

# *In Situ* Proteomic Analysis of Human Breast Cancer Epithelial Cells Using Laser Capture Microdissection: Annotation by Protein Set Enrichment Analysis and Gene Ontology\*<sup>§</sup>

Sangwon Cha<sup>§</sup>, Marcin B. Imielinski<sup>¶</sup>, Tomas Rejtar<sup>§</sup>, Elizabeth A. Richardson<sup>¶</sup>, Dipak Thakur<sup>§</sup>, Dennis C. Sgroi<sup>¶†\*\*</sup>, and Barry L. Karger<sup>†\*\*</sup>

Identification of molecular signatures that allow detection of the transition from normal breast epithelial cells to malignant invasive cells is a critical component in the development of diagnostic, therapeutic, and preventative strategies for human breast cancer. Substantial efforts have been devoted to deciphering breast cancer etiology at the genome level, but only a limited number of studies have appeared at the proteome level. In this work, we compared individual *in situ* proteome profiles of nonpatient matched nine noncancerous, normal breast epithelial (NBE) samples with nine estrogen receptor (ER)-positive (luminal subtype), invasive malignant breast epithelial (MBE) samples by combining laser capture microdissection (LCM) and quantitative shotgun proteomics. A total of 12,970 unique peptides were identified from the 18 samples, and 1623 proteins were selected for quantitative analysis using spectral index (Spl) as a measure of protein abundance. A total of 298 proteins were differentially expressed between NBE and MBE at 95% confidence level, and this differential expression correlated well with immunohistochemistry (IHC) results reported in the Human Protein Atlas (HPA) database. To assess pathway level patterns in the observed expression changes, we developed protein set enrichment analysis (PSEA), a modification of a well-known approach in gene expression analysis, Gene Set Enrichment Analysis (GSEA). Unlike single gene-based functional term enrichment analyses that only examines pathway overrepresentation of proteins above a given significance threshold, PSEA applies a weighted running sum statistic to the entire expression data to discover significantly enriched protein groups. Application of PSEA to the expression data in this study revealed not only well-known ER-dependent and cellular morphology-dependent protein abundance changes, but also significant alterations of downstream targets for multiple transcription factors (TFs), suggesting a role for specific gene regulatory pathways in breast tumorigenesis. A

parallel GOMiner analysis revealed both confirmatory and complementary data to PSEA. The combination of the two annotation approaches yielded extensive biological feature mapping for in depth analysis of the quantitative proteomic data. *Molecular & Cellular Proteomics* 9: 2529–2544, 2010.

Breast cancer is a major health problem that each year affects the lives of millions of women worldwide. In 2008, in the United States alone, ~180,000 women were diagnosed with invasive breast carcinoma (1). The use of high-throughput gene expression technologies applied to the study of human breast cancer has led to the discovery of the “intrinsic gene signatures” that stratify human breast cancers into four subtypes that correlate remarkably well with clinically recognized breast cancer subtypes (2–6). These subtypes include “HER2+,” “basal,” and “luminal A,” “luminal B” breast cancers. HER2+ tumors are most frequently estrogen receptor (ER)-<sup>1</sup>, express proliferation genes, as well as Her-2 and other genes linked to this latter locus. The basal tumors are most commonly ER negative, progesterone receptor negative and Her-2 negative. The luminal A and luminal B tumors express luminal cytokeratins, the estrogen receptor (ER), and trans-acting T-cell-specific transcription factor (GATA3).

The luminal breast cancers (both A and B subtypes) constitute ~70% of all human breast cancers diagnosed worldwide. In general, the luminal breast cancers are associated with favorable prognosis as compared with the HER2+ and basal subtypes. Nevertheless, luminal B tumors have a worse prognosis than luminal A tumors, and recent data suggest

<sup>1</sup> The abbreviations used are: ER, estrogen receptor; ERE, estrogen response element; ES, enrichment score; GO, gene ontology; GSEA, gene set enrichment analysis; HPA, Human Protein Atlas; IHC, immunohistochemistry; LCM, laser capture microdissection; MBE, malignant breast epithelium; MSigDB, the molecular signatures database; NBE, normal breast epithelium; PSEA, protein set enrichment analysis; RSS, running sum statistic; Spl, spectral index; TF, transcription factor; FDR, false discovery rate; SRF, serum response factor; AP-1, activator protein 1; SREBP1, sterol regulatory binding protein 1.

From the <sup>§</sup>Barnett Institute, Northeastern University, 360 Huntington Ave., MA 02115; <sup>¶</sup>Molecular Pathology Unit and Center for Cancer Research, Massachusetts General Hospital, 149 13th Street, Charlestown, MA 02129

Received, May 5, 2010 and in revised form July 23, 2010

Published, MCP Papers in Press, DOI 10.1074/mcp.M110.000398

that luminal A tumors may be adequately treated with anti-hormonal therapy alone, whereas luminal B tumors may benefit from chemotherapy added to antihormonal therapy (7). Despite advances in the gene expression-based stratification of human breast cancer, the molecular basis of luminal breast tumorigenesis and luminal breast cancer clinical heterogeneity is still poorly understood. This gap in knowledge is due, in part, to the well-known limitations associated with gene expression, for it is the gene products, or the proteome, and not the genes themselves that are the biochemical determinants of cell growth and metabolism. Thus, increased knowledge of the proteomic alterations associated with luminal breast cancer tumorigenesis will help advance understanding of human breast cancer and facilitate tailored interventions in select luminal breast cancer patients.

Over the past decade, research has been conducted to study the breast cancer proteome to increase the molecular understanding of breast cancer tumorigenesis beyond that from existing breast cancer gene expression data (8–10). Global proteomic analyses of tumor biopsies, dissected cells, human breast milk and nipple aspirate fluid, cancer cell lines, and sera and plasma provide opportunities for unbiased characterization of protein expression in breast cancer (11, 12). Tumor tissue is likely to be the most informative sample, since proteomic analysis is conducted directly on the sample where the disease resides. However, tumor analysis is challenging, given the heterogeneity of the breast cancer tissue and the limited number of cells generally available.

Highly enriched cell populations can be obtained from heterogeneous samples by laser capture microdissection (LCM) (13, 14). Such sample analysis can lead to detailed proteomic portraits of tumor microenvironments and, importantly, correct for potential confounding effects of stromal contamination (15). Though such analyses are numerous in the world of gene expression studies, proteomic analyses of LCM breast tissue are limited (16–21). Recently, solid tumor heterogeneity at the protein level was assessed by comparing LCM-acquired breast tumor proper and stromal cells from a lymph node containing breast carcinoma (21), and differential proteome profiles according to tamoxifen therapy responsiveness were analyzed from ER+, primary tumor pooled LCM samples (22).

The combination of LCM enrichment and shotgun proteomics has the potential to reveal novel pathogenic pathways and networks by which normal breast epithelial cells progress to invasive malignant epithelial cells. Genomic and gene expression profiling of human breast cancer progression has already shown potential for the identification of molecular alterations that serve as prognostic and predictive biomarkers (2–4, 23). Comprehensive knowledge of proteomic patterns in normal epithelium, as compared with malignant breast epithelium, would be invaluable in providing a more complete picture of the molecular entities associated with breast cancer progression.

In the present study, we performed comparative proteomic profiling of laser capture microdissected normal breast epithelium (NBE) from nine different noncancerous human mammaplasty specimens relative to microdissected malignant breast epithelium (MBE) from nine different human invasive luminal (ER positive) breast cancer specimens. Using only the limited amount of material collected by LCM (60,000 cells), a proteomic profile of each sample was separately acquired by label-free proteomics using our previously developed platform for analysis of LCM samples (24). To the best of our knowledge, this is the most comprehensive global study that compares proteomic signatures of phenotypically normal and ER+ invasive malignant breast epithelial cells in multiple individual samples. For each protein, the spectral index (Spl) value (25) was used as a measure of relative abundance, with differentially abundant proteins between NBE and MBE being identified by employing permutation analysis. Importantly, a number of proteins highly enriched in MBE, including both well-known and novel proteins in the context of breast cancer, were verified to be up-regulated based on immunohistochemical profiles of normal and disease tissues available in the Human Protein Atlas (HPA) (26).

A challenge in all global expression profiling studies is to annotate the large number of observed molecular changes. Functional approaches employing DAVID or GOMiner have been used in the analysis of global proteomic profiling studies (27, 28). Though effective, these approaches rely on strict quantitative thresholds for selecting proteins to be included and ignore the level of expression differences when scoring enriched functional terms. Gene set enrichment analysis (GSEA) was developed in the gene expression field as a powerful approach for determining functional significance of differential expression results by taking into account the quantitative nature of expression correlation (29). Recently, GSEA has been applied, without modification, to proteome datasets to interpret global functional changes in dilated cardiomyopathy (30).

In the present study, we modified GSEA to be suitable for the analysis of label-free quantitative proteomic data in an approach termed Protein Set Enrichment Analysis (PSEA). PSEA takes into account the quantitative level of differential protein expression measured by spectral counts to highlight biologically related proteins with strong and highly concordant expression differences. PSEA of LCM-derived MBE samples revealed significant decreased expression of cytoskeletal proteins and of proteins that are consistently negatively correlated with ER+ human breast cancers. More importantly, many protein sets, representing proteins controlled by common transcription factors (TFs) were found, potentially suggesting a role of estrogen response element (ERE)-independent ER signaling in breast tumorigenesis. In addition, we performed a GOMiner analysis, which requires arbitrary thresholds for differentially abundant proteins, on the proteomic results. For differentially expressed proteins at the

TABLE I  
Patient and tumor characteristics of samples used in this study

Abbreviations: RM, reduction mammoplasty; M, mastectomy; E, excision; Grade, Nottingham combined histological tumor grade; LN, lymph node; ER, estrogen receptor; PR progesterone receptor; HER2, expression determined by immunohistochemistry and/or by fluorescence in situ hybridization; Pos, positive; Neg, negative.

Normal breast specimens				Invasive breast cancer specimens										
Sample ID	Case ID	Age	Specimen type	Sample ID	Case ID	Age	Specimen type	Tumor type	Size (cm)	Grade	LN status	ER	PR	HER2
NBE1	215-4	30	RM	MBE1	155-2	64	E	Ductal	1.1	2	Neg	Pos	Pos	Neg
NBE2	279-1	18	RM	MBE2	153-2	46	M	Ductal	1.6	1	Pos	Pos	Pos	Pos
NBE3	224-4	21	RM	MBE3	128-3	79	M	Lobular	5	2	Pos	Pos	Pos	Neg
NBE4	222-6	24	RM	MBE4	171-1	39	E	Ductal	2.8	2	Neg	Pos	Pos	Neg
NBE5	225-6	30	RM	MBE5	130-2	54	M	Ductal	6.5	2	Pos	Pos	Pos	Neg
NBE6	295-4	25	RM	MBE6	98-2	31	E	Ductal	1.7	3	Pos	Pos	Pos	Neg
NBE7	320-5	38	RM	MBE7	122-3	45	M	Ductal	1.7	2	Neg	Pos	Pos	Neg
NBE8	331-8	22	RM	MBE8	193-2	45	M	Ductal	1.5	1	Pos	Pos	Pos	Neg
NBE9	335-6	34	RM	MBE9	169-1	34	M	Ductal	2.6	3	Pos	Pos	Pos	Neg

95% confidence level, GOMiner analysis added unique biological features such as up-regulation of Golgi vesicle transport activity in MBE. Complementing PSEA with the results of standard GOMiner analysis, we are able to map extensively potential activators and targets for common TFs.

EXPERIMENTAL PROCEDURES

**Materials**—Ammonium bicarbonate, dithiothreitol, iodoacetamide, calcium chloride, hematoxylin, ethanol, xylene, and formic acid were purchased from Sigma (St. Louis, MO). NuPAGE® Novex 4%–12% gradient gels, SimplyBlue SafeStain, and lithium dodecyl sulfate in Tris base buffer were obtained from Invitrogen (Carlsbad, CA), and sequencing grade modified trypsin was from Promega (Madison, WI). liquid chromatography/mass spectrometry-grade water was purchased from Mallinckrodt Baker (Phillipsburg, NJ) and acetonitrile from Fisher Scientific (Pittsburg, PA). Magic C18AQ (3 and 5 μm i.d., 200 Å pore size) particles were obtained from Michrom Bioresources (Auburn, CA), and CapSure HS LCM caps were from Molecular Devices (Sunnyvale, CA).

**Clinical Specimen**—Normal breast tissue samples obtained from mammoplasty reduction specimens from women without a history of breast cancer, and breast cancer tissue samples obtained from either mastectomy or surgical excision specimens were fresh frozen within 20 mins of surgery and stored at –80 °C. All tissues were obtained from the Massachusetts General Hospital between 1999 and 2002. Tissues were selected based on the abundance of target cells (normal breast epithelium and malignant invasive carcinoma cells), ease of microdissection with minimal stromal contamination, and absence of fibrocystic changes in the normal breast lobules. Nonpatient matched, rather than patient-matched, normal samples were used to avoid any potential “field effects,” as previously published data have shown that closely adjacent normal tissue and cancer tissue share common genetic and proteomic abnormalities (31, 32). Diagnostic criteria and tumor grading were described previously (33). Patient and tumor characteristics are listed in Table I. This study was approved by the Massachusetts General Hospital human research committee in accordance with National Institutes of Health human research study guidelines.

**Laser Capture Microdissection**—All tissues were prepared, as previously described (34), except that eosin was not included in the tissue staining protocol. Highly enriched populations of phenotypically normal and malignant epithelial cells were procured by an IR (wavelength, λ at 810 nm) laser capture method using the PixCell Ite LCM system (Molecular Devices, Mountain View, CA), as described (33). Approximately 60,000 cells were procured per tissue sample onto ten LCM caps and subjected to subsequent proteomic analysis.

**Cell Lysis, Protein Separation, and In-Gel Digestion**—Captured epithelial cells were lysed directly on LCM caps with 2% lithium dodecyl sulfate in 140 mM Tris base buffer (pH 8.5) as follows. Cell lysis buffer was added to the first LCM cap, and lysis was conducted for 3 min by gently shaking at room temperature. After lysis on the first cap, the whole volume of the cell lysate was transferred to the next LCM cap, and 3 min lysis was again performed. This process was repeated throughout the ten LCM caps to obtain one complete sample. After cell lysis, all processed LCM caps were washed with fresh buffer, and the wash solution was added to the collected cell extract. The resulting extract (~25 μl in total per sample) was directly loaded onto a NuPAGE® Novex 4%–12% gradient gel, and the protein mixture was separated according to the manufacturer’s instructions. After staining with Coomassie Blue (SimplyBlue SafeStain, Invitrogen), the gel was cut into three sections for the molecular weight ranges of > 50 kDa, 20–50 kDa, and < 20 kDa.

Each gel section was separately subjected to in-gel tryptic digestion as described previously (24). Briefly, proteins were reduced and alkylated, and incubated with trypsin for 18 h at 37 °C. Protein digests were first collected by dehydrating the gel pieces, and the remaining digest in the dried gel was further removed by the extraction buffer (1:2 (v/v) 5% formic acid/acetonitrile). The total extract was dried down in a vacuum centrifuge and stored at –80 °C.

**Liquid Chromatography/Mass Spectrometry**—The protein digests were analyzed by capillary reversed phase liquid chromatography, on-line coupled to a hybrid LTQ-FT MS (Thermo Fisher Scientific, San Jose, CA). The nano-reversed phase liquid chromatography system was composed of an Ultimate 3000 nano-LC pump (Dionex, Sunnyvale, CA), a 75 μm i.d. × 6 cm trap column, packed with 5 μm i.d. Magic C18AQ, 200 Å pore size, and a 75 μm i.d. × 15 cm analytical column packed with 3 μm i.d. Magic C18AQ, 200 Å pore size particles. The columns were configured with a vented system (35, 36) in which the sample was loaded onto the trap column and desalted at a flow rate of 1 μl/min. After sample desalting, the flow was directed to the analytical column at 200 nL/min. Mobile phases A and B were 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively. A three-step linear gradient was employed for the gradient separation (2% B to 30% B in 50 min, 30% B to 60% B in 10 min, and 60% B to 80% B in 5 min).

MS data were collected in the data dependent mode; MS scans were acquired in the FTMS from m/z 400 to 2000 with a mass resolution of 100,000 at m/z 400 with up to 2 × 10<sup>6</sup> ions, which significantly improved sensitivity of analysis while reducing mass accuracy. Up to ten MS/MS scans were collected in the linear ion trap for every FTMS scan with the dynamic exclusion set to 60 s, and the normalized collision energy for collision induced dissociation 35%.

Downloaded from http://www.mcponline.org/ by guest on April 24, 2019



**Data Processing**—The DTA files were generated from acquired MS/MS scans by Extract-MSn (Version 4.0, Thermo Fisher Scientific) and these files were searched against the human SwissProt annotated database (Version 57.9, October, 2009, 20,342 entries including common contaminants) merged with a database containing reversed sequences by using the Sequest algorithm (cluster version 27 rev. 12, Thermo Fisher Scientific) within the Computational Proteomics Analysis System (Version 9.10, Labkey Software Foundation) (37). Cysteine carbamidomethylation was included as a fixed modification, but no variable modification was considered. Mass tolerances were set to 2.5 Da for the precursor ions and 1.0 Da for the fragment ions. Full tryptic enzyme specificity was applied with up to two missed cleavage sites. Identified peptides were filtered by  $X_{\text{corr}} \geq 1.9$  for 1+ ions,  $\geq 2.2$  for 2+ ions, and  $\geq 3.8$  for 3+ ions, PeptideProphet (Institute for Systems Biology, Seattle, WA) probability  $\geq 0.9$ , and after correction for missassigned monoisotopic peaks with the mass accuracy of the precursor ions  $\leq 50$  ppm. Higher mass tolerance was needed to account for a higher number of ions in the FTICR cell; however, run to run mass accuracy was below 5 ppm.

ProteinProphet (Institute for Systems Biology, Seattle WA) was used to assign peptides to protein groups rather than to specific proteins because some peptides could be found in multiple proteins. Only protein groups with probability  $\geq 0.9$  were considered. In order to compare protein groups across multiple samples, a perl script was written to generate a master list of protein groups using peptides identified in all samples. Next, protein groups found in individual samples were matched to the master list. Protein groups identified in individual samples that were found to be subsets of a particular master protein group were considered identical, which simplified assignment of proteins identified in different samples. In rare cases, when the individual sample protein group was not consistent with any group in the master list, e.g. the match was to more than one master protein group; such groups were reported separately and were not used in subsequent analysis. It should be noted that, for simplicity, the term “protein” rather than “protein group” is used throughout the paper.

**Identification of Differentially Abundant Proteins by Spectral Index (Spl)**—Spectral counts, i.e. the total number of MS/MS spectra assigned to a particular protein, in individual samples were normalized by the total number of spectral counts identified in each sample (38). The spectral index (Spl) (25, 39) was next calculated to determine proteins with differential abundance between normal (NBE) and malignant breast cancer (MBE) specimens (see supplemental material for Spl formula). In order to determine Spl values corresponding to statistically significant differences in abundance, permutation analysis was performed (25). Briefly, normalized spectral count data of NBE and MBE samples were randomized 2000 times by reshuffling phenotype class labels, and Spl values for all proteins were calculated for each of the randomized datasets. Then, a frequency distribution of all Spl values, i.e. empirical null distribution, was plotted, and Spl values corresponding to, for example, 95% statistical significance were determined as the 2.5 and the 97.5 percentile of this distribution.

**Protein Set Enrichment Analysis (PSEA)**—Enrichment of specific protein sets, built from annotated gene sets contained in the molecular signatures database (MSigDB v2.5, <http://www.broadinstitute.org/gsea/msigdb/>), were computed from the total list of proteomic results. A GSEA-type enrichment score (ES) using Spl as the abundance metric to weigh ranked proteins with a running sum statistic was implemented, as previously described (29). The ES was taken as the maximum positive or negative deviation of the running sum statistic from zero. Each of the 5452 annotated protein sets in MSigDB was matched against all 1623 proteins detected in our samples. Out of the all protein sets in MSigDB, 876 protein sets matched our data with at least 25 proteins. For each of these matched protein lists, an

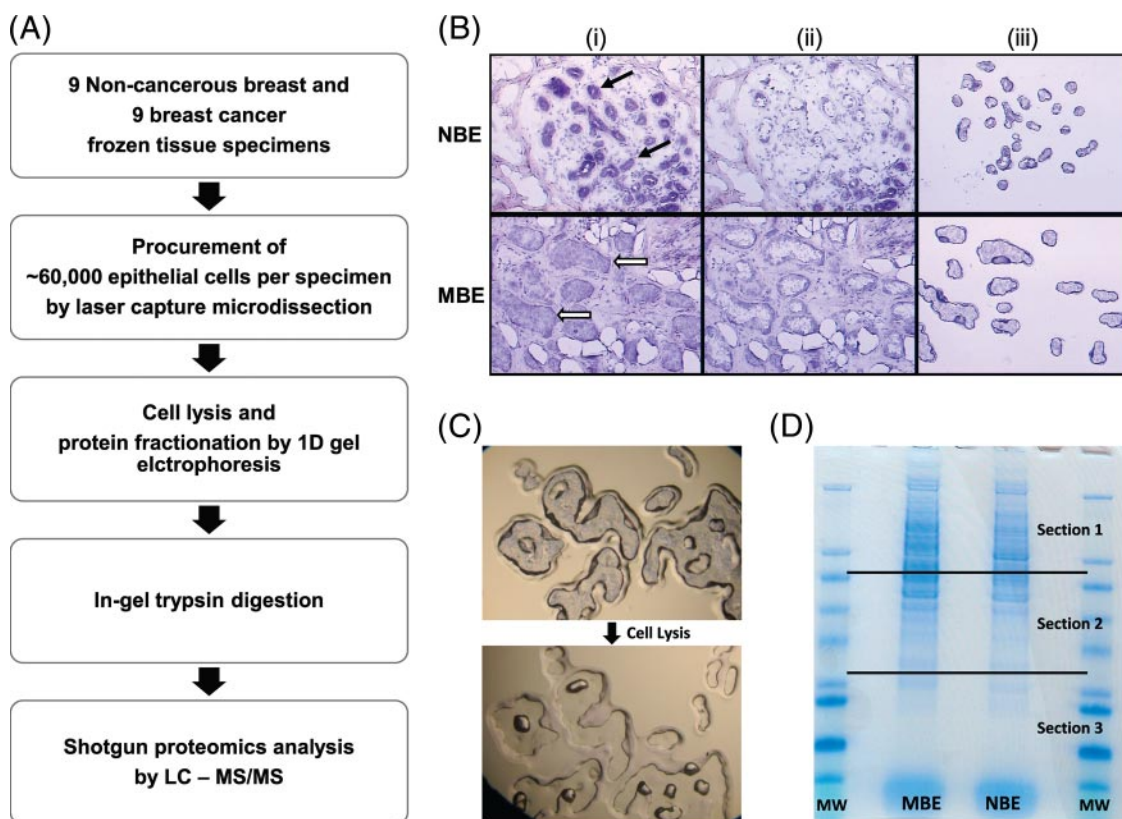
Spl-based ES for the original dataset and 20,000 phenotype class label permutations were computed to facilitate estimation of the false discovery rate (FDR). As described (29), ES values for each protein list were mean normalized across sample label permutations to yield a normalized enrichment score, whose distribution across the observed and permuted datasets was used to compute a q-value for each protein list. Protein sets with q-value  $< 0.25$  (FDR  $< 25\%$ ) were considered as significantly enriched sets. Similarly as in GSEA (29), “Leading edge” protein subsets were computed for each significant protein set by selecting the proteins with more extreme rank than the rank corresponding to the ES value. The PSEA algorithm, which uses the Spl for ES calculation for the proteomics data, was implemented in MATLAB R2009b (Mathworks, Natick, MA). PSEA is available upon request.

**Myoepithelial Marker Analysis**—Myoepithelial marker expression was analyzed in our dataset by constructing a “normal myoepithelial” marker list based on the data from ref (40), where a transcriptomic analysis of flow sorted myoepithelial cells in normal breast tissue using serial analysis of gene expression (SAGE) was conducted. Briefly, 171 genes identified in that study were used as “statistically significantly more abundantly expressed in myoepithelial cells and myofibroblasts than in all other cell types combined.” Because these genes were identified from the analysis of both normal and malignant myoepithelial cells and myofibroblasts, this list was filtered down to 108 genes that displayed at least 10 SAGE counts of gene expression in normal myoepithelial cells, to yield the “normal myoepithelial marker” list. It was found that 29 of these 108 genes had measurable protein expression in our dataset.

**Gene Ontology Analysis Using GOMiner**—Gene ontology (GO) terms for identified proteins were extracted, and overrepresented functional categories for differentially abundant proteins were determined by the high throughput GOMiner tool (National Cancer Institute, <http://discover.nci.nih.gov/gominer/>) (28). All proteins that were subjected to Spl analysis served as the background list, and GO terms with at least five proteins were used for statistical calculations. A *p* value for each term was calculated via the one-sided Fisher’s exact test, and FDR was estimated by permutation analysis using 1000 randomly selected sets of proteins sampled from the background list. Statistically significant (FDR  $< 25\%$ ) GO terms were clustered based on the correlation of associated proteins to minimize potential redundancy in significant GO terms.

## RESULTS AND DISCUSSION

**Proteomics Work Flow**—The proteomic work flow used in this study is summarized in Fig. 1A. Given the cellular heterogeneity of breast tissue, LCM was performed as a means to obtain a highly enriched *in situ* population of epithelial cells. Fig. 1B depicts pre- and postmicrodissected tissue images and procured cells. An infrared laser capture rather than a UV cutting microdissection approach was selected, as the thermal energy required by the latter approach could potentially damage cells and reduce protein harvest yield when microdissecting small tissue structures such as normal breast epithelial acini (14). Approximately 60,000 cells were collected from multiple consecutive tissue sections using 10 LCM caps per sample. After cell lysis and SDS-PAGE separation (Fig. 1C), the gel was cut into three sections and subjected to in-gel digestion, followed by LC-MS (Fig. 1D). Interestingly, after a total of 60 liquid chromatography/mass spectrometry runs, only ~3% of the total unique peptides identified were found in more than one gel section.



**FIG. 1. A, Shotgun proteomics workflow for the analysis of epithelial cells collected from phenotypically normal breast epithelium (NBE) and malignant invasive breast epithelium (MBE).** *B*, NBE (black arrows) and invasive MBE (white arrows) cells were microdissected using infrared laser capture from hematoxylin-stained frozen tissue sections (7  $\mu$ m). Images of premicrodissection (i), postmicrodissection (ii), and isolated normal or malignant epithelial cells (iii) are shown. *C*, Lysis of captured cells was performed directly on LCM caps. Images of the LCM cap before (top) and after (bottom) lysis are shown. *D*, Proteins from cell lysates were separated by gel electrophoresis and three fractions from each sample were collected by cutting gels into three sections according to molecular weight. Coomassie Blue stained gel images for two cases are shown with molecular weight markers.

Because of the minimal amount of available sample, proteomic analysis was conducted using label-free quantitation, thus minimizing potential sample losses during the labeling procedure. Our study focused on determining the protein abundance changes that are consistent across the 9 MBE and 9 NBE samples, rather than determining exact protein abundance ratios between individual MBE and NBE samples. In addition, the spectral counting method used in this work has been shown to be more sensitive to protein abundance changes (38) and to provide a broader dynamic range (41) than label free quantitation methods based on chromatographic peak area.

**Protein Identification**—From the 18 samples, 12,970 unique peptides were identified with a peptide FDR < 3%. In total 2588 proteins were identified, with 1715 proteins found to have at least two unique peptides (1693 proteins) or a single peptide identification with two different charge state assignments (22 proteins) (protein FDR < 0.3%). For subsequent analyses, 1623 proteins from the above 1715, which were detected in at least 2 out of 18 samples, were selected for further examination. Detailed information on the 1623 proteins can be found in Supplemental Tables S1A, B, and C.

As shown in Figs. 2A and B, the distribution of the number of proteins identified across the MBE and NBE samples varied in a narrow range (11.1% relative standard deviation for each group), and the distribution of the spectral counts corresponding to these proteins was also narrow (13.1% and 14.4% relative standard deviation, respectively). Of the 1623 proteins, the number of identified proteins found in both the NBE and MBE samples was 1336 (82.3% overlap), whereas the number of proteins identified exclusively in MBE and NBE were 193 and 94, respectively. The distributions of mean spectral counts for proteins identified exclusively either in the MBE or NBE samples, relative to the number of detection events, *i.e.* the number of samples in which a particular protein was detected, are shown in Fig. 2C. Proteins with a high number of mean spectral counts and high detection occurrences can be regarded as phenotype-specific, potential candidate biomarkers. Log ratios of spectral counts between MBE and NBE for the commonly detected 1336 proteins were plotted according to the number of detection events (Supplemental Fig. S1). Highly differentially abundant proteins can be considered as potential candidate biomarkers.

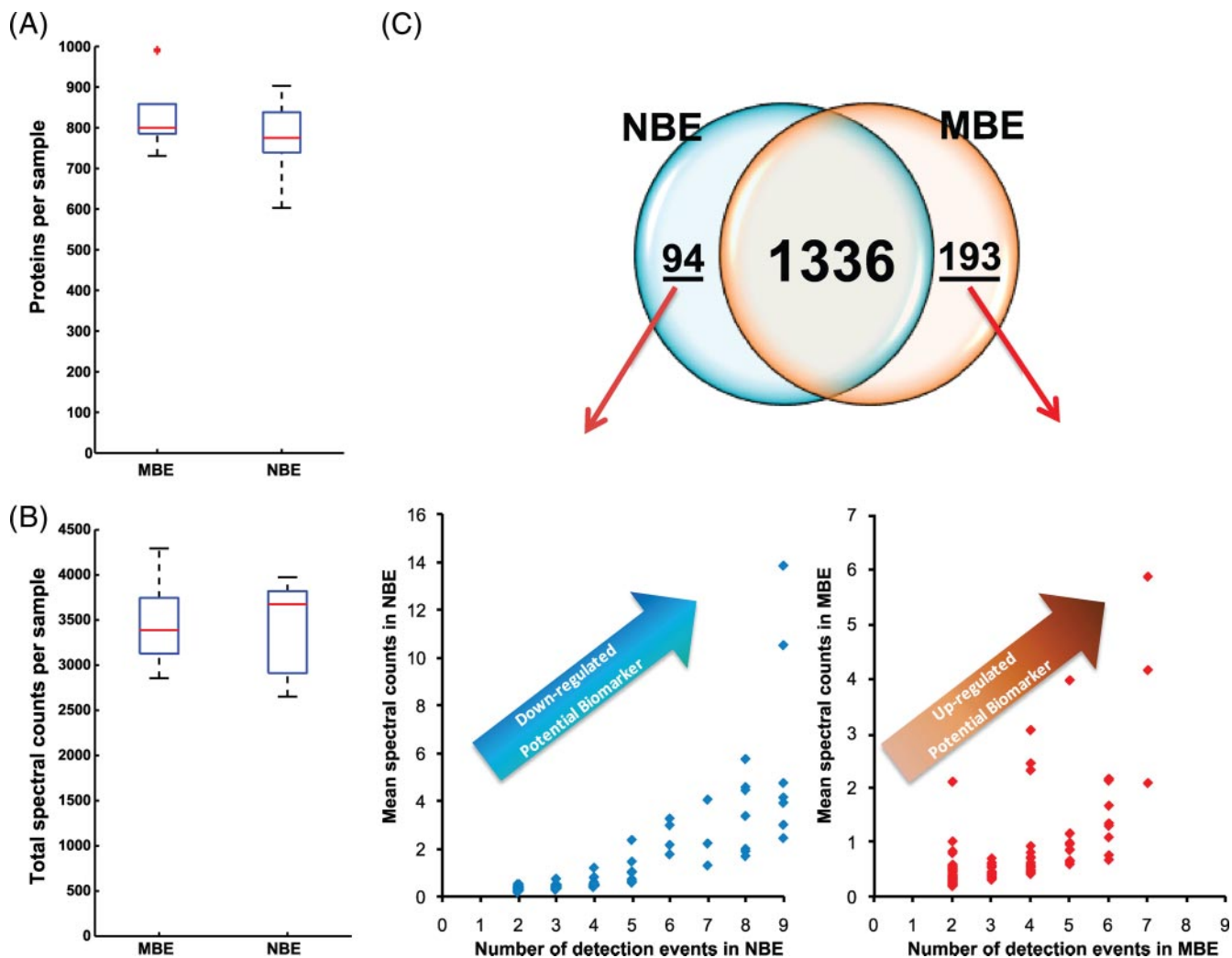


FIG. 2. **A**, Boxplot for the distribution of 1623 proteins across MBE and NBE samples. **B**, Boxplot for the distribution of total spectral counts for 1623 proteins across the MBE and NBE samples. **C**, Venn diagram for the distribution of proteins between the MBE and NBE samples and the mean spectral count distribution of MBE-specific (*lower right*) and of NBE-specific (*lower left*) proteins according to detection occurrences.

In order to evaluate reproducibility of the proteomic analyses in more detail, pairwise comparisons of the spectral counts for proteins identified in individual samples were generated. In order to limit the effect of a few proteins with a high number of spectral counts, log normalized spectral counts were used for these comparisons. Because there were nine samples per group, 153 nonredundant sample pairs could be constructed; Supplemental Fig. S2 shows a subset of this data for 36 plots corresponding to all NBE sample pairs. In order to provide a visualization of similarity of individual samples for all 153 pairs, Pearson correlation was used as the distance metric for clustering of samples (see Supplemental Fig. S3 for the resulting dendrogram). In order to minimize the effect of bias in sample preparation and liquid chromatography-MS, all samples were analyzed blind. Significantly, samples separated into two large clusters corresponding to NBE and MBE. These results demonstrate that

the spectral counts for individual proteins were reproducible across sample groups and that the measure is thus suitable for assessment of protein abundance in this work. In addition, further examination of this data using multidimensional scaling indicated higher similarity for proteins in the NBE than the MBE samples, reflecting greater protein variability in individual MBE samples.

**Determination of Differentially Abundant Proteins by Spectral Index Analysis**—The Spl values were calculated for the 1623 identified proteins, and the Spl distribution is shown in Supplemental Fig. S4A. Proteins with statistically significant Spl values were determined by permutation analysis (25). The frequency distribution of Spl values for 2000 randomly permuted sample phenotype classes (see Supplemental Fig. S4B) and the ranges encompassing 99%, 95%, and 90% of all Spls were determined to be  $Spl < 0.556$ ,  $0.422$ , and  $0.343$ , respectively. A total of 298 proteins were found to be



differentially abundant at 95% confidence level, decreasing to 121 proteins at the 99% level; these proteins are listed in Supplemental Table S2A and B.

Generally, verification of quantitative proteomic results involves immunohistochemical (IHC) analysis with monoclonal antibodies. Because of the expense and effort in developing such IHC only a few proteins are ever examined. Rather than performing IHC analysis on a minimal number of proteins in-house, we took advantage of IHC data available in the Human Protein Atlas (HPA ver. 5.0, 8832 antibodies and 7,334,244 images) (26). As an unbiased approach, we randomly selected proteins from the most significant differentially abundant proteins (121 proteins at the 99% confidence level) and assessed HPA IHC data. Because not all proteins are adequately represented in the HPA with sufficiently high quality image sets or with high quality antibodies, we assessed 44 random proteins in our list to accumulate 25 proteins that met the following IHC quality assessment criteria: (1) more than one normal breast sample reported; (2) eight or more breast cancer samples reported; (3) the HPA antibody verification score was moderate (staining pattern consistent with experimental and/or bioinformatic data) or high (two independent antibodies targeting one protein yielding similar staining patterns consistent with experimental and/or bioinformatic data). For IHC data meeting our quality assessment criteria, HPA images were manually inspected and staining visually quantified to determine (1) whether the protein of interest was overexpressed, unchanged, or underexpressed in MBE relative to NBE and (2) whether differential expression was localized to a particular tissue compartment (e.g. myoepithelial or epithelial cells). Of the 25 randomly selected proteins, 19 (>75%) were found to have concordant Spl-based proteomic differential expression with that determined by IHC (Table II). Five of the six proteins demonstrating apparent discordant results were determined to be down-regulated in tumor samples by Spl whereas unchanged by IHC. There are at least two possible explanations for these apparently discrepant results. First, the five proteins with equivocal expression by IHC may be differentially expressed to a degree that is undetectable with the relatively limited dynamic range of immunoperoxidase IHC staining. Second, it is important to note that the breast cancer subtypes (e.g. luminal, basal etc.) are not specified in the HPA database whereas our cancer tissue samples are of the luminal subtype. Thus, given that human breast cancers demonstrate subtype-specific proteomic expression, it is possible that luminal subtype is underrepresented in the HPA IHC analysis of one or more of the six discordant proteins. Taken together, it can be concluded that the results in Table II are a verification of our Spl-based quantitative proteomic data.

In addition to the randomly selected 25 differentially expressed proteins (99% confidence level) in Table II, we selected several high confidence (99%) differentially expressed proteins for verification that are known to be important in breast cancer. For example, fibronectin (Spl = 0.991) is a

protein known to be up-regulated in primary breast tumors (42, 43), and a recent study showed increased fibronectin gene expression in laser microdissected cells of invasive lobular and ductal carcinoma, relative to normal cells (44). In addition, a fibronectin-dependent mammary epithelial cell morphogenesis study found that fibronectin may play a key role in breast cancer progression by reversing growth arrest, disrupting acinar structure, and reinitiating cell growth (45). Another protein, major vault protein (Spl = 0.865) (included in Table II), showed consistent overexpression in breast cancer tissue in HPA database. Direct interaction among major vault proteins, vault RNA, and the estrogen receptor (ER) was identified in a MCF-7 breast cancer cell line by immunoprecipitation (46). Finally, anterior gradient protein 2 homolog (AGR2, Spl = 0.763) has been shown to be up-regulated in breast cancer at both the RNA and protein levels and its expression was found to be strongly correlated with ER positivity (47–50).

Differential expression of two novel MBE-enriched proteins, isoleucyl-tRNA synthetase, mitochondrial (Spl = 0.673) and  $\Delta(3,5)\text{-}\Delta(2,4)\text{-dienoyl-CoA}$  isomerase, mitochondrial (Spl = 0.564), were also in agreement with IHC profiles in HPA. These two proteins have no previous association with breast cancer. Isoleucyl-tRNA synthetase, mitochondrial catalyzes ATP-involved attachment of isoleucine to its related tRNA. This protein belongs to class Ia aminoacyl-tRNA synthetases that recognize hydrophobic amino acids including Ileu, Leu, Val, Met, Cys, and Arg (51–53). Generally, aminoacyl-tRNA synthetase plays a critical role in translation of genetic information during protein synthesis, and their functional connection to diseases, such as neuronal diseases, cancers, and autoimmune diseases, have been reviewed (54). Turning to  $\Delta(3,5)\text{-}\Delta(2,4)\text{-dienoyl-CoA}$  isomerase, mitochondrial, the protein, known to be localized at both peroxisomes and mitochondria, is involved in fatty acid  $\beta$ -oxidation pathways. In rats and mice,  $\Delta(3,5)\text{-}\Delta(2,4)\text{-dienoyl-CoA}$  isomerase, mitochondrial, has been shown to be regulated by peroxisome proliferator-activated receptor  $\alpha$ , a receptor that which plays a key role in peroxisome proliferator-induced carcinogenesis (55–58). Further study of these two novel proteins in breast cancer is required. IHC profile details of all proteins discussed earlier are listed in Supplemental Table S3.

*Protein Set Enrichment Analysis (PSEA)*—Beyond examining lists of proteins, we sought to obtain biological insight into breast cancer from our data by employing sophisticated annotation tools, namely PSEA. Importantly, unlike enrichment approaches that use the Fisher's exact test to determine statistical significance, PSEA does not require *a priori* specification of differentially abundant proteins based on a significance threshold. Instead, as for GSEA, PSEA uses the *complete* list of identified proteins along with their relative abundances expressed by Spl. As a result, PSEA reduces the chance that a particular protein set is missed because some proteins below the significant threshold may not be

TABLE II

Verification of the relative expression level for 25 proteins, randomly selected from 121 differentially abundant proteins at the 99% confidence level, by IHC profiles found in HPA

Accession number	Protein description	Spl	Antibody	IHC staining intensity				Verification	
				Cellular compartment	Strong	Moderate	Weak		Negative
P55268	Laminin subunit $\beta$ -2	-0.778	CAB000053	Normal myoepithelium	3	0	0	0	Yes
				Tumor epithelium	0	0	0	11	
P53621	Coatomer subunit $\alpha$	0.576	HPA028024	Normal epithelium	0	3	0	0	Yes
				Tumor epithelium	7	4	0	0	
Q14764	Major vault protein	0.865	HPA002321	Normal epithelium	0	0	3	0	Yes
				Tumor epithelium	4	6	1	0	
P53396	ATP-citrate synthase	0.635	HPA022434	Normal epithelium	0	0	3	0	Yes
				Tumor epithelium	1	8	2	0	
P06753	Tropomyosin $\alpha$ -3 chain	0.673	HPA009066	Normal epithelium	0	0	1	1	Yes
				Tumor epithelium	0	1	8	2	
P19971	Thymidine phosphorylase	0.558	HPA001072	Normal epithelium	0	0	0	3	Yes
				Tumor epithelium	2	7	3	0	
P49257	Protein ERGIC-53	0.564	HPA002320	Normal epithelium	0	2	0	0	Yes
				Tumor epithelium	2	7	3	0	
P35222	Catenin $\beta$ -1	-0.631	CAB000108	Normal epithelium	3	0	0	0	No
				Tumor epithelium	6	5	0	0	
Q01995	Transgelin	-0.567	CAB001447	Normal myoepithelium	2	0	0	0	Yes
				Tumor epithelium	0	0	1	7	
Q9BXN1	Asporin	0.667	HPA024230	Normal epithelium	0	2	0	0	No
				Tumor epithelium	0	0	10	1	
Q16822	Phosphoenolpyruvate carboxykinase [GTP], mitochondrial	0.667	CAB018734	Normal epithelium	0	0	0	3	Yes
				Tumor epithelium	0	3	3	5	
P42224	Signal transducer and activator of transcription 1- $\alpha/\beta$	0.574	HPA000931	Normal epithelium	0	0	2	0	Yes
				Tumor epithelium	1	4	3	1	
Q9Y6C2	EMILIN-1	-0.667	HPA002822	Normal myoepithelium	0	3	0	0	Yes
				Tumor epithelium	0	0	0	12	
P51911	Calponin-1	-1.000	CAB000007	Normal myoepithelium	2	0	0	0	Yes
				Tumor epithelium	0	0	2	8	
P19012	Keratin, type I cytoskeletal 15	-0.817	HPA024554	Normal epithelium	2	0	0	0	Yes
				Tumor epithelium	1	0	1	10	
Q12907	Vesicular integral-membrane protein VIP36	0.590	HPA003927	Normal epithelium	0	0	0	3	Yes
				Tumor epithelium	2	6	1	2	
Q9UN36	Protein NDRG2	-0.889	HPA002896	Normal epithelium	2	0	0	0	Yes
				Tumor epithelium	0	1	9	0	
P49419	$\alpha$ -amino adipic semialdehyde dehydrogenase	-0.760	HPA023296	Normal epithelium	0	0	2	0	No
				Tumor epithelium	0	1	9	0	
P09211	Glutathione S-transferase P	-0.922	HPA019779	Normal epithelium	0	3	0	0	No
				Tumor epithelium	2	5	0	4	
Q9HB71	Calcyclin-binding protein	0.667	HPA025753	Normal epithelium	0	2	0	0	Yes
				Tumor epithelium	7	4	0	0	
Q9ULZ3	Apoptosis-associated speck-like protein containing a CARD	0.691	CAB006853	Normal epithelium	0	2	0	0	Yes
				Tumor epithelium	7	3	0	0	
P68133	Actin, $\alpha$ skeletal muscle	0.698	CAB000045	Normal epithelium	0	0	0	3	Yes
				Tumor epithelium	0	1	0	9	
P27338	Amine oxidase [flavin-containing] B	-0.688	HPA002328	Normal epithelium	2	0	0	0	No
				Tumor epithelium	9	3	0	0	
P04083	Annexin A1	-0.743	HPA011271	Normal epithelium	0	2	0	0	Yes
				Tumor epithelium	0	0	0	12	
P12429	Annexin A3	-0.904	HPA013398	Normal epithelium	0	3	0	0	No
				Tumor epithelium	0	8	0	4	

included in the differentially abundant list. PSEA also preferentially weighs functional protein sets with strong (*i.e.* high Spl) and concordant abundance changes (*e.g.* down-regulation, up-regulation) relative to protein sets showing weak and discordant differential abundances. In contrast, commonly used enrichment-based tools, such as DAVID or GOMiner (27, 28), only examine proteins above the significance cutoff level, and do not explicitly take into account the extent or direction of change in abundance of the included proteins.

PSEA was applied to our proteomic data, and representative results are displayed in Fig. 3. Members of the protein set

“V\$SRF\_Q5\_01” (binding motifs for the serum response factor (SRF)) were highly enriched at the negative extreme of the Spl distribution, yielding an ES value of -0.51 (Fig. 3, top panel). As shown in the middle panel of Fig. 3, the statistical significance of the calculated ES value was evaluated using a null distribution of ES constructed by permuting the phenotype class label 20,000 times. The original ES was found to be significant with a *p* value of 0.005 and q-value (FDR) of 0.20 (20%). In addition, as shown in the heat map of normalized spectral counts (Fig. 3, bottom panel), there is a clear difference between the MBE and NBE samples with respect to



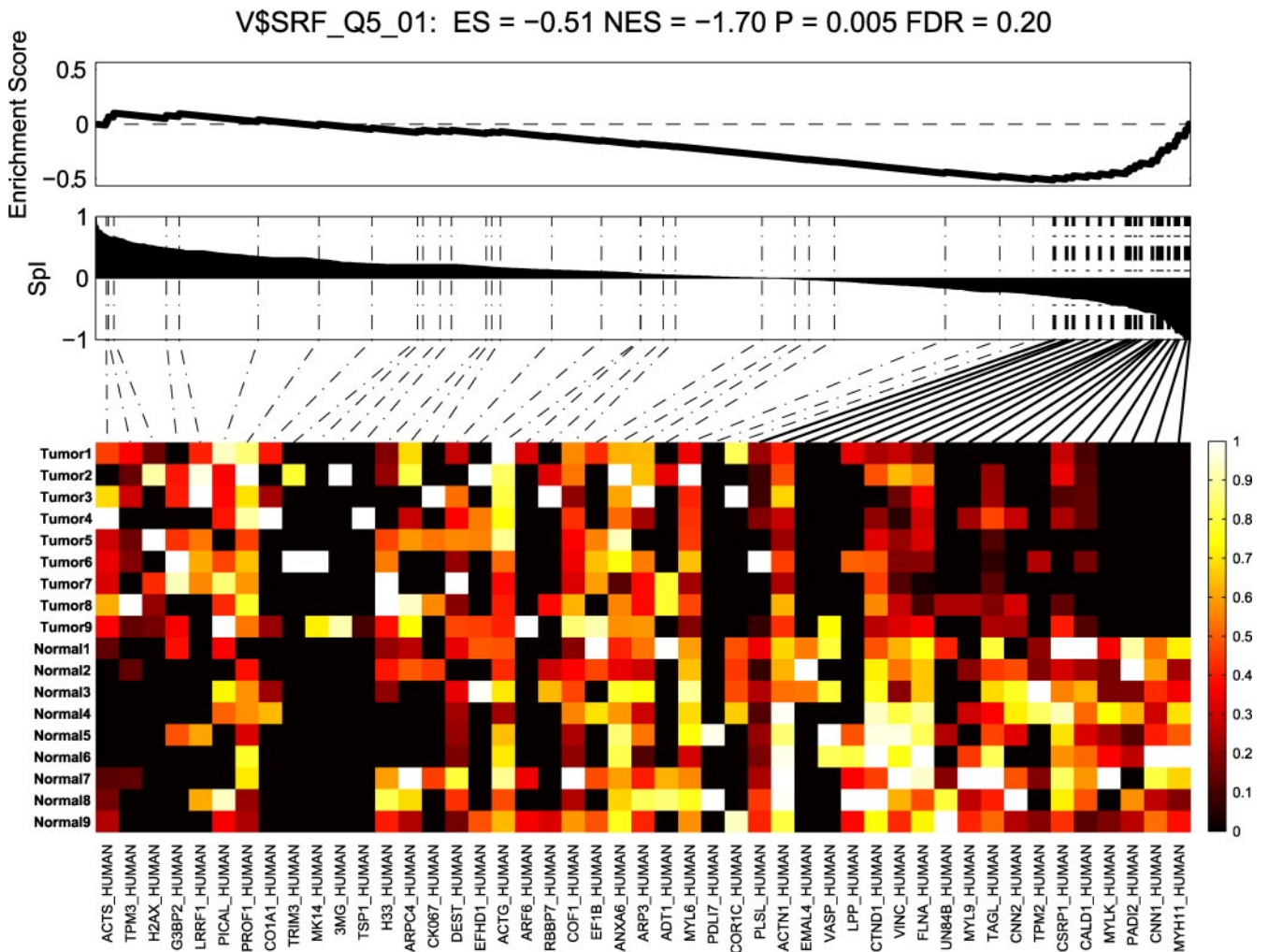


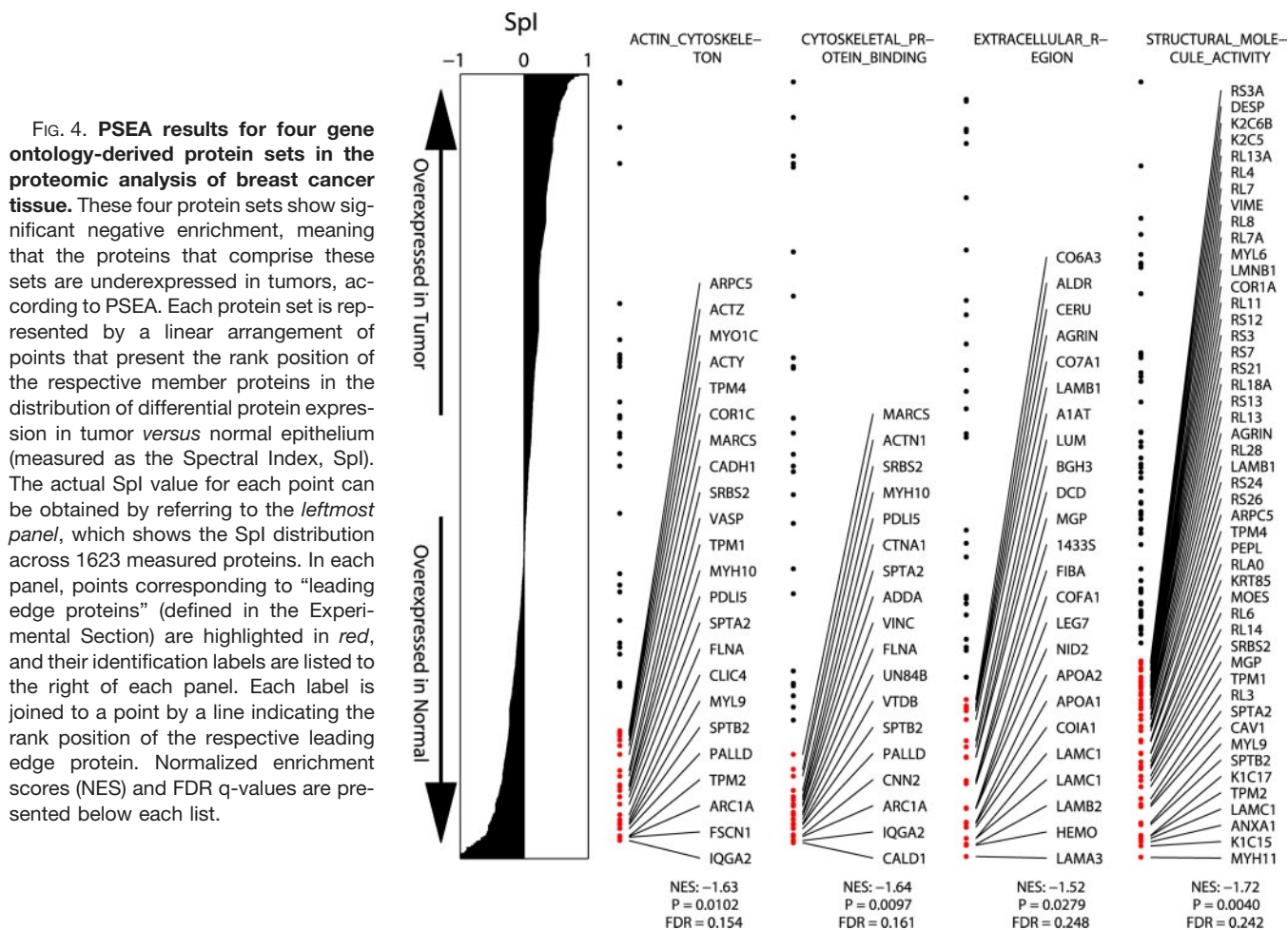
FIG. 3. **Protein set enrichment analysis (PSEA).** An enrichment score for the protein set is computed as the minimum of the running sum statistic, which is plotted in the *top panel*. Vertical notches on the spectral index, Spl, bar plot (*middle panel*) represent members of the protein set that were detected in our study, each of which contribute to the running sum statistic in proportion to their Spl value. The “leading-edge” members of the protein set (see Experimental Section), which drive the enrichment score, are represented by bold notches in the Spl bar plot. Each notch is represented by a column in the protein expression heat map shown in the bottom panel, where “leading edge” proteins are labeled in bold. The heat map intensities for each protein are scaled relative to the highest expression level of that protein across all samples. The PSEA example shown here is for the protein set “V\$SRF\_Q5\_01,” corresponding to genes with promoter regions [-2kb, 2kb] around the transcription start site containing the motif for: serum response factor (c-fos serum response element-binding transcription factor).

leading-edge proteins (labeled in bold), indicating a significant change in abundance for proteins that belong to the V\$SRF\_Q5\_01 set.

Using PSEA, 35 molecular signature database (MSigDB)-derived protein sets were found to be significant in the proteome dataset at the FDR cutoff of 25%, and these groupings were composed of 12 “curated,” 10 “motif-based,” 8 “computational,” and 5 Gene Ontology-based protein sets. Statistical measures, associated proteins, and leading-edge proteins for the significant protein sets are listed in Supplemental Table S4. Intriguingly, all significant protein sets found in our analysis were found to be down-regulated in MBE (*i.e.* associated with an normalized enrichment score <0).

Consistent with previous studies (59–62), breast carcinomas showed decreased expression of structural cytoskeletal proteins (*actin cytoskeleton, structural molecule activity, cytoskeletal protein binding*) (Fig. 4). Tumors were significantly depleted in proteins containing an extracellular region, with collagens (CO6A3, CO7A1, COFA1, and COIA1), laminins (LAMC1, LAMC1, and LAMB2), transforming growth factor  $\beta$ , fibrinogen, serpin peptidase inhibitor, lumican, 14-3-3 protein  $\sigma$ , and hemopexin, forming the leading edge.

Another result in the PSEA of the ER-positive breast tumor samples was the significant down-regulation of proteins encoded by “genes whose expression is consistently negatively correlated with estrogen receptor status in breast cancer” (*BRCA\_NEG\_ER*, Supplemental Table S4 and Fig. S5). We



also found down-regulation of proteins associated with several “computational” protein sets derived from module network analysis of cancer related gene expression data (Supplemental Table S4 and Supplemental Fig. S6). Significant down-regulated sets from this group include “developmental processes” (module\_220) and “adhesion molecules” (module\_122). These results are consistent with processes of de-differentiation and loss of cell adhesion that are known to occur in cancer cells.

Importantly, enrichment of multiple protein sets associated with transcription factor (TF) binding motifs were found (Fig. 5). The most significant were down-regulated protein sets associated with motifs for the TFs, activator protein 1 (AP-1), nuclear factor  $\kappa$ -light-chain-enhancer of activated B cells (NF- $\kappa$ B), sterol regulatory element binding protein1 (SREBP1), serum response factor (SRF), P53, and SMAD. Because these motif-based lists were built purely from sequence analysis, activating or inhibitory relationships have not been determined for most of these target genes (29). Down-regulation of TF-related target genes can result from decreased TF activity or an inhibitory effect of the TF on target genes. It should be stressed that, although we did not directly detect these tran-

scription factors in our proteomic analysis, the proteomic results implicate specific regulatory pathways in breast tumorigenesis based on measurement of downstream targets of these TFs.

The observation of differential abundance of proteins with genes containing AP-1 and NF- $\kappa$ B motifs in ER+ tumors provides an intriguing intersection with previous knowledge regarding ER nuclear signaling. Estrogen classically exerts its transcriptional effects through ligand dependent ER binding of estrogen response elements (ERE) (63). ER also functions through the nonclassical “ERE-independent” estrogen signaling pathway by modulating the activities of AP-1, NF- $\kappa$ B, and other TFs on their cognate transcriptional targets that lack the ERE in their promoter regions (64, 65). ER is specifically known to inhibit the DNA-binding activity of NF- $\kappa$ B (66). Indeed, the majority of leading-edge proteins for both AP-1 (25 out of 34) and NF- $\kappa$ B (four out of five) are not associated with an ERE in the promoter region of their gene. Our proteomic results thus suggest ERE-independent estrogen signaling through these TFs could be a major participant in breast tumorigenesis, strengthening the results of a previous transcriptomic study (67).

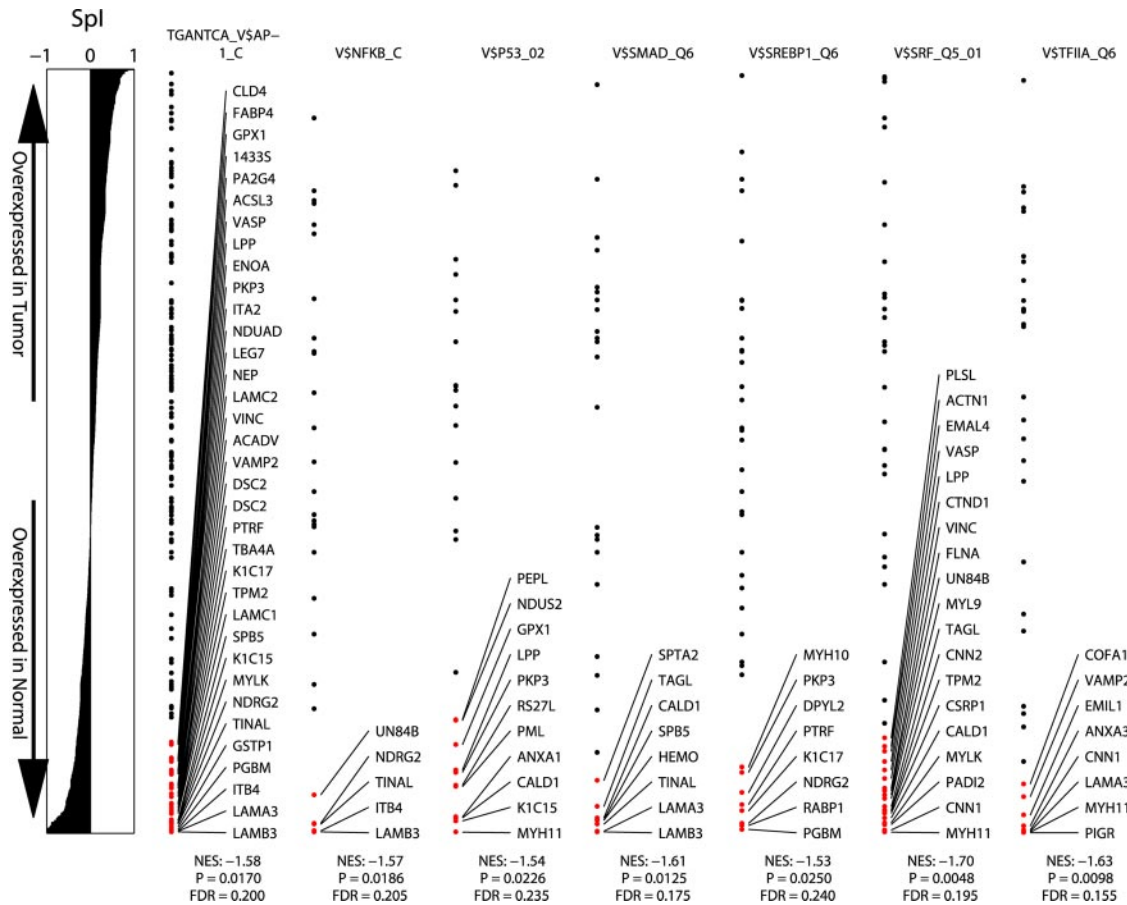


FIG. 5. PSEA results for seven transcription factor binding motif-derived protein sets resulting from the proteomic analysis of invasive breast cancer tissue relative to normal breast tissue. These protein sets represent putative regulatory associations that are derived from analysis of upstream promoter sequences of the corresponding genes. The sets show significant negative enrichment, meaning that the proteins that comprise these sets are underexpressed in tumors relative to normal according to PSEA. Each set is represented by a linear arrangement of points that present the rank position of each member protein in the tumor versus normal epithelium spectral index (Spl) distribution. The actual Spl value for each point can be obtained by referring to the *leftmost* panel, which shows the Spl distribution across 1623 measured proteins and which is aligned to the points. In each panel, points corresponding to “leading-edge proteins” (see Experimental Section) are highlighted in red, and their identification labels are listed to the *right* of each panel. Each label is joined to a point by a line indicating the rank position of the respective leading-edge protein. Normalized enrichment scores (NES) and FDR q-values are presented below each list.

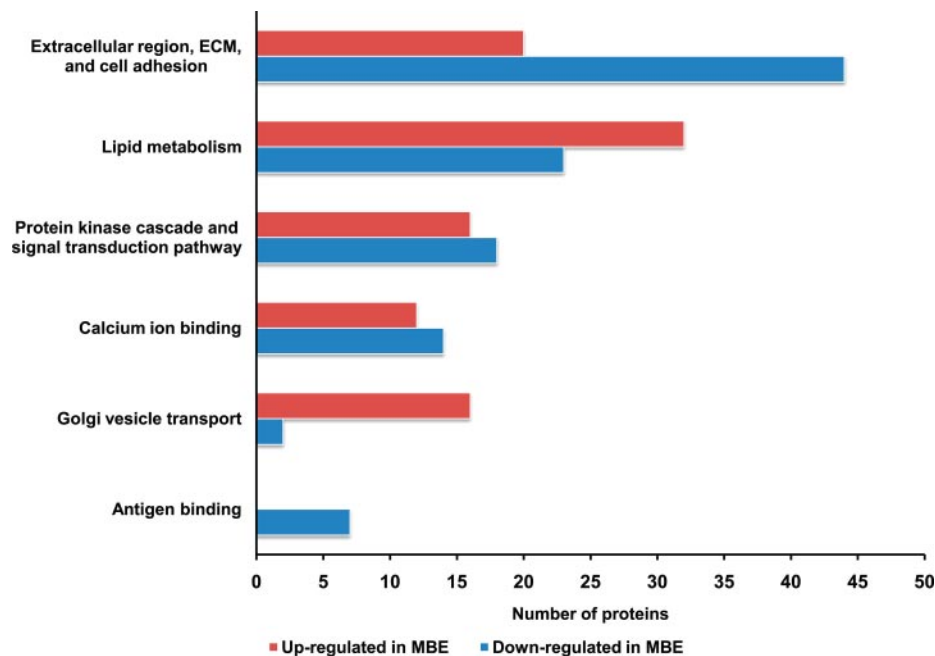
Protein sets corresponding to SRF and SREBP binding motifs were significantly down-regulated in breast tumors in the dataset (Supplemental Table S4). SRF interacts with AP-1 to regulate multiple genes in the early response to growth factor stimulation following serum starvation in cultured cell lines (68). SRF is critical for embryogenesis, mesoderm formation, and skeletal muscle growth and has recently been shown to be necessary for experimental metastasis in a breast cancer xenograft model (69). The differential expression of proteins associated with SRF binding elements may thus be related to ERE-independent ER signaling. SREBP1 is a sterol regulatory element-1 binding factor associated with the transcription of sterol dependent genes, such as the low-density lipoprotein receptor. SREBP1 has been previously shown to bind estrogen receptor (70); however, neither this protein nor its target genes have previously been implicated in breast cancer.

TP53 is frequently mutated and functionally inactivated in many breast cancers (71). Our finding of significant down-regulation of multiple putative TP53 targets in breast cancer, including leading edge proteins such as glutathione peroxidase 1, lipoma-preferred partner, annexin A1, caldesmon, keratin, type I cytoskeletal 15, and myosin-11, suggest that the expression of these proteins may be positively regulated by TP53 in normal cells. Indeed, the transcription of two of these proteins (annexin A1 and keratin, type I cytoskeletal 15) has been shown to be dependent on TP53 in human cells (72, 73).

SMAD3 is known to be inhibited by estrogen receptor activation in human cells (74, 75). Smad gene expression has been shown to decrease in breast cancer, and expression correlates with an invasive phenotype (76). PSEA found negative enrichment of SMAD targets, including leading-edge proteins, spectrin  $\alpha$  chain, brain, transgelin, caldesmon, he-



FIG. 6. **Representative enriched, functional clusters for differentially abundant proteins identified by the GOMiner tool.** The *red* bar indicates the number of proteins enriched in MBE for a given cluster whereas the *blue* bar represents the number of proteins enriched in NBE. All functional clusters identified are listed in Supplemental Table, S6 with statistical significance measures, and proteins associated with these clusters are listed in Supplemental Table, S7 with their Splis.



mopexin (HEMO), and laminin subunit  $\alpha$ -3 and  $\beta$ -3 (LAMA3 and LAMB3). Of these, one (transgelin) has been shown have its transcription directly activated by SMAD (77). Though SMAD3 expression was not measured in our samples, the down-regulation of this (and other) SMAD targets is consistent with inactivation of this pathway in breast tumors. Other significant curated protein sets are discussed in Supplemental Information.

PSEA was also used to explore differential admixture of myoepithelial cells in tumor *versus* normal breast epithelium and examine any contribution of this difference to the enrichment results. LCM is unable to separate normal breast epithelium from admixed myoepithelial cells resulting in potential selective enrichment of myoepithelial proteins in NBE samples as compared with MBE samples. Thus, myoepithelial contamination in NBE samples could potentially lead to enrichment of myoepithelial markers among the down-regulated protein sets. This possibility was investigated by assessing enrichment of myoepithelial specific markers in the proteomic dataset. For a list of myoepithelial specific proteins, we used genes identified in a global transcriptomic analysis of flow cytometry purified subpopulations of the normal breast tumor microenvironment (40). PSEA scoring of this list against our expression data showed significant down-regulation of myoepithelial proteins in MBE (Supplemental Fig. S7). However, these myoepithelial proteins comprised only a small fraction of significant protein sets and their leading edges (Supplemental Table S5). This analysis suggests that the relatively low level of myoepithelial cell contamination in our NBE samples is not impacting the PSEA results.

**Gene Ontology Term Enrichment Analysis**—GOMiner analysis, one of the well-known functional GO term enrichment

analysis tools, was also performed with the present LCM derived dataset. As noted earlier, GOMiner requires a specific list of proteins or genes that statistically discriminate one state from another. In this work, differentially abundant proteins at the 95% confidence level were arbitrarily selected for annotation analysis, and GO terms were then compared with those from the total of 1623 proteins. Statistically significant (FDR <25%), overrepresented 60 GO terms associated with the 298 proteins could be clustered to nine functional categories (Supplemental Table S6 and S7). Six of these categories that are representative of all the results are shown in Fig. 6.

In agreement with the PSEA results, GOMiner annotation finds that the majority of proteins (69%) relating to the extracellular region, extracellular matrix, and cell adhesion were down-regulated in MBE. At the same time, and importantly, GOMiner analysis adds, beyond PSEA, several statistically significant biological features, *e.g.* up-regulation of Golgi vesicle transport, down-regulation of antigen binding, and up- and down-regulation in calcium ion binding related components in MBE. These additional findings revealed by GOMiner are likely related to protein sets that are not included in the MSigDB database that underlies PSEA. Additional complementary findings by GOMiner are discussed in the following.

GOMiner analysis also discovered significant alterations in lipid metabolism between normal and tumor breast epithelium. These changes included proteins involved in lipid binding, transport, and sterol binding and biosynthesis. Most of these significant sets were formed by a mixture of up-regulated and down-regulated proteins, with the exception of “sterol metabolic process” and “sterol binding.” These results provide an intriguing connection with the significantly reduced expression of downstream targets for the sterol-regulatory



process (SREBP1 binding motifs) observed with PSEA. Interestingly, leading-edge proteins of the SREBP1 set have minimal overlap (only one protein, RABP1) with proteins in the lipid metabolism cluster constructed from GOMiner results. These results point to different components of a potentially complex signaling steroid metabolism pathway that appears to be significantly altered in breast carcinoma.

Additional comparison of GOMiner and PSEA analyses points to potentially novel regulatory circuitry connecting altered NF- $\kappa$ B and SMAD signaling to breast tumorigenesis. GOMiner revealed significantly altered expression of “I- $\kappa$ B kinase/NF- $\kappa$ B cascade” in MBE, with the majority of proteins demonstrating up-regulation in MBE. Literature based follow up revealed six of seven of these up-regulated proteins have been implicated in the direct activation of NF- $\kappa$ B. For example, peroxiredoxin-4 is known to activate NF- $\kappa$ B via modulation of I $\kappa$ B- $\alpha$  phosphorylation in the cytoplasm (78), and genes for bone marrow stromal antigen 2 and peptidyl-prolyl *cis-trans* isomerase FKBP1A have been identified as potential activators of NF- $\kappa$ B (79). Interestingly, NF- $\kappa$ B transcriptional targets were found to be significantly down-regulated by PSEA. A similarly relevant signaling-based GO category is SMAD binding, which consists primarily of down-regulated proteins. Interestingly, proteins in this GO category show no overlap with the SMAD motif protein set found to be significantly down-regulated in MBE by PSEA. Taken together, our analyses suggest breast tumors may be associated with reduced SMAD signaling and repression of downstream SMAD signaling pathways. The results from the combined PSEA and GOMiner analysis merit more focused functional studies.

#### CONCLUSIONS

Comparative proteomic analysis of breast epithelium was performed using a combination of LCM and shotgun proteomics. To our knowledge, this work is the most comprehensive study comparing direct proteome signatures obtained from homogeneous cell populations of phenotypically normal and ER+ invasive malignant breast epithelium. The key features of this study included separate processing of each sample, thus providing a measure of biological variability. Furthermore, consistent proteomic profiles across 18 samples were obtained using a robust shotgun proteomics protocol. Nine biological replicates for each phenotype proved to be a good balance between the statistical significance and time necessary for analysis.

Highly MBE-enriched proteins were found to correlate well with tissue IHC profiles deposited in the Human Protein Atlas. This agreement suggest that our proteomics results may not be restricted to ER+ breast cancer samples, but may be generalized to all breast cancer subtypes, given the diversity of breast cancer subtype samples in the HPA database.

Global changes in multiple protein signatures, according to breast epithelial malignancy, were functionally assessed by

employing PSEA, an annotation method that uses the complete quantitative profiles of all identified proteins with no arbitrary cutoffs. PSEA of ER+ breast epithelium not only found many confirmatory biological findings but also revealed significant changes of downstream targets for a number of common transcription factors. This result suggests that ERE-independent ER signaling could be a key pathway in breast cancer progression. GOMiner analysis with 298 differentially abundant proteins at or above the 95% confidence level provided complementary results to PSEA, and the combination of the two approaches revealed extensive insights into the molecular participants involved in signaling cascades.

PSEA is applicable to any quantitative proteome dataset, and the flexibility of PSEA allows exploration of proteome datasets in many different biological contexts. For example, PSEA using the proteome-unique sets, such as protein-protein interaction database, could provide biological insights that are distinguished from those by genome-based annotations. Furthermore, inclusion of all differentially abundant proteins or genes with PSEA will avoid missing possible important biological relationships behind target candidates (e.g. drug targets and biomarkers).

Future studies will include correlation analysis between a previously reported transcriptome dataset (33) and our proteome dataset, which can lead to significant additional protein networks and provide further insight into the biology of breast cancer progression. In addition, comparing proteome profiles of phenotypically normal epithelium in noncancerous and cancerous breast tissues from the same patient could be interesting, because normal tissue in close proximity to breast carcinoma has been known to lose its heterozygosity (31). Such findings should broaden our understanding of the molecular boundary between benign and malignant breast diseases.

*Acknowledgments*— We thank the Enterprise Research IS group at Partners Healthcare for their in-depth support and access to the HPC facilities. Contribution number 962 from the Barnett Institute.

\* This work was supported by the Susan G. Komen Breast Cancer Foundation (BLK and DCS), National Institutes of Health grant GM15847 (BLK), National Institutes of Health grant R01CA112021 (DCS), the NCI National Institutes of Health Specialized Programs of Research Excellence in breast cancer at Massachusetts General Hospital (DCS) and the Avon Foundation (DCS).

§ This article contains supplemental material.

‡ To whom correspondence should be addressed: Barry L. Karger, E-mail: b.karger@neu.edu; Dennis C. Sgroi, E-mail: dsgrói@partners.org.

\*\* These two senior authors contributed equally.

#### REFERENCES

1. Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Murray, T., and Thun, M. J. (2008) Cancer statistics, 2008. *CA-Cancer J. Clin.* **58**, 71–96
2. Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B., Desmedt, C., Larsimont, D., Cardoso, F., Peterse, H., Nuyten, D., Buyse, M., Van de Vijver, M. J., Bergh, J., Piccart, M., and Delorenzi, M. (2006) Gene expression profiling

- in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. *J. Natl. Cancer Inst.* **98**, 262–272
3. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Lønning, P., Eystein, and Børresen-Dale, A. L. (2001) Gene expression patterns of breast carcinoma distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10869–10874
  4. van't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536
  5. Ma, X. J., Wang, Z., Ryan, P. D., Isakoff, S. J., Barmettler, A., Fuller, A., Muir, B., Mohapatra, G., Salunga, R., Tuggle, J. T., Tran, Y., Tran, D., Tassin, A., Amon, P., Wang, W., Enright, E., Stecker, K., Estepa-Sabal, E., Smith, B., Younger, J., Balis, U., Michaelson, J., Bhan, A., Habin, K., Baer, T. M., Brugge, J., Haber, D. A., Erlander, M. G., and Sgroi, D. C. (2004) A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* **5**, 607–616
  6. Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lønning, P. E., Børresen-Dale, A. L., Brown, P. O., and Botstein, D. (2000) Molecular portraits of human breast tumours. *Nature* **406**, 747–752
  7. Brenton, J. D., Carey, L. A., Ahmed, A. A., and Caldas, C. (2005) Molecular classification and molecular forecasting of breast cancer: Ready for clinical application? *J. Clin. Oncol.* **23**, 7350–7360
  8. Hondermarck, H. (2003) Breast cancer - When proteomics challenges biological complexity. *Mol. Cell. Proteomics* **2**, 281–291
  9. Somiari, R. I., Somiari, S., Russell, S., and Shriver, C. D. (2005) Proteomics of breast carcinoma. *J. Chromatogr. B.* **815**, 215–225
  10. Bertucci, F., Birnbaum, D., and Goncalves, A. (2006) Proteomics of breast cancer - Principles and potential clinical applications. *Mol. Cell. Proteomics* **5**, 1772–1786
  11. Hondermarck, H., Tastet, C., El Yazidi-Belkoura, I., Toillon, R. A., and Le Bourhis, X. (2008) Proteomics of breast cancer: The quest for markers and therapeutic targets. *J. Proteome Res.* **7**, 1403–1411
  12. Kulasingam, V., and Diamandis, E. P. (2008) Tissue culture-based breast cancer biomarker discovery platform. *Int. J. Cancer* **123**, 2007–2012
  13. Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., Chuaqui, R. F., Zhuang, Z., Goldstein, S. R., Weiss, R. A., and Liotta, L. A. (1996) Laser Capture Microdissection. *Science* **274**, 998–1001
  14. Espina, V., Wulfschuh, J. D., Calvert, V. S., VanMeter, A., Zhou, W., Coukos, G., Geho, D. H., Petricoin, R. F., 3rd, and Liotta, L. A. (2006) Laser-capture microdissection. *Nat. Protoc.* **1**, 586–603
  15. Ma, X. J., Dahiya, S., Richardson, E., Erlander, M., and Sgroi, D. C. (2009) Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Res.* **11**, R7
  16. Wulfschuh, J. D., Sgroi, D. C., Krutzsch, H., McLean, K., McGarvey, K., Knowlton, M., Chen, S., Shu, H., Sahin, A., Kurek, R., Wallwiener, D., Merino, M. J., Petricoin, R. F., 3rd, Zhao, Y., and Steeg, P. S. (2002) Proteomics of human breast ductal carcinoma in situ. *Cancer Res.* **62**, 6740–6749
  17. Zang, L., Palmer, Toy, D., Hancock, W. S., Sgroi, D. C., and Karger, B. L. (2004) Proteomic analysis of ductal carcinoma of the breast using laser capture microdissection, LC-MS, and O-16/O-18 isotopic labeling. *J. Proteome Res.* **3**, 604–612
  18. Neubauer, H., Clare, S. E., Kurek, R., Fehm, T., Wallwiener, D., Sotlar, K., Nordheim, A., Wozny, W., Schwall, G. P., Poznanović, S., Sastri, C., Hunzinger, C., Stegmann, W., Schratzenholz, A., and Cahill, M. A. (2006) Breast cancer proteomics by laser capture microdissection, sample pooling, 54-cm IPG IEF, and differential iodine radioisotope detection. *Electrophoresis* **27**, 1840–1852
  19. Umar, A., Luider, T. M., Foekens, J. A., and Pasa-Tolic, L. (2007) NanoLC-FT-ICR MS improves proteome coverage attainable for similar to 3000 laser-microdissected breast carcinoma cells. *Proteomics* **7**, 323–329
  20. Sanders, M. E., Dias, E. C., Xu, B. J., Mobley, J. A., Billheimer, D., Roder, H., Grigorieva, J., Dowsett, M., Arteaga, C. L., and Caprioli, R. M. (2008) Differentiating proteomic biomarkers in breast cancer by laser capture microdissection and MALDI MS. *J. Proteome Res.* **7**, 1500–1507
  21. Johann, D. J., Rodriguez-Canales, J., Mukherjee, S., Prieto, D. A., Hanson, J. C., Emmert-Buck, M., and Blonder, J. (2009) Approaching Solid Tumor Heterogeneity on a Cellular Basis by Tissue Proteomics Using Laser Capture Microdissection and Biological Mass Spectrometry. *J. Proteome Res.* **8**, 2310–2318
  22. Umar, A., Kang, H., Timmermans, A. M., Look, M. P., Meijer-van Gelder, M. E., den Bakker, M. A., Jaitly, N., Martens, J. W., Luider, T. M., Foekens, J. A., and Pasa-Tolić, L. (2009) Identification of a Putative Protein Profile Associated with Tamoxifen Therapy Resistance in Breast Cancer. *Mol. Cell. Proteomics* **8**, 1278–1294
  23. Chin, K., DeVries, S., Fridlyand, J., Spellman, P. T., Roydasgupta, R., Kuo, W. L., Lapuk, A., Neve, R. M., Qian, Z., Ryder, T., Chen, F., Feiler, H., Tokuyasu, T., Kingsley, C., Dairkee, S., Meng, Z., Chew, K., Pinkel, D., Jain, A., Ljung, B. M., Esserman, L., Albertson, D. G., Waldman, F. M., and Gray, J. W. (2006) Genomic and transcriptional aberrations linked to breast cancer pathophysiology. *Cancer Cell* **10**, 529–541
  24. Gu, Y., Wu, S. L., Meyer, J. L., Hancock, W. S., Burg, L. J., Linder, J., Hanlon, D. W., and Karger, B. L. (2007) Proteomic analysis of high-grade dysplastic cervical cells obtained from ThinPrep slides using laser capture microdissection and mass spectrometry. *J. Proteome Res.* **6**, 4256–4268
  25. Fu, X., Gharib, S. A., Green, P. S., Aitken, M. L., Frazer, D. A., Park, D. R., Vaisar, T., and Heinecke, J. W. (2008) Spectral index for assessment of differential protein expression in shotgun proteomics. *J. Proteome Res.* **7**, 845–854
  26. Berglund, L., Björling, E., Oksvold, P., Fagerberg, L., Asplund, A., Szgyarto, C. A., Persson, A., Ottosson, J., Wernérus, H., Nilsson, P., Lundberg, E., Sivertsson, A., Navani, S., Wester, K., Kampf, C., Hober, S., Pontén, F., and Uhlen, M. (2008) A gene-centric human protein atlas for expression profiles based on antibodies. *Mol. Cell. Proteomics* **7**, 2019–2027
  27. Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., and Lempicki, R. A. (2003) DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol.* **4**, R60
  28. Zeeberg, B. R., Qin, H., Narasimhan, S., Sunshine, M., Cao, H., Kane, D. W., Reimers, M., Stephens, R. M., Bryant, D., Burt, S. K., Elnekave, E., Hari, D. M., Wynn, T. A., Cunningham-Rundles, C., Stewart, D. M., Nelson, D., and Weinstein, J. N. (2005) High-Throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of Common Variable Immune Deficiency (CVID). *BMC Bioinformatics* **6**, 168
  29. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550
  30. Ruth, I., Daniele, M., Rasoul, A.-K., Anthony, G., Gary, D. B., and Andrew, E. Pathway analysis of dilated cardiomyopathy using global proteomic profiling and enrichment maps. *Proteomics* **10**, 1316–1327
  31. Deng, G., Lu, Y., Zlotnikov, G., Thor, A. D., and Smith, H. S. (1996) Loss of Heterozygosity in Normal Tissue Adjacent to Breast Carcinomas. *Science* **274**, 2057–2059
  32. Oppenheimer, S. R., Mi, D., Sanders, M. E., and Caprioli, R. M. (2010) Molecular Analysis of Tumor Margins by MALDI Mass Spectrometry in Renal Carcinoma. *Journal of Proteome Research* **9**, 2182–2190
  33. Ma, X. J., Salunga, R., Tuggle, J. T., Gaudet, J., Enright, E., McQuary, P., Payette, T., Pistone, M., Stecker, K., Zhang, B. M., Zhou, Y. X., Varnholt, H., Smith, B., Gadd, M., Chatfield, E., Kessler, J., Baer, T. M., Erlander, M. G., and Sgroi, D. C. (2003) Gene expression profiles of human breast cancer progression. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5974–5979
  34. Sgroi, D. C., Teng, S., Robinson, G., LeVangie, R., Hudson, J. R., Jr, and Elkahloun, A. G. (1999) In vivo gene expression profile analysis of human breast cancer progression. *Cancer Res.* **59**, 5656–5661
  35. Licklider, L. J., Thoreen, C. C., Peng, J., and Gygi, S. P. (2002) Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. *Anal. Chem.* **74**, 3076–3083
  36. Guzzetta, A. W., and Chien, A. S. (2005) A double-vented tetraphasic continuous column approach to MuDPIT analysis on long capillary columns demonstrates superior proteomic coverage. *J. Proteome Res.* **4**, 2412–2419
  37. Rauch, A., Bellew, M., Eng, J., Fitzgibbon, M., Holzman, T., Hussey, P., Igra, M., Maclean, B., Lin, C. W., Detter, A., Fang, R., Faca, V., Gafken,

- P., Zhang, H., Whiteaker, J., Whitaker, J., States, D., Hanash, S., Paulovich, A., and McIntosh, M. W. (2006) Computational Proteomics Analysis System (CPAS): An extensible, open-source analytic system for evaluating and publishing proteomic data and high throughput biological experiments. *J. Proteome Res.* **5**, 112–121
38. Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., Resing, K. A., and Ahn, N. G. (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* **4**, 1487–1502
39. Gharib, S. A., Vaisar, T., Aitken, M. L., Park, D. R., Heinecke, J. W., and Fu, X. Y. (2009) Mapping the Lung Proteome in Cystic Fibrosis. *J. Proteome Res.* **8**, 3020–3028
40. Allinen, M., Beroukhi, R., Cai, L., Brennan, C., Lahti-Domenici, J., Huang, H., Porter, D., Hu, M., Chin, L., Richardson, A., Schnitt, S., Sellers, W. R., and Polyak, K. (2004) Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell* **6**, 17–32
41. Zybailov, B., Coleman, M. K., Florens, L., and Washburn, M. P. (2005) Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal. Chem.* **77**, 6218–6224
42. Vasaturo, F., Sallusti, E., Gradiione, A., Malacrino, C., Nardo, T., Avagnano, G., Aglianò, A. M., Granato, T., De Vincenzi, B., Coppotelli, G., Marzullo, A., Soda, G., Simonelli, L., Modesti, M., and Scarpa, S. (2005) Comparison of extracellular matrix and apoptotic markers between benign lesions and carcinomas in human breast. *Int. J. Oncol.* **27**, 1005–1011
43. Hao, X., Sun, B., Hu, L., Lähdesmäki, H., Dunmire, V., Feng, Y., Zhang, S. W., Wang, H., Wu, C., Wang, H., Fuller, G. N., Symmans, W. F., Shmulevich, I., and Zhang, W. (2004) Differential gene and protein expression in primary breast malignancies and their lymph node metastases as revealed by combined cDNA microarray and tissue microarray analysis. *Cancer* **100**, 1110–1122
44. Turashvili, G., Bouchal, J., Baumforth, K., Wei, W., Dziechciarkova, M., Ehrmann, J., Klein, J., Fridman, E., Skarda, J., Srovnal, J., Hajduch, M., Murray, P., and Kolar, Z. (2007) Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis. *BMC Cancer* **7**, 55
45. Williams, C. M., Engler, A. J., Slone, R. D., Galante, L. L., and Schwarzbauer, J. E. (2008) Fibronectin expression modulates mammary epithelial cell proliferation during acinar differentiation. *Cancer Res.* **68**, 3185–3192
46. Abbondanza, C., Rossi, V., Roscigno, A., Gallo, L., Belsito, A., Piluso, G., Medici, N., Nigro, V., Molinari, A. M., Moncharmont, B., and Puca, G. A. (1998) Interaction of vault particles with estrogen receptor in the MCF-7 breast cancer cell. *Journal of Cell Biology* **141**, 1301–1310
47. Fritzsche, F. R., Dahl, E., Pahl, S., Burkhardt, M., Luo, J., Mayordomo, E., Gansukh, T., Dankof, A., Knuechel, R., Denkert, C., Winzer, K. J., Dietel, M., and Kristiansen, G. (2006) Prognostic relevance of AGR2 expression in breast cancer. *Clin. Cancer Res.* **12**, 1728–1734
48. Fletcher, G. C., Patel, S., Tyson, K., Adam, P. J., Schenker, M., Loader, J. A., Daviet, L., Legrain, P., Parekh, R., Harris, A. L., and Terrett, J. A. (2003) hAG-2 and hAG-3, human homologues of genes involved in differentiation, are associated with oestrogen receptor-positive breast tumours and interact with metastasis gene C4.4a and dystroglycan. *Br. J. Cancer* **88**, 579–585
49. Liu, D., Rudland, P. S., Sibson, D. R., Platt-Higgins, A., and Barraclough, R. (2005) Human homologue of cement gland protein, a novel metastasis inducer associated with breast carcinomas. *Cancer Res.* **65**, 3796–3805
50. Thompson, D. A., and Weigel, R. J. (1998) hAG-2, the human homologue of the *Xenopus laevis* cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines. *Biochem. Biophys. Res. Commun.* **251**, 111–116
51. Ribas de Pouplana, L., and Schimmel, P. (2001) Aminoacyl-tRNA synthetases: potential markers of genetic code development. *Trends Biochem. Sci.* **26**, 591–596
52. Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347**, 203–206
53. Sugiura, I., Nureki, O., Ugaji-Yoshikawa, Y., Kuwabara, S., Shimada, A., Tateno, M., Lorber, B., Giege, R., Moras, D., Yokoyama, S., and Konno, M. (2000) The 2.0 angstrom crystal structure of *Thermus thermophilus* methionyl-tRNA synthetase reveals two RNA-binding modules. *Struct. Fold. Des.* **8**, 197–208
54. Park, S. G., Schimmel, P., and Kim, S. (2008) Aminoacyl tRNA synthetases and their connections to disease. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11043–11049
55. Desvergne, B., and Wahli, W. (1999) Peroxisome proliferator-activated receptors: Nuclear control of metabolism. *Endocr. Rev.* **20**, 649–688
56. Hsu, M. H., Savas, U., Griffin, K. J., and Johnson, E. F. (2001) Identification of peroxisome proliferator-responsive human genes by elevated expression of the peroxisome proliferator-activated receptor  $\alpha$  in HepG2 cells. *J. Biol. Chem.* **276**, 27950–27958
57. Berger, J., and Moller, D. E. (2002) The mechanisms of action of PPARs. *Annu. Rev. Med.* **53**, 409–435
58. Peters, J. M., Aoyama, T., Cattle, R. C., Nobumitsu, U., Hashimoto, T., and Gonzalez, F. J. (1998) Role of peroxisome proliferator-activated receptor  $\alpha$  in altered cell cycle regulation in mouse liver. *Carcinogenesis* **19**, 1989–1994
59. Bergstraesser, L. M., Srinivasan, G., Jones, J. C., Stahl, S., and Weitzman, S. A. (1995) Expression of hemidesmosomes and component proteins is lost by invasive breast-cancer cells. *Am. J. Pathol.* **147**, 1823–1839
60. D'Ardenne A. J., Richman P. I., Horton M. A., Mcaulay A. E., and Jordan, S. (1991) Co-ordinate expression of the  $\alpha$ -6 integrin laminin receptor sub-unit and laminin in breast cancer. *J. Pathol.* **165**, 213–220
61. Henning, K., Berndt, A., Katenkamp, D., and Kosmehl, H. (1999) Loss of laminin-5 in the epithelium-stroma interface: an immunohistochemical marker of malignancy in epithelial lesions of the breast. *Histopathology* **34**, 305–309
62. Natali, P. G., Nicotra, M. R., Botti, C., Mottolese, M., Bigotti, A., and Segatto, O. (1992) Changes in expression of  $\alpha$ 6/ $\beta$ 4 integrin heterodimer in primary and metastatic breast cancer. *Br. J. Cancer* **66**, 318–322
63. Heldring, N., Pike, A., Andersson, S., Matthews, J., Cheng, G., Hartman, J., Tujague, M., Ström, A., Treuter, E., Warner, M., and Gustafsson, J. A. (2007) Estrogen receptors: How do they signal and what are their targets. *Physiol. Rev.* **87**, 905–931
64. Hall, J. M., Couse, J. F., and Korach, K. S. (2001) The multifaceted mechanisms of estradiol and estrogen receptor signaling. *J. Biol. Chem.* **276**, 36869–36872
65. Jakacka, M., Ito, M., Weiss, J., Chien, P. Y., Gehm, B. D., and Jameson, J. L. (2001) Estrogen receptor binding to DNA is not required for its activity through the nonclassical AP1 pathway. *J. Biol. Chem.* **276**, 13615–13621
66. Ray, P., Ghosh, S. K., Zhang, D. H., and Ray, A. (1997) Repression of interleukin-6 gene expression by 17  $\beta$ -estradiol: Inhibition of the DNA-binding activity of the transcription factors NF- $\kappa$ B and NF-kappa B by the estrogen receptor. *FEBS Lett.* **409**, 79–85
67. Glidewell-Kenney, C., Weiss, J., Lee, E. J., Pillai, S., Ishikawa, T., Ariazi, E. A., and Jameson, J. L. (2005) ERE-independent ER  $\alpha$  target genes differentially expressed in human breast tumors. *Mol. Cell. Endocrinol.* **245**, 53–59
68. Johansen, F. E., and Prywes, R. (1995) Serum response factor - transcriptional regulation of genes induced by growth-factors and differentiation. *Biochim. Biophys. Acta-Rev. Cancer* **1242**, 1–10
69. Medjkane, S., Perez-Sanchez, C., Gaggioli, C., Sahai, E., and Treisman, R. (2009) Myocardin-related transcription factors and SRF are required for cytoskeletal dynamics and experimental metastasis. *Nat. Cell Biol.* **11**, 257–268
70. Lopez, D., Sanchez, M. D., Shea-Eaton, W., and McLean, M. P. (2002) Estrogen activates the high-density lipoprotein receptor gene via binding to estrogen response elements and interaction with sterol regulatory element binding protein-1A. *Endocrinology* **143**, 2155–2168
71. Børresen-Dale, A. L. (2003) TP53 and breast cancer. *Hum. Mutat.* **21**, 292–300
72. Kostic, C., and Shaw, P. H. (2000) Isolation and characterization of sixteen novel p53 response genes. *Oncogene* **19**, 3978–3987
73. Liu, T., Laurell, C., Selivanova, G., Lundeberg, J., Nilsson, P., and Wiman, K. G. (2007) Hypoxia induces p53-dependent transactivation and Fas/CD95-dependent apoptosis. *Cell Death Differ.* **14**, 411–421
74. Quaedackers, M. E., Van Den Brink, C. E., Wissink, S., Schreurs, R. H., Gustafsson, J. A., Van Der Saag, P. T., and Van Der Burg, B. B. (2001) 4-hydroxytamoxifen trans-represses nuclear factor-kappa B activity in human osteoblastic U2-OS cells through estrogen receptor (ER) $\alpha$ ,

- and not through ER beta. *Endocrinology* **142**, 1156–1166
75. Harnish, D. C., Scicchitano, M. S., Adelman, S. J., Lyttle, C. R., and Karathanasis, S. K. (2000) The role of CBP in estrogen receptor cross-talk with nuclear factor-kappa B in HepG2 cells. *Endocrinology* **141**, 3403–3411
76. Tan, A. R., Alexe, G., and Reiss, M. (2009) Transforming growth factor-beta signaling: emerging stem cell target in metastatic breast cancer? *Breast Cancer Res. Treat.* **115**, 453–495
77. Camoretti-Mercado, B., Fernandes, D. J., Dewundara, S., Churchill, J., Ma, L., Kogut, P. C., McConville, J. F., Parmacek, M. S., and Solway, J. (2006) Inhibition of transforming growth factor beta-enhanced serum response factor-dependent transcription by SMAD7. *J. Biol. Chem.* **281**, 20383–20392
78. Jin, D.Y., Chae, H. Z., Rhee, S. G., and Jeang, K.T. (1997) Regulatory Role for a Novel Human Thioredoxin Peroxidase in NF- $\kappa$ B Activation. *J. Biol. Chem.* **272**, 30952–30961
79. Matsuda, A., Suzuki, Y., Honda, G., Muramatsu, S., Matsuzaki, O., Nagano, Y., Doi, T., Shimotohno, K., Harada, T., Nishida, E., Hayashi, H., and Sugano, S. (2003) Large-scale identification and characterization of human genes that activate NF-kappa B and MAPK signaling pathways. *Oncogene* **22**, 3307–3318