Stanford/Packard researchers find disease genes hidden in discarded data

October 24 2007

Previously hidden obesity-related genes have been uncovered from old experiments by researchers at the Stanford University School of Medicine and Lucile Packard Children's Hospital. The finding suggests that useful information about many medical disorders may be languishing in mountains of discarded data.

"We've devised a fairly simple way to convert large amounts of existing raw data into candidate disease genes for further genetic study," said Atul Butte, MD, PhD, a pediatrician at Packard Children's and director of the hospital's Center for Pediatric Bioinformatics. "When we put the information together, we were not only able to pinpoint those that have already been identified, but we also came up with some very interesting new predictions."

The investigators teased out the existence of more than a dozen new obesity-related genes by comparing the results of 49 independent experiments conducted by other researchers - none of which had yielded similar results on their own.

Butte, who is also an assistant professor of medicine and of pediatrics at the medical school, plans to investigate the biological roles of the new genes soon. The research appears in the Oct. 5 advance access section of the journal Bioinformatics.

Identifying novel genetic culprits for complicated diseases like obesity, diabetes and autism is tricky. Unlike cystic fibrosis, which is caused by a

mutation in just one gene, these conditions are often the result of a "perfect storm" of interacting genes and environmental factors. This complexity leaves researchers with limited time to pursue only their most promising results, leaving other candidates behind.

Managing the unused data can be extremely challenging. Microarray or gene-chip experiments, for example, generate tens of thousands of pieces of information. Because most scientific journals require the authors to submit all of their data to publicly available international databases, Butte estimates that the volume of such data is doubling or tripling each year.

Butte and his colleague, postdoctoral student Sangeeta English, PhD, re-analyzed publicly available data from 49 experiments conducted using different methods in a variety of animals from humans to rats to worms. They cast a wide net: The only thing the studies had in common was that they were each designed to ferret out genes or proteins important to fat storage or body size.

"We don't make any assumptions," said Butte. "We trusted the individual investigators to come up with well-thought-out models for their experiments. What we may lose in precision - by, for example, overlooking species-specific differences - we gain in the ability to generalize. Those genes that we do identify as important are likely to be of fundamental importance."

For example, one experiment focused on an extremely rare pediatric disorder called progeria. Children with the condition appear to age rapidly and usually die in their early teens. They also happen to lose their fat cells. "Now, we don't know if this has something to do with obesity," said Butte. "But if it's at all related to fat metabolism, it may contribute something to our knowledge."

Butte and English mixed and matched pairs and small groups of experiments to identify reliable performers. Their premise was that a gene that is only weakly positive in one experiment may easily be dismissed out of hand. However, if that same gene is weakly positive in two or more experiments, the case against it becomes much stronger - particularly if those experiments used very different methods to generate their results.

The researchers' technique may also be able to pick out even previously non-positive genes for further study by allowing the background "noise," or meaningless variations found in every experiment, to cancel one another out, leaving the true positives standing tall. The effect is much like wearing a pair of noise-cancelling headphones on an airplane in order to hear your favorite symphony.

The researchers knew they were on the right track when they pinpointed about 66 percent of nearly 300 previously identified obesity-related genes. In contrast, none of the individual experiments identified more than 30 percent of the same panel, and the average experiment identified only 2 percent.

Further analysis identified 16 genes that were positive in six or more experiments, and three that were positive in eight experiments. Of those three, one was a known obesity gene. The other two have no known ties to obesity, yet. "Now we can take these two candidates straight to the geneticists and the lab to begin figuring out what they do," said Butte.

The comparison approach should be applicable to many other disorders. "The data are out there," said Butte, who has pioneered ways to categorize and index the vast quantities of biomedical information in preparation for further study. "We need to translate it and make it useful for other researchers and disorders."

Source: Stanford University Medical Center