International team identifies 480 genes that control human cell division

January 7 2008

A team of U.S., Israeli and German scientists used computational biology techniques to discover 480 genes that play a role in human cell division and to identify more than 100 of those genes that have an abnormal pattern of activation in cancer cells.

Malignant cells have lost control of the replication process, so detecting differences in cell cycle gene activation in normal and malignant cells provides important clues about how cancers develop, said Ziv Bar-Joseph, a Carnegie Mellon University computational biologist who led the study. These genes also are potential targets for drug therapy.

Unlike many cancer studies, which seek to identify "missing" genes that might cause cancer, this new research shows that genes can contribute to cancer in less obvious ways. "What we see is that there are many genes that are present and yet still involved in cancer because they are not activated, or expressed, in the way they normally are," said co-lead author Itamar Simon, a molecular biologist at Hebrew University Medical School in Israel. Rather than cycling on and off as normally occurs when cell replication and development proceeds, these genes are expressed in a steady state or not at all.

The findings will be reported in the online Early Edition of the Proceedings of the National Academy of Science during the week of Jan.7.

The genes found to be deregulated in cancer cells include a few, such as

PER2 and HOXA9 that already have been linked to cancer. Most have not, including at least three genes responsible for repairing genetic mutations that occur as DNA is duplicated in the cell.

The failure of the DNA repair genes to cycle in cancer cells raises the possibility that some mutations associated with cancer may not cause cancer. "Some of the mutations may be caused by the non-cycling genes, rather than the other way around," said Bar-Joseph, an assistant professor of computer science and machine learning in the School of Computer Science and a member of Carnegie Mellon's Lane Center for Computational Biology.

Determining if genetic mutations are a side effect of certain cancers rather than a cause will require further investigation, as will identifying which of the 118 genes that do not cycle in cancer cells are most significant.

"These genes seem to be important, but we don't yet know which ones play key roles or might be targets for drug therapy," Simon said. "We have narrowed down the field of candidates. Instead of looking at thousands of genes, now we can concentrate on about 100."

Using conventional techniques even to identify a full complement of human cell cycle genes has been problematic. Molecular biologists have found cell cycle genes in yeast, plants and mice, as well as in a human cancer cell line known as HeLa. But a study that purported to identify cell cycle genes in normal human cells proved flawed and invalid.

The problem that molecular biologists encountered in studying human cells has to do with the fact that the cell development must be arrested so that micro array technology can be used to measure which genes are expressed at each stage of the cell cycle. When the cells are released from arrest, Simon said, some don't resume cycling at all, while others

resume at different intervals.

Why this is a problem in humans and not other species is not understood, Simon noted. But the result is that the cells – and these studies require millions of them – end up scattered among different stages of the cell cycle. Measurements of these unsynchronized cells are hopelessly "noisy."

"People said you couldn't solve this problem," Bar-Joseph said. But a computer science method called deconvolution, which is widely used in such fields as image processing and signal processing, proved effective in eliminating noise from the data.

In experiments, the team arrested and released cells in culture and then measured DNA content to determine which ones had stopped cycling and which ones were at various stages of the cell cycle. This information was used to construct a model of cell behavior that could be used to reanalyze the gene expression data, enabling researchers to combine expression data from cells that are all at the same stage of the cell cycle.

Source: Carnegie Mellon University