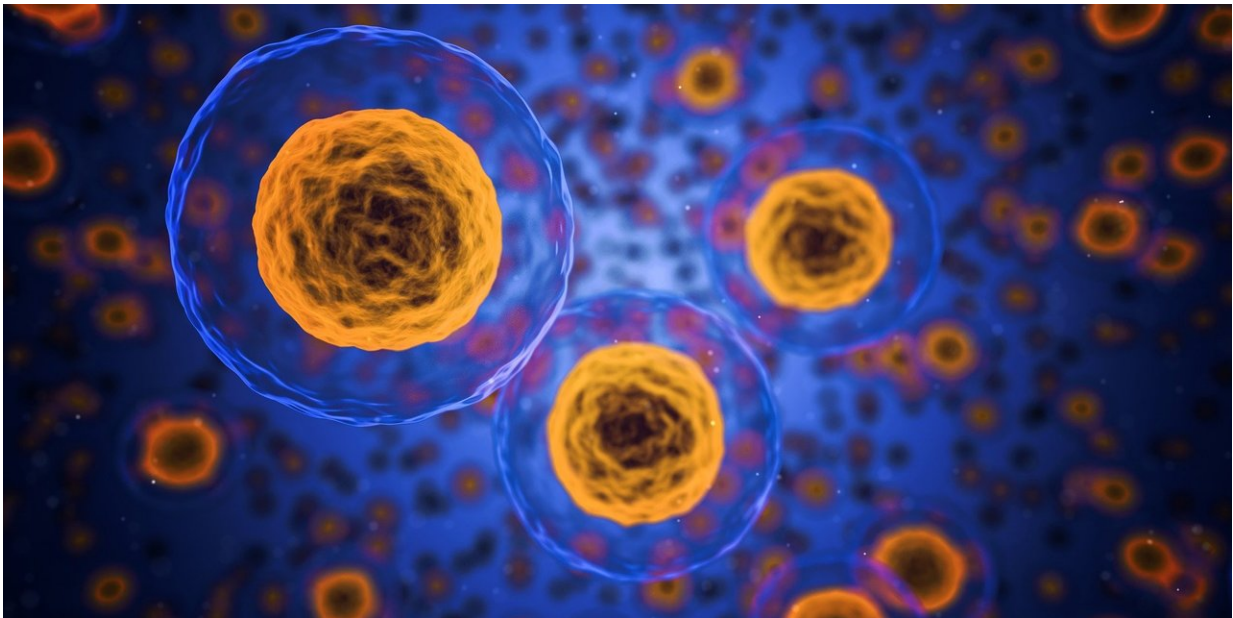


New algorithm enables data integration at single-cell resolution

April 2 2018



Credit: CC0 Public Domain

A team of computational biologists has developed an algorithm that can 'align' multiple sequencing datasets with single-cell resolution. The new method, published today in the journal *Nature Biotechnology*, has implications for better understanding how different groups of cells change during disease progression, in response to drug treatment, or across evolution.

"This approach for data integration will enable the comparison of single-cell datasets and the ability to dissect the differences between them," explains Rahul Satija, the study's senior author, who is an assistant professor in NYU's Center for Genomics and Systems Biology and a core faculty member at the New York Genome Center. "Moreover, these methods will be valuable for the integration of diverse datasets produced across individuals and laboratories—and even for researchers studying the same tissue across different species."

The field of single-cell sequencing is rapidly expanding, with the potential to precisely study how the basic building blocks of life function and evolve. However, significant computational challenges remain, particularly when analyzing multiple datasets. For example, when the team independently analyzed datasets of the same bone-marrow stem cells, produced by two separate labs, they obtained strikingly different results.

"We needed a new [method](#) that could identify and align shared groups of cells present in multiple experiments so that we could integrate the datasets together," says Andrew Butler, a graduate student at NYU and lead author of the study.

To accomplish this, the researchers modified analytical techniques specialized at finding shared patterns across images—for example, to align facial visualizations across different lighting conditions for single-cell sequencing data. When they repeated their bone-marrow analysis, the same cell populations consistently appeared.

"We realized that we could use these methods to learn how cells modify their behavior—for example, in response to [drug treatment](#)," notes Butler.

By analyzing a [dataset](#) of [human immune cells](#) stimulated with

interferon—a signaling protein created in response to pathogens or tumor [cells](#)—the team could precisely identify which genes were switched on in each of 13 responding cell types. Furthermore, they integrated single-cell datasets of pancreatic tissue from humans and mice, thereby identifying 10 cell types that were shared across species and defining the evolutionary changes occurring in each group.

Looking forward, the researchers are applying their approach to study cellular drug responses in clinical samples, but also aim to make their methods widely accessible.

"All of our software is open-source and freely available online," adds Satija. "We hope these methods will help others in the community discover exciting new biological phenomena."

More information: Integrating single-cell transcriptomic data across different conditions, technologies, and species, *Nature Biotechnology* (2018). [nature.com/articles/doi:10.1038/nbt.4096](https://doi.org/10.1038/nbt.4096)

Provided by New York University

Citation: New algorithm enables data integration at single-cell resolution (2018, April 2) retrieved 23 April 2024 from <https://phys.org/news/2018-04-algorithm-enables-single-cell-resolution.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.