Editorial

Defining the Role of Mass Spectrometry in Cancer Diagnostics

O. John Semmes

Department of Microbiology and Cell Biology, Eastern Virginia Medical School, The Virginia Prostate Center, Norfolk, Virginia

The prospect of uncovering the information contained in the human genome garnered such great promise that it attracted a synergistic blend of researchers from multiple disciplines. Notably, this group included molecular biologists, clinician-scientists, computational mathematicians, and industry input via engineers and, in several cases, management expertise in the form of leadership. This synergistic grouping drove tremendous technological advancement that resulted in the completion of this multinational research project with unprecedented success. The promise of proteomics is proving to be an even greater allure than its predecessor, genomics.

Specifically for this discussion, the resurgence of clinical proteomics has energized the stagnating field of cancer biomarker discovery. Although the field of clinical proteomics includes far too many promising technologies to adequately address here, it is safe to say that mass spectrometry (MS), with the great strides in application to protein research, has positioned itself as a key technology platform for biomarker discovery. The promised return of our investment in clinical proteomics is nothing short of precise molecular diagnostics. Again, much as was the case for genomics and indeed inclusive of the field of functional proteomics that is led by many of the same research groups that pushed forward genomics, clinical proteomics is being driven by a diverse group of integrated scientific expertise. Specifically, the active front in the expanding arena of protein biomarker discovery is composed of experts in MS that have been diligently pushing the technical limits of MS, biologists from the postgenomics generation that are applying MS to the new horizon of proteomics and the ever-responsive engineering/industrial segment. In addition, biostatisticians, mathematicians, and computational biologists, already motivated by their achievements in interpreting genomics data, have responded to a similar call for help in analysis of proteomic data. Rounding out this formidable combination of expertise are the clinician-scientists and epidemiologists who provide a critical component in responding to the challenges of developing protein biomarkers for cancer molecular diagnostics.

Unprecedented capabilities in sensitivity and resolution coupled with the development of complimentary enabling computational approaches have allowed for dramatic increase in the ability to obtain sequence iden-

tification from complex solutions, perform fine-scale structure analysis, determine protein-protein interactions, and map post-translational modifications. In spite of the tremendous advances made in protein MS following the completion of the genomes of target organisms, some of the same old obstacles to protein biomarker discovery still exist. The physical hurdles of component complexity, the tremendous range in individual protein concentration, and the dynamic nature of the proteome are still major barriers to overcome. In addition, the added demands associated with targeting proteins that will function as clinical biomarkers of a disease as genetically diverse as cancer present a unique challenge for the field.

It is becoming increasingly clear that the underlying heterogeneity of cancer necessitates that development of accurate diagnostics will be dependent on the discovery of a ''panel'' of proteins that together can discriminate between subtle disease states with population-wide robustness. This requirement for multiple proteins over a single protein biomarker emphasizes the demand for improved technologies, allowing for measurement of the full compliment of the proteome, ''the proteome volume'' (for a useful review of many of the current proteomic technologies, see refs. 1-4). In conceptualizing where the technical challenges lie, it is useful to recall the proteomics equivalent to the Schrödinger equation in that we can know everything about one protein or we can know a little about many proteins. The interrelationship between the complexity of the sample and the accuracy of the MS defines today's technical barrier albeit with a changing battle line. To arise from this quandary, researchers have chosen to develop complimentary ''up-front'' approaches to reduce the complexity prior to the application of high-resolution MS. Essential elements of improved ''up-front'' methods revolve around reducing the tremendous disparity in the concentration of individual proteins. Although automation of the standard practice of selective removal of abundant proteins followed by concentration is a viable approach toward observing the less abundant proteins, there are also nascent efforts under way to employ selective affinity chemistries that will equalize the concentration of the displayed proteins. Alternatively, several ''shotgun'' approaches, which attempt to profile the differentially expressed proteins of large segments of the proteome, are being realized with the help of greatly improved MS resolving power and enabling computational methods. Attacking this problem from the other end is the field of expression profiling in which the state-of-the-art is the ability to reproducibly observe the semiquantitative expression changes of a relatively small number of proteins. The protein expression profiling approach allows for analysis of large patient populations

and thus has the potential for observing robust combinations of protein changes. However, this technique is severely limited with respect to the volume of the proteome, and current and future improvements will surely be forthcoming to address these limitations. In practice, most groups are using a combination of the above approaches in an attempt to mount effective protein biomarker discovery. The need for technical improvements to allow for increased visualization of the proteome coupled with the need for robust biomarkers reflective of the population can be simplistically summarized as ''mining all of the proteome of hundreds of people simultaneously.'' Obviously, we have not yet reached that point technically.

Implied in the concept of a robust biomarker is the requirement of choosing the appropriate study populations for the initial ''discovery'' phase of biomarker research. This point has been driven home again and again in the past as promising new individual protein ''biomarkers'' derived from small sample sets fail to hold up when analyzed in clinical validation. Unfortunately, the patient cohorts that would be ideally suited for discovery efforts are currently only available through large study trials that typically limit their samples to the validation of existing biomarkers. This is an understandable consequence of the sample rarity and cost of generating these studies and is a legitimate restriction on biomarker discovery that in return fulfills the very critical needs of validation. Indeed, there exists now a great unmet need to generate prospectively collected patient cohorts, ideally from population-based studies as well as clinical trials, which are specifically intended to support discovery studies as corollary activities. The need for defined sample cohorts for cancer biomarker discovery should not be underestimated.

As a consequence of the need for defining panels of biomarker proteins that are robust population discriminators, methodologies that facilitate high-throughput analysis are essential to the successful discovery phase. Throughput in this context is strictly limited to the number of individual patient samples that can be analyzed. In response to the demands associated with analyzing many samples simultaneously, the development of automation via robotics has descended on the sample processing steps of biomarker discovery. These procedures typically involve the fractionation, depletion, concentration, and other sample manipulation steps prior to MS analysis. The processing steps should be tightly integrated with the MS analysis to insure maximum reproducibility. Indeed, any successful preanalysis processing step should be fully automatable.

The use of cancer model systems developed in mouse and rat have also been targeted for protein biomarker discovery. The utilization of an appropriate animal model has the advantage of immediately reducing the demand on throughput by analyzing cancers in a syngeneic background as well as serving to better control for potential sample collection biases. Appropriately, these models have been used in the discovery of therapeutic targets and the elucidation of molecular events that define cancer, and the power of these systems is undeniable. However, because organ cancers derive from multiple and heterogeneous genetic events, any clinically useful biomarker will by necessity have to transcend a single defined genetic background. Specifi-

cally, in most animal models, the organism does not normally get cancer and the model specific cancer may not present a proteome comparable with cancer in human. Thus, the big challenge for protein biomarker discovery efforts using animal models is how to integrate the vast amounts of information gained into the development of cancer biomarkers for human application. In this respect, aggressively supporting the building of public repositories for proteomic data, such as that proposed by the Human Proteome Organization, gathered from multiple disciplinary endeavors is warranted. The animal models will undoubtedly contribute to cataloging disease-specific protein classes and at the very least give powerful insight into designing protein class-specific capture approaches that can be applied to human studies.

The ultimate hurdle for any new biomarker is validation. The requirements of validation of a biomarker have been nicely presented by Pepe et al. (5). The points presented herein referencing the need for a biomarker to be reproducible and robust across populations are poignantly represented in fulfilling the subsequent demands of a preclinical and clinical validation. Validation is the great equalizer of all technologies and the sole measure of success. The success of any biomarker as well as the long-term survival of any biomarker discovery approach will eventually be decided at the validation level. In this respect, biomarker accuracy (variously defined as protein identification, relationship of the biomarker to disease, or rather the ability to see what we expect to see) is trumped by biomarker precision (defined as seeing the same disease-specific event again and again). Apropos the recent controversy surrounding protein expression profiling, the identity of the individual components of a biomarker panel is not in itself important so long as the measured panel is reproducibly precise with respect to disease detection. In light of the critical role that validation plays in biomarker development, the need for collaborative research groups that have placed biomarker development as the single overriding goal are essential. These groups could emulate current successful strategies such as the National Cancer Institute's Early Detection Research Network, emerging efforts in corollary research programs within large clinically oriented efforts such as the Southwestern Oncology Group or through increased collaboration of individual SPORE programs with members of either of the aforementioned programs.

In parallel with the incorporation of MS at the discovery phase of biomarkers, there is also emerging success in the application of MS as a diagnostic tool. One example of this is MS-assisted immunoassays, an approach that has been elaborated on elsewhere (6). Central to defining new ground in immunoassay development is the ability of MS to accurately and sensitively measure a specific antigen. Thus, any form of the antigen that can bind a chosen antibody can be measured and the need for a sandwiched second antibody is eliminated. In fact, a useful protein biomarker may be a fragment of a whole protein and would not generate a signal in a classic ELISA that could easily be separated from a signal arising from the uninformative parent protein (7, 8). If the protein fragment biomarker was generated via selective proteolysis, it would be clear that, although the enzyme may be thought of

as the ''true'' biomarker, the fragment is the amplified readout. An example of the application of MS-immunoassay by Wright et al. have shown that the approach holds promise as a reproducible semiquantitative technique (9). Currently, these approaches equal but do not improve the sensitivity and dynamic range of standard ELISA, but as the MS ''reader'' for this immunoassay become more sensitive coupled with the compatibility of MS-immunoassay with multiplexed analysis, we may see more routine application of MS as a diagnostic instrument.

The future success of current efforts in the application of proteomics to clinical diagnostics lies in insuring that both adequate funding and appropriate mix of individual expertise remain as participating components. The synergy provided by strong collaborations among biologists, biostatisticians, chemists, clinicians, engineers, epidemiologists, mathematicians, and physicists, to name a few, will insure this dynamic field continues to move rapidly forward. In addition, it is essential that the connection between initial biomarker discovery and validation be streamlined and that pronouncements on which biomarkers and which biomarker approaches will work and which will not be left to the validation phase. In attempting to push the envelope in applied proteomics, it is essential that ''safe'' methods not predominate ''risky'' methods and that we channel our fears of the novel into the design of appropriate studies that will safeguard from ''fringe'' science. The current dry biomarker pipeline is a call to arms for the development of new approaches that

are rigorously evaluated along the way. Because many of the key enabling technologies will arise from both large proteomic centers and single laboratory researchers, inclusion of all groups with a minimum of undercutting of perceived competitors is a needed ingredient in the current collaborative atmosphere.

References

1. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003;22:198–207.
2. McDonald WH, Yates JR III. Shotgun proteomics: integrating technologies to answer biological questions. Curr Opin Mol Ther 2003;5:302–9.
3. Tyers M, Mann M. From genomics to proteomics. Nature 2003;422:193–7.
4. Yates JR III. Mass spectral analysis in proteomics. Annu Rev Biophys Biomol Struct 2004;33:297–316.
5. Pepe M, Etzioni R, Feng Z, Potter J, Thompson M, Thornquist M. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst 2001;93:1054–61.
6. Petricoin EF, Wulfkuhle J, Espina V, Liotta LA. Clinical proteomics: revolutionizing disease detection and patient tailoring therapy. J Proteome Res 2004;3:209–17.
7. Ye B, Cramer DW, Skates SJ, et al. Haptoglobin-α subunit as potential serum biomarker in ovarian cancer: identification and characterization using proteomic profiling and mass spectrometry. Clin Cancer Res 2003;9:2904–11.
8. Zhang Z, Bast RC Jr, Yu Y, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. Cancer Res 2004;64:5882–90.
9. Xiao Z, Adam BL, Cazares LH, et al. Quantitation of serum prostate-specific membrane antigen by a novel protein biochip immunoassay discriminates benign from malignant prostate disease. Cancer Res 2001;61:6029–33.

Defining the Role of Mass Spectrometry in Cancer Diagnostics

O. John Semmes

Updated version	Access the most recent version of this article at: http://cebp.aacrjournals.org/content/13/10/1555

Cited articles	This article cites 9 articles, 3 of which you can access for free at: http://cebp.aacrjournals.org/content/13/10/1555.full#ref-list-1
Citing articles	This article has been cited by 1 HighWire-hosted articles. Access the articles at: http://cebp.aacrjournals.org/content/13/10/1555.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cebp.aacrjournals.org/content/13/10/1555. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.