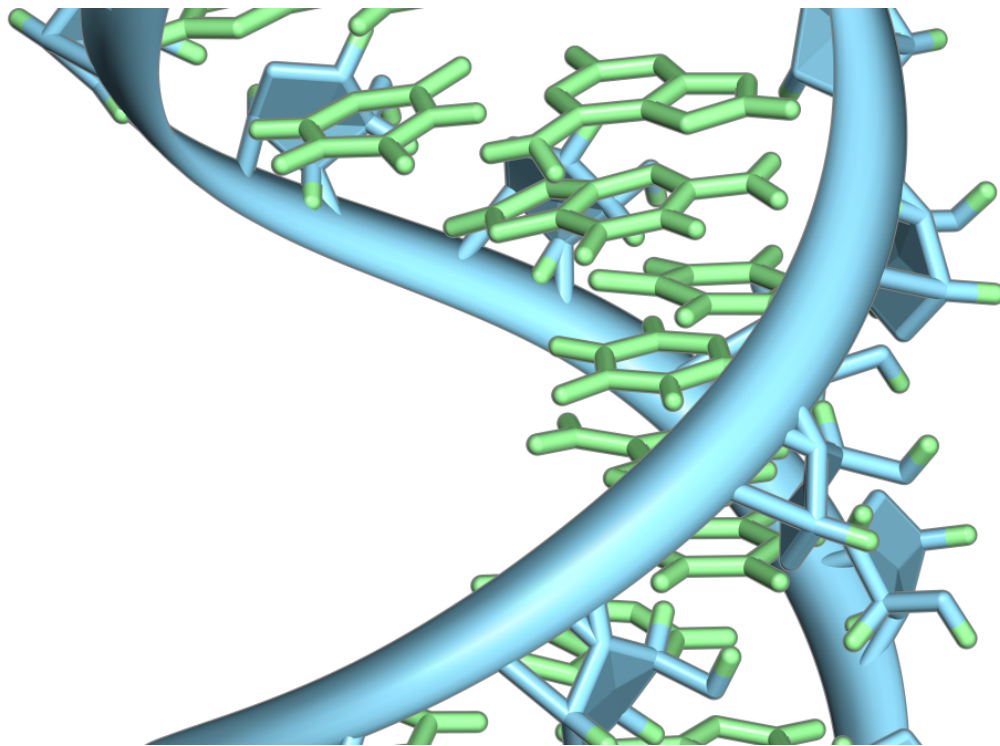


Neural nets supplant marker genes in analyzing single cell RNA sequencing

November 13 2018



A hairpin loop from a pre-mRNA. Highlighted are the nucleobases (green) and the ribose-phosphate backbone (blue). Note that this is a single strand of RNA that folds back upon itself. Credit: Vossman/ Wikipedia

Computer scientists at Carnegie Mellon University say neural networks and supervised machine learning techniques can efficiently characterize cells that have been studied using single cell RNA-sequencing (scRNA-

seq). This finding could help researchers identify new cell subtypes and differentiate between healthy and diseased cells.

Rather than rely on marker genes, which are not available for all cell types, this new automated method analyzes all of the scRNA-seq data to select just those parameters that can differentiate one cell from another. This enables the analysis of all cell types and provides a method for comparative analysis of those [cells](#).

Researchers from CMU's Computational Biology Department explain their method today in the online journal *Nature Communications*. They also describe a web server called scQuery that makes the method usable by all researchers.

Over the past five years, single cell sequencing has become a major tool for cell researchers. In the past, researchers could only obtain DNA or RNA sequence information by processing batches of cells, providing results that only reflected average values of the cells. Analyzing cells one at a time, by contrast, enables researchers to identify subtypes of cells, or to see how a healthy cell differs from a diseased cell, or how a young cell differs from an aged cell.

This type of sequencing will support the National Institutes of Health's new Human BioMolecular Atlas Program (HuBMAP), which is building a 3-D map of the human body that shows how tissues differ on a cellular level. Ziv Bar-Joseph, professor of [computational biology](#) and machine learning and a co-author of today's paper, leads a CMU-based center contributing computational tools to that project.

"With each experiment yielding hundreds of thousands of data points, this is becoming a Big Data problem," said Amir Alavi, a Ph.D. student in computational biology who was co-lead author of the paper with post-doctoral researcher Matthew Ruffalo. "Traditional analysis methods are

insufficient for such large scales."

Alavi, Ruffalo and their colleagues developed an automated pipeline that attempts to download all public scRNA-seq data available for mice—identifying the genes and proteins expressed in each cell—from the largest data repositories, including the NIH's Gene Expression Omnibus (GEO). The cells were then labeled by type and processed via a neural network, a computer system modeled on the human brain. By comparing all of the cells with each other, the neural net identified the parameters that make each cell distinct.

The researchers tested this model using scRNA-seq data from a mouse study of a disease similar to Alzheimer's. As would be expected, the analysis showed similar levels of [brain cells](#) in both healthy and diseased cells, while the [diseased cells](#) included substantially more [immune cells](#), such as macrophages, generated in response to the disease.

The researchers used their pipeline and methods to create scQuery, a web server that can speed comparative analysis of new scRNA-seq data. Once a researcher submits a single cell experiment to the server, the group's [neural networks](#) and matching methods can quickly identify related cell subtypes and identify earlier studies of similar cells.

More information: Amir Alavi et al, A web server for comparative analysis of single-cell RNA-seq data, *Nature Communications* (2018). [DOI: 10.1038/s41467-018-07165-2](https://doi.org/10.1038/s41467-018-07165-2)

Provided by Carnegie Mellon University

Citation: Neural nets supplant marker genes in analyzing single cell RNA sequencing (2018, November 13) retrieved 20 June 2024 from <https://phys.org/news/2018-11-neural-nets-supplant->

[marker-genes.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.