Journal of Mammary Gland Biology and Neoplasia, Vol. 7, No. 4, October 2002 (© 2003)

Proteomics Approaches to Uncover the Repertoire of Circulating Biomarkers for Breast Cancer

Bong Kyung Shin,[1] Hong Wang,[1] and Samir Hanash[1,2]

There is substantial interest in applying proteomics to cancer biomarker discovery. Proteomics is particularly suited for profiling biological fluids and uncovering circulating markers and is currently being applied to the identification of novel biomarkers for breast cancer. Novel proteomic approaches include direct profiling of serum using mass spectrometry, application of a variety of strategies to harness immunity for cancer diagnosis, protein tagging to capture subproteomes rich in diagnostic markers, the use of protein microarrays, and the use of multidimensional liquid protein separation technologies for comprehensive profiling of serum and other biological fluids. It is likely that the application of a wide range of proteomics tools to breast cancer will yield a panel of markers that have utility for breast cancer diagnosis.

KEY WORDS: proteomics; cancer biomarker; breast cancer.

INTRODUCTION

Breast cancer is a relatively common malignancy in women, with an increasing incidence worldwide. Despite the wide use of mammography as a routine screening method for women over the age of 40, less than ~50% of breast cancers are localized at the time of diagnosis (1). There remains much need for reliable biomarkers for breast cancer diagnosis. Several genes or their products, including BRCA, p53, estrogen receptor, and HER2/neu, have been shown to have a predictive role in breast cancer. However, to date, none of the known genes involved in breast cancer pathogenesis have had demonstrable utility for early diagnosis of breast cancer. There remains a need to identify new, predictive biomarkers for breast cancer.

Presently, there is an intense interest in applying proteomics to cancer marker identification, breast cancer included. Proteomic approaches to this end include comparative analysis of protein expression in normal and tumor tissues to identify aberrantly expressed proteins that may represent novel markers, analysis of secreted proteins in cell lines and primary cultures, and direct serum protein profiling to uncover potential markers. Various tools from the classic 2D polyacrylamide gel electrophoresis (2D PAGE), to liquid chromatography and mass spectrometric analysis, Western blotting and immunodetection, and protein chips have been applied to biomarker development. This review emphasizes novel proteomic approaches and their application to the discovery of circulating biomarkers for breast cancer.

DIRECT PROFILING OF SERUM USING MASS SPECTROMETRY

A particular form of matrix-assisted laser desorption ionization (MALDI) referred to as

[1] Department of Pediatrics, University of Michigan, Ann Arbor, Michigan.

[2] To whom correspondence should be addressed at Department of Pediatrics, University of Michigan, 1150 West Medical Center Drive, MSRB 1, A 520, Ann Arbor, Michigan 48109-0656; e-mail: shanash@umich.edu.

Abbreviations used: 2D PAGE, two-dimensional polyacrylamide electrophoresis; MALDI, matrix-assisted laser desorption ionization; SELDI, surface-enhanced laser desorption; 2D, two-dimensional; RS, ribonucleic-acid-binding regulatory subunit; 1D PAGE, one-dimensional polyacrylamide electrophoresis; HLA, human leukocyte antigen; INF, interferon; TRA, tandem repeat array; LC, liquid chromatography.

1083-3021/03/1000-0407/0 © 2003 Plenum Publishing Corporation

surface-enhanced laser desorption ionization (SELDI) has been applied to the analysis of cancer patient sera to identify distinctive protein patterns that may be characteristic of different tumor types, including breast cancer (2). With this approach, microliter quantities of serum from many samples are directly applied to the surface of a protein-binding plate which can bind to a class of proteins of interest. The bound proteins are treated and analyzed by mass spectrometry. The mass spectra patterns obtained for different samples reflect the protein and peptide contents of these samples. Patterns that distinguish between cancer patients and normal subjects with remarkable accuracy have been reported for several types of cancers, including breast, ovary, prostate, and pancreatic cancers (2–5). To screen for potential tumor biomarkers, Li *et al.* used 103 serum samples from breast cancer patients with different clinical stages and 41 samples from healthy women as controls. Diluted serum samples were applied to immobilized metal affinity capture protein chips previously activated with metal ion. Proteins bound to the chelated metal were detected and analyzed with the ProteinChip Reader. A panel of three mass peaks was selected based on the optimal separation between stage 0–1 breast cancer patients and healthy controls. The same separation was observed using independent test data from stage II to stage III cancer patients. Bootstrap cross-validation demonstrated that a sensitivity of 93% for all cancer patients and a specificity of 91% for all controls were achieved by a composite index derived by multivariate logistic regression using the three selected peaks. Appealing aspects of such an approach include a reduced sample requirement relative to 2D gels and a high throughput. The major drawback of direct analysis of serum and other biological fluids by MALDI mass spectrometry is the preferential detection of proteins with a lower molecular mass and the difficulty in identifying the proteins corresponding to the masses observed.

AUTOANTIBODIES TO TUMOR ANTIGENS AS DIAGNOSTIC MARKERS

Molecular studies of precursor breast lesions have provided evidence for genetic alterations that occur early as part of tumor development (6). Although precursor lesions may have similarities in their expression profiles with invasive tumors, detection of circulating markers early during tumor development represents a substantial challenge. A promising approach to cancer marker identification harnesses the immune response that occurs against tumor tissue (7). There is substantial evidence for an immune response to cancer in humans, demonstrated in part by the identification of autoantibodies against a number of intracellular and surface antigens detectable in sera from patients with different cancer types. Harnessing immunity to identify novel cancer markers is attractive because the immune response occurs at an early stage during tumor development, at a time when the tumor may be otherwise undetectable. Analysis of serum for autoantibodies against tumor proteins has been accomplished in several ways. A large number of tumor antigens that elicit autoantibodies have been identified by screening gene expression libraries with patient sera (8,9). Proteomics has provided an alternative approach to expression library screening and is increasingly used for tumor antigen identification. The merit of a proteomic approach to tumor antigen identification is that it allows proteins and peptides in their modified states, as they occur in cells, to be analyzed for antigenicity. Given that proteins are subject to posttranslational modifications that may be immunogenic, notably glycosylation, antibodies to epitopes that result from such posttranslational modifications can be preserved and detected with proteomics techniques. A recent approach has relied on epitope mimicry through the use of a random peptide library to capture autoantibodies and identify corresponding antigens (10).

The standard proteomic tools of 2D gels and mass spectrometry have uncovered a large number of tumor antigens (11). With 2D gels, proteins in tumor cell lysates are first separated and then transferred onto membranes that are incubated with subject sera. Proteins that specifically react with antibodies in cancer patient sera are identified by mass spectrometry. This strategy was applied to the identification of new breast cancer markers by our group (12). Cellular proteins from a human breast cancer cell line were separated by 2D electrophoresis and transferred to membranes that were incubated with individual sera obtained from breast cancer patients or controls. Among the proteins that exhibited specific reactivity with antibodies in breast cancer sera were three proteins that migrated as neighboring spots with an estimated molecular weight of 25 kDa and different PI values, suggestive of three isoforms of the same protein (Fig. 1). Mass spectrometry identified them as RNA-binding protein regulatory subunit (RS) (NCBI accession number 2460318). The protein identity was confirmed using anti-RS rabbit polyclonal antibody.

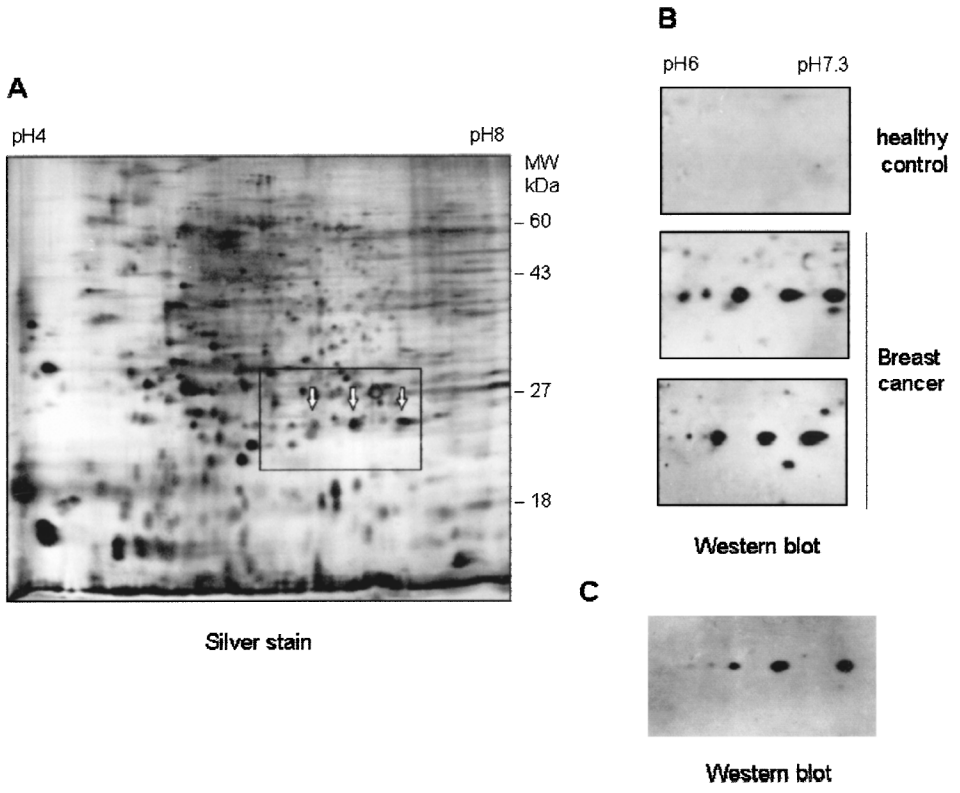

Fig. 1. Screening of autoantibodies and detection of circulating antigens in breast cancer. (A) Proteome profile of SUM-44 cells. The proteins from whole extracts of SUM-44 cells were separated by 2D PAGE and silver-stained. (B) Close-up section showing Western blot performed with lysates from SUM-44 cells and serum from healthy control or from breast cancer patient as primary antibody. An anti-human IgG was used as a secondary antibody. The three spots recognized by sera from patient with breast cancer are labeled with white arrows. They were identified as RS/DJ-1 protein by mass spectrometry and specific reaction with anti-RS and anti-DJ-1 antibodies. (C) Detection of RS/DJ-1 antigen in breast cancer patient sera. Sera containing the protein RS/DJ-1 were separated by 1D PAGE. A band corresponding to RS/DJ-1 was observed with high intensity in breast cancer patient sera.

Because in another previous study it had been suggested that RS was identical to the oncogene DJ-1 (13,14), a 2D Western blot was performed using an anti-DJ-1 antibody for confirmation. The same set of proteins was recognized with both anti-RS and anti-DJ-1 antibodies, confirming that RS was identical to DJ-1. Analysis of concentrated supernatant from the breast cancer cell line utilized for this study (SUM-44) revealed that RS/DJ-1 protein was a secreted protein. RS/DJ-1 was also detected in sera from patients with breast cancer by Western blot analysis using an anti-DJ-1 monoclonal antibody after separation of serum proteins with 1D PAGE. A strong band corresponding to the molecular weight of RS/DJ-1 antigen was observed in sera from 37% of newly diagnosed breast cancer patient sera, while none of the normal sera exhibited a comparably strong band (Fig. 1). Thus, RS/DJ-1 protein was identified as a possible novel circulating tumor antigen in breast cancers, though further studies are needed to determine the specificity of RS/DJ-1 antibodies and circulating antigens to breast cancer.

TUMOR CELL PROFILING TO UNCOVER POTENTIAL MARKERS

Several innovative strategies for comprehensive profiling of tumor cell subproteomes that are rich in tumor antigens are being implemented. An elegant proteomic approach consists of identification of HLA-restricted tumor-specific antigens by transfecting human tumor cell lines, including breast, with truncated genes of HLA-A2 and HAL-B7 (15). Secreted,

soluble HLA peptides were purified by affinity chromatography on pan-HLA class-I-reactive monoclonal antibody W6/32 columns. The peptides were resolved by reversed-phase HPLC and analyzed by an ESI-ion trap mass spectrometer. Identification of peptides was achieved by comparing their masses and collision-induced disintegration (CDI) spectra to the calculated mass and CDI data for peptides derived from all of the human proteins in the data bank. The identified peptides included those derived from γ-inducible protein, β-catenin, CD59, and many other proteins, which were distributed widely among different cancer cell lines from breast, ovary, and prostate cancers. Only a small fraction of peptides was detected in individual cancer types. Examples of unique peptides included a peptide with m/z ratio 947 (MUC1) recovered only from MCF-7 breast cancer cells grown without estrogen, a peptide with m/z 981 (protein tyrosine phosphatase) detected only in the breast cancer cell line MDA-231, and a peptide with m/z 922 (phospholipid transfer protein) only from PC3 prostate cancer cell line. Peptides derived from tumor-associated antigens uniquely presented on cancer cells and presented by HLA can be candidates for development of cancer vaccines. Also, once tumor-specific HLA peptides are identified and their ability to stimulate an immune reaction is confirmed, they can be candidates for adoptive immunotherapy. However, the potential usefulness of identified peptides should be confirmed by the presence of specific T cells against them in patients using standard assays such as cytotoxic T lymphocyte assay.

Since proteins on the tumor cell surface may be shed into circulation and may play a role as circulating markers or as tumor antigens, a strategy for comprehensive profiling of tumor cell surface proteins has been developed. Tumor proteins on the cell surface are of special interest, both as potential diagnostics and as therapeutic targets. For example, the discovery that the gene for the growth factor receptor HER2 is amplified in breast cancers and that its protein product is overexpressed on the cell surface has led to an effective form of therapy for breast cancer utilizing an antibody that targets HER2 (16). Tumor cell surface proteins that are restricted in their expression to specific cancer(s) or that undergo restricted modifications may be detected in circulation or may be utilized for antibody-based therapy, as in the case of HER2, or for vaccine development or other forms of immunotherapy. Recently, Shin *et al.* reported a strategy involving avidin–biotin interaction to purify surface membrane proteins for comprehensive profiling of the surface proteome of cancer cells (17). Their strategy relied on biotinylation of the surface proteins of cancer cells and the affinity capture of tagged proteins by an avidin column, followed by separation as intact proteins by 2D electrophoresis and subsequent identification by mass spectrometry. From analyses of lung, colon, ovary, and leukemia cancer cells, they could identify a subset of proteins showing restricted expression in some cancer(s), as well as more widely expressed proteins including a large number of chaperone proteins. Profiling the cell surface proteome of breast cancer cells resulted in the detection of several hundred proteins that were displayed in 2D gels (Fig. 2).

PROTEIN MICROARRAYS AS A DISCOVERY TOOL FOR THE IDENTIFICATION OF CANCER MARKERS

An alternative to 2D gels to harness the immune response for cancer marker discovery is protein microarray technology. Unlike the use of cDNA expression libraries (18–20) or phage-display libraries (21), which do not readily capture posttranslational modifications that may be restricted to tumor cells, a method that combines liquid phase protein separations of

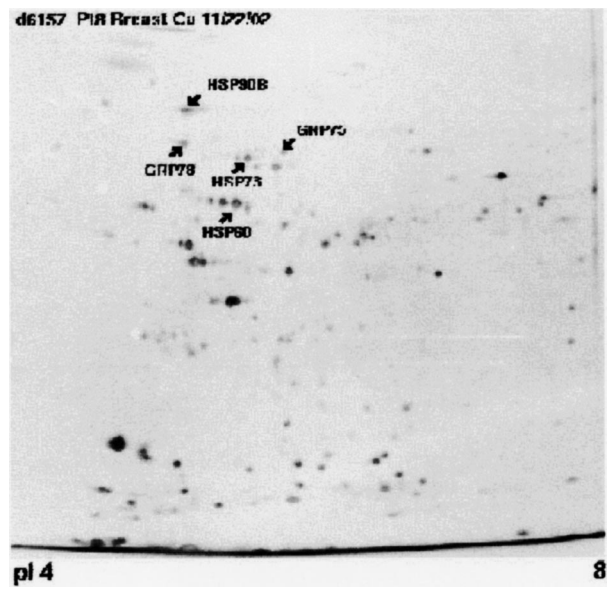

Fig. 2. Profiling of the cell surface proteome of breast cancer cells. Membrane proteins of freshly isolated breast cancer cells from pleural fluid were biotinylated, captured, and enriched by monomeric avidin column and separated by 2D PAGE. Arrows indicate the proteins identified by mass spectrometry.

tumor lysates with microarray technology allows natural proteins in their modified state to be arrayed directly and used as targets for a variety of probes (22). Proteins in cell and tissue lysates or in cellular subfractions can be separated into individual fractions using one of several different separation modes, including ion exchange lipid chromatography (LC), reverse phase LC, carrier ampholyte-based separations (e.g., Rotofor), or affinity-based separations. Each first-dimension fraction obtained by one separation mode can be further resolved using one or more of the other separation modes to yield either purified proteins in solution or liquid fractions with substantially reduced complexity. Liquid-based protein separations are well suited for fractionation of cell and tissue lysates. They provide high efficiency and resolution, and are compatible with automation and integration with a robotic arrayer, given that the proteins remain in a liquid phase throughout the separation. Tumor cell lysates resolved into several thousand individual fractions can be arrayed in less than 1-nL volumes onto specially treated glass slides and used for detection of specific antigens with known antibodies or for screening patient sera for detection of circulating autoantibodies against tumor antigens. Madoz-Gurpide *et al.* used Rotofor fractions obtained from a tumor cell line lysate for protein microarrays and reported that antigens in the range of hundreds of femtograms could be detected with fluorescence-labeled antibodies (22). This strategy is currently being implemented by our group for the detection of novel breast tumor antigens that induce an antibody response in subjects with breast cancer.

COMPREHENSIVE SERUM PROFILING USING MULTIDIMENSIONAL PROTEIN SEPARATION TECHNOLOGIES

Beside their application for the production of protein microarrays, multidimensional protein separation technologies provide a high-resolution analytical tool for the comparative analysis of complex proteomes, followed by the use of mass spectrometry for protein identification. We have implemented four dimensions of separation for serum analysis (23). Over 2000 distinct protein forms could be detected in 100 μL of serum. The multidimensional protein separation scheme is being combined with protein fluorescent tagging for comparative analysis (Fig. 3). One serum sample (or biological fluid) is labeled with 1-(5-carboxypentyl)-1-propylindocarbocyanine halide (Cy3) and another with *N*-hydroxy-succinimidyl ester and 1-(5-carboxypentyl)-1′-methylindodi-carbocyanine halide (Cy5) fluorescent dye (24). The labeled samples are mixed and separated in multiple dimensions and compared (Fig. 3). Interesting protein differences (e.g., between cancer serum and normal serum) are identified by mass spectrometry, although further investigation is necessary to correctly estimate the significance of the differences because protein composition of serum will be influenced by a lot of systemic changes.

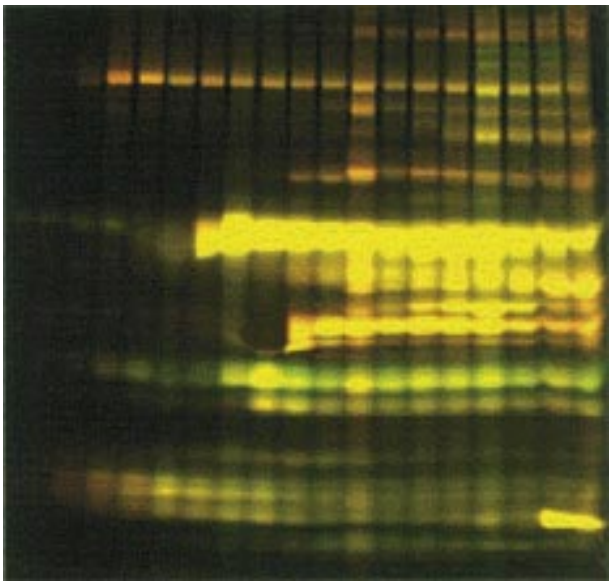

Fig. 3. Profiling of serum proteins by three-dimensional liquid phase separation technique. Sera from a single patient obtained before and after the onset of graft versus host disease were differently labeled with Cy5 (green) and Cy3 (yellow). The serum was first fractionated by Rotofor, then each fraction was further separated by high performance liquid chromatography (HPLC), and then separated by 1D PAGE. This figure represents a 1D PAGE gel containing 28 of the fractions obtained from a Rotofor fraction with pI range 5.2–5.4.

CONCLUDING REMARKS

Table I summarizes the proteomics techniques presented in this paper that have been applied for development of cancer biomarkers. It is likely that the application of novel proteomics tools to breast cancer will yield a panel of markers that have utility for breast cancer diagnosis. It is hoped that this panel includes markers that allow early diagnosis. Particularly promising in this regard is the identification of circulating antigens and/or their corresponding

Table I. Proteomics Tools Applied for Developing Novel Cancer Biomarkers

Proteomics tools	Reference	Potential candidates
SELDI	(2)	Three mass peaks with 93% sensitivity and 91% specificity for detection of breast cancer
2D PAGE and MS	(12)	Circulating RS/DJ-1 antigen and anti-RS/DJ-1 autoantibodies in breast cancer
HLA peptides profiling	(15)	p947 (MUC 1) and p981 (protein tyrosine phosphatase) in breast cancer p922 (phospholipid transfer protein) in prostate cancer
Surface membrane proteome profiling	(17)	Many proteins, including a number of chaperone proteins
Protein microarray using multidimensional protein separation	(22)	Under development
Comprehensive serum profiling using multidimensional protein separation	(23)	Under development

Note. Proteomics techniques presented in the paper are summarized with references and some biomarker candidates uncovered by them. Some methods, such as protein microarrays and comprehensive serum profiling, are very promising but need further work for detecting significant biomarker candidates.

autoantibodies, given that the immune response occurs early during tumor development. However, all of the technologies we have presented, if deployed, have the potential for generating informative markers.

REFERENCES

1. A. Jemal, A. Thomas, T. Murray, and M. Thun (2002). Cancer statistics. *CA. Cancer J. Clin.* **52:**23–57.
2. J. Li, Z. Zhang, J. Rosenzweig, Y. Y. Wang, and D. W. Chan (2002). Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin. Chem.* **48:**1296–1314.
3. A. M. Ardekani, L. A. Liotta, and E. F. Petricoin III (2002). Clinical potential of proteomics in the diagnosis of ovarian cancer. *Expert. Rev. Mol. Diagn.* **2:**312–320.
4. Y. Qu, B. L. Adam, Y. Yasui, M. D. Ward, L. H. Cazares, P. F. Schellhammer, Z. Feng, O. J. Semmes, and G. L. Wright, Jr (2002). Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. *Clin. Chem.* **48:**1835–1843.
5. C. Rosty, L. Christa, S. Kuzdzal, W. M. Baldwin, M. L. Zahurak, D. W. Chan, M. Carnot, K. D. Lillemoe, J. L. Cameron, C. J. Yeo, R. H. Hruban, and M. Goggins (2002). Identification of hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein I as a biomarker for pancreatic ductal adenocarcinoma by protein biochip technology. *Cancer Res.* **62:**1868–1875.
6. S. Krishnamurthy and N. Sneige (2002). Molecular and biologic markers of premalignant lesions of human breast. *Adv. Anat. Pathol.* **9:**185–197.
7. S. Hanash (2003). Harnessing immunity for cancer marker discovery *Nat. Biotechnol.* **21:**37–38.
8. E. Stockert, E. Jager, Y. T. Chen, M. J. Scanlan, I. Gout, J. Karbach, M. Arand, A. Knuth, and L. J. Old (1998). A survey of the humoral immune response of cancer patients to a panel of human tumor antigens. *J. Exp. Med.* **187:**1349–1354.
9. Y. Wang, K. J. Han, X. W. Pang, H. A. Vaughan, W. Qu, X. Y. Dong, J. R. Peng, H. T. Zhao, J. A. Rui, X. Y. Leng, J. Cebon, A. W. Burgess, and W. F. Chen (2002). Large scale identification of human hepatocellular carcinoma-associated antigens by autoantibodies. *J. Immunol.* **169:**1102–1109.
10. P. J. Mintz, J. Kim, K. A. Do, X. Wang, R. G. Zinner, M. Cristofanilli, M. A. Arap, W. K. Hong, P. Troncoso, C. J. Lonothetis, R. Pasqualini, and W. Arap (2003). Fingerprinting the circulating repertoire cancer patients. *Nat. Biotechnol.* **21:**57–63.
11. F. Le Naour (2001). Contribution of proteomics to tumor immunology. *Proteomics* **1:**1295–1302.
12. F. Le Naour, D. E. Misek, M. C. Krause, L. Deneux, T. J. Giordano, S. Scholl, and S. M. Hanash (2001). Proteomics-based identification of RS/DJ-1 as a novel circulating tumor antigen in breast cancer. *Clin. Cancer Res.* **7:**3328–3335.
13. Y. Hod, S. N. Pentyala, T. C. Whyard, and M. R. El-Maghrabi (1999). Identification and characterization of a novel protein that regulates RNA-protein interaction. *J. Cell. Biochem.* **72:**435–444.
14. D. Nagakubo, T. Taira, H. Kitaura, M. Ikeda, K. Tamai, S. M. Iguchi-Ariga, and H. Ariga (1997). DJ-1, a novel oncogene which transforms mouse NIH3T3 cells in cooperation with ras. *Biochem. Biophys. Res. Commun.* **231:**509–513.
15. E. Barnea, I. Beer, R. Patoka, T. Ziv, O. Kessler, E. Tzehoval, L. Eisenbach, N. Zavazava, and A. Admon (2002). Analysis of endogenous peptides bound by soluble MHC class I molecules: A novel approach for identifying tumor-specific antigens. *Eur. J. Immunol.* **32:**213–222.
16. D. J. Slamon, B. Leyland-Jones, S. Shak, H. Fuchs, V. Paton, A. Bajamonde, T. Fleming, W. Eiermann, J. Wolter, M. Pegram, J. Baselga, and L. Norton (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* **344:**783–792.
17. B. K. Shin, H. Wang, A. M. Yim, F. Le Naour, F. Brichory, J. H. Jang, R. Zhao, E. Puravs, J. Tra, C. W. Michael, D. E. Misek, and S. M. Hanash (2003). Global profiling of the cell surface proteome of cancer cells uncovers an abundance of proteins with chaperone function. *J. Biol. Chem.* **278:**7607–7616.
18. K. Bussow, D. Cahill, W. Nietfeld, D. Bancroft, E. Scherzinger, H. Lehrach, and G. Walter (1998). A method for global protein

expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids Res.* **26:**5007–5008.
19. A. Lueking, M. Horn, H. Eickhoff, K. Bussow, H. Lehrach, and G. Walter (1999). Protein microarrays for gene expression and antibody screening. *Anal. Biochem.* **270:**103–111.
20. H. Zhu, J. F. Klemic, S. Chang, P. Bertone, A. Casamayor, K. G. Klemic, D. Smith, M. Gerstein, M. A. Reed, and M. Snyder (2000). Analysis of yeast protein kinases using protein chips. *Nat. Genet.* **26:**283–289.
21. C. A. Rowe, L. M. Tender, M. J. Feldstein, J. P. Golden, B. D. MacCraith, J. J. Cras, and F. S. Ligler (1999). An array immunosensor for simultaneous detection of clinical analytes. *Anal. Chem.* **71:**433–439.
22. J. Madoz-Gurpide, H. Wang, D. E. Misek, F. Brichory, and S. M. Hanash (2001). Protein based microarrays: A tool for probing the proteome of cancer cells and tissues. *Proteomics* **1:**1279–1287.
23. H. Wang and S. M. Hanash (2003). Multi-dimensional liquid based separations in proteomics. *J. Chromatogr. B* **787:**11–18.
24. W. F. Patton (2002). Detection technologies in proteome analysis. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **771:**3–31.