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 (SpI) 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 lead to the discovery of the “intrinsc 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)−1, 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

---

1 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; SpI, spectral index; TF, transcription factor; FDR, false discovery rate; SRF, serum response factor; AP-1, activator protein 1; SREBP1, sterol regulatory binding protein 1.
that luminal A tumors may be adequately treated with anti-
hormonal therapy alone, whereas luminal B tumors may ben-
efit from chemotherapy added to antihormonal therapy (7).
Despite advances in the gene expression-based stratifica-
tion of human breast cancer, the molecular basis of luminal breast
tumorigenesis and luminal breast cancer clinical heterogene-
ity is still poorly understood. This gap in knowledge is due, in
part, to the well-known limitations associated with gene ex-
pression, 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 can-
cer 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 char-
acterization 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 het-
erogeneous samples by laser capture microdissection (LCM)
(13, 14). Such sample analysis can lead to detailed proteomic
portraits of tumor microenvironments and, importantly, cor-
rect for potential confounding effects of stromal contamina-
tion (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-ac-
quired breast tumor proper and stromal cells from a lymph
node containing breast carcinoma (21), and differential pro-
teome profiles according to tamoxifen therapy responsive-
ness were analyzed from ER+, primary tumor pooled LCM
samples (22).

The combination of LCM enrichment and shotgun pro-
teomics has the potential to reveal novel pathogenic path-
ways and networks by which normal breast epithelial cells
progress to invasive malignant epithelial cells. Genomic and
gene expression profiling of human breast cancer progress-
ion has already shown potential for the identification of
molecular alterations that serve as prognostic and predic-
tive biomarkers (2–4, 23). Comprehensive knowledge of
proteomic patterns in normal epithelium, as compared with
malignant breast epithelium, would be invaluable in provid-
ing a more complete picture of the molecular entities asso-
ciated with breast cancer progression.

In the present study, we performed comparative proteomic
profiling of laser capture microdissected normal breast epi-
thelium (NBE) from nine different noncancerous human mam-
maplasty 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 plat-
form 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 in-
dividual samples. For each protein, the spectral index (SpI)
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 immunohistostain-
ing 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 car-
diomyopathy (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 concor-
dant expression differences. PSEA of LCM-derived MBE
samples revealed significant decreased expression of cy-
toskeletal proteins and of proteins that are consistently neg-
atively correlated with ER+ human breast cancers. More im-
portantly, many protein sets, representing proteins controlled
by common transcription factors (TFs) were found, potentially
suggesting a role of estrogen response element (ERE)-inde-
pendent ER signaling in breast tumorigenesis. In addition, we
performed a GOMiner analysis, which requires arbitrary
thresholds for differentially abundant proteins, on the pro-
teomic results. For differentially expressed proteins at the
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 140 mM Tris base buffer were obtained from Invitrogen (Carlsbad, CA), and sequencing grade modified trypsin was from Promega (Madison, WI).

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 trypsic 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., 200 Å pore size particles. 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 2 × 10^5 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%.

**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.

### Table I

**Patient and tumor characteristics of samples used in this study**

| 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  | Ductal        | E          | 1.1       | 2      | Neg      | Pos| Pos| Pos   |
| NBE2          | 279–1   | 18  | RM            | MBE2      | 153–2   | 46  | Lobular       | M          | 1.6       | 1      | Pos      | Pos| Pos| Pos   |
| NBE3          | 224–2   | 21  | RM            | MBE3      | 128–3   | 79  | Ductal        | M          | 1.6       | 1      | Pos      | Pos| Pos| Pos   |
| NBE4          | 222–6   | 24  | RM            | MBE4      | 171–1   | 39  | Ductal        | M          | 2.8       | 2      | Pos      | Neg| Pos| Neg   |
| NBE5          | 225–6   | 30  | RM            | MBE5      | 130–2   | 54  | Ductal        | M          | 6.5       | 2      | Pos      | Pos| Pos| Neg   |
| NBE6          | 295–4   | 25  | RM            | MBE6      | 98–2    | 31  | Ductal        | E          | 1.7       | 3      | Pos      | Pos| Pos| Neg   |
| NBE7          | 320–5   | 38  | RM            | MBE7      | 122–3   | 45  | Ductal        | M          | 1.7       | 2      | Neg      | Pos| Neg| Neg   |
| NBE8          | 331–8   | 22  | RM            | MBE8      | 193–2   | 45  | Ductal        | M          | 1.5       | 1      | Pos      | Pos| Pos| Neg   |
| NBE9          | 335–6   | 34  | RM            | MBE9      | 169–1   | 34  | Ductal        | M          | 2.6       | 3      | Pos      | Pos| Pos| Neg   |
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 Xcorr ≥ 1.9 for 1+ ions, ≥ 2.2 for 2+ ions, and ≥ 3.8 for 3+ ions, PeptideProphet (Institute for Systems Biology, Seattle, WA) probability ≥ 0.9, and after correction for missassigned monoisotopic peaks with the mass accuracy of the precursor ions ±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 ≥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 Spls 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 FDR 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 micro-dissecting 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.
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.
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 SpI values were calculated for the 1623 identified proteins, and the SpI distribution is shown in Supplemental Fig. S4A. Proteins with statistically significant SpI values were determined by permutation analysis (25). The frequency distribution of SpI values for 2000 randomly permuted sample phenotype classes (see Supplemental Fig. S4B) and the ranges encompassing 99%, 95%, and 90% of all SpIs were determined to be SpI < 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 SpI-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 SpI 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 SpI-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 (SpI = 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 (SpI = 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, SpI = 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 (SpI = 0.673) and \( \Delta(3,5)\)-\( \Delta(2,4)\)-dienoyl-CoA isomerase, mitochondrial (SpI = 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 la aminoacyl-tRNA synthetases that recognize hydrophobic amino acids including lleu, 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 \( \Delta(3,5)\)-\( \Delta(2,4)\)-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, \( \Delta(3,5)\)-\( \Delta(2,4)\)-dienoyl-CoA isomerase, mitochondrial, has been shown to be regulated by peroxisome proliferator-activated receptor \( \alpha \), a receptor that 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 SpIs. As a result, PSEA reduces the chance that a particular protein set is missed because some proteins below the significant threshold may not be
included in the differentially abundant list. PSEA also preferentially weighs functional protein sets with strong (i.e. high $SpI$) 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 $SpI$ distribution, yielding an ES value of $H_{11002}^0.51$ (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

| 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 | $SpI$ | Antibody | IHC staining intensity | Verification |
|------------------|---------------------|-------|----------|------------------------|--------------|
|                  |                     |       |          | Cellular compartment   | Strong | Moderate | Weak | Negative |            |             |
| P55268           | Laminin subunit β-2 | -0.778| CAB000053| Normal myoepithelium 3 | 0      | 0        | 0    | Yes       |             |             |
| P53621           | Coatomer subunit α  | 0.576 | HPA028024| Normal epithelium 0    | 0      | 0        | 11   | Yes       |             |             |
| Q14764           | Major vault protein | 0.865 | HPA002321| Normal epithelium 7    | 4      | 0        | 0    | Yes       |             |             |
| Q3396            | ATP-citrate synthase | 0.635 | HPA002243| Normal epithelium 0    | 0      | 0        | 3    | Yes       |             |             |
| P06753           | Tropomyosin α-3 chain| 0.673 | HPA009066| Normal epithelium 1    | 8      | 0        | 2    | Yes       |             |             |
| P11971           | Thymidine phosphorylase| 0.558| HPA001072| Normal epithelium 2    | 7      | 0        | 3    | Yes       |             |             |
| P47567           | Protein ERGIC-53    | 0.564 | HPA002320| Normal epithelium 2    | 7      | 3        | 0    | Yes       |             |             |
| P35222           | Catenin β-1         | -0.631| CAB000108| Normal epithelium 3    | 0      | 0        | 0    | No        |             |             |
| Q81999           | Transgelin          | -0.567| CAB001447| Normal myoepithelium 2| 0      | 0        | 0    | Yes       |             |             |
| Q98X11           | Asporin             | 0.667 | HPA024230| Normal epithelium 0    | 2      | 0        | 7    | No        |             |             |
| Q19822           | Phosphoenolpyruvate | 0.667 | CAB018734| Normal epithelium 0    | 0      | 0        | 3    | Yes       |             |             |
| P42224           | Signal transducer   | 0.574 | HPA000931| Normal epithelium 0    | 0      | 2        | 0    | Yes       |             |             |
| Q9762            | EMLIN-1             | -0.667| HPA002822| Normal myoepithelium 0| 3      | 0        | 0    | Yes       |             |             |
| P59111           | Calponin-1          | -1.000| HPA000007| Normal myoepithelium 2| 0      | 0        | 12   | Yes       |             |             |
| P19701           | Keratin, type I cytoskeletal 15| -0.817| HPA024554| Normal epithelium 2    | 0      | 0        | 8    | Yes       |             |             |
| Q12907           | Vesicular integral-membrane protein VIP36| 0.590| HPA003927| Normal epithelium 2    | 0      | 0        | 3    | Yes       |             |             |
| Q8UN36           | Protein NDRG2       | -0.889| HPA002896| Normal epithelium 2    | 0      | 0        | 0    | Yes       |             |             |
| P49419           | α-a-minoacidic semialdehyde dehydrogenase| -0.760| HPA023296| Normal epithelium 0    | 0      | 0        | 2    | No        |             |             |
| P90211           | Glutathione S-transferase P| -0.922| HPA019779| Normal epithelium 0    | 0      | 1        | 0    | Yes       |             |             |
| Q2871            | Calyculin-binding protein| 0.667| HPA025753| Normal epithelium 2    | 0      | 0        | 4    | Yes       |             |             |
| Q9ULZ3           | Apoptosis-associated speck-like protein containing a CARD | 0.691| HPA006853| Normal epithelium 0    | 2      | 0        | 0    | Yes       |             |             |
| P68133           | Actin, α skeletal muscle| 0.698| HPA000045| Normal myoepithelium 0| 0      | 0        | 3    | Yes       |             |             |
| P27338           | Amine oxidase [flavin-containing] B | -0.688| HPA002328| Normal epithelium 2    | 0      | 0        | 0    | No        |             |             |
| P04063           | Annexin A1          | -0.743| HPA011271| Normal epithelium 9    | 3      | 0        | 0    | No        |             |             |
| P12429           | Annexin A3          | -0.904| HPA013398| Normal epithelium 0    | 3      | 0        | 0    | No        |             |             |

Molecular & Cellular Proteomics 9.11
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 β, fibrinogen, serpin peptidase inhibitor, lumican, 14-3-3 protein σ, 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 κ-light-chain-enhancer of activated B cells (NF-κB), 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 transcription 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-κB 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-κB, 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-κB (66). Indeed, the majority of leading-edge proteins for both AP-1 (25 out of 34) and NF-κB (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).
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 α chain, brain, transgelin, caldesmon, he-
mopexin (HEMO), and laminin subunit α-3 and β-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 extra-cellular 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-κB and SMAD signaling to breast tumorigenesis. GOMiner revealed significantly altered expression of “IκB kinase/NF-κB 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-κB. For example, peroxiredoxin-4 is known to activate NF-κB via modulation of IκB-α 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-κB (79). Interestingly, NF-κB 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 were used to ensure the statistical significance of the results. 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.

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).

** These two senior authors contributed equally.
38. Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., 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., Beroukhim, R., Cai, L., Brennan, C., Lahti-Domenici, J., Huang, H., Poiter, D., Hu, M., Ciba, G., BMC Cancer 7, 384

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., Gradilone, A., Malacrino, C., Nardo, T., Avagnano, G. B., Granato, T., Dili G., Marzulli, A., Soda, G., Simonelli, L., Modioli, 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., Lädhemski, H., Dunrine, V., Fang, Y., Zhang, S. W., Wang, H., Wu, C., Wang, H., Fuller, G. N., Symons, W. F., Shumulevich, 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., Bouchar, 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 mass spectrometry. BMC Cancer 7, 153

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., Belisio, A., Piluso, G., Medici, N., Nigro, V., Molinari, A. M., Moncharmont, B., and Puca, G. A. (1995) Expression of hemidesmosomes and component proteins is lost by invasive breast-cancer cells. Am. J. Pathol. 147, 1823–1839

47. 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

48. Fletcher, G. C., Patel, S., Tyson, K., Adam, P. J., Schenker, M., Loader, J. M., Eaves, J. C., Eaves, C. J., Parmigiani, P., Parchik, R. J. A. (2003) hAG-2 and hAG-3, human homologues of genes involved in cytoskeletal dynamics and experimental metastasis. BMC Cancer Ther. 3, 26

49. Liu, D., Rudland, P. S., Rabinow, D. B., Platt-Higgins, A., and Barracough, R. (2005) Human homologue of cement gland protein, a novel metastasis inducer associated with breast carcinomas. Cancer Res. 65, 3796–3805

50. Thomas, A. D., 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. Erian, G., Delaunay, M., Poch, O., Gangloff, J., and Moras, D. (1990) Partition Res. X. Y. (2009) Mapping the Lung Proteome in Cystic Fibrosis. J. Proteome Res. 8, 3020–3028

53. Sugiura, I., Nureki, O., Ujgai-Yoshikawa, Y., K compact structures 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. Hau, 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 a 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., Cattley, 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, C. J., 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. Ardenne, A. J., Richman P. I., Horton M. A., McAlayu A. E., and Jordan, S. (1991) Co-ordinate expression of the alpha-6 integrin laminin receptor Estramustine and Docetaxel. Br. J. Cancer 66, 318–322

61. Helding, 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

62. 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

63. 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

64. 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-IL6 and NF-kappaB by the estrogen receptor. FEBS Lett. 409, 79–85

65. 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

66. Lieber, A. S., Jozefowicz, A. F. E., and Przewys, R. (1995) Serum response factor - transcriptional regulation of genes induced by growth-factors and differentiation. Biochem. Biophys. Acta-Rev. Cancer 1242, 1–10

67. 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

68. 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 steroid regulatory element binding protein-1A. Endocrinology 143, 2155–2168

69. Barresen-Dale, A. L. (2003) TP53 and breast cancer. Hum. Mutat. 21, 292–300

70. Kostic, C., and Shaw, P. H. (2000) Isolation and characterization of sixteen novel p53 response genes. Oncogene 19, 3978–3987

71. Liu, T., Laurell, C., Selivanova, G., Lundberg, J. N., Wiman, K. G. (2007) Hypoxia induces p53-dependent transactivation and Fas/CD9-dependent apoptosis. Cell Death Differ. 14, 411–421

72. Quaedackers, M. E., Van Den Brink, C. E., Wissink, S., Schreurs, R. H., van Engeland, H., and Dijkstra, P. A. (2003) In Situ Proteomics of ER+ Breast Cancer Cells and PSEA/GO.
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-κB 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