

Free online tool to provide deeper analysis of microarray data

July 18 2012

Imagine listening to a child plinking out a rudimentary tune on the piano. He uses only one octave and one hand — notes he can't reach are skipped. You can pick out the basic melody, but just barely. But you have an inkling of what the full piece could sound like.

Until now, this was the experience of researchers attempting to piece together complex biological processes by analyzing changes in the levels of expression of genes in cells. Limitations with the technology, called DNA microarray analysis, meant that results from separate experiments couldn't be directly compared; scientists could hear only one line of the cellular melody.

Now a new software program designed by scientists at the Stanford University School of Medicine enables researchers to see the whole picture of gene expression in a sample. The program, called the Gene Expression Commons, is publicly available at <https://gexc.stanford.edu/>. Its developers expect it to transform studies of gene expression.

"It is so simple that a researcher can just type in the name of a gene and, within seconds, see the absolute level of the expression of that gene in every cell type in a panel," said professor of pathology Irving Weissman, MD. "We believe that this program will rapidly become the most important tool for discovery in a number of fields, including stem cells, cancer and regenerative medicine." Viewing these results in concert gives a much greater understanding of how the cells develop and function — like an orchestra revealing the complex interplay among the

many musical lines and themes missing in the child's rendition of the piece.

Weissman, who is the director of Stanford's Institute for Stem Cell Biology and Regenerative Medicine, is the senior author of the research, which will be published July 18 in *PLoS ONE*. He is also the Virginia & D.K. Ludwig Professor for Clinical Investigation in Cancer Research at Stanford, and a member of the Stanford Cancer Institute. Jun Seita, MD, PhD, instructor of pathology, is the first author of the study.

The Gene Expression Commons overcomes an inherent shortcoming of microarray technology: the fact that experimental results are delivered as relative differences in gene expression within individual experiments — like lone melodies — rather than absolute values that can be compared in concert among many samples.

"In the past, we've been limited to just comparing relative differences in expression levels among genes, perhaps between a normal and a cancer cell, or a stem or progenitor cell," said Seita. "But if a researcher wanted to profile the absolute levels of expression among many genes in each cell type, it's been very difficult."

Seita and co-author Debashis Sahoo, PhD, hit on the idea of analyzing collections of thousands of publicly available DNA microarray experiments. About 25,000 of the experiments had been performed with human data; 10,000 with data from mice. Individually, each data set suffered from the same drawbacks described above. But together they can be viewed as a continuum, or a stable common reference.

The resulting Gene Expression Commons maps data submitted by the user onto this common reference, and returns absolute expression levels that can then be compared among many combinations of samples.

To test their idea, the Weissman group performed and submitted microarray data from 39 highly purified, distinct cell types in the blood and immune system to the program. Now any researcher can explore the expression pattern of any gene in the system with just a few clicks of a computer mouse.

"The potential analyses are very powerful," said Sahoo.

Microarray technology was developed at Stanford in the 1990s. At its heart, it relies on the fact that single, complementary (or "matching") strands of nucleotides are driven to bind lengthwise to one another, like microscopic zippers. In microarrays, scientists affix thousands of tiny dots of specific nucleotide sequences — each representing a different gene — to glass slides in precise patterns, or arrays. Researchers apply a sample of interest to each slide and then can assess the relative levels of expression of each sequence in these samples. (Binding to a target sequence is indicated by a fluorescent signal that varies in intensity with the number of binding events at that location.)

However, because some sequences will inherently bind to their targets more or less strongly than others, it's not possible to directly compare signal intensities among different spots on the same chip. So researchers could learn that genes X and Y were both expressed at higher levels in one sample than in another, but they had no way of knowing how the absolute levels of X and Y compared.

The distinction can be biologically important. For example, you might not think twice about a doubling in price at the grocery store of your favorite gum from 50 cents to \$1, but an identical increase in the price of hamburger (from \$5 to \$10 per pound), may have a significant impact in your budget and the outcome of your shopping trip. So too it goes with biological systems: a difference in the absolute number of RNA molecules from 10 to 20 would appear similar to an increase from

10,000 to 20,000, although the latter is far more likely to more-profoundly affect cellular function.

"This has been a huge limitation for researchers," said Seita. "We say we're profiling gene expression levels, but that's not really what we've been doing. Instead we've been profiling the relative differences in gene expression, and we had no way of learning more."

Traditional microarray technology also doesn't identify ranges of sensitivity: if expression levels are very high, for example, the signal will appear maximally intense regardless of true differences among samples. The new software overcomes this by merging data from thousands of experiments.

"For each gene, we get about 30,000 values," said Sahoo. "This gives us an idea of the low and high range of expression in many cell types. We can get a sense of the range of the values, and then find where any individual sample fits within this continuum."

"This method takes this raw data from individual microarrays and normalizes them against a global database, or common reference," said Weissman. "It gives an instant readout of absolute expression levels. We can now quickly identify genes that expressed only in blood stem cells with just three clicks of the mouse."

The Gene Expression Commons is designed as an open platform, so new users can input data themselves. A neurologist, for example, may want to analyze data from microarrays conducted on cells of the nervous system. And the Gene Expression Commons is open for more than DNA microarray technology. "This 'common reference' strategy should work well for any type of high-throughput data," said Seita. "We really acknowledge the effort to share published microarray data across the scientific community, so it's time to give back."

Provided by Stanford University Medical Center

Citation: Free online tool to provide deeper analysis of microarray data (2012, July 18) retrieved 26 April 2024 from

<https://medicalxpress.com/news/2012-07-free-online-tool-deeper-analysis.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.