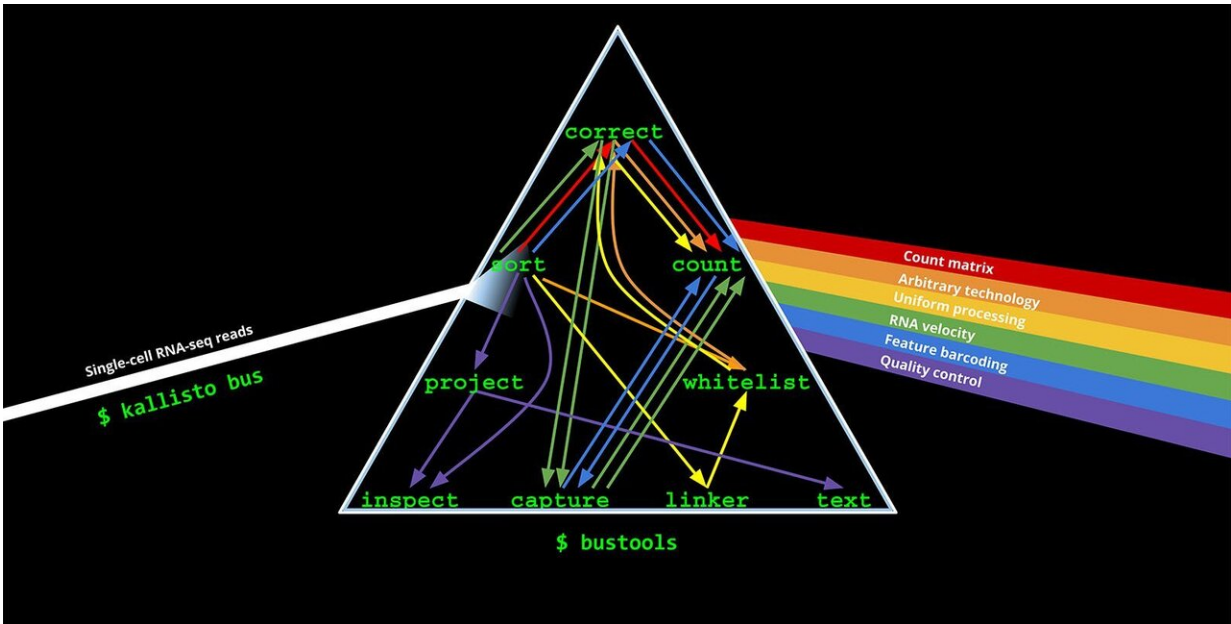


A Swiss army knife for genomic data

April 5 2021



Credit: California Institute of Technology

A good way to find out what a cell is doing—whether it is growing out of control as in cancers, or is under the control of an invading virus, or is simply going about the routine business of a healthy cell—is to look at its gene expression. Though a vast majority of cells in an organism all contain the same genes, how those genes are expressed is what gives rise to different cell types—the difference between a muscle cell and a neuron, for example.

In the last decade, technologies to measure [gene expression](#) in individual cells have revolutionized biology. No longer do biologists need to average out gene expression over many cells within tissues; now they can detect which [genes](#) are active in each cell at any time.

Computational power has struggled to keep up with this explosion of data, however. For example, a single experiment can look at 100,000 [cells](#) and measure information from hundreds of thousands of transcripts (fragments of RNA produced when a gene is active), resulting in tens of billions of sequenced fragments. Genomic data from single-cell sequencing can take up terabytes of space and take hours or days to process on large computing servers.

Now, a new software [tool](#) enables the processing of large sets of [genomic data](#) in about 30 minutes, using the computing power of an average laptop. Like a Swiss Army knife, the tool can be used in myriad ways for different biological needs, and will help ensure the reproducibility of scientific studies.

The tool, which is available online and open for anyone to use, now is being adapted by another research team to study the SARS-CoV-2 virus in samples collected from screening tests.

The research was conducted as a collaboration between the laboratory of Lior Pachter (BS '94), Bren Professor of Computational Biology and Computing and Mathematical Sciences, and Páll Melsted, professor of computer science at the University of Iceland. Melsted is a co-first author along with graduate student Sina Boeshaghi (MS '19). A paper describing the research appears in the journal *Nature Biotechnology* on April 1.

"There are many examples of different groups using different technologies to study the same tissues, for example, the brain," says

Booeshaghi. "Processing all of this data with the same engine—our technique—facilitates integrating the data. Our tool is fast, efficient, and allows for easy reprocessing, which is very important for consistency and reproducibility in science."

Developing this complex software tool "in-house" was important for it to actually address potential users' concerns, because the potential users were right there in the lab.

"The interdisciplinarity of our team was crucial to conceiving of and executing this project," says Pachter. "There are people in the lab who are computer scientists, biologists, engineers. Sina is in the mechanical engineering department and brings the perspective of his design background and engineering; Páll has a strong background in theoretical computer science and software engineering."

The ease-of-use, low cost, and modularity of these tools will enable consistent and reproducible preprocessing of genomic data for large consortiums such as the Human Cell Atlas and the Brain Initiative Cell Census Network.

The paper is titled "Modular, fast, and constant-memory pre-processing of single-cell RNA-seq data."

More information: Páll Melsted et al. Modular, efficient and constant-memory single-cell RNA-seq preprocessing, *Nature Biotechnology* (2021). [DOI: 10.1038/s41587-021-00870-2](https://doi.org/10.1038/s41587-021-00870-2)

Provided by California Institute of Technology

Citation: A Swiss army knife for genomic data (2021, April 5) retrieved 22 March 2023 from

<https://phys.org/news/2021-04-swiss-army-knife-genomic.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.