New method for analyzing genes activity helps predict cancer patients survival

May 28 2019

Konstantin Zaitsev. Credit: ITMO University

An international research team has developed a new method for determining cell types in a tissue sample. The scientists determined the link between the activity of genes in the same cell type and made a

model capable of "recognizing" different cell types in mixed samples based on this relation. This approach works for all tissues, so it can be used to understand the proportions of which cell types are associated with the survival of patients with different types of cancer, for example. The results are published in *Nature Communications*.

The analysis of a transcriptome, or the set of all RNA molecules in set of samples, is widely used in biomedical research. Using this method, researchers may analyze molecular processes in a tissue and characterize the severity of cancer, for instance. However, tissue samples can contain millions of cells that need to be distinguished from each other in order to understand what is happening.

Working on this problem, scientists have developed special deconvolution algorithms. They provide an opportunity to decompose the data and match it to different cell types. It helps to understand what cell types are present in the sample, what their proportion is, and how it affects the transcript. However, when the sample contains many cell types, it becomes difficult to identify all of them without any additional information.

The international research group from ITMO University and the University of Washington in St. Louis, U.S., found a way to overcome this obstacle and proposed the new method for analyzing transcriptome samples. It can determine with high accuracy which cell types the samples contains based on the genes' mutual linearity principle: The expression levels of two genes specific for the same cell type linearly depend on each other. The scientists used this relation to construct networks of such linearly dependent genes. By analyzing such networks, researchers can determine what cells are in the samples.

Scientists have shown that all deconvolution algorithms are subject to the same bias: if different cell types in the sample have different RNA

amounts, all deconvolution algorithms cannot estimate cell type proportions accurately. To test this experimentally, two types of cells with different amounts of RNA were selected and mixed in different predetermined proportions. After that, researchers used deconvolution algorithms to determine cell types ratios.

"We saw that existing algorithms would always be wrong about number of cells since they estimate the amount of RNA in the samples. Yet, if measurements are made after adding a particular amount of artificial RNA to each sample, the predicted cell type proportions can be corrected to become more accurate," explains Konstantin Zaitsev, researcher at the Laboratory of Computer Technologies of ITMO University.

"Our approach is best suited for analyzing mixed samples without sufficient information about their composition. As long as the method doesn't require any hints, it can be used for any tissue type. For example, it can detect differences in cell composition after vaccination in blood samples. Moreover, using the TCGA public database (the Cancer Genome Atlas), we are already trying to identify cell types associated with survival of patients with different cancers," Konstantin concludes.

Provided by ITMO University