For reprint orders, please contact reprints@future-drugs.com

Cancer diagnosis using proteomic patterns

Thomas P Conrads, Ming Zhou, Emmanuel F Petricoin III, Lance Liotta and Timothy D Veenstra[†]

The advent of proteomics has brought with it the hope of discovering novel biomarkers that can be used to diagnose diseases, predict susceptibility and monitor progression. Much of this effort has focused upon the mass spectral identification of the thousands of proteins that populate complex biosystems such as serum and tissues. A revolutionary approach in proteomic pattern analysis has emerged as an effective method for the early diagnosis of diseases such as ovarian cancer. Proteomic pattern analysis relies on the pattern of proteins observed and does not rely on the identification of a traceable biomarker. Hundreds of clinical samples per day can be analyzed utilizing this technology, which has the potential to be a novel, highly sensitive diagnostic tool for the early detection of cancer.

Expert Rev. Mol. Diagn. 3(4), 411–420 (2003)

CONTENTS

[†]*Author for correspondence*
Biomedical Proteomics Program, SAIC-Frederick, Inc., National Cancer Institute at Frederick,PO Box B, Frederick, MD 21702, USA
Tel.: +1 301 846 7286
Fax: +1 301 846 6037
veenstra@ncifcrf.gov

KEYWORDS:
bioinformatics, biomarkers, cancer, early diagnosis, mass spectrometry, proteomic patterns

The term proteomics means different things to different people. Originally, proteomics was coined to describe the large-scale, high-throughput separation and subsequent identification of proteins resolved by 2D polyacrimide gel electrophoresis (2DE) [1]. The field of proteomics has since evolved to include almost any type of technology that focuses upon the wide-scale analysis of proteins. These technologies range from those designed to study a single protein (i.e., mapping of sites of post-translational modifications [PTM]) to those for the analysis of hundreds to thousands of proteins in a single experiment (e.g., protein arrays or isotope-coded affinity tags) [2–4]. In essence, the term proteomics has replaced the use of the phrase protein science. Regardless of the terminology or the scope of the analysis, one of the common requirements of a vast majority of proteomic studies has been the identification of the protein(s) of interest.

Protein identification is central to most proteomic studies [5]. For example, probably the most well known and widely used proteomic technology is the characterization of changes in protein expression between two different samples through comparative 2DE [6]. In such studies, two proteomic samples are resolved and visualized on two separate 2DE gels. Protein spots that are more or less intense on one gel compared with the other are excised and the differentially expressed proteins identified using mass spectrometry (MS). In addition, studies such as phosphorylation mapping would be incomplete without identifying the modified protein, or even better, the specific amino acid that has been modified.

Proteomics as a diagnostic tool

While having an enormous impact in almost every discipline of biomedical science, one of the major focuses is to use the high-throughput capabilities of proteomics in the discovery of novel disease biomarkers [7]. While a biomarker can be defined as any laboratory measurement or physical sign used as a substitute for a clinically meaningful end point that measures directly how a patient feels, functions or survives [8], as applied to proteomics, a biomarker is an identified protein(s) that is unique to a particular disease state. Simply put, the experimental design of a diagnostic proteomic investigation aims to scrutinize clinical samples from healthy and afflicted individuals in a high-throughput manner, allowing for the relative abundance of thousands of proteins from the two histologically distinct samples to be visualized. Proteins that are found to be differentially abundant between the samples are then selected

for identification with the hope that knowledge of their identity will provide the basis for defining a diagnostic biomarker. Unfortunately, this strategy suffers from issues that are in many ways technically overwhelming. Firstly, the proteins being observed in these analyses are generally of high abundance. Therefore, valuable biomarkers expressed at low abundance remain undetected until current analytical technologies become more sensitive. Secondly, discovery of effective biomarkers requires the analysis of hundreds of histologically well-defined samples retrieved from healthy and disease-afflicted individuals. In addition, to be clinically relevant, the biomarker should be present in easily obtainable samples such as serum, plasma or urine. Even ignoring the difficulties in analyzing serum and plasma via MS-based proteomic methods, the natural variability in biofluids obtained from different patients makes the recognition of a single, consistent biomarker in the background of a dynamic proteome extremely challenging. Thirdly, it may be that the presence of a single, definitive biomarker for a particular histological condition, such as human chorionic gonadotropin for pregnancy, is the exception rather than the rule [9,10]. Indeed, many clinical tests that rely on single diagnostic biomarkers, such as cancer-antigen (CA)-125 for ovarian cancer [11] and prostate specific antigen (PSA) for prostate cancer [12], possess positive predictive values (PPV) that are generally quite low.

Biomarker discovery without protein identification

There exists a sobering reality of the lack of success in the discovery of novel diagnostic biomarkers despite the considerable intellectual and financial resources currently invested in the use of conventional proteomic technologies. It is likely that a vast majority of disease states are not the result of a single recognizable change in the abundance or function of a protein. Considering the complexity of an individual cell and the aberrations caused by such disease states as cancer, a vast number of differences between the protein character of healthy and diseased tissues should be observable. Why then has the discovery of disease-specific biomarkers been so elusive? Obviously one of the main reasons is that for a diagnostic marker to be clinically relevant it should be assayed from a sample that can be relatively noninvasively obtained in sufficient quantities from patients. For this reason, the search for biomarkers using proteomic methods largely focus upon plasma and serum. While serum constantly perfuses tissues, hence potentially endowing an archive of histological information, this information is comprised not only of the expected circulatory proteins in serum, such as immunoglobulins, but also of peptides and proteins that are secreted into the blood and species shed from diseased, dying or dead cells present throughout the body [13]. Therefore, the background matrix of biofluids such as serum represents a complex milieu in which to find unique disease-specific biomarkers that are most likely of extremely low abundance. The intrinsic person-to-person variability of the content of biofluids also hampers the identification of a disease-specific biomarker [14]. The identification of a biomarker relies on the comparison of, for example, serum samples from healthy and disease-stricken individuals. The comparison of two distinct serum samples is incredibly laborious using conventional proteomic technologies and the comparison of the hundreds if not thousands needed to validate a biomarker is not routinely possible. More to the point, in the comparison of just two serum samples, a multitude of changes in protein abundance are observed due simply to differences such as age, gender or lifestyle, making the assumption that a particular difference is a result of a specific disease state tenuous at best.

Proteomic pattern technology

A revolutionary proteomic technology has recently been developed that uses the pattern of proteins observed within a clinical sample as a diagnostic fingerprint and does not rely on the identification of the proteins detected. The technology to acquire these so-called proteomic patterns is quite simple, as illustrated in FIGURE 1. In its current state, surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) is the technology used to acquire the proteomic patterns to be used in the diagnostic setting [15,16]. The principle of SELDI-TOF is very simple; proteins of interest are captured, by adsorption, partition, electrostatic interaction or affinity chromatography on a stationary-phase and immobilized in an array format on a chip surface. One of the benefits of this process is that raw biofluids, such as urine, serum and plasma, can be directly applied to the array surface. After a series of binding and washing steps, a matrix is applied to the array surfaces. The species bound to these surfaces can be ionized by matrix-assisted laser desorption/ionization (MALDI) and their mass-to-charge (*m/z*) ratios measured by TOF MS. The result is simply a mass spectrum of the species that bound to and subsequently desorbed from the array surface. While the inherent simplicity of the technology has contributed to the enthusiasm generated for this approach, the implementation of sophisticated bioinformatic tools have enabled the use of SELDI-TOF MS as a potentially revolutionary diagnostic tool.

Application of proteomic patterns for disease diagnosis

The seminal study describing the use proteomic patterns to diagnose ovarian cancer was published in *The Lancet* by Petricoin and coworkers of the US Food and Drug Administration (FDA) and Liotta and coworkers of the National Cancer Institute (NCI) [17]. The aggressive nature of ovarian cancer, the fifth most common cause of cancer-related death in women, makes it a prime example of a disease whose 5-year survival rate would dramatically increase if a more effective means of early (or Stage I) detection could be discovered [17,18]. Unfortunately, almost 80% of women with common epithelial ovarian cancer are not diagnosed until the disease has spread to the upper abdomen (Stage III) or beyond (Stage IV) [19,20]. The 5-year survival rate for these women is only 15–20%, whereas the 5-year survival rate for ovarian cancer at Stage I approaches 95% with surgical intervention. In this study, the proteomic patterns of serum samples from several patients with ovarian cancer were compared with those from control patients. Visual

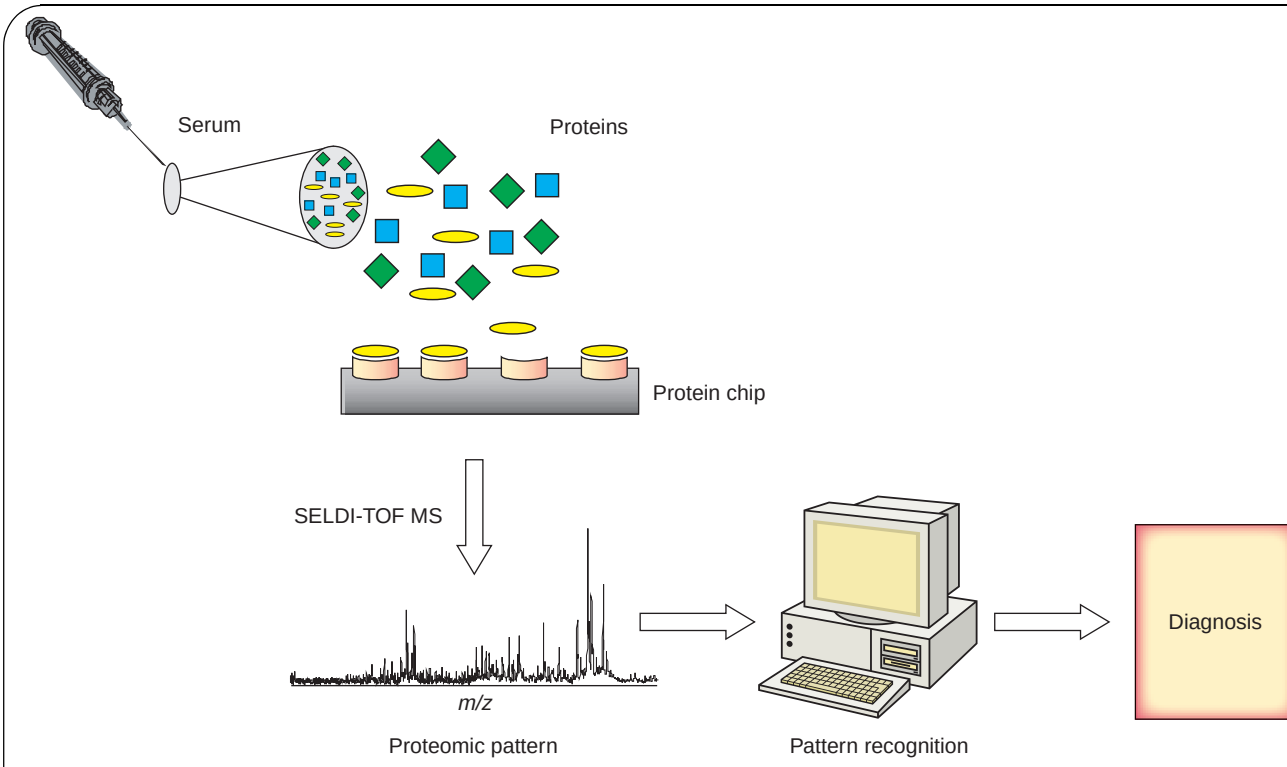

Figure 1. Disease diagnostics using proteomic patterns. The sample drawn from the patient is applied to a protein chip which is made up of a specific chromatographic surface. After several washing steps and the application of an energy-absorbing molecule, the species that are retained on the surface of the chip are analyzed via mass spectrometry. The pattern of peaks within the spectrum is analyzed using sophisticated bioinformatic software to diagnose the source of the biological sample.
m/z: Mass to charge ratio; SELDI-TOF MS: Surface-enhanced laser desorption/ionization time-of-flight mass spectrometry.

inspection of the mass spectra did not reveal mass spectral features unique to the ovarian cancer serum samples. The ability to discriminate patterns generated from serum acquired from healthy individuals and those patients affected with ovarian cancer was only accomplished through application of an artificial intelligence program able to decipher diagnostic patterns within the profiles.

The artificial intelligence program used in this study combined elements of a genetic algorithm with cluster analysis [21–23]. The input files for the analysis were comprised of the *m/z* values on the x-axis along with their corresponding amplitudes on the y-axis. The analysis was divided into two phases, a pattern discovery phase and a pattern matching phase, as illustrated in FIGURE 2. In the pattern discovery phase, a set of mass spectra of serum from both healthy and ovarian cancer-affected individuals (referred to as the training set) was analyzed to identify a subset of *m/z* values and their related amplitudes that are able to completely segregate the data acquired using serum samples from patients with ovarian cancer and unaffected individuals. In the pattern discovery phase, the source of the serum (from healthy or ovarian cancer-affected individuals) was known and is included as part of the data that is provided to the algorithm. The bioinformatic searching process began with hundreds of arbitrary choices of small sets (5–20) of the exact *m/z* values selected along the x-axis of the mass spectra. The diagnostic pattern was formed by plotting the combined y-axis amplitudes of the candidate set of the key *m/z* values in N-dimensional space, where N is equal to the number of *m/z* values found within the training set of spectra. The pattern formed by the relative amplitudes of the spectrum data for this set of chosen *m/z* values was rated for its ability to distinguish the serum mass spectra acquired from the healthy and cancer-affected individuals. Since the aim was to identify the pattern that provides the optimal segregation, the frequency values within the highest rated sets were reshuffled to form new subset candidates and the resultant defined amplitude values were rated iteratively until the set that fully discriminates the preliminary sample sets was revealed.

Once the algorithm recognizes key *m/z* values, the model was tested using a set of masked test spectra in which the optimal pattern recognized in the first phase is tested for its diagnostic capabilities. As opposed to the pattern discovery phase which uses all of the *m/z* values within the entire spectral data set, in the pattern matching phase, only the key subset of the *m/z* values identified in the pattern discovery phase was used to classify the unknown samples as being from healthy or cancer-affected individuals. The pattern formed by the relative amplitudes of the key *m/z* values in each unknown was then matched to the optimum pattern defined in the pattern-matching phase. Each unknown sample was classified based upon the cluster(s) that its feature set populates as an unaffected or cancer-affected individual, or generated a new cluster if it is found not to

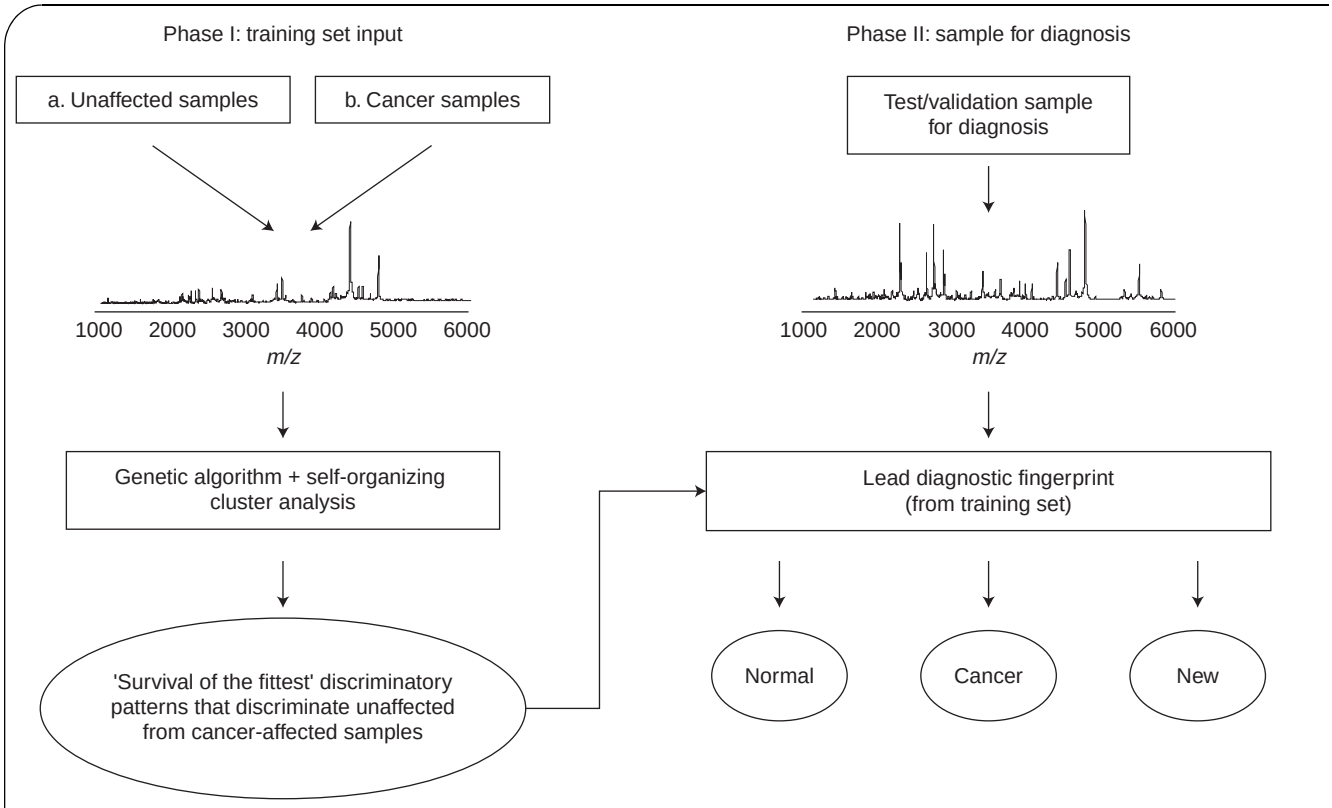

Figure 2. Bioinformatic analysis of proteomic pattern spectra for the determination of the discriminatory patterns in the training and diagnostic (testing and blind validation) phase.
m/z: Mass to charge ratio.

match any of the patterns defined in the pattern discovery phase. If the sample generates a new cluster then the point in N-space of the unknown sample is outside the defined likeness boundaries of the ovarian cancer and unaffected clusters.

After generating the diagnostic model, the diagnostic feature sets defined in training were utilized in a series of test samples in which the source of the serum was blinded. The diagnostic feature set defined in training was able to correctly diagnose the samples as being acquired from either control patients or those suffering from ovarian cancer with a sensitivity of 100% and a specificity of 95%, with an overall PPV of 94% [17]. The success in correctly diagnosing Stage I ovarian cancer suggested that proteomic patterns generated from biofluids may provide a useful indicator of the early onset of a particular disease state.

Improvements in instrumental analysis

Since this original study, several laboratories have acquired and analyzed serum proteomic pattern analysis for the diagnosis of breast [24] and prostate [25–27] cancer. All of these studies used the same analytical platform combined with different sample preparation methods and bioinformatic algorithms. The analytical platform used is the ProteinChip® Biomarker System-II (PBS-II; Ciphergen, CA, USA), a low-resolution TOF MS. Reflective of the success of the ovarian cancer study [17], many of these studies have been able to correctly diagnose serum samples with sensitivities and specificities greater than 90%. The diagnostic accuracy combined with the attributes of the technology (e.g., ease of sample preparation and high-throughput) make proteomic patterns a potentially invaluable screening tool for high-risk populations. Even with this high overall PPV, however, the technology in its present form is not useful as a clinical screening tool for a disease with a low prevalence such as ovarian cancer. While a PPV of 94% as was reported in the ovarian cancer study by Petricoin and colleagues is extremely high, the specificity (95%) of the assay when extrapolated over a large population in which very few patients would actually have ovarian cancer would result in six out of every 100 patients being sent for unnecessary biopsies. This percentage of false-positives would have a tremendous negative impact on the available medical resources. To serve as an effective screening tool, a diagnostic assay screening for ovarian cancer requires a specificity of at least 99.6% [29]. Therefore, while proteomic pattern analysis in its present state represents a useful tool to diagnose cancer, its use as a screening tool for high-risk populations is still limited.

One of the limiting factors in increasing the PPV attributes of proteomic pattern analysis is the PBS-II, which is a simple TOF-MS that is designed to provide for a broad *m/z* range of detection, however, necessarily at the expense of resolution. To assess the increase in diagnostic sensitivity and specificity that would be afforded from higher resolution mass spectra, our laboratory performed a side-by-side comparison of the results obtained analyzing serum samples on a PBS-II and a hybrid

quadrupole (Qq) TOF MS fitted with a SELDI ion source [30]. A schematic comparison of the two instruments is provided in FIGURE 3. The QqTOF can be regarded as a triple Qq MS in which the third quadrupole has been replaced with a reflecting TOF analyzer. This instrument combines the benefits of ion selectivity and tandem MS capabilities of a triple Qq MS with the high mass accuracy and resolution afforded by a reflecting TOF analyzer. The PBS-II, on the other hand, is a relatively simple linear TOF MS. The mass analyzer, though relatively sensitive, provides only low resolution and hence lowmass accuracy data. While time lag focusing is used to increase data resolution and mass accuracy, the achievable mass accuracy is much less than that afforded using more conventional, high-resolution TOF MS instrumentation, such as the QqTOF. An example of a serum sample analyzed using both the PBS-II TOF MS and the QqTOF MS is illustrated in FIGURE 4. While the spectra are qualitatively similar, the resolution obtainable with the QqTOF MS is on the order of 60-fold higher than that obtainable with the PBS-II TOF MS.

To compare the PPV of the results obtained from the two MS platforms, 248 serum samples from either healthy women or those clinically diagnosed with various stages of ovarian cancer were provided from the National Ovarian Cancer Early Detection Program at Northwestern University Hospital (Chicago, Illinois, USA), and processed and analyzed by both instruments. The key to this study is that the identical set of samples were analyzed on the exact same protein chip surface and all experimental variability outside the use of two different instruments was thereby eliminated.

The mass spectra acquired on both the PBS-II and QqTOF MS instruments were analyzed using the ProteomeQuest™ (Correlogic Systems, Inc., MD, USA) bioinformatics tool as illustrated in FIGURE 2. A total of 28 serum samples from unaffected women and 49 women with ovarian cancer were used for the training set. A series of diagnostic models were generated using a variety of different combinations of bioinformatic heuristic parameters. None of these parameters had any affect on the raw MS data, they were simply related to the bioinformatic

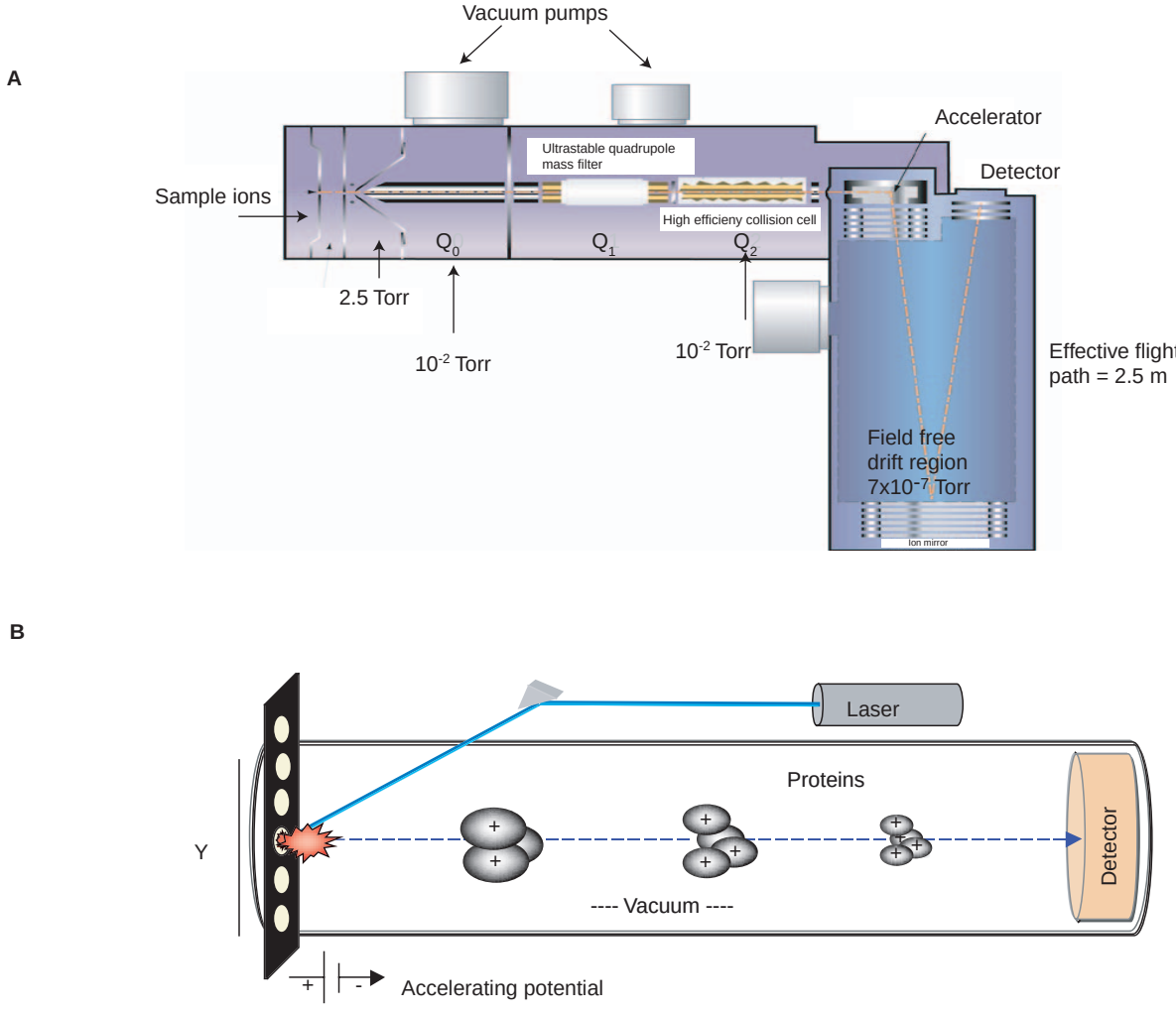

Figure 3. Schematic diagrams comparing the configuration of (A) QqTOF MS and (B) PBS-II TOF MS. The QqTOF combines elements of a triple quadrupole (Qq) MS with a reflecting TOF analyzer, thereby affording higher mass accuracy and resolution than the simple linear TOF of the PBS-II. PBS-II: ProteinChip® Biomarker System-II (Ciphergen, CA, USA); TOF MS: Time-of-flight mass spectrometry.

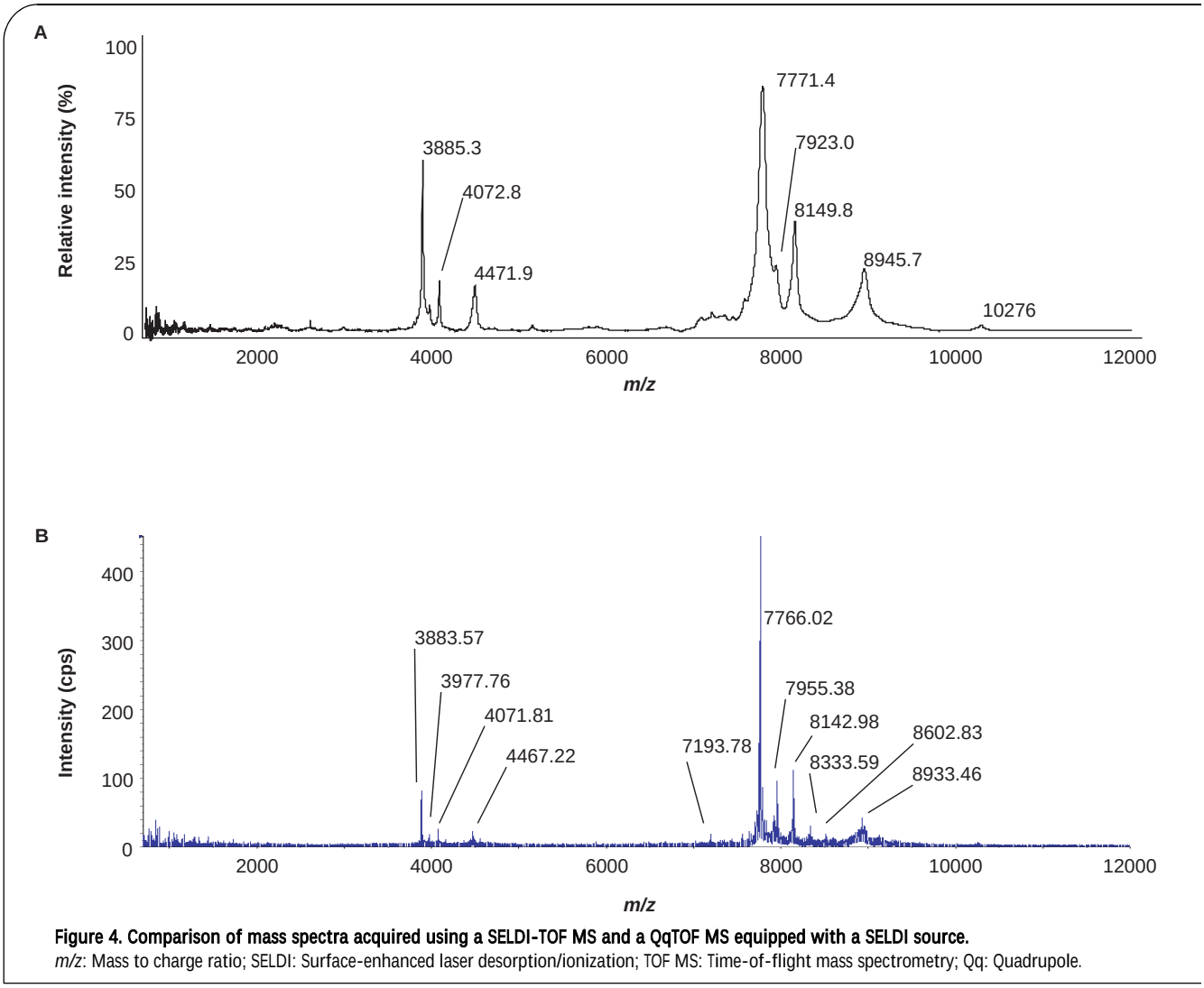

Figure 4. Comparison of mass spectra acquired using a SELDI-TOF MS and a QqTOF MS equipped with a SELDI source.
m/z: Mass to charge ratio; SELDI: Surface-enhanced laser desorption/ionization; TOF MS: Time-of-flight mass spectrometry; Qq: Quadrupole.

process of generating diagnostic models from the raw data and included such factors as the similarity space of likeness for cluster classification, the feature set size of random *m/z* values whose combined intensities comprise each pattern and the learning rate of pattern generation by the genetic algorithm. A total of 108 models were derived and queried with the same set of proteomic pattern spectra in the testing and blind validation phase to assess their sensitivity, specificity and overall PPV.

The models derived from the training sets acquired on both the PBS-II and QqTOF MS were tested using blind serum sample mass spectra obtained from 31 unaffected women and 63 women with ovarian cancer, and further validated using blind serum sample spectra obtained from 37 unaffected women and 40 women with ovarian cancer. The diagnostic models generated from mass spectra acquired using the higher resolution QqTOF MS were statistically superior not only in testing (sensitivity, $p_2 < 0.0001$; specificity, $p_2 < 3 \times 10^{-19}$) but also in validation (sensitivity, $p_2 < 9 \times 10^{-9}$; specificity, $p_2 < 6 \times 10^{-6}$) as evaluated using a two-tailed Cochrane–Armitage test for trend [31].

Four models were found that were both 100% sensitive and specific in their ability to correctly discriminate serum samples taken from unaffected women or those suffering from ovarian cancer. All of these models were generated from data acquired using the QqTOF MS as no models generated from the PBS-II were both 100% sensitive and specific. Just as importantly, and key if this technology is to become a viable screening tool, no false-positive or false-negative classifications occurred using these models, giving each a PPV of 100% using the patient cohort employed in this study.

Another key aspect to this study is that the key *m/z* features that comprise the four diagnostic models that had 100% PPV for ovarian cancer revealed certain consistent features. Although the proteomic patterns generated from both healthy and cancer-affected patients using the QqTOF MS are quite similar (FIGURE 5), careful inspection of the raw mass spectra reveals that peaks at *m/z* values 7060.121 and 8605.678 are indeed differentially abundant in a selection of the serum samples obtained from ovarian cancer patients as compared with unaffected individuals (FIGURE 5, INSETS). The results indicated that these MS peaks originate from species that may be consistent indicators of

the presence of ovarian cancer and represent good candidates for ongoing efforts to identify low molecular weight components in serum that may be key disease progression indicators.

While a number of studies have reported impressive diagnostic success using the lower resolution PBS-II TOF MS to screen for diseases of relatively low prevalence such as ovarian cancer, a minimum level of 99.6% sensitivity and specificity is required [29]. In blinded testing and validation studies, any one of the four best models generated using QqTOF MS data were able to correctly classify 22 of 22 women with Stage I ovarian cancer, 81 of 81 women with Stage II, III and IV ovarian cancer and 68 of 68 benign disease controls. It can be envisioned in the near future that a clinical test would simultaneously employ several combinations of highly accurate diagnostic proteomic patterns which, if taken together, could achieve an even higher degree of accuracy in a screening setting where a diagnostic test will face large population heterogeneity and potential variability in sample quality and handling. Hence, a high-resolution system, such as the QqTOF MS employed in this study, is preferred based upon the present results that serve as a platform for clinical trials of serum proteomic patterns.

Summary & conclusions

One of the overlooked powers of investigating proteomic patterns is the ability to screen hundreds of serum samples in a high-throughput manner and therefore quickly determine targets (key *m/z* values) for further investigation. The inherent variability of serum between individuals makes it impossible to compare and recognize valid disease indicators using the conventional proteomic techniques of protein separation (2DE or multidimensional liquid chromatographic fractionation) and MS identification. The technology used to generate proteomic patterns is highly automated and even an academic laboratory can analyze in excess of 300 samples per day. This throughput allows for key discriminatory features to be distinguished within hundreds of serum or plasma samples over a statistically relevant population in a rapid fashion. It must

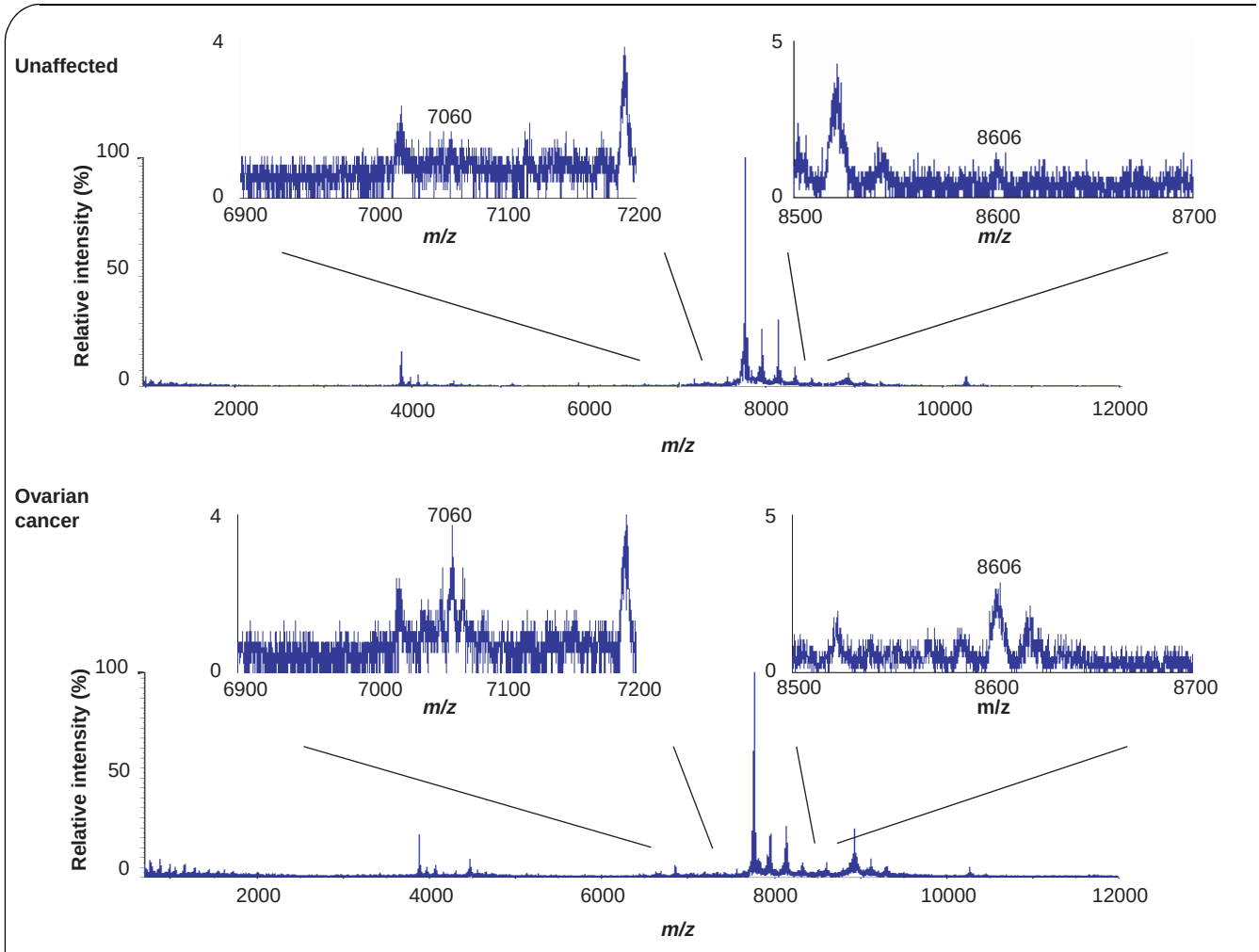

Figure 5. Comparison of SELDI QqTOF mass spectra of serum from an unaffected individual and an ovarian cancer patient. Insets show expanded *m/z* regions highlighting significant intensity differences of the peaks in the *m/z* 7060.121 and 8605.678 identified by the algorithm as belonging to the optimum discriminatory pattern.
m/z: Mass to charge ratio; SELDI QuTOF: Surface-enhanced laser desorption/ionization hybrid quadrupole time-of-flight.

be reiterated that the ability to distinguish sera from an unaffected individual or an individual with ovarian cancer based upon a single serum proteomic *m/z* feature alone is not possible across the entire serum study set. Accurate histological distinction is only possible when the key *m/z* features and their intensities are considered *en masse*.

A limitation of individual cancer biomarkers is the lack of sensitivity and specificity when applied to large heterogeneous populations [29,32]. Biomarker pattern analysis is an emerging technology aimed at overcoming this limitation. While serum proteomic pattern analysis has the potential to provide new tools for early diagnosis, therapeutic monitoring and outcome analysis, the success of this method will depend upon the ability of a selected set of features to transcend the biologic heterogeneity and methodological background noise. Using clinical study sets, progress has been made toward this diagnostic goal by employing a genetic algorithm coupled with a self-organizing cluster analysis to discover diagnostic subsets of *m/z* features and their relative intensities contained within high resolution mass spectral data. One of the consistencies within many of the diagnostic proteomic patterns is that a majority of the key *m/z* values are of low molecular weight, typically less than 10 kDa. The low molecular weight serum proteome is an unexplored archive, even though this is the mass region where MS is best suited for analysis. It is likely that disease-associated species are comprised of low molecular weight peptide/protein species that vary in mass by as little as a few Daltons. Thus, bioinformatic analysis of higher resolution MS data would be expected to discover patterns not discernable within lower resolution MS data.

One major criticism of the use of proteomic patterns for diagnostic purposes is that the identity of the proteins or peptides giving rise to the key *m/z* features is not known [33,34]. At this point, it is debatable as to whether it is worth the effort to identify these features as they may provide little aid in developing an alternative diagnostic platform. For example, many of the key features within the proteomic patterns that account for the diagnostic predictability are of low *m/z* (<10,000 Da) and therefore it is likely that these could be from fragment species generated from larger proteins that are proteolyzed either within the circulatory system or in the tumor/host microenvironment. It would be extremely challenging to generate an affinity reagent with specificity to a peptide fragment without considerable cross-reactivity to its parent protein. In addition, there are many tools in medicine today, such as the electrocardiogram, with which the physician relies solely on a pattern to base his/her diagnosis. Even the identification of a specific biomarker may not provide any direct insight into how a disease may arise or progress. For instance, while PSA is used to indicate the possible presence of a prostatic tumor, its role in cancer development remains unclear. Conversely, there is also the likelihood that these key features may represent proteins that provide exciting insights into the manifestation and progression of cancer. Therefore, identifying these features is most likely a worthwhile effort although the advancement of disease diagnostics using proteomic patterns should not be hindered by this exercise.

Expert opinion

Disease diagnostics using proteomic patterns has rapidly emerged as a potentially revolutionary tool to detect and monitor disease progression or therapeutic response. Its emergence is somewhat analogous to molecular fingerprinting in which the DNA patterns obtained from different tumors have been demonstrated to be unique for each cancer [35]. In molecular fingerprinting, the hope is that, in the future, physicians may be able to use this information to design treatment programs specific for each type of tumor. While the development of molecular fingerprinting has followed the progression of genomic analysis, proteomic pattern analysis, however, represents a complete about-face in proteomic analysis. While the trend in proteomic technology has been to identify and characterize an increasing number of proteins from a particular clinical sample in order to find a disease-specific biomarker, proteomic patterns rely simply on a crude proteomic survey that provides all of the necessary diagnostic information. While the potential is great, much still needs to be learned. The concept of using a proteomic pattern as a diagnostic tool is in its infancy, therefore every step in this analytical process requires optimization. This optimization process will include such aspects as sample acquisition and processing, in addition to pattern acquisition and data analysis. Since the diagnostic power of proteomic patterns relies heavily upon the use of bioinformatics, it is important to discover the biological basis behind the mathematical solution. While the identification of key peaks that are distinguished by the bioinformatic analysis may not provide any clues as to the manifestation or progression of the disease, the hope is they can at least validate the results being provided. While many critics still abound, one simple fact cannot be ignored: the diagnostic models generated from proteomic patterns continue to provide highly sensitive and specific results in testing and blind validation studies, even as the number of samples being analyzed continues to increase.

Five-year view

The next 5 years will be critical in the validation of the use of proteomic patterns in disease detection. Currently, the information present in proteomic patterns may provide an extremely powerful complementary tool to assist physicians in disease diagnosis. The impact of proteomic patterns in disease diagnosis, however, has the potential to be even greater than a complementary tool. While at this point it is not clear as to whether proteomic patterns reveal interindividual differences within the same type of cancer, there is an interest in using proteomic patterns to recognize the best treatment for each afflicted individual. While not fully developed, there is an active interest in determining if proteomic patterns can be used to predict a patient's response to a specific therapy. By combining information from proteomic patterns with that obtained from molecular fingerprints or a histopathological assessment, the optimal treatment for the individual may be more obvious than a simple trial and error regimen. The NCI has invested in a program to garner FDA approval for the use of proteomic patterns in the diagnosis of ovarian cancer in high-risk populations. In addition,

the two largest clinical diagnostic laboratories in the USA, Quest Diagnostics (NJ, USA) and Laboratory Corporation of America (NC, USA), have signed licensing agreements to develop and market the ovarian cancer protein pattern blood test. As with any emerging technology, the niche that proteomics will fill within the field of diagnostic medicine remains to be seen. The most obvious benefit of defined proteomic pattern diagnostic features can provide is in population screening to detect diseases such as cancer at earlier stages to enable more effective medical intervention. The simplicity of the test makes it feasible to screen high-risk populations for a variety of different cancers. The utility of proteomic patterns will be highly dependent upon the level of their inherent sensitivity and specificity. If the sensitivity and specificity can approach 100%, disease diagnosis using proteomic patterns will revolutionize diagnostic medicine as it can be used reliably for the early detection of low prevalence cancers. The detection of cancers at the earliest possible stage will save countless lives and help to meet the goal of the NCI to alleviate the pain and suffering of cancer by the year 2015. Even if this level of sensitivity and specificity is not achieved, proteomic patterns will still provide an invaluable complement to determine the need for a patient biopsy or response to therapy.

Acknowledgments

The authors would like to acknowledge Correlogics Systems for their collaboration on using high resolution mass spectrometry to diagnose ovarian cancer. This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, Contract No. NO1-CO-12400.

Disclaimer

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Key issues

- There is an urgent need for cancer biomarkers with more accurate diagnostic capability, particularly for early-stage disease.
- Conventional proteomic technologies focus upon identifying disease-specific biomarkers, while proteomic pattern analysis uses the overall pattern to diagnose disease states without the need to identify any of the components within the pattern.
- Disease diagnostics using proteomic patterns is a revolutionary method to detect early-stage cancer.
- Raw biofluids, such as serum and plasma, can be used to acquire proteomic patterns with a simple time-of-flight mass spectrometer.
- Bioinformatic software is required to decipher the patterns within the mass spectra that discriminate serum acquired from healthy and cancer-affected individuals.
- Information contained within proteomic patterns has been demonstrated to detect ovarian, breast and prostate cancers with sensitivities and specificities greater than 90%.
- The use of a higher resolution mass spectrometer has demonstrated the potential to provide high enough sensitivity and specificity to enable the use of proteomic patterns as a screening tool for low prevalence cancers.
- Until further blinded validation studies are performed to verify the apparent extraordinary sensitivity and specificity of this approach, the method should be considered investigational and not yet ready for clinical use.

References

Papers of special note have been highlighted as:

- • of interest
- •• of considerable interest

1 Wilkins MR, Sanchez JC, Gooley AA *et al.* Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* 13,19–50 (1996).

2 Phizicky E, Bastiaens PI, Zhu H, Snyder M, Fields S. Protein analysis on a proteomic scale. *Nature* 422(6928), 208–215 (2003).

3 Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 422(6928), 198–207 (2003).

• Describes the current state-of-the-art technology in the characterization of proteomes using methods to identify thousands of proteins in clinical samples.

4 Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. *Nature Genet.* 33(Suppl.), 311–323 (2003).

5 Schieltz DM, Yates JR 3rd. Direct identification of proteins in ultracomplex mixtures. Applications to proteome analysis. *Methods Mol. Biol.* 211, 235–245, (2003).

6 Resing KA. Analysis of signaling pathways using functional proteomics. *Ann. NY Acad. Sci.* 971, 608–614, (2002).

7 Hanash S. Disease proteomics. *Nature* 422(6928), 226–232 (2003).

8 Temple RJ. A regulatory authority's opinion about surrogate end points. In: *Clinical Measurement in Drug Evaluation*. Nimmo WS, Tucker GT (Eds). John Wiley and Sons, Inc., New York, NY, USA (1995).

9 Rippey JH. Pregnancy tests: evaluation of current status. *Crit. Rev. Clin. Lab. Sci.* 19(4), 353–359 (1984).

10 Srivastava S, Gopal-Srivastava R. Biomarkers in cancer screening: a public health perspective. *J. Nutr.* 132(8 Suppl.), 2471S–2475S (2002).

11 Guppy AE, Rustin GJ. CA125 response: can it replace the traditional response criteria in ovarian cancer? *Oncologist* 7(5), 437–443 (2002).

12 Grossklaus DJ, Smith JA, Shappell SB, Coffey CS, Chang SS, Cookson MS. The free/total prostate-specific antigen ratio is the best predictor of tumor involvement in the radical prostatectomy specimen among men with an elevated PSA. *Urol. Oncol.* 7(5), 195–198 (2002).

13 Adkins JN, Varnum SM, Auberry KJ *et al.* Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. *Mol. Cell. Proteomics* 1(12), 947–955 (2002).

14 Pieper R, Su Q, Gatlin CL, Huang ST, Anderson NL, Steiner S. Multi-component immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome. *Proteomics* 3(4), 422–432 (2003).

15 Hutchens TW, Yip TT. New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid. Commun. Mass Spectrom.* 7, 576–580 (1993).

16 Issaq HJ, Conrads TP, Prieto DA, Tirumalai R, Veenstra TD. SELDI-TOF MS for diagnostic proteomics. *Anal. Chem.* 75(7), 148A–155A (2003).

• Describes details of the procedure behind the acquisition of proteomic patterns ranging from the preparation of protein chips to their analysis by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS).

17 Petricoin EF, Ardekani AM, Hitt BA *et al.* Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359(9306), 572–577 (2002).

•• Describes the use of proteomic patterns for the diagnosis of ovarian cancer and provides an in-depth description of the bioinformatic tools used to classify benign from malignant conditions.

18 Jacobs IJ, Skates SJ, MacDonald N *et al.* Screening for ovarian cancer: a pilot randomised controlled trial. *Lancet* 353, 1207–1210 (1999).

19 Cohen LS, Escobar PF, Scharm C, Glimco B, Fishman DA. Three-dimensional power Doppler ultrasound improves the diagnostic accuracy for ovarian cancer prediction. *Gynecol. Oncol.* 82, 40–48 (2001).

20 Ozols RF, Rubin SC, Thomas GM, Robboy SJ. Epithelial ovarian cancer. In: *Principles and Practice of Gynecologic Oncology.* Hoskins WJ, Perez CA, Young RC (Eds),.Lippincott Williams and Wilkins, Philadelphia, PA, USA, 981–1058 (2000).

21 Menon U, Jacobs I. Recent developments in ovarian cancer screening. *Curr. Opin. Obstet. Gynaecol.* 12, 39–42 (2000).

22 Holland JH. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence (3rd edition). MIT Press, Cambridge, MA, USA (1994).

23 Kohonen Y. Self-organizing formation of topologically correct feature maps. *Biol. Cybern.* 43, 59–69 (1982).

24 Kohonen T. The self-organizing map. *Proc. Inst. Electrical Electronics Eng.* 78, 1464–1480 (1990).

25 Li J, Zhang Z, Rosenzweig J, Wang YY, Chan DW. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin. Chem.* 48(8), 1296–1304 (2002).

26 Qu Y, Adam BL, Yasui Y *et al.* Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. *Clin. Chem.* 48(10), 1835–1843 (2002).

27 Adam BL, Qu Y, Davis JW *et al.* Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res.* 62(13), 3609–3614 (2002).

•• Describes the use of proteomic patterns to provide a highly accurate method for the early detection and diagnosis of prostate cancer.

28 Petricoin EF 3rd, Ornstein DK, Paweletz CP *et al.* Serum proteomic patterns for detection of prostate cancer. *J. Natl Cancer Inst.* 94(20), 1576–1578 (2002).

29 Kainz C. Early detection and preoperative diagnosis of ovarian carcinoma. *Wien. Med. Wochenschr.* 146(1–2), 2–7 (1996).

•• Describes the need for high specificity in the screening of women for ovarian cancer.

30 Chernushevich IV, Loboda AV, Thomson BA. An introduction to quadrupole-time-of-flight mass spectrometry. *J. Mass Spectrom.* 36(8), 849–865 (2001).

31 Agresti A. *Categorical Data Analysis.* John Wiley and Sons, New York, NY, USA (1990).

32 Leung GM, Lam T-H, Thach TQ, Hedley AJ. Will screening mammography in the east do more harm than good? *Am. J. Public Health* 92(11), 1841–1846 (2002).

33 Diamandis EP. Proteomic patterns in serum and identification of ovarian cancer. *Lancet* 360(9327), 170 (2002).

• Points out some of the criticisms of using proteomic patterns as a diagnosis for ovarian cancer.

34 Diamandis EP. Serum proteomic patterns for detection of prostate cancer. *J. Natl Cancer Inst.* 95(6), 489–90 (2003).

35 Perou CM, Sorlie T, Eisen MB *et al.* Molecular portraits of human breast tumours. *Nature* 406(6797), 747–752 (2000).

Affiliations

- *Thomas P Conrads, PhD*
  Associate Director
  Mass Spectrometry Center,
  Biomedical Proteomics Program,
  SAIC-Frederick, Inc.,
  National Cancer Institute at Frederick,
  PO Box B,
  Frederick, MD 21702, USA
  Tel.: +1 301 846 7353
  Fax: +1 301 846 6037
  conrads@ncifcrf.gov
- *Ming Zhou, PhD*
  Senior Scientist
  Mass Spectrometry Center,
  Biomedical Proteomics Program,
  SAIC-Frederick, Inc.,
  National Cancer Institute at Frederick,
  PO Box B, Frederick,
  MD 21702, USA
  Tel.: +1 301 846 7199
  Fax: +1 301 846 6037
  mzhou@ncifcrf.gov
- *Emmanuel F Petricoin III*
  Co-Director
  NCI Clinical Proteomics Program,
  Food and Drug Administration-National Cancer Institute
  Clinical Proteomics Program,
  Center for Biologics Evaluation and Research,
  Food and Drug Administration,
  Bethesda, MD 20892, USA
  Tel.: +1 301 827 1753
  Fax: +1 301 480 3256
  petricoin@cber.FDA.gov
- *Lance Liotta*
  Co-Director
  NCI Clinical Proteomics Program,
  Laboratory of Pathology,
  National Cancer Institute,
  Center for Cancer Research,
  Bethesda, MD 20892, USA
  Tel.: +1 301 496 2035
  Fax: +1 301 480 0853
  liottal@nih.gov
- *Timothy D Veenstra, PhD*
  Director
  Analytical Chemistry Laboratory,
  Biomedical Proteomics Program,
  SAIC-Frederick, Inc.,
  National Cancer Institute at Frederick,
  PO Box B, Frederick,
  MD 21702, USA
  Tel.: +1 301 846 7286
  Fax: +1 301 846 6037
  veenstra@ncifcrf.gov