Proteomics and Mass Spectrometry for Cancer Biomarker Discovery

Ming Lu[1,2], Kym F. Faull[3], Julian P. Whitelegge[3], Jianbo He[1,2], Dejun Shen[1,2], Romaine E. Saxton[4] and Helena R. Chang[1,2,4]

[1]Gonda/UCLA Breast Cancer Research Laboratory; [2]Revlon/UCLA Breast Center, Department of Surgery/Oncology, David Geffen School of Medicine, Los Angeles, California; [3]The Pasarow Mass Spectrometry Laboratory, Department of Psychiatry & Biobehavioral and the Neuropsychiatric Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles; [4]Division of Surgical Oncology, Department of Surgery, David Geffen School of Medicine, Los Angeles, California.

Abstract: Proteomics is a rapidly advancing field not only in the field of biology but also in translational cancer research. In recent years, mass spectrometry and associated technologies have been explored to identify proteins or a set of proteins specific to a given disease, for the purpose of disease detection and diagnosis. Such biomarkers are being investigated in samples including cells, tissues, serum/plasma, and other types of body fluids. When sufficiently refined, proteomic technologies may pave the way for early detection of cancer or individualized therapy for cancer. Mass spectrometry approaches coupled with bioinformatic tools are being developed for biomarker discovery and validation. Understanding basic concepts and application of such technology by investigators in the field may accelerate the clinical application of protein biomarkers in disease management.

Keywords: Proteomics, Mass spectrometry, Cancer, Biomarkers.

Abbreviations: 2DE: two-dimensional gel electrophoresis; ABPP: activity-based protein profiling; CEA: carcinoembryonic antigen; CI: confidence interval; ESI: electrospray ionization; FP: fluorophosphonate; HPLC: high performance liquid chromatography; ICAT: isotope coded affinity tags; IEF: isoelectric focusing; iTRAQ: isobaric tags for relative and absolute quantification; LCMS: combined liquid chromatography-mass spectrometry; LCMSMS: liquid chromatography tandem mass spectrometry; LOD: limit of detection; m/z: mass to charge ratio; MALDI: matrix-assisted laser desorption ionization; MS: mass spectrometry; MUDPIT: multidimensional protein identification technology; NAF: nipple aspirate fluid; PMF: peptide mass fingerprinting; PSA: prostate specific antigen; PTMs: post-translational modifications; RPMA: reverse phase protein microarray; SELDI: surface enhanced laser desorption ionization; TOF: time-of-flight.

Introduction

Cancer is a prevalent and serious public health problem not only in the United States and also worldwide. It is a major source of morbidity, and the second leading cause of death in the American population. A total of 1,444,920 new cancer cases and 559,650 deaths are projected in the United States in 2007 (Jemal et al. 2007). When deaths are aggregated by age, cancer has surpassed heart disease as the leading cause of death in individuals 85 years age or younger since 1999 (Jemal et al. 2007).

Before cancer prevention becomes a reality, early detection and better treatment remain the hope for all cancer patients. It is easy to understand that the test for early detection of cancer is best done in blood or other types of body fluids; however, cancer classification and management are frequently accomplished by using tissues, as it truly represent the source for a much more reliable discrimination between different types and subtypes of cancer. Protein biomarkers in blood or tissues may therefore play an important role in cancer detection, monitoring and treatment.

Proteomics is a rapidly advancing field brought about in large part because of the recent developments in mass spectrometry and associated technologies. These developments allow for the fast and reliable detection, identification and relative quantitation of proteins. The measurements can be made from the solid and solution states, and sometimes from complex mixtures. The amount of protein required for these measurements has been continuously reducing, and today many laboratories have demonstrated

Correspondence: Helena R. Chang, M.D., Ph.D., Revlon/UCLA Breast Center, 200 UCLA Medical Plaza, Suite B265, David Geffen School of Medicine, UCLA, Los Angeles, CA 90095. Tel: (310) 794–5640; Fax: (310) 206-2982; Email: hchang@mednet.ucla.edu

the ability to obtain results from samples with protein concentration as low as 5 femtomoles (about 3×10^{-10} grams of albumin, for example). Emerging from this work is a new goal in clinical research to use these techniques to discover and identify proteins that are specific to, and diagnostic for, a given disease. If successful, these developments will ultimately allow for early detection and possibly individualized therapy for the disease. The object of this review is to describe the basic concepts, tools, limitations and progress in this rapidly advancing field. The obvious potential of protein biomarker screening for cancer diagnosis has yet to be fully realized. The clinical application of proteomics to cancer management will not be successful until well controlled investigations with standardized methods of proven accuracy for profiling nanogram level proteins in complex mixture of biological specimens.

Why protein biomakers?

There are many reasons why proteins are important or even to be better biomarker than other molecules such as DNA or RNA. Firstly, proteins are much more diverse than nucleic acids and therefore carry more information. Through alternative splicing and over 100 unique post-translational modifications (PTMs) including covalent modification by attachment of specific functionalities (acetylation, phosphorylation, etc), as well as post-translational trimming and cleavage of proteins by site-specific proteases, tens and possibly hundreds of distinct protein species can result from expression of a single gene. Moreover, many cellular functions are not revealed at the level of the nucleic acids, but are manifested by the PTMs, as well as cleavage of proteins by site-specific proteases. With a limited number of human genes (~20,000) and many more proteins (10^5–10^6) (Stein 2004; Pennisi 2003), the latter are likely to more closely reflect the complex nature of cellullar biology and host physiology. Secondly, fluctuating levels of an mRNA transcript make the correlation with the amount of protein synthesized inconsistent in many cases. For example, when the same cells or tumors were examined by both cDNA arrays and proteomic analysis, the concordance rates were low between the mRNA transcript profiles and corresponding protein expressional patterns (Anderson and Seilhamer, 1997; Alaiya et al. 2000a; Chen et al. 2003; Nishizuka et al. 2003).

Therefore studies limited to the DNA or mRNA analysis may reflect only a partial assessment of cellular function. Documenting proteins and their interactions in cancer cells represents a complementary and much more comprehensive approach to understanding the cellular changes caused by disease processes.

Definitions

A "*biomarker*" is defined as a measurable analyte that correlates with a specific phenotype, such as a normal biological condition, a pathological process, or a pharmacological response to a therapeutic intervention (Group, 2001). A biomarker is unique in that it distinguishes and discriminates between comparative biological conditions such as cancer and non-cancer. "*Cancer proteomics*" is defined as molecular profiling of cancer-associated proteins approached by analyzing the global protein expression patterns of tumor cells or extracellular fluids such as serum from cancer patients (Alaiya et al. 2000b). To date there have been many candidate proteomic biomarkers suggested for various types of cancer. However, none has yet met the stringent requirements for clinical use, such as reproducibility, specificity and sensitivity. A true biomarker should be confirmable across laboratories and technology platforms (Aebersold et al. 2005), in its ability of predicting the clinical state. Today the greatest need in cancer proteomics research lies in improvement in the methodology and technology for the recognition and identification of reliable protein biomarkers specific for the disease or biology of the disease.

Historically, biomarker discovery was dominated by targeted approaches, in which candidates derived from biological knowledge were selectively evaluated for their correlations with clinical conditions (Gillette et al. 2005). However, in the past two decades, the advances in new mass spectrometric ionization techniques for discovering macromolecules has revolutionized biomarker research (Karas and Hillenkamp, 1988). Now for the first time it is possible to display complex protein profiles in convenient one- and two-dimensional formats, and at the same time identify the proteins. More and more studies are employing mass spectrometry technology to search for novel cancer biomarkers. Reflecting this trend, in March 2007, a MedLine search for the words "cancer biomarker" listed in 119,912 publications, while

a similar simultaneous search that included the word "protein" resulted in 96,598 articles. A selection of published representative studies on cancer biomarkers using mass spectrometry technique are summarized in Table 1. However, as reported by Alaiya et al. (2005), the lack of proteomics markers for cancer stands in sharp contrast to the significant progress that has been made in the discovery of gene-based biomarkers during the last 20 years. This disparity also reflects the increased difficulty when working with the proteome as opposed to the genome. Nevertheless, the potential offered by proteomic technology in biomarker discovery draws research support from both government funding agencies and private sector.

A protein biomarker detected by mass spectrometry can be a single entity (with protein identification, therefore identity-based) or a suite of entities (protein signatures, therefore pattern-based). For the biomarker to be clinically useful, the abundance or change in abundance of specific proteins must reflect some aspects of normal physiology/biochemistry or a disease process. Combined liquid chromatography-mass spectrometry (LCMS) is the technique to which most cancer proteomic research is now gravitating. Because of the direct on-line link between chromatography and mass spectrometry, this technique can more easily accommodate complex biological samples such as serum and plasma. In contrast, the enthusiasm which originally greeted other techniques that do not use direct on-line chromatography, such as surface enhanced laser desorption mass spectrometry (SELDI), has now waned. This has occurred as the need for chromatography when dealing with complex samples, such as serum and plasma, has become obvious. In this context then, as defined here, a biomarker may be a *peak or peaks* detected on a chromatogram, a *signal or signals* in a mass spectrum, or a *feature or features* in a three dimensional representation of peak retention times, signal mass to charge (m/z) ratios and intensities. These are three different representations of LCMS data that are routinely acquired in proteomic analyses.

It has been pointed out that the utility of a protein biomarker for disease diagnosis or detection does not necessarily require knowledge of the identity of the protein or proteins involved (Petricoin et al. 2002; Villanueva et al. 2005), so long as the pattern of peaks, signals or features are sufficiently reproducible to reflect the present disease. Strictly speaking, this could be the case.

However, without knowing the identity of the biomarker proteins, information related to the underlying disease process will not be forthcoming. Furthermore, the identity of the protein or proteins involved is necessary for independent validation of the biomarker by different technologies, and for the future development of faster, cheaper and more reliable assays (which probably will not rely on chromatography or mass spectrometry). Consequently, disease-based protein biomarker research involves both recognition of the biomarker signature and identification of the proteins involved.

A dilemma—low- and high-abundance proteins

The components of complex samples (proteomes) are typically present in a wide concentration range. The components of human plasma and serum for example extend across 10–12 orders of magnitude in concentration (Anderson and Anderson 2002). This presents an analytical dilemma because no one technique has a dynamic range that can accommodate such samples. In the case of cancer biomarker screening, while the results of some investigations (Li et al. 2005, Yu et al. 2004) suggest that high abundance proteins such as Complement C3a (desArg) (a 77-amino acid protein) and a C-terminal-truncated form of C3a may be useful for diagnosing breast cancer, and β-globin for colorectal cancer, the low specificity of these tests limits their clinical utility. At the same time, the most clinically useful cancer biomarkers are low-abundance molecules; like prostate specific antigen (PSA) (Wang et al. 1979; Stamey et al. 1987), members of the mucin family of proteins (for example, CA 15.3), carcinoembryonic antigen (CEA) and cytokeratins (i.e. TPA, TPS and Cyfra 21.1; Seregni 2004) being examples. Interestingly, all these established cancer biomarkers are glycoproteins. In addition, the alteration in protein glycosylation which occurs through varying the heterogeneity of glycosylation sites or changing glycan structure of proteins on the cell surface and in body fluids have been shown to correlate with the development of cancer and other disease states (Durand and Seta, 2000). Therefore, the technology for screening protein biomarkers in body fluids must confront the large dynamic range of the concentration of the molecules of interest. Thus to accommodate both low- and high-abundance molecules, there is a need for fractionation of

Table 1. Summary of selected MS-based cancer biomarker studies.

Cancer type	Specimen	Diagnostic sensitivity and specificity or protein ID	MS technology	Reference
Breast	Serum	93%; 91% Complement component $C_{3adesArg}$ and its truncated form	SELDI-TOF, Tandem mass peptide sequencing	Li et al. 2002, 2005
Breast	Ductal lavage	75%	SELDI-TOF	Mendrinos et al. 2005
Breast	Nipple aspirate fluid	Hb-β chain	SELDI-TOF	Sauter et al. 2005
Breast	Serum	High molecular weight kininogen, apolipoprotein A-II.	SELDI-TOF	Heike et al. 2005
Breast	Serum	94%; 81.3%	2DE, LCMSMS	Ru et al. 2006
Breast	Nipple aspirate fluid	Vitamin D binding protein et al..	ICAT MSMS	Pawlik et al. 2006
Breast	Cell line	14-3-3 sigma	MALDI	Liu et al. 2006
Breast	Tissue	Thymosin alpha-1	HPLC and MALDI	Traub et al. 2006
Breast	Nipple aspirate fluid	8 markers with m/z – 5061, 5994, 6001, 10207, 13070, 13436, 13447, 57075	SELDI-TOF MS	He et al. 2007
Breast	Tissue	>1000 protein identified	NanoLC MS	Umar et al. 2007
Ovarian	Serum	100%; 95%	SELDI-TOF	Petricoin et al. 2002
Ovarian	Serum	73–96%; 83–95%	SELDI-TOF	Kozak et al. 2003
Ovarian	Serum	Haptoglobin-alpha subunit	SELDI-TOF	Ye et al. 2003
Ovarian	Serum	Apolipoprotein A1, truncated form of transthyretin, fragment of inter-alpha trypsin inhibitor heavy chain H4. 74%; 97%	SELDI-TOF	Zhang Z et al. 2004
Ovarian	Serum	Transthyretin and its fragment, beta-hemoglobin, apoAI, and transferrin	Micro-LCMSMS	Kozak et al. 2005
Ovarian	Serum	Hemoglobin-alpha and -beta	SELDI-TOF	Woong-Shick et al. 2005
Ovarian	Plasma	94%	SELDI-TOF	Rai et al. 2002
Ovarian	Plasma	Amyloid A1 and its truncated form	SELDI-TOF	Moshkovskii et al. 2005
Ovarian	Plasma	90–96.3%; 100%	SELDI-TOF	Lin et al. 2006
Ovarian	Plasma	fibrinopeptide-A	2DE, MS	Ogata et al. 2006
Endometrial	Tissue	95%, chaperonin 10, a-1-antitrypsin	LCMSMS, Q-STAR	Desouza et al. 2007
Colorectal	Serum	95.2%, 90.0%	MALDI-TOF	De Noo et al. 2006
Colorectal	Cell line	Prothymosin α	SELDI-TOF	Shiwa et al. 2003
Colorectal	Serum	N-terminal fragment of albumin, apoC-I, apoA-I	SELDI-TOF	Engwegen et al. 2006
Colorectal	Serum	β-defensins	SELDI-TOF	Melle et al. 2005
Colorectal	Tissue, serum	β-defensins	SELDI-TOF	Albrethsen et al. 2005
Colorectal	Tissue	Hydrophobic proteins (vimentin)	2DE, MALDI-TOF /TOF, ESI–MSMS	Alvarez-Chaver et al. 2006
Prostate	Serum	Zn-alpha2 glycoprotein (ZAG)	LCMSMS	Bondar et al. 2007

complex samples. While undesirable because it increases the workload and introduces additional error in the measurements, this seems unavoidable. One current research focus is how best to fractionate complex proteomes, and many different strategies are now being compared.

Application of proteomic technologies to cancer

Early detection of cancer

It is well known that early detection of cancer can lower the mortality rate. Cancers diagnosed at an early stage are more likely to be cured by conventional therapy. For many cancers, 5 and even 10-year survival may approach or even exceed 90% if treated early, and only 10% or less when found late (Etzioni et al. 2003). The Pap smear test (named after the inventor Georgios Papanicolaou) for early detection of pre-neoplastic cervical cancer (Shingleton et al. 1995), and colonoscopy for early detection of colon cancer (Winawer et al. 1995; Winawer, 2001), are good examples of how early disease detection can significantly improve cancer outcome. Not surprisingly, the focus therefore in biomarker screening is to identify proteins that can aid in early detection and diagnosis of the disease.

Similar to other cancers, the early detection of breast cancer by serum biomarkers currently remains a research quest. The circulating breast cancer markers now used in clinical practice (CA 15.3, CEA and cytokeratins) are useless in early detection since they are all associated with a large tumor burden at late stage of the disease (Seregni et al. 2004). It is estimated that a minimum number of 10^9 tumor cells are required for breast cancer detection by the current methods such as mammography or clinical breast examination (Dummin et al. 2006). Therefore the challenge is to develop techniques that can detect specific disease-related changes from cells fewer than 10^9 cells or a total mass equivalent to milligrams of tissues. Fortunately, the field of cancer biomarker screening is evolving rapidly, and recent developments in the technological, biological and statistical realms show promise for uncovering markers that meet this goal.

In 2002, Petricoin et al. made a sensational report of mass spectrometry/proteomics-based methods for the detection of ovarian cancer. In this work SELDI mass spectrometry was used to profile serum proteins, the pattern of which could be used to separate cancer from non-cancer patients. One claim was that the technology could be used for early detection of ovarian cancer. In the study, a preliminary "training" data set, consisting of spectra derived from the analysis of serum from 50 unaffected women, and 50 patients with ovarian cancer, was analyzed by an interactive pattern-recognition algorithm. A pattern was identified that discriminated cancer from non-cancer. The pattern was then used to classify an independent set of spectra obtained from 116 masked serum samples collected from 50 women with ovarian cancer, and 66 unaffected women. The pattern identified in the training set successfully segregated cancer from non-cancer in the second set of samples. The validation set correctly identified all 50 ovarian cancer cases, including all 18 stage I cases. Of 66 cases of non-malignant disease, 63 were classified correctly, giving a sensitivity of 100% (95% confidence interval (CI) 93–100%), a specificity of 95% (CI of 87–99%), and a high positive predictive value of 94% (CI of 84–99%).

Since publication of this report, there have been several attempts to repeat, confirm and extend these findings to other types of cancer. These newer investigations have used similar proteomic profiling approaches with body fluids from patients with various types of cancer (Adam et al. 2002; Li et al. 2002; Borozdenkova et al. 2004; Ebert et al. 2004, Li et al. 2005; Mendrinos et al. 2005; Villanueva et al. 2005). Unique biomarker patterns have been reported for early detection of ovary, prostate, breast, and thyroid cancer. The enthusiasm led to the postulation that a new era in cancer diagnostics had come, in which serum proteomic profiling would meet the goal of early cancer detection.

However, several methodological and bioinformatic artifacts and biases have been identified, which challenged the validity of the published results (Baggerly et al. 2005; Diamandis and van der Merwe, 2005). While it appears that all have high diagnostic sensitivities and specificities, few of the reported biomarkers have adequate reproducibility (Diamandis, 2002). In retrospect it appears that the limitations of these initial efforts to identify reliable diagnostic cancer biomarkers resulted from differences in methodology among the investigators.

Like almost every newly emerged technology, through these early efforts including errors and

fails in searching disease associated biomarkers, much has been learned to pave the way for true biomarker discovery. The expectation to the new technique frequently exceeds the capacity at the moment. On the other hand, the unfulfilled expectations or the aware limitations of the technique are the forces to drive the research forward.

In the foreseeable future, the field will be continuously challenged by the heterogeneous nature of the disease, the specimen and the technology. It is quite often the case that the heterogenesis of patients, such as gender, age, genetics, ethnicity, body mass, medication, the presence of other conditions and diseases, emotional stress, menopausal status, and dietary/nutritional preferences, as well as sample preparation, may all introduce artifacts. Also, given the small quantity of the human specimens and the limited availability of quality samples, and the extraordinary cost of proteomics technologies, many studies are based on small numbers of samples or even pooled samples. Thus, with small sample numbers, it is reasonably expected that the overlap of findings among different groups would be low. As progress being made rapidly, the findings would eventually grow together. To make the matter even more complex, the mass spectrometry platforms and methodology used drastically could be different among research groups. Even if the instruments are same, they could be operated in very different modes for which those details are not disclosed in publications. Thus, the MS methodologies used are as heterogenous as cancer disease itself. Taking together, it becomes clear that consistent findings amongst laboratories can only accomplished after.

Another fundamental problem, as already mentioned, lies in the enormous number and 10^9 concentration range of proteins in tissues and body fluids and the complexity of these fluids. Compared with MS techniques, current cancer markers identified by immunological techniques are at subnanogram levels, and have no cross-reactivity in the presence of a huge excess of other unrelated proteins. Immunological techniques avoid the need for removal of high abundance proteins before nanoscale detection of the target analyte. However, compared with immunological techniques, the unique promise of proteomics results from its potential to simultaneously resolve and compare thousands of proteins for qualitative and quantitative differences prior to identification

by mass spectrometry. This technology will become a powerful diagnostic tool once methods are developed to first separate high, medium, and low abundance proteins before proteomics analysis of each of these three subsets using the more than 1,000-fold dynamic range capability of mass spectrometry.

Proteomics in cancer prognostics and monitoring response to therapy

Increased serum levels of proteins like CEA and PSA are used to detect re-growth of some common malignancies after conventional therapy including surgery, radiation, and chemotherapy. Progression of nearly all tumors results from induction of inflammatory cytokines, growth factors, angiogenic peptides like VEGF, and release of high abundance coagulation proteins in the blood that stimulate cancer cell proliferation and metastasis. Many of these cancer-promoting peptides are transported by plasma lipoproteins that have been identified recently by a combination of proteomic approaches including 1-DE and 2-DE MALDI-TOF, isotope-coded affinity tag and Western blot analysis (Rezaee et al. 2006). This provides an unprecedented opportunity to apply current proteomics technology to cancer prognosis by monitoring serum and plasma protein levels after primary or adjuvant therapy.

In addition to the effort investigated in early detection, chemotherapy has been widely used in the treatment of various cancers for reducing cancer mortality. However, unpredictable effectiveness and treatment toxicity continue to plague the chemotherapy. The novel treatments that are focused on specific targets in the signal transduction and/or metabolic pathways are extremely effective, and have less treatment associated toxicity. Herceptin® (trastuzumab), a humanized monoclonal antibody which targets tumors over-expressing the Her-2/neu receptor protein (Baselga et al. 2004; Ross et al. 2004ab), and Gleevec® (imatinib mesylate), targeting cancers with bcr-abl gene translocation in chronic myeloid leukemia (Druker, 2004), are two of the best examples. The success of Herceptin and Gleevec has stimulated the development of pathway specific treatments for other types of cancer. Protein biomarker screening offers potential for subtyping cancers according to their unique protein profiles. Such signatures could be used to guide tailored

treatment, avoid unnecessary toxicity, and reduce cost because of the selective nature of the treatment. Signatures with biological significance may not only improve cancer sub-classification but also lead to the development of novel treatments (Hunt and Keyomarsi, 2005).

Clinical proteomic approaches and platforms

Specimens for protein profiling of cancer

Many types of biological specimens have been used in cancer proteomic research, such as cell lines, tumor tissue, serum, plasma, urine, salvia and nipple aspirate fluid (NAF) for breast cancer (He et al. 2007; Shau et al. 2003). Tumor tissue may be an ideal source to study cancer proteomic signatures. Unfortunately tumor tissue can be difficult to obtain and to preserve for proteomic analysis. Serum or plasma on the other hand has been most commonly used for biomarker research because sufficient amounts of these samples are easily attainable by relatively non-invasive means, and frequently contain tumor markers, albeit maybe in low abundance. It is known that not only do tumors leak or secrete proteins into circulation, but also the surrounding stroma releases proteases and other mediators of tissue injury in response to the nearby tumor growth. The currently used tumor markers such as PSA, CA125, CEA, and alpha-fetoprotein (AFP), are all examples of useful low abundance circulating cancer biomarkers that probably arise from the tumor and/or the surrounding stroma (Bast et al. 2001; Dhanasekaran et al. 2001).

However, serum/plasma profiling is not without its challenges. As already alluded to, one major obstacle is the wide concentration range of the candidate markers that exceeds the dynamic range of any single analytical method or instrument (Domon and Aebersold, 2006). For example, it has been estimated that the concentration range of various serum proteins exceeds 10 orders of magnitude (Anderson and Anderson, 2002). Although this represents a daunting hurdle, recent technological developments have increased the dynamic range of analytical instrumentation, and new methodological developments have reduced the threshold for protein detection to sub-nanogram levels (Domon and Aebersold, 2006).

Fractionation of plasma and serum proteome

It is generally accepted that any method for detection of disease biomarkers in complex proteomes such as plasma and serum will require fractionation to reduce the complexity of the sample. This is considered essential even after the major abundant proteins, such as albumin, hemoglobin, immunoglobulin, transferrin, complement, haptoglobin and others, have been depleted from the sample (Echan et al. 2005). A reflection of this trend can be seen in a number of methods already reported for accomplishing this purpose. Numerous strategies have been developed including ion exchange based separations using the combination of strong cation exchange chromatography with reversed phase separations such as the multidimensional protein identification technology (MUDPIT) developed by Yates and co-workers (Washburn et al. 2001), molecular weight based filtration (Hu et al. 2005), hydrophobicity based separation such as the use of reverse phase separations (Wang and Hanash, 2005), the capture of cysteine-containing peptides with biotinylated thiol reagents (Schrimpf et al. 2005), immobilizing of phosphorylated peptides by metal affinity chromatography (Corthals et al. 2005), dendrimer conjugation chemistry (Tao et al. 2005), glycopeptide capture (Zhang H. et al. 2005), activity-based protein profiling (ABPP) (Speers and Cravatt, 2004), fluorophosphonate (FP)-based ABPP (Jessani et al. 2005) targeting serine hydrolase, monolithic columns and Beckman's PF2D fractionation instrument (Ratnayake et al. 2000). Also, there has been extensive work on microseparations for froteomic studies by Cheng Lee's team (Li et al. 2003) and Smith's group (Smith, 2006). Finally, Wall et al. (2000) developed a unique two-dimensional all liquid-phase method combined with MS for protein profile analysis. With this method, proteins are fractionated by pI using isoelectric focusing (IEF) in the Rotofor cell and then separated by hydrophobicity using reverse phase-HPLC in the second dimension. All these separation or depletion procedures may help to remove interfering proteins and allow detection of biomarkers with lower concentrations for the analysis of cancer proteome.

Methodology and proteomic platforms

The basic principles of proteomics methods currently used in the application of cancer research including two-dimensional gel electrophoresis

(2DE) and mass spectrometry (MS). Although 2DE is still a currently used tool for proteomic analysis, it has disadvantages regarding to its throughput, reproducibility, sensitivity and limited range of MW (<200 kDa) and p*I* (between 4 and 10), which limit the use of this method (Jenkins and Pennington, 2001). This review will focus on proteomic techniques based on mass spectrometry for cancer biomarker research.

Since the discovery of new mass spectrometric ionization techniques for macromolecules almost 20 years ago (Karas and Hillenkamp, 1988), many MS instruments have been developed and used in cancer biomarker research, for example, MALDI-TOF MS, SELDI-TOF MS, LCMSMS and Qstar.

The term matrix-assisted laser desorption ionization (MALDI) was coined in 1985 by Franz Hillenkamp, Michael Karas and their colleagues (Karas et al. 1985). A matrix is used to protect biomolecules from being destroyed by direct laser beam. Using laser and matrix combination, ionization of large biomolecules is possible (Karas and Hillenkamp, 1988). Further improvements were realized through the use of a 355 nm laser and the cinnamic acid derivatives ferulic acid, caffeic acid and sinapinic acid as the matrix (Beavis et al. 1989). Today, MALDI-TOF MS became a popular and versatile method to analyze macromolecules from biological origin. In combination with 1DE and 2DE separation, MALDI-TOF is used to discover disease markers. For example, peptide mass fingerprint (PMF) is a protein identification method based on the specificity of a mass spectrum of the peptide mixture resulting from the digestion of a protein by an enzyme (e.g. trypsin). After proteolysis with a specific protease, proteins of different amino acid sequence produce a series of peptides masses, which can be detected by MALDI. The spectrum of identified peptide masses is unique for a specific protein and is known as a mass fingerprint (Marvin et al. 2003). Searching the selected masses from the fingerprint against databases of known protein sequences (e.g. SwissProt-TrEMBL) enables the identification of most proteins.

SELDI technology was developed by Hutchens at Baylor College of Medicine in 1993 (Hutchens and Yip 1993). The technology was commercialized by Ciphergen Biosystems Inc. in 1997 as the ProteinChip system. In this method, proteins are captured directly on a chemically derivatized MALDI plate. SELDI-TOF MS is similar to MALDI-TOF. They differ in how the matrix, or energy-absorbing molecule, is mixed with the protein sample. In MALDI, a protein or peptide sample is mixed with the matrix molecule in solution. Small amounts of the mixture are "spotted" on a surface and allowed to dry. The peptide sample and matrix co-crystallizes as the solvent evaporates. In SELDI the protein mixture is spotted on a surface modified with some chemical functionality. Some proteins in the sample bind to the surface, while the others are removed by washing. After washing the spotted sample, matrix is applied to the surface and allowed to crystallize with the sample peptides. Binding to the SELDI surface acts as a chromatography step and the subset of proteins that bind to the surface are easier to analyze. Common surfaces include CM10 (weak-cation exchanger), H50 (hydrophobic surface, similar to C6-C12), IMAC30 (metal-binding surface), and Q10 (strong anion exchanger). Surfaces can also be functionalized with antibodies, other proteins, or DNA.

SELDI allows the discrimination of peptides based on mass over charge ratio and provides a semi-quantitative evaluation, but cannot identify these peptides (Fung et al. 2001), which is the major disadvantage of this technology. So each protein of interest has to be purified or enriched for identification purpose. The application of SELDI led to many exciting results (Petricoin et al. 2002), although results were not always reproducible (Editorial, 2004; Sorace and Zhan, 2003; Baggerly et al. 2004). In Table 1, a list of several representative SELDI publications in cancer biomarker research was posted to reflect the historical interest in this technology during the time period. While the enthusiasm in traditional SELDI declining, MALDI-TOF MS which is a particle-counting method that responds to molar abundance, still represent a useful tool for surveying small proteins and peptides. It is complementary to techniques such as electrophoresis and HPLC, which have an advantage for detecting larger molecules (Hortin, 2006).

The methods used to explore cancer biomarkers will be discussed to facilitate the understanding of different techniques involved in different strategies. This can be simply categorized into two groups: the "bottom up" (peptide level) and "top down" (intact protein level) approaches (Figure 1, from Dr. Weinberger of GenNext Technologies™ Inc. with permission). The "bottom up" approach

involves protein digestion followed by mass measurement of the resulting peptides and subsequent determination of partial sequence of the peptides. This data set is then compared against a data set composed of theoretical peptides, their masses and sequences for prediction of the identity of the peptide or protein in question. The result of this comparison is a report of the matches between the measured peptides and the theoretical peptides, usually listed in order of decreasing strength of the match. The "top down" approach first involves mass measurement of the intact protein followed by attempts to identify the protein on the basis of this value and the measured molecular weights and sequences of the peptides derived from each digested protein. Careful and thorough examination of plasma/serum proteomes for disease biomarkers identification should embrace both approaches because each has their own advantages and disadvantages. For example, the bottom-up approach has been successfully adapted to high throughput screening of complex proteomes but it lacks the molecular weight information of the intact protein and is less effective in recognizing and identifying the presence of PTMs (Bogdanov and Smith, 2005).

Few of the known protein biomarkers of cancer are exclusive for a specific malignancy, and most are also found in non-malignant conditions. Current cancer biomarkers are associated with an abnormal temporal, quantitative or conformational presentations (i.e. in incorrectly spliced or post-translationally modified forms), or combinations thereof. Although some of these associations are mere reflections of abnormalities downstream of the real pathogenesis of the disease, it does not detract from their value as markers for cancer detection or stratification. Consequently, the most useful cancer biomarkers are likely to be a suite of proteins that change in relative abundance during the disease process and during treatment. The implication of this is that the technology to detect these biomarkers must be both qualitative and sufficiently quantitative to record subtle changes in the plasma/serum proteome in the face of a large number of other proteins or after depletion procedures.

To this end, significant investments have already been made in the exploration of the serum/plasma proteome to diagnose disease (Conrads et al. 2005, 2006). Electrospray (Whitehouse et al. 1985; Mora et al. 2000; Fenn, 2003) and laser ionization mass
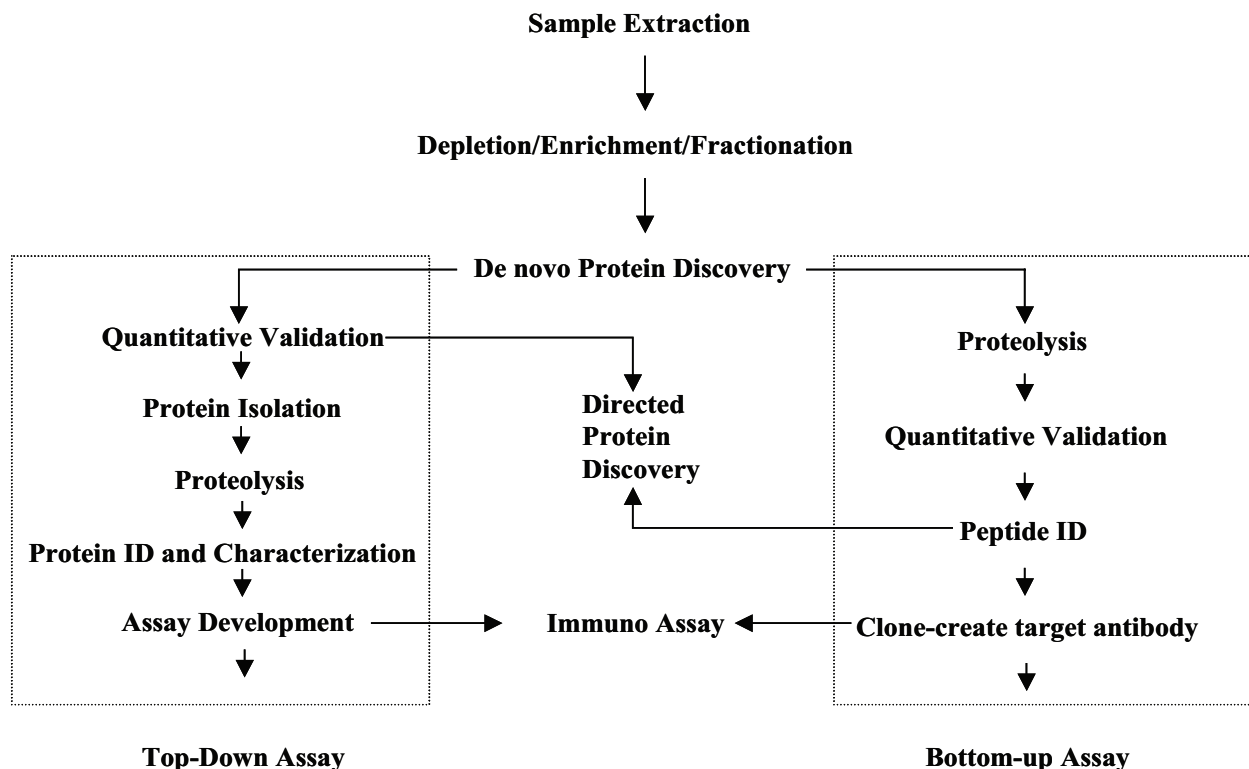

Figure 1. Proteomic biomarker discovery: integration of Top-Down and Bottom-Up approaches.

spectrometry have become important in this process as the most robust methods for ionizing a wide range of proteins and peptides prior to their mass spectrometric analysis. Several other important techniques have been used effectively to probe complex proteomes. Several important techniques are useful in this effort. The MUDPIT technology (Washburn et al. 2001) has already been described for the successful fractionation of complex proteomes. Tryptic digestion of the mixture is followed by two or more steps of sequential liquid chromatography coupled with ESI mass spectrometry. Isotope coded affinity tags (ICAT, Gygi et al. 1999) and the newer, more successful technique of isobaric tags for relative and absolute quantification (iTRAQ, Zhang Y. et al. 2005) both use stable isotope- labeled reagents for relatively quantitative comparisons of the proteomes of two or more samples. These strategies are based on the reaction of protein mixtures using reagents with specificity toward certain functional groups such as the free sulfhydryl groups on cysteine residues. The reagents contain components with different molecular weights and an affinity tag. Samples are separately reacted using the reagents with different molecular weights. The samples are then mixed. The same peptides present in multiple samples differ in molecular weight and can be distinguished by MS, and their relative signal intensities accurately reflect their relative abundances in the original samples. Extrapolation of such approaches for a comparison of relatively large numbers of samples, such as will be collected in the course of clinical trials, is an issue with which the field has yet to grapple. Also iTRAQ technique labels all peptides in a sample making affinity separations unnecessary. Similarly, iTRAQ reagents are all isobaric, the difference in mass only becomes apparent due to the "reporter ion" in the MS/MS dimension. Finally, Wall et al. (2000) developed a unique two-dimensional all liquid-phase method combined with MS for protein profile analysis. With this method, proteins are fractionated by p*I* using isoelectric focusing (IEF) in the Rotofor cell and then separated by hydrophobicity using reverse phase-HPLC in the second dimension.

Reverse phase protein microarray (RPMA) in clinical phosphoproteomic profiling

Besides mass spectrometry technology, protein microarray technology has also been widely used in proteomic studies. Protein microarrays can be divided into two broad categories: forward phase microarrays (FPMA) and reverse phase microarrays (RPMA). In the FPMA format, the analyte is captured from solution using a "bait" molecule immobilized on the array substrate. In contrast, in the RPMA format, the analyte is immobilized directly on the array substrate before being probed with an analyte-specific ligand, usually an antibody (Liotta et al. 2003; Templin et al. 2002; Mundinger et al. 2006).

RPMA technology, first introduced in 2001, is well suited to clinical proteomic research of oncology. One of the advantages of RPMA is to measure multiple analytes simultaneously from relatively smaller numbers of cells than required by mass spectrometry tools (Mundinger et al. 2006). Another advantage is to study the phosphorylation and dephosphorylation events mediated by protein kinases, which are critical in transduction networks and their aberrancies in cancer constitutes an exciting frontier in oncology (Pawson, 2002, Mundinger et al. 2006). Moreover, this method is highly sensitive with detection capabilities of 50 fg/l, or 1,000 to 5,000 molecules per spot (Liotta et al. 2003; Paweletz et al. 2001; Geho et al. 2005). However, the limitations of this analytical tool are that the quality of the arrays depends, as in all immunoassays, on the performance of the primary antibody used. The endogenous biotin, immunoglobins, peroxidases, alkaline phosphatases, or fluorescent proteins, contained in biological samples, have the potential to interfere with amplification methods currently used for array detection (Mundinger et al. 2006). And finally since RPMA is dependent on antibodies, it cannot identify novel protein species, although it may identify novel network interactions among previously characterized proteins (Jones et al. 2006).

To date, several studies have utilized RPMA in the analysis of cancer signaling pathway alterations with clinical prostate (Paweletz et al. 2001; Grubb et al. 2003; Herrmann et al. 2003), ovarian (Wulfkuhle et al. 2003; Sheehan et al. 2005), colorectal (Belluco et al. 2005), breast cancer (Zhang DH et al. 2005ab; Cowherd et al. 2004, Wulfkuhle et al. 2002; Kersting et al. 2004) and lymphoma (Zha et al. 2004; Gulmann et al. 2005) specimens using RPMA. In these studies, pro-survival, pro-mitogenic, and cell-cycle regulatory proteins, including phospho-Akt, PI3K, ERK, MAPK, PKCα, p38, STAT1, GSK3-β, cytochrome

c oxidase, c-erbB2, and c-erbB1, phospho-PTEN, EGFR, have been studied.

Advantages and pitfalls of current proteomic technology

Proteomic technology enables high throughput analysis of protein biomarkers, and therefore provides an opportunity to identify and evaluate all potential protein biomarkers for early detection and to predict various tumor behaviors, including response to chemotherapy. This approach is gaining popularity among cancer researchers in their quest for cancer biomarkers with high diagnostic, prognostic and predictive accuracy. However, despite initial excitement, skepticism about the methodology and the lack of concordance of results among labs and even within the same laboratories has eroded confidence in this technology (Diamandis 2004a, 2004b; Baggerly et al. 2004; Check, 2004; Villanueva and Tempst, 2004; Ransohoff, 2005). The simple fact that different research groups have found different discriminatory markers when analyzing similar samples suggests that the methodology needs to be standardized to improve comparability, reproducibility and reliability of the findings.

Apart from reproducibility and reliability of the mass spectrometric technology, other factors may also complicate the analytical effort. For example, sample collection, storage, and processing procedures can produce proteomic artifacts that could overshadow those representing of cancer (Villanueva et al. 2005; Banks et al. 2005; Karsan et al. 2005). Also the protein turnover rates in serum may be affected by liver and kidney function of each individual. It is known that cancer-associated biomarkers are also released by conditions other than cancer, which lowers specificity of the approach. Therefore, extreme sensitivity may be required because crucially important proteins often exist at low concentration in a particular type of cancer at its early stage.

Detection of proteins released from tumor cells into the circulation is a challenging task, so prudent choices need to be made when selecting the proper technology and strategy. Depending on the experiment and the analyte, the sensitivity of proteomic LCMS experiments is now down to the 50 fmol to 10 attomol of analyte-injected range. This translates to 5 ngm to 1 pgm of a 10 kDa protein. If the analyte were recovered from 1 ml

of specimen, this limit of detection (LOD) is equivalent to or below the concentrations of most known circulating tumor markers, for example PSA (Maattanen et al. 2001). Therefore, mass spectrometric detection is now at the level of mid-range circulating biomarkers. Although further improvements are needed before the low-range circulating biomarkers become accessible, the improvements in mass spectrometric LOD's in the last few years have been remarkable. Nevertheless, discriminatory peaks may include acute-phase reactants (i.e. molecules whose serum concentrations are increased with acute or chronic inflammatory conditions) or other proteins or protein fragments that are released into the circulation by large organs, such as the liver, in response to the presence of a tumor or cancer, but not from cancer itself. Such epiphenomena can be mimicked by condition other than cancer including infection, inflammation, or malnutrition (Diamandis 2003a, 2003b, 2004a, 2004b).

In addition, numerous patient and environment-related variables such as gender, age, genetics, ethnicity, body mass, medication, the presence of other conditions and diseases, psychological distress, menopausal status, and dietary/nutritional preferences, may all introduce artifacts, and the effect of these factors on the serum/plasma proteome have yet to be systematically investigated.

In summary, designing a protein/peptide profiling study must be rigorous to control for all important variables. Furthermore, standardized and optimized methodology is essential for achieving accurate measurement and meaningful analysis. This includes all involved steps extending from experimental design, specimen collection, storage and handling, throughout all methods used in the analytical chemistry and MS signal processing. Proper bioinformatics including analytical tools, data storage and sharing are required for data mining and validation.

References

Adam, B.L., Qu, Y., Davis, J.W. et al. 2002. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res.*, 62:3609–14.

Aebersold, R., Anderson, L., Caprioli, R. et al. 2005. Perspective: a program to improve protein biomarker discovery for cancer. *J. Proteome. Res.*, 4:1104–9.

Alaiya, A., Al-Mohanna, M. and Linder, S. 2005. Clinical cancer proteomics: promises and pitfalls. *J. Proteome. Res.*, 4:1213–22.

Alaiya, A., Roblick, U., Egevad, L. et al. 2000b. Polypeptide expression in prostate hyperplasia and prostate adenocarcinoma. *Anal. Cell. Pathol.*, 21:1–9.

Alaiya, A.A., Franzen, B., Auer, G. et al. 2000a. Cancer proteomics: from identification of novel markers to creation of artifical learning models for tumor classification. *Electrophoresis*, 21:1210–7.

Albrethsen, J., Bogebo, R., Gammeltoft, S. et al. 2005. Upregulated expression of human neutrophil peptides 1, 2 and 3 (HNP 1–3) in colon cancer serum and tumours: a biomarker study. *BMC Cancer*, 9:5–8.

Alvarez-Chaver, P., Rodriguez-Pineiro, A.M., Rodriguez-Berrocal, F.J. et al. 2006. Identification of hydrophobic proteins as biomarker candidates for colorectal cancer. *Int. J. Biochem. Cell. Biol.*, 39:529–40.

Anderson, L. and Seilhamer, J. 1997. A comparison of selected mRNA and protein abundancies in human liver. *Electrophoresis*, 18:533–7.

Anderson, N.L. and Anderson, N.G. 2002. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics*, 1:845–67.

Baggerly, K.A., Morris, J.S. and Coombes, K.R. 2004. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics*, 20:777–85.

Baggerly, K.A., Morris, J.S., Edmonson, S.R. et al. 2005. Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. *J. Natl. Cancer Inst.*, 97:307–9.

Banks, R.E., Stanley, A.J., Cairns, D.A. et al. 2005. Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. *Clin Chem.*, 51:1637–49.

Baselga, J., Gianni, L., Geyer, C. et al. 2004. Future options with trastuzumab for primary systemic and adjuvant therapy. *Semin. Oncol.*, 31:51–7.

Bast, R.C. Jr, Ravdin, P., Hayes, D.F. et al. 2001. 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. *J. Clin. Oncol.*, 19:1865–78.

Beavis, R.C., Chait, B.T. and Standing, K.G. 1989. Matrix-assisted laser-desorption mass spectrometry using 355 nm radiation. Rapid Commun Mass *Spectrom*, 3:436–9.

Belluco, C., Mammano, E., Petricoin, E. et al. 2005. Kinase substrate protein microarray analysis of human colon cancer and hepatic metastasis. *Clin. Chim. Acta.*, 357:180–3.

Bogdanov, B. and Smith, R.D. 2005. Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom Rev.*, 24:168–200.

Bondar, O.P., Barnidge, D.R., Klee, E.W. et al. 2007. LC-MS/MS Quantification of Zn-{alpha}2 Glycoprotein: A Potential Serum Biomarker for Prostate Cancer. *Clin. Chem.*, 53:673–8.

Borozdenkova, S., Westbrook, J.A., Patel, V. et al. 2004. Use of proteomics to discover novel markers of cardiac allograft rejection. *Proteome Res.*, 3:282–8.

Check, E. 2004. Proteomics and cancer: running before we can walk? *Nature*, 429:496–7.

Chen, G., Gharib, T.G., Wang, H. et al. 2003. Protein profiles associated with survival in lung adenocarcinoma. *Proc. Natl. Acad. Sci. U.S.A.*, 100:13537–42.

Conrads, T.P., Hood, B.L., Petricoin, EF 3rd. et al. 2005. Cancer proteomics: many technologies, one goal. *Expert Rev Proteomics*, 2:693–703.

Conrads, T.P., Hood, B.L., Veenstra, T.D. 2006. Sampling and analytical strategies for biomarker discovery using mass spectrometry. *Biotechniques*, 406:799–805.

Corthals, G.L., Aebersold, R. and Goodlett, D.R. 2005. Identification of phosphorylation sites using microimmobilized metal affinity chromatography. *Methods Enzymol.*, 405:66–81.

Cowherd, S.M., Espina, V.A., Petricoin, E.F. et al. 2004. Proteomic analysis of human breast cancer tissue with laser-capture microdissection and reverse-phase protein microarrays. *Clin. Breast. Cancer*, 5:385–92.

de Noo, M.E., Mertens, B.J., Ozalp, A. et al. 2006. Detection of colorectal cancer using MALDI-TOF serum protein profiling. *Eur. J. Cancer*, 42:1068–76.

Desouza, L.V., Grigull, J., Ghanny, S. et al. 2007. Endometrial carcinoma biomarker discovery and verification using differentially tagged clinical samples with multidimensional liquid chromatography and tandem mass spectrometry. *Mol. Cell. Proteomics*. [Epub ahead of print].

Dhanasekaran, S.M., Barrette, T.R., Ghosh, D. et al. 2001. Delineation of prognostic biomarkers in prostate cancer. *Nature*, 412:822–6.

Diamandis, E.P., van der and Merwe, D.E. 2005. Plasma protein profiling by mass spectrometry for cancer diagnosis: opportunities and limitations. *Clin. Cancer Res.*, 11:963–5.

Diamandis, E.P. 2002. Proteomic patterns in serum and identification of ovarian cancer. *Lancet*, 360:170.

Diamandis, E.P. 2003a. Point: Proteomic patterns in biological fluids: do they represent the future of cancer diagnostics? *Clin. Chem.*, 49:1272–5.

Diamandis EP. 2003b. Serum proteomic patterns for detection of prostate cancer. *J. Natl. Cancer Inst.*, 95:489–90.

Diamandis EP. 2004a. Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J. Natl. Cancer, Inst.*, 96:353–6.

Diamandis EP. 2004b. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Mol. Cell. Proteomics*, 3:367–78.

Domon, B. and Aebersold, R. 2006. Mass spectrometry and protein analysis. *Science*, 312:212–7.

Druker, B.J. 2004. Imatinib as a paradigm of targeted therapies. *Adv. Cancer Res.*, 91:1–30.

Dummin, L.J., Cox, M., Plant L. 2007. Prediction of breast tumor size by mammography and sonography-A breast screen experience. *Breast*, 16:38–46.

Durand, G. and Seta, N. 2000. Protein glycosylation and diseases: blood and urinary oligosaccharides as markers for diagnosis and therapeutic monitoring. *Clin. Chem.*, 46:795–805.

Ebert, M.P., Meuer, J., Wiemer, J.C. et al. 2004. Identification of gastric cancer patients by serum protein profiling. *J. Proteome. Res.*, 3:1261–6.

Echan, L.A., Tang, H.Y., Ali-Khan, N. et al. 2005. Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. *Proteomics*, 5:3292–303.

Editorial. 2004. Proteomic diagnostics tested. *Nature*, 429:487.

Engwegen, J.Y., Helgason, H.H., Cats, A. et al. 2006. Identification of serum proteins discriminating colorectal cancer patients and healthy controls using surface-enhanced laser desorption ionisation-time of flight mass spectrometry. *World. J. Gastroenterol*, 12:1536–44.

Etzioni, R., Urban, N., Ramsey, S. et al. 2003. The case for early detection. *Nat. Rev. Cancer*, 3:243–52.

Fenn, J.B. 2003. Electrospray wings for molecular elephants (Nobel lecture). *Angew. Chem. Int. Ed. Engl.*, 42:3871–94.

Fung, E.T., Thulasiraman, V., Weinberger, S.R. et al. 2001. Protein biochips for differential profiling. *Curr. Opin. Biotechnol.*, 12:65–9.

Geho, D., Lahar, N., Gurnani, P. et al. 2005. Pegylated, steptavidin-conjugated quantum dots are effective detection elements for reverse-phase protein microarrays. *Bioconjug, Chem.*, 16:559–66.

Gillette, M.A., Mani, D.R., Carr, SA. 2005. Place of pattern in proteomic biomarker discovery. *J. Proteome Res.*, 4:1143–54.

Group, Biomarkers Definitions Working Group. 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther, 69:89–95.

Grubb, R.L., Calvert, V.S., Wulkuhle, J.D. et al. 2003. Signal pathway profiling of prostate cancer using reverse phase protein arrays. *Proteomics*, 3:2142–6.

Gulmann, C., Espina, V., Petricoin, E. et al. 2005. Proteomic analysis of apoptotic pathways reveals prognostic factors in follicular lymphoma. *Clin. Cancer Res.*, 11:5847–55.

Gygi, S.P., Han, D.K., Gingras, A.C. et al. 1999. Protein analysis by mass spectrometry and sequence database searching: tools for cancer research in the post-genomic era. *Electrophoresis*, 20:310–9.

He, J., Gornbein, J., Shen, D. et al. 2007. Detection of breast cancer biomarkers in nipple aspirate fluid by SELDI-TOF and their identification by combined liquid chromatography-tandem mass spectrometry. *Int. J. Oncol.*, 30:145–54.

Heike, Y., Hosokawa, M., Osumi, S. et al. 2005. Identification of serum proteins related to adverse effects induced by docetaxel infusion from protein expression profiles of serum using SELDI ProteinChip system. *Anticancer Res.*, 25:1197–203.

Herrmann, P.C., Gillespie, J.W., Charboneau, L. et al. 2003. Mitochondrial proteome: altered cytochrome c oxidase subunit levels in prostate cancer. *Proteomics*, 3:1801–10.

Hortin, G.L. 2006. The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. *Clin. Chem.*, 52:1223–37.

Hu, S., Xie, Y., Ramachandran, P. et al. 2005. Large-scale identification of proteins in human salivary proteome by liquid chromatography/mass spectrometry and two-dimensional gel electrophoresis-mass spectrometry. *Proteomics*, 5:1714–28.

Hunt, K.K. and Keyomarsi, K. 2005. Cyclin E as a prognostic and predictive marker in breast cancer. *Semin. Cancer Biol.*, 15:319–26.

Hutchens, T.W. and Yip, T.T. 1993. New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun Mass Spectrom*, 7: 576–80.

Jemal, A., Siegel, R., Ward, E. et al. 2006. Cancer statistics, 2006. CA *Cancer. J. Clin.*, 56:106–30.

Jemal, A., Siegel, R., Ward, E., Murray, T., Xu, J., Thun, M.J. Cancer statistics, 2007. *CA Cancer J. Clin.*, 2007, 57:43–66.

Jenkins, R.E., Pennington, S.R. 2001. Arrays for protein expression profiling: towards a viable alternative to two-dimensional gel electrophoresis?. *Proteomics*, 1:13–29.

Jessani, N., Niessen, S., Wei, B.Q. et al. 2005. A streamlined platform for high-content functional proteomics of primary human specimens. *Nat. Methods.*, 2:691–7.

Jones, R.B., Gordus, A., Krall, J.A. et al. 2006. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature*, 439:168–74.

Karas, M., Bachmann, D. and Hillenkamp, F. 1985. Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Anal. Chem.*, 57:2935–9.

Karas, M. and Hillenkamp, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.*, 60:2299–301.

Karsan, A., Eigl, B.J., Flibotte, S. et al. 2005. Analytical and preanalytical biases in serum proteomic pattern analysis for breast cancer diagnosis. *Clin. Chem.*, 51:1525–8.

Kersting, C., Tidow, N., Schmidt, H. et al. 2004. Gene dosage PCR and fluorescence in situ hybridization reveal low frequency of EGFR amplifications despite protein overexpression in invasive breast carcinoma. *Lab Invest*, 84:582–7.

Kozak, K.R., Amneus, M.W., Pusey, S.M. et al. 2003. Identification of biomarkers for ovarian cancer using strong anion-exchange Proteinchips: Potential use in diagnosis and prognosis. *Proc. Natl. Acad. Sci., U.S.A.*, 100:12343 –8.

Kozak, K.R., Su, F., Whitelegge, J.P. et al. 2005. Characterization of serum biomarkers for detection of early stage ovarian cancer. *Proteomics*, 5:4589–96.

Li, J., Orlandi, R., White, C.N. et al. 2005. Independent validation of candidate breast cancer serum biomarkers identified by mass spectrometry. *Clin. Chem.*, 51:2229–35.

Li, J., Zhang, Z., Rosenzweig, J. et al. 2002. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin. Chem.*, 48:1296–304.

Li, Y., De Voe, D.L. and Lee, C.S. 2003. Dynamic analyte introduction and focusing in plastic microfluidic devices for proteomic analysis. *Electrophoresis*, 24:193–9.

Lin, Y.W., Lin, C.Y., Lai, H.C. et al. 2006. Plasma proteomic pattern as biomarkers for ovarian cancer. *Int. J.Gynecol. Cancer*, 16 Suppl 1:139–46.

Liotta, L.A., Espina, V., Mehta, A.I.et al. 2003. Protein microarrays: meeting analytical challenges for clinical applications. *Cancer Cell* 3:317–25.

Liu, Y., Liu, H., Han, B. et al. 2006. Identification of 14–3–3sigma as a contributor to drug resistance in human breast cancer cells using functional proteomic analysis. *Cancer Res.*, 66:3248–55.

Maattanen, L., Auvinen, A., Stenman, U.H. et al. 2001. Three-year results of the Finnish prostate cancer screening trial. *J. Natl. Cancer. Inst.*, 937:552–3.

Marvin, L.F., Roberts, M.A. and Fay, L.B. 2003. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry. *Clin. Chim. Acta.*, 337:11–21.

Melle, C., Ernst, G., Schimmel, B. et al. 2005. Discovery and identification of alpha-defensins as low abundant, tumor-derived serum markers in colorectal cancer. *Gastroenterology*, 129:66–73.

Mendrinos, S., Nolen, J.D., Styblo, T. et al. 2005. Cytologic findings and protein expression profiles associated with ductal carcinoma of the breast in ductal lavage specimens using surface-enhanced laser desorption and ionization-time of flight mass spectrometry. *Cancer*, 105:178–83.

Mora, J.F., Van Berkel, G.J., Enke, C.G. et al. 2000. Electrochemical processes in electrospray ionization mass spectrometry. *J. Mass. Spectrom.*, 35:939–52.

Moshkovskii, S.A., Serebryakova, M.V., Kuteykin-Teplyakov, K.B. et al. 2005. Ovarian cancer marker of 11.7 kDa detected by proteomics is a serum amyloid A1. *Proteomics*, 5:3790–7.

Mundinger, M.S., Espina, V., Liotta, L.A. et al. 2006. Clinical phosphoproteomic profiling for personalized targeted medicine using reverse phase protein microarray. *Targeted Oncology*, 1:151–67.

Nishizuka, S., Charboneau, L., Young, L. et al. 2003. Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc. Natl. Acad. Sci., U.S.A.*, 100:14229–34.

Ogata, Y., Hepplmann, C.J., Charlesworth, M.C. et al. 2006. Elevated levels of phosphorylated fibrinogen-alpha-isoforms and differential expression of other post-translationally modified proteins in the plasma of ovarian cancer patients. *J. Proteome. Res.*, 5:3318–25.

Paweletz, C.P., Charboneau, L., Bichsel, V.E. et al. 2001. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene.*, 20:1981–9.

Pawson, T. 2002. Regulation and targets of receptor tyrosine kinases. *Eur. J. Cancer.*, 38 Suppl 5:S3–S10.

Pawlik, T.M., Hawke, D.H., Liu, Y. et al. 2006. Proteomic analysis of nipple aspirate fluid from women with early-stage breast cancer using isotope-coded affinity tags and tandem mass spectrometry reveals differential expression of vitamin D binding protein. *BMC Cancer*, 6:68.

Pennisi, E. 2003. Bioinformatics. Gene counters struggle to get the right answer. *Science*, 301:1040–1.

Petricoin, E.F., Ardekani, A.M., Hitt, B.A. et al. 2002. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*, 359:572–7.

Rai, A.J., Zhang, Z., Rosenzweig, J. et al. 2002. Proteomic approaches to tumor marker discovery. *Arch. Pathol. Lab. Med.*, 126:1518–26.

Ransohoff, D.F. 2005. Bias as a threat to the validity of cancer molecular-marker research. *Nat. Rev. Cancer*, 5:142–9.

Ratnayake, C.K., Oh, C.S. and Henry, M.P. 2000. Characteristics of particle-loaded monolithic sol-gel columns for capillary electrochromatography. I. Structural, electrical and band-broadening properties. *J. Chromatogr, A.*, 887:277–85.

Rezaee, F., Casetta, B., Levels, J.H. et al. 2006. Proteomic analysis of high-density lipoprotein. *Proteomics*, 6:721–30.

Ross, J.S., Fletcher, J.A., Bloom, K.J. et al. 2004a. Targeted therapy in breast cancer: the HER-2/neu gene and protein. *Mol. Cell. Proteomics.*, 3:379–98.

Ross, J.S., Linette, G.P. Stec, J. et al. 2004b. Breast cancer biomarkers and molecular medicine: part II. *Expert. Rev. Mol. Diagn.*, 4:169–88.

Ru, Q.C., Zhu, L.A., Silberman, J. et al. 2006. Label-free semi-quantitative peptide feature profiling of human breast cancer and breast disease sera via two-dimensional liquid chromatography-mass spectrometry. *Mol. Cell. Proteomics.*, 5:1095–104.

Sauter, E.R., Shan, S., Hewett, J.E. et al. 2005. Proteomic analysis of nipple aspirate fluid using SELDI-TOF-MS. *Int. J. Cancer*, 114:791–6.

Schrimpf, S.P., Meskenaite, V., Brunner, E. et al. 2005. Proteomic analysis of synaptosomes using isotope-coded affinity tags and mass spectrometry. *Proteomics*, 5:2531–41.

Seregni, E., Coli, A., Mazzucca, N. et al. 2004. Circulating tumour markers in breast cancer. *Eur. J. Nucl. Med. Mol. Imaging.*, 31 Suppl 1: S15–22.

Shau, H., Chandler, G.S., Whitelegge, J.P. et al. 2003. Proteomic profiling of cancer biomarkers. *Brief. Funct. Genomic. Proteomic.*, 2:147–58.

Sheehan, K.M., Calvert, V.S., Kay, E.W. et al. 2005. Use of reverse phase protein microarrays and reference standard development for molecular network analysis of metastatic ovarian carcinoma. *Mol. Cell. Proteomics.*, 4:346–55.

Shingleton, H.M., Patrick, R.L., Johnston, W.W. et al. 1995. The current status of the Papanicolau smear. *CA Cancer J. Clin.*, 45:305–20.

Shiwa, M., Nishimura, Y., Wakatabe, R. et al. 2003. Rapid discovery and identification of a tissue-specific tumor biomarker from 39 human cancer cell lines using the SELDI ProteinChip platform. *Biochem Biophys. Res. Commun.*, 309:18–25.

Smith, R.D. 2006. Future directions for electrospray ionization for biological analysis using mass spectrometry. *Biotechniques*, 41:147–8.

Sorace, J.M. and Zhan, M. 2003. A data review and re-assessment of ovarian cancer serum proteomic profiling, *BMC. Bioinformatics*, 4:24.

Speers, A.E., Cravatt, B.F. 2004. Profiling enzyme activities in vivo using click chemistry methods. *Chem. Biol.*, 11:535–46.

Stamey, T.A., Yang, N., Hay, A.R. et al. 1987. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. *N. Engl. J. Med.*, 317: 909–16.

Stein, L.D. 2004. Human genome: end of the beginning. *Nature*, 431:915–6.

Tao, W.A., Wollscheid, B., O'Brien, R. et al. 2005. Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry. *Nat. Methods.*, 2:591–8.

Templin, M.F., Stoll, D., Schrenk, M. et al. 2002. Protein microarray technology. *Trends Biotechnol.*, 20:160–6.

Traub, F., Jost, M., Hess, R. et al. 2006. Peptidomic analysis of breast cancer reveals a putative surrogate marker for estrogen receptor-negative carcinomas. *Lab. Invest.*, 86:246–53.

Umar, A., Luider, T.M., Foekens, J.A. et al. 2007. NanoLC-FT-ICR MS improves proteome coverage attainable for approximately 3000 laser-microdissected breast carcinoma cells. *Proteomics*, 7:323–9.

Villanueva, J., Philip, J., Chaparro, C.A. et al. 2005. Correcting common errors in identifying cancer-specific serum peptide signatures. *J. Proteome. Res.*, 4:1060–72.

Villanueva, J. Tempst, P. 2004. OvaCheck: let's not dismiss the concept. Nature, 430:611.

Wall, D.B., Kachman, M.T., Gong, S. et al. 2000. Isoelectric focusing nonporous RP HPLC: a two-dimensional liquid-phase separation method for mapping of cellular proteins with identification using MALDI-TOF mass spectrometry. *Anal. Chem.*, 72:1099–111.

Wang, H., Hanash, S. 2005. Intact-protein based sample preparation strategies for proteome analysis in combination with mass spectrometry. *Mass. Spectrom. Rev.*, 24:413–26.

Wang, M.C., Valenzuela, L.A., Murphy, G.P. et al. 1979. Purification of a human prostate specific antigen. *Invest, Urol.*, 17: 159–63.

Washburn, M.P., Wolters, D., Yates, J.R. 3rd. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.*, 19:242–247.

Whitehouse, C.M., Dreyer, R.N., Yamashita, M. et al. 1985. Electrospray interface for liquid chromatographs and mass spectrometers. *Anal. Chem.*, 57:675–9.

Winawer, S.J., St. John, D.J., Bond, J.H. et al. 1995. Prevention of colorectal cancer: guidelines based on new data. WHO Collaborating Center for the Prevention of Colorectal Cancer. *Bull. World, Health, Organ.*, 73:7–10.

Winawer, S.J. 2001. A quarter century of colorectal cancer screening: progress and prospects. *J Clin Oncol.*, 19(18 [Medline] Suppl):6S–12S.

Woong-Shick, A., Sung-Pil, P., Su-Mi, B. et al. 2005. Identification of hemoglobin-alpha and -beta subunits as potential serum biomarkers for the diagnosis and prognosis of ovarian cancer. *Cancer Sci.*, 96:197–201.

Wulfkuhle, J.D., Aquino, J.A., Calvert, V.S. et al. 2003. Signal pathway profiling of ovarian cancer from human tissue specimens using reverse-phase protein microarrays. *Proteomics*, 3:2085–90.

Wulfkuhle, J.D., Sgroi, D.C., Krutzsch, H. et al. 2002. Proteomics of human breast ductal carcinoma in situ. *Cancer Res.*, 62:6740–9.

Ye, B., Cramer, D.W., Skates, S.J. et al. 2003. Haptoglobin-α subunit as potential serum biomarker in ovarian cancer: identification and characterization using proteomic profiling and mass spectrometry. *Clin. Cancer Res.*, 9:2904–11.

Yu, B., Li, S.Y., An, P. et al. 2004. Comparative study of proteome between primary cancer and hepatic metastatic tumor in colorectal cancer. *World. J. Gastroenterol.*, 10:2652–6.

Zha, H., Raffeld, M., Charboneau, L. et al. 2004. Similarities of prosurvival signals in Bcl-2-positive and Bcl-2-negative follicular lymphomas identified by reverse phase protein microarray. *Lab. Invest.*, 84:235–44.

Zhang, D.H., Tai, L.K., Wong, L.L. et al. 2005a. Proteomics of breast cancer: enhanced expression of cytokeratin19 in human epidermal growth factor receptor type 2 positive breast tumors. *Proteomics*, 5:1797–805.

Zhang, D.H., Wong, L.L., Tai, L.K. et al. 2005b. Overexpression of CC3/TIP30 is associated with HER-2/neu status in breast cancer. *J. Cancer Res. Clin. Oncol.*, 131:603–8.

Zhang, H., Yi, E.C., Li, X.J. et al. 2005. High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. *Mol. Cell. Proteomics*, 4:144–55.

Zhang, Y., Wolf-Yadlin, A., Ross, P.L. et al. 2005. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell. Proteomics.*, 4:1240–50.

Zhang, Z., Bast, R.C. Jr, Yu, Y. et al. 2004. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.*, 64:5882–90.