

Editorial

Proteomic and Genomic Technologies for Biomarker Discovery

Where Are the Biomarkers?

Vathany Kulasingam and Eleftherios P. Diamandis*

Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada
and Department of Pathology and Laboratory Medicine, Mount Sinai Hospital,
600 University Ave., Toronto, ON M5G 1X5, Canada

Introduction

Recently, we have witnessed the emergence of the “-omics” era with proteomics, peptidomics, degradomics, metabolomics, fractionomics, transcriptomics, interactomics, and so forth. However, the only “-omics” other than genomics to become a “buzzword” is *proteomics*, a term used for studying the proteome of an organism. With the completion of the Human Genome Project, optimistic views were expressed that novel disease-specific biomarkers will be discovered through various high-throughput techniques, such as microarrays and mass spectrometry (MS). Did the current technologies deliver the goods? In the following, we discuss briefly the current status and future of proteomics and genomics with respect to

cancer, as it is a model disease for studying such applications.

Current Cancer Biomarkers

Despite tremendous progress in our understanding of the molecular basis of many diseases, cancer continues to be a major cause of morbidity and mortality among men and women. There is convincing evidence that early detection of cancer can lead to better patient outcomes through early administration of therapy (1). Also, better clinical outcomes can be achieved with individualized treatments, which are more effective in subgroups of patients, rather than in the total patient population. One of the best ways to diagnose cancer early, or to predict therapeutic responses, is by using serum or tissue biomarkers.

*Author to whom all correspondence and reprint requests should be addressed:

Eleftherios P. Diamandis, Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada and Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Ave., Toronto, ON M5G 1X5, Canada.

E-mail: ediamandis@mtsinai.on.ca.

Currently, a number of cancer biomarkers are used clinically, including prostate-specific antigen (PSA; for prostate cancer), carcinoembryonic antigen (CEA; for colon, lung, and breast cancer), the carbohydrate antigens (CA) 125 (for ovarian cancer), CA 15.3 (for breast cancer), CA 19.9 (for gastrointestinal cancer), α -fetoprotein (AFP; for liver and testicular cancer), and human chorionic gonadotropin (hCG; for testicular and gestational cancers). All of these markers, despite being generally *not suitable* for population screening (because of low diagnostic sensitivity and specificity), play a major role in aiding diagnosis and for monitoring patients during therapy. Unfortunately, these cancer biomarkers have proven to be ineffective in detecting the early stages of disease and many suffer from low positive predictive values.

Renewed Interest in Cancer Biomarkers

Historically, every era of cancer biomarker discovery was associated with the emergence of a powerful analytical technology. For example, CEA was discovered in the 1960s mainly as a result of the introduction of novel and relatively sensitive immunological techniques (such as radial immunodiffusion), which allowed for the detection of the antigen in cancer tissues with high specificity and reasonable sensitivity. CA 125 and other CAs were developed, in the late 1970s, because of the establishment of the monoclonal antibody technology. Most of these tumor markers were discovered by using cell lines or tumor extracts as immunogens and then selecting specific hybridoma clones that recognized these tumor antigens (2). With the completion of the Human Genome Project and the introduction of technologies that enabled simultaneous examination of thousands of proteins and genes at any given time, such as MS and protein and DNA arrays, a renewed interest was taken by the

proteomics and genomics community to discover novel biomarkers.

Some Proteomic and Genomic Approaches to Biomarker Discovery

A number of different proteomic- and genomic-based approaches have been utilized to date to discover disease-specific biomarkers. Some of the approaches and strategies are briefly described here along with an example of how these methodologies were applied to aid in biomarker discovery or to develop patterns for early diagnosis.

Two-Dimensional Gel Electrophoresis and Mass Spectrometry

Two-dimensional polyacrylamide gel electrophoresis, known traditionally as the gold standard, discovery-based tool for proteomics, aims to resolve and visualize protein spots that are differentially expressed in healthy vs diseased samples. The spots can be excised from the gel, digested, and subsequently identified via MS. Using this methodology, Celis et al. (3) developed a comprehensive database for bladder cancer profiles of both transitional and squamous cell carcinomas. Furthermore, Kageyama et al. (4) used a differential display method of bladder cancer vs healthy urothelial tissue and MS to identify proteins, which are increased in cancer tissues. Calreticulin, a potential tumor marker, was identified in this study as it was found in urine of patients with bladder cancer. The authors confirmed their data with Western blot analysis, immunoprecipitation, and immunohistochemistry. They reported a sensitivity of 73% and a specificity of 86%. However, a caveat to this technology is that the identified proteins are usually of high abundance and the method is labor intensive. Similar strategies have been used by many other groups.

Protein Arrays

Another area of proteomic research that is gaining popularity is the use of protein arrays to identify novel biomarkers. Chinnaiyan and colleagues recently published data suggesting that autoantibody signatures may improve the early detection of prostate cancer (5). Using a combination of phage-display technology and protein microarrays, they identified new autoantibody-binding peptides derived from prostate cancer tissue. Serum samples from 119 patients with prostate cancer and 138 control patients were analyzed using this methodology. Their results were impressive. A 22-phage peptide detector had 88% specificity and 82% sensitivity in discriminating between the two groups. This panel of peptides seemed to perform better than PSA in distinguishing between prostate cancer and the control group.

Proteomic Pattern Analysis (SELDI-TOF)

The ability to identify diseases through proteomic patterns in biological fluids appeared to be a new paradigm shift in the use of MS-based methodologies. This approach attracted considerable attention over the past few years. Developed by Petricoin, Liotta, and colleagues, this potentially revolutionary methodology involves the use of a minute amount of unfractionated serum sample added to a "protein-chip," which is subsequently analyzed by surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) spectrometry to generate a proteomic signature of the serum. These patterns reflect part of the blood proteome, without knowledge of the actual identity of the proteins that make up the proteome. The potential of proteomic pattern analysis was first demonstrated in the diagnosis of ovarian cancer (6). In this study, a preliminary training set of mass spectra, generated by SELDI-TOF, from 50 unaffected women and 50 ovarian cancer patients was obtained. Searching algorithms were used to identify a proteomic pattern that discriminated cancer from

noncancer. This pattern was used to classify 116 blinded serum samples. The results were exceptional with a sensitivity of 100% at 95% specificity. These numbers are far superior to the sensitivities and specificities obtained with current serologic cancer biomarkers.

MS-Imaging (MALDI-MS)

Another proteomic pattern analysis approach involves the use of matrix-assisted laser desorption/ionization (MALDI)-MS on small amounts of fresh frozen tissue sections. In a recent application of this methodology, termed MS imaging, pioneered by Caprioli and colleagues (7), the authors profiled the protein expression of surgically resected lung tumor tissues (8). To detect proteomic patterns in lung tumors, the protein expression profiles of 50 tissue samples (42 lung tumors and 8 normal lung samples) in a training cohort was assessed. Mass spectra were obtained directly from 1-mm regions of single frozen tissue sections. More than 1600 distinct protein species were observed, of which 82 discriminating MS signals were selected. A class-prediction model was built and it was able to correctly classify all 42 lung tumors and the 8 normal samples. This model was further validated in an independent blinded test cohort yielding a phenomenal result of perfect classification of the samples.

Quantitative Proteomics

Quantitative proteomics aims to measure quantitatively the relative or absolute abundance of proteins across samples. The emergence of experimental protocols, such as isotope-coded affinity tags (ICAT), iTRAQ™, stable isotope labeling by amino acids in cell culture (and SILAC), now allow for quantitative proteomic approaches to biomarker discovery. As a result, quantitative differential analysis can be performed on an MS signal—the end result being the identification of the relative abundance of the molecular species

detected in the sample. In a study by Martin et al. (9), a cell culture model system was used to identify secreted and cell-surface proteins from neoplastic prostate epithelium using ICAT reagents and tandem MS. Proteomic analysis of the media identified more than 600 proteins, of which 524 were quantified. This study demonstrated the feasibility of using high-throughput quantitative proteomics to identify proteins in conditioned media.

Genomic Approaches

Genomic microarrays represent a highly powerful technology for genome-wide studies on gene expression, single-nucleotide polymorphism identification, mutational analysis, and so forth. Microarray experiments are usually performed with DNA or RNA isolated from tissues. Thus, unlike serologic markers, this methodology does not lend itself to minimally invasive diagnostics. However, tissue biopsies can be very useful for confirming diagnosis, for tumor subclassification and for predicting therapeutic response. It was recently hypothesized that the phenotypic differences of tumors from different patients may be associated with unique gene expression profiles, which can be captured by using DNA microarrays (10). The proof-of-principle for the cancer subclassification hypothesis has been provided for various malignancies, such as breast cancer, leukemias, and many other cancer types (11–18). Some of the most celebrated studies on cancer subclassification and subsequent selection of therapeutic intervention have created excitement in the media (19,20). Other approaches, based on the same hypothesis, used alternative technologies, such as quantitative reverse transcriptase-polymerase chain reaction or fluorescence *in situ* hybridization to derive prognostic and predictive information (21,22). Some of these tests have now been submitted to the Food

and Drug Administration for approval (Table 1).

Did Proteomic and Genomic Methodologies Deliver the Goods?

Rapid advances in proteomic and genomic technologies have created optimistic views that many more cancer biomarkers will be discovered through various high-throughput techniques. However, these predictions have yet to come true. Very few, if any, new serum cancer biomarkers have been introduced at the clinic over the last 15 yr. Organizations like the American Society of Clinical Oncology do not encourage the widespread use of tumor markers, unless they affect patient outcome measures (23). On the other hand, there is general agreement that a combination of multiple biomarkers may increase the diagnostic sensitivity and specificity of individual markers. This is particularly important in relation to the recent development of powerful bioinformatic algorithms (such as artificial neural networks and logistic regression), which can interpret multiple parameters much more efficiently than the average clinician (24,25).

Most of the serum proteomic pattern analysis technologies (such as SELDI-TOF), although promising, have not as yet been evaluated at the clinic and a number of important limitations have already been identified (26,27). These limitations include the possibility of bias stemming from artifacts related to the clinical samples used to generate published data, the inherent qualitative nature of MS and/or the bioinformatic analysis applied to the raw data (28–31). To illustrate this point, a meta-analysis of prostate cancer proteomic data from four articles by three different research groups was compiled (27). The discriminatory peaks identified in the four papers using SELDI-TOF were very different. Interestingly, two papers published among the same group using the same experimental data but different bioinformatic tools yielded different discriminatory peaks.

Table 1
New and Emerging “-omic” Cancer Tests

Company	Product	Approval status	Characteristics
Agendia	MammaPrint	Pending at FDA	<ul style="list-style-type: none"> ➤ A 70-gene expression signature measured on an Agilent DNA microarray ➤ Predicts prognosis for certain breast cancer patients
Correlogic systems	OvaCheck	Pending at FDA	<ul style="list-style-type: none"> ➤ Protein expression pattern analysis ➤ More than 150,000 starting data points are reduced to 5–10 for final evaluation
Exagen	Breast cancer prognosis test	Submission for FDA for PMA anticipated by end of 2005	<ul style="list-style-type: none"> ➤ Traditional FISH test, derived from gene expression as well as other data
Genomic health	Oncotype Dx	Not an FDA-regulated product. Launched as a laboratory-developed test in California, January 2004	<ul style="list-style-type: none"> ➤ A 21-gene expression assay done using RT-PCR for prognosis in breast cancer
Roche	Leukemia array	Launch anticipated 2006–2007	<ul style="list-style-type: none"> ➤ A 300-gene expression microarray test

Modified from ref. 35.

FDA, Food and Drug Administration; FISH, fluorescent *in situ* hybridization; PMA, pre-market approval; RT-PCR, reverse transcriptase-polymerase chain reaction.

This raises the problem of data overfitting and the lack of reproducibility with some of these technologies.

In a study by Gygi and coworkers, six companies were asked to analyze the same protein extract from 10,000 human cells by liquid chromatography-tandem MS after trypsin digestion (32). Only 52 proteins, representing 3% of all the data, were identified by all six participants. Furthermore, about two-thirds of the proteins were identified by only one participant. This exemplifies a very important point—different MS platforms identify different proteins in the same mixture. In addition, it should be noted that serum is a complex mixture and it is possible that new candidate biomarkers have not yet been identified

because their concentrations in serum and/or biological fluids are too low and therefore cannot be measured or the proteins cannot be purified, unless specific immunological reagents and highly sensitive enzyme-linked immunosorbent assay methods are available. Thus, in the initial discovery phase for novel cancer biomarkers, a less complex sample (elimination of high-abundant proteins) is essential. Therefore, proteomic technologies have yet to deliver the “goods” because they are more difficult to find than originally anticipated.

On the genomics side, Michiels et al. (33) recently performed a meta-analysis of seven of the most prominent studies on cancer prognosis, by using microarray-based expression

profiling. Surprisingly, in five of the seven studies on cancer prognosis, the original data could not be reproduced (34). The other two studies showed much weaker prognostic information than the original data. The meta-analysis also indicated that the list of genes identified as predictors of prognosis was highly unstable and that the molecular signatures were strongly dependent on the selection of patients in the training sets. This article suggested that the results of the aforementioned studies, as published, are highly over optimistic and that they need careful validation before conclusions can be drawn.

Where Do We Stand Today?

The emergence of the “-omics” era is an exciting time for the scientific community. The technologies and tools needed to decipher the proteome of organisms and to dig deeper into the biomarker mine are available. There is no doubt that, if these new technological advances prove to be successful in identifying cancer biomarkers for early cancer detection, the clinical benefits are likely to be enormous. What is not clear is which technologies will be the most successful as it is too early to say if we have indeed succeeded in delivering the “goods.” We should also not forget that other diagnostic modalities, such as imaging, may eventually become so sensitive and specific that the use of biomarkers for patient diagnosis and management may become redundant. Have we come upon a new paradigm shift in the use of proteomic and genomic technologies for biomarker discovery? Only meticulous validation steps and time will tell.

References

1. Etzioni, R., Urban, N., Ramsey, S., et al. (2003) The case for early detection. *Nat. Rev. Cancer* **3**, 243–252.
2. Bast, R. C., Jr., Klug, T. L., St. John, E., et al. (1983) A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. *N. Engl. J. Med.* **309**, 883–887.
3. Celis, A., Rasmussen, H. H., Celis, P., et al. (1999) Short-term culturing of low-grade superficial bladder transitional cell carcinomas leads to changes in the expression levels of several proteins involved in key cellular activities. *Electrophoresis* **20**, 355–361.
4. Kageyama, S., Isono, T., Iwaki, H., et al. (2004) Identification by proteomic analysis of calreticulin as a marker for bladder cancer and evaluation of the diagnostic accuracy of its detection in urine. *Clin. Chem.* **50**, 857–866.
5. Wang, X., Yu, J., Sreekumar, A., et al. (2005) Autoantibody signatures in prostate cancer. *N. Engl. J. Med.* **353**, 1224–1235.
6. Petricoin, E. F., Ardekani, A. M., Hitt, B. A., et al. (2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**, 572–577.
7. Caprioli, R. M. (2005) Deciphering protein molecular signatures in cancer tissues to aid in diagnosis, prognosis, and therapy. *Cancer Res.* **65**, 10,642–10,645.
8. Yanagisawa, K., Shyr, Y., Xu, B. J., et al. (2003) Proteomic patterns of tumor subsets in non-small-cell lung cancer. *Lancet* **362**, 433–439.
9. Martin, D. B., Gifford, D. R., Wright, M. E., et al. (2004) Quantitative proteomic analysis of proteins released by neoplastic prostate epithelium. *Cancer Res.* **64**, 347–355.
10. Perou, C. M., Sorlie, T., Eisen, M. B., et al. (2000) Molecular portraits of human breast tumors. *Nature* **406**, 747–752.
11. Alizadeh, A. A., Ross, D. T., Perous, C. M., and van de Rijn, M. (2001) Towards a novel classification of human malignancies based on gene expression patterns. *J. Pathol.* **195**, 41–52.
12. Weigelt, B., Hu, Z., He, X., et al. (2005) Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. *Cancer Res.* **65**, 9144–9158.
13. Alizadeh, A. A., Eisen, M. B., Davis, R. E., et al. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511.
14. Rosenwald, A., Wright, G., Chan, W. C., et al. (2002) Lymphoma/leukemia molecular profiling project: The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med.* **346**, 1937–1947.

15. Bhattacharjee, A., Richards, W. G., Staunton, J., et al. (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* **98**, 13,790–13,795.
16. Ramaswamy, S., Ross, K. N., Lander, E. S., and Golub, T. R. (2003) A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33**, 49–54.
17. Pomeroy, S. L., Tamayo, P., Gaasenbeek, M., et al. (2002) Prediction of central nervous system embryonal tumor outcome based on gene expression. *Nature* **415**, 436–442.
18. Iizuka, N., Hammoto, Y., and Oka, M. (2004) Predicting individual outcomes in hepatocellular carcinoma. *Lancet* **364**, 1837–1839.
19. van de Vijver, M. J., He, Y. D., van't Veer, I. J., et al. (2002) A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009.
20. van't Veer, I. J. and Weigelt, B. (2003) Road map to metastasis. *Nat. Med.* **9**, 999–1000.
21. Lossos, I. S., Czerwinski, D. K., Alizadeh, A. A., et al. (2004) Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N. Engl. J. Med.* **350**, 1828–1837.
22. Cronin, M., Pho, M., Dutta, D., et al. (2004) Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *Am. J. Pathol.* **164**, 35–42.
23. Bast, R. C., Jr., Ravdin, P., Hayes, D. F., et al. (2000) Update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. *J. Clin. Oncol.* **19**, 1865–1878.
24. Finne, P., Finne, R., Auvinen, A., et al. (2000) Predicting the outcome of prostate biopsy in screen-positive men by a multilayer perceptron network. *Urology* **56**, 418–422.
25. Stephan, C., Cammann, H., Semjonow, A., et al. (2002) Multicenter evaluation of an artificial neural network to increase the prostate cancer detection rate and reduce unnecessary biopsies. *Clin. Chem.* **48**, 1279–1287.
26. Diamandis, E. P. (2003) Point: proteomic patterns in biological fluids: do they represent the future of cancer diagnostics? *Clin. Chem.* **49**, 1272–1275.
27. Diamandis, E. P. (2003) Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J. Natl. Cancer Inst.* **96**, 353–356.
28. Karsan, A., Eigl, B. J., Fibotte, S., et al. (2005) Analytical and preanalytical biases in serum proteomic pattern analysis for breast cancer diagnosis. *Clin. Chem.* **51**, 1525–1528.
29. Banks, R. E., Stanley, A. J., Cairns, D. A., et al. (2005) Influences of blood sample processing on low-molecular-weight-proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. *Clin. Chem.* **51**, 1637–1649.
30. Baggerly, K. A., Morris, J. S., Edmonson, S. R., and Coombes, K. R. (2005) Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. *J. Natl. Cancer Inst.* **97**, 307–309.
31. Ransohoff, D. F. (2005) Lessons from controversy: ovarian cancer screening and serum proteomics. *J. Natl. Cancer Inst.* **97**, 315–319.
32. Elias, J. E., Haas, W., Faherty, B. K., and Gygi, S. P. (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods* **2**, 667–675.
33. Michiels, S., Koscielny, S., and Hill, C. (2005) Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* **365**, 488–492.
34. Ioannidis, J. P. (2005) Microarrays and molecular research: noise discovery? *Lancet* **365**, 454–455.
35. Branca, M. A. (2005) Omic diagnostics trip up on way to clinic. *Nat Biotechnol.* **23**, 769.