

Feature Articles : [Sep 15 2010 \(Vol. 30, No. 16\)](#)

Finding Promising Protein Biomarkers

Discovery Aided by Availability of Large Patient Databases and Robust Data-Mining Practices

Catherine Shaffer

There is an urgent need for robust biomarkers in clinical and companion diagnostics. As a result, there is tremendous interest in improving the processes by which protein biomarkers are identified and validated. Researchers gathered at CHI's "Next Generation Diagnostics Summit" last month in Washington, D.C., to discuss, among other things, the need for larger databases of patient samples with clinical outcomes and smarter bioinformatics and data-mining tools.

Stephen A. Williams, M.D., Ph.D., CMO at SomaLogic, talked about his firm's SOMAscan biomarker discovery platform. Based on specially designed nucleic acids called SOMAmers, the platform can scan a 15 μ L sample for up to 1,000 different proteins, identifying individual biomarkers of interest or profiles that can be used to diagnose disease.

The menu of approximately 1,000 proteins includes cytokines, growth factors, receptors, proteases, protease inhibitors, kinases, structural proteins, and hormones that could potentially be of interest in human disease origins.

SOMAmers are created by an in vitro competitive process where a single protein is exposed to roughly a quadrillion different synthetic nucleic acids. The strongest binders are isolated and amplified, and the process is repeated until a "winner" is found. This aptamer and protein pair then becomes part of the menu offered in a proteomic profiling scan.

"We can measure a thousand proteins simultaneously without any fractionation of the sample of serum plasma, CSF, or tissue homogenate," noted Dr. Williams. The co-efficients of variation are about 5%, and the sensitivity is in the low femtomolar range. "It's like running a thousand ELISAs simultaneously on a single sample."

In order to develop a clinical assay in lung cancer, SomaLogic researchers used archived sample sets, screening 1,400 samples in eight days to find a possible protein signature for lung cancer. The analysis turned up 61 proteins that showed a difference between the diseased and nondiseased state.

Dr. Williams and his team then whittled that number down to the 12 best markers. "Our demands were that it work just as well in stage 1 and stage 3, and that it work just as well in people with COPD as people with normal lung function."

SomaLogic has created similar screens for pancreatic cancer and mesothelioma, and has a number of other screens in the pipeline.

Multireaction Monitoring

Christoph Borchers, Ph.D., facility director of the University of Victoria Genome BC Proteomics Centre, presented his concept for developing biomarkers that predict risk for inflammatory bowel disease (IBD) and cardiovascular disease (CVD) using multireaction monitoring (MRM) for quantitative measurement of proteins.

The range of clinically significant proteins in human blood and serum spans as much as 12 orders of magnitude and is a serious challenge to most protein biomarker screening technologies. MRM can provide accurate quantitation of proteins over a large dynamic range within a single sample. Absolute quantitation of proteins is made possible by the use of isotopically coded peptides as standards.

According to Dr. Borchers, each protein can have many peptides being used for MRM analysis associated with it. By developing an assay that includes the standard peptide spiked in at a ratio of approximately 1:1 to the endogenous peptide, it is possible to cover a large portion of the range of abundance of proteins in the human blood or serum proteome. Dr. Borchers and his colleagues then created cocktails of these standards tailored to specific diseases for large-scale proteomic analysis.

In the case of cardiovascular disease, Dr. Borchers identified approximately 80 proteins that are known to be potential cardiovascular risk biomarkers. In a preliminary study using a standard cocktail, he narrowed the list down to five markers that could stratify patient groups with and without CVD with 80–90% accuracy.

Taking these techniques into IBD, Dr. Borchers is looking at approximately 100 proteins that are potential markers of risk for the disease. Diagnosis of IBD is complicated by the flare and remission cycle, which obscures endpoints and at times generates large placebo effects in clinical trials. Reliable biomarkers and companion diagnostics would eliminate much of the uncertainty in IBD therapeutic discovery and development.

Progress in biomarker discovery in areas such as CVD and IBD depends on rapid, accurate methods for screening samples. “We’re talking about more than three million people being screened every year,” said Dr. Borchers.

“We need to have high throughput in addition to absolute quantitation. For this purpose we are now setting up an assay based on the immuno MALDI (iMALDI) technology we have pioneered. iMALDI combines immuno-affinity enrichment of peptides, MALDI mass spectrometry, and MRM. It is capable of analyzing a sample in a few seconds, or thousands of samples in one day, with absolute quantitative results for proteins.”

Breast Cancer

Ginette Serrero, Ph.D., CEO of A&G Pharmaceutical, talked about A&G’s quest to find a breast cancer biomarker that could be used as a theranostic target with diagnostic and therapeutic

application, similar to Genentech's Her-2 and Herceptin.

A wide-scale biological screen to mine for functional targets turned up GP88, which is a secreted protein that is overexpressed by breast cancer tumors. It is also involved in breast tumorigenesis and resistance to therapy. GP88 is not expressed in normal tissue. Because it is secreted, it can be found in the extracellular environment, making it a good biomarker candidate.

Dr. Serrero has studied GP88 expression in tumor biopsies and correlated results with breast cancer patient survival. Researchers at A&G are developing two tests for GP88. One is a tissue test to be used in tumor biopsies. The second is a serum test. The firm also has an ongoing prospective study measuring GP88 in breast cancer patients that includes correlation with clinical outcomes.

The GP88 marker has been shown in biological studies to be associated with resistance to anti-estrogen therapy. "That means that this biomarker could have application in the current standard of care not just as a companion diagnostic to its own therapy," said Dr. Serrero.

One of the biggest challenges in discovering GP88 was in tracking down quality clinical samples. "You have a lot of banked tissue samples that don't necessarily have clinical outcomes associated with them. It took an enormous amount of time to find the right samples to establish the clinical utility of GP88."

Not Reinventing the Wheel

Difficulty finding enough good patient data for screening was also a main theme for Stephen Suh, Ph.D., scientific director for the Cancer Research Program at The John Theurer Cancer Center at Hackensack University Medical Center.

Dr. Suh noted that 85% of Americans receive their medical care at regional or community medical centers rather than academic medical centers. These regional and community medical centers do not always have the infrastructure required to procure samples from patients. This was one of the first observations Dr. Suh made while mining for data on ovarian cancer biomarkers in various scientific databases.

Dr. Suh used the BioXM software platform from Sophic Alliance to mine the NCI Cancer Gene Index (CGI) for potential cancer biomarkers. The CGI contains 6,955 manually curated cancer genes, and 2,200 of them met the NCI Thesaurus criteria as "biomarker genes."

The CGI was created through a collaborative effort by Sophic, Biomax Informatics, and the NCI. The massive project took five years and entailed mining 18 million medline abstracts and manually applying millions of classifiers to the 6,955 candidate cancer genes. The final list of highest potential biomarker genes was generated by using the evidence codes for genomic databases that Sophic organized into confidence tiers.

"By using a bioinformatics approach to identify putative biomarkers, for example lymphoma or ovarian cancer, we didn't have to reinvent the wheel by performing multiple omics experiments," said Dr. Suh. "Sometimes, it is efficient and productive to just mine the data that is out there and

validate the candidate biomarkers with patient samples to identify clinically relevant biomarkers.”

During his investigation, Dr. Suh pursued only the most robust and consistent biomarkers. This keeps the total number of biomarkers small enough to be manageable in a clinic. “For a typical biomarker project, we find less than a half dozen for diagnostic purposes, about a half dozen for prognostic purposes, and another half dozen predictive biomarkers.”

In projects focusing on mantle cell lymphoma and Hodgkin lymphoma, this process identified markers such as CD23, IGHE, and CCND1. In studies to predict clinical outcomes in a small sample set, biomarkers identified in this way performed well.

“For screening, we only work with patient samples with known clinical outcomes. So we only select markers that are different between poor to favorable outcomes,” explained Dr. Suh.

This means that because he constructed the clinical database first, a few of the selected markers would be expected to retain robustness in predicting outcomes. This is the opposite of many biomarker data-mining projects where the biomarkers are identified first and applied prospectively to clinical samples.

Scientists are gradually moving closer to the goal of having strong diagnostic biomarkers for challenging diseases like cancer. Innovative screening approaches and better data-mining techniques can facilitate the search for new biomarkers, while mass spectrometry and MRM continue to be leading technologies for overcoming the problem of dynamic range. Success in developing biomarkers depends on large quantities of quality clinical sample data and the application of intelligence and ingenuity in the search.