

Review

High-throughput genomic technology in research and clinical management of breast cancer

Plasma-based proteomics in early detection and therapy

Michael A Davis and Samir Hanash

Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA

Correspondence: Samir Hanash, shanash@fhcrc.org

Published: 18 December 2006

This article is online at <http://breast-cancer-research.com/content/8/6/217>

© 2006 BioMed Central Ltd

Breast Cancer Research 2006, **8**:217 (doi:10.1186/bcr1619)

Abstract

Protein-based breast cancer biomarkers are a promising resource for breast cancer detection at the earliest and most treatable stages of the disease. Plasma is well suited to proteomic-based methods of biomarker discovery because it is easily obtained, is routinely used in the diagnosis of many diseases, and has a rich proteome. However, due to the vast dynamic range in protein concentration and the often uncertain tissue and cellular origin of plasma proteins, proteomic analysis of plasma requires special consideration compared with tissue and cultured cells. This review briefly touches on the search for plasma-based protein biomarkers for the early detection and treatment of breast cancer.

Introduction

Early detection decreases breast cancer-related mortality [1], and breast cancer biomarkers offer a promising means to detect this disease at the earliest and most treatable stages. Both plasma and serum (collectively referred to as 'plasma' in the following discussion for simplicity) are excellent sources of clinically relevant sample material for the early detection of breast cancer. Plasma is easily obtained, is routinely used in the diagnosis of many diseases, and has a rich proteome [2]. Thus, plasma is well suited to proteomics based methods of biomarker discovery and may be a rich source of protein-based biomarkers for the early detection of cancer. Examples of such biomarkers include prostate-specific antigen and CA-125, which are used to detect prostate and ovarian cancers, respectively [3,4]. However, while there are hundreds of unvalidated candidate biomarkers for the detection and treatment of breast cancer, there are currently no validated plasma markers in clinical use for the early detection of breast cancer. Furthermore, only a handful of biomarkers are used in its diagnosis and prognosis; examples include HER-2/neu, estrogen receptor, and progesterone receptor. Therefore instead of discussing numerous unvalidated candidates, this review is intended as a brief conceptual introduction to the

proteomic search for plasma-based biomarkers that may be used in the early detection and therapy of breast cancer.

Proteomic analysis of plasma

In general, there are two approaches to proteomic biomarker discovery: target specific and global/nondirected (Table 1) [5]. Target-specific approaches often use antibodies to screen specific proteins through western blot analysis, enzyme-linked immunosorbent assays, and antibody arrays, to name a few. While these techniques are clinically applicable, they are generally low-throughput with regard to the number of proteins that can be surveyed at any one time. Thus, they may not be ideal for biomarker discovery. In contrast, global/nondirected approaches may be better suited for biomarker discovery because they are relatively unbiased, high-throughput screens. Nondirected approaches can also be divided into two groups: those studies that rely on profiling of unidentified proteins and those that generate profiles of identified proteins [6].

Protein profiling of unidentified proteins is often, although not exclusively, accomplished through matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectrometry (MS) or surface-enhanced laser desorption/ionization (SELDI)-TOF-MS [6]. In MALDI-TOF-MS, small protein fragments or peptides are crystallized within a solid matrix, which is bombarded with a laser to ionize the peptides. The ionized peptides are separated by TOF, in which smaller, less charged particles move through a flight tube faster than do larger ones with greater charge. These particles are analyzed via MS to generate a spectrum of mass/charge peaks. The peak intensities in case and control samples are then compared in order to define a pattern of peaks that can segregate case from control samples. SELDI-TOF-MS differs from MALDI-TOF-MS in that peptides are

MS = mass spectrometry; MALDI = matrix-assisted laser desorption/ionization; SELDI = surface enhanced laser desorption/ionization; TOF = time-of-flight.

Table 1

Summary of proteomic approaches used to analyze plasma for breast cancer biomarkers

Approaches		Examples	Pros	Cons
Nondirected/global	Identification based	Tandem MS	Unbiased searches Protein identities are determined Highly reproducible	Requires fractionation and/or depletion/enrichment to overcome limited dynamic range Time intensive
	Pattern based	MALDI-TOF-MS	High throughput Potentially applicable for clinical use Unbiased searches	Protein identities not determined, making validation difficult Identified proteins have been high-abundance immune and clotting factors Low reproducibility between laboratories
Target specific		Western/ELISA/ antibody arrays	Relatively high throughput Clinically applicable Identity of proteins is known	Limited antibody availability Untargeted proteins will remain undetected

ELISA, enzyme-linked immunosorbent assay; MALDI, matrix-assisted laser desorption/ionization; MS, mass spectrometry; TOF, time-of-flight.

bound to a solid support through such mechanisms as electrostatic or hydrophobic/hydrophilic forces, rather than being immobilized within a crystalline matrix. Then, nonbound peptides are washed away, thereby cleaning and simplifying the samples to be analyzed. As a result, the number of proteins identified may actually be increased (see below) as compared with MALDI.

The main advantage of MALDI and SELDI is their speed. Many samples can be processed in a short period of time, thus making them attractive tools for clinical screening. Because peptides of interest are not typically identified, validation through other techniques is difficult. However, with additional steps, peaks (proteins) of interest can be identified.

Proteins are typically identified using tandem MS, which separates and fragments peptides in multiple stages. The MS-derived fragmentation patterns are used to determine the exact amino acid sequence for a peptide of interest. Before the human genome had been sequenced, peptide sequences were determined manually based on generated mass/charge spectra; this *de novo* sequencing was extremely time consuming and error prone. In the postgenomic era, however, we can compare the spectra of observed peptides with those of peptides that are determined from the theoretical gene products of the entire genome. This process of peptide identification is much quicker and more accurate than *de novo* sequencing and makes high-throughput proteomics a reality. Although it is much faster than *de novo* sequencing, protein identification is still slower than the profiling-based proteomic studies mentioned above, which do not rely on peptide fragmentation or database searches to identify peaks of interest. However, it must be noted that protein profiles are also generated from proteins identified through tandem MS based analysis of a sample of interest. This is a laborious process but can provide much information.

MS-based proteomic methods offer an unbiased view of a sample's proteome, but they suffer a significant limitation. Specifically, even the best mass spectrometers have an analytical dynamic range of only a few orders of magnitude. Therefore, in a single analytical run it is difficult to detect proteins within the microgram/milliliter range, where many biomarkers are thought to reside. This is because of the fact that plasma contains proteins, such as albumin, that are 50,000 times more abundant than the potential biomarkers in the microgram/milliliter range [2]. Furthermore, because plasma may contain proteins with concentrations covering 10 orders of magnitude [2], a significant fraction of proteins will remain undetected. To overcome the limitations associated with the small analytical dynamic range of mass spectrometers, much work is aimed at reducing the complexity of samples before analysis. This has been accomplished most often through depletion/enrichment and fractionation/separation.

Sample depletion/enrichment involves the specific removal or purification of a subset of the proteome. For example, Agilent's Multiple Affinity Removal System (MARS) is a reusable immunodepletion column that uses antibodies to remove the six most abundant plasma proteins: albumin, IgG, IgA, transferrin, haptoglobin, and antitrypsin. These six proteins account for 90% of plasma protein content by mass. Their removal effectively decreases the concentration at which we can detect plasma proteins by two orders of magnitude. Alternatively, phosphoproteins can be enriched by capturing them with phospho-specific antibodies [7], and glycoproteins can be enriched/depleted with a lectin-based pull down system [8]. The latter was designed to reduce plasma complexity because many plasma constituents are glycoproteins. Of possible concern, sample depletion may remove potential biomarkers. For instance, albumin is a known carrier of plasma proteins and may bind potential

biomarkers. However, when combined with fractionation, the number of additional proteins identified after depletion of albumin far outweighs the number of proteins that are potentially lost. Furthermore, the pool of depleted proteins can be analyzed for the presence of candidate biomarkers that have bound the captured proteins. Therefore, little to no data should be lost through depletion.

Sample fractionation/separation for proteomic analysis has most often been accomplished through two-dimensional gel electrophoresis, in which samples are separated electrophoretically based on isoelectric point and size. Samples are visualized as spots using protein stains, autoradiography, or fluorescent tags. There are many variations of this technique, but traditionally gels/membranes from cases are compared with those from controls to identify differences between the two. Proteins of interest are excised, digested with a protease such as trypsin, and then analyzed by MS. Of note, two-dimensional gel electrophoresis can separate the individual species of a protein, such as isoforms, fragments, and modifications. However, other techniques may be necessary to determine which particular species have been identified. One drawback of two-dimensional gel electrophoresis is its low sensitivity, which is limited by the ability to visualize a protein on the gel/membrane. Thus, two-dimensional gel electrophoresis suffers from a small analytical dynamic range. In addition, it also suffers from the limited range of protein species that can be resolved in a single run due to physicochemical properties such as size, charge, or post-translational modifications.

Due to the limitations of two-dimensional gel electrophoresis, a growing number of researchers are using liquid chromatography to fractionate proteins before MS analysis. In this method, proteins are first digested into peptides. The peptides are separate based on net positive or net negative charge using a strong cation or anion exchange column, respectively. Alternatively, peptides are separated by hydrophobicity using a reversed-phase column. The beauty of this system is the ability to combine these two separation methods for multidimensional peptide separation. For example, fractions collected from the cation/anion exchange column are further separated by reversed-phase chromatography. Each fraction is then analyzed by MS. Because each fraction is significantly less complex than whole plasma, many more proteins and more low-abundance proteins are identified. However, this method also significantly increases the time needed for MS analysis. Abundance differences in protein isoforms, degradation fragments, or post-translational modifications may serve as biomarkers for early detection of breast cancer, but it is impossible to discern one protein species from another when proteins are digested before separation. Thus, some groups, including our own, have opted to separate intact proteins before digestion [9,10]. In this manner, the information regarding changes in individual protein species is conserved, and we can routinely identify

well over 1000 proteins and their protein subspecies (isoforms, cleavage fragments, and so on).

Tumor and tissue specificity

One goal of biomarker research is to identify markers that are tumor specific. Unfortunately, plasma-based proteomics suffers from the fact that proteins within plasma originate from numerous tissues. For example, when studies report an upregulation of interleukin-6 in the serum of breast cancer patients compared with control individuals, it is difficult to know whether this protein is released directly from the tumor or whether interleukin-6 upregulation is a systemic reaction to the tumor and released by nontumor tissues [11].

Therefore, when searching for tumor-specific biomarkers, it is necessary to consider how tumor proteins appear in the blood. Possibilities include the following: increased expression of secreted and membrane-bound proteins; loss of polarity, resulting in apical secretion of basal proteins and *vice versa*; increased cleavage of matrix and membrane-bound proteins caused by increased protease expression and/or activation; and release of cytoplasmic proteins from cells that have died. The latter possibility may be unlikely, given the findings of a recent report presenting a proteomic analysis of tissue interstitial fluid from breast tumors [12]. That study identified few of the nuclear proteins that are so often identified in proteomic analyses of cell lines and whole tissues. Based on these observations, the authors hypothesized that cell lysis does not significantly contribute to the proteomic content of tumor tissue interstitial fluid. As an extension, it may be likely that cell lysis also contributes little to the blood proteome of cancer patients. Therefore, secreted proteins, matrix proteins, and cleaved membrane-bound proteins may be the most probable source of tumor-specific biomarkers. Additionally, as stated above, tumor-specific plasma markers may result from altered localization of proteins and protein fragments rather than from increased expression. As a result, proteomic studies may not correlate well with RNA expression studies.

Crucial to early detection, the biomarker field is seeking markers that are tissue specific in addition to being tumor specific. If we can detect the cancer but not the tissue of origin, then we may actually do more harm than good, since searching for a suspected tumor will add undue stress to the patient and increased cost to the treatment. Finding tissue-specific tumor markers has thus far proven difficult. Many candidate biomarkers have been concurrently identified in numerous tumor types. This likely reflects that fact that 90% of all cancers are of epithelial origin and thus express many of the same proteins [13]. It is probable that a panel of markers will be required to establish tissue specificity rather than a single protein; this panel may or may not be independent of a tumor-specific panel of biomarkers. In addition, early detection markers may need to be used in conjunction with other screening methods, such as mammography, where the tissue origin is not in question.

Sample-dependent considerations

Although plasma is the specimen of choice for early diagnosis, proteomic-based biomarker studies also rely on cells grown in culture and tissue samples. Cells grown *in vitro* provide ample material for analysis and are easily manipulated with therapeutic agents. However, cells grown *in vitro* behave quite differently from those *in vivo* [14], and because cell lines are derived from a single person, it can be difficult to generalize results to the population as a whole. Tissue is also frequently used and is generally collected via biopsy or mastectomy. Often, tissue sections with more than 50% or more than 70% tumor tissue are compared with normal tissue sections. However, care must be taken because most breast cancers are of epithelial origin, and normal tissue sections with more than 50% epithelium may be difficult to find. Thus, differences observed in many studies may simply reflect differences in epithelial:stromal ratios rather than differences between cancer and normal tissues. Finally, tissue is heterogeneous, containing epithelium, fibroblasts, fat cells, endothelium, immune cells, and so on. These differences can be difficult to control. To circumvent these problems, many groups use laser capture microdissection [15] to specifically capture equal numbers of cancer and normal cells for analysis. Alternatively, imaging MS may allow differentiation of stroma from epithelium within breast tissue, although the 50 μm resolution of imaging MS may be too low to distinguish epithelial ducts from the milieu of stroma that comprises much of the mammary tissue [16,17].

When blood is used as sample material, the cells are removed before analysis. This is accomplished with the addition of anticoagulants such as EDTA followed by centrifugation (plasma) or by allowing the blood to clot followed by removal of the supernatant (serum). Some groups prefer plasma over serum because the clotting process in serum preparation may introduce inconsistencies and significant sample-to-sample variation [2].

Conclusion

Disease-related proteomics is fueled by the hope that we might literally save hundreds of thousands of lives per year with markers of early detection and with markers that allow optimized treatment for each individual. There are hundreds of identified candidate biomarkers, but these must be validated to prove their specificity and clinical relevance. Thus, for breast cancer, we do not, as yet, have those golden

markers so actively sought. However, proteomics has come far in the past decade, and numerous candidates are now progressing through validation studies.

Competing interests

The authors declare that they have no competing interests.

References

1. American Cancer Society: **After the Tests: Staging** [http://www.cancer.org/docroot/CRI/content/CRI_2_2_3X_After_the_tests_Staging_5.asp?sitearea=]
2. Anderson NL, Anderson NG: **The human plasma proteome: history, character, and diagnostic prospects.** *Mol Cell Proteomics* 2002, **1**:845-867.
3. Barry MJ: **Clinical practice. Prostate-specific-antigen testing for early diagnosis of prostate cancer.** *N Engl J Med* 2001, **344**:1373-1377.
4. Rustin GJ, Bast RC Jr, Kelloff GJ, Barrett JC, Carter SK, Nisen PD, Sigman CC, Parkinson DR, Ruddon RW: **Use of CA-125 in clinical trial evaluation of new therapeutic drugs for ovarian cancer.** *Clin Cancer Res* 2004, **10**:3919-3926.
5. Srivastava S, Srivastava RG: **Proteomics in the forefront of cancer biomarker discovery.** *J Proteome Res* 2005, **4**:1098-1103.
6. Gillette MA, Mani DR, Carr SA: **Place of pattern in proteomic biomarker discovery.** *J Proteome Res* 2005, **4**:1143-1154.
7. Mann M, Jensen ON: **Proteomic analysis of post-translational modifications.** *Nat Biotechnol* 2003, **21**:255-261.
8. 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.
9. Misek DE, Kuick R, Wang H, Galchev V, Deng B, Zhao R, Tra J, Pisano MR, Amunugama R, Allen D, 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.
10. Wang H, Clouthier SG, Galchev V, Misek DE, Duffner U, Min CK, Zhao R, Tra J, Omenn GS, Ferrara JL, Hanash SM: **Intact-protein-based high-resolution three-dimensional quantitative analysis system for proteome profiling of biological fluids.** *Mol Cell Proteomics* 2005, **4**:618-625.
11. Celis JE, Moreira JM, Gromova I, Cabezon T, Ralfkiaer U, Guldborg P, Straten PT, Mouridsen H, Friis E, Holm D, et al.: **Towards discovery-driven translational research in breast cancer.** *FEBS J* 2005, **272**:2-15.
12. Celis JE, Gromov P, Cabezon T, Moreira JM, Ambartsumian N, Sandelin K, Rank F, Gromova I: **Proteomic characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery.** *Mol Cell Proteomics* 2004, **3**:327-344.
13. Birchmeier C, Birchmeier W, Brand-Saberi B: **Epithelial-mesenchymal transitions in cancer progression.** *Acta Anat (Basel)* 1996, **156**:217-226.
14. Birgersdotter A, Sandberg R, Ernberg I: **Gene expression perturbation in vitro—a growing case for three-dimensional (3D) culture systems.** *Semin Cancer Biol* 2005, **15**:405-412.
15. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA, Liotta LA: **Laser capture microdissection.** *Science* 1996, **274**:998-1001.
16. Caldwell RL, Caprioli RM: **Tissue profiling by mass spectrometry: a review of methodology and applications.** *Mol Cell Proteomics* 2005, **4**:394-401.
17. Chaurand P, Schwartz SA, Reyzer ML, Caprioli RM: **Imaging mass spectrometry: principles and potentials.** *Toxicol Pathol* 2005, **33**:92-101.

This article is part of a review series on *High-throughput genomic technology in research and clinical management of breast cancer*, edited by Yudi Pawitan and Per Hall.

Other articles in the series can be found online at http://breast-cancer-research.com/articles/review-series.asp?series=bcr_Genomic