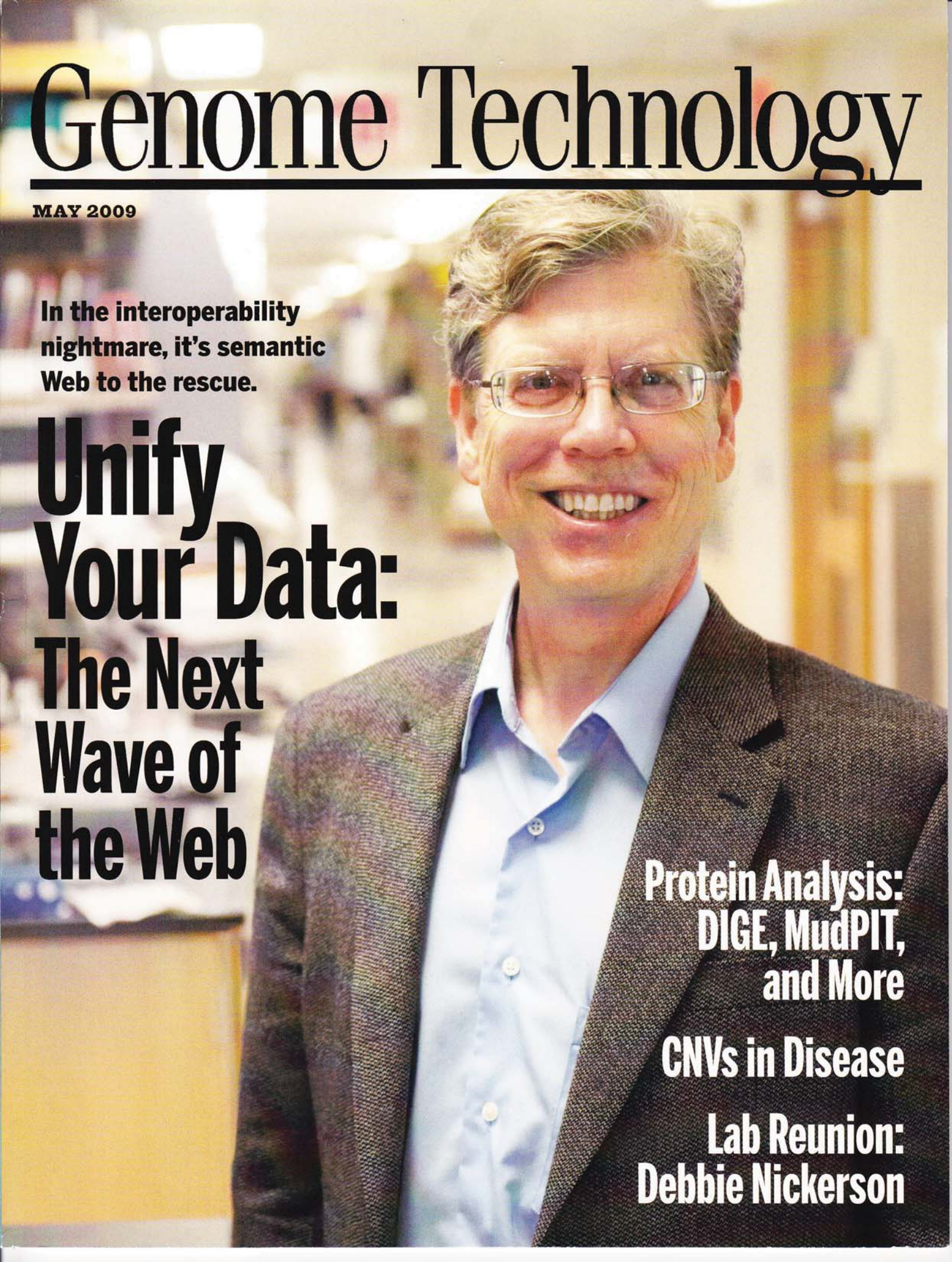
Genome Technology

MAY 2009



In the interoperability nightmare, it's semantic Web to the rescue.

Unify Your Data: The Next Wave of the Web

Protein Analysis: DIGE, MudPIT, and More

CNVs in Disease

Lab Reunion: Debbie Nickerson

sequences," he says.

Margulies and his colleagues then looked into how structural changes could have an impact on biological function. They examined disease-associated SNPs in the ENCODE database and found that those SNPs are more likely to have dramatic structural changes than neutrally evolving SNPs. From protein-binding assays, he also notes that "base changes that affected structure more dramatically also affected protein binding" giving a possible way to connect structure to disease. "There is this whole world trying to figure out how noncoding SNPs [can be] causative in disease and not just links to disease. One way is through protein binding," Margulies says.

— *Ciara Curtin*

Cell analysis: New Computational Model Predicts Timing, Roadblocks in Transcription Factor Binding

Scientists at Uppsala University and Harvard have built a computational model that addresses the effects of transcription factor crowding and DNA looping on the process of a transcription factor finding its target site. Their model can better predict the effect that roadblocks, or other macromolecules that bind to DNA, have on TF diffusion as well as suggest a role for DNA looping in speeding up the time it takes a TF to find it target site.

Models to describe the diffusion of a TF to its target site have been around for decades, but it's fairly recent that biologists could actually watch the process take place. In work appearing in *Science* in 2007, Gene-Wei Li along with Uppsala's Jonah Elf at and Harvard's Sunney Xie performed single molecule imaging to measure how long it took fluorescently labeled *lac* repressors to find their target sites in *E. coli*. Their work implicated macromolecular crowding on DNA in slowing TF binding times.

Since previous models were based on *in vitro* observations, they didn't take into account this crowding effect. Li's new model predicts that these roadblocks will greatly decrease search speed. "Those models dealt with a system that is very simple, where you have protein and DNA and that's it," Li says. While the team found that it takes only a few minutes for an individual protein molecule to find the right place to bind, this is slower than what is predicted by the established theoretical model. "If you really think about what's happening in living cells, you have all kinds of proteins that are coating the DNA as well," Li adds. "In other words, these transcription factors cannot diffuse freely along DNA as is the case *in vitro*. The models were built upon those simplified components."

GENE-WEI LI

Based on observations in *E. coli*, the new model also incorporates a new finding: that there is an optimal, and total, number of DNA binding proteins allowed in a cell. While it might seem intuitive that the more transcription factors there are searching for the same target site, the faster the search will be, "in reality, you cannot have all the different transcription factors increasing their copy number indefinitely because that would just further occupy the chromosome more and more," Li says. "So there's a limit."

Just how the TF binds quickly in light of these constraints might be explained by the effect of DNA looping. Li adds this to his equation as well. DNA looping may speed up the process because if one TF has two different possible binding sites, they could act independently if roadblocks are correctly positioned, Li says. For instance, a TF that can bind to two sites separated by a few hundred base pairs sees them as one target because they can be reached within one round of sliding. However, the existence of roadblocks makes these two sites appear to be independent "so for the TF that's trying to search for the site, it's twice as fast to find any of them," Li adds.

What has made this work possible is single cell analysis, and Uppsala's Elf is planning collaborative studies to test the new model *in vivo*. A more detailed understanding of gene regulation is important to better understand diseases related to cellular dysfunction, like cancer, as well as systems that are highly regulated by TFs, like stem cells. "In essence, all transcription factors have to go through this kind of search problem, and there are times when cells really need to respond to environmental changes through gene regulation," Li says. "This is where this model comes in; it tells us how fast a cell can respond by changing its gene activity."

— *Jeanene Swanson*