New AI-based tool can find rare cell populations in large single-cell datasets

May 20 2021

Credit: CC0 Public Domain

Researchers at The University of Texas MD Anderson Cancer Center have developed a first-of-its-kind artificial intelligence (AI)-based tool that can accurately identify rare groups of biologically important cells

from single-cell datasets, which often contain gene or protein expression data from thousands of cells. The research was published today in *Nature Computational Science*.

This computational tool, called SCMER (Single-Cell Manifold presERving feature selection), can help researchers sort through the noise of complex datasets to study cells that would likely not be identifiable otherwise.

SCMER may be used broadly for many applications in oncology and beyond, explained senior author Ken Chen, Ph.D., associate professor of Bioinformatics & Computational Biology, including the study of minimal residual disease, drug resistance and distinct populations of immune cells.

"Modern techniques can generate lots of data, but it has become harder to determine which genes or proteins actually are important in those contexts," Chen said. "Small groups of cells can have important features that may play a role in drug resistance, for example, but those features may not be sufficient to distinguish them from more common cells. It's become very important in analyzing single-cell datasets to be able to detect these rare cells and their unique molecular features."

Developing methods to effectively study small or rare cell populations in cancer research is a direct response to one of the provocative questions posed by the National Cancer Institute (NCI) in 2020, designating this an important and underexplored research area. SCMER was designed to address the issue and to enable researchers to get the most out of increasingly complex datasets.

Rather than the traditional approach of sorting cells into clusters based on all data contained in a dataset, SCMER takes an unbiased look to detect the most meaningful distinguishing features that define unique

groups of cells. This allows researchers not only to detect rare cell populations, but to generate a compact set of genes or proteins that can be used to detect those cells among many others. To highlight the utility of SCMER, the research team applied it to analyze several published single-cell datasets and found it compared favorably to currently available computational approaches.

In a reanalysis of more than 4,500 melanoma cells, SCMER was able to distinguish the cell types present using the expression of just 75 genes. The results also pointed to a number of genes involved in tumor development and drug resistance that were not identified as meaningful in the original study.

In a complex dataset of nearly 40,000 gastrointestinal immune cells, SCMER separated cells using only 250 distinct features. This analysis identified all the original cell types detected in the original study, but in many cases further defined subgroups of rare cells that were not previously identified.

Finally, the research team applied SCMER to study more than 1,400 lung cancer cells taken at various points in time after drug treatment. Using just 80 genes, the tool was able to accurately distinguish cells based on treatment responses and pointed to possible novel drivers of therapeutic resistance.

"Using state-of-the-art AI techniques, we have developed an efficient and user-friendly tool capable of uncovering new biological insights from rare cell populations," Chen said. "SCMER offers researchers the ability to reduce highly dimensional, complex datasets into a compact set of actionable features with biological significance."

  More information: Single-cell manifold-preserving feature selection for detecting rare cell populations, *Nature Computational Science*, DOI:

Provided by University of Texas M. D. Anderson Cancer Center

Citation: New AI-based tool can find rare cell populations in large single-cell datasets (2021, May 20) retrieved 26 June 2024 from https://phys.org/news/2021-05-ai-based-tool-rare-cell-populations.html