

Computational feat speeds finding of genes to milliseconds instead of years

March 15 2010

Like a magician who says, "Pick a card, any card," Stanford University computer scientist Debashis Sahoo, PhD, seemed to be offering some kind of trick when he asked researchers at the Stanford Institute for Stem Cell Biology and Regenerative Medicine to pick any two genes already known to be involved in stem cell development. Finding such genes can take years and hundreds of thousands of dollars, but Sahoo was promising the skeptical stem cell scientists that, in a fraction of a second and for practically zero cost, he could find new genes involved in the same developmental pathway as the two genes provided.

Sahoo went on to show that this amazing feat could actually be performed. The proof-of-principle for his idea, to be published online March 15 in the [Proceedings of the National Academy of Sciences](#), opens a powerful, mathematical route for conducting stem cell research and shows the power of interdisciplinary collaborations in science. It also demonstrates that using computers to mine existing databases can radically accelerate research in the laboratory. Ultimately, it may lead to advances in diverse areas of medicine such as disease diagnosis or [cancer therapy](#).

Biologists have long used math and statistics in their work. In the simplest case, when looking for [genes](#) involved in a certain biological process, they look for genes that have a symmetrical correlation. For instance, if they know gene A is involved in a certain process, they try to determine if gene C is correlated with gene A during the same process.

Four years ago, while studying for his doctorate in electrical engineering with advisor David Dill, PhD, professor of computer science, and co-advisor Sylvia Plevritis, PhD, associate professor of radiology, Sahoo took an immunology class and realized that many of the relationships in biology are not symmetric, but asymmetric. As an analogy, Sahoo noted that trees bearing fruit almost certainly have leaves, but trees outside of the fruiting season may or may not have leaves, depending on the time of year.

Sahoo and Dill realized that these asymmetric relationships could be found by applying Boolean logic, in which the researchers established a series of if/then rules and then searched data for candidates that satisfied all the rules. For example, scientists might know that gene A is very active at the beginning of cell development, and gene C is active much later. By screening large public databases, Sahoo can find the genes that are almost never active when A is active, and almost always active when C is active, in many other types of cells. Researchers can then test to determine whether these genes become active between the early and late stages of development.

In the paper, lead author Sahoo looked at gene expression patterns in the development of an immunological cell called a B cell. Starting with two known B-cell genes, Sahoo searched through databases with thousands of gene products in milliseconds and found 62 genes that matched the patterns he would expect to see for genes that got turned on in between the activation of the two genes he started with. He then examined databases involving 41 strains of laboratory mice that had been engineered to be deficient in one or more of the 62 genes. Of those 41 strains, 26 had defects in B cell development.

"This was the validation of the method," Sahoo said. "Biologists are really amazed that, with just a computer algorithm, in milliseconds I can find genes that it takes them a really long time to isolate in the lab." He

added that he was especially gratified that the information comes from databases that are widely available and from which other scientists have already culled information.

Sahoo is now using the technique to find new genes that play a role in developing cancers.

"This shows that computational analysis of existing data can provide clues about where researchers should look next," he said. "This is something that could have an impact on cancer. It's exciting."

Provided by Stanford University

Citation: Computational feat speeds finding of genes to milliseconds instead of years (2010, March 15) retrieved 30 April 2024 from <https://medicalxpress.com/news/2010-03-feat-genes-milliseconds-years.html>

<p>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.</p>
--