

New Genome Functions Pile On

Researchers are developing new models of genome function to investigate regulation and DNA repair, as well as to better understand disease.

BY CHRISTIE RIZK

Genomics isn't just about sequencing anymore. With the cost of genome sequencing rapidly dropping, and data from sequencing projects piling up, the field of functional genomics is coming into its own. Researchers are mining that sequence data to extend their studies to mapping interactions between genes and understanding how the genome is regulated. Instead of focusing on DNA sequences and protein structures, researchers are plunging into the ever-changing world of gene transcription, translation, and protein interactions, with an eye toward finding ways to regulate or knock down genes that cause cancer or neurodegenerative diseases. They're asking themselves how damaged DNA is repaired, how mutations can be silenced, and how chemicals interact with target genes to create useful therapeutics. More and more, genome function researchers are making new discoveries, getting grant money to start new projects, and developing new and improved methods to advance their understanding of the complex network of gene interactions that make up the human genome.

"Genetics is at the very beginning — or even in the process — of a new

scientific breakthrough," says Vladimir Teif of the German Cancer Research Center's Genome Organization and Function research group in Heidelberg. Teif compares the pace of discovery with the development of physics 100 years ago, when experimental data was accumulating more quickly than researchers could produce theories to explain it all. "The analogy with genetics nowadays is straightforward," he says. Physics eventually got Einstein's quantum theory, and, 10 years ago, genetics got the first draft of the human genome. "Now, when the price of genome sequencing has decreased dramatically, experimental data are being accumulated orders of magnitude faster," Teif says. Entire families are having their genomes sequenced, allowing researchers to find mutations associated with genetic diseases, he adds.

New high-throughput technologies are being developed to map not only protein-DNA and protein-protein interactions, but also to map histone modifications at the nucleosomes across the entire genome, and to map a three-dimensional view of the human genome as well.

Computational models of function

Such physical models of the genome are invaluable to researchers. But because of the rapid development in high-throughput technology, Teif says, data is accumulating much faster than anyone can understand it. Computational models are being developed to accompany physical models, aiding researchers' efforts to interpret data to uncover the keys to gene function and regulation. "Now the large amounts of biological data can be treated with sophisticated mathematical methods — and many purely mathematical models are being developed," he says.

Researchers from the University of Manchester in the UK, Aalto University in Finland, and the European Molecular Biology Laboratory in Germany have developed a new computational approach to gene function that they say identifies targets of regulator genes. Aalto's Antti Honkela, one of the collaborators on the project and first author on the subsequent study published in *PNAS*, says the method is based on "combining a



VLADIMIR TEIF



ANTTI HONKELA

simple linear dynamic model of gene regulation with a probabilistic model of the [time series gene expression] data." This model allows researchers to test hypotheses about transcription factors and their regulatory effects on a particular candidate gene by evaluating how well the data for the candidate gene fits the model, Honkela adds.

Because the systems being studied are so complex, purely physical models may not be enough for researchers to glean all the useful information they can from them. Honkela hopes this computational model can be combined with other methods and tools to create a complete gene function research toolkit. "On an abstract level, the model can only establish a sort of correlation, not causation, and hence cannot alone resolve regulatory relationships conclusively," he says. "However, the method can be a very useful part of the toolkit for analysis of time series expression measurements in generating hypotheses that can then be verified by other means."

Despite the model's limitations, Honkela says, the general method allows for a wide range of applications and, in the end, has a place in genome function research. "Uncovering the regulatory network could have potentially huge implications to all biological sciences," he says — not the least of which is finding more specific alternatives to drugs targeting specific transcription factors. Using his new method, researchers could screen the targets of certain transcription factors to see if directly targeting one of them would have the same benefits as drugs, with fewer side effects.

Another aspect of his work involves comparing modeling results with transcription factor binding from ChIP experiments, Honkela says. According to the computational model, strong binding sites often are next

to genes whose expression profiles show no sign of regulation, indicating that the location of the binding site has no bearing on which genes it regulates. Results also show genes that, while farther away from the binding site, fit the model better. "This supports the view that mapping of enhancers to genes they affect is non-obvious from the linear genome sequence, even in relatively simple organisms," he says.

To make the model more efficient, Honkela and his collaborators are working to extend their computational approach to generate more complex models, including combinatorial regulation by multiple transcription factors. "In terms of genomic measurement technologies, I see our method more as a technology enabling better utilization of available, often very limited measurements," Honkela says. He and his collaborators have recently received funding through the ERASysBio initiative to use the model to increase researchers' understanding of nuclear receptors, as they affect many physiological and pathological processes. The project will also involve a combination of time series expression data from RNA-seq experiments, together with nuclear receptor binding from ChIP-seq and epigenetic measurements. "An integrated view of all of these could shed new light, for example, on the role of histone marks in gene regulation," Honkela adds.

New ideas

Computational models are improved and enhanced when they are combined with physical models of gene function. And in this ever-growing field of study, new discoveries



JAY SAGE



KENDALL KNIGHT

are being made. Kendall Knight at the University of Massachusetts Medical School and his graduate student Jay Sage recently published their findings on a novel function for the homologous recombination protein, human Rad51, in the *Journal of Biological Chemistry*. Most researchers have focused on the nuclear function of Rad51, Knight says, as the protein's main function is to catalyze strand exchange between broken chromosomes. But as Knight and his team studied it, they found a lot of cytosolic Rad51. They developed a

fractionation scheme to cleanly isolate mitochondria from other subcellular components and found Rad51 and its associated HR proteins — Rad51C and Xrcc3 — in the mitochondria.

Although this was novel, Knight and Sage weren't sure how it was significant and what Rad51's function in the mitochondria really was. To determine whether Rad51's presence there was as a function of DNA damage, Sage developed a way to damage mitochondrial DNA material with glucose oxidase. "The question was, 'Does glucose oxidase and the resulting hydrogen peroxide also increase the levels of Rad51 in the mitochondria?'" Knight says. And the answer was, "Yes, it does."

When Sage performed a ChIP assay using mitochondrial DNA, he was able to show that more and more Rad51 is physically associated with mitochondrial DNA as a function of increased glucose oxidase exposure. "So they're in the mitochondria, and no one's ever seen this before," Knight says.

Knight and Sage were able to show that without Rad51, the mitochondria loses large amounts of mitochon-

drial DNA after some type of damage. When they hit cells with glucose oxidase, there was an immediate increase in mitochondrial DNA, which then dropped back to baseline levels after a time and remained steady. But when the researchers knocked down Rad51 and then gave cells the glucose oxidase treatment, there was no initial increase in mitochondrial DNA; instead, levels plummeted. "After about six to eight hours, we were down to about 40 percent of the total amount of mitochondrial DNA that was there before the damage," Knight says. "So there's definitely something that Rad51 is doing to maintain the level of mitochondrial DNA."

Knight is planning to do more work with human Rad51. He intends to find out whether the fact that the protein helps mitochondrial DNA replication forks push through damage — helping to decrease mutation levels and help the mitochondria recover from oxidative damage — is another one of its overlooked functions. The idea is to see whether Rad51 can be used to reduce the incidence of mutations in the human mitochondrial genome and thereby elucidate treatments for disease. For example, one of the major functions of the BRCA2 protein is

"The regulation of gene expression is far from being solved ... There's a lot yet to be figured out."

to help Rad51 get to the site of a break in DNA so it can repair the damage, and a mutation in BRCA2 is known to be a cause of breast cancer. Knight is also looking at mitochondrial depletion syndrome, which can lead to a debilitating, and sometimes fatal, neurodegenerative disease. "If Rad51 is indeed playing a role there," he says, "that's going to be important."

The German Cancer Research Cen-

ter's Teif is taking a different tack than Sage and Knight; instead of investigating a single protein, Teif's most recent study, published in the *Biophysical Journal*, focused on deciphering the gene regulation functions of an entire organism. He attempted to predict gene regulatory functions by calculating the logical systems operation of genetic switching in simple systems: bacteriophages. "Scientists have somehow assured themselves that the logic of biological systems is the same simple logic as in computer programs," he says. "It is not."

Teif was able to show that the logic of the bacteriophage gene switching system isn't Boolean. "This is bad news because plenty of things have been developed in the last decades using the approximation of Boolean logic, and it's unfortunate if biological systems appear to use a different logic," he says. Researchers may yet have to contend with a new logic system and a new way of thinking. However, he adds, it's still possible to develop predictive models for gene regulation using a direct mechanistic understanding about what happens at the molecular level, and not using any prior assumptions about the logic.

Teif plans to develop a human model of his mechanistic bacteriophage gene function model. This is going to be much more complex, he says, as gene expression in humans is dependent not only on the DNA sequence, but also on other factors like DNA compaction in the chromatin. Medical studies confirm that cancer and other diseases depend on these molecular constituents of gene regulation to flourish, so they have to be taken into account if gene regulation is to be truly understood.

Meanwhile, in Canada, the University of Toronto just received a \$24.7

million grant from the Ontario Ministry of Research and Innovation to be split among five researchers from various disciplines. Charles Boone of the school's Banting and Best Department of Medical Research plans to use a significant portion of his funds to finish a reference map he and his team are developing of the functional network of yeast, and to extend that research into mammalian cells, particularly focusing on networks that are relevant to cancer. Boone, whose lab has been developing this network analysis for the last nine years, says with this funding, his group should be able to identify new targets for therapies. The researchers will also be doing chemical genetic interactions as part of the grant, which could elucidate the links between compounds and targets and how they affect cellular function and physiology.

A growing field of study

Despite new tools and methods, the work of deciphering genome function has barely begun. UMass's Knight says the community is only just seeing the tip of the iceberg. The game changed with the sequencing of the human genome, he says. "The regulation of gene expression is far from being solved given the whole chromatin code issue and different chromatin modifications and things like that," he adds. "There's a lot yet to be figured out."

But the pace is increasing. More and more, technologies are being developed and researchers are becoming interested in genome function. As problems are solved, more questions will open up, Knight says, and more people will come forward to work toward answering them.

"This is a very exciting field," Teif says. "Almost every week a new paper appears which makes you say 'Wow,' and a new brick is added to the building that we are trying to build." ■