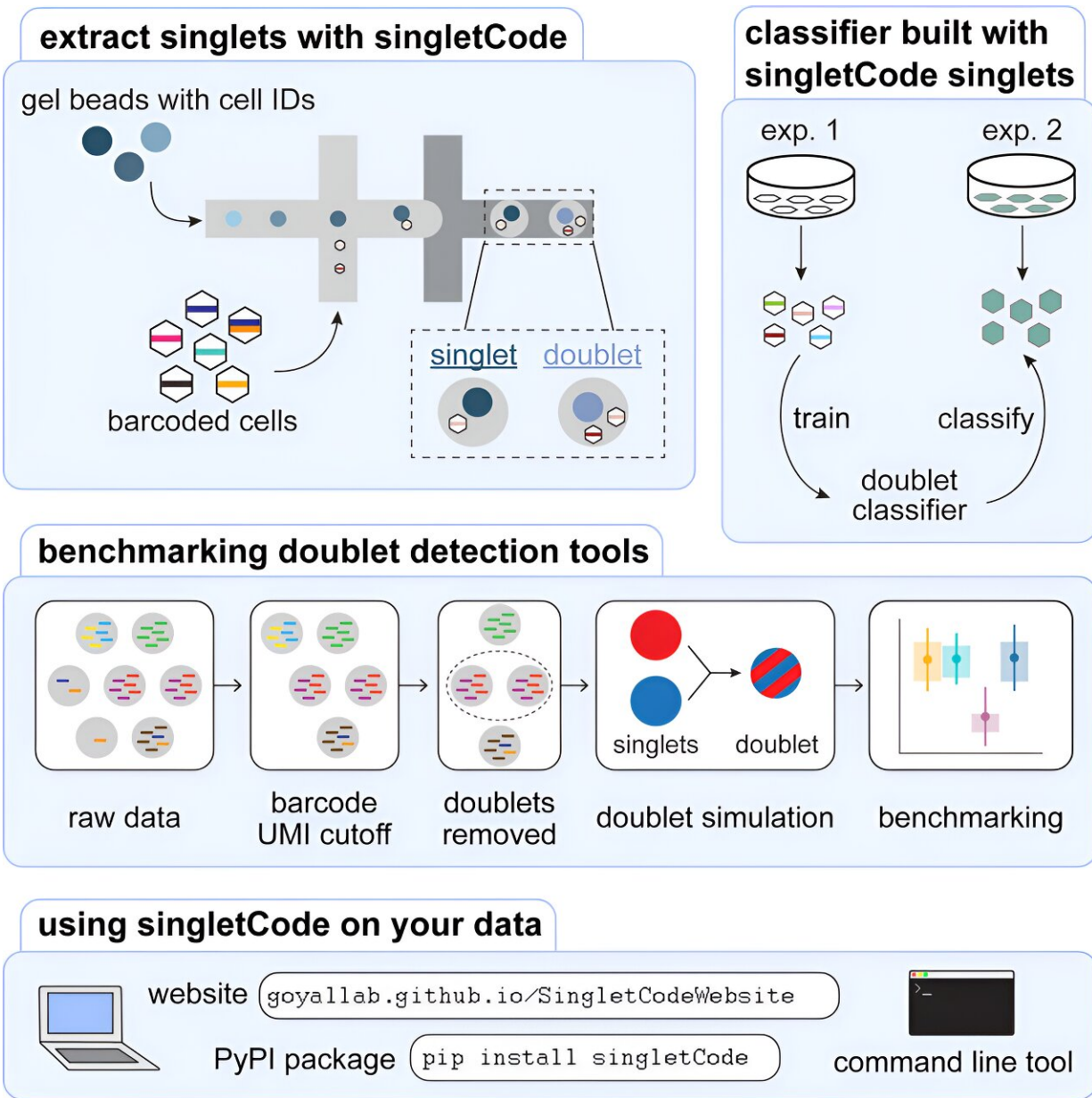


Researchers improve measurement of gene expression in single cells

June 26 2024, by Olivia Dimmer



Credit: *Cell Genomics* (2024). DOI: 10.1016/j.xgen.2024.100592

Northwestern Medicine scientists have developed a new technique to identify individual cells for RNA sequencing, which will empower scientists to gather more accurate and precise scientific data, according to details [published](#) in *Cell Genomics*.

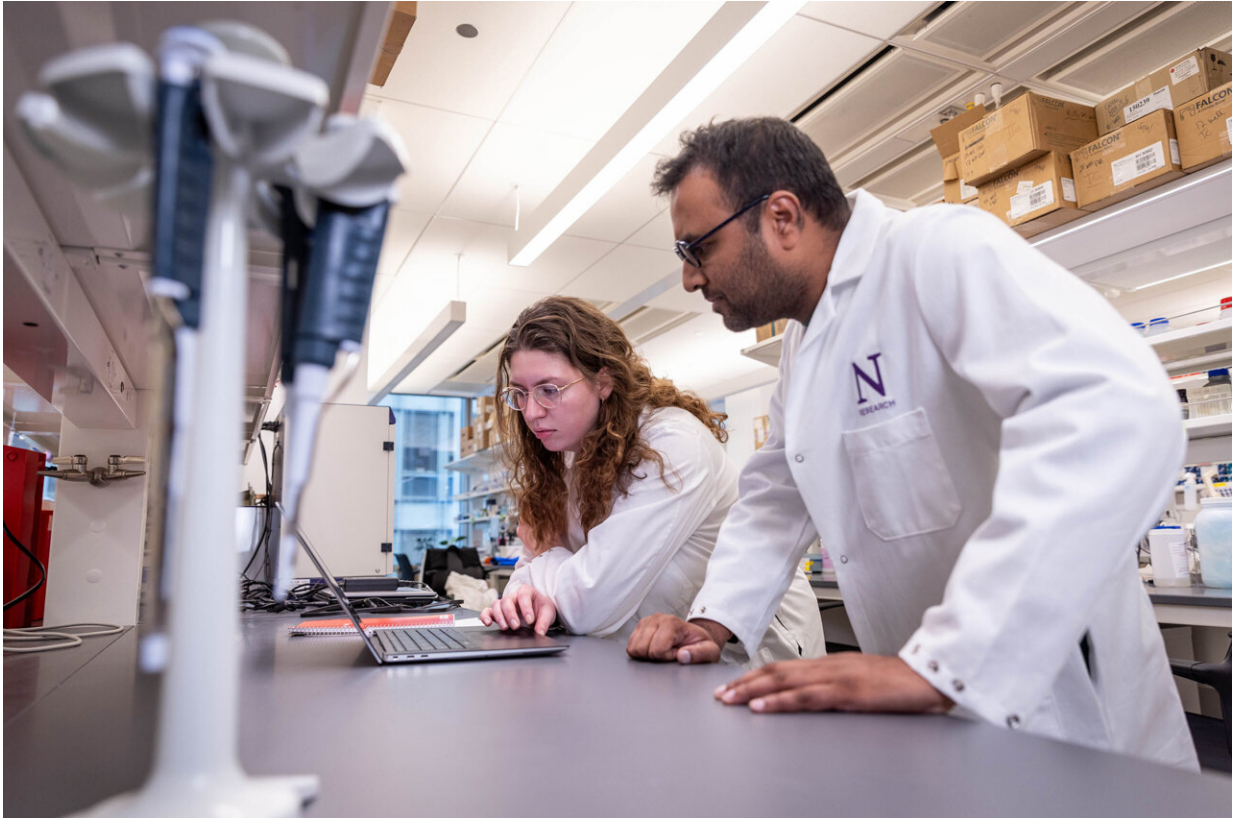
RNA sequencing, designed to reveal the quantity of RNA molecules in a biological sample and give scientists a snapshot of gene expression, has quickly become an essential tool in [scientific research](#), said Yogesh Goyal, Ph.D., assistant professor of Cell and Developmental Biology and senior author of the study.

"Single-cell RNA sequencing has really transformed the world of biomedicine," said Goyal, who is also a member of the Center for Synthetic Biology. "But one of the fundamental limitations of this technology is trying to isolate a single cell to pass through a microfluidic device. It's very easy to have more than one cell inside each sample. It leads to a lot of false-positives and false-negatives."

To address this, Goyal and his collaborators first employed a barcoding technique, in which individual cells (singlets) are labeled with unique nucleic acid sequences so they can be more easily tracked throughout an experiment.

Next, the investigators tested various existing machine learning algorithms to see how accurately they could distinguish synthetically barcoded single cells from groups of cells (doublets). They found that existing algorithms could not accurately differentiate singlets from doublets following RNA sequencing, according to the study.

Finally, Goyal and his laboratory developed their own machine learning algorithm designed to identify true singlets. By training it on barcoded singlet data from expansive dataset libraries, investigators were able to use the algorithm to successfully distinguish doublets from singlets more accurately than previous methods, according to the findings.



Scientists in the laboratory of Yogesh Goyal, PhD, assistant professor of Cell and Developmental Biology, have developed a new technique to identify individual cells for RNA sequencing. Credit: Northwestern University Global Marketing and Communications

"When you get multiple cells captured together, called a doublet, this can cause problems in your downstream analysis," said Madeline Melzer, a

Ph.D. student in the Driskill Graduate Program in Life Sciences (DGP) and co-first author of the study. "What we have done here is use barcodes introduced into individual single cells before they go in the sequencer so we can later identify when two cells have been captured together."

The algorithm, which is [open source and available](#) to other scientists, will hopefully empower investigators at Northwestern and elsewhere to produce more accurate RNA sequencing results, said Ziyang Zhang, a Ph.D. student in the DGP program and a co-first author of the study.

"For the longest time, most technologies did not allow us to know which cells are truly singlets. The beauty of this barcoding technology is that we leverage these unique nucleic acid sequences to recover singlets from a significant chunk of the data. That allows us to feed better data into a machine learning classifier and we show that it will achieve better performance," Zhang said. "That's something significant that we hope people will recognize and eventually adopt into their studies."

Next, Goyal and his laboratory will attempt to use the technology to measure [gene expression](#) in a sample and map where the activity is occurring.

"Going from single-cell RNA sequencing to spatial transcriptomics has become a big part of scientific discovery and Northwestern has invested quite a lot in this area," said Goyal, who is also a member of the Robert H. Lurie Comprehensive Cancer Center of Northwestern University.

"We are trying to understand how we can use these barcoding systems to interpret spatial biology."

More information: Ziyang Zhang et al, Synthetic DNA barcodes identify singlets in scRNA-seq datasets and evaluate doublet algorithms, *Cell Genomics* (2024). [DOI: 10.1016/j.xgen.2024.100592](https://doi.org/10.1016/j.xgen.2024.100592)

Provided by Northwestern University

Citation: Researchers improve measurement of gene expression in single cells (2024, June 26)
retrieved 29 June 2024 from <https://phys.org/news/2024-06-gene-cells.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.