Journal of Zhejiang University SCIENCE B
ISSN 1673-1581 (Print); ISSN 1862-1783 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Review:

Proteomic technology for biomarker profiling in cancer: an update*

ALAOUI-JAMALI Moulay A.†, XU Ying-jie

(*Lady Davis Institute for Medical Research and Segal Comprehensive Cancer Center of the Sir Mortimer B. Davis Jewish General Hospital, Department of Oncology and Medicine, McGill University, Montreal, Que. H3T 1E2, Canada*)

†E-mail: moulay.alaoui-jamali@mcgill.ca

Received Apr. 10, 2006; revision accepted Apr. 16, 2006

Abstract: The progress in the understanding of cancer progression and early detection has been slow and frustrating due to the complex multifactorial nature and heterogeneity of the cancer syndrome. To date, no effective treatment is available for advanced cancers, which remain a major cause of morbidity and mortality. Clearly, there is urgent need to unravel novel biomarkers for early detection.

Most of the functional information of the cancer-associated genes resides in the proteome. The later is an exceptionally complex biological system involving several proteins that function through posttranslational modifications and dynamic intermolecular collisions with partners. These protein complexes can be regulated by signals emanating from cancer cells, their surrounding tissue microenvironment, and/or from the host. Some proteins are secreted and/or cleaved into the extracellular milieu and may represent valuable serum biomarkers for diagnosis purpose. It is estimated that the cancer proteome may include over 1.5 million proteins as a result of posttranslational processing and modifications. Such complexity clearly highlights the need for ultra-high resolution proteomic technology for robust quantitative protein measurements and data acquisition. This review is to update the current research efforts in high-resolution proteomic technology for discovery and monitoring cancer biomarkers.

Key words: Cancer, Biomarkers, Proteomics
doi:10.1631/jzus.2006.B0411 **Document code:** A **CLC number:** R73

CANCER BIOMARKERS: CURRENT STATUS

Despite advances in diagnostic imaging technology, surgical management, and therapeutic modalities, cancer remains a major cause of mortality worldwide. Currently, it is estimated that cancer kills over 6 million people per year worldwide, with over 10 million new cases being diagnosed every year. Mortality is mainly attributed to dissemination of primary cancer to distant organs, on which no effective treatment is available. The dilemma in oncology practice is the large number of patients presenting macro- or micro-metastases at primary diagnosis, and the fact that in some cases (e.g. breast cancer) metastases are seen in patients free of metastasis-positive axillary lymph nodes (Braun *et al*., 2000).

A small number of clinically approved biomarkers are available for early diagnosis and/or for successful monitoring of treatment and relapses (Table 1), so that the contributed significantly to reduced mortality rates and increased overall survival for cancers such as prostate cancer (Ludwig and Weinstein, 2005; Sidransky, 2002). However, most solid tumors, the absence of selective biomarkers hampers efforts to improve early detection and therapeutic management.

In the past years, significant progress in microarray technology has provided encouraging results on genetic patterns, or "fingerprints", that can predict cancer progression, drug response, and/or help cancer staging (Hoheisel, 2006; Balmain *et al*., 2003; van't Veer *et al*., 2002; van de Vijver *et al*., 2002; Staunton *et al*., 2001; Ziauddin and Sabatini, 2001). Nevertheless, comparative transcriptional profiling alone is unlikely to fully identify "the gold standard" biomarkers that can mirror the cancer phenotype. It is evident that most of the functional information on the genes resides in the proteome, which is the sum of

* Project supported by the National Cancer Institute of Canada

multiple dynamic processes that include protein phosphorylation, protein trafficking and localization, and protein-protein interactions with secondary effectors (Fig.1). Functional proteomics is potentially more predictive than genomics. Unlike genomics, however, acquiring proteomic data faces technical challenges, particularly with regard to lack of standardized methodologies, sensitivity and reproducibility, and intra- and inter-individual tumor heterogeneity. Nevertheless, these obstacles can be overcome with the current research efforts to develop high-resolution proteomic instrumentation for high-throughput monitoring of protein changes that occur in cancer (Tyers and Mann, 2003; Hanash, 2003).

The cancer proteome is an exceptionally complex biological sample containing information on perhaps every biological process that takes place in cancer cells, cancer tissue microenvironment, and cancer cell-host interaction. Cancer cells release protein biomarkers into the extracellular fluid through secretion of intact or cleaved peptides. In addition, cancer-associated circulating markers can be contributed by the tumor microenvironment, e.g. surrounding host cells such as fibroblasts and macrophages. Some of these products can end up in the bloodstream and hence serve as potential serum biomarkers. Therefore, studying the cancer proteome is the logical starting point for identifying diagnostic biomarkers and therapeutic targets for cancer. Already, several putative cancer biomarkers with potential clinical applications have been reported using proteomic technology (Table 2), and with the current trend many others are expected to come on board. This will result in a heavy load for biologists, pharmacologists, and oncologists for biomarker validation and translation into clinical applications.

This review will discuss the strengths of proteomic technology for high-resolution and high-

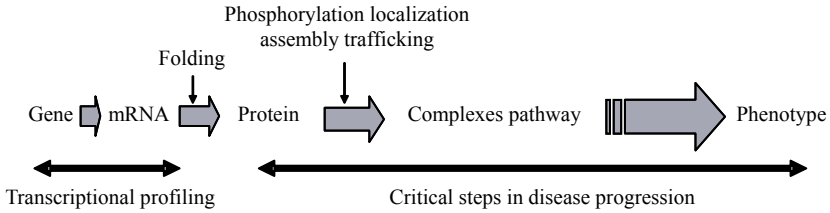

Fig.1 Illustration of the information flow from genotype to phenotype. The framework for the regulatory network would encompass knowledge from gene sequence to protein sequence, to posttranslational modifications, to protein cellular and sub-cellular distribution and trafficking, and the protein-protein interaction pattern

Table 1 Common serum cancer markers used in primary care

Biomarker	Cancer type	Specificity	Example of non-cancer pathology	Primary clinical use
α-fetoprotein	Hepatocellular, non-seminomatous testicular	Moderate	Prostatitis	Staging
Human chrionic gonadotropin-β	Testicular, ovarian	Low	Pregnancy	Staging
CA15-3	Breast	Poor	Cirrhosis, benign diseases of ovaries and breast	Disease monitoring
CA19-9	Gastro, pancreatic, stomach	Poor	Gastritis	Disease monitoring
CA125	Ovarian, cervical, uterine, fallopian tube	Moderate	Pancreatitis, kidney or liver disease	Disease monitoring
CA27-29	Breast			Disease monitoring
CEA	Colorectal, pancreas, lung, breast, medullary thyroid	Low	Non-malignant disorders	Disease monitoring
Epidermal growth factor receptor	Colon, non-small cell lung cancer	Low	Non malignant disorders, such as benign prostatic hyperplasia	Selection of therapy
Her2/Neu	Breast, ovarian	Moderate	Benign breast disease	Disease monitoring; selection of therapy
PSA	Prostate	High	Benign prostatic hyperplasia	Screening; disease monitoring
Thyroglobulin	Thyroid	Poor	Grave's disease thyroiditis	Disease monitoring

CA: Cancer antigen; CEA: Carcinoembryonic antigen; PSA: Prostate-specific antigen

Table 2 Examples of putative cancer biomarkers identified by mass spectrometry as potentially useful for diagnosis

Biomarker	Cancer type	References
Apolipoprotein A1	Ovarian, pancreatic	Zhang *et al*., 2004; Kozak *et al*., 2005
Heptaglobin α-subunit	Ovarian, pancreatic, lung	Ye *et al*., 2003
Transthyretin fragment	Ovarian	Kozak *et al*., 2005
Inter-alpha-trypsin inhibitor fragment	Ovarian, pancreatic	Zhang *et al*., 2004
Vitamin D-binding protein	Prostate, breast	Corder *et al*., 1993; Pawlik *et al*., 2006
Serum amyloid A	Nasopharyngeal, pancreatic, ovarian	Orchekowski *et al*., 2005; Moshkovskii *et al*., 2005
α1-antitrypsin and α1-antichymotrypsin	Pancreatic	Orchekowski *et al*., 2005; Yu *et al*., 2005
Osteopontin	Ovarian, prostate	Khodavirdi *et al*., 2006

throughput identification of cancer biomarkers with special emphasis given to clinical applications. For technical aspects related to instrumentation we refer the reader to several review articles that provide an exhaustive description of proteomic technology (Liotta and Petricoin, 2006; Conrads *et al*., 2005; Jessani *et al*., 2005; Wu *et al*., 2003; Celis and Gromov, 2003; Wulfkuhle *et al*., 2003; Phizicky *et al*., 2003; Aebersold and Mann, 2003; Petricoin *et al*., 2002a; Bichsel *et al*., 2001).

FISHING FOR "GOLD STANDARD" CANCER BIOMARKERS: GREAT EXPECTATION FROM NEW DEVELOPMENTS IN PROTEOMIC TECHNOLOGY

It is estimated that the 45000 human genes generate approximately 250000 spliced variants of RNA, which are translated into over 1.5 million proteins as a result of posttranslational processing and modifications (Ozier *et al*., 2003). Such complexity clearly highlights the need for assembling the protein circuits that operate in cancer cells and their surrounding microenvironment to achieve meaningful clinical prediction of the disease. Assuming that the proteome is the global representative of all biological processes that take place in cancer cells, then the discovery of specific biomarkers in the midst of such biological complexity would seem difficult in the absence of ultra-high resolution analytical techniques for quantitative measurement of tens to hundreds of thousands of components, and robust data acquisition and analysis techniques to efficiently and reliably process these large datasets.

Current progress in proteomics has been largely due to recent developments in mass spectrometry (MS)-based technologies (Aebersold and Goodlett, 2001). Particularly, new techniques for the ionization of proteins and peptides, such as matrix-assisted laser desorption-ionization (MALDI) and electrospray ionization (ESI) combined with time-of-flight (TOF), as well as new hybrid mass spectrometers, are now becoming the tools of choice for protein characterization. These advances have been highly recognized by the scientific community to include two mass spectrometrists, Drs. John B. Fenn and Koichi Tanaka as co-recipients (with the developer of NMR Dr. Kurt Wüthrich) recipient of the 2002 Nobel Prize for chemistry. These techniques have also been accompanied, although with a significant lag, by dramatic improvements in bioinformatic tools for analysis of complex datasets. In addition, powerful multi-dimensional chromatographic and sample labelling techniques have been developed to further benefit from the improvements in mass spectrometry (Kachman *et al*., 2002; Wolters *et al*., 2001; McCormack *et al*., 1997).

The standard proteomic approach for biomarker research consists of isolation of cell proteins from clinical specimens (tissue or biological fluids such as serum, ascites, saliva, etc.), digestion with proteases such as trypsin, and separation of the resulting mixture by two-dimensional (2D) electrophoresis or liquid chromatography (LC). The desired spots (2D) or protein fractions (LC) are isolated, digested, and peptides are separated by LC and depending on the sample complexity, the low-molecular weight fractions may be further fractionated by ion-exchange chromatography. The peptides are then subjected to electrospray or MALDI mass spectrometry (MS) or MS/MS analysis for qualitative and quantitative

comparisons. The mass spectrometer measures the mass to charge ratio of the components and, if required, can then perform more detailed experiments to identify the components. In the case of peptides, one obtains a partial sequence that is usually sufficient for identification of the peptide or parent protein from which the peptide was derived. Therefore, for each sample, one can obtain a three dimensional dataset (retention time, mass to charge ratio (m/z) and intensity).

Each of these technical approaches has advantages and limitations. For example, 2D gel electrophoresis suffers from low sensitivity, low dynamic range (10^3), difficulty in resolving proteins with extreme masses or isoelectric points, and the inability to resolve proteins with low solubility (Ong and Mann, 2005; Hanash, 2001). Some improvements were obtained when 2D methods are combined with prefractionation chromatography techniques (Righetti *et al*., 2003). The introduction of differential gel electrophoresis (DIGE) has provided improvement over standard 2D. DIGE involves tagging the two comparison protein solutions with different fluorescent dyes (Cy2, Cy3 or Cy5), then mixing the solutions and running them on the same gel. Computer software (DeCyder) allows comparison of these protein mixtures and, by using a pooled standard labelled with the third dye, the software can reliably compare quantitative differences in protein levels across multiple gels (Lilley and Friedman, 2004). However, this methods still suffers some of the disadvantages seen in 2D gel electrophoresis, in particular limited sensitivity and resolution to detect low molecular weight proteins (Chakravarti *et al*., 2004). In contrast, multidimensional protein identification technology (MudPIT) is more widely used as a sensitive method (Bayer *et al*., 2006; Chen *et al*., 2006; Guzzetta and Chien *et al*., 2005; Durr *et al*., 2004). In MudPIT a complex protein mixture is digested with a specific protease (most commonly trypsin). The peptide fragments then undergo two-dimensional liquid chromatography, which separates the peptides initially by charge, then by hydrophobicity to allow adequate separation of peptides. The column elutes into an electrospray mass spectroscopy unit that identifies peptides as they elute, one by one, based on their tandem mass spectroscopy spectrum. Databases of peptides can then be searched to identify the proteins based on the mass spectra of the peptides. This technique has exquisite sensitivity, reported to be in the femtomolar (10^{-15}) range, with a dynamic range of 10000 to 1, and is reproducible to within 0.5% (Wang *et al*., 2005). One of the major weaknesses of MudPIT is in identifying quantitative differences in protein expression across protein mixtures (Washburn *et al*., 2003). To overcome this limitation, alternative variants have been developed. For example, one sample is labelled differently from another and the two samples are mixed prior to analysis; through mass spectroscopy it becomes feasible to detect small differences in protein concentrations. The two options are in vivo labelling (using N^{15} isotope) (Blagoev *et al*., 2004) or in vitro labelling using techniques such as ICAT (Ramus *et al*., 2006; Schrimpf *et al*., 2005; Kaji *et al*., 2003), MCAT (Cagney and Emili, 2002), or iTRAQ (Wu *et al*., 2006; Shadforth *et al*., 2005). The latter is particularly interesting and uses trypsin digestion, iTRAQ labelling, and automated in-line two-dimensional column chromatography and tandem mass spectroscopy. In the case of iTRAQ, the trypsin-digested samples are labelled with different iTRAQ reagents to allow detection of relative quantitative differences between samples. The sample passes through a nano-column packed with a strong cation exchange resin and a reverse phase resin, which sequentially separates peptides based on charge and hydrophobicity. The elution passes directly into an electrospray mass spectroscopy (Q TRAP MS/MS) unit, where the individual peptides are initially characterized by their mass to charge ratio. The peptides are then hit with a high velocity inert gas causing them to further fragment in a predictable pattern. By comparing the counts of peptides labelled with each iTRAQ reagent, one can be able to detect quantitative differences. Peptides with the greatest quantitative differences can be identified by searching databases of known peptide mass spectra.

Examples of exciting advances in biomarker discovery include the use of SELDI (surface enhanced laser desorption) mass spectroscopy to search for cancer biomarkers (Ricolleau *et al*., 2006; Henderson and Steele, 2005). In these studies, protein profiles were compared between healthy individuals and patients with advanced cancer and to those with benign disease. Statistical algorithms were used to identify a set of peaks with the greatest sensitivity and

specificity for disease, with reports on sensitivity approaching 100% and specificity of 95% for ovarian cancer (Conrads *et al*., 2005). Another interesting study was aimed at discovering novel biomarkers for the early detection and classification of ovarian cancer by SELDI-TOF mass spectrometry in 2002 (Petricoin *et al*., 2002b). This seminal research, which is both an excellent example of the potential benefits and pitfalls of biomarker research, rapidly became the subject of an act of Congress in the United States, the basis for a commercially available diagnostic test (Correlogic Inc.) and the topic of vigorous debate in the scientific community. Early criticism of the research was directed at both the egregious errors in statistical analysis and the lack of details regarding these biomarkers, specifically their identity. While the statistical errors have been corrected, the lack of identity of the putative biomarkers remains unaddressed. However, in a recent study the authors expanded their analysis to include the use of higher resolution Qq-TOF mass spectrometry in place of SELDI-TOF mass spectrometry, resulting in a dramatic increase in performance (100% sensitivity and specificity). Yet, the identity of the biomarkers remains unknown. The technique has also been used to study pancreatic cancer, prostate cancer, and recently head and neck cancer (Menard *et al*., 2006; Kozak *et al*., 2005; Clarke *et al*., 2005; Nishizuka *et al*., 2003; Paweletz *et al*., 2001). This technique has some reported disadvantages, including protein concentrations responsible for the 'peaks' in SELDI that are in the μg/ml range, much higher than known biomarkers such as PSA (prostate specific antigen).

New development for detailed analysis of the proteome can be achieved by using high resolution analytical techniques such as LC-FTICR-MS followed by LC-Qq-TOF-MS/MS. RPLC-FTICR-MS provides resolving power (the ability to distinguish biomolecules that are very close in mass) that is roughly 100 times higher than SELDI-TOF-MS and 20 times higher than Qq-TOF-MS. RPLC has a peak capacity of roughly 50. Therefore, one can calculate that this analytical procedure will be able to characterize a theoretical maximum of ~9 million individual components (20 ion exchange fractions×50 fractions by RPLC×9000 measurements per fraction by FTICR MS) (Ebanks *et al*., 2005; 2006; Melanson *et al*., 2006). In practice the number of components measured is much lower the theoretical maximum; the sample components do not conveniently distribute themselves evenly in the various fractions and the high-abundance peptides can suppress signals from the low-abundance peptides through ion-suppression. A reasonable estimate is 600 peaks in each FTICR MS spectrum, which would result in 600 000 components measured (Masselon *et al*., 2005; Shen *et al*., 2005; Nemeth-Cawley *et al*., 2003).

In contrast, the LC-Qq-TOF-MS techniques can provide approximately the same number of potential data points (4000) for every chromatographic peak (LC typically provides≥30 peaks). Therefore, the LC-Qq-TOF-MS technique can theoretically measure 120000 components. In similar fashion, LC-FTICR-MS provides very high resolution in terms of mass measurements resulting in a theoretical data density of $>10^6$ components per sample. In practice, samples do not typically contain such a large number of components; however, it is not unusual for complex proteomic samples to exceed 10^5 components. Therefore, techniques that can handle high sample complexity are logically preferable over low-density analytical techniques.

DATA ANALYSIS

Following the development of a suitable analytical platform for sample analysis, a rigorous evaluation of analytical reproducibility is essential to fully characterize the analytical variability. Although the analytical variability is likely to be small relative to biological variability, its assessment is necessary to avoid the inclusion of artifacts arising from variations in analytical processing in the list of putative biomarkers (Boguski and McIntosh, 2003).

One of the most widely used data analysis technique is principal component analysis (PCA) tries to describe the variation between classes of samples (e.g. diseased vs controls) by quantitatively comparing data in a pair-wise manner. The advantage of this technique is that one can process high dimensionality datasets and reduce the data to a few principal components.

In principle, a subset of standard peptides can be added to a sample prior to sample preparation and analysis to assess the variability of preparative,

chromatographic and mass spectrometric steps. These errors can be characterized by calculating the measurement error covariance matrix from replicate preparations and injections. Errors arising from variations in sample processing can be quantified by the addition of known amounts of internal standards (molecules with known chromatographic and mass spectral properties and which cover the expected range of biomarkers sought) at various points in the sample preparation and sample analysis steps. Specifically, internal standards can be added to the samples to correct for: (1) variations in fractionation of the raw plasma prior to LC-MS analysis, (2) correction of chromatographic variations (retention time shifts, injection volume variations), and (3) correction of mass spectrometry variations (drift in mass calibration and ionization efficiency).

Following the establishment of normalization methods, it will be necessary to subject the raw (normalized) data to a data reduction step. This is required for several reasons. First, the huge volume of LC-MS data makes direct analysis of the data computationally impractical. Perhaps more importantly, however, registration in the temporal and mass-spectral domains is not consistent, i.e. the mass-chromatograms do not overlay one another perfectly due to experimental variations. Data reduction will therefore be accomplished through a preliminary feature reduction based on the fact that many peaks are close to background noise or can be assigned to known contaminants (solvents, buffers, detergents, etc.). Using representative test and control samples, selected chromatographic peaks can be used to establish thresholds values below which mass spectral peaks can be ignored.

Although tedious, the sheer size (up to 10^6 data points per second) and complexity of the data requires the establishment of detailed data processing procedures. In addition, peptides give rise to several peaks in high-resolution mass spectrometry. Therefore, without knowledge of isotopic patterns of radiolabelled peptides, automated peak-finding algorithms could assign an isotope artifact as a unique feature that is only found in cancer. These potential artifacts are especially problematic for low-level components whose signals are near the instrument's detection limit. This situation is crucial since these low-level components are the most likely source of biomarkers for early detection. Therefore, intelligent algorithms for peak finding and data reduction are needed that will avoid the introduction of artifacts yet be sensitive enough to detect components in low amounts.

The next stage of the data analysis is to characterize the biological variability of the selected features within the test and control groups by analyzing multiple samples within each group. This can be accomplished through the use of error covariance and correlation matrices. Not only will this allow the variance of individual LC-MS features to be assessed, but the relationships among features (their covariance) can be examined. This characterization is important for excluding normal biological variation from the features that distinguish the test and control samples.

VALIDATION OF THE BIOMARKERS

The end result of proteomic analysis is to have appropriate validation before the marker can reach clinical applications. Once a putative biomarker has been identified, validation using additional measurements and compound identification is necessary (Hammond and Taube, 2002; Anderson and Anderson, 2002). For example, one can repeat the analysis at additional time points and determine the temporal correlation of putative biomarkers with the progression of the disease. Determining the timing of the appearance of a biomarker has been shown to be important in assessing a biomarker's prognostic utility. For example, ovarian cancer is accompanied by elevated serum osteopontin levels (Bast *et al*., 2005), though its diagnostic utility is inferior to CA125 (the only currently validated biomarker for ovarian cancer) as a biomarker for postoperative monitoring. However, serum osteopontin was shown to rise earlier than CA125 in patients suffering from recurrent ovarian cancer. This fact illustrates the utility of simultaneous measurement of multiple biomarkers.

Critical issues that need to be addressed for the validation studies include the specificity and reproducibility of the marker. In the case of cancer tissues and biological fluids, this is further complicated by intra- and inter-cell heterogeneity. The use of tumor tissues or needle biopsies may be problematic because multiple and representative tissue sampling is not always feasible, e.g. tumors in non-accessible

sites. Furthermore, protein profiling can differ between a given area of the tumor mass. To address these problems, a proteomic study by Celis *et al*.(2002) on the most prevalent transitional cell carcinoma of the bladder has combined proteomics and immunohistochemistry with clinical data gathered over a 5-year period and correlated the protein database to cancer cell heterogeneity within normal urothelium with recurrence. The study reported that tumor lesions displaying phenotypic alterations in the basal, proliferative compartment of the urothelium have the highest incidence of recurrence.

More exciting methods are being developed that can compare proteins present in easily accessible biological fluids from patients, which are predictive of disease progression and/or therapeutic response. This approach has been reported for body fluids including serum (Petricoin *et al*., 2004; Pusztai and Gianni, 2004; Bachelot *et al*., 2003). However, most current published studies are very preliminary and were conducted in a very small number of samples with no specific marker being carefully validated. As with genomics, proteomic technology is still evolving to improve its predictive value (Knezevic *et al*., 2001).

CONCLUSION AND PERSPECTIVES

Current clinical and pathological markers poorly predict early disease development and response to treatment. Standard diagnostic methods, including tissue histopathology are now shifting rapidly toward molecular diagnosis due to the rapid progress in proteomic instrumentation. This powerful technology can identify all proteins and their posttranslational modifications in disease conditions, and hence will greatly accelerate progress toward novel diagnostic and predictive tools to track early disease and tailor treatments to specific patients.

ACKNOWLEDGEMENT

Dr. Alaoui-Jamali is a Scholar of the "Fond de Recherches Scientifiques du Quebec (FRSQ)" and a recipient of Dundi and Lyon Sachs Distinguished Scientist Award. His laboratory has been supported by the Canadian Breast Cancer Research Alliance of the National Cancer Institute, the National Cancer Institute of Canada, the Canadian Institutes for Health Research (CIHR), the Cancer Research Society Inc., and the US Army Research Program. Ying-jie Xu is supported by a CIHR Training Award.

References

Aebersold, R., Goodlett, D.R., 2001. Mass spectrometry in proteomics. *Chem. Rev.*, **101**(2):269-295. [doi:10.1021/cr990076h]

Aebersold, R., Mann, M., 2003. Mass spectrometry-based proteomics. *Nature*, **422**(6928):198-207. [doi:10.1038/nature01511]

Anderson, N.L., Anderson, N.G., 2002. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics*, **1**(11):845-867. [doi:10.1074/mcp.R200007-MCP200]

Bachelot, T., Ray-Coquard, I., Menetrier-Caux, C., Rastkha, M., Duc, A., Blay, J.Y., 2003. Prognostic value of serum levels of interleukin 6 and of serum and plasma levels of vascular endothelial growth factor in hormone-refractory metastatic breast cancer patients. *Br. J. Cancer*, **88**(11):1721-1726. [doi:10.1038/sj.bjc.6600956]

Balmain, A., Gray, J., Ponder, B., 2003. The genetics and genomics of cancer. *Nature Genet.*, **33**(Suppl. 3):238-244. [doi:10.1038/ng1107]

Bast, R.C.Jr, Badgwell, D., Lu, Z., Marquez, R., Rosen, D., Liu, J., Baggerly, K.A., Atkinson, E.N., Skates, S., Zhang, Z., *et al*., 2005. New tumor markers: CA125 and beyond. *Int. J. Gynecol. Cancer*, **15**(Suppl. 3):274-281. [doi:10.1111/j.1525-1438.2005.00441.x]

Bayer, E.M., Bottrill, A.R., Walshaw, J., Vigouroux, M., Naldrett, M.J., Thomas, C.L., Maule, A.J., 2006. Arabidopsis cell wall proteome defined using multidimensional protein identification technology. *Proteomics*, **6**(1):301-311. [doi:10.1002/pmic.200500046]

Bichsel, V.E., Liotta, L.A., Petricoin, E.F., 2001. Cancer proteomics: from biomarker discovery to signal pathway profiling. *Cancer J.*, **7**:69-78.

Blagoev, B., Ong, S.E., Kratchmarova, I., Mann, M., 2004. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat. Biotechnol.*, **22**(9):1139-1145. [doi:10.1038/nbt1005]

Boguski, M.S., McIntosh, M.W., 2003. Biomedical informatics for proteomics. *Nature*, **422**(6928):233-237. [doi:10.1038/nature01515]

Braun, S., Pantel, K., Muller, P., Janni, W., Hepp, F., Kentenich, C.R., Gastroph, S., Wischnik, A., Dimpfl, T., Kindermann, G., *et al*., 2000. Cytokeratin-positive cells in the bone marrow and survival of patients with stage I, II, or III breast cancer. *N. Engl. J. Med*., **342**(8):525-533. [doi:10.1056/NEJM200002243420801]

Cagney, G., Emili, A., 2002. De novo peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging. *Nat. Biotechnol.*,

20(2):163-170. [doi:10.1038/nbt0202-163]

Celis, J.E., Gromov, P., 2003. Proteomics in translational cancer research: toward an integrated approach. *Cancer Cell*, **3**(1):9-15. [doi:10.1016/S1535-6108(02)00242-8]

Celis, J.E., Celis, P., Palsdottir, H., Ostergaard, M., Gromov, P., Primdahl, H., Orntoft, T.F., Wolf, H., Celis, A., Gromova, I., 2002. Proteomic strategies to reveal tumor heterogeneity among urothelial papillomas. *Mol. Cell Proteomics*, **1**(4):269-279. [doi:10.1074/mcp.M100031-MCP200]

Chakravarti, D.N., Gallagher, S., Chakravarti, B., 2004. Difference gel electrophoresis: application in quantitative proteomics research. *Current Proteomics*, **1**(4):261-271.

Chen, E.I., Hewel, J., Felding-Habermann, B., Yates, J.R.3rd, 2006. Large scale protein profiling by combination of protein fractionation and multidimensional protein identification technology (MudPIT). *Mol. Cell Proteomics*, **5**(1):53-56. [doi:10.1074/mcp.T500013-MCP200]

Clarke, C.H., Buckley, J.A., Fung, E.T., 2005. SELDI-TOF-MS proteomics of breast cancer. *Clin. Chem. Lab. Med.*, **43**(12):1314-1320. [doi:10.1515/CCLM.2005.225]

Conrads, T.P., Hood, B.L., Petricoin III, E.F., Liotta, L.A., Veenstra, T.D., 2005. Cancer proteomics: many technologies, one goal. *Expert Rev. Proteomics*, **2**(5):693-703. [doi:10.1586/14789450.2.5.693]

Corder, E.H., Guess, H.A., Hulka, B.S., Friedman, G.D., Sadler, M., Vollmer, R.T., Lobaugh, B., Drezner, M.K., Vogelman, J.H., Orentreich, N., 1993. Vitamin D and prostate cancer: a prediagnostic study with stored sera. *Cancer Epidemiol., Biomarkers & Prev.*, **2**:467-472.

Durr, E., Yu, J., Krasinska, K.M., Carver, L.A., Yates, J.R., Testa, J.E., Oh, P., Schnitzer, J.E., 2004. Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture. *Nat. Biotechnol.*, **22**(8):985-992. [doi:10.1038/nbt993]

Ebanks, R.O., Goguen, M., McKinnon, S., Pinto, D.M., Ross, N.W., 2005. Identification of the major outer membrane proteins of Aeromonas salmonicida. *Dis. Aquat. Organ.*, **68**(1):29-38.

Ebanks, R.O., Chisholm, K., McKinnon, S., Whiteway, M., Pinto, D.M., 2006. Proteomic analysis of Candida albicans yeast and hyphal cell wall and associated proteins. *Proteomics*, **6**(7):2147-2156. [doi:10.1002/pmic.200500100]

Guzzetta, A.W., Chien, A.S., 2005. A double-vented tetraphasic continuous column approach to MudPIT analysis on long capillary columns demonstrates superior proteomic coverage. *J. Proteome Res.*, **4**(6):2412-2419. [doi:10.1021/pr050209h]

Hammond, M.E., Taube, S.E., 2002. Issues and barriers to development of clinically useful tumor markers: a development pathway proposal. *Semin. Oncol.*, **29**(3):213-221. [doi:10.1053/sonc.2002.32896]

Hanash, S., 2001. 2-D or not 2-D: is there a future for 2-D gels in proteomics? Insights from the York proteomics meeting. *Proteomics*, **1**:635-637.

Hanash, S., 2003. Disease proteomics. *Nature*, **422**(6928):226-232. [doi:10.1038/nature01514]

Henderson, N.A., Steele, R.J., 2005. SELDI-TOF proteomic analysis and cancer detection. *Surgeon*, **3**(6):383-390.

Hoheisel, J.D., 2006. Microarray technology: beyond transcript profiling and genotype analysis. *Nat. Rev. Genet.*, **7**(3):200-210. [doi:10.1038/nrg1809]

Jessani, N., Niessen, S., Wei, B.Q., Nicolau, M., Humphrey, M., Ji, Y., Han, W., Noh, D.Y., Yates, J.R.3rd, Jeffrey, S.S., *et al*., 2005. A streamlined platform for high-content functional proteomics of primary human specimens. *Nat. Methods*, **2**(9):691-697. [doi:10.1038/nmeth778]

Kachman, M.T., Wang, H., Schwartz, D.R., Cho, K.R., Lubman, D.M., 2002. A 2-D liquid separations/mass mapping method for interlysate comparison of ovarian cancers. *Anal. Chem.*, **74**(8):1779-1791. [doi:10.1021/ac011159c]

Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi, N., Isobe, T., 2003. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotechnol.*, **21**(6):667-672. [doi:10.1038/nbt829]

Khodavirdi, A.C., Song, Z., Yang, S., Zhong, C., Wang, S., Wu, H., Pritchard, C., Nelson, P.S., Roy-Burman, P., 2006. Increased expression of osteopontin contributes to the progression of prostate cancer. *Cancer Res.*, **66**(2):883-888. [doi:10.1158/0008-5472.CAN-05-2816]

Knezevic, V., Leethanakul, C., Bichsel, V.E., Worth, J.M., Prabhu, V.V., Gutkind, J.S., Liotta, L.A., Munson, P.J., Petricoin III, E.F., Krizman, D.B., 2001. Proteomic profiling of the cancer microenvironment by antibody arrays. *Proteomics*, **1**(10):1271-1278. [doi:10.1002/1615-9861(200110)1:10<1271::AID-PROT1271>3.0.CO;2-6]

Kozak, K.R., Su, F., Whitelegge, J.P., Faull, K., Reddy, S., Farias-Eisner, R., 2005. Characterization of serum biomarkers for detection of early stage ovarian cancer. *Proteomics*, **5**(17):4589-4596. [doi:10.1002/pmic.200500093]

Lilley, K.S., Friedman, D.B., 2004. All about DIGE: quantification technology for differential-display 2D-gel proteomics. *Exp. Rev. Proteomics*, **1**(4):401-409. [doi:10.1586/14789450.1.4.401]

Liotta, L.A., Petricoin, E.F., 2006. Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. *J. Clin. Invest.*, **116**(1):26-30. [doi:10.1172/JCI27467]

Ludwig, J.A., Weinstein, J.N., 2005. Biomarkers in cancer staging, prognosis and treatment selection. *Nature Reviews Cancer*, **5**(11):845-856. [doi:10.1038/nrc1739]

Masselon, C., Pasa-Tolic, L., Tolic, N., Anderson, G.A., Bogdanov, B., Vilkov, A.N., Shen, Y., Zhao, R., Qian, W.J., Lipton, M.S., *et al*., 2005. Targeted comparative proteomics by liquid chromatography-tandem fourier ion cyclotron resonance mass spectrometry. *Anal. Chem.*, **77**(2):400-406. [doi:10.1021/ac049043e]

McCormack, A.L., Schieltz, D.M., Goode, B., Yang, S., Barnes, G., Drubin, D., Yates, J.R.3rd, 1997. Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-Femtomole level. *Anal. Chem.*, **69**(4):767-776. [doi:10.1021/ac960799q]

Melanson, J.E., Chisholm, K.A., Pinto, D.M., 2006. Targeted

comparative proteomics by liquid chromatography/matrix-assisted laser desorption/ionization triple-quadrupole mass spectrometry. *Rapid Commun. Mass Spectrom.*, **20**(5):904-910. [doi:10.1002/rcm.2391]

Menard, C., Johann, D., Lowenthal, M., Muanza, T., Sproull, M., Ross, S., Gulley, J., Petricoin, E., Coleman, C.N., Whiteley, G., *et al*., 2006. Discovering clinical biomarkers of ionizing radiation exposure with serum proteomic analysis. *Cancer Res.*, **66**(3):1844-1850. [doi:10.1158/0008-5472.CAN-05-3466]

Moshkovskii, S.A., Serebryakova, M.V., Kuteykin-Teplyakov, K.B., Tikhonova, O.V., Goufman, E.I., Zgoda, V.G., Taranets, I.N., Makarov, O.V., Archakov, A.I., 2005. Ovarian cancer marker of 11.7 kDa detected by proteomics is a serum amyloid A1. *Proteomics*, **5**(14):3790-3797. [doi:10.1002/pmic.200401205]

Nemeth-Cawley, J.F., Tangarone, B.S., Rouse, J.C., 2003. "Top Down" characterization is a complementary technique to peptide sequencing for identifying protein species in complex mixtures. *J. Proteome Res.*, **2**(5):495-505. [doi:10.1021/pr034008u]

Nishizuka, S., Charboneau, L., Young, L., Major, S., Reinhold, W.C., Waltham, M., Kouros-Mehr, H., Bussey, K.J., Lee, J.K., Espina, V., *et al*., 2003. Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc. Natl. Acad. Sci. (USA)*, **100**:14229-14234.

Ong, S.E., Mann, M., 2005. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.*, **1**(5): 252-262. [doi:10.1038/nchembio736]

Orchekowski, R., Hamelinck, D., Li, L., Gliwa, E., van-Brocklin, M., Marrero, J.A., Vande Woude, G.F., Feng, Z., Brand, R., Haab, B.B., 2005. Antibody microarray profiling reveals individual and combined serum proteins associated with pancreatic cancer. *Cancer Res.*, **65**(23):11193-11202. [doi:10.1158/0008-5472.CAN-05-1436]

Ozier, O., Amin, N., Ideker, T., 2003. Global architecture of genetic interactions on the protein network. *Nat. Biotechnol.*, **21**(5):490-491.

Paweletz, C.P., Liotta, L.A., Petricoin, E.F., 2001. New technologies for biomarker analysis of prostate cancer progression: laser capture microdissection and tissue proteomics. *Urology*, **57**(4):160-163. [doi:10.1016/S0090-4295(00)00964-X]

Pawlik, T.M., Hawke, D.H., Liu, Y., Krishnamurthy, S., Fritsche, H., Hunt, K.K., Kuerer, H.M., 2006. Proteomic analysis of nipple aspirate fluid from women with early-stage breast cancer using isotope-coded affinity tags and tandem mass spectrometry reveals differential expression of vitamin D binding protein. *BMC Cancer*, **6**(1):68.

Petricoin, E.F., Zoon, K.C., Kohn, E.C., Barrett, J.C., Liotta, L.A., 2002a. Clinical proteomics: translating benchside promise into bedside reality. *Nature Rev. Drug Discov.*, **1**(9):683-695. [doi:10.1038/nrd891]

Petricoin III, E., Ardekani, A., Hitt, B., Levine, P., Fusaro, V., Steinberg, S., Mills, G., Simone, C., Fishman, D., Kohn, E., *et al*., 2002b. Use of proteomic patterns in serum to identify ovarian cancer. *The Lancet*, **359**(9306):572-577. [doi:10.1016/S0140-6736(02)07746-2]

Petricoin, E.F., Ornstein, D.K., Liotta, L.A., 2004. Clinical proteomics: applications for prostate cancer biomarker discovery and detection. *Urol. Oncol.*, **22**(4):322-328.

Phizicky, E., Bastiaens, P.I.H., Zhu, H., Snyder, M., Fields, S., 2003. Protein analysis on a proteomic scale. *Nature*, **422**(6928):208-215. [doi:10.1038/nature01512]

Pusztai, L., Gianni, L., 2004. Technology insight: emerging techniques to predict response to preoperative chemotherapy in breast cancer. *Nat. Clin. Pract. Oncol.*, **1**(1):44-50. [doi:10.1038/ncponc0025]

Ramus, C., de Peredo, A.G., Dahout, C., Gallagher, M., Garin, J., 2006. An optimized strategy for ICAT quantification of membrane proteins. *Mol. Cell Proteomics*, **5**(1):68-78. [doi:10.1074/mcp.M500205-MCP200]

Ricolleau, G., Charbonnel, C., Lode, L., Loussouarn, D., Joalland, M.P., Bogumil, R., Jourdain, S., Minvielle, S., Campone, M., Deporte-Fety, R., *et al*., 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**(6):1963-1975. [doi:10.1002/pmic.200500283]

Righetti, P.G., Castagna, A., Herbert, B., Reymond, F., Rossier, J.S., 2003. Prefractionation techniques in proteome analysis. *Proteomics*, **3**(8):1397-1407. [doi:10.1002/pmic.200300472]

Schrimpf, S.P., Meskenaite, V., Brunner, E., Rutishauser, D., Walther, P., Eng, J., Aebersold, R., Sonderegger, P., 2005. Proteomic analysis of synaptosomes using isotope-coded affinity tags and mass spectrometry. *Proteomics*, **5**(10):2531-2541. [doi:10.1002/pmic.200401198]

Shadforth, I.P., Dunkley, T.P., Lilley, K.S., Bessant, C., 2005. i-Tracker: for quantitative proteomics using iTRAQ. *BMC Genomics*, **6**(1):145. [doi:10.1186/1471-2164-6-145]

Shen, Y., Strittmatter, E.F., Zhang, R., Metz, T.O., Moore, R.J., Li, F., Udseth, H.R., Smith, R.D., Unger, K.K., Kumar, D., *et al*., 2005. Making broad proteome protein measurements in 1~5 min using high-speed RPLC separations and high-accuracy mass measurements. *Anal. Chem.*, **77**(23):7763-7773. [doi:10.1021/ac051257o]

Sidransky, D., 2002. Emerging molecular markers of cancer. *Nature Rev. Cancer*, **2**(3):210-219. [doi:10.1038/nrc755]

Staunton, J.E., Slonim, D.K., Coller, H.A., Tamayo, P., Angelo, M.J., Park, J., Scherf, U., Lee, J.K., Reinhold, W.O., Weinstein, J.N., Mesirov, J.P., *et al*., 2001. Chemosensitivity prediction by transcriptional profiling. *Proc. Natl. Acad. Sci. (USA)*, **98**:10787-10792.

Tyers, M., Mann, M., 2003. From genomics to proteomics. *Nature*, **422**(6928):193-197. [doi:10.1038/nature01510]

van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., *et al*., 2002. A gene-expression

signature as a predictor of survival in breast cancer. *N. Engl. J. Med.*, **347**(25):1999-2009. [doi:10.1056/NEJMoa021967]

van't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., *et al*., 2002. Gene-expression profiling predicts clinical outcome of breast cancer. *Nature*, **415**(6871):530-536. [doi:10.1038/415530a]

Wang, H., Clouthier, S.G., Galchev, V., Misek, D.E., Duffner, U., Min, C.K., Zhao, R., Tra, J., Omenn, G.S., Ferrara, J.L., *et al*., 2005. Intact-protein-based high-resolution three-dimensional quantitative analysis system for proteome profiling of biological fluids. *Molecular & Cellular Proteomics*, **4**(5):618-625. [doi:10.1074/mcp.M400126-MCP200]

Washburn, M.P., Ulaszek, R.R., Yates, J.R.3rd, 2003. Reproducibility of quantitative proteomic analyses of complex biological mixtures by multidimensional protein identification technology. *Anal. Chem.*, **75**(19):5054-5061. [doi:10.1021/ac034120b]

Wolters, D.A., Washburn, M.P., Yates, J.R.3rd, 2001. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.*, **73**(23):5683-5690. [doi:10.1021/ac010617e]

Wu, C.C., MacCoss, M.J., Howell, K.E., Yates, J.R.3rd, 2003. A method for the comprehensive proteomic analysis of membrane proteins. *Nat. Biotechnol.*, **21**(5):532-538. [doi:10.1038/nbt819]

Wu, W.W., Wang, G., Baek, S.J., Shen, R.F., 2006. Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. *J. Proteome Res.*, **5**(3):651-658. [doi:10.1021/pr050405o]

Wulfkuhle, J.D., Liotta, L.A., Petricoin, E.F., 2003. Proteomic applications for the early detection of cancer. *Nature Rev. Cancer*, **3**(4):267-275. [doi:10.1038/nrc1043]

Ye, B., Cramer, D.W., Skates, S.J., Gygi, S.P., Pratomo, V., Fu, L., Horick, N.K., Licklider, L.J., Schorge, J.O., Berkowitz, R.S., *et al*., 2003. Haptoglobin-alpha subunit as potential serum biomarker in ovarian cancer: identification and characterization using proteomic profiling and mass spectrometry. *Clin. Cancer Res.*, **9**(8):2904-2911.

Yu, K.H., Rustgi, A.K., Blair, I.A., 2005. Characterization of proteins in human pancreatic cancer serum using differential gel electrophoresis and tandem mass spectrometry. *J. Proteome Res.*, **4**(5):1742-1751. [doi:10.1021/pr050174l]

Zhang, Z., Bast, R.C.Jr, Yu, Y., Li, J., Sokoll, L.J., Rai, A.J., Rosenzweig, J.M., Cameron, B., Wang, Y.Y., Meng, X.Y., *et al*., 2004. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.*, **64**(16):5882-5890. [doi:10.1158/0008-5472.CAN-04-0746]

Ziauddin, J., Sabatini, D.M., 2001. Microarrays of cells expressing defined cDNAs. *Nature*, **411**(6833):107-110. [doi:10.1038/35075114]