

Clinical Proteomics: The Promises and Challenges of Mass Spectrometry–Based Biomarker Discovery

Imelda E. deVera, Jonathan E. Katz, PhD, and David B. Agus, MD

Ms. deVera is at the Center for Applied Molecular Medicine at Cedars Sinai Medical Center, in Los Angeles, where Dr. Katz is Head, Mass Spectrometry Group, and Dr. Agus is Director.

Address correspondence to:
Dr. David B. Agus, Center for Applied Molecular Medicine (CAM²), Cedars Sinai Medical Center, 8631 Third Street, Suite 1001E, Los Angeles, CA 90048; Tel: 310-423-7600; Fax: 310-423-1998; E-mail: david.agus@cshs.org.

Abstract: The lack of effective surrogates or biomarkers for biologic and disease state is a major hindrance to effective cancer care. Many laboratories have been using mass spectrometry as the means to discover new biomarkers. The success of these efforts relies on effective collaboration between clinicians and mass spectroscopists. This review describes many of the common proteomic experimental methodologies and evaluates the current state of proteomics-based biomarker discovery—starting from sample collection through sample processing and ultimately sample measurement. Guidelines are given for the evaluation of experimental design and for the gauging of result significance. Particular emphasis is placed on those aspects strongly associated with reproducibility from both clinical and laboratory perspectives.

The primary yield of the human genome project is the raw nucleotide sequences of a putative 30,000 genes. The scientific community has been empowered by this knowledge to engage in large-scope proteomics experiments to determine the function of the proteins these genes encode. Previously unidentifiable “peaks” and “bands” can now be directly assigned to the proteins that represent the source of these signals. It is the proteins and their actions that are responsible for the functional diversity of cells, and it is proteins that perform most biological processes.

Proteomics is the study of proteins on a large scale. Through proteomics, scientists link differences in protein expression with underlying biological processes, including those that govern disease.¹ Our rapidly growing repertoire of techniques allows us to identify and study protein expression and modification either singly or collectively. There is no doubt that clinical research will benefit enormously from these methodologies through the development of better diagnostic tests, the identification of new therapeutic targets, and, ultimately, personalized patient therapy.² Unfortunately, it has been common in recent years to ascribe unreasonable expectations to the possible outcomes from proteomics-based experiments. The following will address many of the concerns, strengths, and weaknesses of proteomics with regard to both experimental design and proteomics-based analyses.

Keywords

Proteomics, mass spectrometry, profiling, plasma, serum

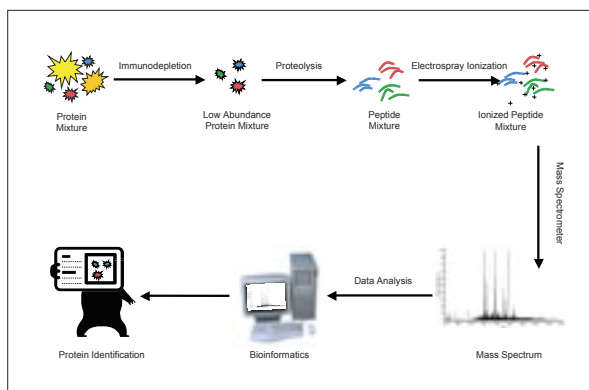


Figure 1. Schematic depiction of a typical protein identification/characterization experiment. Protein mixture (eg, plasma) is immunodepleted to remove high abundance proteins. The remaining proteins are proteolysed (eg, trypsin). Resulting peptide fragments are ionized (in this case via electrospray) and injected into the mass spectrometer, yielding a spectrum of mass-to-charge ratio (m/z). Subsequently these spectra are searched against existing databases yielding protein identifications.

Although other techniques exist to perform proteomic analysis, this review will focus on mass spectrometry (MS). Simply, MS represents a collection of techniques, the end result of which is to report the mass-to-charge ratios (m/z) of all the analytes placed in the instrumentation (Figure 1). For reviews of MS see Katz and colleagues³ and Anderson and Anderson.⁴

The Common Proteomic Experiments

There are several core experiments that utilize MS for proteomic analysis, the first being protein identification. Given a reasonably pure protein or protein fragment (from a gel, liquid chromatography [LC], or even a well-isolated m/z signal within a mass spectrum), it is a fairly straightforward process to isolate the species, energetically or chemically fragment it, and compare the resultant mass fingerprint with theoretical protein/peptide libraries (eg, MASCOT, SwissProt).⁵ A slightly derivative experiment is the identification of post-translational modifications (PTMs). Having identified the analyte, it is possible to consider the sources of anomalous signals and attempt to ascribe them to various modifications such as methylation, phosphorylation, sumoylation, etc. Identification experiments can be done in a very high-throughput fashion; in shotgun proteomics, complex mixtures are analyzed and the mass spectrometer attempts to isolate and fragment every m/z signal for subsequent database searching. It is not uncommon to be able to identify 500 or more distinct proteins from a single LC/MS analysis.

Beyond identification, quantitation of analytes can also be performed by looking at the relative intensity response from the mass spectrometer. MS-based quantitation has a strong advantage over antibody- or nucleic acid-based methods in that there is no need to have prior knowledge of the identity of the analyte, and unlike spectrophotometric or gel-based techniques, MS-based quantitation remains specific to a single analyte.

An extension of quantitation is differential analysis, in which two or more samples are compared for discriminating features. These features are often subsequently analyzed to determine the identity of the parent protein in order to enable studies of biologic relevance. Differential analysis can be used as part of a biomarker discovery pipeline, dividing samples into clinically relevant groups with the hope that the features that are found to be discriminating between groups will have clinical utility.

Current Biomarker Progress

Although the possibilities proteomics provides in identifying biomarkers are vast, there are still major obstacles. In fact, as of yet, no clinically useful biomarkers have been discovered through proteomics. What has happened? In 2002, Petricoin and colleagues⁶ published a seminal article in *The Lancet* that brought serum proteomics to the forefront of biomarker discovery paradigms. They suggested that a methodical searching algorithm could identify unique serum proteomic patterns that would completely discriminate cancer from noncancer samples in their surface enhanced laser desorption ionization (SELDI)-derived profiles. Their discrimination between benign and early-stage ovarian cancer used five key “peaks” identified from a set of 116 masked serum samples (50 from women with ovarian cancer and 66 from unaffected women). After publication of these data, several other groups came out with data that supported the notion that Ciphergen’s iterative learning algorithm could similarly distinguish between other cancers.^{7,8} This pushed the field of serum proteomics into the spotlight of biomarker discovery. Unfortunately, it quickly became a spotlight of suspicion as initial successes with this method came under question.

Almost immediately, these studies were scrutinized because the peaks were not identified.⁹ Additionally, later studies challenged the actual significance of the conclusions of the study on the presumed basis of technical flaws in both instrument operation and the execution of experimental protocol.¹⁰ Reanalysis of the data from the Petricoin lab suggested that inconsistencies existed due to improper mass calibration.¹⁰ Although the original authors addressed many of these criticisms, overall, the field of biomarker-discovery proteomics is currently

viewed with some skepticism with regard to significance and reproducibility.¹¹ Perhaps the ultimate conclusion to be drawn from these initial studies is a reaffirmation of the importance of standardization of sample handling, a rigor in data analysis, and the importance of open data, data standards, and communication.

Proteomic Analysis of Blood

Our initial successes and failings have become the guideposts for proteomic experimentation. Of the many possible human materials to profile (eg, saliva, urine, cerebrospinal fluid, synovial fluid), the current trend is to use blood. It is readily obtainable and, because it circulates throughout the entire body, should represent a holistic sampling of the individual.³ The question of measuring via plasma or serum remains. Although nonclinicians often use the two terms interchangeably, the differences between plasma and serum have ramifications for proteomic analysis. Plasma is the noncellular liquid component of blood collected in the presence of an anticoagulant (eg, heparin, sodium citrate). Conversely, serum is the noncellular liquid component of blood that remains after allowing the blood to clot. Misek and associates¹² have recently studied some of the difference in protein content of plasma and serum. As expected, the main difference, by mass, is in the level of coagulation proteins like fibrinogen and thrombin. When deciding whether to use serum or plasma, a clinician will usually opt for the specimen most suitable for subsequent laboratory tests; because the anticoagulants used in plasma collection can interfere with many laboratory tests, serum tends to be more versatile. However, some proteomicists have become wary about using serum. There are several sources of variability that manifest in a mass spectral profiling experiment including clotting time, incubation temperature, and differential loss of proteins associated with the clotting factors.¹³ For good biomarker discovery data—for both plasma and serum—it is important to remove as much variability as possible and to create and follow protocols that limit leeway in sample handling (eg, always collect the same amount of blood in the same brand of tubes; let them incubate at the same temperature for the same duration before performing a constant, fixed-time centrifugation).¹⁴

Metaphorically, hunting for differential biomarkers using proteomics is very similar to panning for gold. Although human serum contains no silt, it does contain a small number of proteins that make up the bulk of mass within a blood-derived sample, which many consider to be worthless for biomarker discovery. Albumin, immunoglobulins, haptoglobin, α 1-anti-trypsin, and transferrin alone represent 85% of the total protein mass.¹⁵ Even worse, these few proteins measured in the mg/mL range

represent one tail of a very wide dynamic range of proteins present within the blood stream with known cytokines (pg/mL) as the other.¹⁶ This dynamic range, spanning more than 12 orders of magnitude, is beyond the capabilities of even the best of current mass spectrometric technologies to analyze within one scan.

Although MS-based techniques are indeed powerful, they have limitations. Contrary to a common belief (and to what one might expect from the price), mass spectrometers are not the best tools for all scientific questions. For example, if the identity of the protein in question is known, enzyme-linked immunosorbent assays (ELISAs) and radioassays are both significantly cheaper and more sensitive than mass spectrometers for the purposes of detection and quantitation.

Consider, for example, the ELISA-based measurement of prostate-specific antigen (PSA) levels in the ng/mL range for the management of prostate cancer. Measurements of antibody and chemoluminescence reagent levels are cost effective. Assay readers can perform hundreds of measurements per minute. There is no reason to replace these types of assays with a lower-throughput, higher-cost mass spectrometric-based solution. On the other hand, if one wishes to analyze all the possible yet currently unknown PTMs to PSA, ELISAs are unambiguously worthless. Additionally, in the case where the researcher is searching for novel differences between samples (ie, biomarker discovery) MS arguably becomes the first tool of choice; there is no practical protocol for the development of an immunoassay targeting a modified site if neither the site nor the modification is known. This is the type of problem where MS excels.

In spite of the ability of mass spectrometers to work within the realm of analyzing unknowns, targeted questions are still easier to answer than profiling questions. For instance, if you are looking for a 10% difference in a random subset of signals, it will be difficult to make conclusions of statistical significance. On the other hand, if one wishes to identify a protein in a given band of a given gel, or within a given spectrum using peptide mass fingerprinting, it is easier to obtain convincing evidence to support your claim (Tables 1 and 2).

Biomarker Discovery at the Clinic

Successful use of proteomics to identify novel biomarkers will require close collaborations between clinical, basic, and statistical/computational research teams. Samples come from clinics, usually in the context of a clinical trial. The efforts of many people are required before any results are seen: the phlebotomist drawing the blood, the technician centrifuging the blood, the courier transporting the plasma/serum, the technician who prepares the sample for

Table 1. Common Problems and Solutions Encountered in Biomarker Discovery Using Proteomics via Mass Spectrometry

Common Problems	Common Solutions
High abundance protein signals overshadowing potential lower abundance biomarker signals	Immunodepletion and multidimensional chromatography
Sample collection variability	Randomization, limited number of freeze/thaw cycles, consistency, and strict protocols
Sample handling errors	Randomize samples and run replicates for each sample
Overfitting the data	Increase sample count
Machine fluctuation	Regular machine calibration and frequent quality-assurance sample analysis

Table 2. Basic Rules for Applying Proteomic Techniques

1. If you know your target, mass spectrometry is probably more expensive and less sensitive than alternative techniques.
 - ELISA and radioassays are more sensitive than a mass spectrometer.
2. If you do not know your target, mass spectrometry is probably the only viable technique.
 - It would be difficult to develop an immunoassay for a specific post-translational modification, if its site was unknown.
3. Targeted questions are easier than profiling questions.
 - It is easier to identify an isolated protein than to find statistically significant differences in a forest of signals.

ELISA = enzyme-linked immunoabsorbant assay.

analysis, the technician that performs the analysis, and the bioinformaticist that distills the measurements into statistical observations. Below, this process will be explored in detail in order to discuss the points of significance that are required for a successful collaboration.

Consider an example in which the difference in the abundance of relevant biomarkers in an arbitrary example of clinical utility is on the order of 15%. Then consider how biological samples are collected. There is no single standardization in sample collection; nevertheless, standardization is a minimal requirement for obtaining significant results. It is important to consider how the blood was drawn. For how long did it clot? At room temperature or on ice?¹⁷ For how long was the sample centrifuged? Were samples immediately frozen after separation or was there a delay?¹³ All of these differences have been shown

to alter the resultant profiles obtained from MS. Even seemingly innocuous factors such as whether the patient is lying down or sitting, patient stress associated with the blood collection, type of collection (venipuncture, arterial puncture, skin puncture), tourniquet technique, order of sampling (first, second, third tube), and quantity of blood withdrawn affect the composition of the specimen and have been shown to affect the resultant profiles.¹⁸ After the blood is processed it is typically stored in large aliquots and frozen. Whenever a request for material is made, one of these large aliquots will be thawed, the requested volume removed, and the aliquot refrozen. Studies have shown, however, that freeze/thaw cycles can cause significant variations in the profile spectra.¹⁸ It would be optimal, at the time of collection, to fractionate a sample into numerous aliquots and freeze them separately for individual use. Because of the many pressures involved in running a clinical laboratory, it is sometimes not feasible to implement all the levels of rigor that one would desire for an optimal level of reproducibility and sample fidelity; nevertheless, these are important issues that should be considered.

Randomization is necessary to guard against observational bias. If, as a researcher, you are attempting to identify a biomarker, you do not want to accidentally identify the banal: if all your “healthy” samples come from hospital A and all your “diseased” samples come from hospital B then, almost assuredly, when the results are validated, the conclusions will include a marker of which hospital produced the sample.¹⁹ Within a clinical trial design, and even within any single center, it is important to avoid any source of systematic bias.

Biomarker Discovery at the Wet Lab

After a sample arrives in the research laboratory from the clinic, it needs to be prepared for analysis. Depending on the questions being asked and the problems that need to be addressed, many of these steps will be common across laboratories (eg, a core “denature, reduce, alkylate, digest” protocol) while other steps will vary greatly (eg, choice of chromatography protocols). It is broadly accepted that the known proteins in blood span a concentration range of more than 12 orders of magnitude, which creates problems when trying to profile all of these species simultaneously.⁴ What is less accepted is what, if anything, should be done to reduce this complexity prior to analysis. For researchers that have decided to do preanalysis sample simplification, there are two common approaches, depletion and enrichment.

The purpose of depletion protocols is to reduce the large bulk of proteins from a sample with a concomitant increase in sensitivity for the remaining analytes. There are

several common biochemical methods for depleting (eg, precipitation of larger species with high concentrations of organic solvents, somewhat specific depletion of albumin by the use of a Cibacron blue), but the techniques that seem to be gaining the most acceptance currently are those based on immunodepletion. There are commercially available immunodepletion reagents that can remove up to 10 of the most highly abundant proteins from serum.²⁰ By mass, the top 10 proteins represent over 99% of the protein in plasma/serum. The arguments against depletion frequently entail issues regarding reproducibility and the lack of absolute specificity toward the targeted proteins. Albumin, for example, is known to associate with other proteins and peptides²¹; therefore, when the albumin “sponge” is removed, some of the low-abundance species that are trying specifically under analysis can be lost.²² This is particularly true for peptides and other smaller analytes that otherwise might be filtered out of the blood stream by the kidneys. Recent studies have shown that there is a group of low-abundance peptide biomarkers produced from specific and ongoing tumorigenic processes including apoptosis, tumor-stromal interaction, vascularization, immune cell infiltration, and antigenic processing that exist in a sequestered state, complexed to highly abundant resident blood proteins such as albumin.^{16,23,24} Thus, attempts to deplete high-abundance proteins may also remove these low molecular weight molecules, which exist in circulation as protein-protein complexes.²³ Studies comparing IgG- and IgY-based immunodepletions find that the IgG-based system provides a more effective removal of targeted proteins, with minimum carryover, high longevity, and minimal nonspecific binding.²⁵

To aid in protein/peptide quantification, one can chemically modify the sample of interest and mix it with a control sample that has been altered with a chemically identical yet isotopically different reagent. The isotope-coded affinity tag (ICAT) is perhaps the best-known of these reagents for stable isotope labeling and serves as a good paradigm for understanding similar reagents.²⁶ The ICAT reagent consists of three components: a reactive group that reacts with the free thiol functionality of cysteine residues; a linker in which stable isotopes have been incorporated (“light” with 4 hydrogens or “heavy” with 4 deuteriums); and a biotin tag for affinity isolation.²⁷ Because the ICAT reagent is specific for cysteines (other reagents can be specific for amino groups or other reactive moieties in addition to sulfhydryls), the complexity of the original peptide mixture is greatly reduced as only a few tryptic peptides in any protein contain cysteines. Peptides labeled with ICAT will copurify with their matched controls and will be visualized as peak pairs separated by approximately 4 Da on the mass spectrometer. By calculating the peak areas for matched peptide peaks

labeled with the light and heavy versions of the reagent, the relative abundance of that peptide in each sample can be determined, which is directly related to the abundance of the parent protein.²⁷

These methods are, unfortunately, not universally applicable. Although tagging/capturing technologies work well in narrowing down the proteomic dynamic range, large portions of the proteome that may be proven useful in later studies are discarded. ICAT, for example, is not only blind to the significant number of proteins lacking cysteine residues, but it is blind to any noncysteinated peptide as well, making it ill-suited for the analysis of PTMs or protein isoforms generated by alternative mRNA splicing.²⁸ It has also been noted that the ICAT tag is a relatively large molecule compared to the small peptides it labels, which may result in interference with peptide ionization and inaccurate mass spectra.²⁹ Additionally, a problem with all techniques that use isotopic heavy/light labels for differential expression analysis is that they rely on the comparison of two (or more) isotope envelopes. If one condition involves the complete repression of a protein, that will result in the presence of only a single envelope, which may be ignored depending on the sophistication of the analysis software; ironically, the effect is to be blinded to conditions that involve the transition from expression to complete repression.

Enrichments can also be performed on peptide modifications. For example, several glycopeptide enrichment techniques—such as lectin-affinity chromatography,³⁰ hydrophilic interaction solid-phase extraction,³¹ and covalent immobilization of glycopeptides using hydrazide chemistry³²—have been developed to isolate glycopeptides resulting from tryptic digests of complex glycoprotein mixtures.³³ In addition to the direct link between the expression of the glycoconjugate biosynthetic genes and pathways involved in oncogenic transformation,³⁴ glyco-capture typically leads to a large reduction in sample complexity.³² However, if you are going to add additional steps, reagents, incubations, and processes to your sample preparation, it is imperative that the benefit to sample simplification outweigh the additional sources of randomization that have been introduced. For example, one laboratory that has been performing routine hydrazide-based glycopeptide enrichment for over two years recently had a dramatic drop in yield, which they ultimately traced to a change in a reagent supplier’s method of protein purification (P. Mallick, written communication; 2006.)

As another possible means of increasing reproducibility, many researchers have considered the addition of protease inhibitors to their samples as a preservative during sample processing. However, this too has been met with controversy as some putative biomarkers are hypothesized to be exoprotease fragments derived from

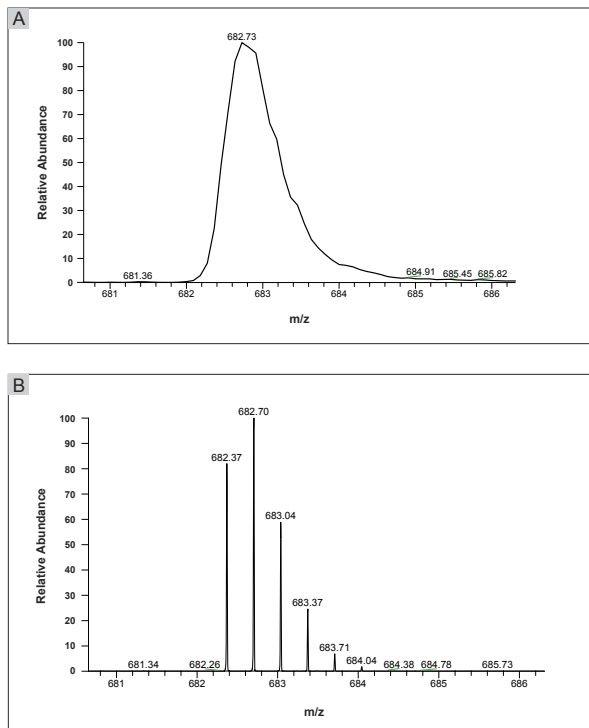


Figure 2. Advances in mass spectrometry technology. The top spectrum (A) is from a linear ion trap (LTQ) with unit resolution—similar to a quadrupole mass spectrometer. (B) is the mass spectrum of the same analyte as measured by a Fourier transform ion cyclotron resonance mass spectrometer (FTICRMS), 7.0 Tesla, 0.72 s acquisition. Where the ion trap sees only a single peak, the FTICRMS is able to resolve that peak into multiple distinct species of similar m/z .

larger plasma proteins.³⁵ Exoproteases are a heterogeneous group of enzymes that play a role in the regulation of biologically active peptides.³⁶ They are part of a larger set of proteases recently referred to as the “degradome.”³⁷ Protease inhibitors added at the time of sample collection will typically block the activity of these inhibitors, which have been considered “background noise” in peptide marker discovery efforts.³⁸ However, studies conducted by Villanueva and coworkers³⁸ have shown that a small subset of serum peptides created as a result of exoprotease activity can provide accurate class discrimination between prostate cancer, breast cancer, bladder cancer, and nondiseased states—peptides that are lost by the addition of protease inhibitors. Proteomics studies concerning peptide patterns of the degradome may allow mechanistic insights into specific activities of tumor-associated proteases.³⁸

After sample preparation, one must consider the possible variability that can be introduced through machine operation. A common proteomic pipeline will consist of loading the sample into an LC flow coupled to a mass spec-

trometer (LC/MS).³ Briefly, the sample is separated over time based on a chemical characteristic such as hydrophobicity. During this separation the mass spectrometer samples and produces spectra for the analytes that are eluting at that particular moment. If machine fluctuations dwarf the biologically relevant signals, they will remain unidentified. In the field of MS, mass accuracy is fundamental to correctly identifying and classifying proteins/peptides. When data are compared across runs, it is important to identify the corresponding features: does the peak with an m/z of 500.1234 that elutes at 20 minutes in one run represent the same or a different species than a peak with an m/z of 500.1338 that elutes at 22 minutes in a different run? For the LTQ-FT (Thermo Electron Corporation), a hybrid mass spectrometer comprised of a linear ion trap coupled in tandem to a Fourier transform ion cyclotron resonance mass spectrometer, the vendor publishes a mass accuracy of better than 2 ppm with external calibration, a resolving power greater than 500,000, and sub-fmol sensitivity. If one assumes the machine is operating within tolerance, then one can conclude that the 500.1234 peak represents a different species than the 500.1338 peak. If, however, mass calibration was not maintained, then it is quite possible that these two signals might reflect the same species (see Figure 2 for sample data).

In addition to the above-mentioned sources of variability, the importance of randomization and blinding after the samples are received by the research laboratory cannot be stressed enough. Similar to the biologic factors that may be discovered based on handling in the clinic, after samples reach the laboratory it is imperative that we warrant against discovering markers that identify biologic factors, such as which operator prepared the sample, which machine was used to analyze the sample, which lot number of trypsin was used, and what day the machine was calibrated. Further, by coding the samples, the operators and researchers are unable to allow their biases to affect the experimental outcome.¹⁰

Biomarker Discovery With the Bioinformaticist

After all of the samples have been collected, prepared, and analyzed via LC/MS, because vast quantities of data have inevitably been produced, the analysis of these data is usually performed through automated means. Unfortunately, multivariate statistical analysis and machine learning algorithms will always produce results, but not necessarily significant ones. Further, it is not always obvious how to assess the level of significance. Perhaps the most common artifact of such analyses is overfitting.¹⁹ As explained by Ludwig and Weinstein,³⁹

In multivariate predictive analysis, a statistical model can be overfitted if it has too many free parameters for the number and type of cases in the training set. The result can be a model that fits the training data set very well but does poorly when applied to other data.

For example, if 10 people come into your office and you ask them their shoe size, where they were born, how many vowels are in their name, and 1,000 other such arbitrary questions, you will inevitably find some coincidences such as, “if the people’s names contains an even number of letters, then they are male, otherwise they are female” or “if their names begin with a vowel, then they drive red cars.” If specificity is calculated as the ratio of the number of negative samples (nondiseased) correctly identified to the total number of true negative samples in the population, and sensitivity is calculated as the ratio of the number of samples correctly classified as positive (diseased) to the total number of samples known to be from diseased individuals,⁴⁰ it is frequently possible, with a set of proteomic data, to predict with 100% sensitivity and 100% specificity whether the patient was wearing a hat or not when their blood was collected.* Although these examples would immediately raise suspicion, readers frequently accept conclusions with identical measures of significance when the claim is the presence of cancer rather than the presence of a hat.

Oftentimes the problem of overfitting arises from the assignment of training and validation sets. If you have a lot of variables to measure relative to the number of samples, a common statistical trap is to assume that you can predict with 100% accuracy something about the samples. To correct for this, many studies divide the initial sample into training and validation sets. The training data are used for selecting features and training a classifier. Therefore, some measurements are made in part of the total sample (training set) to derive a decision rule that is then applied on the reserved samples (validation set). When the number of samples is small, it is important to ensure that the data used to test the classifier are not part of the data used to train it. This gives the resubstitution estimate of the true error, which is known to give falsely low error estimates for small samples.⁴¹

The initial papers on proteomics published results based solely on mass spectral peaks without any link to the underlying biological source. This, along with a changing landscape in equipment and a lack of standards for inter-laboratory validation, severely hampered progress in the field of proteomics. Perhaps because of these negative early experiences, there is now very strong pres-

* Analogy courtesy of Parag Mallick.

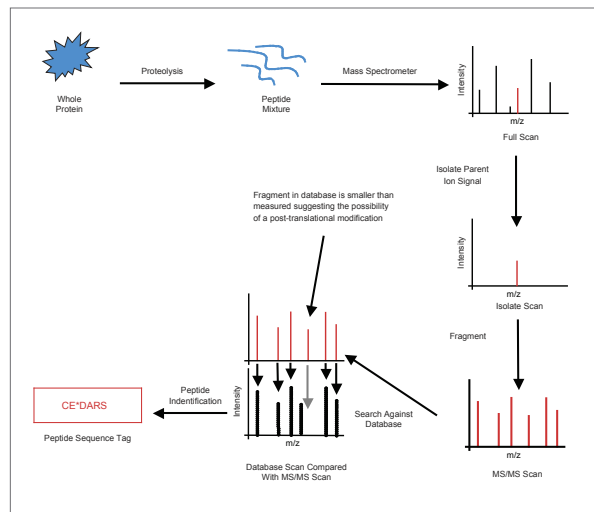


Figure 3. Peptide identification by MS/MS sequencing. Whole protein is proteolyzed into peptide fragments. These fragments are analyzed by the mass spectrometer and a single m/z signal is selected as the parent ion. This parent ion is isolated and fragmented, creating daughter ions which are measured in an MS/MS scan. These MS/MS scans are compared against predicted values for all the possible fragments based on the human genome project. Matches are scored for significance and reported. Differences can be further analyzed for possible post-translational modifications.

MS/MS = tandem mass spectrometry.

sure to identify any peaks of relevance (although CA125 was considered an effective biomarker when the scope of its characterization was as a somewhat effective antigen in a radioimmunoassay).⁴² Bioinformatics has provided the bridge between observations and identifications. As defined by Luscombe and associates, bioinformatics conceptualizes biology in terms of macromolecules and applies “informatics” techniques to understand and organize the information associated with these molecules on a large scale.⁴³

The utility of bioinformatics-based techniques in the analysis of proteomic data depended on the completion of the sequencing of the human genome. The importance of this comes in the validation of fragments identified during differential profiling experiments. Identification of the peaks of difference is the first step to a biological validation. Databases of DNA sequences are translated *in silico*, creating “cyberproteins” that are subsequently “digested” and the resultant putative masses are used for comparison to actual measured signals. When measured masses match predicted masses, calculations of likelihood can be performed to assign a confidence to the identifica-

tion. After identification, scientists are enabled to study the biological relevance and function of these proteins.

Tandem mass spectrometry (MS/MS) involves the isolation of a single peak in a spectrum (the parent), fragmenting it, and measuring the resultant fragments (Figure 3). These fragments, or daughter ions, can, if desired, be further fragmented into smaller peptides. For the purposes of protein identification, these fragment scans combined with the original peptide masses can be used to search a database of values for known proteins. Increasingly protein identification from MS/MS data is being fully or partially automated.⁵ It should be noted, however, that peak identification does not warrant against poor analysis—all the comments about proper data and sample handling still apply whether peak identifications have been made or not. For example, the claim that peak 537 indicates the patient was wearing a hat can be just as insignificant as the claim that increased hexokinase means the patient was wearing a hat. What identification does provide is additional biological confirmation (does it make sense that hexokinase would be linked to hat-wearing?) as well as a strong lead for additional confirmation (do hexokinase knock-out mice wear fewer hats?) and biological/drug research (if I design a therapy that makes mice produce more hexokinase, can I restore their hat-wearing to a normal level?). Additionally, mass-spectrometric techniques are relatively expensive and slow compared to most high-throughput screens, so when a peak is identified it facilitates the development of high-throughput screens based on inexpensive technologies such as ELISAs. That said, routine high-throughput newborn screenings based on MS/MS have proven very feasible and of exceptional value in diagnosing many metabolic deficiencies that would otherwise have only been diagnosed after the manifestation of clinical symptoms.⁴⁴

Standardization and the Reporting of Data

Preliminary guidelines for the publication of proteomics data were recently published.⁴⁵ Due to the rapidly growing number of datasets obtained from MS-derived protein identifications, significant but undefined numbers of the proteins reported as “identified” in proteomics articles are likely to be false positives.⁴⁶ The month after the publication of the preliminary guidelines, 30 participants met in Paris to further discuss and elaborate on these guidelines. (These drafted guidelines can be found at www.mcponline.org.)⁴⁷ The guidelines can be divided into recommendations for methodologies and requirements for the reporting of data. For example, if peptide identifications are reported, authors must document what search engine was used and how peptide and protein assignments were made using that software.⁴⁵ Additionally, it is strongly

encouraged that MS/MS data mentioned in any paper be submitted as supplemental material. These guidelines also take into account the loss of connectivity between peptides and proteins that occurs with proteolytic digests of whole proteins. When doing peptide identification after proteolysis, one cannot tell if multiple identical fragments are from more than one protein sequence (bioinformatics redundancy) or if more than one precursor was in the original protein mix (physical redundancy) (www.mcponline.org/misc/ParisReport_Final.shtml). Scientists are encouraged to post datasets on the internet, as it is an invitation for others to test the validity of their claims. Although these guidelines are not accepted universally, efforts are being made to revise and amend these guidelines, as well as implementing them in scientific journals.

Conclusions

We are not yet at the point of personalized medicine. Biomarkers such as PSA tend to be remarkably poor surrogates for the actual disease. In diagnosis, 15% of individuals found to be positive for prostate cancer have PSA levels below 4 ng/mL, ie, in the normal zone.⁴⁸ Yet even with its poor utility, PSA is still commonly used as a diagnostic tool for the detection of prostate cancer. Often, it is PSA levels that prompt aggressive therapy in patients. Additionally, although much of biomarker discovery has focused on diagnosis, and although diagnosis of a disease is clearly important in the management of patients, there are little data to demonstrate that earlier diagnosis for many diseases yields a more favorable outcome.⁴⁹ It is clear that clinicians need better biomarkers. Our laboratory is specifically and initially concentrating on advanced disease and the outcome of therapeutic intervention for our biomarker discovery projects. It is important that we choose datasets that will help a clinician make a decision. The process of validating early diagnostics will take many years to demonstrate an effect on outcome (if, in fact, there is one). Ultimately, the oncologic community wants to use biomarkers for the detection, diagnosis, and prognosis of our patients and all three are integral to appropriate patient management and optimal clinical outcomes. It is only through rigorous application of these promising technologies that we will achieve this goal.

References

1. Kavallaris M, Marshall GM. Proteomics and disease: opportunities and challenges. *Med J Aust.* 2005;182:575-579.
2. Petricoin E, Wulfkuhle J, Espina V, Liotta LA. Clinical proteomics: revolutionizing disease detection and patient tailoring therapy. *J Proteome Res.* 2004;3:209-217.
3. Katz JE, Mallick P, Agus DB. A perspective on protein profiling of blood. *BJU Int.* 2005;96:477-482.
4. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics.* 2002;1:845-867.

5. Barrett J, Brophy PM, Hamilton JV. Analysing proteomic data. *Int J Parasitol.* 2005;35:543-553.
6. Petricoin EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet.* 2002;359:572-577.
7. Adam BL, Qu Y, Davis JW, et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res.* 2002;62:3609-3614.
8. Li J, Zhang Z, Rosenzweig J, Wang YY, Chan DW. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem.* 2002;48:1296-1304.
9. Diamandis EP. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Mol Cell Proteomics.* 2004;3:367-378.
10. Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics.* 2004;20:777-785.
11. Liotta LA, Lowenthal M, Mehta A, et al. Importance of communication between producers and consumers of publicly available experimental data. *J Natl Cancer Inst.* 2005;97:310-314.
12. Misek DE, Kuick R, Wang H, et al. A wide range of protein isoforms in serum and plasma uncovered by a quantitative intact protein analysis system. *Proteomics.* 2005;5:3343-3352.
13. Banks RE, Stanley AJ, Cairns DA, et al. Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. *Clin Chem.* 2005;51:1637-1649.
14. Tammen H, Schulte I, Hess R, Menzel C, Kellmann M, Schulz-Knappe P. Prerequisites for peptidomic analysis of blood samples: I. Evaluation of blood specimen qualities and determination of technical performance characteristics. *Comb Chem High Throughput Screen.* 2005;8:725-733.
15. Chromy BA, Gonzales AD, Perkins J, et al. Proteomic analysis of human serum by two-dimensional differential gel electrophoresis after depletion of high-abundant proteins. *J Proteome Res.* 2004;3:1120-1127.
16. Tirumalai RS, Chan KC, Prieto DA, Issaq HJ, Conrads TP, Veenstra TD. Characterization of the low molecular weight human serum proteome. *Mol Cell Proteomics.* 2003;2:1096-1103.
17. Hulmes JD, Bethea D, Ho K, et al. An Investigation of Plasma Collection, Stabilization, and Storage Procedures for Proteomic Analysis of Clinical Samples. *Clin Proteomics J.* 2004;1:17-32.
18. Vitzthum F, Behrens F, Anderson NL, Shaw JH. Proteomics: from basic research to diagnostic application: a review of requirements & needs. *J Proteome Res.* 2005;4:1086-1097.
19. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer.* 2004;4:309-314.
20. Pieper R, Su Q, Gatlin CL, Huang ST, Anderson NL, Steiner S. Multi-component immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome. *Proteomics.* 2003;3:422-432.
21. Fasano M, Curry S, Terreno E, et al. The extraordinary ligand binding properties of human serum albumin. *IUBMB Life.* 2005;57:787-796.
22. Zhou M, Lucas DA, Chan KC, et al. An investigation into the human serum "interactome". *Electrophoresis.* 2004;25:1289-1298.
23. Liotta LA, Petricoin EF. Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. *J Clin Invest.* 2006;116:26-30.
24. Lowenthal MS, Mehta AI, Frogale K, et al. Analysis of albumin-associated peptides and proteins from ovarian cancer patients. *Clin Chem.* 2005;51:1933-1945.
25. Zolotarjova N, Martosella J, Nicol G, Bailey J, Boyes BE, Barrett WC. Differences among techniques for high-abundant protein depletion. *Proteomics.* 2005;5:3304-3313.
26. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol.* 1999;17:994-999.
27. Li J, Steen H, Gygi SP. Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents: the yeast salinity stress response. *Mol Cell Proteomics.* 2003;2:1198-1204.
28. Patton WF, Schulenberg B, Steinberg TH. Two-dimensional gel electrophoresis; better than a poke in the ICAT? *Curr Opin Biotechnol.* 2002;13:321-328.
29. Haynes PA, Yates JR, 3rd. Proteome profiling-pitfalls and progress. *Yeast.* 2000;17:81-87.
30. Kaji H, Saito H, Yamauchi Y, et al. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol.* 2003;21:667-672.
31. Hagglund P, Bunkenborg J, Elortza F, Jensen ON, Roepstorff P. A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J Proteome Res.* 2004;3:556-566.
32. Zhang H, Li XJ, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol.* 2003;21:660-666.
33. Alvarez-Manilla G, Atwood J 3rd, Guo Y, Warren NL, Orlando R, Pierce M. Tools for glycoproteomic analysis: size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites. *J Proteome Res.* 2006;5:701-708.
34. Orntoft TF, Vestergaard EM. Clinical aspects of altered glycosylation of glycoproteins in cancer. *Electrophoresis.* 1999;20:362-371.
35. Mattrisian LM, Sledge GW Jr, Mohla S. Extracellular proteolysis and cancer: meeting summary and future directions. *Cancer Res.* 2003;63:6105-6109.
36. Silveira PF, Gil J, Casis L, Irazusta J. Peptide metabolism and the control of body fluid homeostasis. *Curr Med Chem Cardiovasc Hematol Agents.* 2004;2:219-238.
37. Lopez-Otin C, Overall CM. Protease degradomics: a new challenge for proteomics. *Nat Rev Mol Cell Biol.* 2002;3:509-519.
38. Villanueva J, Shaffer DR, Philip J, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest.* 2006;116:271-284.
39. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer.* 2005;5:845-856.
40. Roboz J. Mass spectrometry in diagnostic oncoproteomics. *Cancer Invest.* 2005;23:465-478.
41. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics.* 2006;7:91.
42. Bast RC Jr, Klug TL, St John E, et al. A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. *N Engl J Med.* 1983;309:883-887.
43. Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. *Methods Inf Med.* 2001;40:346-358.
44. Frazier DM, Millington DS, McCandless SE, et al. The tandem mass spectrometry newborn screening experience in North Carolina: 1997-2005. *J Inherit Metab Dis.* 2006;29:76-85.
45. Carr S, Aebersold R, Baldwin M, Burlingame A, Clauser K, Nesvizhskii A. The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol Cell Proteomics.* 2004;3:531-533.
46. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem.* 2002;74:5383-5392.
47. Bradshaw RA, Burlingame AL. From proteins to proteomics. *IUBMB Life.* 2005;57:267-272.
48. Thompson IM, Pauler DK, Goodman PJ, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. *N Engl J Med.* 2004;350:2239-2246.
49. Ransohoff DF, McNaughton Collins M, Fowler FJ. Why is prostate cancer screening so common when the evidence is so uncertain? A system without negative feedback. *Am J Med.* 2002;113:663-667.