

A single number helps data scientists find most dangerous cancer cells

January 24 2020, by Christopher Vaughan



Credit: CC0 Public Domain

Stanford data scientists have shown that figuring out a single number can help them find the most dangerous cancer cells.

Biomedical data scientists at the Stanford University School of Medicine

have shown that the number of genes a cell uses to make RNA is a reliable indicator of how developed the cell is, a finding that could make it easier to target [cancer](#)-causing genes.

Cells that initiate cancer are thought to be stem cells, which are hard-to-find cells that can reproduce themselves and develop, or differentiate, into more specialized tissue, such as skin or muscle—or, when they go bad, into cancer.

"Right now, targeted therapies are focused on [specific genes](#) or molecules, the vast majority of which may not be specific to cancer stem cells," said Aaron Newman, Ph.D., assistant professor of biomedical data science and a member of the Institute for Stem Cell Biology and Regenerative Medicine. "Usually these therapies don't work for very long. But if you can identify the least-differentiated cells and then look for markers specific to them, it's no longer a guessing game to find the genes to target."

The study's finding is also significant because identifying stem cells of various tissue types is an important step toward regenerating damaged or malfunctioning tissues.

What the scientists showed is that as stem cells become more differentiated and more like adult cells, they express fewer and fewer genes. Previously, other researchers had noticed this correlation and thought it might be an interesting coincidence. But Newman and his colleagues were the first to sort through thousands of single-cell genetic tests in public databases and prove this pattern was consistent and reliable.

Newman and MD-Ph.D. student Gunsagar Gulati combined the measurement of the number of [genes](#) expressed in a cell with the measurement of the number of RNA copies created per gene as the basis

for a computer algorithm, CytoTRACE, designed to determine how developmentally advanced cells are.

A paper describing the research has been published online today in *Science*. Newman is the senior author. Gulati and Shaheen Sikandar, Ph.D., an instructor at the institute, share lead authorship.

Tumor cells are diverse

Cancerous tumors can contain many millions of cells, each of which may have thousands of gene mutations. The cells in a tumor are diverse. Most will be differentiated cells that die out naturally on their own, while relatively few are the more dangerous cancer stem cells, or tumor-initiating cells. These cells are hard to find and therefore hard to characterize using current methods, but far easier to find with CytoTRACE.

"As a cancer researcher, what I find most exciting is that this tool helps us find the tumor-initiating cells that have long been known to be responsible for resistance to treatment, metastasis and relapse after treatment," Sikandar said.

Michael Clarke, MD, one of the authors of the paper, was the first researcher to identify cancer stem cells in a solid tumor. A professor of medicine at Stanford, Clarke said that CytoTRACE, which analyzes data on all the RNA created in a single cell, can quickly recapitulate research that takes years using traditional methods. "The way that we currently find cell markers for cancer stem cells is to make educated guesses about which markers will likely be important, then sort those cells and look for stem cell activity," said Clarke, the Karel H. and Avice N. Beekhuis Professor in Cancer Biology and associate director of the Institute for Stem Cell Biology and Regenerative Medicine.

Researchers can look at relatively few markers at a time, so it takes a lot of sorting and analysis, and in the end, they will likely be only partially successful in finding good markers of the stem cells they are looking for, he said. "What CytoTRACE allows us to do is first find the stem or progenitor cells, then look at what unique markers they have on them."

In the paper, the researchers describe using CytoTRACE to query single-cell RNA data for triple-negative breast cancer, a type of tumor that is rarer but more dangerous because tumor growth doesn't rely on the biochemical pathways that physicians usually target to treat breast cancer. Not only did CytoTRACE identify known markers of [cancer stem cells](#), it also spotted a marker that had not been previously been thought to be important. "This one gene looks like it has amazing potential as a therapeutic," Clarke said.

Potential tool for hunting other disease-linked stem cells

CytoTRACE also has the potential to transform how researchers hunt for stem cells associated with other diseases, Newman said. "This tool could also be useful in finding treatments for disorders such as Alzheimer's or other degenerative diseases where loss of stem cell function might be part of the disease process," he said.

Regenerative medicine, in which diseased or damaged tissue is repaired through the activity of stem cells, requires the ability to isolate purified populations of stem cells specific to a given tissue. To regrow bone, the heart or the eyes, for example, researchers must first find the stem cells responsible for regrowing those organs. Finding the markers that are specific to these normal stem [cells](#) has been much like the process for finding cancer stem cell markers, the researchers say—that is, the product of educated guesses, luck and a lot of work in the lab.

CytoTRACE could significantly shorten that process.

"One of the main motivations behind developing CytoTRACE was to create a tool for rapid and accurate identification of [stem cells](#) in humans," Gulati said. "But another important question we hope to answer is how the inner workings of a cell change as the cell transforms from one state to another. This research opens up a whole new avenue of research to study how global changes in gene expression and DNA structure influence a cell's state."

Overall, Newman said, the study shows the power and promise of using big data to advance biology and medicine through computer research that complements discoveries made in the lab.

"It wouldn't have been possible to gather all this data in our lab, but by using public databases and asking the right questions, it's more and more possible to make fundamental discoveries in biology and medicine," he said.

More information: Gunsagar S. Gulati et al. Single-cell transcriptional diversity is a hallmark of developmental potential, *Science* (2020). [DOI: 10.1126/science.aax0249](https://doi.org/10.1126/science.aax0249)

Provided by Stanford University Medical Center

Citation: A single number helps data scientists find most dangerous cancer cells (2020, January 24) retrieved 13 March 2024 from <https://medicalxpress.com/news/2020-01-scientists-dangerous-cancer-cells.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.
