

cate. “Signaling networks are so complicated right now that common sense doesn’t always hold true,” Yaffe says.

“The thing that makes me really stop and pay attention is the methodology, which I found of special note,” says **Raphael Levine, PhD**, distinguished professor of chemistry at the University of California, Los Angeles. “Instead of trying to see if the model can predict something new, they tried to drive it to say something which they know it shouldn’t say. As a result, they were successful in finding some new biology.”

—By **Kayvon Sharghi**

Diagnosing Cell Circuitry

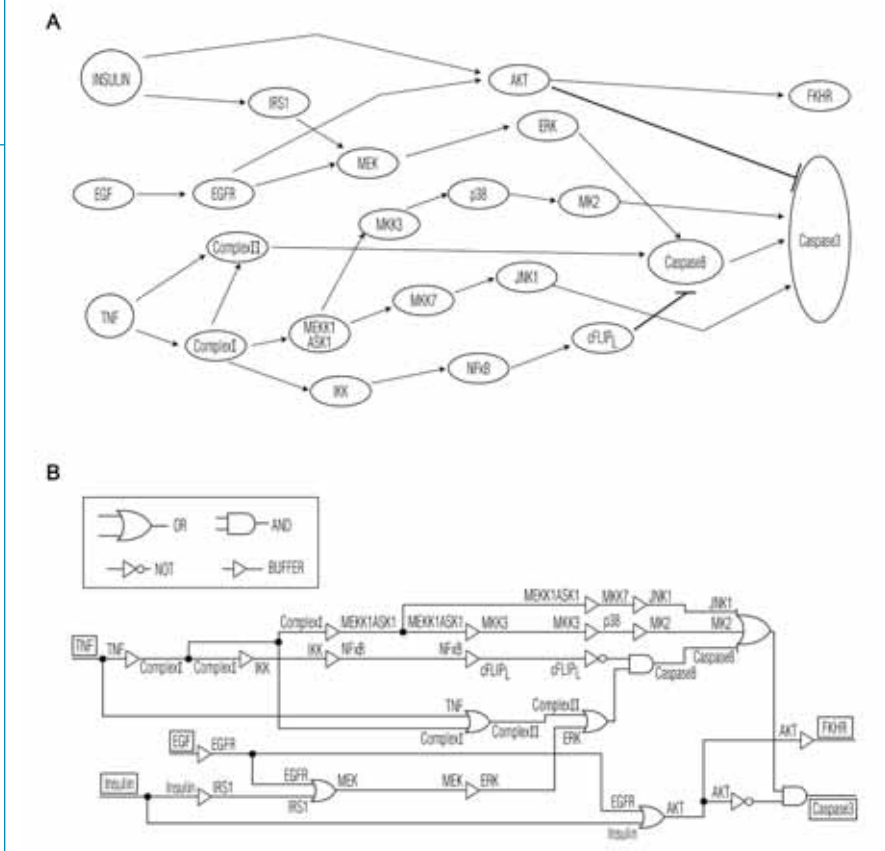
To biologists, a computer’s motherboard may just look like highways of circuitry connecting various chips. But if they focus harder, they might see a model for disease, according to new research.

Just as a single corrupt circuit can foul a computer’s operation, a faulty molecule can upset a healthy body. “If your body is not functioning correctly, then the molecules inside your cells are causing the problem,” says **Effat Emamian, MD**, president and CEO of Advanced Technologies for Novel Therapeutics in New Jersey.

The parallels between signal transduction pathways in a cell and circuit networking in a motherboard inspired Emamian’s team to identify defective cell pathways in the same way that engineers inspect faulty circuits. This technique, known as fault diagnosis, can pinpoint the molecules that are most critical to a cell’s function.

Such an accurate assessment may lead to more precise medicines. Most new drugs in trial are toxic, Emamian says, because they often target molecules essential for cell function. Fault diagnosis can reveal safer molecules to target. The work appears in the October 21, 2008 issue of *Science Signaling*.

Lead author **Ali Abdi, PhD**, associate professor of electrical and computer engineering at the New Jersey Institute of Technology, helped test Emamian’s theory. Abdi re-envisioned three previously studied cell pathways as electronic circuits: tumor suppressor p53, cell



A simple model of the caspase3 network (top) shows the various regulatory molecules and their relationships to each other. Depending on which regulatory molecules are active or inactive, caspase3 will induce cell death. This network can be re-envisioned (below) as an electronic circuit after organizing previous knowledge of the molecules’ relationships using Boolean logic. Algorithms applied to this circuit can predict molecules to which a pathway’s signal is most vulnerable. Reprinted with permission from Abdi A, et al., *Fault Diagnosis Engineering of Digital Circuits Can Identify Vulnerable Molecules in Complex Cellular Pathways*, *Science Signaling*, (2008) 1(42):ra10.

death regulator caspase3, and a nerve-cell network called CREB. His reconstructions used binary language to characterize a molecule’s state in its pathway as “active” or “inactive.” Relationships between molecules were organized into decision-making operations using Boolean logic where each relationship contains only two possible values—on or off. This allowed the researchers to write algorithms predicting which molecules were critical to a pathway’s smooth functioning. The algorithms confirmed what was known about p53 and caspase3, but they also revealed new critical molecules in the CREB network.

The approach is a good start for quickly identifying essential points in cell networks, says **Kevin Janes, PhD**, assistant professor of biomedical engineering at the University of Virginia. But while Boolean logic can make good approximations, it may oversimplify the relationships for some networks, he says. For example, Emamian’s approach doesn’t allow consideration for graded responses between “active” and “inactive.” “But it’s

not a fundamental flaw,” Janes adds.

The team acknowledges these limitations in its *Science Signaling* paper. The next step, Emamian says, is to focus on larger networks, and not necessarily just signaling pathways. “We can analyze metabolic pathways, or pathways that also have several critical enzymes playing in the whole game.”

—By **Emmanuel Romero**

Cancer’s Signature—Written in Blood

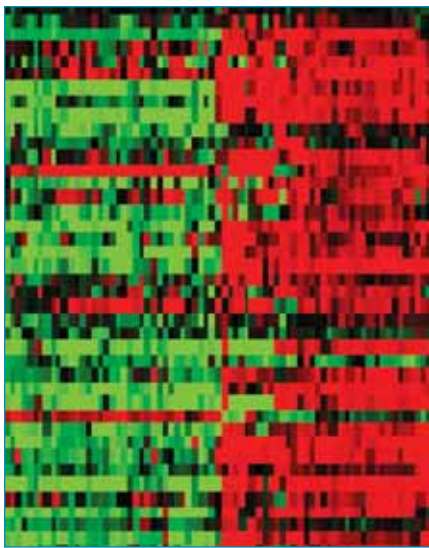
When it comes to deciphering the health of the body, the blood carries a potential mother lode of protein clues. Given the ease of extracting blood, such proteins could serve as efficient health barometers. But it’s tough to distinguish between the multitude of proteins naturally found in blood and those that are secreted into the blood—including those secreted by diseased tissue such as cancer. Their signal may get swamped by the many other proteins present in blood, thwarting efforts to discover useful infor-

“Figuring out which proteins are secreted into the blood is like searching for a needle in a big, big haystack,” says Ying Xu, PhD. “This [algorithm] sorts through all that hay.”

mation. Now, scientists have developed an algorithm that sorts through the multitude, expediting the search for blood-based cancer biomarkers.

“Figuring out which proteins are secreted into the blood is like searching for a needle in a big, big haystack,” says Ying Xu, PhD, professor of bioinformatics and computational biology at the University of Georgia. “This [algorithm] sorts through all that hay.”

To develop their algorithm, Xu and his colleagues began by scouring the literature for all proteins known to be secreted into the blood, regardless of their origins. They then analyzed the amino-acid sequences of these proteins to identify common features, such as signal peptides, transmembrane domains, solubility, and secondary structure. They discovered 18 features that were powerful predictors of blood secretion, and used them to train a computerized classifier.



This microarray shows genes that differ in regulation between cancerous and non-cancerous lung tissue. Ying Xu’s classifier can predict which of the proteins made by these genes may be useful as blood-based biomarkers. Courtesy of Ying Xu.

When the researchers applied the classifier to other data sets, it could distinguish proteins secreted into the blood from all other proteins in the blood with more than 80 percent accuracy. The results appear in the October 2008 issue of *Bioinformatics*.

Xu and his colleagues are now using microarrays to identify differences in gene expression levels between cancerous and non-cancerous stomach tissue. Using their classifier, they can then sift through the data to zero in on genes that produce proteins that are most likely to be secreted into the blood, followed by validation with mass spectrometry.

“We’ve already identified proteins that are elevated during different stages of stomach cancer,” Xu says. “Typically, in order to find out what stage it’s in, you’d have to actually cut the patients open and do a biopsy. Our markers could be the first markers to provide information about cancer stage.”

By applying his biomarker discovery pipeline to a range of cancers, Xu ultimately hopes to identify general biomarkers that apply to any cancer. He envisions doctors detecting various cancers at early stages with a simple blood test.

Bo Huang, PhD, a post-doctoral fellow at Vanderbilt University, hopes to use Xu’s classifier to find biomarkers for breast cancer. “These results provide a powerful method to discover potential biomarkers, not only for cancers but also for many other diseases,” Huang says.

—By Lizzie Buchen

Blurring Data for Privacy and Usefulness

Hospitals with research agendas share a common problem: how to use medical records for research while protecting patient privacy. One approach—the data-protection equivalent of blurring the face of an anonymous source on television—has now been tested using

real-world data. The results, which show promise for protecting privacy without rendering the data set useless, appear in the September/October 2008 issue of the *Journal of the American Medical Informatics Association*.

“It’s not a theoretical problem,” says Khaled El Emam, PhD, associate professor at the University of Ottawa and Canada Research Chair in electronic health information, who collaborated with Fida Kamal Dankar, PhD, on the paper. “We’re trying to protect privacy, but we need the tools.”

Just as the nightly news renders the faces of anonymous sources unrecognizable, the approach known as *k*-anonymity blurs distinctive variables to reduce the risk that someone could trace patients with distinctive characteristics. For example, the approach might cut birthdates down to birth years. And easily identifiable outliers—the octogenarian in a college town, the teenager in a retirement community—are omitted. The remaining information contains at least *k* data points that look identical, where $1/k$ is deemed an acceptable level of risk.

