CLINICAL PROTEOMICS:
COMING TO TERMS WITH COMPLEXITY

Proteomics can explore what is going on (or going wrong) in a particular tissue at a particular time. Such knowledge can help to develop diagnostics and to select compounds and doses for clinical trials. But selecting the information to use is a conundrum. Protein content varies from cell to cell and minute to minute. Abundant proteins mask the presence of rare but key proteins. It is hard enough to find reliable differences between disease and health or between patients likely to be helped by a drug and those likely to be harmed by it. Even harder is homing in on the differences that really matter. By Monya Baker

The early days of proteomics were characterized by optimistic naiveté. Large-scale study of the body's proteins promised to reveal new drug targets plus markers that would diagnose disease and reveal when and for whom a drug was working. The ease of generating data, paired with inconvenient or inadequate techniques for controlling variation, spawned poor experimental design.

Eventually, researchers looking for protein differences between experimental and disease groups began to realize that what they were finding either could not be replicated or could not be tied directly to diseases or drug mechanisms. In 2004, sales of an early proteomics test to detect ovarian cancer were halted following criticism that the protein profiling patterns could not be tied down to a consistent set of proteins. Just last year, the journal Proteomics issued a set of standards for submitted research and acknowledged that the rapid expansion of the field had led to publications of questionable quality (Wilkins, M.R. et al. Guidelines for the next 10 years of proteomics. Proteomics 6:4–8, 2006).

“There is a lot of exuberance when technologies do things that were previously not attainable,” says Scot Weinberg, former head of proteomics at Ciphergen and now CEO of GenNext Technologies, a company that trains scientists for translational research. He thinks proteomics was harmed early on by “blinded enthusiasm” that has only recently started to listen to common sense.

For proteomics studies to yield applications for clinical research, companies must master the science of smart trade-offs. Researchers must simplify complex mixtures of proteins so that they can be analyzed reliably, but not so much that they eliminate the very differences they hope to find. Conversely, scientists must ensure techniques that collect data on hundreds of thousands of peptides don't pull out differences that don't actually exist. And once protein differences are identified, teams developing drugs and diagnostics must decide how, and how well, to characterize them.

Lowered Expectations
Proteomics advocates never tire of describing how a protein-level view of biology can reveal more than a survey of gene transcripts. “The mRNA levels might change, but not be consistent with the protein changes,” says Yuqiao (Jerry) Shen, vice-president of research and development at Applied Biomics. He estimates that transcripts and proteins are correlated less than half the time. “Proteins are harder to analyze,” he says, “but they give you more direct information.”

“There is a lot of exuberance when technologies do things that were previously not attainable.”

Look for these Upcoming Articles
Genomics — March 16
Cell Signaling — April 6
Stem Cells — April 20
RNAi — June 1

continued ›
Ultimately, researchers often choose, not the technique that enriches for a biologically interesting set of proteins, but one they can trust or afford.

The technology of proteomics is less mature than that of its cousin genomics. Commercially available DNA microarrays can interrogate the expression of every known human gene even in tiny samples, but the most optimistic advocates estimate that a proteomics survey will detect only one-fifth to one-third of the proteins in a sample. If researchers want to monitor a set of proteins from sample to sample, those fractions are even smaller.

One of the most significant advances in using proteomics to guide drug and diagnostics development is that researchers have stopped trying to be as complete as they can be with genomics, says Jenny Harry, deputy CEO of Proteome Systems in Sydney, Australia. “Initially, people used to try to make the technology display every protein in the proteome. That was an impossible task.” Now, researchers tend to ask more limited, pertinent questions.

Clinical proteomics works best as a “practical science,” says Howard Schulman, vice-president of biomarker discovery sciences at PPD, a large contract research organization. “You don’t have to define the entire systems biology,” he says. “You have to find enough things that change reproducibly that turn out to be useful.” To do so, researchers must first imagine how protein profiles might be altered when a patient has a disease or when a drug affects a particular pathway. Then, they look for the samples and proteins that could reveal these differences.

Sample Collection and Protein Enrichment
Proteomics faces a strategic dilemma even before samples get near a mass spectrometer, the workhorse of proteomics research. In general, researchers must choose between less-informative samples from many patients or more-informative ones from fewer patients. Heart tissue might be ideal for studying cardiovascular disease, but collecting biopsies is difficult. Blood and urine are the most accessible but tend to be far removed from disease. Any proteins that leak from diseased tissue into the bloodstream are vastly diluted within a patient’s body. Their presence will be masked further by much more abundant plasma proteins, and their characteristics modified by proteases and other substances in the blood or urine.

Even if diseased tissue can be sampled directly, as in a tumor biopsy, tissues are heterogeneous mixtures of cells; a tumor may contain highly vascularized areas as well as other areas suffering from hypoxia. Consequently, the signal from the most relevant cells might be too faint to detect. Research teams can use a variety of tools to refine what they collect. Flow cytometry can select for certain cell types. In some techniques, healthy and cancerous cells taken from the same biopsy (and thus the same individual) can even be compared; or laser capture microdissection can pull specific organelles from cells, allowing researchers to focus on proteins associated with mitochondria, lysosomes, or other cellular fractions.

But these techniques are costly and time consuming. It can be difficult both to collect enough tissue from one patient and to collect samples from enough patients. Scott Patterson, executive director of medical sciences at Amgen and a co-author of the recently published proteomics guidelines (Proteomics 2006 as above), says that the difficulty of studying the individual cell types within tissues is the biggest barrier for using clinical samples in proteomics.

Researchers steeped in molecular biology don’t always realize just how difficult it can be to find rare proteins with mass spectrometry (MS), says Ruth Vanbogelen, who is in charge of helping researchers at Pfizer incorporate proteomics studies into research programs across the company. Since no equivalent of PCR exists for proteins, they can only be detected if isolated in sufficient levels at the outset. The difference in concentration between the most and least abundant proteins in blood is around 10^12.

Researchers deplete abundant proteins to boost signal from less abundant proteins, but rare proteins are still hard to see. “If you remove 99 percent of the serum albumin, then most likely your most abundant species is still going to be albumin,” observes Amgen’s Patterson. Moreover, the affinity columns typically used to remove these abundant proteins introduce new sources of variation. Even if the amount of abundant proteins removed from different samples varies by only a fraction of a percent of the original amount, the concentrations left in the samples can vary by 50-fold. Worse, techniques to remove albumin and other abundant proteins tend to remove less abundant ones as well, often in irreproducible ways.

After depletion, researchers need to concentrate a particular class of proteins within a sample. Ideally, these techniques enrich proteins based on hypotheses of which proteins are likely to be biologically important. Techniques can pull out proteins that have been glycosylated, polyubiquinated, or phosphorylated, with varying degrees of reliability and cost. Ultimately, researchers often choose, not the technique that enriches for a biologically interesting set of proteins, but one they can trust or afford. Additional variation is introduced during the analysis stage, since mass spectrometry cannot simultaneously detect and identify every chemical species in a sample.

Bringing Proteomics to the Masses
Standardized techniques could reduce variation across experiments. So it’s no surprise that kit and reagent companies are beefing up their offerings. David Smoller, vice-president of R&D at Sigma-Aldrich, estimates that his company’s investment in proteomics has probably quadrupled over the past five years. Bio-Rad acquired the proteomics side of Ciphergen last year and hopes to “democratize” the technology so that biologists rather than MS specialists can run many of the experiments.

Part of Vanbogelen’s job at Pfizer is to work with commercial suppliers and even competitors to disseminate proteomics tools. “Clinical proteomics is one area where intellectual property is not important and the freedom-to-operate concept is a key driver,” she says. In addition to making benchwork easier for users, gold-standard kits and reagents would mean regulators are more likely to trust results from proteomics experiments. Amgen’s Patterson continued ›
says that kits could be very useful, particularly for standardizing sample preparation, but that researchers have quite a wait before such products will be bought and trusted by drug companies.

Vanbogelen is more optimistic. “How much is available right now is limited,” she admits, “but by 2010, we’ll have seen that big shift.” Still, she thinks it might take until 2050 before there’s a way to measure 40,000 proteins, comparable to the number of transcripts that can be measured today on commercially available DNA microarrays.

Saved by Statistics
In the early days of proteomics, few researchers understood just how standardized studies needed to be during the discovery stage. Now, they are realizing that even rigorously standardizing techniques won’t eliminate variation. That does not make analysis impossible, but it does mean that researchers must make sure that the biologic variation can still be seen despite variation introduced by experimental techniques.

To avoid finding artificial differences created by protein processing and analysis, researchers can label proteins in samples differently, then mix, process, and analyze them together. The ratios of the differentially labeled proteins should reveal biological variation between the samples. A variety of techniques exists. Applied Biomics, a certified service provider for GE Healthcare, labels proteins in samples with differently colored fluorescent dyes, and dissimilarities between samples show up readily in 2-dimensional difference gel electrophoresis (Dige). Other techniques include PerkinElmer’s ExacTag and Proteome Sciences’ Chemical Mass Tags, Sigma’s AQUA, and Applied Biosystems’ ICAT, and can be used with gel or gel-free separation systems. Still other methods require heavy isotopes to be incorporated into proteins as they are produced, and so are not always practical for work on human samples.

Labeling samples can make statistical analysis possible, says Pfizer’s Vanbogelen. Label-free methods, she says, tend to create what her team calls “holey spreadsheets,” where information for thousands of different proteins is simply missing. With a label-free method, she says, you can run the same sample through a mass spectrometer six times, and only about 5 percent of the proteins will show up in all six runs, a rate too low to run a statistical analysis to find differences between groups. But if samples are labeled, several can be run at once, and this allows enough datapoints to be collected for multivariate statistics. However, better software to identify proteins from MS data are making label-free methods more useful, a development that will open up proteomics to broader groups of researchers.

In fact, the improvement of statistical techniques for both labeled and nonlabeled proteins is one of the most substantial advances the field has made, says Stephen Kingsmore, president of the National Center for Genome Resources in Santa Fe, New Mexico. Particularly useful are methods that can correct for the false discoveries that crop up whenever computers crunch through mounds of data. New enthusiasts don’t always appreciate that they need to design experiments that collect the right data to answer their questions. “The idea that you can pull a mass spec into the room and have it magically deliver the goods is naive,” says Kingsmore. “The error that people make all the time is they get their data set and then think, ‘Now what am I going to do with it?’”

Bringing in Biology
PPD’s Schulman says that he’s seen research and development groups get more sophisticated over his years working for proteomics providers. Early on, he says, he would meet with protein chemists who knew how to separate components in a mixture but were unfamiliar with surveying large numbers of proteins within a sample. Now, says Schulman, he’s more likely to meet with broader biomarker teams of experts who approach the problem from a clinical and biological vantage, not simply a technical one. They consider what kinds of samples can be collected and what classes of proteins should be enriched, not just how to generate the data.

Proteomic studies can indicate that a drug or disease has changed the concentrations of particular proteins, but not why. Ideally, additional literature and laboratory research will suggest biological reasons for these results along with further ideas for study. In any case, to translate proteomics studies into clinical decisions and other applications, most researchers believe they must shift from profiling proteins to validating the relevance of just a few.

That shift requires another exercise in trade-offs. Multiplexed antibody tests are relatively easy to perform, results vary little from test to test, and technologies to measure several proteins at once are improving. However, creating such tests can take several months, and cross-reactivity is a problem. On the other hand, a new advance in MS, multiplexed reaction monitoring, is more sensitive and less variable than other MS profiling techniques. Yet, it only works for small and highly abundant proteins and is still more variable than reagent-based tests.

But such debates are a sign of progress. Proteomics is littered with work that stopped at the discovery phase, says Peter Schulz-Knappe, chief scientific officer of Proteome Sciences. “The need is not just to do discovery but to commit yourself after discovery to do proper assays.” In other words, proteomics research can become useful only when the study of the proteome shifts to the study of proteins.

Monya Baker is a freelance science writer based in San Francisco.
Human Protein Microarray
The ProtoArray Human Protein Microarray version 4.0 enables rapid profiling of thousands of biochemical interactions in as little as one day. Users can screen biological samples or labeled probes against 8,000 unique human proteins from multiple gene families in a single, rapid experiment. Full-length, native human proteins from a broad range of protein classes are arrayed in duplicate on nitrocellulose-coated glass slides. These high-content protein microarrays can be used for identifying disease-specific protein biomarkers, mapping protein-protein interactions, profiling antibody specificity, and performing target discovery and validation. | Qiagen For information +49-2103-29-11710
www.qiagen.com
|
Proteomics-Grade Detergents
The Variety packs are high-purity detergents suitable for membrane and total protein extraction. These proteomics-grade detergents are purified to reduce levels of peroxidases, aldehydes, carbonyls, heavy metals, and other contaminants. They are packed in a glass vial under inert gas to maintain quality. The Variety packs include 10 ml each of 10 nonionic, ready-to-use detergents. | G-Biosciences/Genotech For information 314-991-6034
www.GBiosciences.com
|
Proteomics Standard
The Universal Proteomics Standard is a complex mixture of 48 human proteins designed to enable researchers to better assess proteomics strategies, troubleshoot protocols, and normalize results from day to day and lab to lab. The protein mixture was evaluated by more than 100 independent proteomics laboratories worldwide. | Sigma-Aldrich For information 314-286-7616
www.SigmaAldrich.com
|
Gel Documentation System
The CSL MicroDOC is a compact, transportable, and inexpensive gel documentation system that produces high-quality images. Its integral 16-bit charge-coupled device digital camera provides resolution of 8.0 megapixels. A wide variety of images from agarose, other fluorescent gels, colorimetric gels, autoradiographic film, and blotting membranes can be viewed from the system’s high-definition, 8-inch, thin-film transistor liquid crystal display. Files can be saved in a variety of formats, such as TIF, JPEG, RAW image, or QuickTime, ready for transfer to computer systems for storage and further analysis using the included USB cable. Images can also be printed directly from the MicroDOC. A 55-mm ethidium bromide filter includes a built-in safety switch to turn off the ultraviolet transilluminator when the door is open. | Cleaver Scientific For information +44 (0) 1788 565 300
www.cleaverscientific.com

Sample Processing Platform
QIAcube is a compact platform that incorporates novel and proprietary technologies to allow users to fully automate the processing of almost all Qiagen consumable products. The platform can be used with more than 100 spin column–based protocols for DNA, RNA, and protein sample processing. The protocols are based on the same Qiagen consumable products that are used manually in more than 40,000 laboratories around the world, enabling effortless migration to the automated system. There are several hundred proprietary technologies incorporated into these Qiagen consumables, so by running them on the QIAcube, users can free up time and enjoy even higher levels of quality and performance. This compact system is designed for the low- to medium-throughput range for laboratories engaged in research, applied testing, and molecular diagnostics. | Biacore For information 800-955-6288
www.biacore.com
|
Advertisers of Proteomics-related Products:
Biacore www.biacore.com
Takara www.takara.com
Reichert www.reichertai.com

Crystal Imaging for Protein Crystallography
Designed to fit on a lab bench, the Desktop Minstrel is a modular and expandable protein crystal imaging and analysis system. The Desktop Minstrel can be configured to meet a range of applications, from low throughput research to high throughput drug discovery programs. The system automatically images crystallization experiments and links images with crystallization conditions. The data are captured in CrystalTrak, a complete virtual crystallization laboratory that provides a chemical and crystallization database and data analysis tools. CrystalTrak provides a powerful interface for viewing images and conditions for evaluation and scoring. With a high-resolution imaging system that can visualize hanging drop, sitting drop, microbatch, and free interface diffusion experiments across most commercially available plate types, the Desktop Minstrel facilitates scoring and reporting as well as experimental design and project management. The instrument is cold-room compatible and expandable to a 160-plate capacity. | Rigaku For information 281-362-2300
www.Rigaku.com
|
Dual Channel Surface Plasmon Resonance
Introducing the SR7000DC Dual Channel SPR Instrument, a powerful, flexible and affordable instrument for determining biomolecular interactions. The SR7000DC provides real-time, label-free, simultaneous monitoring of sample and reference channels and is useful in proteomics, genomics, and antibody studies. Its modular design is easily configured to do customized research. Fluidic configurations range from manual to fully automated with an optional autosampler. It is able to do kinetic studies of small molecules up to 200 Daltons. Temperatures are programmable, micro or macro flow cells are available for different surface work, and flow rates can be run fast or slow. | Reichert For more information 888-849-8955
www.reichertai.com

1154