

New search engine for single cell atlases

March 3 2021



Credit: CC0 Public Domain

A new software tool allows researchers to quickly query datasets generated from single-cell sequencing. Users can identify which cell types any combination of genes are active in. Published in *Nature Methods* on 1st March, the open-access 'scfind' software enables swift analysis of multiple datasets containing millions of cells by a wide range of users, on a standard computer.

Processing times for such datasets are just a few seconds, saving time and computing costs. The tool, developed by researchers at the Wellcome Sanger Institute, can be used much like a [search engine](#), as users can input free text as well as gene names.

Techniques to sequence the genetic material from an individual cell have advanced rapidly over the last 10 years. Single-cell RNA sequencing (scRNAseq), used to assess which genes are active in individual [cells](#), can be used on millions of cells at once and generates vast amounts of data (2.2 GB for the Human Kidney Atlas). Projects including the Human Cell Atlas and the Malaria Cell Atlas are using such techniques to uncover and characterize all of the cell types present in an organism or population. Data must be easy to access and query, by a wide range of researchers, to get the most value from them.

To allow for fast and efficient access, a new software tool called scfind uses a two-step strategy to compress data ~100-fold. Efficient decompression makes it possible to quickly query the data. Developed by researchers at the Wellcome Sanger Institute, scfind can perform large scale analysis of datasets involving millions of cells on a standard computer without special hardware. Queries that used to take days to return a result, now take seconds.

The new tool can also be used for analyzes of multi-omics data, for example by combining single-cell ATAC-seq data, which measures epigenetic activity, with scRNAseq data.

Dr. Jimmy Lee, Postdoctoral Fellow at the Wellcome Sanger Institute, and lead author of the research, said: "The advances of multiomics methods have opened up an unprecedented opportunity to appreciate the landscape and dynamics of gene regulatory networks. Scfind will help us identify the genomic regions that regulate gene activity—even if those regions are distant from their targets."

Scfind can also be used to identify new genetic markers that are associated with, or define, a cell type. The researchers show that scfind is a more accurate and precise method to do this, compared with manually curated databases or other computational methods available.

To make scfind more user friendly, it incorporates techniques from natural language processing to allow for arbitrary queries.

Dr. Martin Hemberg, former Group Leader at the Wellcome Sanger Institute, now at Harvard Medical School and Brigham and Women's Hospital, said: "Analysis of single-cell datasets usually requires basic programming skills and expertise in genetics and genomics. To ensure that large single-cell datasets can be accessed by a wide range of users, we developed a tool that can function like a search engine—allowing users to input any query and find relevant cell types."

Dr. Jonah Cool, Science Program Officer at the Chan Zuckerberg Initiative, said: "New, faster analysis methods are crucial for finding promising insights in single-cell data, including in the Human Cell Atlas. User-friendly tools like scfind are accelerating the pace of science and the ability of researchers to build off of each other's work, and the Chan Zuckerberg Initiative is proud to support the team that developed this technology."

More information: Jimmy Tsz Hang Lee et al, Fast searches of large collections of single-cell data using scfind, *Nature Methods* (2021). [DOI: 10.1038/s41592-021-01076-9](https://doi.org/10.1038/s41592-021-01076-9)

Provided by Wellcome Trust Sanger Institute

Citation: New search engine for single cell atlases (2021, March 3) retrieved 28 April 2024 from

<https://phys.org/news/2021-03-cell-atlases.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.