Researchers develop a better method to compare gene expression in single cells

June 20 2018

Computational biologist and corresponding author Xiang Chen, Ph.D., (right) of St. Jude, and his colleagues (first author Wenan Chen, Ph.D., pictured left) developed NBID to take better advantage of single-cell RNA sequencing to track differences in gene expression in individual cells. Credit: St. Jude Children's Research Hospital

Efforts to capitalize on next-generation sequencing to compare gene

expression in individual cells for clues about cancer's origins, progression or relapse just got a boost. St. Jude Children's Research Hospital researchers have developed an algorithm that provides a more accurate and sensitive method of identifying differences in gene expression in individual cells.

The algorithm is called negative binomial model with independent dispersions or NBID. St. Jude is providing NBID at no charge to researchers worldwide. Computational biologist and corresponding author Xiang Chen, Ph.D., of St. Jude, and his colleagues developed NBID to take better advantage of single-cell RNA sequencing to track differences in gene expression in individual cells. Their work appeared online recently in the journal *Genome Biology*.

Single-cell RNA sequencing has emerged in the last decade and gained popularity for the study of cancer and development of the immune system and other organs. By comparing gene expression in different cells, researchers aim to improve our understanding of cancer genetics. Scientists use the technology to find tumor cell subpopulations that are chemotherapy resistant or that represent rare subtypes. The information may also reveal corresponding marker genes, which are defined as genes with different expression levels between populations. Such information would aid efforts to develop precision medicines and more sensitive diagnostic tests.

"Numerous studies now employ single-cell RNA sequencing techniques, but statistical methods to characterize the data lag," said Chen, an assistant member of the St. Jude Department of Computational Biology. "We created NBID, a software package developed specifically for analyzing single-cell RNA sequencing data. We showed that NBID provides a more accurate and sensitive analysis of differential gene expression compared to other software packages developed for analyzing single-cell RNA sequencing data.

"We believe NBID will prove useful in identifying biomarkers for other in-depth sequencing data evaluation as well."

Challenges

The human genome includes 20,000 to 25,000 genes that carry instructions for making specific proteins that do most of the work in cells. The process requires DNA to be copied by messenger RNA, from which it is translated into a specific protein.

Single-cell RNA sequencing requires researchers to capture messenger RNA within single cells, use the messenger RNA to assemble the complementary strand of DNA, which is then copied (amplified) and analyzed.

Gene expression varies widely and fluctuates within cells. Capturing messenger RNA for genes with low- to-moderate expression in individual cells is particularly challenging. Another challenge is data sparsity or low signal and high noise, which requires identifying data of interest, in this case RNA, in a sea of noise. Examples include "drop-out" events in which genes expressed at relatively high levels in a subset of cells are undetectable in other cells.

Chen and his colleagues used molecular "barcodes" to track gene expression by tagging and then tallying messenger RNAs using a process called unique molecular identifier (UMI) counting.

"The advantages of UMI counts over another method, read counts, in quantification of RNA have been well documented. The statistical difference between these two schemes had been underappreciated," Chen said. "Upon extensive evaluation of single-cell RNA sequencing data, we revealed that these two approaches should be modelled differently and UMI count could be approximated by the negative

binomial model."

NBID allowed gene-specific and group-specific negative binomial models, resulting in better performance. In comparison tests, NBID proved more sensitive and more accurate in recognizing differences in gene expression between different groups of cells. For example, NBID helped researchers identify marker genes that can be used to separate subpopulations of rhabdomyosarcoma cells with distinct gene expression patterns, which suggested a potentially novel mechanism of the solid tumor progression.

More information: Wenan Chen et al. UMI-count modeling and differential expression analysis for single-cell RNA sequencing, *Genome Biology* (2018). DOI: 10.1186/s13059-018-1438-9

Provided by St. Jude Children's Research Hospital

Citation: Researchers develop a better method to compare gene expression in single cells (2018, June 20) retrieved 18 April 2024 from https://phys.org/news/2018-06-method-gene-cells.html