

Proteomics Discovery of Disease Biomarkers

Mamoun Ahram¹ and Emanuel F. Petricoin²

¹Department of Pharmacology and Physiology, Faculty of Medicine, Mutah University, Mutah, Jordan.

²Center for Applied Proteomics and Molecular Medicine, George Mason University, Virginia, U.S.A.

Abstract: Recent technological developments in proteomics have shown promising initiatives in identifying novel biomarkers of various diseases. Such technologies are capable of investigating multiple samples and generating large amount of data end-points. Examples of two promising proteomics technologies are mass spectrometry, including an instrument based on surface enhanced laser desorption/ionization, and protein microarrays. Proteomics data must, however, undergo analytical processing using bioinformatics. Due to limitations in proteomics tools including shortcomings in bioinformatics analysis, predictive bioinformatics can be utilized as an alternative strategy prior to performing elaborate, high-throughput proteomics procedures. This review describes mass spectrometry, protein microarrays, and bioinformatics and their roles in biomarker discovery, and highlights the significance of integration between proteomics and bioinformatics.

Keywords: proteomics, mass spectrometry, protein microarrays, surface enhanced laser desorption/ionization, bioinformatics

Introduction

Discovery of biomarkers constitutes an essential part of biomedical research. The association of biomarkers to diseases advances understanding of cellular and molecular mechanisms of diseases since biomarkers can be direct causes of diseases, secondary players in disease initiation and progression, or mere signals of pathological conditions. In addition, the appearance of molecular biomarkers distinctly in certain pathologies will greatly assist in disease detection. More specifically, the presence of molecular biomarkers in specific stages of diseases will enable their use in disease diagnosis and prognosis. Molecular biomarkers can also play an important role in therapy as drug targets. Otherwise, such biomarkers can be used to assess the efficiency of therapeutic strategies, whereby the presence of certain molecules could be indicative of treatment goals and/or toxic effects of drugs during the course of therapy.

Research initiatives targeting biomarkers had been slow at various stages, in particular discovery and validation. Many research efforts had engaged the evaluation of very few molecules in a limited number of samples at one time. However, the initiation of molecular profiling studies aided by the completion of Human Genome Project and the recent development of biotechnological tools has revolutionized the search for disease biomarkers. Recent studies can, therefore, generate enormous amount of data end-points in a short period of time, thus termed “high-throughput.” High-throughput studies not only would enhance the discovery of novel biomarkers, but would also elucidate molecular and cellular interactions.

It has been calculated earlier that proteins constitute the main bulk of therapeutic targets accounting for more than 98% of drug targets (Drews, 2000). It has also been recently estimated that almost 10% of the genome is directly involved in pathogenesis with a fraction of these being potential targets of therapeutic intervention (Hopkins and Groom, 2002; Betz, 2005). These observations have legitimized the emergence of proteomics as a result of the biological significance of proteins. Among the many goals of proteomics is understanding all aspects of proteins including their expression, function, interaction, and structure. It is hoped that proteomics analyses lead to discovery of novel disease biomarkers that can be utilized for detection, prognostication, and treatment of diseases. In order to reach this goal, two important challenges must be encountered in terms of studying cellular proteomes, or complete sets of proteins in cells. The first is development of new technologies that would allow simultaneous investigation of numerous samples and multiple target proteins. The second challenge is the development of bioinformatics tools for the purpose of handling and analyzing the large data output.

Correspondence: Dr. Mamoun Ahram, Faculty of Medicine, Mutah University, P. O. Box 7, Mutah, Karak 61710, Jordan. Tel: 962-79-556-7779; Fax: 962-3-238-6175; Email: Dr.Ahram@gmail.com

Emanuel F. Petricoin, Center for Applied Proteomics and Molecular Medicine, College of Arts and Sciences, George Mason University, 10900 University Boulevard, MS 4E3, Room 181A, Manassas, Virginia 20110, U.S.A. Tel: (703) 993-9444; Fax: (703) 993-4288; Email: epetrico@gmu.edu



Copyright in this article, its metadata, and any supplementary data is held by its author or authors. It is published under the Creative Commons Attribution By licence. For further information go to: <http://creativecommons.org/licenses/by/3.0/>.

Role of Bioinformatics in Proteomics

With the explosion of data generated from proteomics studies, it has become a bottleneck dealing with this tremendous amount of information. Hence, the integration of bioinformatics into proteomics has become greatly needed to transform this information into meaningful knowledge. Such knowledge can be used to attain better understanding of biological systems as a whole, let alone an individual signaling pathway or a biological mechanism. A significant part of bioinformatics in proteomics, as well as other high-throughput studies, entails data management and mining. Management of proteomics data include the ability to store, search, query, and retrieve certain information; functions that have largely been achieved mainly by three major protein databases: SWISS-PROT, TrEMBL, and NCBI. These databases allow investigators access large amount of data from different species with variable redundancy and annotations in protein sequences. Nevertheless, a major shortcoming of protein databases is the existence of large amount of hypothetical and unknown proteins (Ahram and Springer, 2004).

Data mining in proteomics involves the ability to analyze and interpret the data generated by proteomics technologies. Bioinformatics tools used in proteomics can be divided into three main categories: data interpretation, pattern recognition, and predictive analysis. These tools will be discussed throughout this review. For some of these computational methods, the databases provided in SWISS-PROT and NCBI can be integrated in order to transform proteomics data into informative output.

In this review, two proteomics technologies that have been developed with high level of innovation in order to fit into the promises of proteomics, namely mass spectrometry (MS) and protein microarrays, will be discussed. As will be illustrated, both of these technologies have been successfully used in analyzing biological and clinical samples in search for disease biomarkers. The integration of analytical bioinformatics in proteomics and their limitations will also be discussed. In addition, the growing need and development of predictive bioinformatics will be presented.

Mass Spectrometry

Currently, mass spectrometry (MS) is the most widely used method in high-throughput proteomics

studies. A mass spectrometer can measure the masses of small molecules such as peptides by converting them into ions and sorting them via a stream of electrical fields according to their mass/charge (m/z) ratio. Basic MS instruments are composed of three components: an ionization source that converts particles into ions, a mass analyzer that sorts ions according to their m/z , and an ion detector that measures m/z . Recent and continuing development and improvement of these instruments have made them greatly suitable for high-throughput proteomics studies including the invention of soft ionization methods and the innovation of hybrid instruments composed of two mass analyzers.

Two common types of soft ionization methods exist: Matrix-Assisted Laser Desorption/Ionization (MALDI) and electrospray ionization (ESI). In MALDI, samples containing peptides are embedded into specific matrix molecules. The matrix absorbs the ionization laser beam and transfers the energy into the analyte. Sample analysis by ESI, on the other hand, involves direct injection of analyte into the ionizing chamber converting peptides into smaller ions. In both approaches, ionized peptides are directed via a mass analyzer towards a detector, which generates MS spectra with each peak representing a m/z ratio of an ion.

Different mass analyzers can be combined with ESI and MALDI ionization sources. Time-of-flight (TOF) analyzer is usually associated with MALDI ion sources. In contrast, ESI can be integrated with wider variety of mass analyzers including ion trap and quadrupole. Among ion-trap mass analyzer is Fourier transform ion cyclotron resonance (FTICR), which is a special type of ion traps where ions are trapped in a magnetic field rather than an electrical one. FTICR is a powerful mass analyzer providing the highest sensitivity, resolution, and mass accuracy. For example, it has been reported that FTICR-MS can identify peptides at concentrations as low as zeptomoles (10^{-21} moles) (Belov et al. 2000).

The generated MS spectra can then be analyzed by search programs that computationally compare the actual MS spectra to hypothetical spectra. The simplest method by which proteins can be identified is via protein mass fingerprinting (PMF). This method is based on the fact that since proteins generate peptides of distinct lengths when digested by a specific protease, the identity of proteins can be determined according to their PMF. PMF works

best when the analyzed sample is composed of a purified protein. Protein identification can also be performed in case of a simple mixture of proteins where database searching can be conducted repeatedly with successive removal of peptides assigned to a conclusive match (Jensen et al. 1997). A good example is the identification of protein spots in a two dimensional electrophoretic gel. Such spots commonly contain more than one protein that either possess similar molecular weight and charge (Gygi et al. 2000) or are protein contaminants such as cytokeratins (Shevchenko et al. 1996).

The introduction of tandem mass spectrometry (MS/MS) instruments has greatly improved MS technologies. These instruments are composed of two mass analyzers where following determination of peptide masses by the first mass analyzer, few peptide ions are individually selected and fragmented by collision-induced dissociation (CID) yielding even smaller ions. These ions are analyzed further by a second mass analyzer. Hybrid MS instruments include innovative combinations of mass analyzers, which can be of the same or different type. Examples include MALDI TOF-TOF where both mass analyzers are TOF, and MALDI-Qq-TOF that is composed of a quadrupole as the first mass analyzer and TOF as the second one. The dual mass analysis leads to determination of partial amino acid sequences of proteins resulting in more accurate identification of proteins than PMF only. Another major advantage of dual MS instruments is the ability to start with complex samples and the generation of amino acid sequences independently of sequence databases, although an informative database is still required for highly accurate results.

Interpretation of MS/MS data output is a rate-limiting step in accurate peptide identification. Several limitations in data analysis in accurate identification of proteins exist. In fact, Resing and Ahn (2005) have mentioned that only up to 25% of MS data could be interpreted accurately. These limitations are related to the MS instrument itself, the sample, and/or the database. MS instruments differ in their resolution and sensitivity of detection. For example, whereas ion-trap MS is of limited resolution, FTICR MS possesses the highest resolution and mass accuracy and is the most sensitive MS instrument (Domon and Aebersold, 2006). In terms of the database used for data interpretation, highly accurate results are obtained when the protein sequences in the utilized database are

nearly complete. In addition, the use of large protein database can result in higher level of false-positive identifications (Resing et al. 2004; Kapp et al. 2005).

The analyzed proteins may also severely hinder accurate identification. Protein complexity may stem from their synthesis in different isoforms or their modification *in vivo* or *in vitro*. The minimal level of modification of prokaryotic proteins enables better identification than for eukaryotic proteins (Lipton et al. 2002). Amino acid differences in a protein sequence resulting from DNA polymorphism among individuals may contribute to protein complexity and, hence, prevent protein identification. Nonspecific peptides can be generated either by intracellular proteases or during sample preparation. Importantly, peptides can be chemically modified *in vitro* during sample preparation where, for example, methionines can be oxidized (Johnson et al. 2005). In fact, search programs have difficulty differentiating certain amino acids since they can carry identical or similar molecular weights such as leucine and isoleucine (Johnson et al. 2005). This is also true for phenylalanine and oxidized methionine. Pairs of amino acids may also be mistaken for a single residue such as two consecutive glycines from asparagines.

Following peptide identification by search programs, validation of these results is absolutely needed. This can be done by considering multiple factors such as missed cleavage, peptide mass, peptide modification, and number of peptides identified. Another approach is to utilize discriminant analyses such as Peptideprophet, which is a Bayesian statistical computational program that ranks peptides according to probability scores. Such a computational method enhances the performance and accuracy of search programs by reducing the rate of false-positive identifications (Kapp et al. 2005). Another discriminant approach facilitates the information provided by peptide sequences in predicting chromatographic elution time (Petritis et al. 2003). Strattmatter et al. (2004) have developed a function based on a combination of the scores generated by search programs, peptide mass, the observed versus calculated peptide mass, the cleaved nature of the peptide, and the actual versus estimated chromatographic retention time for each peptide. This method was utilized in order to identify proteins shed from the extracellular surface of hamster cells using mouse and human

protein databases with limited false-positive assignments (Ahram et al. 2005).

Surface Enhanced Laser Desorption/Ionization (SELDI)

Complex diseases cannot be labeled by a single protein biomarker. Hence, a considerable portion, if not the whole proteome, needs to be scanned in search for rather a particular profile of the particular diseases. A powerful and more versatile technology named Surface Enhanced Laser Desorption/Ionization (SELDI) branched out of MALDI-TOF mass spectrometry has been developed and can provide a solution for high-throughput analyses of cellular proteomes (Yip and Hutchens, 1992). This technology enables researchers to search for single biomarkers, a group of biomarkers, or a proteome profile. In SELDI, a sample is applied on surface of a chip rather than mixed with a matrix molecule as conducted in MALDI. The chip is then placed in a vacuum chamber of the SELDI instrument where peptides and small proteins are ionized and travel towards a detector inversely according to their masses.

With SELDI, multiple samples can be simultaneously analyzed generating numerous data points making this instrument a true high-throughput proteomics instrument. Two major advantages of SELDI is the ability to analyze highly complex samples, and the low volume of needed for analysis. The versatility of SELDI stems from the fact that chip surface can be made of a defined chemical property (e.g. hydrophobic, cationic, and anionic) allowing certain classes of proteins to adsorb. Otherwise, the chip can be coated with antibodies to capture specific antigens as has been reported earlier in measuring prostate-specific antigen and prostate-specific membrane antigen (Wright et al. 2000; Xiao et al. 2001; Adam et al. 2002).

SELDI has been utilized in search for biomarkers for Alzheimer's disease (Austen et al. 2000; Carrette et al. 2003) as well as for cancers of the prostate (Pawelczak et al. 2001; Liu et al. 2003; Lehrer et al. 2005), bladder (Vlahou et al. 2001), colon (Engwegen et al. 2006), and breast (Ricolleau et al. 2006). Although direct determination of proteins represented as mass peaks is not possible, different means can be utilized to reveal the identity of specific peaks. In a recent report, SELDI analysis of cerebrospinal fluid (CSF) samples of patients with multiple sclerosis revealed the

presence of a differential peak when compared to subjects with other diseases (Irani et al. 2006). This peak was identified by further MS analyses as cystatin C, an inhibitor of the lysosomal cysteine protease cathepsin B. Although burdensome and elaborate, proteins represented by specific SELDI spectra peaks can also be identified by a series of liquid chromatography fractionation as has been illustrated by Diamond et al. (2003), Sanchez et al. (2004), and Yang et al. (2004).

In one study, proteins extracted from LCM-microdissected prostate normal and tumor cells were analyzed by SELDI. The mass spectra patterns of the proteins revealed several remarkable alterations as compared to those of matched normal samples (Petricoin et al. 2002). However, due to the dynamic heterogeneity of proteomes even within the same individual, consistent detection of differential peaks is not always feasible. This complexity has prompted the group of Petricoin and Liotta to integrate an artificial neural network algorithm to search for "hidden" patterns. In a prominent study, the group has been able to differentiate ovarian cancer patients from normal subjects and patients with other ovarian diseases with unprecedented sensitivity of 100% and specificity of 95% (Petricoin et al. 2002). This very similar approach has been utilized in detecting gastric cancers with high sensitivity and specificity in differentiating the disease (Ebert et al. 2004).

Although these results are promising, serious concerns have been raised in terms of the robustness and reproducibility of the approach. One such concern is the inability of the current SELDI-TOF instruments to directly sequence and identify the peptides/proteins that generate the discriminatory peaks, and thus be able to independently validate the markers by other analytical approaches. Some investigators suggest that these differential low-molecular-weight (LMW) products may not be produced by the diseased cells themselves, but rather they may be generated by epiphenomena within the microenvironment (Diamandis, 2003; Diamandis, 2004; Seibert et al. 2005; Poon, 2007). However, recent reports (Lowenthal et al. 2005; Lopez et al. 2007) indicate that the LMW fragments constitute unique disease-specific protein fragment isoforms that appear to emanate from low abundant tumor cell input. As we transition from patterns of unknown analytes to fingerprints of multiplexed known markers, there is agreement of the importance of quality control, quality assurance,

and the development of high operating standards in order to minimize potential bias that can result from sample collection, handling and processing (Liotta and Petricoin, 2008).

Protein Microarrays

The success of DNA microarrays has encouraged scientists to invent a similar technology for proteins, hence termed protein microarrays. Different types of protein microarrays have been introduced and can be categorized according to their end-point purpose. Similar to SELDI, protein microarrays can aid in search for single biomarkers, a group of biomarkers, or a proteome profile, depending on the type of microarray. In general, they can be categorized into three groups: expression-based, function-based, and interaction-based microarrays. Expression-based protein microarrays are more common and better developed. It aims to investigate protein expression within a sample. Expression-based microarrays can also be of two types: forward phase and reverse phase microarrays. Forward phase microarrays entail spotting thousands of bait molecules on a glass or membrane-coated slide. Each spot would then represent a specific bait for a single protein. Usually the bait molecule is an antibody (Haab, 2001), although other capture molecules such as aptamers (small DNA or RNA molecules) or phage lysates have been reported (Choi et al. 2005). By incubating a sample containing mixed populations of proteins onto the spotted slides, protein molecules would bind specifically to the corresponding bait molecule. Captured proteins can then be detected by directly labeling the proteins before applying them onto the slide. This direct labeling method has been utilized in identifying biomarkers of prostate cancer (Miller et al. 2003) and radiation-regulated proteins (Sreekumar et al. 2001). With the direct labeling method, it is possible to perform comparative expression analysis of two or more samples with proteins in each sample labeled with a distinct tag (Haab, 2005).

Otherwise, an indirect labeling method, also known as sandwich immunoassay, can be used where bound proteins are targeted by a second bait such as a different antibody that targets a different domain. This method is limited to analyzing the expression of proteins within a single sample. Although the requirement for two independent bait molecules may limit multiplexing, this method can

be more specific and sensitive than the direct labeling approach since two bait molecules are required to target the same proteins rather than one molecule. The sandwich immunoassay has been illustrated in measuring the expression the epidermal growth factor receptor and ERB2 and monitoring EGF-dependent phosphorylation in human tumor cells (Nielsen et al. 2003).

Forward-phase protein microarrays are in contrast to another design of expression microarray technology termed “reverse-phase protein microarrays” (Paweletz et al. 2001). This technology involves spotting the analytes (i.e. protein extracts) rather than bait molecules, with each spot representing a single test sample. Since proteins are expressed at a wide range of orders of magnitude, lysates can be spotted at different dilutions providing an internal standard curve and an opportunity for quantitative measurement. It is worthy to mention that spotted lysates can be obtained from microdissected cells from tissues allowing for studies of pure cell populations. For example, differential protein expression in microdissected prostate cancer cells has been compared to that in patient-matched normal and premalignant cells from the same tissue samples (Paweletz et al. 2001). In addition to expression pattern, signal transduction circuitry can be studied using reverse-phase protein microarrays (Paweletz et al. 2001; Nishizuka et al. 2003; Petricoin and Liotta, 2003). Two reports have shown that the activation of Akt pathway is responsible for cell viability in both ovarian and prostate cancers as cancer cells progress from premalignancy to malignancy (Paweletz et al. 2001; Wulfkühle et al. 2003). Whereas with forward-phase proteins microarrays, one can analyze multiple proteins in a few samples per one microarray, reverse-phase protein microarrays enable investigation of one or various proteins in multiple samples per one microarray.

Both forward-phase and reverse-phase protein microarrays are hampered, though, by the availability of a specific bait molecule. A huge endeavor of the HUPO Antibody Initiative has been undertaken by the scientists of the Swedish Royal Institute of Technology led by Mattihas Uhlén to generate, validate, catalog, and annotate antibodies that target human proteins with high specificity and low cross-reactivity to other proteins (Uhlén et al. 2005). The main objective of this initiative is to create a protein atlas for localized expression of proteins in human tissues. Recently, the National

Institutes of Health have established a similar initiative termed Protein Capture Tools that emphasizes on monoclonal antibodies (Haab et al. 2006; Hober and Uhlén, 2008). The European Union has also funded another program, Proteome Binder, with special focus not only antibodies and antibody related reagents like single chain antibodies (scFv), but also on other affinity reagents such as nucleic acid aptamers, protein scaffolds, peptides and chemical entities (Taussig et al. 2007).

The second type of protein microarrays, functional microarrays, aims to assign function to proteins. This can be achieved by conducting a micro-scale enzyme assay whereby product formation can be measured. The enzymatic activities of purified 119 yeast proteins predicted to be tyrosine kinases have been investigated using a functional microarray where kinase substrates are arrayed on a solid surface (Zhu H et al. 2000). Purified proteins are then added individually to the microarrays in the presence of $^{32}\text{P}\gamma\text{-ATP}$. The ability of 27 of the 119 proteins to phosphorylate certain substrates has confirmed their kinase activity.

Interaction protein microarrays can play an important role in determining protein function as well in therapeutics. Interaction arrays can be designed to investigate interaction of certain proteins to various types of molecules including other proteins, peptides, nucleic acids, lipids, carbohydrates, and small molecules (MacBeath and Schreiber, 2000; Ge et al. 2000; Iyer et al. 2001; Linnell et al. 2001; Zhu et al. 2001). For example, by attaching over 90% of yeast proteins onto a microarray, Zhu and colleagues were able to identify protein-phospholipid interactions as well as new calmodulin-interacting proteins (Zhu et al. 2001). In addition, effects of DNA mutations and polymorphism on DNA-protein binding have been studied using interaction protein microarrays (Boutell et al. 2004). Both functional arrays and interaction arrays can also prove invaluable in drug discovery where binding of small ligands to protein targets can be multiplexed, in addition to analyzing the effect of different small molecules on enzyme activity.

Predictive Bioinformatics in Proteomics

As mentioned earlier, bioinformatics is expected to play a major role in analyzing proteomics data. However, it is still a long shot to understand what

these data mean and how they can be useful. A major reason for functional deficiency is the existence of large number of unidentified protein sequences in protein databases accounting for almost 60% in one report (Ahram and Springer, 2004). A branch of bioinformatics that can play an informative role in proteomics is based on predicative computational tools. Predictive bioinformatics can overcome the technical limitations of proteomics by contributing to the annotation of proteins and determination of their function and structure. Several tools have been reported by which protein localization, function and structure can be examined theoretically before moving into the experimental arena. Identification of homologous regions of proteins can lead to predicting protein function and localization. In addition, protein localization can be predicted based on presence of specific amino acid sequence. Many of these tools are offered in the SWISS-PROT and some of them are reviewed elsewhere (Emanuelsson and von Heijne, 2001).

Understanding protein topology is critical for determining protein structure and function and, hence, developing novel therapeutics. Although membrane proteins are thought to constitute 20–30% of annotated genomes (Steven and Arkin, 2000; Wallin and von Heijne, 1998; Krogh et al. 2001), the 3D structures of only 1% of these proteins have been determined (Melen et al. 2003). A class of proteins of special interest is membrane proteins, in particular plasma membrane proteins. A significant role of these proteins is that they constitute more than 45% of current drug targets (Drews, 2000) with 25%–30% of drugs targeting G-protein coupled receptors (Hopkins and Groom, 2002). Thus, construction of computational tools that predict protein topology is imperative for large-scale proteomics studies. These tools operate by predicting the presence of transmembrane segments. However, a major pitfall of all methods is the erroneous prediction of the N-terminal signal peptides as transmembrane segments as a result of their hydrophobic nature (Ahram and Springer, 2004).

In a recent study, a human proteome database was analyzed using five predictive computational methods in search for membrane proteins (Ahram et al. 2006). These five methods are commonly used and based on different computational approaches. In order to eliminate false-positive predictions, a sixth method, SignalP, which discriminates between

signal peptides and transmembrane segments, have also been utilized. Based on these analyses, the ratio of human proteins with transmembrane segments is estimated to fall between 15% and 39% with a consensus of 13%. Such a broad range of prediction depends on the selectivity of the individual method in predicting integral membrane proteins. These methods can play a critical role in determining protein structure and, hence, identifying suitable drug targets in humans.

Another major effort was conducted in search for novel secreted proteins using both biological approach and a computational strategy (Clark et al. 2003). The latter was based on analysis of genomic and expressed sequence tags (ESTs). This study resulted in the isolation of over 1000 cDNA clones with 25% of them representing novel genes. An important conclusion of this study is the significance of applying multi-directed approaches in the identification of proteins.

Conclusion

Exciting advances have been posted in proteomics. These advances have mainly been observed in the technical field with the development and improvement of technologies such as mass spectrometry, SELDI, and protein microarrays. However, these technologies are still limited in many areas including specificity and sensitivity of detection. In addition, data mining can also see significant attention for better accuracy of protein analysis. It is the integration of these technologies as well as the development of bioinformatics tools that can speed up the discovery of protein targets for therapeutics and lead to more accurate and safer drugs.

References

Adam, B.L., Qu, Y., Davis, J.W., Ward, M.D., Clements, M.A., Cazares, L.H., Semmes, O.J., Schellhammer, P.F., Yasui, Y., Feng, Z. and Wright, G.L., Jr. 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.

Ahram, M. and Springer, D.L. 2004. Large-scale proteomics analysis of membrane proteins. *Expert Rev. Proteomics*, 1:293–302.

Ahram, M., Strittmatter, E.F., Monroe, M.E., Adkins, J.N., Hunter, J.C., Miller, J.H. and Springer, D.L. 2005. Identification of shed proteins from Chinese hamster ovary cells: application of statistical confidence using human and mouse protein databases. *Proteomics*, 5:1815–26.

Ahram, M., Zhou, L., Fang, R. and Al-Tawalbeh, G. 2006. Estimation of membrane proteins in the human proteome. *In Silico Bio.*, 6:0036.

Arnaud, M.C., Gazarian, T., Rodriguez, Y.P., Gazarian, K. and Sakanyan, V. 2004. Array assessment of phage-displayed peptide mimics of Human Immunodeficiency Virus type 1 gp41 immunodominant epitope: binding to antibodies of infected individuals. *Proteomics*, 4:1959–64.

Austen, B.M., Frears, E.R. and Davies, H. 2000. The use of seldi proteinchip arrays to monitor production of Alzheimer's betaamyloid in transfected cells. *J. Pept Sci.*, 6:459–69.

Belov, M.E., Gorshkov, M.V., Udseth, H.R., Anderson, G.A. and Smith, R.D. 2000. Zeptomole-sensitivity electrospray ionization–Fourier transform ion cyclotron resonance mass spectrometry of proteins. *Anal Chem.*, 72:2271–9.

Betz, U.A. 2005. How many genomics targets can a portfolio afford? *Drug Discov Today*, 10:1057–63.

Boutell, J.M., Hart, D.J., Godber, B.L., Kozlowski, R.Z. and Blackburn, J.M. 2004. Functional protein microarrays for parallel characterisation of p53 mutants. *Proteomics*, 4:1950–58.

Carrette, O., Demalte, I., Scherl, A., Yalkinoglu, O., Corthals, G., Burkhard, P., Hochstrasser, D.F. and Sanchez, J.C. 2003. A panel of cerebrospinal fluid potential biomarkers for the diagnosis of Alzheimer's disease. *Proteomics*, 3:1486–94.

Choi, J.K., Yu, U., Yoo, O.J. and Kim, S. 2005. Differential coexpression analysis using microarray data and its application to human cancer. *Bioinformatics*, 21:4348–55.

Clark, H.F., Gurney, A.L., Abaya, E., Baker, K., Baldwin, D., Brush, J., Chen, J., Chow, B., Chui, C., Crowley, C., Currell, B., Deuel, B., Dowd, P., Eaton, D., Foster, J., Grimaldi, C., Gu, Q., Hass, P.E., Heldens, S., Huang, A., Kim, H.S., Klimowski, L., Jin, Y., Johnson, S., Lee, J., Lewis, L., Liao, D., Mark, M., Robbie, E., Sanchez, C., Schoenfeld, J., Seshagiri, S., Simmons, L., Singh, J., Smith, V., Stinson, J., Vagts, A., Vandlen, R., Watanabe, C., Wieand, D., Woods, K., Xie, M.H., Yansura, D., Yi, S., Yu, G., Yuan, J., Zhang, M., Zhang, Z., Goddard, A., Wood, W.L., Godowski, P. and Gray, A. 2003. The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment. *Genome Res.*, 13:2265–70.

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

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

Diamond, D.L., Zhang, Y., Gaiger, A., Smithgall, M., Vedvick, T.S. and Carter, D. 2003. Use of ProteinChip array surface enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) to identify thymosin beta-4, a differentially secreted protein from lymphoblastoid cell lines. *J. Am. Soc. Mass Spectrom.*, 14:760–5.

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

Drews, J. 2000. Drug discovery: a historical perspective. *Science*, 287:1960–4.

Ebert, M.P., Meuer, J., Wiemer, J.C., Schulz, H.U., Reymond, M.A., Traugott, U., Malfertheiner, P. and Rocken, C. 2004. Identification of gastric cancer patients by serum protein profiling. *J. Proteome Res.*, 3:1261–6.

Emanuelsson, O. and von Heijne, G. 2001. Prediction of organellar targeting signals. *Biochim. Biophys. Acta.*, 1541:114–9.

Engwegen, J.Y., Helgason, H.H., Cats, A., Harris, N., Bonfrer, J.M., Schellens, J.H. and Beijnen, J.H. 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.

Ge, H.U.P.A. 2000. A universal protein array system for quantitative detection of protein–protein, protein–DNA, protein–RNA and protein–ligand interactions. *Nucl Acids Res.*, 28:3.

Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y. and Aebersold, R. 2000. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci., U.S.A.*, 97:9390–5.

Haab, B.B. 2001. Advances in protein microarray technology for protein expression and interaction profiling. *Curr. Opin. Drug Discov. Devel.*, 4:116–23.

Haab, B.B. 2005. Antibody arrays in cancer research. *Mol. Cell. Proteomics*, 4:377–83.

- Haab, B.B. 2006. Applications of antibody array platforms. *Curr. Opin. Biotechnol.*, 17:415–21.
- Haab, B.B., Paulovich, A.G., Anderson, N.L., Clark, A.M., Downing, G.J., Hermjakob, H., Labaer, J. and Uhlen, M. 2006. A reagent resource to identify proteins and peptides of interest for the cancer community: a workshop report. *Mol. Cell. Proteomics*, 5:1996–2007.
- Hober, S. and Uhlén, M. 2008. Human protein atlas and the use of microarray technologies. *Curr. Opin. Biotechnol.*, 19:30–5.
- Hirsch, J., Hansen, K.C., Burlingame, A.L. and Matthay, M.A. 2004. Proteomics: current techniques and potential applications to lung disease. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, 287:L1–23.
- Hopkins, A.L. and Groom, C.R. 2002. The druggable genome. *Nat. Rev. Drug Discov.*, 1:727–30.
- Irani, D.N., Anderson, C., Gundry, R., Cotter, R., Moore, S., Kerr, D.A., McArthur, J.C., Sacktor, N., Pardo, C.A., Jones, M., Calabresi, P.A. and Nath, A. 2006. Cleavage of cystatin C in the cerebrospinal fluid of patients with multiple sclerosis. *Ann. Neurol.*, 59:237–47.
- Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M. and Brown, P.O. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, 409:533–38.
- Johnson, R.S., Davis, M.T., Taylor, J.A. and Patterson, S.D. 2005. Informatics for protein identification by mass spectrometry. *Methods*, 35:223–36.
- Kapp, E.A., Schutz, F., Connolly, L.M., Chakel, J.A., Meza, J.E., Miller, C.A., Fenyo, D., Eng, J.K., Adkins, J.N., Omenn, G.S. and Simpson, R.J. 2005. An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis. *Proteomics*, 5:3475–90.
- Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, 305:567–80.
- Lehrer, S., Roboz, J., Ding, H., Zhao, S., Diamond, E.J., Holland, J.F., Stone, N.N., Droller, M.J. and Stock, R.G. 2003. Putative protein markers in the sera of men with prostatic neoplasms. *BJU Int.*, 92:223–25.
- Linnell, J., Groeger, G. and Hassan, A.B. 2001. Real time kinetics of insulin-like growth factor II (IGF-II) interaction with the IGF-II/mannose 6-phosphate receptor: the effects of domain 13 and pH. *J. Biol. Chem.*, 276:23986–91.
- Liotta, L.A. and Petricoin, E.F. 2008. Putting the “bio” back into biomarkers: orienting proteomic discovery toward biology and away from the measurement platform. *Clin. Chem.*, 54:3–5.
- Lipton, M.S., Pasa-Tolic, L., Anderson, G.A., Anderson, D.J., Auberry, D.L., Battista, J.R., Daly, M.J., Fredrickson, J., Hixson, K.K., Kostandarithes, H., Masselon, C., Markillie, L.M., Moore, R.J., Romine, M.F., Shen, Y., Strittmatter, E., Tolic, N., Udseth, H.R., Venkateswaran, A., Wong, K.K., Zhao, R. and Smith, R.D. 2002. Global analysis of the *Deinococcus radiodurans* proteome by using accurate mass tags. *Proc. Natl. Acad. Sci., U.S.A.*, 99:11049–54.
- Liu, A.Y., Zhang, H., Sorensen, C.M. and Diamond, D.L. 2005. Analysis of prostate cancer by proteomics using tissue specimens. *J. Urol.*, 173:73–8.
- Lopez, M.F., Mikulskis, A., Kuzdzal, S., Golenko, E., Petricoin, E.F., Liotta, L.A., Patton, W.F., Whiteley, G.R., Rosenblatt, K., Gurnani, P., Nandi, A., Neill, S., Cullen, S., O’Gorman, M., Sarracino, D., Lynch, C., Johnson, A., Mckenzie, W. and Fishman, D. 2007. A novel, high-throughput workflow for discovery and identification of serum carrier protein-bound peptide biomarker candidates in ovarian cancer samples. *Clin. Chem.*, 53:1067–74.
- Lowenthal, M.S., Mehta, A.I., Frogale, K., Bandle, R.W., Araujo, R.P., Hood, B.L., Veenstra, T.D., Conrads, T.P., Goldsmith, P., Fishman, D., Petricoin, E.F. and Liotta, L.A. 2005. Analysis of albumin-associated peptides and proteins from ovarian cancer patients. *Clin. Chem.*, 51:1933–45.
- MacBeath, G. and Schreiber, S.L. 2000. Printing proteins as microarrays for high-throughput function determination. *Science*, 289:1760–3.
- Melen, K., Krogh, A. and von Heijne, G. 2003. Reliability measures for membrane protein topology prediction algorithms. *J. Mol. Biol.*, 327:735–44.
- Miller, J.C., Zhou, H., Kwekel, J., Cavallo, R., Burke, J., Butler, E.B., The, B.S. and Haab, B.B. 2003. Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers. *Proteomics*, 3:56–63.
- Nielsen, U.B., Cardone, M.H., Sinskey, A.J., MacBeath, G. and Sorger, P.K. 2003. Profiling receptor tyrosine kinase activation by using Ab microarrays. *Proc. Natl. Acad. Sci., U.S.A.*, 100:9330–5.
- Nishizuka, S., Charboneau, L., Young, L., Major, S., Reinhold, W.C., Waltham, M., Kouros-Mehr, H., Bussey, K.J., Lee, J.K., Espina, V., Munson, P.J., Petricoin, E. 3rd, Liotta, L.A. and Weinstein, J.N. 2003. Proteomics 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.
- Paweletz, C.P., Charboneau, L., Bichsel, V.E., Simone, N.L., Chen, T., Gillespie, J.W., Emmert-Buck, M.R., Roth, M.J., Petricoin III, E.F. and Liotta, L.A. 2001. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene*, 20:1981–9.
- Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C. and Liotta, L.A. 2002. Use of proteomics patterns in serum to identify ovarian cancer. *Lancet*, 359:572–7.
- Petricoin, E.F. and Liotta, L.A. 2003. Clinical applications of proteomics. *J. Nutr.*, 133:2476S–4S.
- Petricoin, E.F., Ornstein, D.K., Paweletz, C.P., Ardekani, A., Hackett, P.S., Hitt, B.A., Velasco, A., Trucco, C., Wiegand, L., Wood, K., Simone, C.B., Levine, P.J., Linehan, W.M., Emmert-Buck, M.R., Steinberg, S.M., Kohn, E.C. and Liotta, L.A. 2002. Serum proteomics patterns for detection of prostate cancer. *J. Natl. Cancer Inst.*, 94:1576–8.
- Petritis, K., Kangas, L.J., Ferguson, P.L., Anderson, G.A., Pasa-Tolic, L., Lipton, M.S., Auberry, K.J., Strittmatter, E.F., Shen, Y., Zhao, R. and Smith, R.D. 2003. Use of artificial neural networks for the accurate prediction of peptide liquid chromatography elution times in proteome analyses. *Anal. Chem.*, 75:1039–48.
- Poon, T.C. 2007. Opportunities and limitations of SELDI-TOF-MS in biomedical research: practical advices. *Expert RevProteomics*, 4:51–65.
- Resing, K.A. and Ahn, N.G. 2005. Proteomics strategies for protein identification. *FEBS Lett.*, 579:885–9.
- Resing, K.A., Meyer-Arendt, K., Mendoza, A.M., Aveline-Wolf, L.D., Jonscher, K.R., Pierce, K.G., Old, W.M., Cheung, H.T., Russell, S., Wattawa, J.L., Goehle, G.R., Knight, R.D. and Ahn, N.G. 2004. Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. *Anal. Chem.*, 76:3556–68.
- Ricolleau, G., Charbonnel, C., Lode, L., Loussouarn, D., Joalland, M.P., Bogumil, R., Jourdain, S., Minvielle, S., Campone, M., Deporte-Fety, R., Campion, L. and Jezequel, P. 2006. Surface-enhanced laser desorption/ionization time of flight mass spectrometry protein profiling identifies ubiquitin and ferritin light chain as prognostic biomarkers in node-negative breast cancer tumors. *Proteomics*, 6:1963–75.
- Sanchez, J.C., Guillaume, E., Lescuyer, P., Allard, L., Carrette, O., Scherl, A., Burgess, J., Corthals, G.L., Burkhard, P.R. and Hochstrasser, D.F.J. 2004. Cystatin C as a potential cerebrospinal fluid marker for the diagnosis of Creutzfeldt-Jakob disease. *Proteomics*, 4:2229–33.
- Seibert, V., Ebert, M.P. and Buschmann, T. 2005. Advances in clinical cancer proteomics: SELDI-ToF-mass spectrometry and biomarker discovery. *Brief Funct Genomic Proteomic*, 4:16–26.
- Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H. and Mann, M. 1996. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci., U.S.A.*, 93:14440–5.
- Sreekumar, A., Nyati, M.K., Varambally, S., Barrette, T.R., Ghosh, D., Lawrence, T.S. and Chinnaiyan, A.M. 2001. Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Res.*, 61:7585–93.

- Stevens, T.J. and Arkin, I.T. 2000. Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins*, 39:417–20.
- Strittmatter, E.F., Kangas, L.J., Petritis, K., Mottaz, H.M., Anderson, G.A., Shen, Y., Jacobs, J.M., Camp, D.G. 2nd and Smith, R.D. 2004. Application of peptide LC retention time information in a discriminant function for peptide identification by tandem mass spectrometry. *J. Proteome Res.*, 3:760–9.
- Taussig, M.J., Stoevesandt, O., Borrebaeck, C.A., Bradbury, A.R., Cahill, D., Cambillau, C., de Daruvar, A., Dübel, S., Eichler, J., Frank, R., Gibson, T.J., Gloriam, D., Gold, L., Herberg, F.W., Hermjakob, H., Hoheisel, J.D., Joos, T.O., Kallioniemi, O., Koegl, M., Konthur, Z., Korn, B., Kremmer, E., Krobitsch, S., Landegren, U., van der Maarel, S., McCafferty, J., Muylderms, S., Nygren, P.A., Palcy, S., Plückerthun, A., Polic, B., Przybylski, M., Saviranta, P., Sawyer, A., Sherman, D.J., Skerra, A., Templin, M., Ueffing, M., Uhlén, M. 2007. ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nat. Methods*, 4:13–7.
- Uhlén, M., Björling, E., Agaton, C., et al. 2005. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics*, 4:1920–32.
- Vlahou, A., Schellhammer, P.F., Mendrinos, S., Patel, K., Kondylis, F.I., Gong, L., Nasim, S. and Wright, G.L., Jr. 2001. Development of a novel proteomics approach for the detection of transitional cell carcinoma of the bladder in urine. *Am. J. Pathol.*, 158:1491–502.
- Wallin, E. and von Heijne, G. 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.*, 7:1029–38.
- Wright, G.L., Jr. Beckett, M.L., Newhall, K.R., Adam, B.L., Cazares, L.H., Cartwright, S.L., Xiao, Z., Gong, L. and Schellhammer, P.F. 2000. Identification of a superimmunoglobulin gene family member over-expressed in benign prostatic hyperplasia. *Prostate*, 42:230–8.
- Wulfkühle, J.D., Aquino, J.A., Calvert, V.S., Fishman, D.A., Coukos, G., Liotta, L.A. and Petricoin, E.F. 2003. Signal pathway profiling of ovarian cancer from human tissue specimens using reverse-phase protein microarrays. *Proteomics*, 3:2085–90.
- Xiao, Z., Adam, B.L., Cazares, L.H., Clements, M.A., Davis, J.W., Schellhammer, P.F., Dalmaso, E.A. and Wright, G.L., Jr. 2001. Quantitation of serum prostate-specific membrane antigen by a novel protein biochip microarrays. *Proteomics*, 3:2085–90.
- Yang, E.C., Guo, J., Diehl, G., DeSouza, L., Rodrigues, M.J., Romaschin, A.D., Colgan, T.J. and Siu, K.W. 2004. Protein expression profiling of endometrial malignancies reveals a new tumor marker: chaperonin 10. *Proteome Res.*, 3:636–43.
- Yip, T.T. and Hutchens, T.W. 1992. Mapping and sequence-specific identification of phosphopeptides in unfractionated protein digest mixtures by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *FEBS Lett.*, 308:149–53.
- Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R.A., Gerstein, M. and Snyder, M. 2001. Global analysis of protein activities using proteome chips. *Science*, 293:2101–5.
- Zhu, H., Klemic, J.F., Chang, S., Bertone, P., Casamayor, A., Klemic, K.G., Smith, D., Gerstein, M., Reed, M.A. and Snyder, M. 2000. Analysis of yeast protein kinases using protein chips. *Nat. Genet.*, 26:283–9.