Proteomic analysis for the early detection and rational treatment of cancer—realistic hope?

E. M. Posadas¹,²*, F. Simpkins², L. A. Liotta¹,², C. MacDonald² & E. C. Kohn¹,²

¹Laboratory of Pathology, National Cancer Institute and ²NCI/FDA Clinical Proteomics Program, Bethesda, MD, USA

Received 30 April 2004; revised 19 July 2004; accepted 23 July 2004

Proteomics is an emerging field in medical science focused on the library of proteins specific to a given biosystem, the proteome, and understanding relationships therein. This field incorporates technologies that can be applied to serum and tissue in order to extract important biological information to aid clinicians and scientists in understanding the dynamic biology of their system of interest, such as a patient with cancer. These tools include laser capture microdissection, tissue lysate arrays and mass spectrometry approaches. These new technologies are more potent coupled with advanced bioinformatics analysis. They are used to characterize the content of, and changes in, the proteome induced by physiological changes, benign and pathologic. The application of these tools has assisted in the discovery of new biomarkers and may lead to new diagnostic tests and improvements in therapeutics. These tools additionally can provide a molecular characterization of cancers, which may allow for individualized molecular therapy. Understanding the basic concepts and tools used will illustrate how best to apply these technologies for patient benefit for the early detection of cancer and improved patient care.

Key words: cancer, early detection, laser capture microdissection, proteomics, surface-enhanced laser desorption ionization, tissue lysate arrays

Introduction

Prevention, early detection and early intervention are the primary aims of oncologists and cancer biologists. The available therapeutic measures for most malignant conditions yield cures only when applied at early stages, when disease bulk is both minimal and localized. Multiple studies have shown that early intervention improves outcome. This has driven physicians and scientists to seek accurate and cost-effective means of early diagnosis, preferably in the preinvasive and/or premalignant phases. There are few malignancies for which we have highly sensitive and specific screening procedures at this time. The frequency of identification of premalignant and preinvasive disease remains inadequate, even in those cancers for which we have screening modalities in place, such as breast, prostate and colon cancers. The information from screening may be misleading due to its inadequate sensitivity and specificity, leading the physician and patient to a false sense of confidence or to unnecessary invasive procedures.

The field of proteomics encompasses tools, technologies and approaches targeted at studying variations within the proteome of a given biosystem. In clinical applications, the proteome encompasses the protein content in a patient. Cancer is commonly described as a genetic disease. A gene alone, however, is only potential information that must be put into a functional form. The DNA is transcribed into RNA then translated into protein, the manifestation of the genetic information. There is information in the proteome that cannot be predicted simply from its related nucleic acid sequence. A number of alterations or modifications can occur at transcriptional, translational and posttranslational levels to affect function. These include gene amplification, alternative RNA splicing, co-translational modification, posttranslational modifications, and differential stability and secretion of proteins [1]. Laboratory testing used in the clinical setting is targeted at expressed proteins and not genomic information. Therapeutics are now being developed that work at the level of the proteome, such as targeted small molecules and recombinant humanized monoclonal antibodies (rhuMabs). Examples of these agents include imatinib mesylate (Gleevec), gefitinib (Iressa), sorafenib (BAY 43-906), bevacizumab (rhuMab against vascular endothelial growth factor), trastuzumab (Herceptin; rhuMab against Her2) and cetuximab (Erbitux; rhuMab against epidermal growth factor receptor).

The field of proteomics studies proteins in an effort to catalog them and to understand their role in biology and pathology so they may be applied to early diagnosis and to optimizing treatments. The applications of proteomic technologies may benefit the oncologic community in several areas related
to biomarker discovery and treatment: serum screening and tissue sample analysis for the early detection of malignancy and molecular signal mapping to implement rational pharmacoproteomic therapeutic interventions.

**Proteomic technologies**

Specific protein content is traditionally measured using low-throughput techniques such as western blotting, in situ hybridization and immunohistochemical staining [2]. Two-dimensional gel electrophoresis has been the standard discovery tool for proteomics [3, 4]. This technology has been used to identify several important biomarkers by virtue of separating proteins on the basis of the size and charge. Data output from this type of analysis can be slightly cumbersome, and analysis is limited to low-throughput means, preventing it from being applied to large volumes of samples in a limited time frame. Immunohistochemistry (IHC) has long been used as an adjunctive diagnostic tool in a variety of cancers. It has provided clinicians with correlative insight into potential prognosis differential diagnosis. IHC, although commonly employed, lacks uniformity of technique, appropriate controls, and standardization of antibodies and grading techniques, making it difficult to compare results across institutions, laboratories and experiments. Limitations of IHC have been addressed by other techniques, including isotopic labeling and in situ hybridization, which allow for more quantitative analysis of variations in protein expression. This technique has been applied both at the mRNA label and the protein level. Proteins and peptides that have been modified by introducing non-canonical amino acids, fluorescence tags, spin resonance labels or cross-linking agents have been used to quantify protein expression, identify protein–protein interactions and elucidate protein structures [5]. This particular technique, while more quantitative than traditional IHC, is still limited by availability of labeled compounds to perform the hybridization necessary for quantitation. Some groups have employed methods of body fluid enrichment using separation and purification techniques, including magnetic bead-based separations. While these approaches are at times capable of enhancing desired signals, important proteomic information can be lost or obscured during the purification process. New technologies are now available that can overcome these limitations and can be adapted for high-throughput processing with quantitative and reproducible results.

**Laser capture microdissection**

Laser capture microdissection (LCM) is a technique that allows for procurement of pure cell populations from heterogeneous tissue sections under direct microscopic visualization in a clean fashion [6]. LCM can be used to isolate tumor and stroma from a single core of tissue, providing the opportunity for independent analysis of the tumor and its local microenvironment. It employs a pulsed infrared laser to activate a thermoplastic film placed over the cells of interest, causing the film to become fused to the cells. Laser shots measure 7.5–50 μm in diameter, and are repeated until all cells of interest are collected onto a film-coated plastic cap. The cap is then lifted away from the tissue and the plastic-fused cells are removed en bloc from the tissue specimen. These can be placed into lysis buffer to liberate DNA, RNA or protein [7, 8].

Selection of cells by this method allows clean separation of malignant, in situ and various normal cell subpopulations within a single biopsy specimen. Application of this technology improves the ability of other tools to detect DNA, RNA or protein signals that may be of relatively low abundance by diminishing confounding or diluting protein signals [9]. Most importantly, collection of cells with LCM preserves the molecular composition and architecture of the cells so that direct comparisons of transcriptional and translational messages can be made between tissue microcompartments of the same sample, providing a snapshot of a tumor’s in vivo biological and physiological properties.

**Tissue lysate arrays**

Protein microarrays can be applied to serum [10], and lyzed tissue samples [11, 12] with acceptable reliability. Protein aliquots are printed onto nitrocellulose-coated glass slides using a pin arrayer. Each spot on the array can contain a homogeneous or heterogeneous set of proteins [13–15]. Protein microarray formats fall into two major classes, forward phase arrays (FPAs) and reverse phase arrays [RPAs; also called tissue lysate arrays (TLAs)], depending on whether the analyte...
is captured from the solution phase or bound to the solid phase (Figure 1).

FPAs utilize capture molecules, such as an antibody to a specific protein, which are fixed onto the nitrocellulose substratum. When the substratum is exposed to the solubilized proteome, the capture molecules serve to act as bait for the protein of interest. Each spot on the array contains one type of immobilized antibody or bait protein and each array is incubated with one test sample such as a cellular lysate from one treatment condition. Multiple analytes are measured at once [10, 11]. The major weaknesses of this approach are low sensitivity and the between-run variability [16, 17]. These arrays may be used as biological pathway screening approaches, and are limited by the fact that all antibodies printed may not be in optimal concentrations for the antigens in the protein or lysate queried.

TLAs are RPAs wherein the lysate or protein of interest is arrayed without selection via a capture molecule. This array can then be queried with an optimized antibody or ligand probe, or an unknown biological component. The TLA format immobilizes an individual test sample in each array spot, such that an array can be comprised of a variety of different patient samples or cellular lysates. Each array is incubated with one detection protein or antibody, and a single endpoint is measured across the arrayed cohort and can be directly compared across multiple samples. Replicates can be reproducibly printed at a given sitting, increasing quality control over a series of queried arrays. TLAs are now being applied to basic, translational and clinical research, including clinical trials. Clinical trials have been designed by our group in which we obtain 18-gauge core needle biopsies of sentinel tumor masses

**Figure 2.** Clinical application of laser capture microdissection (LCM) tissue lysate array (TLA) technology. A core needle biopsy can be obtained from a patient’s tumor and subjected to proteomic analysis. The core is sectioned and subjected to LCM to separate tumor and stroma. The isolated populations are then lysed and printed onto TLAs, which are then analyzed.
before and during molecular targeted drug therapy [18]. These samples are flash-frozen until needed. Frozen sections are cut and subjected to LCM for isolation of cell populations. Captured cells are lysed, and arrays are then produced (Figure 2). The printed arrays have a shelf-life of 12–24 months. TLAs are being used to study the changes in the total protein content and activation state of identified molecular targets associated with signaling cascades affecting survival, apoptosis and angiogenesis pathways that are putatively downstream or transactivated by the pharmacologically targeted molecules.

Such a tool has powerful implications for the growing field of molecular therapeutics. The trials seek to determine the biochemical efficacy of targeted agents and compare biochemical and clinical activity. The resultant data will be compiled into a database creating the background to build signal pathway diagrams for different patients, cancers and pathway effects by molecularly targeted agents. This will allow development of an understanding of the pharmacoproteomic effects of these agents, from which we may rationally employ the agents in cancer therapeutics. These data will also allow for transition from hypothesis generation to prospective individualized combinatorial therapy (Figure 3).

Mass spectrometry and protein signatures

New methods for rapid identification of both known and unknown proteins are under development. Matrix-assisted laser desorption and ionization with time-of-flight detection mass spectrometry (MS) (MALDI-TOF) and surface-enhanced laser desorption and ionization with time-of-flight spectrometry (SELDI-TOF) are two of the methods currently being employed. MALDI techniques immobilize protein samples in an energy absorbing matrix (chemical) on a chip or plate. The entire repertoire of proteins in the sample interacts with the matrix from which a selected subset of proteins is bound to, a function of the composition of the selected matrix. MALDI analysis is well suited for resolution of proteins <20 kDa, the low molecular weight proteome, a heretofore poorly dissected information reserve. Conversely, SELDI technology uses selective surfaces for binding a subset of proteins based on absorption, partition, electrostatic interaction or affinity chromatography on a solid-phase protein chip surface [19–21]. In SELDI, like MALDI, the protein-bound chip is pulsed with laser energy causing proteins or protein fragments to ionize, and fly from the chip surface down a vacuum tube to the detector plate. The time of flight is affected by the mass of the particle and the charge it bears (m/z ratio). The detector plate records the intensity of the signal at a given m/z value, and a spectrum is generated. The different peaks in the spectrum correspond to different m/z protein species. This datastream of information can be coupled with datastreams from a series of test subjects and complex bioinformatics to define discriminating for cancer detection. Since MALDI-TOF and SELDI-TOF analyses generate streams of data comprised of hundreds of thousands of data points, complex computational systems are being developed to discover changes in protein expression patterns that are predictive of cancer. A variety of artificial intelligence bioinformatic tools have been demonstrated to successfully develop discriminating signatures for different cancers [22], and early work is ongoing to demonstrate the applicability of this concept to other diseases.

Clinical applications of proteomics

The application of proteomic technologies is currently limited to research purposes. Translational scientists are exploring the merits of proteomic applications in a number of disease processes. The goals of these efforts are to improve diagnostic methods by either discovering new serological tests or biomarkers, or to improve pathological analysis using tissue proteomics.

Ovarian cancer

Ovarian cancer continues to be the leading cause of death from gynecological malignancies in the USA and Europe.
Oncologists attribute this to the inability to identify ovarian cancer at a stage when surgical therapy may be curative. Current diagnostic measures are limited in efficacy and too often establish diagnosis when the disease is advanced. This has driven a search for newer more sensitive and specific methods of detection of early ovarian cancers.

Numerous groups are pursuing serological approaches to ovarian cancer diagnosis. The most common approach to date is enzyme-linked immunosorbent assay-based quantitative techniques measuring one or several proteins at a time, and then using various forms of one or several proteins at the same time [23]. Investigators have been able to identify characteristics of SELDI-TOF and MALDI-TOF mass spectral patterns in sera of patients that may yield a sensitive and specific signature for ovarian cancer [22]. Our group used spectral patterns generated from unfractionated serum samples from patients with cancer and a series of controls. A pattern-matching heuristic algorithm was trained with spectra from known patients to recognize key features, a protein signature, allowing it to discriminate between unaffected women and patients with cancer. The signature was tested on a set of blinded patients and was able to discriminate all cancers correctly, including all stage I cases. The study yielded an overall sensitivity of 100%, with a specificity of 95%. The specificity needed to yield a 10% positive predictive value for screening ovarian cancer in the general population is 99.6% [24, 25]. This technology continues to undergo development at multiple centers.

Progress has been made since that initial report. Several groups are using similar tools for ovarian cancer and other cancers. Our group has further refined the technique to use a quadrupole tandem liquid chromatography–MS–MS system. This system has higher resolution than that used previously, and is highly reproducible. The sensitivity and specificity results of models now being generated are consistently 100% and 100% [26]. Importantly, this new approach allows isolation of the individual features for peptide sequence analysis and identification. Over 1000 proteins and peptides that may be selective or specific to ovarian cancer have now been sequenced from serum samples from women with early- and advanced-stage ovarian cancer. This allows both the biomarker component of the field and the etiologic study of ovarian cancer to advance simultaneously.

**Prostate cancer**

Prostate cancer is now the most commonly diagnosed cancer in American men and the second leading cause of cancer-related deaths in men. Current screening procedures include clinical examination and serum prostate-specific antigen (PSA) measurement. Absolute PSA levels are difficult to interpret, as non-malignant prostatic diseases can also increase PSA levels [27]. Furthermore, there are emerging data from the Prostate Cancer Prevention Trial and other sources showing that a significant number of men with PSA scores <4 ng/ml may have undetected prostate cancer [28]. This underscores the need to find better screening measures than PSA alone that are deployable in a high-throughput and affordable fashion, and that identify men with high-grade aggressive disease for whom intervention has a clinical impact on survival.

The use of LCM for analysis of prostatic tissue has been described by several groups of investigators [8, 12, 29–31] and has been proposed as a method to discover new biomarkers for diagnosis. Zheng et al. [32] have employed a combination of LCM and SELDI to analyze prostatic tissue from patients with both non-malignant benign prostatic hyperplasia (BPH) and invasive prostate cancers. They were able to identify a protein present in 94% of prostate cancers (n = 17) examined that was not present in normal epithelium (n = 17) or BPH samples (n = 17). The function and identity of this protein is still being determined; however, the situation reflects the power of these technologies to find such discriminatory information even with limited user input. Other groups have tried mass spectral techniques on serum samples [20, 31, 33–35]. Our group [20] examined blinded serum samples from 266 men, which were analyzed using SELDI with a trained bioinformatics algorithm. The test was able to correctly identify 36/38 men (95%) with prostate cancer (PSA >4 ng/ml with biopsy-proven disease), while 177/228 (78%) of men without known disease (PSA <1 ng/ml) were identified.

Lehrer et al. [34] examined serum from three sets of men: patients with prostate cancer (n = 11) or BPH (n = 12), or unaffected controls (n = 12), and were able to identify three protein peaks, using SELDI-TOF technology, present in cancer samples that were absent in both BPH and normal tissue. Li et al. [35] examined archival serum samples of 345 men from a single institution: 246 underwent radical retropubic prostatectomy and 99 had no evidence of prostate cancer on biopsy. Three peaks were identified with the ability to discriminate patients with cancer from unaffected individuals. The markers were compared individually and then as a composite. Their composite biomarker test yielded sensitivity and specificity of 67% and 65%, respectively, at the point of optimal efficiency. PSA testing alone in the same sample set yielded sensitivity and specificity of 38%. Thus, despite the relative lack of sensitivity and specificity, this approach was superior to standard PSA testing. In addition, in a set of seven patients, a signature obtained preoperatively for organ-confined disease changed in all cases to a non-cancerous spectrum. This signature was not altered at 6 weeks, showing the ability of this approach to robustly detect changes.

**Pancreatic cancer**

Pancreatic adenocarcinoma has one of the lowest survival rates for any solid cancer [36]. The poor prognosis is attributed to the fact that most patients do not develop overt symptoms until the disease has disseminated or caused local organ dysfunction [36, 37]. CA 19-9 is currently the accepted serum marker for pancreatic cancer, but is Food and Drug Administration-approved only for monitoring treatment response. Current methods of diagnosis including CA 19-9 are ineffective for identifying small, surgically resectable cancers. Preliminary studies of serum protein profiling differentiated between
patients with surgically resectable pancreatic cancer and patients with non-malignant pancreatic disease and healthy controls [36]. The two most discriminating protein peaks in one study had a sensitivity of 78% and specificity of 97%, and were even more accurate when used in conjunction with CA 19-9 [36]. Proteins differentially expressed in the pancreatic fluid have also been studied as potential biomarkers for pancreatic adenocarcinoma.

Hepatocarcinoma—intestine—pancreas/pancreatitiss-associated protein 1 (HIP/PAP-1), a protein released from pancreatic acini during acute pancreatitis and overexpressed in hepatocellular carcinoma, was identified by SELDI. It is significantly elevated in patients with pancreatic adenocarcinoma over controls [37], and may serve as a new biomarker for this disease. Current efforts at serum-based approaches using SELDI also have yielded promising results. In a study set of 116 patients, 61 with cancer and 55 controls, pancreatic cancer was detected with 95% sensitivity and 97% specificity.

Breast cancer

Breast cancer is the second leading cause of cancer death in American women. Despite advances in understanding the biology of this disease, early diagnosis and intervention is the most important factor affecting survival. Mammography is a significant advancement, but is inadequate in detection of non-calcified, premalignant and non-invasive disease. Nipple-aspirate fluid initially appeared a promising source for the detection of potential biomarkers, as it a direct sampling of breast epithelial cells. However, no clinically useful results are available [38]. Serum proteomic pattern diagnostics are now being applied to this disease. Li et al. [39] have identified three protein peaks using SELDI-TOF mass spectrometry that discriminate between stage 0–1 cancer patients and non-cancer controls. Other recently identified breast cancer biomarkers using SELDI include Hsp27, 14-3-3 sigma, and mammaglobin/lipophilin B complex [40, 41]. Application of SELDI-based technologies to serum screening in breast cancer screening is promising. Preliminary results show pattern recognizing algorithms were able to identify cancer patients with 90% sensitivity and 71% specificity in a series of 317 patient samples: 142 patients without cancer and 165 with cancer.

Conclusions

The field of proteomics has yielded a set of technologies and analytical techniques that are significantly advancing the field of cancer diagnostics. These technologies allow for efficient means of identifying new biomarkers for the early detection of cancer, and promise hope of new serological screening methods for diagnosis. Limitations remain for which the tools and techniques are being improved, and carefully tested and refined. The potential outcomes of for this technology are manifold. The physical nature of the instrumentation associated with mass spectrometric analysis can adapted to both cost-effective high-throughput applications and input into multiple models built for discrimination of different disease states. Refinements in serum- and tissue-based proteomics will allow for earlier and more accurate clinical diagnosis. Rather than relying solely upon histopathological classification of tumors, physicians may be able to further classify cancers at the molecular level, resulting in a true molecular pathological classification. The oncologist can then individualize therapeutic interventions specifically aimed at the key signaling pathways in a given patient’s tumor. There is a promise of important diagnostic and therapeutic improvements on the horizon, but further developments and validations are required before such technologies are available for clinicians. These proteomic tools have developed to the point where their applications are reaching beyond the laboratory and into the clinic in experimental clinical trials. Understanding and appreciating their promise will be of importance to both cancer biologists and oncologists in the future.

References


