Deciphering Protein Molecular Signatures in Cancer Tissues to Aid in Diagnosis, Prognosis, and Therapy

Richard M. Caprioli

Department of Biochemistry and the Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, Tennessee

Abstract
Protein signatures in disease, a portion of the protein complement of cells that directly reflects disease related changes, provide a unique data set that may be correlated with or be an aid to more effective diagnosis, prognosis, and response to therapy. A recent article in this Journal from the author's laboratory described the protein analysis of glioma tissue and the discovery of protein signatures for assessing the stage of disease as well as their correlation with patient survival. This investigation used direct proteomic analysis of glioma biopsies, the discovery of molecular signatures for stage and outcome, and identification of specific proteins of this signature group. This minireview discusses the background and present state of the technology used in this work and the role that direct tissue analysis can play in the discovery of high-quality protein signatures. (Cancer Res 2005; 65(23): 10642-5)

Introduction
The windfall of information ushered in by the extensive knowledge of the human genome has not only provided a leap in our ability to assess the relationship between genes and disease but, perhaps as importantly, has enabled a revolution in proteomics. Proteins take part in a multitude of molecular processes ongoing in the tissue, and their expression levels and molecular forms are a consequence of genomic factors, structural modifications, regulatory processes involving cellular integration and balance (or imbalance), environmental factors, temporal processes, etc. The net sum of these gives rise to a proteome expression level and distribution that reflects the integrated metabolic state of the cells in that tissue at any given point in time.

Protein Signatures: The Combinatorial Advantage
It is highly unlikely that a single protein marker will provide the sensitivity and specificity required for disease detection and prognosis; thus, emphasis has shifted to the discovery of combinations of such markers directly related to disease processes. A multiplicity of such markers make up a molecular "signature" of a given disease phenotype and, further, may be a primary descriptor in the early detection of disease. Especially in cancer, it is clear that early detection is critical to successful management and markedly increased survival. The signature may be made up of some 10 to ≥50 individual protein markers, and as a multivariant ensemble, individual markers may in fact vary for a variety of reasons, but at a statistically significant level, an "effective" signature would remain robust. Of prime importance is the establishment of a high degree of reproducibility in measuring such patterns and the rigorous validation of these with respect to the clinical condition they are meant to address.

This concept has been embraced by a number of investigators, and recent discoveries have shown proof of concept that protein patterns in disease contain both diagnostic and prognostic information (1), and that protein expression profiles and their change with time can provide clues to the understanding of the molecular evolution of disease. Furthermore, changes in protein expression may be useful in treatment decisions where the use of signatures can be predictive of drug efficacy (2). The latter is quite an exciting possibility because it suggests that protein profiling in some relatively accessible tissue could be used to determine the efficacy of drug treatment of an individual patient and moreover could be used to help titrate therapy in that individual. This would provide a vital bridge between simple blood level measurements and overall patient reaction, the latter being a long-term process with, in many cases, highly undesirable side effects. Measurements of proteomic changes early in drug therapy could predict eventual efficacy long before physical changes in the disease are manifest.

There are a number of "proofs of concept" reported using dozens or even several hundreds of patients samples. What are needed are much larger studies with many thousands of patients with known diverse genetic and environmental backgrounds to validate the robustness of these early studies in wide populations. In such an arena, we need to assess to what extent does the temporal state of a protein signature in disease predict the course and outcome at a high confidence level. Similarly, to what extent can the clinician use this information with high confidence to predict the presence of the disease at an early stage and also predict the effectivenss of a given therapy? Can protein signatures predict the risk of disease in nominally healthy individuals? Most importantly, can such proteomic information provide an effective inroad for chemoprevention and help direct and titrate chemotherapy in the individual patient?

Quest for Markers in Serum and Plasma: A Formidable Challenge
Serum and plasma analysis for biomarkers that indicate the presence of disease has been the focus of investigators for many years, because samples are relatively easy to obtain (compared with biopsies), and molecular distributions in these fluids should reflect ongoing metabolic processes of the individual. Serum markers as indicators of disease or risk of disease can be of immense value, provided they are reliable and false negatives are minimal. Again, patterns or signatures are of great importance because single biomarkers have proven to be of limited value. For
example, prostate-specific antigen (PSA) is widely used to screen for prostate cancer (3). Although it is a sensitive assay for a prostate protein that can be elevated in cancer, it is not specific for cancer because other diseases, such as infection of the prostate, benign prostate hypertrophy, and other conditions, can also produce elevated levels of this protein. Similarly, some patients with prostate cancer do not have abnormally high levels of PSA.

Several investigators have used a global approach for the discovery of groups of biomarkers employing technologies, such as two-dimensional gel electrophoresis, to identify differences between normal and disease through comparative proteomics of serum and plasma (4, 5). This has led to the identification of potential markers, some of which may be useful for early detection of disease. Overall, despite the potential value of this approach, significant problems remain. The task to specifically assay a few high-quality biomarkers at exceeding low concentrations, especially in very early onset of disease, is formidable. Moreover, global screening of serum may lead to the identification of markers that reflect non-disease-specific systemic responses of the patient or other processes secondary to the disease. These may include effects of other medical problems, diet, smoking and drinking status, and other factors that can lead to complex and rapidly changing protein patterns not directly related to the primary disease. Work over the past few years has not been abundantly successful despite extreme early excitement and enthusiasm. Several investigators and companies, well intentioned for the most part, were swept up in a wave of enthusiasm and promised rewards far beyond what was achievable at the time. Indeed, such early “hype” is no stranger to human endeavors as evidenced in the recent past in the application of genomics to disease.

The field of proteomics faces formidable challenges because of the enormity of the proteome and its dynamic state. These challenges lie in the area of the dynamic range of proteins in body fluids and solid tissues, the multiple forms present due to modifications and protease activity, the sheer number (estimated to be a million or more distinct molecular forms), the present limitations of current technologies, the lack of rigorous validation for all areas of this process, and the slow progress of global coordination of ongoing efforts. These notwithstanding, many laboratories are working diligently in an attempt to begin to define molecular signatures in disease, and indeed, progress has been made. Some have continued to work directly on serum and plasma hoping to discover those disease-related patterns that one reasonably expects to be present. Others have chosen to study solid human tissues, whereas others use animal models and cell culture techniques.

Molecular Signatures from Tissues

One approach for the discovery of high-quality protein signatures first involves the identification of protein markers directly in tissue biopsy that correlate with the primary disease. This will, for the most part, circumvent problems of discovery of protein markers and patterns that are due to secondary and tertiary conditions of the patient, such as secondary health problems and environmental and lifestyle conditions. The discovery process involves the analysis of hundreds to thousands of tissue samples from each disease phenotype or subtype to provide molecular weight–annotated protein patterns for each specimen and then through a computational approach, a list of proteins or signature unique to the clinical question being assessed (Fig. 1). Direct analysis of a tissue section using matrix-assisted laser desorption ionization mass spectrometry (MALDI MS) technology has already been shown to be a fast and effective means to view a window of many hundreds of protein signals over a wide molecular weight range (6, 7). Many replicate analyses can be obtained from extremely small pieces of tissue because the laser spot size is typically about 50 μm in diameter. Each spot (or pixel) produced by irradiation of such a spot on tissue by the laser produces a spectrum of proteins desorbed just from that area. “Profiling” of that tissue section then may involve analysis of one or more spots from various areas of interest determined from histology. For a more complete information on the distribution of signals with the tissue, imaging of the tissue is done by analysis of an array of spots to give hundreds to thousands of pixels from a single biopsy specimen. One can display a mass spectrum for each pixel, covering proteins from molecular weight of a few thousand to >100,000. A plot of the relative intensity of any molecular weight species in each pixel over the area imaged thus produces a molecular weight–specific image of the tissue. Lasers operating at 1 kHz or faster can be used and so an analysis of 10 spots on a biopsy where 200 laser shots is acquired per spot can be accomplished in just several seconds, and data acquired from a target plate holding a hundred samples can be obtained in <10 minutes. Mass measurement accuracies are achievable below the 50 to 100 ppm (0.005-0.01%) range.

Acquiring such data-rich spectral patterns necessitates advanced computational approaches to data mining and interpretation and represents a critical part of the process of discovery of protein signatures. Although it is not the purpose of this article to review this aspect, it cannot be overstated that validations of several types are critical to ensure that data are correctly fitted and assignments made at high confidence. Through this process, molecular signatures can be discovered for a wide variety of clinically relevant questions. Hierarchical lists of protein molecular weights are produced that are the result of a given query followed by identification of the specific proteins involved using well-established MS methods. This often involves the use of electrospray ionization liquid chromatography tandem MS technology with fractionation of proteins from a tissue extract, protease digestion, peptide sequencing, and database-matching protocols (8).

Protein Signatures in Gliomas

Direct (in situ) protein analyses of human glioma biopsies were done to identify signatures that are specific to tumor development, glioma grade, and correlation with patient survival (9, 10). Patients having gliomas are characterized by short survival when diagnosed with higher-grade tumors, and these are generally nonresponsive to most anticancer therapies. Here, as in many cancers, the need for early detection and treatment is compelling. In the cited work, in situ protein analysis was accomplished using MALDI MS for the analysis of 162 biopsies from 127 patients, consisting of 26 nontumor, 35 grade 2, 28 grade 3, and 73 grade 4 biopsies. Data for statistical analysis were separated into training and testing sets, consisting of two thirds and one third of the samples per classification, respectively. From these, a total of 1,053 mass spectra were acquired and used as the data set for statistical analyses. Typically, 300 to 500 individual protein signals were detected in the range of 2,000 to 70,000 Da.

A major goal of this early work was to first assess whether or not one could find reproducible signatures for the different progressive...
states of the disease and whether or not these could be useful as prognostic indicators. Validation of these protein signatures was accomplished through correlation with traditional grading based on histopathology. We were also interested in identification of individual proteins that made up these signatures to better understand the biology of tumor progression and to identify potential drug targets and pathways.

Comparisons were done on the training set to identify a subset of differentially expressed proteins that were able to discriminate nontumor from grades 2, 3, and 4; nontumor from each individual tumor grade; and various tumor grades from each other. Two independent statistical methods were used to produce a model that best classified biopsies in the training data set. The results showed conclusively that protein patterns could readily distinguish between tumor and nontumor tissues, including mapping to individual tumor grades and comparing favorably to studies of interclass observer variability in pathology and neuropathology (11, 12). The full data set of 108 glioma patients was then evaluated, averaged by patient, to identify biomarker patterns that correlate to patient survival trends. A pattern of 24 distinct MS signals distinguished patients based on survival trends from the time of pathologic diagnosis into either a short-term survival (STS; mean survival, <15 months) and a long-term survival (LTS; mean survival, >90 months) group. Fifty-two patients were placed in the STS prognostic group, and 56 patients were laced in the LTS prognostic group, with \( P < 0.0001 \). Most importantly, the protein pattern served as an independent indicator of patient survival.

**Perspective**

Molecular weight-annotated protein signatures represent a unique data set with which to classify and correlate clinically relevant information and outcomes with changing molecular events ongoing in the progression and treatment of disease. A global approach that may use several selection and identification technologies can lead to the identification of scores of potential markers. When taken as a multivariate ensemble, the predictive value of such protein profiles may be extraordinarily high because no single marker is solely responsible for a specific indication. Serum or plasma markers of high quality can be obtained by targeted searches for those markers, or protease fragments of these, previously identified in the primary disease tissue. Such in depth proteomic studies will allow investigators to assess important clinical aspects, such as the stage of disease, the rate of progression, prognosis, and selection of appropriate therapeutic options. Indeed, this would provide the clinician with insight into
predicting the aggressiveness of an individual patient’s disease, with molecular means to follow the course of the disease and the effect of therapeutic intervention. The powerful combination of genomics and proteomics will help enable the management of patients at the molecular level and will provide an important gateway into the era of individualized medicine.

**Acknowledgments**

Received 10/4/2005; accepted 10/20/2005.

**Grant support:** NIH/National Cancer Institute grants NIGMS 5R01 GM58008 and NCI 5R33 CA86243 and T.J. Martell Foundation and Robert J. Kleberg Jr. and Helen C. Kleberg Foundation, Vanderbilt-Ingram Cancer Center.

I thank Sarah Schwartz, Robert Weil, Reid Thompson, Yu Shyr, Jason Moore, Steven Toms, and Mahlon Johnson.

**References**


Deciphering Protein Molecular Signatures in Cancer Tissues to Aid in Diagnosis, Prognosis, and Therapy

Richard M. Caprioli


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/65/23/10642

Cited articles
This article cites 10 articles, 2 of which you can access for free at:
http://cancerres.aacrjournals.org/content/65/23/10642.full#ref-list-1

Citing articles
This article has been cited by 6 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/65/23/10642.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, contact the AACR Publications Department at permissions@aacr.org.