The Path to Progress

Giving biological context to biomarker discovery efforts can lead to faster identification of proteins. Pathway mapping and visualisation techniques are more likely to bring success to the drug development process.

Protein biomarkers have the potential to not only transform the diagnosis, treatment and prevention of disease, but also to accelerate and refine the drug discovery process. They are vital as tools for target discovery and elucidation of the mechanism of action for a drug. Protein biomarkers can be used to target responders, monitor clinical response and adverse effects, and provide indicators of the efficacy of a drug. They may provide the key to disease prevention and treatment which is personalised to the individual patient.

Protein biomarkers have the potential to not only transform the diagnosis, treatment and prevention of disease, but also to accelerate and refine the drug discovery process. They are vital as tools for target discovery and elucidation of the mechanism of action for a drug. Protein biomarkers can be used to target responders, monitor clinical response and adverse effects, and provide indicators of the efficacy of a drug. They may provide the key to disease prevention and treatment which is personalised to the individual patient.

The promise of protein biomarkers relies on the ability to identify a specific protein or, perhaps more importantly, panels of proteins. This is useful for early detection of disease, measurement of disease progress, characterisation of the biochemical process related to the disease, and the monitoring of treatment efficacy. Obtaining this information is critical to the drug development process, since proteins represent the majority of current drug targets.

Proteomic Challenges

Proteomic approaches are most commonly used to identify protein biomarkers. Proteomics endeavours to characterise the entire protein complement of an organism and use that information to understand the organism’s biology. Many techniques attempt to accomplish this herculean task, but the most promising employs liquid chromatography and tandem mass spectrometry analysis (LC/MS/MS). This approach involves the wholesale extraction and digestion of an organism’s proteins, followed by LC/MS/MS to identify specific peptides that can then identify and quantify specific proteins, using sophisticated software.

The inherent difficulties with proteomic approaches derive from the complexity of the protein makeup of an organism. There are obviously thousands of human proteins that may be involved with disease processes. In addition, accurate measurement of changes in protein levels that occur during disease progression is complicated by the enormous dynamic range of protein concentrations in biological fluids, which can be of at least 12 orders of magnitude. Therefore, signal-to-noise ratio becomes an inherent limitation of proteomic approaches to identifying useful protein biomarkers.

Untargeted proteomics approaches using LC/MS/MS strive to identify every protein in an organism that may undergo a significant change in concentration during a disease state. An inherent limitation in this approach is that not every protein involved in a disease process may be identified, since many proteins may have a very low concentration, relative to the others and the background noise in the experiment. As a result, some important biomarkers may be missed, and others that are identified may not be subsequently validated as useful for drug development and disease diagnosis.

Biology-Driven Strategies

Knowledge of the biological relevance of a protein can greatly increase the probability of a protein’s usefulness as a biomarker for disease. This knowledge can be provided by known human biochemical pathways. That is, once potential biomarkers are identified using LC/MS/MS, they can be mapped to pathways that may be involved in disease progression. Pathway information is available in many online databases. Pathway mapping will identify other proteins involved that may be potential biomarkers; a follow-on experiment can then be designed to target only the proteins involved in the impacted pathways, greatly increasing the chances of identifying useful biomarkers.

Another means of determining biological relevance is to cross-reference the proteomics results to those from other ‘-omics’ approaches. While one laboratory may...
use only proteomics to study a particular disease, another may have profiled the same disease using genomics, transcriptomics or metabolomics. Key to making these orthogonal approaches relevant to the proteomics data are, again, biological pathways. For example, metabolomic data can be mapped to pathways as well, identifying biochemical processes that may be involved in disease progression. The proteomics laboratory can use this information to identify proteins involved in these pathways that could be potential biomarkers, and then design targeted LC/MS/MS experiments to determine changes in their concentration in the disease state.

Pathway-Enhanced Discovery

Sophisticated MS instrumentation and software are essential to performing effective proteomics-based protein biomarker discovery. Quadrupole time of flight (Q-TOF) MS is the most efficacious for this application, due to its accurate mass capabilities that maximise the potential for accurate protein identification. Software is required for identifying peptides and the proteins from which they were derived, and statistical analysis software can then identify changes in protein concentration that are statistically significant for disease progression. Finally, biochemical pathway analysis software is required to map those proteins to relevant pathways.

Figure 1 illustrates the workflow for a pathway-enhanced protein discovery experiment using LC/MS/MS, showing the mass spectrometer and software packages required. As proof of principle, yeast was used as a model organism to identify proteins involved in drug-induced biological changes. The protein discovery workflow summarised in Figure 1 was used with the digested lysates from the treated yeast cultures. A Spectrum Mill software search resulted in a total of 3,446 distinct proteins and 13,616 unique peptides being identified in the sample set, using validation criteria of 1.2 per cent false discovery rate at the spectral level.

Protein Abundances

In order to see true differences in protein level between the culture conditions, the same amount of total protein was injected for each condition. In addition, the Spectrum Mill protein-protein comparison summary was used to verify that the abundance of some typical housekeeping proteins (heat shock proteins, elongation factor and enolase) were at roughly equivalent levels across the culture conditions and across several orders of magnitude in abundance.

The protein abundances were then exported to Mass Profiler Professional (MPP) software and assessed for statistical significance to distinguish proteins that were differentially expressed under the various drug treatment regimes. Principal component analysis (PCA) of the differential proteins showed clear separation of the culture treatments, indicating that the identified proteins were in fact indicative of changes in protein abundance caused by the drug treatments, making them candidates for protein biomarkers.

The differentially expressed proteins were also mapped by Pathway Architect software to the curated yeast pathways from BioCyc, a collection of 2,988 pathway/genome databases (PGDBs) (1). Each PGDB in the BioCyc collection describes the genome and metabolic pathways of a single organism.
By mapping these proteins and the abundance results to the yeast metabolomic pathways, it is possible to identify pathways impacted by the treatment. For example, 13 of the 22 proteins in the superpathway of ergosterol biosynthesis were found to be significantly down-regulated in the FK506 drug-treated culture. The individual heat strips adjacent to each differential protein summarise the log2 transformed abundances for each condition, thus allowing a quick view of the experiment. In addition, a heat map table for all proteins in the pathway is matched with the BioCyc pathway protein list.

-Omics-Targeted Discovery

The -omics-targeted protein biomarker discovery workflow was demonstrated using previously acquired metabolomics results for the same yeast samples as the starting point for generating the list of target proteins (see Figure 2) (2). All protein accession numbers were then exported for the following three pathways, based on the previous identification of metabolic pathways perturbed by the drug treatment:

- Salvage pathways of adenine, hypoxanthine and their nucleosides
- Superpathway of purine nucleosides salvage
- Guanosine ribonucleotides de novo biosynthesis

This list of 15 proteins was then used to generate proteotypic peptides and possible multiple reaction monitoring (MRM) transitions for the triple quadrupole proteomics method. As a first step, the proteomics discovery data from the untargeted Q-TOF experiment was used to create MRM transitions where possible, to be used in a targeted triple quadrupole MS/MS experiment looking for these 15 proteins. The Spectrum Mill results were also exported as a spectral library for import into Skyline – a freely-available Windows client application for building selected reaction monitoring (SRM), multiple reaction monitoring (MRM) and parallel reaction monitoring (3).

The SRMAtlas is a compendium of targeted proteomics assays used to detect and quantify proteins in complex proteome digests by MS (4). It results from high-quality measurements of natural and synthetic peptides conducted on a triple quadrupole mass spectrometer, and is intended as a resource for SRM/MRM-based proteomic workflows.
The advantage of a targeted workflow – as opposed to an extensive fractionation-based discovery workflow – is speed. Using data from a metabolomics experiment and pathway information to focus the targeted experiment gives greater coverage of these biologically interesting proteins.

The final results were reviewed in Skyline, and areas were exported for evaluation in MPP. Peak areas were summed at the protein level prior to import, then subjected to statistical significance testing and visualisation. The PCA plot shows that the culture conditions are well separated, and all 15 proteins were found to be significant statistically (see Figure 3).

**Targeted Workflow**

The MPP 12.0 software suite has introduced multi-omics capability that enables two different -omics experiments to be mapped and seen on the same pathway. Figure 4 shows the pathway results of a multi-omics experiment created using the metabolomics and targeted proteomics data for the salvage pathways of adenine, hypoxanthine and their nucleosides. Differentially detected metabolites and proteins are highlighted in colour and have an adjacent heat strip for the relative abundances across the culture conditions. In addition, separate heat map tables are produced for the metabolomic and proteomic data.

The advantage of a targeted workflow – as opposed to an extensive fractionation-based discovery workflow – is speed. Using data from a metabolomics experiment and pathway information to focus the targeted experiment gives greater coverage of these biologically interesting proteins. For the three targeted pathways, the initial proteomics experiment only confidently identified three of the 15 proteins.

The SRMAtlas for yeast provided possible peptides and transitions for those proteins not identified in the Q-TOF experiment.

Skyline software was used to further develop and refine the MRM method for the 15 proteins, including *in silico* prediction of peptides and transitions, peak identification, and comparison with the spectral library. The final retention time-scheduled dynamic multiple reaction monitoring (DMRM) method had two to four peptides per protein (43 total peptides) and three to four transitions per peptide (166 total transitions).

This DMRM method was used to perform replicate Q-TOF LC/MS/MS analyses for each yeast drug treatment condition.

---

**Figure 3:** Principal component analysis of the differential proteins (ANOVA) between the culture conditions based on protein abundances from targeted triple quadrupole results. The results show clear separation of the four conditions. CA: calcium control; CY: Cyclosporin A treatment; FK: FK506 treatment; WT: wildtype control.
Focused Research

Mapping the results of a proteomics biomarker discovery experiment onto pathways focuses the research to active biological areas. Pathway mapping and visualisation enable identification of proteins that were missed in the untargeted discovery experiment. Using -omics data can bring additional focus to the pertinent biology, increasing the chances of identifying biomarkers that will be proven to be relevant to drug development, disease diagnosis and monitoring.

A metabolomics (or genomics or transcriptomics) experiment can be the starting point for the next study after the initial untargeted proteomics experiment, using pathway mapping and visualisation to propose a targeted proteomics study. The results can then be merged with the previously obtained metabolomics experiment to produce a more complete view of the biology. This powerful biology-driven approach is made possible by sophisticated MS instrumentation and software suites for data acquisition, statistical analysis and pathway mapping.

References
3. MacCoss Lab Software, Department of Genome Sciences, University of Washington School of Medicine. Visit: https://skyline.gs.washington.edu/labkey/project/home/software/skyline/begin.view

Christine Miller is a Senior Application Scientist for discovery and targeted proteomics in the LC/MS group at Agilent Technologies. She joined Agilent (then Hewlett Packard) 24 years ago as a field application specialist for LC and capillary electrophoresis. For the past 17 years, she has been involved in the development and use of MS instrumentation and software, including the triple quadrupole, Q-TOF, single quadrupole and ion trap mass spectrometers, as well as nanoflow LC/MS.

Enquiries: susan_berg1@agilent.com