Cancer Biomarker Discovery in Plasma Using a Tissue-targeted Proteomic Approach

Hui Zhang and Daniel W. Chan
Department of Pathology, Johns Hopkins University, Baltimore, Maryland

Cancer Biomarkers and Their Potential Uses

A cancer biomarker is a molecular signature that indicates the physiologic and pathologic changes in a particular tissue or cell type during cancer development. When such molecular signatures can be detected in the plasma (in this article, the term plasma is used to indicate plasma or serum), cancer biomarkers can have a significant effect on clinical outcomes. First, they can be used to detect cancer at an early stage, when treatment is potentially curative. Second, plasma biomarkers can be used to select patients for targeted interventions which hold the promise of improved efficacy and/or safety. Third, they can be used to monitor disease progression and response to treatment (1, 2).

Current Methods for Cancer Biomarker Discovery in Plasma and Their Limitations. Recent advances in proteomic technologies (3), intriguing initial results from analyzing plasma protein patterns using mass spectrometry (MS; ref. 4), and the clinical validation of a number of cancer markers in plasma, such as CA125 for ovarian cancer, prostate-specific antigen for prostate cancer, and carcinoembryonic antigen for colon, breast, pancreatic, and lung cancers have fueled interest in the identification of other proteins associated with cancer (5, 6). Although some studies have examined other body fluids such as cerebrospinal fluid, urine, breast nipple aspirate fluid, saliva, lung lavage fluid, and pancreatic juice, most have targeted commonly available plasma (7-9).

One method for cancer biomarker discovery is profiling plasma proteomic patterns using MS. This allows the detection of protein or peptide peaks that differ in their mass/charge ratio in patients with cancer compared with healthy individuals. However, these findings should be treated with caution. This type of study is often affected by nonbiological biases related to methods of specimen collection, processing, and storage. The other criticism of this approach is that most of these protein peaks were identified as well known, high-abundance classical plasma proteins (10). Nonetheless, identifying discriminatory peaks as peptides from high-abundance proteins indicates that at least some proteins in the plasma differ between cancer patients and normal individuals, and that these differences can be detected. Although these proteins may be indicators of interesting biology and have been shown to associate with different types of cancer (11), they are probably indicators of a systemic response to cancer or other diseases and are not likely to be derived directly from cancer tissue. Cancer-specific markers in plasma which may be useful for cancer detection will include proteins released in smaller amounts from cancer tissues (tissue-specific proteins) or as a result of structural changes in the microenvironment surrounding cancer cells, as well as indicators of an immune response to cancer cells (auto antibodies to cancer-specific proteins). For these reasons, most cancer markers currently in clinical use are low-abundance proteins with concentrations in the nanogram per milliliter range in plasma.

Another approach to cancer biomarker discovery is to separate the plasma proteome into fractions and to analyze each fraction separately using MS. This permits bypassing the most abundant proteins in order to focus on the less abundant proteins derived from specific cancer tissue/cells. Fractionation strategies based on the physicochemical properties of proteins and peptides such as size, charge, and hydrophobicity, in addition to the depletion of plasma for highly expressed proteins, has been successfully applied for biomarker discovery and has produced an extensive plasma protein catalogue. The Plasma Proteome Project of the Human Proteome Organization (12, 13), has identified thousands of proteins, including low-abundance proteins, and has established a plasma proteome database. Nonetheless, most of these methods involve extensive processing or separation and often lack the necessary reproducibility, throughput, and quantitative accuracy for near-term translation into the clinical realm.

The Challenges of Cancer-Specific Marker Discovery by Profiling Plasma Samples

It is imperative that the method used for biomarker discovery have sufficient sensitivity, reproducibility, and throughput to reliably detect low-abundance...
cancer-specific proteins in plasma. Some authors have argued that the current MS-based technologies are too insensitive to detect such proteins (14). We can estimate that, with a 100-amol detection sensitivity, which is readily achievable in a modern mass spectrometer, to detect prostate-specific antigen at a concentration of 1 ng/mL in plasma, one would need to analyze at least 3 μL of plasma sample (~200 μg of plasma proteins), sample loading that can be readily achieved in some proteomic discovery platforms (15). However, one of the major barriers is the extreme complexity of the plasma proteome. In addition, the constituent proteins span a concentration range of at least 10 orders of magnitude (16). The genetic variation among individuals in a population (17) and dynamic changes in the plasma proteome as a function of a multitude of factors, including sex, age, health status, lifestyle influences (18, 19), and physiologic and pathologic conditions other than the putative cancer of interest, could further complicate the proteomic patterns. The cumulative effect of different modifications and genetic variations produces hundreds, if not thousands, of protein species for each of the high-abundance proteins. These high-abundance proteins/peptides, together with limited separation resolution, readily mask the low-abundance proteins/peptides derived from cancer cells. Many plasma analysis methods are able to profile only the most abundant protein species.

With the completion of the human genome sequence and the application of high-throughput gene expression analysis to normal and cancer tissues or cells, specific, but protein expression is often known or predictable from gene expression data. Proteins that are expressed only in specific cancer tissues provide a set of candidates for targeted screening of plasma samples. Nonetheless, protein abundance is not always correlated with gene transcripts. Furthermore, protein modifications have been shown to associate with cancer development and cannot be predicted from genomic data. Although the use of immunoassays in the clinical setting, with known targets, provides a cost-effective approach to diagnosis, the detection of cancer markers in plasma using an immunoassay during the discovery phase, generally requires the development and production of antibodies, a time-consuming and costly process.

**Solutions to these Challenges: Detection of Cancer-Specific Proteins in Plasma Using a Cancer Tissue–Targeted Proteomic Approach**

There are a variety of approaches to reducing the complexity of the plasma proteome and proteins or their modified forms from specific cancers could be identified. These cancer-specific proteins are then detected in plasma, in an attempt to identify these proteins in plasma, identifying low-abundance tissue-specific proteins. In pursuit of these goals, glycoprotein analysis has several advantages. First, most cell surface and secreted proteins are glycosylated, and disease-associated glycoproteins (secreted by cells, shed from their surface, or otherwise released) are likely to enter into the bloodstream, and thus, represent a rich source of potential disease markers. Furthermore, the reduction in complexity achieved by focusing on the glycosylated plasma proteins and peptides translates into more favorable limits of detection, thus increasing the likelihood that the same polypeptide will be detectable in both tissue and plasma (15, 20, 21). However, the most significant benefits come from the selective analysis of N-linked glycopeptides. The number of N-linked glycosylation sites in the human proteome is modest, known in principle, and identifiable with current technology. The benefits of navigating in a mapped space as opposed to *de novo* discovery have been shown for the genome. Because proteins/peptides generated by this method are identified for each glycopeptide, MS-based methods could be developed for the specific and sensitive detection of these peptides in plasma (22, 23). The same pool of N-linked glycopeptides could be explored to identify potential biomarkers at the cell surface and in tissues, and for the targeted search for such biomarkers in plasma; this will markedly reduce the challenge of identifying biomarkers from global plasma protein profiles (21).

Currently, there are only a handful of cancer markers in clinical use; however, this does suggest that cancer-specific markers could be detected in blood using targeted approaches (24). To avoid the need for generating antibodies in the discovery and identification stages, MS-based methods have been applied. In the first approach, a specific peptide with a certain mass from the candidate protein is selected by MS (selection of precursor ion). After fragmentation of the selected peptide, a specific ion fragment of the target peptide is selectively detected (selection of fragment ion by tandem MS). This two-step analysis is referred to as a selected reaction monitoring (or multiple reaction monitoring if several precursors are monitored simultaneously; ref. 25). In a second approach, a stable isotope–labeled peptide is synthesized as a standard for quantitative analysis of each candidate protein in plasma (22). After mixing the stable isotope–labeled peptides with peptides from a clinical sample, the peptide mixture is separated with chromatography. The mass spectrometer is then used to detect and quantify the specific peptide from plasma by comparison with the heavier isotope–labeled peptide eluted simultaneously.

**Implementation and Next Steps**

To identify cancer-specific proteins in plasma, the first step is to identify which proteins are expressed in the cancer tissue. The next step is to perform a quantitative analysis comparing the specific cancer with normal tissues and other cancer tissues in order to identify cancer-specific, and even cancer subtype–specific, proteins. Using targeted MS-based approaches, potential markers can be evaluated by determining which proteins, or their modified forms from cancer tissues/cells, can be detected in plasma and quantifying their relative abundance in patients with the specific cancer, in patients with other types of cancers, and in a healthy control group. This may involve extensive separation of plasma proteins, depletion of high-abundance proteins, analysis of specific fractions of the proteome (26), or a combination of these methods. For markers that can be both detected in plasma and identified with a specific cancer, reagents and assays will need to be developed to...
enrich and detect these proteins. Finally, clinical validation in a well-defined target population is conducted for the discovered biomarkers to establish their appropriate translation into clinical use (27).

References

Cancer Biomarker Discovery in Plasma Using a Tissue-targeted Proteomic Approach

Hui Zhang and Daniel W. Chan


Updated version
Access the most recent version of this article at:
http://cebp.aacrjournals.org/content/16/10/1915

Cited articles
This article cites 26 articles, 14 of which you can access for free at:
http://cebp.aacrjournals.org/content/16/10/1915.full#ref-list-1

Citing articles
This article has been cited by 2 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/16/10/1915.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.