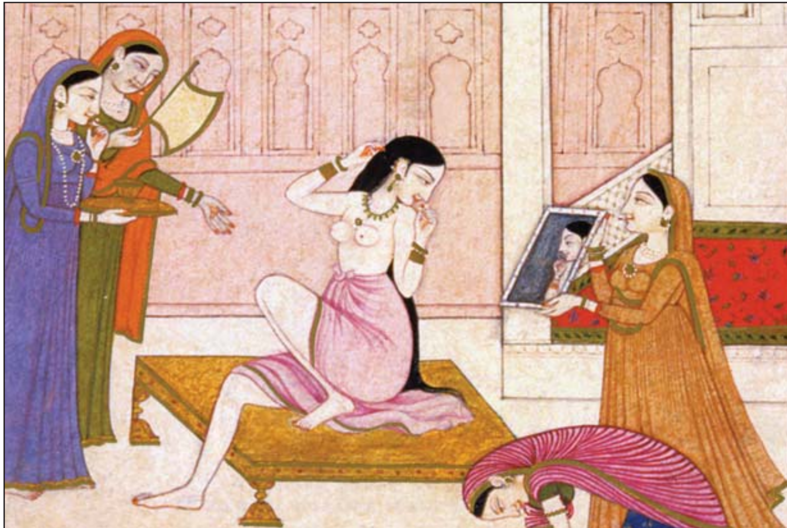

A royal lady admiring her youthful beauty. Kangra miniature, circa 18th century. Reproduced by the courtesy of the Bharat Kala Bhavan, Banaras Hindu University.

Proteomic studies can clarify the heterogeneity of breast cancer and expand on advances gained from genomics research.

Proteomic Approach to Breast Cancer

Christine Laronga, MD, FACS, and Richard R. Drake, PhD

Background: *Breast cancer is the most common cancer affecting women worldwide. Despite tremendous advances in screening, diagnosis, and treatment, the causes of this disease remain elusive and complex. Proteomics is a rapidly developing field that can explore the heterogeneity of breast cancer and supplement the wealth of information gained from genomics.*

Methods: *This article serves as an overview of the application of matrix-assisted laser desorption/ionization source with a time-of-flight (MALDI-TOF) proteomic techniques as applied to breast cancer. Examples of the clinical applicability of MALDI-TOF mass spectrometry are provided but represent only a fraction of the potential uses yet to be discovered. In addition, a brief summary of the bioinformatics issues that surround proteomics is included.*

Results: *Mass spectrometry has provided new proteomic approaches to unravel the complexities of clinical specimens relevant to breast cancer diagnostics. In particular, MALDI-TOF mass spectrometry analysis has been used to differentiate cancer profiles from benign profiles in samples from sera, plasma, tissue, nipple fluid, and ductal lavage. Some discriminating proteins have subsequently been identified.*

Conclusions: *Mass spectrometry applications to breast cancer diagnostics continue to be developed but are evolving faster than bioinformatics/statistical analysis can adapt. The future of these techniques in terms of clinical investigation is limitless, but in terms of general applicability, these applications are currently cost-prohibitive.*

From the Department of Surgery (CL), and the Department of Microbiology and Molecular Cell Biology (RRD) at the Eastern Virginia Medical School, Norfolk, Virginia.

Submitted February 20, 2007; accepted June 29, 2007.

Address correspondence to Christine Laronga, MD, FACS, Comprehensive Breast Program, H. Lee Moffitt Cancer Center & Research Institute, 12902 Magnolia Drive, MCC-BRPROG, Tampa, FL 33612. E-mail: Christine.Laronga@moffitt.org

No significant relationship exists between the authors and the companies/organizations whose products or services may be referenced in this article.

The editor of Cancer Control, John Horton, MB, ChB, FACP, has nothing to disclose.

Dr Laronga is now with the Comprehensive Breast Program at the H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida.

Abbreviations used in this paper: MALDI-TOF = matrix-assisted laser desorption/ionization source with a time-of-flight, SELDI = surface-enhanced laser desorption/ionization, MS = mass spectrometry, WCX = weak cation exchange.

Introduction

Breast cancer remains one of the leading cancers diagnosed in women worldwide. Great strides have been made over the last 20 years in early detection and treatment of breast cancer as evidenced by the earlier stage distribution at diagnosis and improved overall survival. With increased awareness and public education, screening mammography according to recommended guidelines is achieved in 65% to 70% of the US population.[1] Computer-assisted detection (CAD) and digital imaging have been pioneered in the last few years to improve cancer detection on mammograms by the radiologist. Additionally, technical advances in breast ultrasonography and breast MRI provide physicians with a full armamentarium of supplemental diagnostic imaging modalities for earlier detection of abnormalities found on clinical breast examination or mammography. Once an abnormality is identified, a tissue biopsy is performed that traditionally involved a surgical procedure. However, minimally invasive core biopsies now offer equal diagnostic accuracy without the need for surgery and its inherent risks and costs.

Once breast cancer is diagnosed, today's woman has many treatment choices. Mastectomy with an axillary node dissection is no longer her only surgical option. Breast conservation and sentinel lymph node biopsies can allow for accurate staging of her disease and equivalent survival to mastectomy with axillary node dissection.[2,3] Chemotherapy with hormone therapy was introduced as medical management for breast cancer and was shown to be beneficial in terms of overall survival, disease-free survival, and time to progression. Now a whole host of different chemotherapeutic agents as well as hormonal agents are available to choose from, depending on the woman's clinical and pathologic data. Even more recently, we have seen the dawn of targeted therapies for breast cancer such as monoclonal antibodies against the HER-2/neu receptor and vascular endothelial growth factor (VEGF).[4]

During this same timeframe mutations in two breast cancer genes, BRCA-1 and BRCA-2, were discovered and account for 5% to 10% of all breast cancer cases.[5] Genetic testing can now be recommended to women at increased risk for being a BRCA-1/BRCA-2 mutation carrier. If found to be genetic mutation carriers, these women can be offered chemoprevention or prophylactic surgeries for risk reduction.

Despite these advances in breast cancer detection and treatment, we seem no closer to eradicating the disease, perhaps because breast cancer is more complex than the current clinicopathologic parameters allow us to appreciate. We require a better understanding of the subtle nuances differentiating cancers in one woman from another and the evolution of a normal breast cell into a cancer cell. The fields of genomics and proteomics are just beginning to investigate the possibilities of individualized molecular signatures (combinations of individual molecular markers) as a way to better delineate these nuances. Proteomics has several advantages over genomics in that (1) proteins are more reflective of the current state of the cell's microenvironment, (2) mRNA levels do not necessarily correlate to corresponding protein levels, (3) proteins can undergo a vast cadre of posttranslational modifications affecting protein stability, localization, interaction, and functions, and (4) proteins represent more accessible and relevant therapeutic targets.[6] As such, many studies are underway using various proteomic approaches for detection of breast cancer or differentiation of breast cancer nuances (such as racial differences, gene mutation vs sporadic disease, failure of therapy, and development of local or distant recurrences). These studies include utilization of sera or plasma, nipple aspirate, ductal lavage, fine-needle aspiration, and breast tissue. This article reviews various proteomic approaches utilized in breast cancer and discusses some of the preliminary findings to date.

Protein Biomarkers and Mass Spectrometry

The premise for this type of proteomics is that all cells have unique identifiable characteristics (signatures) related to their role in the body and that during transformation into cancer cells, the signature changes. This change then becomes a unique "fingerprint" of the presence and character of cancer. The ability to detect this unique "fingerprint" could facilitate earlier detection and treatment of tumors, thereby affecting patient outcome. Although genomic analysis would provide insight into genetic predisposition for neoplastic development, it does not pinpoint the timeframe when the signature changes. A proteomic approach gives a more accurate real-time evaluation of the current status of the cells or tissue and thus reflect when the "fingerprint" changes. We also know that most cancer biomarkers are proteins (the ultimate regulators of cell function and the largest target for therapeutic interventions). This approach can identify these "fingerprint" proteins from a plethora of sources such as sera, nipple aspirates, ductal lavage samples, and even breast tissues. By comparing the "fingerprint" profiles of cancerous to noncancerous samples, we can assess their clinical utility for early detection of breast cancer, prognosis, response to treatment, risk of recurrence, and identification of potential therapeutic targets.

This was traditionally done using gel electrophoresis and column chromatography separations of proteins present in clinical samples. More recently, these separation methods have been coupled in some form to mass spectrometry analysis of proteins and peptides. The studies include analysis of protein-protein interactions

via affinity-based isolations on a small and proteome-wide scale, the mapping of numerous organelles, and the generation of quantitative protein profiles from diverse species.[6] Mass spectrometers share at least three common features: an ionization source, a mass analyzer, and a detector. For analysis of proteins, particularly from clinical samples, the two most commonly used mass spectrometers involve electrospray-ionization source coupled to an ion-trap mass analyzer (ESI-MS) and a matrix-assisted laser desorption/ionization (MALDI) source with a time-of-flight (TOF) mass analyzer. Because ESI-MS involves the generation of multiply-charged ions in solution prior to mass analysis, interpretation of the data can be complicated without extensive up-front purification and can be more complex in terms of protein-dense clinical samples. The MALDI-TOF approach is more amenable to the higher throughput analysis of many clinical specimens, as the laser desorption process results in primarily single ion species, allowing for a profiling approach of multiple species. While ESI-MS is a critical and useful tool in the analysis of proteins from clinical samples, we have focused our review primarily on the applications of the MALDI-TOF approach in relation to developing new diagnostic or surveillance assays for breast cancer.

ProteinChip SELDI Mass Spectrometry

A relatively high-throughput and once-popular approach to mass spectrometry-based proteomics is surface-enhanced laser desorption/ionization (SELDI) ProteinChip mass spectrometry (ProteinChip, Ciphergen BioSystems, Fremont, Calif) a derivation of MALDI that is based on protein chip arrays (Fig 1).[7,8] Modular, interlocking, finger-length strips are coated with an active chemical substance with classical chromatographic properties. In this way proteins incubated with the chip surfaces adhere based on solid-phase ion exchange and immobilized metal chelation. In addition, "biologic" chip arrays with preactivated surfaces allow covalent immobilization of antibodies, receptors, DNA, glycoproteins, etc, which provide for affinity capture of molecules in the sample. Energy-absorbing matrix molecules are subsequently overlaid with the bound proteins. The chip is placed in the SELDI mass reader, it is irradiated with a laser pulse (337 mm), the UV-absorbing matrix ions are ejected, carrying with them the retained protein molecules (analytes). The gas phase proton transfer that occurs between matrix and analytes mainly produces singly charged analyte ions. The ions generated are analyzed by a TOF mass spectrometer (MS) that separates them based on velocity. The TOF spectrum for different ions is recorded and converted into a mass spectrum. Each array has 100 addressable regions, and each addressable region is irradiated multiple times by the laser. With the aid of SELDI software, a retentate map is generated depicting the mass/charge, which corresponds to the molecular weight. Spectra are combined or compared to elucidate changes in protein profiles ("fingerprints") between different samples, such as cancer vs normal.

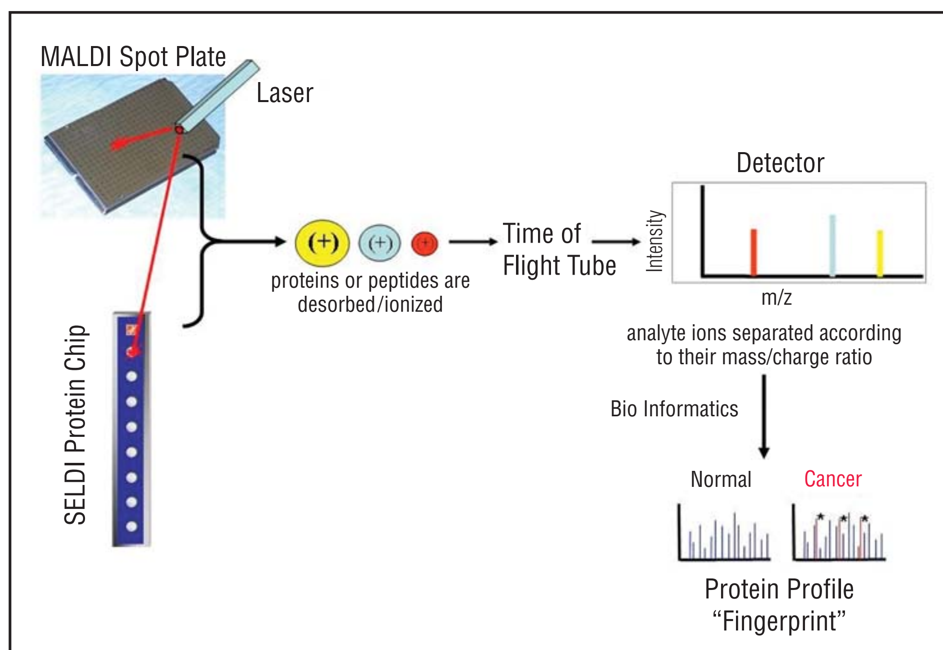

Fig 1. — Schematic illustrating mass spectrometry via MALDI-TOF and SELDI-TOF.

Previous Studies

SELDI-TOF Protein Profiling for the Detection of Cancer

Extensive studies have been conducted by our laboratory at Eastern Virginia Medical School (EVMS) and others to test the efficiency of the SELDI technology in protein profiling.[7-12] The proteomics laboratory at EVMS was the first to show that SELDI-TOF protein profiling of serum coupled with a learning classification algorithm could effectively differentiate prostate cancer from benign disease and healthy men.[9] One of the criticisms of SELDI-TOF technology has been reproducibility on a large scale (most studies to date have utilized small sample sizes) and reproducibility between

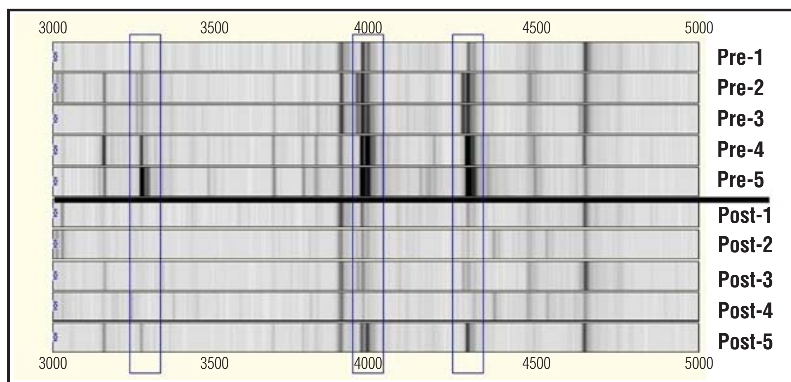

Fig 2. — Spectral analysis of sera samples from 5 women with breast cancer obtained before any treatment (pre 1–5) and 6 to 9 months after surgery (post 1–5) but currently receiving no therapy.

institutions. To that end, EVMS has been a lead coordinating group involved with successful studies that demonstrated reproducibility and correct classification of the same samples analyzed at six separate sites.[10] The first successful study for analysis of breast cancer at EVMS utilized the IMAC-Cu chip and included sera samples from 50 patients with breast cancer (stages 0–III), 42 patients with benign breast diseases, and 47 normal women.[12] In the cross-validation analysis, the specificity in cancer diagnosis was 90% with a sensitivity of 93%. These results led to a series of four pilot studies investigating subset populations within the breast cancer cohort to identify utility in the realm of detection and treatment: (1) women with BRCA-1 gene mutations vs women with sporadic breast cancer, (2) pre- and posttreatment women, (3) women who had a sentinel node biopsy as part of their surgical treatment, and (4) most recently, racial differences between Caucasian and African American women.

In the first study, involving women with BRCA-1 gene mutations vs women with sporadic breast cancer, the aim was to determine if sera profiles from all breast cancers were alike. Sporadic breast cancer is a hodgepodge of different histologic types with varying tumors sizes, node status, and receptor status but increasing in incidence with age. BRCA-1 gene mutations account for approximately 5% of all breast cancers, are mostly estrogen- and progesterone-negative, and are found in young women (<50 years of age). Pretreatment serum was collected from women newly diagnosed with sporadic breast cancer and compared with that of women known to be BRCA-1 mutation carriers. At the time of their serum draw, the BRCA-1 carriers were not yet diagnosed with breast cancer. They were followed for 7 years and then the samples were retrospectively labeled "cancer" or "carrier" as a result of the interim development of breast cancer. SELDI-TOF analysis revealed differentially expressed proteins ($P<.05$) between the BRCA-1 cancer women, BRCA-1 carriers without cancer, and women with sporadic breast cancer such that 13 of 15 BRCA-1 cancers vs BRCA-1 carriers were correctly identified (sensitivity/specificity of 87% and 87%), and 14 of 15 BRCA-1 cancers vs sporadic breast cancers were correctly identified (sensitivity/specificity 94% and 100%).[13]

In the second study, involving pretreatment and posttreatment women, paired serum samples were obtained from women with breast cancers before and after surgery (6 to 9 months) (Fig 2).[14] Our questions were simple: If a woman has a cancer profile and the cancer is eradicated, shouldn't her "fingerprint" profile convert back to normal? If it does convert, how long does it take? In 14 of 16 posttreatment patients, SELDI-TOF protein profiles could distinguish these samples from the pretreatment samples. When compared to serum samples from normal women, 11 of these posttreatment samples retained global protein profiles not found in healthy women. Thus, the profile conversion is not immediate, and for the majority of women this progression towards "normalcy" is greater than 9 months. Representative results from this study are summarized in Figs 2 and 3.

The sentinel lymph node biopsy study evaluated the ability of SELDI-TOF to distinguish between the serum profiles from sentinel lymph node positive and sentinel lymph node negative patients.[14] The goal of a

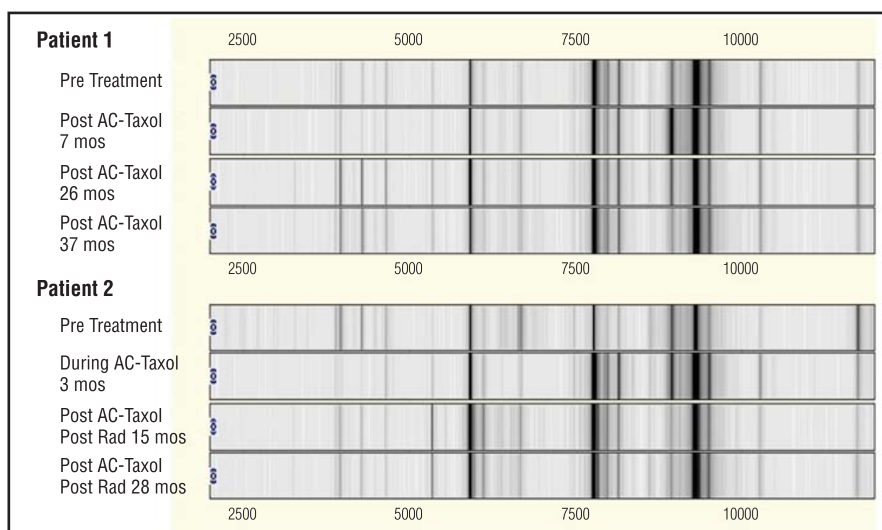

Fig 3. — Spectral analysis of sera samples from 2 women with breast cancer obtained before any treatment and longitudinally after surgery. Follow-up samples were drawn during and after doxorubicin/cyclophosphamide and paclitaxel (AC-Taxol) chemotherapy ± radiation therapy.

sentinel lymph node biopsy is to provide accurate staging of the woman's breast cancer, which offers important prognostic information. The benefits of sentinel lymph node biopsy over traditional axillary node dissection are plentiful but not without any risk of morbidity. Additionally, the false-positive rate (identifying the wrong node) is still approximately 5%. A serum test that could not only minimize the false-negative rate or potentially eliminate the operative morbidity of a sentinel lymph node biopsy but also maintain staging accuracy would be most welcome. Our results showed that in the sentinel lymph node positive samples, 22 of 27 were correctly classified, while in the sentinel lymph node negative samples, 55 of 71 were correctly classified. The lower sensitivity and specificity may be in small part related to the false-negative rate at our institution (approximately 6%), but for the most part, SELDI-TOF was not equivalent to and did not improve our staging of the axilla over the sentinel lymph node biopsy itself.

Detection of Breast Cancer Biomarkers in Plasma

Plasma can be used instead of sera for SELDI-TOF analysis. In a study we conducted at EVMS in collaboration with the M.D. Anderson Cancer Center, proteomic changes were examined to determine response to chemotherapy.[15] Sixty-nine women receiving either 5-fluorouracil, doxorubicin, and cyclophosphamide (FAC) or paclitaxel chemotherapy had their plasma drawn just before initiation of therapy and again 3 days later. Plasma was also drawn concurrently from healthy volunteers to determine if any cyclical variations occur in protein profiles. A single chemotherapy-inducible peak and 5 other peaks were able to distinguish plasma obtained from breast cancer women from the normal healthy volunteers.

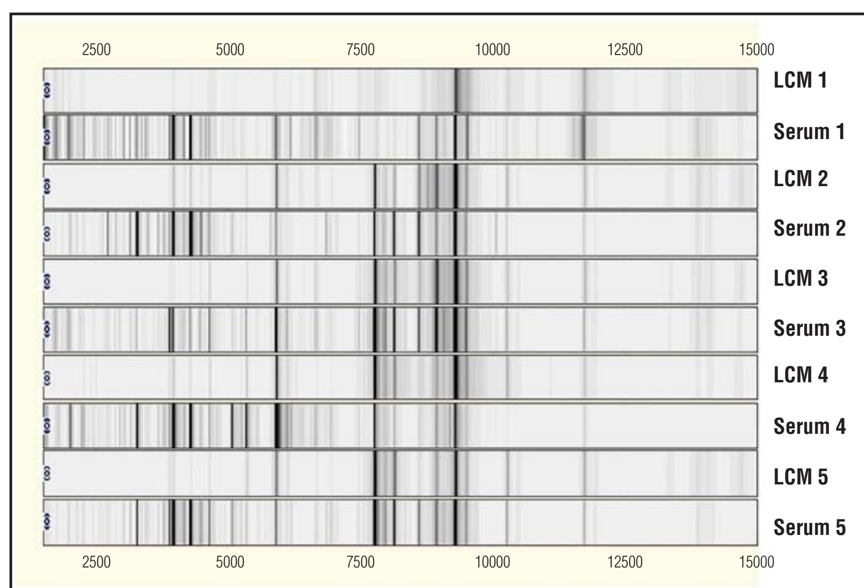

Fig 4. — Spectral analysis of sera samples from 5 women with breast cancer and their corresponding laser capture microdissected (LCM) breast tumor cell lysates.

Detection of Breast Cancer Biomarkers in Tissue

Sera and plasma are more plentiful and readily available for proteomic research but are also more complex biological specimens. One is looking for tumor biomarkers or host response proteins to a tumor in a background of potentially competing medical states. Utilization of breast tissue may eliminate some of these competing factors. A pilot study using laser-captured microdissected ductal epithelial cells was compared to sera from 16 Caucasian and 19 African American women with stage I/II infiltrating ductal cancer (Fig 4).[16] The samples were not matched for age, body mass index, or receptor status due to the small sample size. However, SELDI-TOF analysis of the cell lysates was able to differentiate between the races with a sensitivity and specificity of 79% and 81%, respectively. Surprisingly, the corresponding sera showed even better separation with a sensitivity and specificity of 80% and 88%, respectively. This study needs to be repeated on a large sample size and matched for as many factors (eg, age, body mass index) before any conclusions can be drawn about racial proteomic differences in breast cancer. If differences can be identified, then treatment choices might be better tailored for each race.

Detection of Breast Cancer Biomarkers in Nipple Aspirate/Ductal Lavage

The majority of breast cancers originate in the ductal system of the breast, and nipple fluid is a window to the underlying microenvironment. Eight to 10 milk ducts exit the nipple, and nipple aspirates can be obtained using various noninvasive techniques in the majority of women. To obtain larger volumes of ductal epithelial cells, ductal lavage can be performed using small microcatheters to cannulate individual milk ducts. Not all ducts can be cannulated, and thus the lavage sample is representative of only the portion of the breast from which that ductal tree arises. In a pilot study, ductal lavage fluid was collected from 6 normal volunteers and 2 high-risk women for a total of 13 samples.[17] High-risk women were defined as having a Gail risk greater than 1.67% at 5 years but no personal history of breast cancer. The lavage specimens were spun down and separated into a supernatant fluid portion and the cell pellet portion. Four SELDI chip surfaces were initially evaluated for feasibility using the lavage fluid and lysates of the cell pellets. All chip surfaces were able to generate protein profiles with reproducibility between the same ducts but variability between ductal trees of the same breast or same patient (right

and left breasts) or different patients. The cell pellets had a high homology rate (80%) with the corresponding lavage fluid. This study demonstrated the ability of SELDI to generate protein profiles from ductal lavage specimens including the fluid phase while also showing the heterogeneity of the breast microenvironment. A follow-up study was performed in collaboration with Emory University using ductal lavage samples from 12 high-risk women and 18 women with documented breast cancer.[18] All women had both breasts lavaged if nipple aspirate fluid was identified. Supernatant samples were applied to a chip surface with weak cation exchange (WCX) binding affinity. Chip selection was based on the fact that four chip surfaces were used in the aforementioned pilot study with the WCX chip surface yielding the most discriminatory results. SELDI-TOF mass spectrometry was then performed and a decision tree algorithm was generated using Biomarker Patterns Software (Bio-Rad Laboratories Inc, Hercules, Calif) for the classification of the high-risk and cancer groups. Three protein peaks distinguished samples as cancer or high-risk. Cross validation studies showed a cancer diagnosis sensitivity and specificity of 81% and 91%, respectively.

Ductal lavage is invasive, requires technical expertise to perform, and is costly. Although nipple aspiration may produce lower volumes, it is noninvasive, inexpensive, and easier to obtain. Several proteomic studies have been performed on nipple aspiration fluid (NAF). One example is the study by Pawlik et al[19] in which isotope-coded affinity tag (ICAT) tandem mass spectrometry was performed on paired NAF samples from 18 women with early-stage breast cancer and 4 healthy volunteers. A total of 353 peptides were identified with equal numbers of peptides overexpressed and underexpressed in the cancer-bearing breasts, such as overexpression of vitamin D-binding protein in the cancer breasts. In this technology, samples are isotope-labeled with tags for ^{12}C (light) and ^{13}C (heavy) that bind covalently to cysteine residues.[7] In addition to ICAT tandem mass spectrometry, Pawlik et al[20] have performed SELDI-TOF analysis of NAF using WCX and IMAC-Cu chips on 23 women with early-stage breast cancer and 5 healthy volunteers. A total of 463 distinct peaks were identified, but no differences could be found between the right and left breasts of the same woman despite one of the two breasts containing cancer. However, differences were seen between NAF samples from the healthy volunteers vs women with breast cancer regardless of whether the NAF sample was from the cancer-containing breast or the contralateral uninvolved normal breast. Their results suggest a more general global "fingerprint" change reflected equally in both breasts when a cancer arises in one breast. This interesting observation was not seen in the other aforementioned NAF or ductal lavage studies that demonstrated a heterogeneous milieu within the ductal systems of each breast.

The Next Generation: Bead-Based Expression Profiling by MALDI-TOF

The high-throughput capabilities and surface capture utility of SELDI-TOF for complex clinical samples has spurred the development of more sensitive MALDI-TOF instrumentation coupled with derivatized-magnetic bead surfaces. The magnetic beads allow robotic processing and automation of the samples onto the MALDI spot plates. MALDI spot plates differ from SELDI chips in that they generally contain 384 spots per steel plate (with roughly the same dimensions as a standard 96 well plate). The other difference is that the beads provide the surface capture component and are separate from the spot plate that is inserted into the instrument. The beads provide a scalability function to allow more of the sample to be interrogated than the defined SELDI chip surface area. Like the SELDI chip surface, these beads can be derivatized with multiple chemical affinity properties or biological capture components, like antibodies. Eluted samples from the beads are dried on the plate, coated with matrix, and irradiated with a laser for mass analysis by TOF, as described for the SELDI process. The peak spectra generated are analyzed with the same types of algorithms that have been applied to SELDI data.

Detection of Breast Cancer Biomarkers in Serum

Using the cumulative experience of our previous SELDI studies, we next designed a study around women with a score of 4 on a Breast Imaging Reporting and Data System (BIRADS) mammogram report that, if successful, would have more direct clinical utility, and analysis would be done with the more sensitive MALDI-TOF bead-based approach. Our BIRADS 4 study was designed to select women with a BIRADS 4 mammogram who were about to have a breast biopsy and obtain prebiopsy serum. The BIRADS system is a reporting system used nationally to standardize the interpretation of mammograms. A score of 4 means that a suspicious abnormality is identified and a biopsy should be considered. Of all the women undergoing biopsy, 80% will have a benign histology diagnosis. A blood test obtained prior to biopsy and subjected to expression profiling analysis may help discriminate those women who are more likely to have benign disease and therefore not warrant a tissue biopsy. The study was opened, and during the accrual phase, robotics was introduced into the laboratory to eliminate human error in sample processing, minimize consumption of samples, and increase throughput. After 14 months, a sufficient number of samples had been accrued for an initial pilot study.

For this BIRADS 4 study, 46 cancer serum samples and 46 benign samples were incubated with WCX magnetic beads using a ClinProt robotic platform (Bruker Daltonics Inc, Billerica, Mass).[21] The bound proteins were eluted and spotted on an AnchorChip sample tar-

get platform. Spectra were generated on an UltraFlex MALDI-TOF and normalized with the ClinProt software version 2.0. A *k*-nearest neighbor genetic algorithm contained in this software was used to identify statistically significant differences in protein peaks between the cancer and benign samples. The MALDI-TOF analysis yielded 273 peaks, of which 14 were differentially expressed. The most robust model used 12 peaks for a sensitivity of 88.3% for correct cancer classification and a specificity of 85.8% for classification of benign disease. As shown in Fig 5, the most significant peak was at 8,949 m/z (mass/charge ratio). Studies are underway to identify this protein as a potential biomarker to distinguish benign from cancerous lesions.

Another study has also used serum and MALDI-TOF for breast cancer discrimination, and in this case, C8 magnetic beads were used to discriminate breast cancer from controls.[22] In this study, preoperative serum was obtained from 78 cancer patients and 29 controls. Linear discriminant analysis was used to classify the protein profiles. A total recognition rate of 99%, a sensitivity rate of 100%, and a specificity rate of 97% were achieved for the cancer cohort. The area under the curve for the first classifier in this analysis was 98.3%, suggesting that the classification could be attributed to actual information in the protein profiles rather than to chance. A different study using C8 beads and MALDI-TOF was done with multiple sera cancer cohorts, including breast cancers.[23] In this study, 61 separate low-mass peptides were determined to be differentially expressed in the different cancer cohorts and healthy normal controls. These peptides were isolated and sequenced and their identities determined. These proteins were all fragments of much larger serum proteins, and evidence of extensive proteolysis by exoproteases and endoproteases in serum were demonstrated.[23]

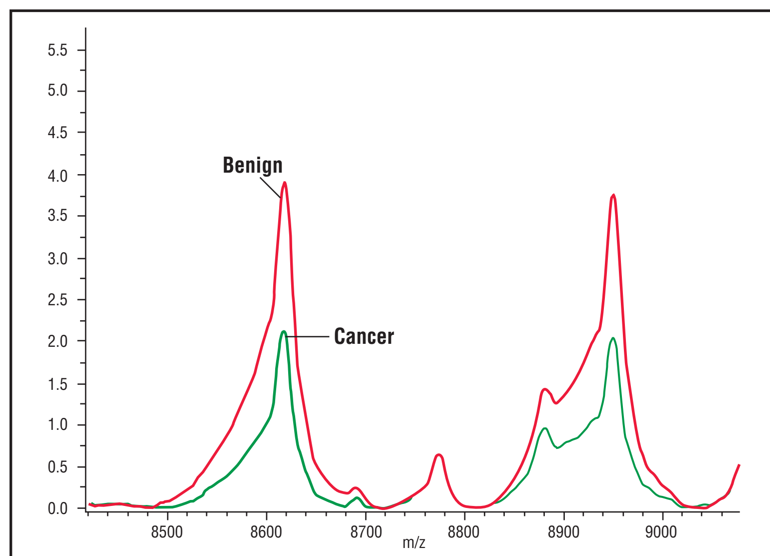

Fig 5. — The averaged spectral analysis of all sera samples from women with breast cancer and benign breast disease in the BIRADS 4 study.

Instead of using beads, MALDI-TOF can also be readily coupled to gel-based separations using 2-D electrophoresis. In 2-D electrophoresis, the proteins are separated according to two independent physicochemical parameters — isoelectric focusing and molecular weight.[7] The separated proteins on the gel can be cut out of the gel and subjected to MALDI-TOF. Dual fluorescence in-gel electrophoresis (2-D DIGE) is an improvement over conventional 2-D electrophoresis and uses covalently labeled fluorescent dyes such as Cy3 and Cy5. This allows quantification of the protein based on differential expression of the dyes and reduces gel-gel variability. An example is a study by Huang et al[24] in which serum samples from 39 cancer patients and 35 controls were used. They found that proapolipoprotein A-1, transferrin, and hemoglobin were upregulated in cancer and that apolipoprotein A-1, apolipoprotein C-III, and haptoglobin alpha2 were downregulated in cancer. Where isoform-specific antibodies exist, they were able to confirm their results by clinically available immunoassays.

MALDI-MS Imaging and Detection of Breast Cancer Biomarkers in Tissue

Proteomic analysis of cells derived directly from breast cancer tissues represents an increasingly attractive target as methods of isolation improve and sensitivities of mass spectrometers increase. Two recent studies highlight this trend and also illustrate the extension of MALDI-TOF applications to breast cancer tissues.[25,26]

One newly emerging application is that of MALDI-MS imaging of frozen tissue samples prepped directly on MALDI plates.[25] In an application to breast tumor tissues, a robotic spotter was used to deposit micron-sized droplets of matrix directly onto normal epithelial cells, stromal areas, ductal carcinoma in situ cells, and invasive cancer cells at defined locations. The spectra generated from each defined matrix spot were compared to each other, as well as the underlying pathology of the cell type present at each spot. Instead of being represented as peak heights, an individual m/z value is assigned a pixel intensity and color. Therefore, the larger the peak intensity at a given m/z, the brighter the pixel color in the image at that location. All of the many different cell types present in a heterogeneous breast tumor tissue can thus be profiled simultaneously for each individual patient, and all of these data are linked integrally to any pathology and cytology assessments. For breast cancers, this MALDI-MS imaging approach has the potential to improve biopsy diagnoses, better define tumor margins, and aid in multiple prognostic and treatment decision-making processes.

Laser capture microscopy (LCM) of cells within breast tumor tissues is also a viable method for identifying differentially expressed tumor biomarkers. Cornett et al[26] described a sample preparation method that generates tryptic peptides for profiling derived from small numbers of laser micro-dissected breast cancer cells. This approach had two advantages over previous studies. The generation of peptides as the profiled material requires less cell numbers than those used in most LCM proteomic studies, and the tryptic peptides can be directly sequenced on the MALDI-TOF, providing protein identification of differentially expressed peaks.[26]

Bioinformatics

The aforementioned studies highlight the different sources (eg, sera, plasma) of samples used and some of the avenues of investigation in which proteomics might have applicability in breast cancer. These studies were introductions rather than comprehensive reviews of this exploding field. The majority of these studies were designed to be exploratory and, due to small sample size, have limited clinical utility. In fact, questions of reproducibility and reliability of peak quantifications, overall lack of sensitivity and specificity of tumor specific fingerprints, and handling of the extensive amount of data generated by MALDI-TOF/SELDI-TOF analysis have ignited the development of a specialized field of bioinformatics, novel algorithms to process the spectra, and new analytical models to analyze the data.

Due to the large number of mass variables (thousands) in unprocessed MS spectra compared to the number of subjects (hundreds), the selection of a few disease-specific biomarkers or features presents a statistical challenge. Preprocessing steps such as baseline removal, normalization, alignment, denoising, and peak detection are important to the reproducibility of data results. These important issues were recently summarized by Hilario et al.[27] Additionally, supervised feature selection has emerged as a necessary dimension reduction step for proteomic profiling data. Intelligent reduction of variables is required to avoid overfitting and statistical instability that can lead to false discoveries and irreproducible results. Previous use of unsupervised "black-box" analysis, whereby experimental design and signal processing were decoupled from statistical pattern recognition, plagued some early TOF-MS survey studies with a lack of reproducibility.[28,29] Most proteomic profiling studies reported use of default instrumental algorithms for signal processing to detect peaks, followed by identification of diagnostic patterns using standard statistical and pattern recognition tools. These pattern recognition tools include support vector machines for analysis of variance,[11,23] regression trees,[9,13] *t* tests,[15,30] and neural networks.[21,22,31] We have worked closely with mathematicians and biostatisticians who have developed multiple approaches to dimension reduction and signal processing of MALDI-TOF serum profiles.[32,33] Their efforts have focused on minimizing instrumental noise from the spectra; once these noise signatures are removed, the variables are reduced from tens of thousands of m/z values to hundreds. Other groups have described use of an undecimated or discrete wavelet transform to denoise spectra and improve reproducibility between spectra from the same sample.[34,35] Overall, these noise and feature reduction steps make the data more applicable to the strengths of most standard classification strategies mentioned above. Lastly, the need for common data models and standardized software infrastructure to enable more efficient access and sharing of distributed computational resources in cancer research has been recognized by the National Cancer Institute. To address this, a national effort termed the cancer Biomedical Informatics Grid (caBIG) was launched to develop an accessible network of research information systems across the United States.[36]

Conclusions

For the first time in years, the incidence of breast cancer seems to be on a decline, but the disease is far from cured. The fields of genomics and proteomics are showing promise in unraveling some of the complexities of cancer that have eluded researchers for centuries. Both fields are still in their infancy and have yet to reach their full potential. Proteomics has several advantages over genomics in that proteins are good markers of the current state of the cells in a particular person. Proteins make good targets for specific therapies, much like antibiotics. Physicians commonly treat first with a broad spectrum antibiotic until culture results have been obtained. More targeted antibiotics can then be chosen with better efficacy and less chance of resistance to other families of antibiotics. Similarly, proteomic-based assays can be easily coupled to other functional clinical tests.

Mass spectrometry techniques, such as MALDI-TOF and SELDI-TOF, allow for differentiation and classification of samples for a given disease state. In breast cancer, initial studies were aimed at distinguishing benign sera, plasma, tissue, nipple fluid, or ductal lavage from its cancerous counterparts. By generating proteomic profiles of a given sample, the computer can be trained to recognize the differences between cancer and benign states. This technology is akin to the UPC barcodes on items we buy at a store. The scanner can recognize the difference between a can of green beans and the box of cereal because it has already been taught what the codes for those items are. Since breast cancer is heterogeneous, we must

advance the technologies to better discriminate within the cancer cohort, much like the UPC code can correctly identify different brands of cereal. Pilot studies in that vein are underway (eg, BRCA-1/2 vs sporadic breast cancer, node-positive vs node-negative, pretreatment vs posttreatment). Moving these studies to the next level will require large sample allocations, standardization of sample processing and storage conditions, calibration of instrumentation between institutions, development of software to store and analyze large complex datasets, adaption of statistical methods specifically for proteomics technology, and finally, a large investment of money.

These technologies can lead to identification of discriminatory proteins that can then be used as markers for screening (as is the hope for the BIRADS 4 study[21]), diagnosis, prognosis, monitoring of treatment response, interval development of local or distant recurrences, and design of new molecularly targeted therapies personalizing and tailoring treatment of breast cancer. In the future, biomarker proteins might be identified that represent the initiation of transition of a benign cell to a malignant cell. If those proteins can be identified and drug therapies are created to reverse or halt the process before it becomes cancer, then breast cancer can finally be eradicated. These technologies hold that potential and promise.

References

1. National Breast and Cervical Cancer Early Detection Program: Screening Program Data. Department of Health and Human Services. Centers for DiseaseControl and Prevention. Available at http://www.cdc.gov/cancer/nbccedp/data/. Accessed on July 26, 2007.
2. Fisher B, Anderson S, Redmond CK, et al. Reanalysis and results after 12 years of follow-up in a randomized clinical trial comparing total mastectomy with lumpectomy with or without irradiation in the treatment of breast cancer. *N Engl J Med*. 1995;333:1456-1461.
3. Kelley MC, Hansen N, McMasters KM. Lymphatic mapping and sentinel lymphadenectomy for breast cancer. *Am J Surg*. 2004;188:49-61.
4. Altundag K, Esteva FJ, Arun B. Monoclonal antibody-based targeted therapy in breast cancer. *Curr Med Chem Anticancer Agents*. 2005;5:99-106.
5. Chen S, Iversen ES, Friebel T, et al. Characterization of BRCA1 and BRCA2 mutations in a large United States sample. *J Clin Oncol*. 2006;24:863-871.
6. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003;422:198-207.
7. Bertucci F, Birnbaum D, Goncalves A. Proteomics of breast cancer: principles and potential clinical applications. *Mol Cell Proteomics*. 2006;5:1772-1786. Epub 2006 May 29.
8. Wright GL. SELDI proteinchip MS: a platform for biomarker discovery and cancer diagnosis. *Expert Rev Mol Diagn*. 2002;2:549-563.
9. Adam BL, Qu Y, Davis JW, et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res*. 2002;62:3609-3614.
10. Semmes OJ, Feng Z, Adam BL, et al. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. assessment of platform reproducibility. *Clin Chem*. 2005;51:102-112.
11. Li J, Zhang Z, Rosenzweig J, et al. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem*. 2002;48:1296-1304.
12. Vlahou A, Laronga C, Wilson L, et al. A novel approach toward development of a rapid blood test for breast cancer. *Clin Breast Cancer*. 2003;4:203-209.
13. Becker S, Cazares LH, Watson P, et al. Surfaced-enhanced laser desorption/ionization-time of flight (SELDI-TOF) differentiation of serum protein profiles of BRCA-1 and sporadic breast cancer. *Ann Surg Oncol*. 2004;11:907-914. Epub 2004 Sep 20.
14. Laronga C, Becker S, Watson P, et al. SELDI-TOF serum profiling for prognostic and diagnostic classification of breast cancers. *Dis Markers*. 2003-2004;19(4-5):229-238.
15. Pusztai L, Gregory BW, Baggerly KA, et al. Pharmacoproteomic analysis of prechemotherapy and postchemotherapy plasma samples from patients receiving neoadjuvant or adjuvant chemotherapy for breast carcinoma. *Cancer*. 2004;100:1814-1822.
16. Shin S, Fournier K, Schneider H, et al. Racial disparities in protein profiles of breast cancer ductal epithelial cells/sera. Presented at the Society of Surgical Oncology 58th Annual Cancer Symposium. Atlanta, Georgia. March 3-6, 2005.
17. Fournier K, Vlahou A, Gregory B, et al. Application of SELDI-TOF mass spectrometry to breast ductal lavage fluid. Presented at the American Association of Cancer Research 94th Annual Meeting. Washington, DC. July 11-14, 2003.
18. Fournier K, Mendrinos S, Ritchie J, et al. Feasibility of protein evaluation of ductal lavage fluid for breast diseases. Presented at the 26th Annual San Antonio Breast Cancer Symposium. San Antonio, Texas. December 3-6, 2003.
19. Pawlik TM, Hawke DH, Liu Y, et al. 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*. 2006;6:68.
20. Pawlik TM, Fritsche H, Coombs KR, et al. Significant differences in nipple aspirate fluid protein expression between healthy women and those with breast cancer demonstrated by time-of-flight mass spectrometry. *Breast Cancer Res Treat*. 2005;89:149-157.
21. Shin S, Cazares L, Schneider H, et al. Serum biomarkers to differentiate benign and malignant mammographic lesions. *J Am Coll Surg*. 2007;204:1065-1071; discussion 1071-1073.
22. de Noo ME, Deelder A, van der Werff M, et al. MALDI-TOF serum protein profiling for the detection of breast cancer. *Onkologie*. 2006;29:501-506. Epub 2006 Oct 25.
23. Villanueva J, Shaffer DR, Philip J, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest*. 2006;116:271-284.
24. Huang HL, Stasyk T, Morandell S, et al. Biomarker discovery in breast cancer serum using 2-D differential gel electrophoresis/maldi-TOF/TOF and data validation by routine clinical assays. *Electrophoresis*. 2006;27:1641-1650.
25. Umar A, Dalebout JC, Timmermans AM, et al. Method optimisation for peptide profiling of microdissected breast carcinoma tissue by matrix-assisted laser desorption/ionisation-time of flight and matrix-assisted laser desorption/ionisation-time of flight/time of flight-mass spectrometry. *Proteomics*. 2005;5:2680-2688.
26. Cornett DS, Mobley JA, Dias EC, et al. A novel histology-directed strategy for MALDI-MS tissue profiling that improves throughput and cellular specificity in human breast cancer. *Mol Cell Proteomics*. 2006;5:1975-1983. Epub 2006 Jul 18.
27. Hilario M, Kalousis A, Pellegrini C, et al. Processing and classification of protein mass spectra. *Mass Spectrom Rev*. 2006;25:409-449.
28. Petricoin EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*. 2002;359:572-577.
29. Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics*. 2004;20:777-785. Epub 2004 Jan 29.
30. Mathelin C, Cromer A, Wendling C, et al. Serum biomarkers for detection of breast cancer: a prospective study. *Breast Cancer Res Treat*. 2006;96:83-90.
31. Hu Y, Zhang S, Yu J, et al. SELDI-TOF-MS: the proteomics and bioinformatics approaches in the diagnosis of breast cancer. *Breast*. 2005;14:250-255.
32. Malyarenko DI, Cooke WE, Adam BL, et al. Enhancement of sensitivity and resolution of surface-enhanced laser desorption/ionization time-of-flight mass spectrometric records for serum peptides using time-series analysis techniques. *Clin Chem*. 2005;51:65-74. Epub 2004 Nov 18.
33. Malyarenko DI, Cooke WE, Tracy ER, et al. Resampling and deconvolution of linear time-of-flight records for enhanced protein profiling. *Rapid Commun Mass Spectrom*. 2006;20:1670-1678.
34. Coombes KR, Tsavachidis S, Morris JS, et al. Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. *Proteomics*. 2005;5:4107-4117.
35. Du P, Kibbe WA, Lin SM. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*. 2006;22:2059-2065. Epub 2006 Jul 4.
36. Saltz J, Oster S, Hastings S, et al. caGrid: design and implementation of the core architecture of the cancer biomedical informatics grid. *Bioinformatics*. 2006;22:1910-1916. Epub 2006 Jun 9.