Published in final edited form as:

J Proteome Res. 2008 April ; 7(4): 1500–1507. doi:10.1021/pr7008109.

Differentiating Proteomic Biomarkers in Breast Cancer by Laser Capture Microdissection and MALDI MS

Melinda E. Sanders^{*,†,○}, Eduardo C. Dias[‡], Baogang J. Xu^{§,†}, James A. Mobley^{§,†}, Dean Billheimer^{||}, Heinrich Roder[⊥], Julia Grigorieva[⊥], Mitchell Dowsett[#], Carlos L. Arteaga^{‡,¶,○}, and Richard M. Caprioli^{*,§,○,†}

[†] Department of Pathology, Vanderbilt University, Nashville, Tennessee 37232

[○] Breast Cancer Research Program, Vanderbilt University, Nashville, Tennessee 37232

[‡] Department of Medicine, Vanderbilt University, Nashville, Tennessee 37232

[§] Department of Biochemistry, Vanderbilt University, Nashville, Tennessee 37232

⁺ Mass Spectrometry Research Center, Vanderbilt University, Nashville, Tennessee 37232

^{||} Department of Statistics, Vanderbilt University, Nashville, Tennessee 37232

[⊥] Biodesix, Inc. Steamboat Springs, Colorado 80477

[#] Royal Marsden Hospital and Institute of Cancer Research, London, U.K.

[¶] Department of Cancer Biology, Vanderbilt University, Nashville, Tennessee 37232

Abstract

We assessed proteomic patterns in breast cancer using MALDI MS and laser capture microdissected cells. Protein and peptide expression in invasive mammary carcinoma versus normal mammary epithelium and estrogen-receptor positive versus estrogen-receptor negative tumors were compared. Biomarker candidates were identified by statistical analysis and classifiers were developed and validated in blinded test sets. Several of the *m/z* features used in the classifiers were identified by LC-MS/MS and two were confirmed by immunohistochemistry.

Keywords

MALDI MS; laser capture microdissection; breast cancer

© 2008 American Chemical Society

*To whom correspondence should be addressed. Melinda E. Sanders, M.D., Vanderbilt University Medical Center, 23rd and Pierce Ave., 4918-B TVC Bldg., Nashville, TN 37232. Phones, 615-322-1410(office), 615-343-9060(secretary); fax, 615-343-9563; e-mail, melinda.sanders@vanderbilt.edu. Richard M. Caprioli, Ph.D., Mass Spectrometry Research Center, 465 21st Avenue South, 9160 MRB-III, Vanderbilt University School of Medicine, Nashville, TN 37232, Phone, 615-322-4336, fax, 615-343-8372, e-mail, r.caprioli@vanderbilt.edu..

Supporting Information Available: (1) Detailed protein identification methods, (2) reproducibility analyses, (3) interinstitutional sample variability, (4) variability of samples with age, (5) complete list of features defined for the IMC vs NME comparison, (6) subset of features used for the IMC vs NME classifier, (7) IMC vs NME classification LOOCV error depending on selection of *k*-value for the KNN algorithm, (8) complete list of features defined for the ER+ vs ER- comparison, (9) subset of features used for the ER+ vs ER- classifier, (10) ER+ vs ER- classification LOOCV error depending on selection of *k*-value for the KNN algorithm, (11) detailed protein identification data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Introduction

Breast cancer is the leading cause of cancer the USA for women and ranks second in cancer deaths, with an estimated 182 460 new cases of invasive cancer, 67 770 cases of noninvasive breast cancer, and 40 480 deaths in 2008. Rather than one disease, it is a heterogeneous group of neoplasms, some of which are locally aggressive and may metastasize early, while other forms proliferate slowly and may be cured by surgical excision alone. Among the subset of “special type” carcinomas¹ an excellent prognosis may be indicated by histology alone (e.g., Pure tubular carcinoma); however, these represent less than 15% of all breast cancers. The majority of “no special type” (aka. ductal carcinomas) have by definition no distinctive features. Clinical decision making including the need for systemic adjuvant therapy²⁻⁵ is currently based on a combination of estrogen (ER) and progesterone (PR) receptor status and expression levels, presence or absence of Her2-neu gene amplification, tumor size, grade, proliferative rate, and stage.⁶

Retrospective patient analyses including gene expression profiling suggest that differences in intrinsic biology of individual tumors have important implications for therapy and prognosis and that these differences are often not discernible on a histological basis. Subsequent predictors of prognosis in breast cancer based on cDNA expression⁷⁻¹⁰ have been developed, some of which are in use in clinical trials.¹¹⁻¹³ However, one would expect there ultimately to be limits to their predictive power because mRNA expression is poorly correlated with the functional protein component. Accordingly, proteomics which studies the active mediators of cellular processes is a required complement to gene expression analyses. Proteomic expression among breast cancer subtypes is largely unexplored and should prove to be an important complement to microarray studies and an excellent mechanism for further understanding these different phenotypes.

Matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS) can profile proteins at high sensitivity up to 50 kDa in tissues.¹⁴ This technology can directly measure many peptides and proteins in tumor tissue sections and can also be used for high resolution imaging of individual biomolecules present in tissue sections.¹⁵⁻¹⁷ Coupled with laser capture microdissection (LCM), MALDI MS is an ideal approach for generation of separate protein profiles of the invasive tumor and normal epithelial components of breast tumors and tissues. In addition, epithelial elements usually compose only 5–15% of normal breast tissue making LCM mandatory in most cases to ensure a dominantly epithelial sample for evaluation. We aimed to use MALDI MS to assess protein expression profiles in approximately 2000 cells from frozen sections of surgically resected breast tumors and reduction mammoplasty tissue, and to assess the resulting data using ProTS Marker software (Biodesix, Inc., Steamboat Springs, CO). The goal of this project was to provide a distinctive protein profile of each tumor and assess the ability of our analysis algorithms to classify the tumors into ER-positive and ER-negative subgroups based on differences in these patterns. Recent results have shown that MALDI MS-based diagnostics can be highly reproducible across different laboratories, and may overcome some of the ambiguities arising from other techniques.¹⁸

Experimental Procedures

Tissue Collection and Evaluation

A total of 122 invasive mammary carcinomas (IMC) and normal mammary epithelium (NME) from 167 reduction mammoplasty specimens were analyzed in this study. These samples were derived from 289 women. These tissue samples were collected and distributed to our laboratory in a deidentified fashion by the four divisions of the Cooperative Human Tissue Network and the Royal Marsden Institute, U.K. (M.D.). Table 1. details the distribution of clinical and pathologic characteristics across centers. None of these women had received preoperative

hormonal, chemo-, or radiation therapy. These tissues were obtained at the time of the woman's primary surgery, snap-frozen in liquid nitrogen within 30 min after removal from the patient, and stored at -80°C until analyzed. The presence of tumor or NME was confirmed by a board-certified pathologist who specializes in breast disease (M.E.S.) who examined a frozen section of each tissue block and subsequently performed LCM on appropriate areas.

Tissue Sample Preparation

Sections for microdissection were prepared according to our previously developed protocol.¹⁹ In brief, using a cryostat, $7\text{ }\mu\text{m}$ frozen tissue sections were mounted on uncharged glass slides without the use of embedding media and placed immediately in 70% ethanol for 1 min. Subsequent dehydration was achieved using graded alcohols and xylene treatments as follows: 95% ethanol for 30 s (2 times), 100% ethanol for 30 s (2 times), and xylene for 5 min (2 times). Slides were then dried in a laminar flow hood for 10 min prior to microdissection.

Laser Capture Microdissection

LCM was performed using the PixCell IIe LCM system (Arcturus, Mountain View, CA). Depending on the size of the lesion, 500–1000 shots using the 7.5 or $15\text{ }\mu\text{m}$ infrared laser beam were utilized to obtain an average of 2000 cells. All samples were microdissected in duplicate.

Preparation of Microdissected Cells for MALDI MS

MALDI MS was performed directly on the LCM acquired cells. After LCM, the thermoplastic film was removed from the LCM cap using forceps and placed onto the MALDI plate using conductive double-sided tape. A finely pulled glass capillary was employed to deposit as little as 10 nL of matrix solution as required to cover the captured cells under microscopic visualization. The matrix solution consisted of sinapinic acid at 20 mg/mL in 6/4/0.01 (v/v/v) acetonitrile/water/TFA.

MALDI MS Analysis

MALDI MS analysis was performed using a Voyager DE-STR MALDI time-of-flight mass spectrometer (Applied Biosystems, Framingham, MA) with a 337 nm nitrogen laser. Acquisition was achieved in the linear positive ion mode under optimized delayed extraction conditions as described previously.^{14,15,20–24} Approximately 750 laser shots were averaged to create a single spectrum from the captured cells. In most cases, we generated three spectra per sample which were then combined to create one average spectrum which was used in the statistical analysis. In a subset of cases where the number of tumor cells was very small, we were able to generate only one spectrum. In this analysis, signals in the mass-to-charge (m/z) range from 2000 to 30 000 Da were considered.

Immunohistochemistry

Estrogen receptor (ER) and progesterone receptor (PR) expression were evaluated using the Zymed PREDILUTED ER antibody 6F11 (S., San Francisco, CA) and the DAKO PgR636 PR antibody (Carpinteria, CA) each at a 1:100 dilution and incubated for 1 h at room temperature. Immunohistochemistry was performed on frozen sections from the same tissue block on which LCM was performed. The results were scored as the percentage of tumor cells with nuclear staining. Tumors were considered to be ER-positive or PR-positive when $>10\%$ nuclear staining was observed at any intensity.

Data Processing and Statistical Analysis

1. Spectral Pre-processing—Preprocessing of spectra is necessary to render spectra comparable and to ensure the reproducibility of the statistical analysis procedure.

Preprocessing was performed using proprietary analysis tools developed by Biodesix, Inc., CO. A detailed description can be found at <http://www.biodesix.com> in the technology section. Raw spectra were sent to Biodesix (Steamboat Springs, CO) for analysis. Mass spectra generated over 2 years time, although run by the same personnel and on the same instrument, may exhibit variation. To enable analysis of these spectra, we applied a suite of preprocessing procedures²⁵⁻²⁹ and developed some additional procedures. In brief, the background and noise were estimated and then subtracted from each spectrum based on local noise estimators using a local (in m/z) robust asymmetric estimator²⁵⁻²⁹ and normalized to total ion current (TIC). Peaks were detected using a signal-to-noise ratio (S/N) cutoff of 4.0, which was found to be a good compromise between over detection and sensitivity.

Spectra were aligned sequentially in three steps. The set of common peaks appear in a plurality of spectra whose m/z values differ by $< \pm 5$ Da was used for the alignment process using a polynomial up to quadratic terms. The sets of points for the alignment were selected based on the criteria of communality (present in at least 2/3 of all spectra), S/N ($S/N > 5$ in the first alignment, $S/N > 4$ in the second and third alignment), and roughly even distribution along the m/z range. The following set of common peaks was used in the first phase of the alignment process: m/z 3373.5, 4939.7, 5359.1, 5654.4, 6548.4, 7670.7, 8570.2, 10091.7, 10839.3, 11309.6, 11349.3, 11650.0, 12346.6, 13781.1, 14005.6, 15346.4, 15860.2, 17885.1, 17926.1. Spectra were saved after the first alignment, reloaded, aligned again using a second set of peaks, and saved (m/z 3449.0, 4939.8, 5360.7, 5653.9, 6175.2, 6546.9, 6650.0, 6890.2, 7004.7, 8091.9, 9154.7, 10093.9, 10841.5, 11306.4, 113450.0, 11653.78, 12347.9, 13780.2, 14009.9, 15342.6, 17893.3, 20945.4). Replicate spectra after the second alignment were used to create an average spectrum for each individual LCM sample. The third alignment was performed on the averaged spectra using a third set of alignment peaks (m/z 4047.0, 4937.7, 5358.3, 5652.8, 5939.5, 6175.8, 6277.6, 6547.7, 6665.1, 6890.6, 7005.8, 8413.1, 8568.4, 9155.0, 9971.6, 10094.7, 10142.6, 10843.6, 11309.1, 11653.4, 12349.0, 12652.4, 13780.1, 14010.0, 15341.1, 15867.9, 17897.5, 20758.4, 26600.1). The preprocessing procedure was optimized using the training set and held fixed for the classification of all test sets.

2. Feature Definition and Selection—Each MALDI spectrum is characterized by a set of features, which are defined as integrated, background-subtracted, and normalized spectral intensities integrated over a chosen m/z range containing a peak. The m/z range for each feature was calculated from the alignment error and the local peak width of each spectrum. The features were predefined from a set of peaks that were common (within a predefined tolerance of 0.5 Da) to at least 3 spectra of each clinical group. A combination of a selected subset of features and of the algorithm (and its parameters), which assigns a clinical label to a spectrum, constitutes a classifier. Candidate features for the classification algorithms were identified as differentially expressed m/z values from spectra from IMC versus NME and ER-positive versus ER-negative tumors. The initial selection of significant features for the classifier was based on a calculation of univariate p -values from Mann-Whitney U-tests (Wilcoxon rank sum test). Then a variant of a floating search method was applied.³⁰ The floating search looks iteratively at combinations of the significant features to see how they perform as a classifier. As an optimization criterion, we used the leave-one-out cross validation (LOOCV) error on the training set. Finally, a visual inspection of features was carried out using the graphs of the group averaged spectra in the ProTS Marker software. In some cases, the feature widths were manually adjusted during the training process to take asymmetric peak shapes into account.

3. Classification Algorithm—The classification algorithm used was a straightforward implementation of a k -nearest-neighbor (KNN) algorithm.³¹ KNN requires as parameters a set of representative and labeled instances (i.e., list of selected feature values from a training sample set). Samples from each clinical group of interest (IMC vs NME and ER-positive vs ER-negative tumors) were randomly split into training and test sets. Aligned average spectra

were used for each sample instance, both in the training and in the test sets. To classify a new spectrum, the KNN algorithm first calculates the Euclidean distance of the feature values of the new spectrum to those of the training set spectra. This calculation yields a list of distances from the test spectrum to each representative spectrum. For the k -nearest neighbors (those with the k smallest distances), the labels are compared. Finally, the assigned label is a simple majority vote over the k -nearest neighbor labels. The number “ k ” of the nearest neighbors was the only classifier parameter used.

4. Cross-Validation of a Classifier and Validation—Cross-validation is an important tool in the assessment of classifier performance during the training phase. A fixed number (one for LOOCV or a prescribed number N for LNOCV) of instances was removed from the training set, a classifier was generated using the remaining instances, and the performance of this classifier was evaluated by applying it to the left-out instances and comparing classifier and true labels. Ideally, this analysis is performed for all possible permutations of left-out and kept instances in the training set. The average of the classification performance over these permutations is an estimate of the expected performance of the classification. Classifier parameters and the set of selected features were optimized using these cross-validation procedures. After training and optimization, all parameters were frozen. No changes in the classification algorithm were allowed during the validation performed on the independently and randomly selected test spectra.

5. Assignment of Training and Test Groups—Average spectra of IMC samples (122 instances) and NME samples (167 instances) were split into test and training sets. A random numbers generator was used to assign each sample a number, then all samples were sorted by their numbers; the first 62 IMC instances and 84 NME instances were assigned to the training set and used for generation of a classifier to distinguish IMC from NME. The remaining IMC and NME spectra were assigned to the test set. The same approach was used to randomly split the group of 117 IMC spectra with known ER status into ER+ and ER− training and test sets. The training set consisted of 36 ER+ and 25 ER− tumors, while the test set consisted of the remaining 32 ER+ and 24 ER− tumors.

Protein Identification

Once candidate molecular weight markers were selected by the class prediction model, we utilized a combination of MS techniques previously validated by the VUMSRC^{15,32} to isolate and identify several of the protein species of interest. The remainder of each specimen used for LCM was kept frozen to enable subsequent protein extraction and identification. Tissues with the highest relative intensity of the features of interest were selected for use in the protein identification. A total of 12 tissue fragments, 5 IMC and 7 normal, were subsequently selected from this list because the MALDI spectra generated from these samples contained the largest total number of the significant differentially expressed peaks. Protein extracts were prepared using 3 to 4 mm³ portions of mammary tissue in 1:20 (w/v) Tissue Protein Extraction Reagent (T-PER) plus Halt Protease Inhibitor Cocktail (10 μ L/mL) (Pierce, Rockford, IL). Protein extracts were subjected to a cleanup step using a disposable hand-packed C18 preparative column from Waters (Milford, MA). Samples were then fractionated with a Vydac (Hesperia, CA) 208TP5315 reversed-phase C8 polymeric column at 40 °C using a Waters Alliance HPLC system (Milford, MA) using a flow rate of 0.5 mL/min. Thirty second fractions were collected into a 96 well plate using a Gilson Fraction Collector (Middleton, WI) and further analyzed by MALDI MS and Flex Analysis software (Bruker-Daltonics, Billerica, MA) to determine the fractions containing the peaks of interest. The remaining volume of the fractions containing peaks of interest was separated by one-dimensional gel electrophoresis. Bands were excised corresponding to the m/z peaks of interest. Bands were in-gel-digested with trypsin, and subjected to LC-MS/MS analysis on a Thermo LTQ linear ion trap instrument equipped with

a Thermo nanoelectrospray source, Surveyor LC system, and autosampler (Thermo Fisher, San Jose, CA). Tandem MS spectra were search against the UniRef human database using SEQUEST (Thermo Electron, San Jose, CA) and data filtered based upon the following filtering criteria: cross correlation (Xcorr) value of >1.9 for singly charged ions, > 2.2 for doubly charged ions, and >3.75 for triply charged ions. A RSp (ranking of primary score) value of <4 and a dCN value of ≥ 0.1 were also required for positive peptide identifications.³³ A more detailed description of the protein identification methods is given in the Supporting Information.

Validation of Differentially Expressed Features

Two of the differentially expressed features were confirmed using two commercially available tissue microarrays composed of paraffin-embedded tissue. The AcuMax A202IV array (ISU Abxis Co.) contained 45 breast cancers and 4 adjacent normal tissues in duplicate cores. The US Biomax array consisted of 24 breast tumors with self-matching adjacent tissue and normal tissue (Ijamsville, MD), and was accompanied by ER, PR, and Her2/neu immunohistochemistry results and staging information. Each array was stained with the Sigma calcyclin/Anti-S100A6 antibody (St. Louis, MO) at a 1:125 dilution for 1 h at room temperature and the DAKO calgranulin A/clone MAC 387 antibody (Carpinteria, CA) overnight at 4 °C. Antigen retrieval for both antibodies was performed with proteinase K. The results were scored based on the intensity of cytoplasmic staining on a scale of 0 to 3+ and percentage of cells staining positively. The Goodman Kruskal Γ was used as a measure of association between the ordered categorical calcyclin and calgranulin A staining and the clinicopathological variables. For all tests, differences were considered significant for *P*-values less than 0.05. In this exploratory study, we seek to identify potentially interesting relationships (with $p < 0.05$), rather than control for experiment-wise error rate by using a reduced significance level for individual tests. Representative photomicrographs showing staining with each antibody were taken with an Olympus DP-70 digital camera attached to an Olympus BX40 microscope with a 22 $\times$ ocular and a 20 $\times$ lens.

Results and Discussion

Reproducibility and Influence of Clinical Covariates

The intersample reproducibility of spectra is illustrated in Supporting Information. Spectral replicas (after ProTS Marker preprocessing see above) were highly reproducible with a slightly higher variance of spectra in cancer samples than in normal samples. We have also compared groups of samples originating from different research centers using methods described above to evaluate possible differences attributable to specimen handling at the different centers. In subsequent analyses, we ensured that none of the features that had statistical significance in interinstitutional sample comparisons were used in the classification of clinical groups. Details of this analysis are reported in the Supporting Information (Table 1). We performed similar comparisons of the NME from women 25–35 years and greater than 45 years to examine the possibility that age differences between samples may bias the classifier. In subsequent analyses, we also ensured that none of the features that had statistical significance in the age comparisons were used in the classification of clinical groups. Details of this analysis are reported in the Supporting Information (Table 2).

Cancer versus Normal

To detect proteomic patterns in breast tumors, we assessed the protein expression profiles of 122 IMC and 167 examples of NME from reduction mammoplasty specimens utilizing laser capture microdissection and MALDI MS. Spectra were obtained from an average of 2000 cells dissected from frozen breast tissue by a breast pathologist (M.E.S.) using a serial hematoxylin

and eosin stained section as a guide. Using wrapper methodology and cross-validation as a criterion, we created a classifier optimized for the correct classification of the training set.

From 88 features considered, a set of 14 features was selected to minimize the LOOCV error in the analysis of the training set comparing IMC and NME (Table 2). The details of this feature selection and the LOOCV error for various k -values are presented in the Supporting Information (Tables 3–5). Using the class prediction model based on the selected signals and a $k = 7$, we were able to distinguish IMC versus NME with 97% accuracy in the training cohort and 94% accuracy in the testing cohort with a sensitivity and specificity of 89% and 98%, respectively (Table 3). Discriminating features subsequently identified by the protein identification studies are shown in Figure 1A. The complete set of discriminating features is shown in the Supporting Information (Table 2). While tumor can be distinguished from normal mammary epithelium based on histology alone, we believe that demonstrating a high accuracy for a classifier distinguishing these groups is a necessary proof of concept. The next task is correlation of the proteomic profiles with known biomarkers such as estrogen receptor status.

ER Status

Sufficient residual tissue was available for 117 tumors to evaluate ER status by immunohistochemistry (Table 1). Of these, 61 tumors were ER+ and 56 tumors were ER-. On the basis of a training set of 36 ER+ tumors and 25 ER- tumors, a set of 4 features from among 94 features considered was selected to minimize the LOOCV error (Table 2). The details of this feature selection and the classification LOOCV error for various k -values are shown in the Supporting Information (Tables 6–8). Using the class prediction model based on the selected signals, a k -value of 5, and the remaining 32 ER+ and 24 ER- tumors, we were able to distinguish ER+ and ER- tumors with 85% concordance with immunohistochemistry in the training cohort and 66.1% accuracy in the testing cohort with a sensitivity and specificity of 53% and 87.5%, respectively (Table 3). One discriminating feature subsequently identified by the protein identification studies is shown in Figure 1B.

To our knowledge, this is the first study to examine differences in proteomic expression among ER+ and ER- tumors utilizing MALDI MS of LCM acquired tumor cells. The study of protein expression within these tumor subsets should provide insights complementing those indicated by gene array studies because mRNA expression cannot always indicate which proteins are actually expressed and how their activity might be modulated after translation.^{34,35} Therefore, analysis of the proteome directly from tumor tissue may provide a better molecular snapshot of the pathological status of the cancer than gene expression patterns. As MALDI MS is an easy to use, reproducible, high-throughput technology, it may in the long run provide a cheaper and faster alternative to genetic and immunohistochemical approaches.

Classification accuracy for the ER+ and ER- tumors was less accurate than anticipated. There are likely several factors contributing to this phenomenon. First, these subsets are themselves heterogeneous groups as is well-documented by gene array studies.³⁶ Second, we were limited by the small number of samples available. Having 32 spectra in ER+ and 24 spectra in ER- groups of the training cohort was apparently not enough to create a robust classifier. Thus, the results for ER-status classification should be considered as preliminary. Still, identification of Calgranulin A (see below) as a significant discriminator indicates that even with a limited amount of samples we were able to obtain valuable information using our approach. We expect the classification to be noticeably improved and the protein identities of other significant features identified once we have increased the sample size.

An advantage of our approach was the use of LCM-acquired cells. The genomic studies classifying these tumor types used whole tumor tissues; thus, the presence of stroma, inflammatory cells, and small vessels also contribute to the tumor expression profiles.

Protein Identification

Although the spectral profiles may be useful for classification and prognosis, clues to the underlying biology of neoplastic transformation and progression can be obtained from identification and functional investigation of these peptides and proteins. For protein identification, we prepared a tissue extract containing portions of 12 tissues (5 IMC and 7 normal) remaining after LCM and MALDI analysis and then fractionated the proteins by HPLC. The tissues were selected based on the fact that they had the highest relative intensity of the features of interest. Monitoring of the fractions by MALDI MS permitted identification of the fractions containing the peaks of interest. Fractions of interest were then run on a tryptic gel, and in-gel trypsin digests were performed on bands or molecular weight regions of interest. The resulting peptide extracts were subjected to LC-MS/MS analysis. By this methodology, we were able to identify 126 proteins by 2 or more unique peptides; however, only 3 were among the statistically significant discriminator peaks, ubiquitin m/z 8568, calcyclin (S100A6) 10094 m/z , and calgranulin A (S100A8) 10842 m/z (Table 2).

Two of these classifiers, calcyclin and calgranulin A, were confirmed by immunohistochemistry using two different commercially available TMAs, an AccuMax array containing duplicate cores of 45 tumors which were accompanied by ER, PR, and Her2/neu staining results and a US Biomax TMA which contained 24 tumors and matched normal tissues. Figure 2 shows the spectrum of staining intensity observed with both antibodies. By our MALDI MS analysis, calcyclin expression was greater in normal compared to tumor tissues in the IMC versus NME comparison (Table 2). Correspondingly, on the US Biomax TMA, the number of normal tissues and the intensity of expression of calcyclin was significantly greater than in tumor tissues (Table 4; $p = 0.02$). Interestingly, in a subset of cases, the myoepithelial component also stained positively for calcyclin; however, this relationship did not reach statistical significance. By our MALDI MS analysis, calgranulin A showed increased expression in IMC relative to NME and served as a classifier in the IMC versus NME comparison (Table 2), but we found no statistically significant difference in the staining intensities on the Biomax array (Table 4).

By our MALDI MS analysis, calgranulin A served as one of the features constituting the classifier for ER-negative tumors (Table 2). Calgranulin A staining is negatively associated with ER status for tumors on the AccuMax array ($\Gamma = -0.53$, $p = 0.007$, see Table 4). Although not serving as a discriminator by our MALDI analysis, calcyclin expression is also negatively correlated with ER status ($\Gamma = -0.67$, $p < 0.0001$), which might be improved with performance of a true multivariate analysis as in the KNN classifier based on protein profiles.

The S100 Ca(2+)-binding proteins, a subfamily of EF-hand Ca(2+)-binding proteins, recently became of major interest because of their differential expression in neoplastic tissues, their involvement in metastatic processes, and the clustered organization of at least 10 S100 genes on human chromosome 1q21, a region frequently rearranged in several tumors. Calcyclin has implied roles in the regulation of cell growth and division, exhibits deregulated expression in association with cell transformation, and is found in high abundance in certain breast cancer cell lines. In an immunohistochemical survey of S100 protein expression in 28 tissue types and 21 tumor types, Cross et al. found expression of S100A6 and S100A8 in 12% and 29% of breast cancers.³⁷ Our findings are consistent with those of Carlsson et al. who found down-regulation of S100A6 regardless of pathological stage and up-regulation of S100A8 in breast cancer using serial analysis of gene expression (SAGE).³⁸ Kennedy et al. have shown that BRCA1 is capable of repressing several of the members of the S100A family including S100A8 and that functional BRCA1 is required for this repression.³⁹ It is interesting that in our study the ER- and Her2- tumors showed the highest expression of S100A8 and these “triple negative” tumors are the subtype which contains tumors from patients carrying BRCA-1 mutations.^{36,40} Mutant forms of BRCA-1 may be incapable of S100A8 repression. Finally,

Ohuchida et al. have shown that inhibition of S100A6 decreased proliferation and invasiveness of pancreatic cancer cell lines⁴¹ and Vimalachandran et al. demonstrated the high expression of S100A6 (Calcyclin) is significantly associated with poor survival in pancreatic cancer patients.⁴² While we do not have specific follow-up information on the patients in this study, it is striking that the tumor types with the highest expression of calcyclin in our study, triple negative and Her2-overexpressing, are known to have a poor prognosis.

Conclusions

The work reported here represents the first stage in the analysis of proteomic expression in human breast tumors utilizing MALDI MS and LCM-acquired cells from frozen tissue sections. Following spectral alignment and processing, biomarker candidates were identified by statistical analysis. We developed classifiers for distinction of breast tumor versus normal mammary epithelium and ER+ versus ER- tumors using test sets and then successfully used these classifiers in blinded test sets. Two of the *m/z* features, 10094 and 10842, were subsequently identified by LC-MS/MS as calcyclin (S100A6) and calgranulin A (S100A8) and confirmed by immunohistochemistry. Additional studies to characterize the function and biological role of these proteins in breast cancer will be undertaken.

Proteomic expression in breast cancer was evaluated by MALDI MS of laser captured microdissected cells. Protein and peptide expression in cancer versus normal mammary epithelium and estrogen receptor-positive versus estrogen receptor-negative tumors were compared. Biomarker candidates were identified by statistical analysis and classifiers developed using a training set and validated in independent test sets. Several *m/z* features used in the classifiers were identified by LC-MS/MS and two were confirmed by immunohistochemistry.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

Dr. Sanders is the recipient of a Komen Foundation Award. Dr. Dias is the recipient of an AVON-AACR Scholarship Award. This project was supported in part by The Vanderbilt Breast Cancer Specialized Program in Research Excellence (3P50 CA098131), NIH RO1 CA80195 (C.L.A.), and Cancer Center Support Grant P30 CA68485.

References

1. Page DL. Special types of invasive breast cancer, with clinical implications. *Am. J. Surg. Pathol* 2003;27(6):832–5. [PubMed: 12766589]
2. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomized trials. *Lancet* 2005;365(9472):1687–717. [PubMed: 15894097]
3. International breast Cancer Study (IBCSG). Endocrine responsiveness and tailoring adjuvant therapy for postmenopausal lymph node-negative breast cancer: a randomized trial. *J. Natl. Cancer Inst* 2002;94(14):1054–65. [PubMed: 12122096]
4. Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N. Engl. J. Med* 2005;353(16):1673–84. [PubMed: 16236738]
5. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N. Engl. J. Med* 2005;353(16):1659–72. [PubMed: 16236737]
6. Carlson RW, Brown E, Burstein HJ, et al. NCCN task force report: adjuvant therapy for breast cancer. *J. Natl. Compr. Cancer Network* 2006;4:S1–26.

7. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumors. *Nature* 2000;406(6797):747–52. [PubMed: 10963602]
8. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med* 2002;347(25):1999–2009. [PubMed: 12490681]
9. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005;365(9460):671–9. [PubMed: 15721472]
10. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med* 2004;351(27):2817–26. [PubMed: 15591335]
11. Bogaerts J, Cardoso F, Buyse M, et al. Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. *Nat. Clin. Pract. Oncol* 2006;3(10):540–51. [PubMed: 17019432]
12. Mauriac L, Debled M, MacGrogan G. When will more useful predictive factors be ready for use. *Breast* 2005;14(6):617–23. [PubMed: 16169224]
13. Sparano JA. TAILORx: trial assigning individualized options for treatment (Rx). *Clin. Breast Cancer* 2006;7(4):347–50. [PubMed: 17092406]
14. Caprioli RM, Farmer TB, Gile J. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Anal. Chem* 1997;69(23):4751–60. [PubMed: 9406525]
15. Chaurand P, DaGue BB, Pearsall RS, Threadgill DW, Caprioli RM. Profiling proteins from azoxymethane-induced colon tumors at the molecular level by matrix-assisted laser desorption/ionization mass spectrometry. *Proteomics* 2001;1(10):1320–6. [PubMed: 11721643]
16. Chaurand P, Schwartz SA, Caprioli RM. Imaging mass spectrometry: a new tool to investigate the spatial organization of peptides and proteins in mammalian tissue sections. *Curr. Opin. Chem. Biol* 2002;6(5):676–81. [PubMed: 12413553]
17. Stoeckli M, Chaurand P, Hallahan DE, Caprioli RM. Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nat. Med* 2001;7(4):493–6. [PubMed: 11283679]
18. Taguchi F, Solomon B, Gregorc V, et al. Mass spectrometry to classify non-small-cell lung cancer patients for clinical outcome after treatment with epidermal growth factor receptor tyrosine kinase inhibitors: a multicohort cross-institutional study. *J. Natl. Cancer Inst* 2007;99(11):838–46. [PubMed: 17551144]
19. Xu BJ, Caprioli RM, Sanders ME, Jensen RA. Direct analysis of laser capture microdissected cells by MALDI mass spectrometry. *J. Am. Soc. Mass Spectrom* 2002;13(11):1292–7. [PubMed: 12443019]
20. Cazares LH, Adam BL, Ward MD, et al. Normal, benign, preneoplastic, and malignant prostate cells have distinct protein expression profiles resolved by surface enhanced laser desorption/ionization mass spectrometry. *Clin. Cancer Res* 2002;8(8):2541–52. [PubMed: 12171882]
21. Chalmers MJ, Gaskell SJ. Advances in mass spectrometry for proteome analysis. *Curr. Opin. Biotechnol* 2000;11(4):384–90. [PubMed: 10975458]
22. Chaurand P, Stoeckli M, Caprioli RM. Direct profiling of proteins in biological tissue sections by MALDI mass spectrometry. *Anal. Chem* 1999;71(23):5263–70. [PubMed: 10596208]
23. Stoeckli M, Farmer TB, Caprioli RM. Automated mass spectrometry imaging with a matrix-assisted laser desorption ionization time-of-flight instrument. *J. Am. Soc. Mass Spectrom* 1999;10(1):67–71. [PubMed: 9888186]
24. Todd PJ, Schaaff TG, Chaurand P, Caprioli RM. Organic ion imaging of biological tissue with secondary ion mass spectrometry and matrix-assisted laser desorption/ionization. *J. Mass Spectrom* 2001;36(4):355–69. [PubMed: 11333438]
25. Amann JM, Chaurand P, Gonzalez A, et al. Selective profiling of proteins in lung cancer cells from fine-needle aspirates by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin. Cancer Res* 2006;12(17):5142–50. [PubMed: 16951232]
26. Chaurand P, Norris JL, Cornett DS, Mobley JA, Caprioli RM. New developments in profiling and imaging of proteins from tissue sections by MALDI mass spectrometry. *J. Proteome Res* 2006;5(11):2889–900. [PubMed: 17081040]
27. 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(10):1975–83. [PubMed: 16849436]

28. Meistermann H, Norris JL, Aerni HR, et al. Biomarker discovery by imaging mass spectrometry: transthyretin is a biomarker for gentamicin-induced nephrotoxicity in rat. *Mol. Cell. Proteomics* 2006;5(10):1876–86. [PubMed: 16705188]
29. Roder, H.; Gringorieva, J.; Tsylin, M. The use of mass spectra for cancer biomarker detection. Biodesix; 2005. Available at: <http://www.biodesix.com/Documents/MarkerWhitePaper.pdf>
30. Theodoritis, S.; Koutroumbas, K. Pattern Recognition. Vol. 3rd ed.. Academic Press; Amsterdam, Boston: 2006.
31. Webb, A. Statistical Pattern Recognition. Wiley; Chichester, U.K.; 2002.
32. Yanagisawa K, Shyr Y, Xu BJ, et al. Proteomic patterns of tumor subsets in non-small-cell lung cancer. *Lancet* 2003;362(9382):433–9. [PubMed: 12927430]
33. Zimmerman LJ, Wernke GR, Caprioli RM, Liebler DC. Identification of protein fragments as pattern features in MALDIMS analyses of serum. *J. Proteome Res* 2005;4(5):1672–80. [PubMed: 16212420]
34. Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 1997;18(3–4):533–7. [PubMed: 9150937]
35. Wilkins MR, Sanchez JC, Williams KL, Hochstrasser DF. Current challenges and future applications for protein maps and post-translational vector maps in proteome projects. *Electrophoresis* 1996;17(5):830–8. [PubMed: 8783009]
36. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U.S.A* 2001;98(19):10869–74. [PubMed: 11553815]
37. Cross SS, Hamdy FC, Deloulme JC, Rehman I. Expression of S100 proteins in normal human tissues and common cancers using tissue microarrays: S100A6, S100A8, S100A9 and S100A11 are all overexpressed in common cancers. *Histopathology* 2005;46(3):256–69. [PubMed: 15720411]
38. Carlsson H, Petersson S, Enerback C. Cluster analysis of S100 gene expression and genes correlating to psoriasin (S100A7) expression at different stages of breast cancer development. *Int. J. Oncol* 2005;27(6):1473–81. [PubMed: 16273201]
39. Kennedy RD, Gorski JJ, Quinn JE, et al. BRCA1 and c-Myc associate to transcriptionally repress psoriasin, a DNA damage-inducible gene. *Cancer Res* 2005;65(22):10265–72. [PubMed: 16288014]
40. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. U.S.A* 2003;100(14):8418–23. [PubMed: 12829800]
41. Ohuchida K, Mizumoto K, Ishikawa N, et al. The role of S100A6 in pancreatic cancer development and its clinical implication as a diagnostic marker and therapeutic target. *Clin. Cancer Res* 2005;11(21):7785–93. [PubMed: 16278400]
42. Vimalachandran D, Greenhalf W, Thompson C, et al. High nuclear S100A6 (Calcyclin) is significantly associated with poor survival in pancreatic cancer patients. *Cancer Res* 2005;65(8):3218–25. [PubMed: 15833853]
43. 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(13):3609–14. [PubMed: 12097261]
44. Paweletz CP, Trock B, Pennanen M, et al. Proteomic patterns of nipple aspirate fluids obtained by SELDI-TOF: potential for new biomarkers to aid in the diagnosis of breast cancer. *Dis. Markers* 2001;17(4):301–7. [PubMed: 11790897]
45. Petricoin EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002;359(9306):572–7. [PubMed: 11867112]
46. Petricoin EF III, Ornstein DK, Paweletz CP, et al. Serum proteomic patterns for detection of prostate cancer. *J. Natl. Cancer Inst* 2002;94(20):1576–8. [PubMed: 12381711]

Figure 1.

Representative features selected for classifiers. (A) Three representative features from the IMC vs NME classifier. The features m/z 8568, m/z 10094, and m/z 10842 in spectra generated from IMC are shown in blue and from NME in red. (B) A representative feature for the ER-positive vs ER-negative classifier. The feature 10842 m/z from ER-positive tumors is shown in blue and the feature for ER-negative tumors is shown in red. The spectra in the left column represent median spectra from the two groups. Bold lines represent the median spectrum for each group, and the thin lines represent 25th and 75th percentile. The spectra in the right column show the specific peak as it appears in each individual spectrum.

Figure 2. Calcyclin and calgranulin A immunohistochemistry. (A) Calcyclin (upper panel). The photomicrographs show representative examples of calcyclin staining seen in tumor on the AccuMax tissue microarray. The staining intensity ranged from 3+ (1), 2+ (2), 1+ (3), to 0 (4). All photos were shot with a 22× ocular and 20× lens with an Olympus DP-70 digital camera. (B) Calgranulin A (lower panel). The photomicrographs show representative examples of calgranulin A staining seen in tumor on the AccuMax tissue microarray. The staining intensity ranged from 3+ (1), 2+ (2), 1+ (3), to 0 (4). All photos are shot with a 22× ocular and 20× lens with an Olympus DP-70 digital camera.

Table 1
Clinical and Pathologic Characteristics of Samples Across Centers^a

	centers					Totals
	1	2	3	4	5	
IMC	24	9	13	3	73	122
Average age yrs (no.) ^b	66 (n = 22)	57 (n = 4)	65 (n = 13)	49 (n = 3)	53 (n = 73)	
Grade						
Low	7	1	0	0	8	16
Intermediate	15	7	2	0	27	51
High	2	1	11	3	38	55
Histologic type						
No Special type	23	6	13	3	70	115
Special type ^c	1	3	0	0	3	7
Stage ^b						
I	5	0	1	0	11	17
II	6	0	4	0	19	31
III	5	2	3	2	15	27
Hormone Receptor Status						
ER+/PR+	10	6	5	0	20	41
ER+/PR-	11	1	3	0	14	29
ER-/PR+	0	0	0	0	0	0
ER-/PR-	3	2	5	3	34	47
NME	0	26	91	31	19	167
Average age (yrs)	N/A	36	33	33	31	

^aCenters: (1) Royal Marsden Hospital, U.K.; (2) CHTN Eastern Division, University of Pennsylvania, Philadelphia, PA; (3) CHTN Midwestern Division, Ohio State University; (4) CHTN Southern Division, University of Alabama, Birmingham, AL; (5) CHTN Western Division, Vanderbilt University.

^bInformation was not available for all women. No reduction mammoplasty specimens were obtained from center 1.

^cSpecial type cancers: center 1, lobular-2; center 2, mucinous carcinoma-1, lobular carcinoma-2; center 5, tubular-1, mucinous carcinoma-2.

Table 2Features Used in Classifiers^a

Classifier	IMC vs NME	
<i>Greatest Relative expression</i>	IMC	NME
<i>m/z</i>	4205, 4938, 5421, 5827, 7176, 8435, 8568 , 10842 , 11654	6891, 7651, 10094 , 13782, 22602
Classifier	ER+ vs ER−	
<i>Greatest Relative expression</i>	ER+ tumors	ER− tumors
<i>m/z</i>	7177	6548, 9155, 10842

^aThe *m/z* values in boldface were subsequently identified as ubiquitin (*m/z* 8568), calcyclin (*m/z* 10094), and calgranulin A (*m/z* 10842).

Table 3Error Rates for the Classification of the Independent Test Sets^a

test set		Classifier		
		IMC (<i>n</i> = 61)	NME (<i>n</i> = 83)	total (<i>n</i> = 144)
IMC vs NME	Correct	54	81	135
	Error	7	2	9

test set		Classifier		
		ER+ (<i>n</i> = 32)	ER− (<i>n</i> = 24)	total (<i>n</i> = 56)
ER+ vs ER−	Correct	17	21	37
	Error	15	3	19

^aSensitivity, specificity, and accuracy for the IMC vs NME comparison are 88.5%, 97.6%, and 93.7%, respectively. Sensitivity, specificity, and accuracy for the ER+ vs ER− comparison are 53.0%, 87.5%, and 66.1%, respectively.

US Immunohistochemistry Staining Results for Tissue Microarrays

Table 4

US Biomax TMA Breast Cancer Cases vs Normal Breast						
antigen	interpretable cores (%)	invasive tumor (%)	luminal epithelium	myoepithelium		
Calcyclin+	37 (77)	11	14	5		
Calcyclin-		8	4	13		
Total		19	18	18		
P-value			0.02 ^a	0.20 ^a		
Calgranulin+	36 (75)	7	5	0		
Calgranulin-		12	12	0		
Total		19	17	0		
P-value			0.51 ^a	0.06 ^a		

AccuMax TMA Breast Cancer Cases and Tumor Subsets						
antigen	interpretable cores (%)	positive staining	ER+	ER-	Her2+	Her2-
Calcyclin+	41 (91)	32	22	10	11	21
Calcyclin-		8	8	0	4	4
Total		40	30	10	15	25
P-value			<0.0001 ^b		<0.14 ^b	
Calgranulin+	41 (91)	25	16	9	10	15
Calgranulin-		16	14	2	5	11
Total		41	30	11	15	26
P-value			0.007 ^b		0.51 ^b	

^aValues are calculated with respect to invasive tumor.

^bHypothesis tests are computed based on the IHC scoring from 0 to 3+, but are collapsed to positive and negative staining for this table.