Optimizing Biomarker Use in Discovery and Development
Biomarker BREAKTHROUGHS
Cambridge Healthtech Institute

Thursday, July 12, 2007

Case Study

Biomarker BREAKTHROUGHS

Building a Google for Bioinformatics: IO Informatics and CLDA Team Up to Crack Genomic Data Boondoggle

By Malorye Allison

It's not getting the data but making sense of it that seems to be the hard part in genomics. "One of the big challenges is combining data sets, such as metabolomic and gene expression data. You have to do a lot of manual manipulation," says Alan Higgins, senior director of Translational Medicine at the Cogenics division (formerly Icoria) of Clinical Data, Inc. (CLDA). On top of that, there's the complexity of adding information from outside sources, which may be in different formats. Given how voluminous these data sets are, the problem is further compounded.

This challenge has been dogging genomics for years, and there has seemed to be no solution except for biologists to either learn how to do all this data handling themselves, or lean on their IT guys for it.

That's one reason CLDA's Cogenics unit teamed up with IO Informatics on an $11.7 million, five-year, Advanced Technology Program grant they received in 2002 from the National Institute of Science and Technology. CLDA was looking at multiple data types to find biomarkers that predict disease or response to therapy. But they wanted to be able to integrate and analyze that data without all the mucking around. IO informatics brought something new to the table – intelligent multidimensional object (IMO) database records.

IMOs are based on the same principles underlying the Semantic Web project Tim Berners-Lee and others have recently been promoting. (See Bio-IT World's Masters of the Semantic Web.) Like Adobe Acrobat PDFs, IMOs are portable and can be easily shared. But unlike PDFs, IMOs are created so that specific data types are turned into freeform relational objects: Within the platform, the discrete data types are still recognized and distinguished from each other, but now they can also be manipulated, integrated, and compared. Users can thus work with specific data within records as well as pass the whole record easily between them.

CLDA's researchers became collaborators and beta testers for a new software platform built by IO Informatics, providing queries and other input to the product's development. The result is

Sentient, which lets researchers "Look at all the data related to their field of interest, all at the same time, all in the same place, and regardless of type of information or where the data are located," says IO Informatics CEO Robert Stanley. In short, people can ask complex research questions in a "Google-like" environment.

The platform was built to accommodate the entire range of data types that make up what is now known as systems biology. IO Informatics has also added many features to assist scientists in doing the range of analyses they might want to try. Whether represented in a spreadsheet or as a complex image, the data can be easily moved, integrated, and analyzed. Data can be viewed through a variety of means, including the Web Query, a browser that lets researchers quickly peek at a variety of data types, or the Knowledge Explorer, which lets them search and relate data.

For example, researchers can select an interesting data set, then drill down on that data to a finer level and integrate it with other data. Scientists can dart around into different databases, while focusing on the genes, proteins, or compounds of interest. Because of Sentient's semantics standards approach, scientists can also "Easily fit data from their own systems silos into internal or published pathways, interactions, or other correlation networks," says Stanley.

Tox Markers of Alcohol-Related Damage

For Higgins and the team at CLDA, who are currently hunting for toxicity biomarkers related to alcohol and other chemicals, that meant being able to combine data from metabolomic and gene expression studies with digitized histopathology images. "This software gives me the ability, for the first time, to ask more complex questions," says Higgins. "If I am looking at an alcohol study, and seeing effects in liver and brain, I can now ask if that's happening in other studies, what is common between rats and humans, and what is common to acetaminophen and alcohol."

Higgins concedes he could do similar things with other tools, "But it will take much longer and you have to do most of it manually. This is a key enabling tool."

Some of CLDA's work centers on biomarkers of liver disease. As part of the National Institute of Environmental Health Sciences' Compendium Study, the company is trying to find markers of specific types of liver toxicity. They are doing gene expression, metabolomic, histopathology, blood, and urine studies in animals on a variety of chemicals at different doses and time points to see if they can correlate specific signs, such as lobular necrosis, to particular markers. Ideally, those markers will be in blood or urine. In another study, they are examining mechanisms of alcohol toxicity in rat liver and brain.

"One of the key things Sentient lets us do is to ask exactly the same set of questions about different compounds," says Higgins. While the studies are not yet completed, Higgins reports that they are seeing some correlation of genomic and metabolomic data. In addition, certain common features are starting to emerge. "Oxidative stress is clearly important in a variety of organ toxicities," says Higgins. "And our data are bearing this out as well as other findings."
