

## Gene expression abnormalities in histologically normal breast epithelium of breast cancer patients

Anusri Tripathi<sup>1</sup>, Chialin King<sup>1</sup>, Antonio de la Morenas<sup>2</sup>, Victoria Kristina Perry<sup>2</sup>, Bohdana Burke<sup>2</sup>, Gregory A. Antoine<sup>3</sup>, Erwin F. Hirsch<sup>3</sup>, Maureen Kavanah<sup>3</sup>, Jane Mendez<sup>3</sup>, Michael Stone<sup>3</sup>, Norman P. Gerry<sup>4</sup>, Marc E. Lenburg<sup>4</sup> and Carol L. Rosenberg<sup>1,2\*</sup>

<sup>1</sup>Department of Medicine, Boston University School of Medicine and Boston Medical Center, Boston, MA

<sup>2</sup>Department of Pathology, Boston University School of Medicine and Boston Medical Center, Boston, MA

<sup>3</sup>Department of Surgery, Boston University School of Medicine and Boston Medical Center, Boston, MA

<sup>4</sup>Department of Genetics and Genomics, Boston University School of Medicine and Boston Medical Center, Boston, MA

Normal-appearing epithelium of cancer patients can harbor occult genetic abnormalities. Data comprehensively comparing gene expression between histologically normal breast epithelium of breast cancer patients and cancer-free controls are limited. The present study compares global gene expression between these groups. We performed microarrays using RNA from microdissected histologically normal terminal ductal-lobular units (TDLU) from 2 groups: (i) cancer normal (CN) (TDLUs adjacent to untreated ER+ breast cancers ( $n = 14$ )) and (ii) reduction mammoplasty (RM) (TDLUs of age-matched women without breast disease ( $n = 15$ )). Cyber-T identified differentially expressed genes. Quantitative RT-PCR (qRT-PCR), immunohistochemistry (IHC), and comparison to independent microarray data including 6 carcinomas *in situ* (CIS), validated the results. Gene ontology (GO), UniProt and published literature evaluated gene function. About 127 probesets, corresponding to 105 genes, were differentially expressed between CN and RM ( $p < 0.0009$ , corresponding to FDR  $< 0.10$ ). 104/127 (82%) probesets were also differentially expressed between CIS and RM, nearly always (102/104 (98%)) in the same direction as in CN vs. RM. Two-thirds of the 105 genes were implicated previously in carcinogenesis. Overrepresented functional groups included transcription, G-protein coupled and chemokine receptor activity, the MAPK cascade and immediate early genes. Most genes in these categories were under-expressed in CN vs. RM. We conclude that global gene expression abnormalities exist in normal epithelium of breast cancer patients and are also present in early cancers. Thus, cancer-related pathways may be perturbed in normal epithelium. These abnormalities could be markers of disease risk, occult disease, or the tissue's response to an existing tumor.

© 2007 Wiley-Liss, Inc.

**Key words:** global gene expression; normal breast epithelium; early genetic changes

Breast cancer arises as genetic aberrations accumulate in precursor epithelial cells. Considerable information is available about the molecular alterations characterizing breast cancers, but knowledge of alterations in earlier lesions is limited. Recently, abnormalities have been appreciated in histologically normal breast epithelium. These abnormalities include allelic imbalance or loss of heterozygosity,<sup>1–8</sup> aberrant methylation of p16<sup>INK4</sup><sup>9</sup> and of RASSF1A,<sup>10</sup> cytogenetic changes,<sup>11,12</sup> telomere shortening,<sup>13</sup> loss of IGF2 imprinting,<sup>14</sup> aberrant response to estrogen,<sup>15</sup> loss of RAR $\beta$  expression,<sup>16</sup> aberrant phosphorylation of p38,<sup>17</sup> upregulation of EZH2.<sup>18</sup> Some of these abnormalities have been detected in normal-appearing tissue adjacent to the tumor, and others have been found at a distance from it. Some abnormalities are concordant, and others are discordant, with abnormalities in the tumors themselves.

Despite the evidence supporting the existence of occult abnormalities in normal-appearing breast epithelium of breast cancer patients, the roles these abnormalities play in carcinogenesis is poorly understood. One approach to better understand their significance is to compare histologically normal breast epithelium of breast cancer patients to normal breast epithelium of women without breast cancer. The few studies comparing these groups have examined allelic imbalance, aneuploidy and methylation or

expression of specific proteins, and have found abnormalities more frequently in patients with cancer than in controls.<sup>7,8,10,12,17–19</sup> We hypothesized that by taking a comprehensive gene expression approach, we might detect consistent abnormalities in the normal-appearing epithelium of breast cancer patients, compared to controls. These abnormalities might suggest mechanisms predisposing to cancer, or activated early in carcinogenesis. If this hypothesis were true, then elucidating these abnormalities could enhance understanding of important functional alterations present early in carcinogenesis, suggest targets for cancer prevention and improve cancer risk assessment. To begin testing this hypothesis, we undertook the present study to identify gene expression differences in the histologically normal breast epithelium of breast cancer patients, compared to reduction mammoplasty controls.

### Material and methods

#### Tissue samples

After obtaining institutional review board approval, deidentified tissues not needed for pathological diagnosis were collected from breast cancer surgeries and reduction mammoplasties performed at Boston Medical Center. To preserve RNA quality, tissues were obtained within 1 hr of surgery and immediately snap-frozen in liquid nitrogen, embedded in OCT medium and stored at  $-80^{\circ}\text{C}$ . Tissues from 2 groups were examined: (i) CN and (ii) RM.

#### Laser-capture microdissection, RNA isolation and amplification

These procedures were carried out to obtain RNA from homogeneous populations of breast epithelial cells of normal-appearing terminal ductal-lobular units (TDLU) of breast cancer patients and disease-free individuals, as described previously.<sup>20</sup> Figure 1 is a representative photograph of a microdissection. Supplemental File 4 presents photographs of TDLUs from multiple cases. To obtain enough RNA, 2–3 TDLUs were microdissected per case. TDLUs from the CN group were “tumor-adjacent”, *i.e.*, located 1–2 cm from the tumor, on blocks lacking malignant cells. Total RNA was extracted from the captured cells (Picopure RNA isolation kit, Arcturus Engineering) and 100 ng was used for T7-based RNA amplification (MessageAmp aRNA kit, Ambion, Austin, TX). To obtain enough amplified RNA (aRNA), a second round of RNA amplification was performed as described previously,<sup>20</sup> except

This article contains supplementary material available via the Internet at <http://www.interscience.wiley.com/jpages/0020-7136/suppmat>.

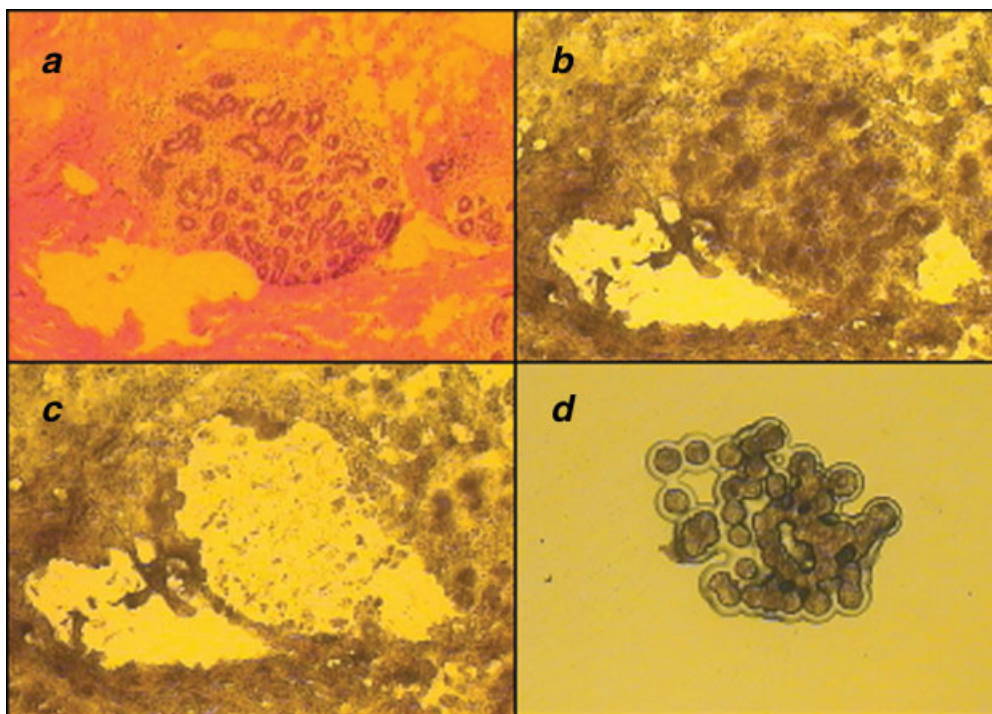
Grant sponsor: Department of Defense Breast Cancer Research Program; Grant number: DAMD17-01-1-0159; Grant sponsor: NIH; Grant numbers: RO1 CA081078, S10 RR021211.

\*Correspondence to: Departments of Medicine and Pathology, Boston University Medical Center, EBRC-413, 650 Albany Street, Boston, MA 02118. E-mail: crosenbe@bu.edu

Received 5 June 2007; Accepted after revision 25 September 2007

DOI 10.1002/ijc.23267

Published online 30 November 2007 in Wiley InterScience (www.interscience.wiley.com).



**FIGURE 1** – Laser capture microdissection of histologically normal, snap-frozen epithelium. Terminal ductal lobular units (TDLU) microdissected from consecutive 10- $\mu$ m thick tissue sections stained with dilute (33%) hematoxylin (H) and dilute (10%) eosin (E) using the PixCell II (Arcturus Engineering, Mountain View, CA). (a) TDLU stained with standard H&E, (b) TDLU stained with dilute H&E, (c) post-capture stromal compartment, (d) captured epithelial compartment.  $\times 40$  magnification.

using 500 ng of aRNA from the first round amplification as starting material.

#### Hybridization, microarray data quantification and normalization

These procedures were carried out as described before.<sup>20</sup> For each hybridization, 10  $\mu$ g of fragmented, biotin-labeled aRNA were hybridized to U133A GeneChip arrays (Affymetrix, Santa Clara, CA), then washed, stained and scanned according to standard protocols (Affymetrix). The scanned arrays were quantified and scaled using the GCOS software package (Affymetrix). Each probeset's expression level was determined from the hybridization intensities of the 22 constituent probes using the Affymetrix Microarray Suite 5.0 (MAS5) algorithm. After removing probesets that lacked sequence-specific hybridization intensity in any sample, the final dataset included hybridization intensities for 14,681 probesets. These were log-transformed and used in subsequent analyses. The transcript interrogated by each probeset was determined using the NetAffx database.<sup>21</sup>

#### Identification of differentially expressed genes

Identification of genes differentially expressed between CN and RM samples was performed with Cyber-T,<sup>22</sup> which combines Student's *t* test with a Bayesian estimate of the intragroup variance obtained from the observed variance of probesets at a similar expression level. We have used this approach previously<sup>23</sup> as it provides increased sensitivity for the identification of differentially expressed probesets with differential hybridization intensity without inflating the false-positive error rate. For the Cyber-T analysis, we set the Sliding Window Size parameter at 101 and the Bayes Confidence Estimate parameter at 10. To identify probesets with significant differential hybridization intensity between groups, we ranked probesets by their Cyber-T *p*-value and calculated a False Discovery Rate statistic.<sup>24</sup> Hybridization intensities for the differentially expressed probesets were each *z*-score normalized (mean = 0, standard deviation = 1) and organized by hierarchical clustering using a Euclidean Distance Measure in DecisionSite for Functional Genomics (Spotfire, Somerville, MA).

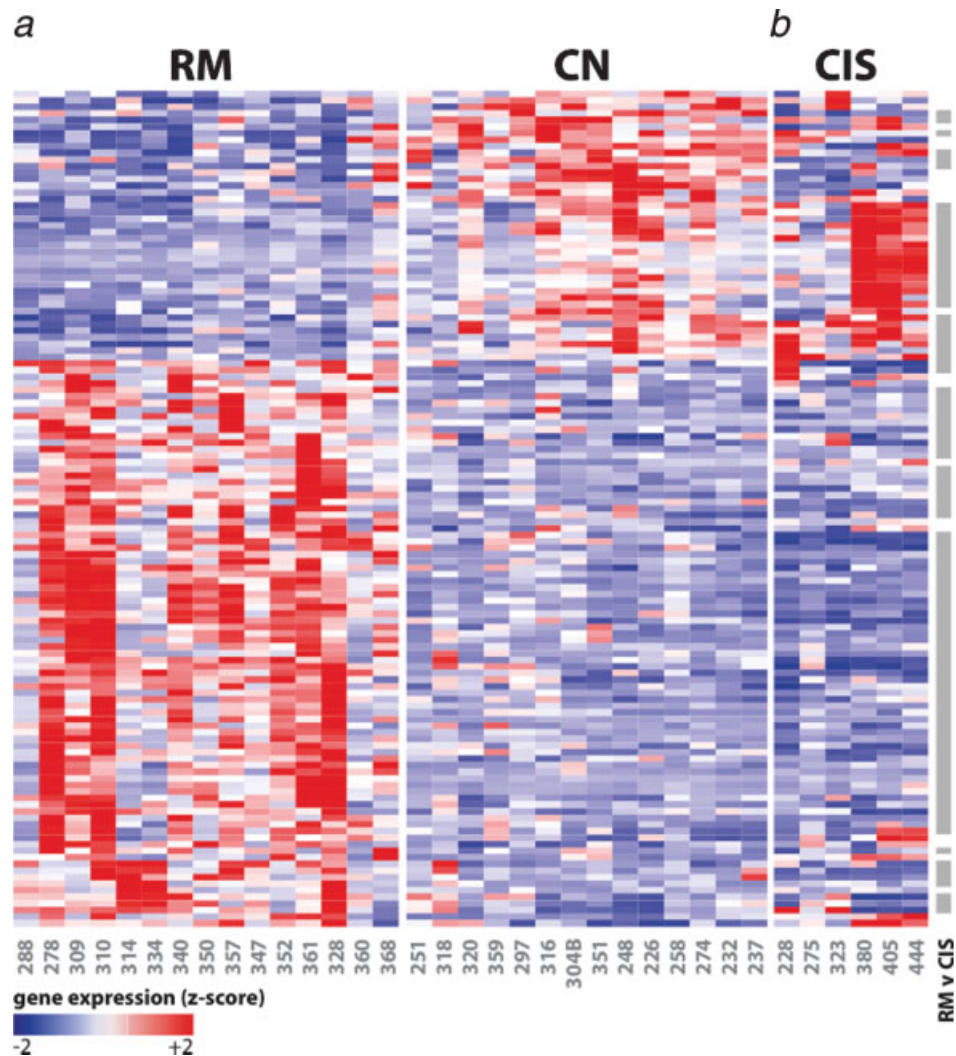
#### Validation of microarray data

We used qRT-PCR to confirm gene expression levels of 4 of the 105 genes (CXCL2, FOS, FOSB, KLF6). Each gene was examined in 6 samples (3 CN and 3 RM). A total of 16 independent samples (8 RM and 8 CN) had sufficient remaining unamplified RNA for additional studies, and each was examined with 1–3 test genes, plus the control. For each qRT-PCR validation, 5 ng of total unamplified RNA were reverse transcribed (Multiscript RT and TaqMan RT reagent kit, Applied Biosystems, Fostercity, CA). The RT reaction was performed using random hexamer, in a total volume of 25  $\mu$ l and carried out at 25°C  $\times$  10 min, 37°C  $\times$  60 min, 95°C  $\times$  5 min. The PCR reaction was performed in a 25  $\mu$ l volume, which included 11.25  $\mu$ l cDNA solution, 12.5  $\mu$ l Universal Mastermix (ABI) and 1.25  $\mu$ l TaqMan gene expression assay (ABI) of the gene to be validated. PCR was performed at 95°C  $\times$  10 min followed by 40 cycles of 95°C  $\times$  15 sec and 60°C  $\times$  1 min. Amplifications for each of the 4 test and 1 control gene (GUS $\beta$ )<sup>25</sup> were conducted in duplicate and monitored (TaqMan Gene assays and Prism 7000 Sequence Detector System, ABI). Standard curves for each gene were generated from samples of known concentration. From these curves, the absolute quantity of each gene was determined for each sample, and the relative quantity of each test gene was calculated after normalization with GUS $\beta$ .

For immunohistochemical (IHC) corroboration of the microarray data, 5- $\mu$ m serial sections were cut from paraffin blocks of both RM and CN samples, and mounted on Shandon Colomark Plus slides, deparaffinized in xylene and rehydrated in graded alcohol to water. The slides were steamed in Vector Retrieval Buffer for 60 min, blocked for 20 min with 10% normal goat serum in PBS and 30 min with Vector Avidin/Biotin blocking kit. The primary antibody (FosB rabbit monoclonal antibody, cell signaling, 17  $\mu$ g/ml) at 1:50 dilution was incubated overnight at 4–6°C. Rabbit IgG (Vector, 5 mg/mL) was used as the negative control at the same concentration as the primary antibody. Using Vector Vectastain Elite ABC Kit, slides were then reacted with a biotinylated secondary antibody (goat anti-rabbit) and incubated with preformed avidin–biotin–peroxidase complex (ABC reagent). Diaminobenzidine (DAB) was used as a substrate. Sections were



**FIGURE 2** – Genes that are differentially expressed in histologically normal breast epithelium of patients with breast cancer (CN) compared to without breast cancer (RM). (a) The 127 probesets differentially expressed in histologically normal breast epithelium of CN compared to RM patients were z-score normalized and organized from top to bottom by hierarchical clustering. Overexpression is colored red, under-expression is blue. Case numbers are below each column. (b) The expression of the probesets in Panel A was examined in 6 independent CIS samples from another dataset. The analysis of differential expression was repeated comparing the RM and CIS samples. Probesets exhibiting differential expression (FDR < 0.10) are labeled with a dark grey box to the right of the heat map. Supplemental File 2 includes additional information for each of the 127 probesets.



counterstained with hematoxylin, dehydrated, and mounted. Computer assisted morphometric analysis was performed (iVision Automated Digital Image Analysis System with proprietary software, BioGenex San Ramon, CA). The pathologist was blinded to which group each sample belonged. The analysis was performed on 5 regions of interest for each case, selected to include glandular tissue and avoid stroma. The results were averaged and expressed as percent positive staining.

#### Molecular and functional analysis of the microarray data

For classification of genes into biological categories, we used the EASE program (<http://david.niaid.nih.gov/david/ease.htm>), which calculates overrepresentation of GO<sup>26</sup> categories among genes on the gene list, compared to all genes on the chip used to generate the list, using Fisher's Exact Test. These analyses were supplemented by queries to UniProt ([www.expasy.uniprot.org](http://www.expasy.uniprot.org)) and the literature. Placement of genes onto biological pathways was performed using the KEGG pathways (<http://www.genome.jp/kegg/pathway.html>) and 2 commercially available tools (iPATH, which maps genes to 225 well-established signaling and metabolic pathways based on the literature (<http://escience.invitrogen.com/ipath/index.jsp>); and Ingenuity Pathway Analysis (<http://www.ingenuity.com>) which connects a gene list to hypothetical networks of interacting genes derived from the literature).

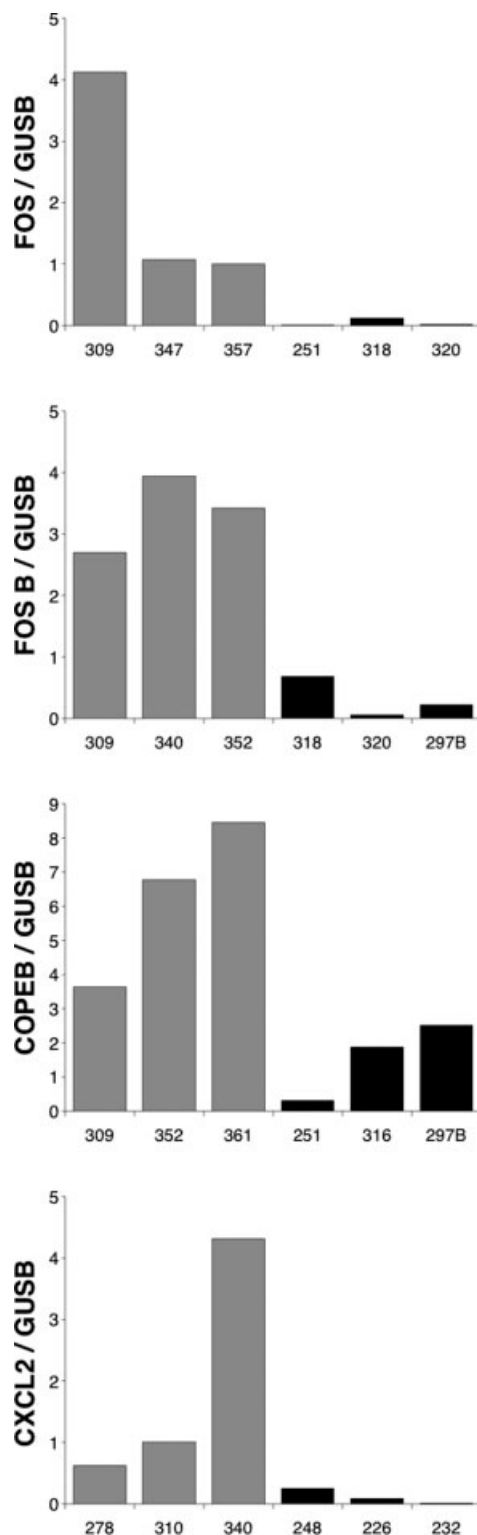
## Results

### Samples and patients

Microarray analyses were performed on 29 samples from 29 patients belonging to 2 groups: (i) the CN group, consisting of 14 samples of histologically normal TDLUs microdissected from 14 patients with ER+ ductal breast cancers undergoing surgery (median age = 49 years, range: 34–65); (ii) the RM or control group, consisting of 15 samples of histologically normal TDLUs microdissected from 15 patients at usual risk of breast cancer, undergoing breast reduction surgeries (median age = 47 years, range: 41–60). No patient had received chemo- or radiation therapy. Although no genotyping of the breast cancer cases was done, the subjects' available histories and the tumors' immunophenotypes<sup>27</sup> suggest that only a small proportion were likely to represent BRCA-associated tumors (see Supplemental File 1). The microarray data from these samples, including the raw probe-level hybridization intensities, are freely available from the NCBI Gene Expression Omnibus under accession GSE9574.

### Genes differentially expressed between the RM and CN groups

We analyzed the probeset hybridization intensities for differences between the RM and CN samples, using the Cyber-T-test and identified 127 probesets with a  $p$ -value < 0.0009, corresponding to a false-discovery rate < 0.10. Figure 2a shows the relative in-



**FIGURE 3** – Validation of microarray data by qRT-PCR. Each panel shows results for 1 test gene (top to bottom: FOS, FOSB, KLF6 (= COPEB), CXCL2). The horizontal axis indicates the particular sample tested, indicated by a case number. RM samples are on the left side of each panel and are depicted with grey bars; CN samples are on the right side of each panel and are depicted with black bars. The vertical axis reflects absolute transcript quantity. The height of each bar is the ratio of the absolute quantities of the gene of interest (FOS, FOSB, COPEB, or CXCL2) to the control gene (GUSB) in each sample. Results from duplicate PCRs are averaged.

tensity of these probesets in each of the 29 samples. Among the 127 probesets are 7 that represent ESTs not yet assigned to a genetic locus, and 25 that represent a total of 10 genes that are detected by multiple probesets (range: 2–4 probesets per gene). When these were accounted for, 105 distinct locus-assigned genes were differentially expressed. Forty of 105 (38%) were overexpressed and 65 of 105 (62%) were underexpressed in CN compared to RM. Additional information about these probesets is provided in Supplemental File 2.

#### Validation of microarray data

We used qRT-PCR to confirm gene expression levels of 4 of the 105 genes (CXCL2, FOS, FOSB, KLF6) selected based upon consistent expression levels in the 29 RM and CN samples. Each gene was tested in 6 independent samples (3 RM and 3 CN). A total of 16 independent samples (8 RM and 8 CN) had sufficient remaining unamplified RNA, and each sample was examined with 1–3 test genes plus the control gene. As shown in Figure 3, we found that the relative abundance of every test gene recapitulated the relative expression levels on the microarray.

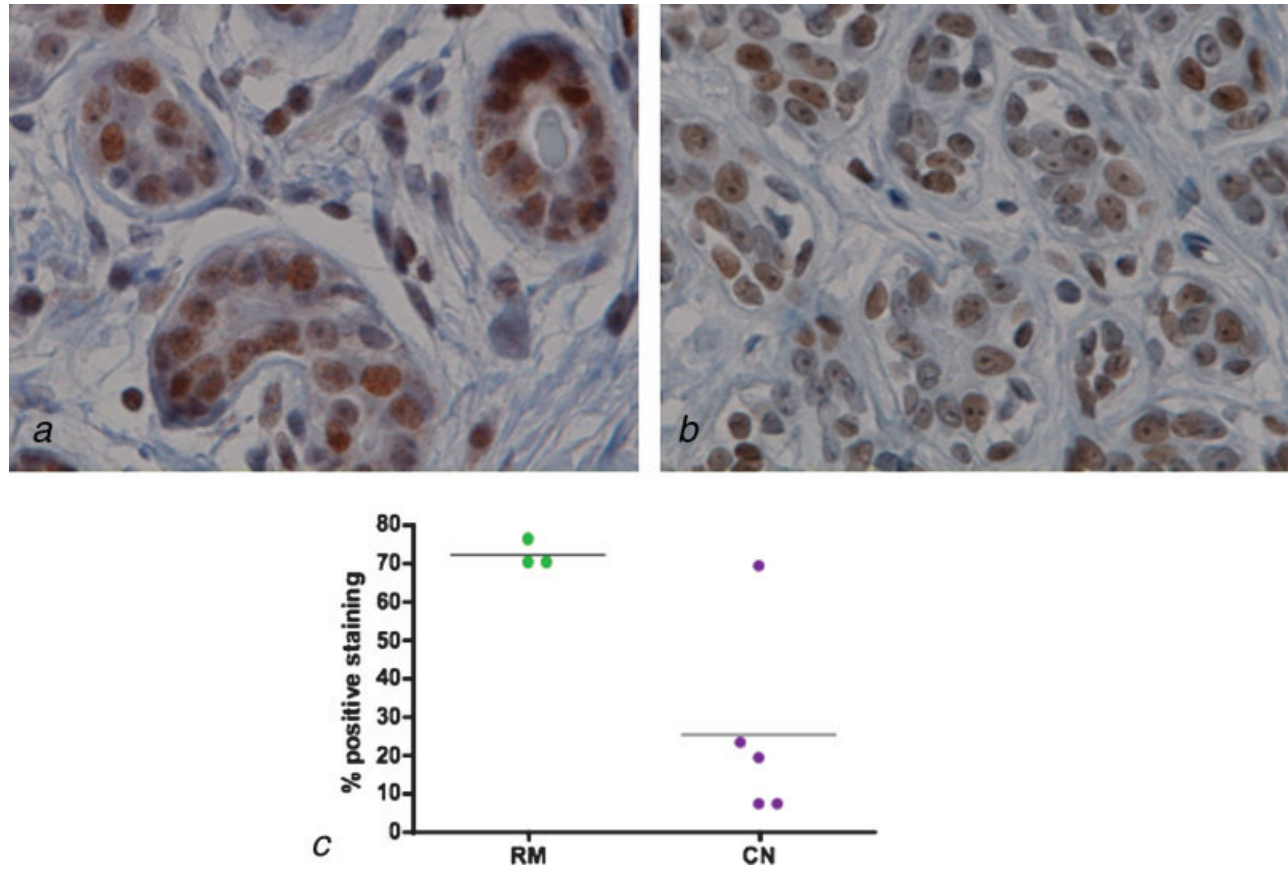
In a different approach to confirming the microarray data, we examined protein expression of FOSB in 8 microarray cases (3 RM and 5 CN) by IHC. FOSB was chosen because it had been evaluated by qRT-PCR, a reliable antibody was commercially available for use in formalin-fixed, paraffin-embedded tissue, and the protein's nuclear location makes quantitative automated image analysis feasible. One RM and 3 CN had been examined by qRT-PCR for FOSB transcript expression and the others had not. As shown in Figure 4, IHC corroborated what was seen in the microarray and by qRT-PCR. In all 3 RM cases, ~70% of the ductal epithelial cell area stained for FOSB, whereas in 4 of 5 CN cases, ~20% or less of the ductal cell area stained for FOSB. In the 5th CN case, FOSB staining resembled the level seen in RM tissues (69%).

To evaluate if the CN vs. RM expression data were relevant to breast cancer, we examined gene expression in a set of 6 CIS from independent patients with ER+ breast cancers collected using the same protocols as part of a separate study (manuscript in preparation). Despite the CIS patients being substantially older than the CN or RM patients (median age = 76, range 48–92), we found that the majority of the genes differentially expressed between CN and RM samples were also differentially expressed between CIS and RM samples. Specifically, 104 of the 127 (82%) probesets were also differentially expressed between the CIS and RM samples (Cyber-T-derived FDR < 0.10). For 102/104 (98%) probesets, the direction of differential expression in CIS was the same as in CN ( $\chi^2$   $p$ -value < 0.0001). These data are summarized in Figure 2b and Table 1.

To further evaluate the relevance of our CN vs. RM results to breast cancer, we examined an independent dataset of genes differentially expressed between invasive ductal carcinomas and normal luminal epithelium cultured from reduction mammoplasties.<sup>28</sup> We found that 75 of the 105 genes in our list were also differentially expressed between cancer and normal epithelium in that study, and that 80% of these 75 genes showed the same direction of change in both datasets, indicating significant concordance ( $\chi^2$  test;  $p$  = 0.0002) and demonstrating the similar differential expression of the majority of the 105 genes in an independent data set (see Supplemental File 2).

#### Functional analysis of microarray data

We took several approaches to identify the potential functional significance of the 105 differentially expressed genes. We used gene ontology (GO) to classify each differentially expressed gene into functional categories and then to determine if any categories were overrepresented compared to all genes on the array. The most overrepresented GO molecular function and biological process categories relate to DNA binding and various types of transcriptional activity (see Table 2). This is reflected in the numerous transcription factors that are differentially expressed, including AP-1 components, Kruppel-like factors, nuclear hormone recep-



**FIGURE 4** – Immunohistochemistry of FOSB protein. Normal appearing breast epithelium from 3 RM and 5 CN patients was stained for FOSB protein and scored using image analysis. Representative sections of staining in RM (panel *a*) and CN (panel *b*) epithelium are shown ( $\times 400$  magnification). Each case's percent positivity is graphed (panel *c*).

tors and zinc-finger proteins (see Table 1). Other significant GO-defined categories included G-protein coupled- and chemokine-receptor binding and activity, and cell proliferation, metabolism and response to various stimuli (see Table 2).

We also evaluated the pathways linking the differentially expressed genes. Using the KEGG pathway databases, as well as the iPATH and Ingenuity programs, we found that the MAPK signaling cascade contained the most genes from the list (DUSP1, DUSP2, FOS, GADD45 $\beta$ , JUN, JUND, NR4A1). The pathways containing the next largest number of genes were the cytokine-cytokine receptor interaction pathway (CCL2, CXCL1, CXCL2) and the calcium-signaling pathway (GNAS, ATP2B2, ADCY2). The genes noted above (except ADCY2) were underexpressed in CN epithelium compared to RM (see Table 1). Any functional connections among the 40 genes that were overexpressed in CN epithelium remain to be discovered.

Finally, we reviewed putative functions and categorization of the encoded proteins in the Uniprot database and the published literature. We noted a large number ( $n = 16$ ) of immediate early (IE) genes. We also noted that at least 32 of the 105 genes (31%) had been implicated previously in breast carcinogenesis and 34 additional genes (32%) had been implicated in other cancers, leaving 39 genes (37%) not previously reported to be associated with cancer (see Table 1). Some belong to functional categories implicated in cancer, and others are genes currently of unknown function.

## Discussion

The current understanding of events that initiate or predispose to breast carcinogenesis is limited. Therefore, the present study eval-

uated global gene expression in tumor-adjacent, histologically normal breast TDLUs microdissected from patients with untreated ER+ breast cancers, compared to TDLUs from control patients with no increased breast cancer risk. We identified differences in 127 probesets, corresponding to 105 genes. Most differences were maintained in a set of CIS. The 105 genes included a large group of transcriptional regulators, IE genes, and members of signaling pathways. The majority of these genes were expressed at lower levels in epithelium from women with cancer. One-third of the genes have been implicated previously in breast cancer, another third have been implicated in other cancers, and a final third have not been associated with cancer before. We cannot determine if these changes represent an effect of the tumor or an occult premalignant condition. But taken together, the data suggest that perturbations of key cellular functions are identifiable prior to the development of any histological abnormality, and that these perturbations may play important roles in the early stages of breast carcinogenesis.

Several potential objections could be raised to our study. The number of patients investigated is small, due to practical limits on the number of samples that can be investigated meticulously. However, a counterbalancing strength of the study is its use of primary uncultured epithelium, which eliminates introduction of artifacts inherent in cultured cells. We used amplified RNA, because only nanogram quantities are available from microdissected epithelium; however, we (and others) have shown that this approach yields reliable and reproducible data in which the biological variation between samples is greater than the technical variation between replicates.<sup>20</sup> The data may not be generalizable to ER-breast cancers, but that would not be unexpected, given breast cancers' considerable intrinsic heterogeneity.

**TABLE 1 – FUNCTIONAL CLASSIFICATION OF GENES DIFFERENTIALLY EXPRESSED IN EPITHELIUM FROM ER+ BREAST CANCERS (CN) COMPARED TO REDUCTION MAMMOPLASTY CONTROLS (RM)**

| Genes and functional classes         | CN vs.<br>RM     | Fold chg:<br>CN vs. RM | CIS vs.<br>RM | Fold chg:<br>CIS vs. RM | Implicated in cancer<br>previously |            |
|--------------------------------------|------------------|------------------------|---------------|-------------------------|------------------------------------|------------|
|                                      |                  |                        |               |                         | Breast                             | Other type |
| Transcription factors and regulators |                  |                        |               |                         |                                    |            |
| AP-1 components                      |                  |                        |               |                         |                                    |            |
| FOS <sup>1</sup>                     | ↓                | −10.0                  | ↓             | −13.5                   | ✓ (1s) <sup>2</sup>                | ✓          |
| FOSB <sup>1</sup>                    | ↓                | −21.2                  | ↓             | −75.9                   | ✓ (1s)                             | ✓          |
| JUN <sup>1</sup>                     | ↓                | −2.5                   | ↓             | −4.7                    | ✓ (1s)                             | ✓          |
| JUNB <sup>1</sup>                    | ↓                | −2.2                   |               |                         | ✓ (1s)                             | ✓          |
| JUND <sup>1</sup>                    | ↓                | −1.7                   | ↓             | −1.8                    | ✓ (26s)                            |            |
| ATF3                                 | ↓                | −5.8                   | ↓             | −16.9                   | ✓ (1s)                             |            |
| Kruppel-like factors                 |                  |                        |               |                         |                                    |            |
| KLF4                                 | ↓                | −3.8                   | ↓             | −2.9                    | ✓ (2s)                             | ✓          |
| KLF6                                 | ↓                | −3.6                   | ↓             | −6.8                    |                                    | ✓          |
| KLF10 <sup>1</sup>                   | ↓                | −1.8                   | ↓             | −2.2                    |                                    | ✓          |
| KLF11 <sup>1</sup>                   | ↓                | −1.5                   | ↓             | −2.4                    |                                    | ✓          |
| Nuclear hormone receptors            |                  |                        |               |                         |                                    |            |
| NR4A1 <sup>1</sup>                   | ↓                | −4.1                   | ↓             | −4.4                    |                                    | ✓          |
| NR4A2 <sup>1</sup>                   | ↓                | −6.4                   |               | −5.0                    |                                    | ✓          |
| NR4A3 <sup>1</sup>                   | ↓                | −3.4                   | ↓             | −2.9                    |                                    | ✓          |
| Zinc finger proteins                 |                  |                        |               |                         |                                    |            |
| ZFP36                                | ↓                | −3.1                   | ↓             | −4.5                    |                                    |            |
| ZFP36L2                              |                  | −1.5                   |               | −3.0                    |                                    |            |
| ZNF232                               | ↑                | 1.6                    | ↑             | 1.2                     |                                    |            |
| ZNF451                               | ↑                | 1.2                    |               |                         |                                    |            |
| Other                                |                  |                        |               |                         |                                    |            |
| BHLHB2                               | ↓                | −2.1                   |               |                         | ✓ (3s)                             | ✓          |
| BTG2 <sup>1</sup>                    | ↓                | −3.8                   | ↓             | −3.6                    | ✓ (5s)                             | ✓          |
| EGR1 <sup>1</sup>                    | ↓                | −2.4                   | ↓             | −9.9                    | ✓ (4s)                             | ✓          |
| HEY2                                 | ↓                | −2.1                   | ↓             | −2.0                    |                                    |            |
| ILF3                                 | ↑                | 1.6                    | ↑             | 1.6                     | ✓ (6s)                             |            |
| NFIL3                                | ↓                | −1.9                   | ↓             | −4.7                    |                                    | ✓          |
| PNRC1                                | ↓                | −1.8                   | ↓             | −2.1                    | ✓ (20s)                            |            |
| SAFB2                                | ↑                | 1.6                    | ↑             | 2.1                     | ✓ (8s)                             |            |
| SFPQ                                 | ↓                | −1.7                   | ↓             | −1.4                    |                                    |            |
| SPDEF                                | ↑                | 2.4                    | ↑             | 9.6                     | ✓ (7s)                             |            |
| Cytokines                            |                  |                        |               |                         |                                    |            |
| CD69                                 | ↓                | −2.5                   | ↓             | −2.5                    | ✓ (14s)                            | ✓          |
| CXCL1                                | ↓                | −2.0                   | ↓             | −5.4                    | ✓ (9s)                             | ✓          |
| CXCL2                                | ↓                | −3.4                   | ↓             | −25.6                   | ✓ (10s)                            | ✓          |
| CCL2                                 | ↓                | −5.9                   | ↓             | −8.5                    | ✓ (9s)                             |            |
| PBEF1                                | ↓                | −1.8                   | ↓             | −2.4                    |                                    | ✓          |
| Nucleosome components (histones)     |                  |                        |               |                         |                                    |            |
| HIST2H2BE                            | ↓                | −2.4                   |               |                         |                                    |            |
| H3F3B                                | ↓                | −3.0                   | ↓             | −3.1                    |                                    | ✓          |
| Translation factors                  |                  |                        |               |                         |                                    |            |
| EIF1                                 | ↓                | −3.3                   | ↓             | −6.0                    |                                    |            |
| EIF5                                 | ↓/↑ <sup>3</sup> | −2.9/1.5               | ↑             | −3.8/1.5                |                                    | ✓          |
| EIF4A1                               | ↓                | −1.8                   | ↓             | −2.9                    |                                    | ✓          |
| Phosphatases                         |                  |                        |               |                         |                                    |            |
| DUSP1                                | ↓                | −3.7                   | ↓             | −5.7                    |                                    | ✓          |
| DUSP2                                | ↓                | −4.4                   | ↓             | −5.6                    |                                    |            |
| PTP4A1 (also stimulates G1-S)        | ↓                | −2.5                   | ↓             | −2.0                    |                                    | ✓          |
| DNA damage and repair                |                  |                        |               |                         |                                    |            |
| FANCG                                | ↑                | 2.3                    | ↑             | 4.2                     | ✓ (11s)                            | ✓          |
| GADD45B                              | ↓                | −2.9                   |               |                         | ✓ (9s)                             | ✓          |
| Cell cycle and division              |                  |                        |               |                         |                                    |            |
| YWHAZ                                | ↓                | −1.9                   | ↓             | −1.4                    | ✓ (25s)                            | ✓          |
| CCNL1                                | ↓                | −2.3                   | ↓             | −2.8                    |                                    | ✓          |
| CEBPD                                | ↓                | −2.6                   |               | −2.8                    | ✓ (15s)                            |            |
| YPEL5                                | ↓                | −1.7                   | ↓             | −2.1                    |                                    |            |
| Metabolic or enzymatic activity      |                  |                        |               |                         |                                    |            |
| ADCY2                                | ↑                | 2.2                    |               |                         |                                    |            |
| AKR7A2                               | ↑                | 1.6                    | ↑             | 2.2                     |                                    | ✓          |
| ALDH1A3                              | ↓                | −2.0                   | ↓             | −9.0                    |                                    | ✓          |
| CANT1                                | ↑                | 1.8                    | ↑             | 3.1                     |                                    |            |
| CNDP2                                | ↑                | 1.7                    | ↑             | 2.3                     |                                    |            |
| GBA                                  | ↑                | 2.7                    | ↑             | 7.2                     |                                    |            |
| ODC1                                 | ↓                | −1.7                   | ↓             | −2.6                    |                                    | ✓          |
| Cell adhesion                        |                  |                        |               |                         |                                    |            |
| CLDN1                                | ↓                | −3.6                   | ↓             | −2.4                    | ✓ (16s)                            | ✓          |
| CYR61                                | ↓                | −1.8                   | ↓             | −5.5                    | ✓ (18s)                            | ✓          |
| FLRT2                                | ↑                | 1.8                    |               | −2.0                    |                                    |            |
| TRO                                  | ↑                | 2.4                    | ↓             | −3.2                    |                                    | ✓          |



**TABLE 1 – FUNCTIONAL CLASSIFICATION OF GENES DIFFERENTIALLY EXPRESSED IN EPITHELIUM FROM ER+ BREAST CANCERS (CN) COMPARED TO REDUCTION MAMMOPLASTY CONTROLS (RM) (CONTINUED)**

| Genes and functional classes                            | CN vs. RM | Fold chg: CN vs. RM | CIS vs. RM | Fold chg: CIS vs. RM | Implicated in cancer previously |            |
|---|-----------|---------------------|------------|----------------------|---------------------------------|------------|
|   |           |                     |            |                      | Breast                          | Other type |
| <b>Apoptosis-related</b>                                |           |                     |            |                      |                                 |            |
| MCL1  | ↓         | −1.8                | ↓          | −1.9                 | ✓ (19s)                         | ✓          |
| HBEGF   | ↓         | −2.5                | ↓          | −2.7                 |                                 | ✓          |
| PMAIP1 <sup>1</sup>                                     | ↓         | −2.7                |            |                      |                                 | ✓          |
| <b>Protein transport and folding</b>                    |           |                     |            |                      |                                 |            |
| CLGN  | ↑         | 2.3                 | ↑          | 7.8                  |                                 | ✓          |
| LMAN2L  | ↑         | 1.7                 | ↑          | 1.9                  |                                 |            |
| <b>Inflammation</b>                                     |           |                     |            |                      |                                 |            |
| ANXA1   | ↓         | −1.9                | ↓          | −5.4                 | ✓ (12s)                         | ✓          |
| PTGS2   | ↓         | −2.9                | ↓          | −7.1                 | ✓ (21s)                         |            |
| <b>G-protein, RhoGAP and GTPase related</b>             |           |                     |            |                      |                                 |            |
| ARHGAP8   | ↑         | 1.7                 | ↑          | 3.2                  | ✓ (13s)                         | ✓          |
| ARHGEF2   | ↑         | 1.8                 | ↑          | 1.9                  |                                 | ✓          |
| GNAS  | ↓         | −1.9                |            |                      |                                 | ✓          |
| HMHA1   | ↑         | 2.4                 | ↑          | 4.3                  |                                 | ✓          |
| RGS1 <sup>1</sup>                                       | ↓         | −2.5                |            |                      |                                 | ✓          |
| RGS2 <sup>1</sup>                                       | ↓         | −2.6                | ↓          | −6.5                 |                                 | ✓          |
| RHOB  | ↓         | −2.2                |            |                      | ✓ (22s)                         | ✓          |
| <b>Channel or pump proteins</b>                         |           |                     |            |                      |                                 |            |
| ATP2B2  | ↓         | −2.3                | ↓          | −2.3                 |                                 |            |
| CLNS1A  | ↑         | 1.6                 |            |                      | ✓ (17s)                         |            |
| <b>Ubiquitin-related</b>                                |           |                     |            |                      |                                 |            |
| CYLD  | ↓         | −1.7                | ↓          | −2.1                 |                                 | ✓          |
| RNF7  | ↑         | 1.7                 | ↑          | 2.3                  |                                 | ✓          |
| TRIM9   | ↑         | 2.1                 |            |                      |                                 | ✓          |
| <b>RNA binding</b>                                      |           |                     |            |                      |                                 |            |
| CSDE1   | ↓         | −1.6                | ↓          | −2.4                 |                                 |            |
| HNRPA1  | ↓         | −1.5                | ↓          | −1.3                 |                                 | ✓          |
| <b>Roles in signaling</b>                               |           |                     |            |                      |                                 |            |
| PELI1 (IL-1 receptor associated kinase)                 | ↓         | −1.6                | ↓          | −5.9                 |                                 |            |
| SOSTDC1 (antagonizes BMP signaling)                     | ↓         | −2.3                | ↓          | −17.6                |                                 |            |
| TACSTD2 (cell surface receptor)                         | ↓         | −3.8                | ↓          | −3.1                 | ✓ (23s)                         |            |
| TRIB1 (MAPK regulator)                                  | ↓         | −2.1                | ↓          | −2.5                 |                                 | ✓          |
| WNT5B (Wnt pathway)                                     | ↑         | 2.3                 |            |                      | ✓ (24s)                         | ✓          |
| <b>Others</b>   |           |                     |            |                      |                                 |            |
| IER2 <sup>1</sup>                                       | ↓         | −4.4                | ↓          | −8.2                 |                                 |            |
| LSR (binds lipids)                                      | ↑         | 2.9                 | ↑          | 6.7                  |                                 |            |
| SNF1LK (kinase; differentiation)                        | ↑         | −3.6                | ↓          | −3.7                 |                                 | ✓          |
| <b>Genes of unknown function and hypothetical genes</b> |           |                     |            |                      |                                 |            |
| EBI2  | ↓         | −2.4                | ↓          | −2.3                 |                                 |            |
| HD  | ↑         | 2.0                 | ↑          | 3.9                  |                                 | ✓          |
| JOSD3   | ↑         | 1.5                 | ↑          | 1.2                  |                                 |            |
| LAGE3 = DXS9879E  | ↑         | 1.7                 | ↑          | 5.2                  |                                 |            |
| MORC2   | ↑         | 2.1                 | ↑          | 4.9                  |                                 |            |
| MSTP9   | ↑         | 1.5                 |            |                      |                                 |            |
| MUM1  | ↑         | 1.9                 | ↑          | 2.5                  |                                 | ✓          |
| NOL2 = MGC3731  | ↑         | 3.8                 | ↑          | 3.2                  |                                 |            |
| SIDT2   | ↑         | 1.6                 |            |                      |                                 |            |
| ZNHIT1  | ↑         | 2.1                 | ↑          | 5.2                  |                                 |            |
| FLJ10213  | ↑         | 1.7                 | ↑          | 1.6                  |                                 |            |
| FLJ35348  | ↑         | 2.4                 | ↑          | 2.8                  |                                 |            |
| FLJ20699  | ↑         | 2.0                 |            |                      |                                 |            |
| KIAA0690  | ↓         | −1.9                |            |                      |                                 |            |
| MGC5139   | ↑         | 1.5                 |            |                      |                                 |            |
| SLC35E2   | ↑         | 2.6                 | ↑          | 2.9                  |                                 |            |
| C2orf28   | ↑         | 1.6                 | ↑          | 2.2                  |                                 |            |
| C9orf3  | ↑         | 1.9                 |            |                      |                                 |            |
| C9orf26   | ↓         | −2.3                | ↓          | −9.7                 |                                 |            |

Each gene is listed only once, although genes may have multiple functions. The direction of differential expression is indicated with an arrow: higher is an upwards arrow, lower is a downwards arrow. A similar scheme is used for genes with significant differential expression in carcinoma *in situ* (CIS) relative to RM.

<sup>1</sup>Immediate early (IE) gene. <sup>2</sup>Numbers in parentheses refer to references listed in Supplemental File 3. <sup>3</sup>For this gene, two probesets yielded contradictory results. For all other genes represented by multiple probesets, all probesets showed the same direction of change.

**TABLE II – FUNCTIONAL CLASSIFICATION OF THE 105 DIFFERENTIALLY EXPRESSED GENES, BY GO, AND THEIR OVERREPRESENTATION (BY EASE)**

| System             | Category   | No. of genes | EASE score |
|--------------------|--|--------------|------------|
| Molecular function | DNA binding  | 31           | 0.00000129 |
|                    | Transcription regulator activity                               | 23           | 0.0000114  |
|                    | Transcription factor activity                                  | 19           | 0.0000202  |
|                    | Nucleic acid binding   | 33           | 0.000183   |
|                    | Transcription corepressor activity                             | 5            | 0.00493    |
|                    | Binding  | 58           | 0.0159     |
|                    | G-protein-coupled receptor binding                             | 3            | 0.0488     |
|                    | Chemokine receptor binding                                     | 3            | 0.0488     |
| Biological process | Chemokine activity   | 3            | 0.0488     |
|                    | Regulation of transcription, DNA-dependent                     | 23           | 0.000839   |
|                    | Regulation of transcription                                    | 23           | 0.00114    |
|                    | Transcription, DNA-dependent                                   | 23           | 0.00157    |
|                    | Transcription  | 23           | 0.00258    |
|                    | Negative regulation of transcription, DNA-dependent            | 5            | 0.00356    |
|                    | Nucleobase, nucleoside, nucleotide and nucleic acid metabolism | 30           | 0.00448    |
|                    | Cell proliferation   | 17           | 0.00452    |
|                    | Regulation of cellular process                                 | 10           | 0.00542    |
|                    | Negative regulation of transcription                           | 5            | 0.00882    |
|                    | Regulation of transcription from Pol II promoter               | 7            | 0.0114     |
|                    | Response to stimulus   | 21           | 0.0118     |
|                    | Negative regulation of transcription from Pol II promoter      | 4            | 0.0124     |
|                    | Response to biotic stimulus                                    | 14           | 0.0153     |
|                    | Regulation of cell proliferation                               | 7            | 0.0163     |
|                    | Response to external stimulus                                  | 18           | 0.0187     |
|                    | Regulation of biological process                               | 10           | 0.0231     |
|                    | Negative regulation of cell proliferation                      | 5            | 0.0239     |
|                    | Cell growth and/or maintenance                                 | 33           | 0.0349     |
|                    | Defense response   | 12           | 0.0402     |
|                    | Physiological process  | 37           | 0.0434     |
|                    | Metabolism   | 51           | 0.0465     |
|                    | Immune response  | 11           | 0.047      |
| Cellular component | Nucleus  | 34           | 0.00017    |
|                    | Intracellular  | 46           | 0.0349     |
|                    | Chromosome   | 5            | 0.049      |

All categories with EASE scores < 0.05 are listed.

Despite these potential objections, the data raise several points for consideration. First, how do our results compare to existing expression data from human breast tissue? Most existing breast tissue expression signatures were derived to predict tumor subtype<sup>29,30</sup> or disease outcome,<sup>31–35</sup> or to distinguish luminal from myoepithelial cells in RM tissue,<sup>28,36</sup> as opposed to distinguishing between patients with and without breast cancer, and so are not directly comparable to our data. It is therefore not surprising that few of the genes that we find to vary between CN and RM epithelium have been useful in predicting tumor subtype, disease outcome, or epithelial cell type (analyses not shown).

#### Other studies are more comparable to ours

One found no gene expression differences between RM and tumor-adjacent normal epithelium by unsupervised hierarchical clustering.<sup>37</sup> We also could not discern differences between CN and RM by unsupervised hierarchical clustering or principal component analysis of all genes (results not shown). This may be due to the presence of genes that vary from patient to patient and obscure the consistent differences between CN and RM that we see when comparing directly these 2 sample types. In contrast, there is overlap between our results and those reported to distinguish TDLUs from an early hyperplastic breast cancer precursor.<sup>38</sup> There is also overlap between our results and those reported in a study comparing luminal epithelium from RM and cancers.<sup>28</sup> These reports, combined with the fact that the CN vs. RM differences are largely preserved in the independent CIS samples we examined, suggest that the CN vs. RM differences are authentic alterations reflecting a breast cancer related process.

Second, although the CN vs. RM differences appear authentic, we cannot distinguish whether they represent cause or effect, *i.e.*, an occult premalignant condition, or secondary changes due to the tumor or its surrounding stroma. We favor the former explanation, because of the similarity of the CN vs. RM differences to cancer microarray data. However, we cannot determine how far the affected area might extend geographically, since all CN TDLUs were tumor-adjacent. Tissue that is adjacent to a breast tumor may harbor more, or different, genomic abnormalities than tissue that is more distant.<sup>4</sup>

If the CN vs. RM differences represent a primary abnormality, then the identification of genes whose expression varies in normal epithelium from patients with breast cancer, compared to controls, suggests mechanisms that may predispose to breast cancer development or are active early in carcinogenesis. If the CN vs. RM differences represent a secondary abnormality, then they can illuminate direct or paracrine effects occurring *in vivo*. Regardless, the largest functional category among the 105 genes is transcription factors and regulators, especially members of the composite transcription factor AP-1. Transcription regulators are implicated frequently in breast carcinogenesis (for review see Ref. 39). Many transcription-related genes (23/29 (79%)) were underexpressed in CN (and CIS) samples, which may reflect a generalized decrease in transcriptional activity, rather than involvement of a specific family. Also notable among the 105 genes was a large group ( $n = 16$ ) of IE genes, which are rapidly induced upon cell stimulation and whose transcription is not dependent on protein synthesis. The IE genes were also underexpressed in CN (and CIS) epithelium. In addition, many of the 105 genes participate in signaling pathways. The largest number participates directly in the MAPK pathway, and others may affect MAPK signaling more peripherally. Consid-



erable evidence supports the involvement of MAPK in breast cancer (for reviews see Refs. 39 and 40). Increased MAPK activity is usually reported, but the tumors examined have been mainly ER-negative and ERBB2-overexpressing.<sup>41,42</sup> In contrast, we found decreased expression of MAPK components, which could be related to using tissue from ER-positive tumors, or may reflect an initial step in the pathway's perturbation.

A final consideration is how can these data be utilized. If validated in future studies, the genes or pathways implicated here could identify new targets for chemoprevention, or help prioritize those already being studied.<sup>43</sup> If differential expression of these genes can be detected in women without evident breast cancer, and associated with future disease, then they may be pertinent to risk assessment, since breast cancer risk is not thought to be uniform across all women.<sup>44,45</sup> DNA structural variants<sup>46–48</sup> or single nucleotide polymorphisms might alter RNA expression<sup>49</sup> and be associated with risk of disease.

To our knowledge, this is the first study to find gene expression differences between histologically normal epithelium of breast cancer patients and breast-cancer free controls. Our findings suggest that cancer-related pathways are already perturbed in normal epithelium of breast cancer patients. These perturbations could be markers of disease risk, of occult disease, or of the tissue's response to an existing tumor. Future studies should expand upon these results by examining expression of these genes in additional samples from breast cancer patients and controls, manipulating these genes' expression in model systems and developing clinically useful disease and risk classifiers.

### Acknowledgement

This work was supported by grants from the Department of Defense Breast Cancer Research Program (DAMD17-01-1-0159) and the NIH (RO1 CA081078, S10 RR021211) to CLR.

### References

- Cavalli LR, Singh B, Isaacs C, Dickson RB, Haddad BR. Loss of heterozygosity in normal breast epithelial tissue and benign breast lesions in BRCA1/2 carriers with breast cancer. *Cancer Genet Cytogenet* 2004;149:38–43.
- Clarke CL, Sandle J, Jones AA, Sofronis A, Patani NR, Lakhani SR. Mapping loss of heterozygosity in normal human breast cells from BRCA1/2 carriers. *Br J Cancer* 2006;95:515–9.
- Deng G, Lu Y, Zlotnikov G, Thor AD, Smith HS. Loss of heterozygosity in normal tissue adjacent to breast carcinomas. *Science* (New York) 1996;274:2057–9.
- Ellsworth DL, Ellsworth RE, Love B, Deyarmin B, Lubert SM, Mittal V, Shriver CD. Genomic patterns of allelic imbalance in disease free tissue adjacent to primary breast carcinomas. *Breast Cancer Res Treat* 2004;88:131–9.
- Lakhani SR, Chaggar R, Davies S, Jones C, Collins N, Odel C, Stratton MR, O'Hare MJ. Genetic alterations in "normal" luminal and myoepithelial cells of the breast. *J Pathol* 1999;189:496–503.
- Larson PS, de las Morenas A, Bennett SR, Cupples LA, Rosenberg CL. Loss of heterozygosity or allele imbalance in histologically normal breast epithelium is distinct from loss of heterozygosity or allele imbalance in co-existing carcinomas. *Am J Pathol* 2002;161:283–90.
- Larson PS, de las Morenas A, Cupples LA, Huang K, Rosenberg CL. Genetically abnormal clones in histologically normal breast tissue. *Am J Pathol* 1998;152:1591–8.
- Larson PS, Schlechter BL, de las Morenas A, Garber JE, Cupples LA, Rosenberg CL. Allele imbalance, or loss of heterozygosity, in normal breast epithelium of sporadic breast cancer cases and BRCA1 gene mutation carriers is increased compared with reduction mammaplasty tissues. *J Clin Oncol* 2005;23:8613–9.
- Holst CR, Nuovo GJ, Esteller M, Chew K, Baylin SB, Herman JG, Tlsty TD. Methylation of p16(INK4a) promoters occurs in vivo in histologically normal human mammary epithelia. *Cancer Res* 2003;63:1596–601.
- Yan PS, Venkataramu C, Ibrahim A, Liu JC, Shen RZ, Diaz NM, Centeno B, Weber F, Leu YW, Shapiro CL, Eng C, Yeatman TJ, et al. Mapping geographic zones of cancer risk with epigenetic biomarkers in normal breast tissue. *Clin Cancer Res* 2006;12:6626–36.
- Cianciulli AM, Pescatore B, Bovani R, Coletta AM, Gandolfo GM, Greco C, Botti C. Aneusomy of chromosomes 1 and 17 in normal tissue adjacent to breast carcinomas. *Eur J Histochem* 1997;41(Suppl 2):159–60.
- Steinarsdottir M, Jonasson JG, Vidarsson H, Juliusdottir H, Hauksdottir H, Ogmundsdottir HM. Cytogenetic changes in nonmalignant breast tissue. *Genes Chromosomes Cancer* 2004;41:47–55.
- Meeker AK, Hicks JL, Gabrielson E, Strauss WM, De Marzo AM, Argani P. Telomere shortening occurs in subsets of normal breast epithelium as well as in situ and invasive carcinoma. *Am J Pathol* 2004;164:925–35.
- van Roozendaal CE, Gillis AJ, Klijn JG, van Ooijen B, Claassen CJ, Eggermont AM, Henzen-Logmans SC, Oosterhuis JW, Foekens JA, Looijenga LH. Loss of imprinting of IGF2 and not H19 in breast cancer, adjacent normal tissue and derived fibroblast cultures. *FEBS Lett* 1998;437:107–11.
- Khan SA, Sachdeva A, Naim S, Meguid MM, Marx W, Simon H, Halverson JD, Numann PJ. The normal breast epithelium of women with breast cancer displays an aberrant response to estradiol. *Cancer Epidemiol Biomarkers Prev* 1999;8:867–72.
- Widschwendter M, Berger J, Daxenbichler G, Muller-Holzner E, Widschwendter A, Mayr A, Marth C, Zeimet AG. Loss of retinoic acid receptor beta expression in breast cancer and morphologically normal adjacent tissue but not in the normal breast tissue distant from the cancer. *Cancer Res* 1997;57:4158–61.
- Gauthier ML, Pickering CR, Miller CJ, Fordyce CA, Chew KL, Berman HK, Tlsty TD. p38 regulates cyclooxygenase-2 in human mammary epithelial cells and is activated in premalignant tissue. *Cancer Res* 2005;65:1792–9.
- Ding L, Erdmann C, Chinnaiyan AM, Merajver SD, Kleer CG. Identification of EZH2 as a molecular marker for a precancerous state in morphologically normal breast tissues. *Cancer Res* 2006;66:4095–9.
- Botti C, Pescatore B, Mottolese M, Sciarretta F, Greco C, Di Filippo F, Gandolfo GM, Cavaliere F, Bovani R, Varanese A, Cianciulli AM. Incidence of chromosomes 1 and 17 aneusomy in breast cancer and adjacent tissue: an interphase cytogenetic study. *J Am Coll Surg* 2000;190:530–9.
- King C, Guo N, Frampton GM, Gerry NP, Lenburg ME, Rosenberg CL. Reliability and reproducibility of gene expression measurements using amplified RNA from laser-microdissected primary breast tissue with oligonucleotide arrays. *J Mol Diagn* 2005;7:57–64.
- Liu G, Loraine AE, Shigeta R, Cline M, Cheng J, Valmeekam V, Sun S, Kulp D, Siani-Rose MA. NetAffx: affymetrix probesets and annotations. *Nucleic Acids Res* 2003;31:82–6.
- Baldi P, Long AD. A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics* 2001;17:509–19.
- Carson JP, Zhang N, Frampton GM, Gerry NP, Lenburg ME, Christman MF. Pharmacogenomic identification of targets for adjuvant therapy with the topoisomerase poison camptothecin. *Cancer Res* 2004;64:2096–104.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995;57:289–300.
- Tricarico C, Pinzani P, Bianchi S, Paglierani M, Distanti V, Pazzagli M, Bustin SA, Orlando C. Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single house-keeping genes is inappropriate for human tissue biopsies. *Anal Biochem* 2002;309:293–300.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25–9.
- Lakhani SR, Van De Vijver MJ, Jacquemier J, Anderson TJ, Osin PP, McGuffog L, Easton DF. The pathology of familial breast cancer: predictive value of immunohistochemical markers estrogen receptor, progesterone receptor, HER-2, and p53 in patients with mutations in BRCA1 and BRCA2. *J Clin Oncol* 2002;20:2310–8.
- Grigoriadis A, Mackay A, Reis-Filho JS, Steele D, Iseli C, Stevenson BJ, Jongeneel CV, Valgeirsson H, Fenwick K, Iravani M, Leao M, Simpson AJ, et al. Establishment of the epithelial-specific transcriptome of normal and malignant human breast cells based on MPSS and array expression data. *Breast Cancer Res* 2006;8:R56.
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, et al. Gene expression patterns of breast carcinomas distinguish tumor sub-classes with clinical implications. *Proc Natl Acad Sci USA* 2001;98:10869–74.
- Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 2003;100:8418–23.

31. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoe T, Berns EM, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005;365: 671–9.
32. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;347:1999–2009.
33. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–6.
34. Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, Montgomery K, Chi JT, van de Rijn M, Botstein D, Brown PO. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* 2004;2:E7.
35. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, et al. A multi-gene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817–26.
36. Jones C, Mackay A, Grigoriadis A, Cossu A, Reis-Filho JS, Fulford L, Dexter T, Davies S, Bulmer K, Ford E, Parry S, Budroni M, et al. Expression profiling of purified normal human luminal and myoepithelial breast cells: identification of novel prognostic markers for breast cancer. *Cancer Res* 2004;64:3037–45.
37. Finak G, Sadekova S, Pepin F, Hallett M, Meterissian S, Halwani F, Khetani K, Souleimanova M, Zabolotny B, Omeroglu A, Park M. Gene expression signatures of morphologically normal breast tissue identify basal-like tumors. *Breast Cancer Res* 2006;8:R58.
38. Lee S, Medina D, Tsimelzon A, Mohsin SK, Mao S, Wu Y, Allred DC. Alterations of gene expression in the development of early hyperplastic precursors of breast cancer. *Am J Pathol* 2007;171: 252–62.
39. Shen Q, Brown PH. Novel agents for the prevention of breast cancer: targeting transcription factors and signal transduction pathways. *J Mammary Gland Biol Neoplasia* 2003;8:45–73.
40. Dunn KL, Espino PS, Drobnic B, He S, Davie JR. The Ras-MAPK signal transduction pathway, cancer and chromatin remodeling. *Biochem Cell Biol* 2005;83:1–14.
41. Janes PW, Daly RJ, deFazio A, Sutherland RL. Activation of the Ras signalling pathway in human breast cancer cells overexpressing erbB-2. *Oncogene* 1994;9:3601–8.
42. Creighton CJ, Hilger AM, Murthy S, Rae JM, Chinnaiyan AM, El-Ashry D. Activation of mitogen-activated protein kinase in estrogen receptor alpha-positive breast cancer cells in vitro induces an in vivo molecular phenotype of estrogen receptor alpha-negative human breast tumors. *Cancer Res* 2006;66:3903–11.
43. Shen Q, Brown PH. Transgenic mouse models for the prevention of breast cancer. *Mutat Res* 2005;576:93–110.
44. Peto J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer. *Nat Genet* 2000;26:411–4.
45. Antoniou AC, Pharoah PD, McMullan G, Day NE, Stratton MR, Peto J, Ponder BJ, Easton DF. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer* 2002;86:76–83.
46. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. *Nat Genet* 2004;36:949–51.
47. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, et al. Large-scale copy number polymorphism in the human genome. *Science (New York)* 2004;305:525–8.
48. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE. Fine-scale structural variation of the human genome. *Nat Genet* 2005;37:727–32.
49. Stranger BE, Dermitzakis ET. The genetics of regulatory variation in the human genome. *Hum Genomics* 2005;2:126–31.