By Kristin Sainani, PhD

Several biomedical computing projects received big money in the fall of 2012. If there's one clear winner, it's "Big Data": three of the grants focus on building new computational infrastructure and tools for dealing with massive biological datasets. A fourth grant focuses on building new tools for multiscale modeling.

Advancing Bioinformatics in Africa

Genomics research in Africa will receive a major boost thanks to a $10 million grant to establish a sustainable bioinformatics network on the continent. The project, H3ABioNet, is part of the H3Africa initiative—a joint venture of the National Institutes of Health (NIH) and the Wellcome Trust to promote large-scale genomics research in Africa.

"It's very important not only to generate data but also to have the infrastructure and know-how to make sense of them," says **Victor Jongeneel**, **PhD**, director of the High-Performance Biological Computing program at the University of Illinois. "The idea is to make sure that the data produced by H3Africa projects are analyzed in Africa and not shipped out for analysis to groups in Europe or the U.S. We have a moral imperative to make sure that the benefits of the research and the credit for the work are reaped locally." The H3ABioNet team includes researchers from the University of Illinois, Harvard, and the University of Cape Town.

H3ABioNet will help set up the computational infrastructure needed to analyze high-throughput sequence data, including establishing and providing access to high-performance computing centers, installing analysis software, and facilitating data storage. H3ABioNet will also train local scientists how to handle and analyze high-throughput data, including establishing internships to provide practical, hands-on experience. Centers in the H3ABioNet network will go through an accreditation process that will ensure their proficiency in analyzing human genomic data.

Though the focus of H3Africa is on human genomics, the H3ABioNet network could provide support for any kind of computationally intensive biological research in Africa. "I'm hoping that the capacity devel-

> "It's very important not only to generate data [in Africa] but also to have the infrastructure and know-how to make sense of them," says Victor Jongeneel.

opment that will be funded in this project will also have an impact on fields other than human genetics," Jongeneel says.

Making Sense of Metabolomics Data

Metabolomics (the systematic study of the end products of the cellular processes that allow cells to grow and reproduce) is coming of age—leaving scientists with a mountain of new "omics" data to decode. But a new data repository and coordination center at the University of California, San Diego, will help deal with the data deluge, thanks to a $6 million grant from the NIH.

"In the last few years, mass spectrometry technology has matured to a point where one can do reasonably robust medium-throughput to high-throughput metabolomics," says principal investigator **Shankar Subramaniam**, **PhD**, professor of bioengineering. "We've been funded to help figure out, 'what do we do with these data?'"

The new center will serve as the data hub for research cores financed through the NIH Common Funds Metabolomics Program and similar metabolomics research initiatives. The center will provide a national repository for metabolomics data and will provide publicly available, user-friendly tools for data access and analysis. Among the challenges, Subramaniam's team will develop strict standards for data and metadata. "Data without meta-data is almost al-
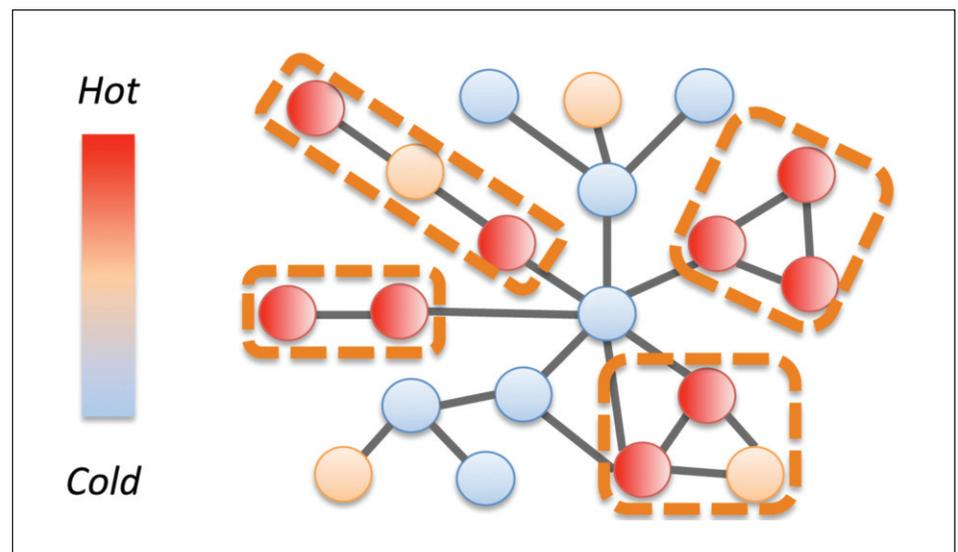
Pattern Extraction. *HotNet is an example of a statistically robust algorithm for extracting information from big genomic datasets. "Hot" gene networks are more correlated with clinical phenotype. Courtesy of: Eli Upfal, Brown University.*

ways useless," Subramaniam says.

Subramaniam's team is developing robust statistical methods for "metabotyping" diseases—identifying metabolic patterns that correlate with disease states. They are also working on complex algorithms for reconstructing metabolomics networks and integrating metabolomics data with data from proteomics and transcriptomics. "This data integration is a very complex research task; there's no plug and play application that you can buy off the shelf," he says.

Rigorously Mining Genomic Data

From personal and cancer genomes to social networks to online buying habits, we have entered the era of Big Data. Computer scientists have developed efficient and successful machine learning algorithms to mine these data for patterns. But no one quite knows: how reliable are the results? So, a team from Brown University has received a $1.5-million "Big Data" grant from the National Science Foundation and the NIH to develop rigorous statistical tools that answer this question. Their work focuses on data from the Cancer Genome Atlas.

"Machine learning is very popular today. But we don't have statistical guarantees on the quality of our results," says **Eli Upfal**, **PhD**, professor of computer science and principal investigator on the project. "We want to build techniques that will still be efficient and practical but also would quantify the quality of the results." Fellow computer science professors **Ben Raphael**, **PhD**, and **Fabio Vandin**, **PhD**, will help lead the effort.

Genomic data are large, complex, and noisy on many levels, so they are a good prototype for testing Big Data tools, Upfal says. His team has already built a tool called HotNet that helps identify—with high statistical confidence—pathways of mutated proteins that are involved in cancer. Upfal's team also aims to develop statistical tools to answer the question: how big of a sample does one need to answer a particular question in genomics? Ultimately the tools will have applications in many domains beyond genomics, Upfal says.

Bridging Gaps in Multiscale Modeling

Multiscale models in neurobiology can help scientists understand the brain from the small molecule to the tissue level. But to realize the promise of these models, scientists first must bridge critical gaps between biologists and computer scientists, as well as between computer scientists working at different scales. A new $9.3-million NIH center at the University of Pittsburgh—the Biomedical Technology Research Center—aims to do just that.

"Our first goal is to bridge the gap between experiments and computations. A first step toward this goal is to reach out to those people already doing experiments and assist them in solving their neurobiological problems," says principal investigator **Ivet Bahar**, **PhD**, professor of computational and systems biology. The center, which is a joint collaboration between the University of Pittsburgh, Carnegie Mellon University, Pittsburgh Su-

percomputing Center, and the Salk Institute, will develop multiscale models for five different driving biomedical projects across several institutions. "The tools we build will be tailored to their needs," Bahar says. The projects focus on understanding brain signaling and may lead to new treatments for neurological and behavioral disorders.
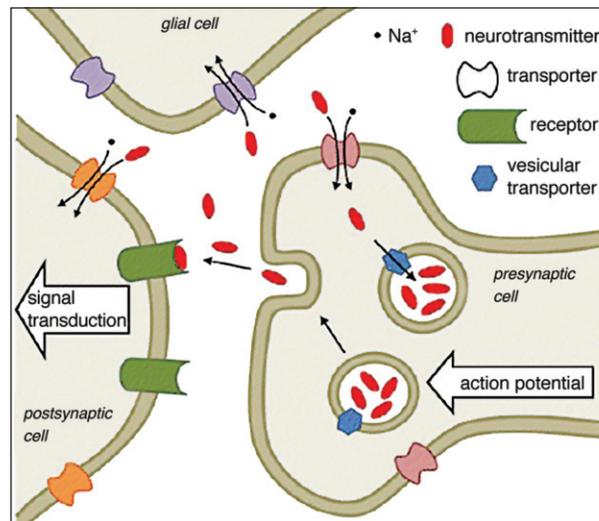
Modeling the Synapse. *The new Biomedical Technology Research Center at the University of Pittsburgh aims to combine computational models and experimental data to understand how messages are sent across synapses; and how that affects the brain. Courtesy of: Ivet Bahar, University of Pittsburgh.*

The center will also foster collaborations among modelers at different scales, such as the molecular, synapse, and brain tissue levels, she says. Currently, she says, "There is a serious disconnect. We would really like to overcome that," Bahar says. For example, her team will use molecular simulations to calculate parameters that can be fed into cell simulations created by other researchers. Those cell modelers will in turn provide feedback that can improve her team's molecular models. □

Cancer's Drivers. *Scientists applied HotNet to The Cancer Genome Atlas data for glioblastoma multiforme (a type of brain tumor). The algorithm extracted subnetworks of highly mutated genes. Courtesy of: Eli Upfal, Brown University.*