# Epithelial to Mesenchymal Transition in Human Breast Epithelial Cells Transformed by 17β-Estradiol

Yong Huang,[1] Sandra V. Fernandez,[2] Shirlean Goodwin,[1] Patricia A. Russo,[2] Irma H. Russo,[2] Thomas R. Sutter,[1] and Jose Russo[2]

[1]W. Harry Feinstone Center for Genomic Research, University of Memphis, Memphis, Tennessee and [2]Breast Cancer Research Laboratory, Fox Chase Cancer Center, Philadelphia, Pennsylvania

## Abstract

**The estrogen dependence of breast cancer has long been recognized; however, the role of 17β-estradiol ($E_2$) in cancer initiation was not known until we showed that it induces complete neoplastic transformation of the human breast epithelial cells MCF-10F. $E_2$ treatment of MCF-10F cells progressively induced high colony efficiency and loss of ductulogenesis in early transformed (trMCF) cells and invasiveness in Matrigel invasion chambers. The cells that crossed the chamber membrane were collected and identified as bsMCF; their subclones were designated bcMCF; and the cells harvested from carcinoma formation in severe combined immunodeficient mice were designated caMCF. These phenotypes correlated with gene dysregulation during the progression of the transformation. The highest number of dysregulated genes was observed in caMCF, being slightly lower in bcMCF, and lowest in trMCF. This order was consistent with the extent of chromosome aberrations (caMCF > bcMCF >>> trMCF). Chromosomal amplifications were found in 1p36.12-pter, 5q21.1-qter, and 13q21.31-qter. Losses of the complete chromosome 4 and 8p11.21-23.1 were found only in tumorigenic cells. In tumor-derived cell lines, additional losses were found in 3p12.1-14.1, 9p22.1-pter, and 18q11.21-qter. Functional profiling of dysregulated genes revealed progressive changes in the integrin signaling pathway, inhibition of apoptosis, acquisition of tumorigenic cell surface markers, and epithelial-mesenchymal transition. In tumorigenic cells, the levels of E-cadherin, epithelial membrane antigen, and various keratins were low and CD44E/CD24 were negative, whereas SNAI2, vimentin, S100A4, FN1, HRAS, transforming growth factor β1, and CD44H were high. The phenotypic and genomic changes triggered by estrogen exposure that lead normal cells to tumorigenesis confirm the role of this steroid hormone in cancer initiation.** [Cancer Res 2007;67(23):11147–57]

## Introduction

Breast cancer is a malignancy whose dependence on ovarian function was shown by Beatson (1), who induced regression of advanced cancer in premenopausal women by surgically removing the ovaries. Thereafter, the same procedure was also proven to control the progression of metastatic disease (2). The identification of estrogen ($E_2$) production by the ovaries, the isolation of the estrogen receptor (ER) protein, and the greater incidence of ERα-positive tumors observed in postmenopausal women led to the identification of a strong association between estrogen exposure with increased breast cancer risk (3). Despite the epidemiologic and clinical evidence linking cumulative and sustained exposure to estrogens with increased risk of developing breast cancer (3), the ultimate mechanisms by which estrogens induce cancer and the specific cells they act upon for initiating malignant transformation have not been fully identified. Among the mechanisms of estrogen action, the most widely acknowledged is the binding of the hormone to its specific nuclear ERα, initiating a signal that is potently mitogenic (4). However, the fact that ERα knockout mice expressing the Wnt-1 oncogene (*ERKO/Wnt-1*) develop mammary tumors in response to treatment with estrogen provides direct evidence that $E_2$ may cause breast cancer through a genotoxic, non-ERα–mediated mechanism (5). This postulate is further supported by the observation that when ovariectomized mice are supplemented with estrogen, they develop a higher tumor incidence with shorter latency time than control animals, even in the presence of the pure antiestrogen ICI-182,780. Experimental studies on estrogen metabolism (6), formation of DNA adducts (7), carcinogenicity (8, 9), mutagenicity (10), and cell transformation (11, 12) have supported the hypothesis that reaction of specific estrogen metabolites, namely catechol estrogen-3,4-quinones (CE-3,4-Q) and, to a much lesser extent, CE-2,3-Q, can generate critical DNA mutations that initiate breast, prostate, and other cancers (13–15).

Our observations that ductal carcinomas originate in lobules type 1 (Lob.1) of the immature breast (16), which are the structures with the highest proliferative activity and highest percentage of ERα and progesterone receptor–positive cells, provide a mechanistic explanation for the higher susceptibility of these structures to undergo neoplastic transformation when exposed to chemical carcinogens, as shown by *in vitro* experiments (17). However, the role of ERα-positive and ERα-negative cells in the initiation of breast cancer is not clear. The fact that the cells that do proliferate in culture are ERα-negative suggests that the stem cells that originate cancer are the ERα-negative proliferating cells. This idea is further supported by our observations that MCF-10F, a spontaneously immortalized ERα-negative human breast epithelial cell line derived from breast tissues containing Lob.1 and Lob.2 (18), becomes malignant after exposure to the chemical carcinogen benz(*a*)pyrene (17) and 17β-estradiol ($E_2$; refs. 12, 19).

Breast cancer has been subdivided into five major subtypes: basal-like, Her2 (ERBB2)–overexpressing, normal breast tissue-like, and two subtypes of luminal-like, luminal A and luminal B (20). The

**Note:** Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/), and http://www.fccc.edu/research/pid/russo-j/index.html.

Present address for Y. Huang: Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL 60637.

**Requests for reprints:** Jose Russo, Breast Cancer Research Laboratory, Fox Chase Cancer Center, 333 Cottman Avenue, Philadelphia, PA 19111-2497. Phone: 215-728-4782; Fax: 215-728-2180; E-mail: Jose.Russo@fccc.edu.



doi:10.1158/0008-5472.CAN-07-1371

luminal-like subtypes display moderate to high expression of ERα and luminal cytokeratins, whereas the basal-like subtype is negative for both ERα and ERBB2, with high expression of basal cytokeratins 5 and 17. The ERBB2-overexpressing subtype is also ERα negative and, like the basal-like tumors, is associated with poorer prognosis as measured as time to development of distal metastasis (20–22). Altogether, these data support the concept that ER-positive and ER-negative tumors may originate from two different cell populations, as postulated earlier (14). In addition to differences inherent to the type of cell in which cancer originates, neoplastically initiated cells lose specific characteristics of epithelial differentiation as the result of their progression toward malignancy. As the epithelial cells lose their polarity and cell-to-cell junctions, regulated in part by the expression of E-cadherin, they acquire characteristics of mesenchymal cells, which lack stable intercellular junctions (21). This epithelial to mesenchymal transition (EMT) leads to exacerbation of motility and invasiveness in many cell types and is often considered a prerequisite for tumor infiltration and metastasis (21).

To outline the pathways through which estrogen acts as carcinogen in the human breast (i.e., either through the receptor pathway or through a genotoxic effect in a specific cell type of the breast), we used an *in vitro-in vivo* system in which the spontaneously immortalized ERα-negative human breast epithelial cell (HBEC) line MCF-10F was transformed by treatment with $E_2$ (23). $E_2$-transformed cells progressively express phenotypes of *in vitro* cell transformation, including colony formation in agar methocel, decreased ductulogenesis, increased invasiveness in a Matrigel invasion system, and tumorigenesis in a heterologous host. Tumors formed in severe combined immunodeficient (SCID) mice by invasive cells and by cell lines derived from those tumors were poorly differentiated ERα-, progesterone receptor–, and ERBB2-negative adenocarcinomas (24). These characteristics are similar to those of basal cell type primary carcinomas previously described (25). To better understand the molecular events associated with the progressive phenotypic changes that were observed during estrogen-mediated malignant cell transformation, we performed Affymetrix 100k single nucleotide polymorphism (SNP) arrays to measure chromosomal copy number and loss of heterozygosity (LOH), and HG-U133_Plus_2 array for analyzing mRNA expression in MCF-10F cells at different stages of cell transformation. By integrating these data, we were able to identify associations between copy number changes, LOH, and tumorigenic phenotype, as well as the related changes in transcript expression. Functional analyses of these data identified several dysregulated pathways associated with progressive tumorigenic and invasive capacity.

## Materials and Methods

**Experimental model of transformation of MCF-10F cells by $E_2$ treatment.** These studies were performed using the spontaneously immortalized ERα-negative human breast epithelial cell line MCF-10F. As it was described in a previous publication, cells cultured and treated with 70 nmol/L $E_2$ were collected 24 h after the last treatment and maintained in culture for 10 additional passages (24). Thereafter, control and $E_2$-treated MCF-10F cells were evaluated for colony formation in agar-methocel, or colony efficiency, ductulogenic capacity in collagen-matrix, invasiveness in Matrigel invasion chambers, and tumorigenic assay in SCID mice, as previously described (24). $E_2$-treated MCF-10F cells formed colonies in agar methocel, exhibiting a high colony efficiency, and lost their ductulogenic capacity in collagen-matrix, forming instead spherical masses; cells exhibiting these characteristics were classified as transformed (trMCF). Briefly, as it was described, trMCF cells were seeded onto Matrigel invasion chambers and at the end of a 22-h incubation period, cells that had crossed the membrane were collected and identified as bsMCF cells (24). The tumorigenic ability of control MCF-10F, trMCF, and bsMCF was tested by injecting them into the mammary fat pad of 45-day-old female SCID mice (24). MCF-10F and trMCF cells did not induce tumors, whereas bsMCF formed tumors. From the tumors formed by the bsMCF cells, four cancer cell lines were derived and identified as caMCF1, caMCF2, caMCF3, and caMCF4 (these cells were previously called $E_2$-70-C5-A1-T1, $E_2$-70-C5-A4-T4, $E_2$-70-C5-A8-T8, and $E_2$-70-C5-A6-T6, respectively; ref. 24). The caMCF1, caMCF2, caMCF3, and caMCF4 were isolated from four tumors from four different animals (24).

**Ring cloning and tumorigenesis in mice.** The invasive bsMCF cells were plated at low density and cell colonies were isolated using cloning rings. The cells were cultured in DMEM:F12 medium containing 1.05 mmol/L calcium, antimycotics, hormones, growth factors, and equine serum (24). After trypsinization and plating, the clones obtained were identified as bcMCF-1, bcMCF-2, bcMCF-3, bcMCF-4, bcMCF-5, bcMCF-6, and bcMCF-7. The tumorigenic ability of the bcMCF subclones was tested by injecting them into the mammary fat pad of 45-day-old female SCID mice (see Supplementary Table S1).

**Affymetrix microarray expression and genotyping assay.** MCF-10F cells at three different passages, designated MCF-10F1, MCF-10F2, and MCF-10F3 (passages 135, 137, and 138, respectively); trMCF cells at three different passages, designated trMCF1, trMCF2, and trMCF3 (passages 20, 22, and 23, respectively); bcMCF clones 1, 2, and 3; and caMCF 1, 2, and 3 cells were used for microarray expression and genotyping assay (Fig. 1*A*). Total RNA and high molecular weight genomic DNA were isolated using Stat-60 (Tel-Test, Inc.) and the DNA isolation protocol previously described (24). After hybridization, the chips were scanned using GeneChip Scanner 3000. The SNP genotype calls (heterozygous or homozygous) were determined by GTYPE v 4.0. MCF-10F served as the diploid reference for detection of copy number changes by dChip. The human genome release v17 was used to generate the genome information for SNP data analysis. Hierarchical clustering of samples based on copy number was performed using the default parameters in dChip.

The intensities of probe sets were calculated by dChip with Perfect-match/Mismatch difference model after invariant-set normalization. Three lists of differentially expressed genes (dysregulated genes) were generated by pairwise comparison using MCF-10F cells as reference. The significance of differential expression was determined by the following combined criteria: (*a*) the gene in all three samples of at least one group in pairwise comparison is expressed (determined by GCOS "Present" call); (*b*) difference between the average value of the gene in the two compared groups $\geq 50$; (*c*) fold change $\geq 1.7$; (*d*) unpaired *t* test using log-transformed intensities, followed by Benjamini-Hochberg procedure controlling false discovery rate ($q \leq 0.08$). For probe sets representing the same Entrez Gene or UniGene ID, only the one with lowest $q$ value was included in the gene list.

**Identification of chromosomes and Gene Ontology categories enriched with differentially expressed genes.** The functional profiles were represented by the biological processes in the Gene Ontology (GO) database (26). The number of dysregulated genes in each chromosome or GO category was compared with that of all genes in the HG-U133_Plus_2 chip to determine the significance of the chromosome or GO category. The analysis was performed using Onto-Express,[3] with the default selection of statistical method (hypergeometric distribution followed by false discovery rate correction). The three lists of genes dysregulated in trMCF, bcMCF, or caMCF were uploaded into Onto-Express to identify significant GO categories ($q \leq 0.05$ with five or more genes). The up- or down-regulated genes were uploaded into Onto-Express separately to identify the individual chromosomes enriched ($q \leq 0.0001$) with these genes.

[3] http://www.geneontology.org/GO.tools.microarray.shtml#onto-e

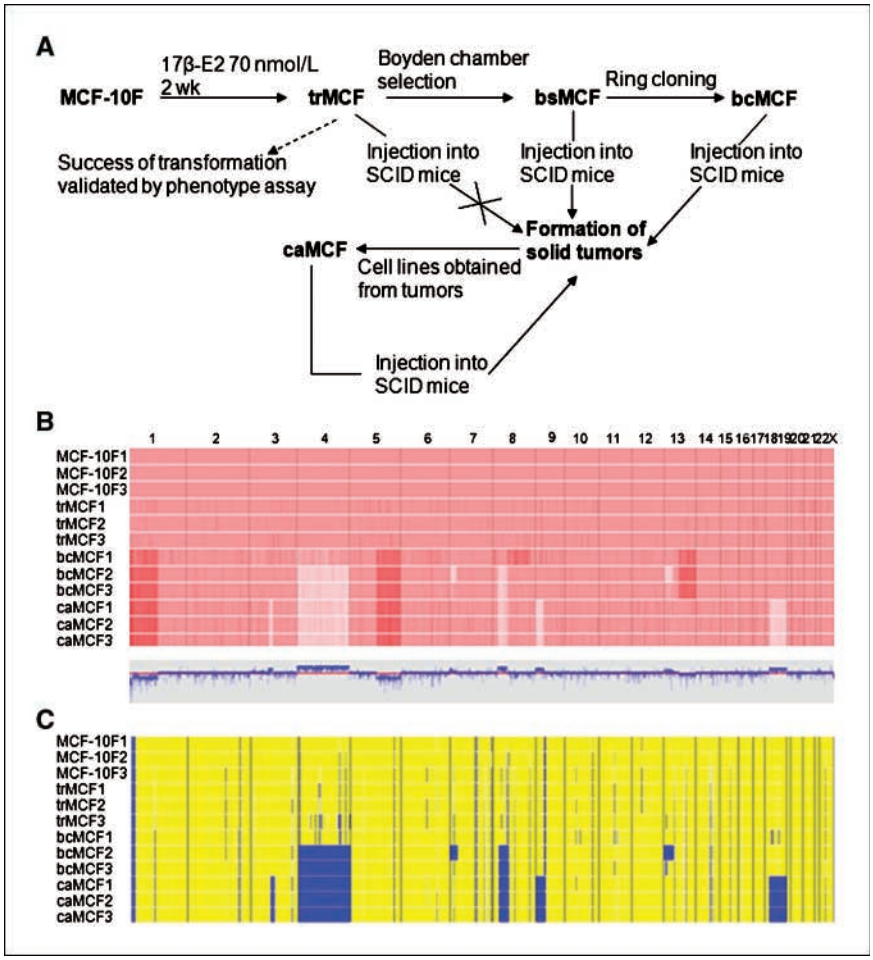


**Figure 1.** *A,* transformation of MCF-10F cells by $E_2$ treatment. Experimental protocol: MCF-10F cells treated with 70nmol/L $E_2$ that expressed high colony efficiency and loss of ductulogenic capacity in collagen-matrix were classified as transformed (*trMCF*). Transformed cells that were invasive in a Matrigel Boyden-type invasion chambers were selected (*bsMCF*) and plated at low density for cloning (*bcMCF*). MCF-10F, trMCF, bsMCF, and bcMCF were tested for carcinogenicity by injecting them into the mammary fat pad of 45-d-old female SCID mice. MCF-10F and trMCF cells did not induce tumors (*canceled arrow*); bsMCF and bcMCF formed solid tumors from which four cell lines, identified as caMCF, were derived and proven to be tumorigenic in SCID mice. *B* and *C,* chromosome copy number analysis using Affymetrix 100k SNP chips and dChip software. *B,* display of inferred copy number. MCF-10F at three different passages serves as diploid reference. *Pink shade,* diploidy; *darker red* and *lighter pink,* regions of copy number amplification and deletion, respectively. *Gray box,* range from 0 to 4 copies (*blue curve*); *red line,* baseline for diploidy. *C,* complete genome view of LOH: *yellow,* retention of heterozygosity; *blue,* LOH; *white,* no information due to lack of SNPs. Each column in *B* and *C* represent the different chromosomes; the chromosome number, from 1 to 22 and chromosome X are at the top of *B* and, at the left of these panels, the different cells are indicated.

**Ingenuity pathway analysis.** The differentially expressed genes were uploaded into the Ingenuity Pathway Analysis software.[4]

**Immunocytochemical analysis.** The cells were scraped into the medium, centrifuged, and the cell pellets were suspended in 10% phosphate-buffered formalin. After overnight fixation at 4°C, the pellets were embedded in 2% agarose, postfixed in 10% phosphate-buffered formalin, and embedded in paraffin. The following mouse monoclonal antibodies were used: E-cadherin, clone 36 (BD Transduction Laboratories), epithelial membrane antigen (EMA) clone E29 and vimentin, clone V9, both from DakoCytomation, Inc. All sections were lightly counterstained with hematoxylin.

**Expression of *CD24* and *CD44*.** The microarray expression levels of *CD24* and *CD44* were validated and further explored by reverse transcription-PCR (RT-PCR) and flow cytometry analysis [fluorescence-activated cell sorting (FACS)] that was performed at the Cell Sorting Facility of the Fox Chase Cancer Center. For RT-PCR, 40 to 80 ng of cDNA were used for PCR amplification in GeneAmp PCR System 9700 (Applied Biosystems). PCR primers were designed to span exons 2 to 18 of *CD44*, to characterize the expression of alternative splicing variants. The sequence of forward (F) and reverse (R) primers are as follows: *CD24*, F-GCCAGTCTCTTCGTGGTCTC; R-CTCCATTCCACAATCCCATC; *CD44*, F-GAGCATCGGATTTGAGACCTG; R-AGCTCCATTGCCACTGTTGAT; β-actin, F-ACCCACACTGTGCCCATCTACGA; R-AGCTGGAAGCAGCCGTGGCCAT. The density of the bands was quantified using a scanner (Expression 836XL, EPSON) and measurement software (Image J, W.S. Rasband, NIH, MD).[5]

FACS was performed using FITC mouse anti-human CD24 and phycoerythrin mouse anti-human CD44 antibodies from BD Biosciences. Cells ($10^6$) were stained and analyzed by FACS using a LSR II system (BD Biosciences). These experiments were repeated twice.

## Results

**Analysis of chromosome copy number and LOH in neoplastically transformed MCF-10F cells.** MCF-10F cells that after treatment with $E_2$ expressed high colony efficiency and loss of ductulogenic capacity in collagen-matrix represented the first level of *in vitro* transformation. Cells expressing these two variables were classified as transformed (trMCF), which after further selection for invasiveness in a Matrigel invasion chamber originated the second level of transformation: the invasive (bsMCF) and the cloned (bcMCF) cells (Fig. 1*A*). The bsMCF cells formed tumors in SCID mice from which four cell lines, caMCF, were derived (Fig. 1*A*). By ring cloning, seven subclones were isolated from the invasive bsMCF cells: bcMCF-1, bcMCF-2, bcMCF-3, bcMCF-4, bcMCF-5, bcMCF-6, and bcMCF-7. All the bcMCF subclones produced invasive poorly differentiated tumors in SCID mice with different morphologic phenotypes: spindle cell type (bcMCF-1 and bcMCF-4), epithelial cell type (bcMCF-2, bcMCF-6, and bcMCF-7), and with mix features of spindle and epithelial type (bcMCF-3 and bcMCF-5; Supplementary Table S1). As it was previously reported, MCF-10F cells were seeded on Boyden chamber as control; cells that passed through the membrane were selected, expanded, and injected in SCID mice; these cells did not produce tumors (24).

[4] http://www.ingenuity.com
[5] http://rsb.info.nih.gov/ij

Using the 100k SNP GeneChip Mapping Array set, the DNA from MCF-10F, trMCF, bcMCF, and caMCF were analyzed for the structure of chromosomes 1 to 22 and X at very high resolution (Fig. 1*B* and *C*). Changes in copy number (Fig. 1*B*) and LOH (Fig. 1*C*) were progressive in these cells. The three different passages of MCF-10F exhibited identical copy number, and these data served as the reference for copy number analysis. Only small and scattered chromosome gains were observed in trMCF cells and most alterations were observed in the bcMCF and caMCF (Fig. 1*B*). Large fragments of chromosome gain were observed only in the telomeric ends at 1p36.12-pter and 5q21.1-qter in both bcMCF and caMCF, and 13q21.31-qter in bcMCF (Fig. 1*B, dark pink areas*). Almost no chromosome loss was detected in the trMCF cells and very few in bcMCF1 clone, whereas bcMCF2 and bcMCF3 clones and the three caMCF cell lines exhibited loss of the whole chromosome 4 and loss of 8p11.21-23.1 (Fig. 1*B, pale pink areas*). Loss of 3p12.1-14.1, 9p22.1-pter, and 18q11.2-qter was observed in the three lines of caMCF (Fig. 1*B*). Deletion was also observed in 7pter and 13pter of bcMCF2, but not in the rest of tumorigenic cells, indicating that such deletion is not required for the expression of tumorigenic capacity.

The subclone bcMCF1 displayed higher frequency of chromosome 8 amplification (Supplementary Fig. S1*A*) and lower frequency of chromosome 4 deletion when compared with the subclones bcMCF2 and bcMCF3 (Supplementary Fig. S1*B*). Also, subclones bcMCF2 and bcMCF3 showed a small deletion on chromosome 8 that bcMCF1 did not have (Fig. 1*B, pale pink area in bcMCF2 and bcMCF3*). However, sample clustering based on copy number profile grouped the bcMCF1 in the tumorigenic class together with bcMCF2; bcMCF3; and caMCF1, caMCF2, and caMCF3 (Supplementary Fig. S1*C*), and this was confirmed by sample clustering based on gene expression profile (see below EMT phenotype). The cells trMCF1, trMCF2, and trMCF3 were clustered as nontumorigenic (Supplementary Fig. S1*C*). Furthermore, all bcMCF clones were shown to form tumors in SCID mice (Supplementary Table S1). Therefore, the isolation of subclones may afford the possibility to detect minimal regions of copy number change associated with neoplastic cell transformation.

This cell transformation model involves treatment of immortal MCF-10F cells and selection for a transformed phenotype followed by further selection for invasiveness and then tumorigenesis (Fig. 1*A*). As major copy number changes and LOH were observed in chromosome 4 (Fig. 1*B* and *C*), we present this chromosome at higher resolution. Interestingly, the copy number loss in chromosome 4 detected in bcMCF can be observed in trMCF cells (Supplementary Fig. S2, *arrows*), albeit at a level that does not reach statistical significance.



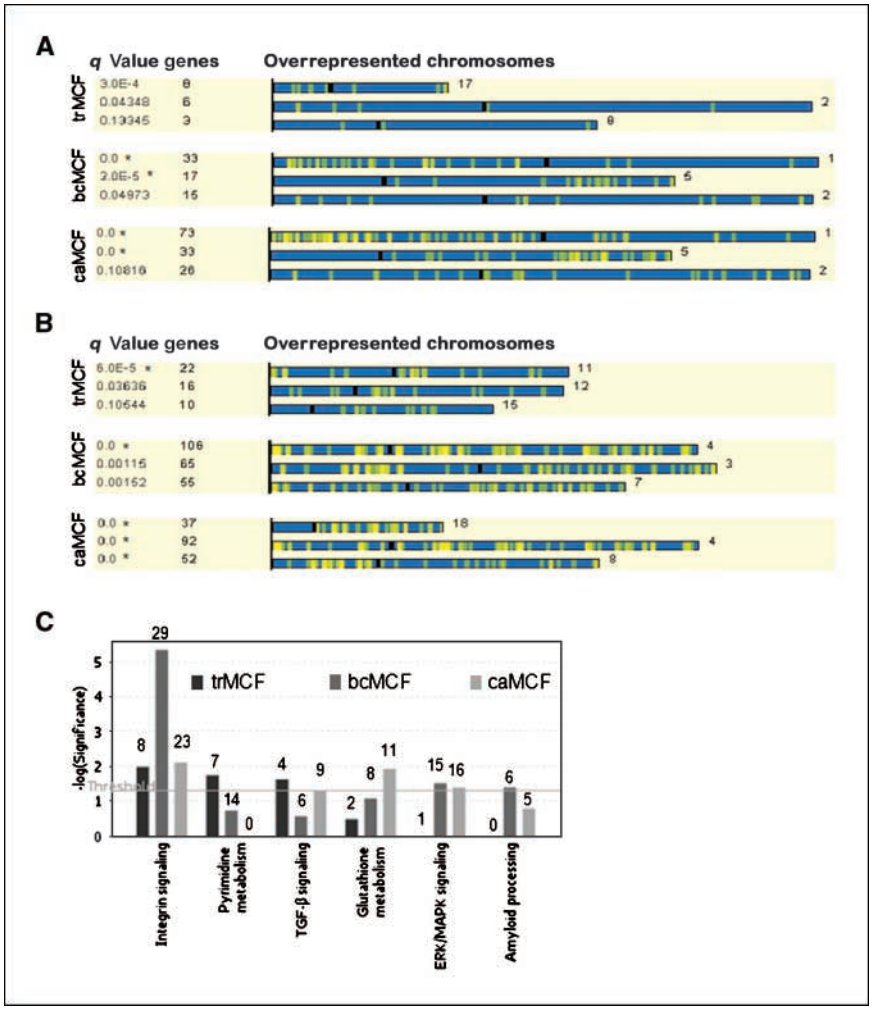


**Figure 2.** *A* and *B,* chromosomes enriched with differentially expressed genes. The top three chromosomes (ranked by *q* value) in trMCF, bcMCF, and caMCF cell lines enriched with up-regulated (*A*) or down-regulated (*B*) genes. *, the chromosomes overrepresented ($q \leq 0.0001$) by up- or down-regulated genes. A *q* value of 0.0 indicates a value smaller than $10^{-5}$. *C,* canonical pathways altered in progressive malignant cell transformation. The pathways significantly enriched with the dysregulated genes are identified by Ingenuity Pathway Analysis. The number of differentially expressed genes in each pathway is labeled above the corresponding cell. The genes identifying the integrin signaling pathway are displayed in Table 2; those in the other five canonical pathways are displayed in Supplementary Table S3. *ERK/MAPK,* extracellular signal-regulated kinase/mitogen-activated protein kinase.

**Table 1.** Biological processes enriched with differentially expressed genes in tumorigenic and nontumorigenic cells

| GO* ID | Function name | trMCF | | bcMCF | | caMCF | |
|---|---|---|---|---|---|---|---|
| | | Genes | *q* | Genes | *q* | Genes | *q* |
| In all tumorigenic and nontumorigenic cell lines | | | | | | | |
| GO:0007155 | Cell adhesion | 14 | 6.4E−04 | 43 | 3.7E−04 | 40 | 5.9E−03 |
| GO:0008283 | Cell proliferation | 10 | 1.5E−03 | 31 | 9.5E−04 | 42 | 4.5E−08 |
| GO:0007267 | Cell-cell signaling | 5 | 5.0E−02 | 26 | 5.4E−03 | 32 | 1.0E−04 |
| GO:0008544 | Epidermis development | 7 | 6.3E−06 | 20 | 2.1E−09 | 19 | 2.2E−09 |
| GO:0006955 | Immune response | 11 | 6.1E−04 | 26 | 1.5E−02 | 30 | 4.0E−03 |
| GO:0008152 | Metabolism | 10 | 8.8E−03 | 36 | 6.6E−03 | 35 | 1.5E−02 |
| GO:0006508 | Proteolysis and peptidolysis | 12 | 2.3E−03 | 32 | 4.5E−02 | 40 | 3.8E−03 |
| GO:0000074 | Regulation of cell cycle | 11 | 3.4E−04 | 23 | 2.3E−02 | 28 | 2.9E−03 |
| Only in nontumorigenic cell lines | | | | | | | |
| GO:0006260 | DNA replication | 6 | 1.4E−03 | 1 | NS | 6 | NS |
| GO:0006334 | Nucleosome assembly | 5 | 5.4E−04 | 2 | NS | 3 | NS |
| Only in tumorigenic cell lines | | | | | | | |
| GO:0006915 | Apoptosis | 4 | NS | 28 | 5.0E−03 | 27 | 1.7E−02 |
| GO:0005975 | Carbohydrate metabolism | 2 | NS | 20 | 2.0E−02 | 24 | 3.5E−03 |
| GO:0007166 | Cell surface receptor linked signal transduction | 2 | NS | 14 | 4.2E−02 | 15 | 3.0E−02 |
| GO:0006935 | Chemotaxis | 2 | NS | 11 | 1.6E−02 | 11 | 2.1E−02 |
| GO:0006897 | Endocytosis | 0 | NS | 11 | 2.6E−02 | 11 | 2.7E−02 |
| GO:0006629 | Lipid metabolism | 1 | NS | 21 | 1.1E−02 | 21 | 1.8E−02 |
| GO:0043123 | Positive regulation of IKB/NFKB cascade | 1 | NS | 16 | 5.0E−04 | 13 | 1.5E−02 |
| GO:0019538 | Protein metabolism | 0 | NS | 5 | 1.2E−02 | 5 | 1.4E−02 |
| GO:0006986 | Response to unfolded protein | 1 | NS | 9 | 4.7E−03 | 9 | 5.8E−03 |
| GO:0016192 | Vesicle-mediated transport | 1 | NS | 10 | 6.6E−04 | 12 | 1.2E−03 |
| GO:0007169 | Transmembrane receptor protein tyrosine kinase signaling pathway | 2 | NS | 12 | 4.6E−04 | 12 | 1.5E−02 |

Abbreviation: NS, nonsignificant GO category.
*GO categories were identified by Gene Ontology tool, Onto-Express.

**Chromosomes enriched with the dysregulated genes.** Three gene lists were generated by pairwise comparison of trMCF, bcMCF, or caMCF with the reference MCF-10F.[6] The highest number of total dysregulated genes was observed in caMCF cells (1,306 genes, 340 up-regulated versus 966 down-regulated), slightly lower in bcMCF (1,236 genes, 160 up-regulated versus 1,076 down-regulated), and the lowest in trMCF cells (260 genes, 45 up-regulated versus 215 down-regulated). This order was consistent with the extent of chromosome aberrations: number of regions and size in caMCF > bcMCF >>> trMCF (Fig. 1*B*).

To further investigate the relationship between transcript levels and copy number, we analyzed the list of up- and down-regulated genes using Onto-Express software, to identify chromosomes significantly enriched (or overrepresented) with dysregulated genes. The chromosomes enriched with up-regulated (Fig. 2*A*) or down-regulated (Fig. 2*B*) genes also showed large regions of amplification or deletion (Fig. 1*B*), respectively. For example, chromosomes 18, 4, and 8 were significantly overrepresented by down-regulated genes in caMCF cells (Fig. 2*B, caMCF*), and, correspondently, showed large regions of deletion in caMCF (Fig. 1*B*). Similarly, chromosome 4 was overrepresented by down-regulated genes in bcMCF cells (Fig. 2*B, bcMCF*), and showed the largest copy number loss in these cells (Fig. 1*B*). Chromosomes 1 and 5 were significantly overrepresented by up-regulated genes in both bcMCF and caMCF cells (Fig. 2*A, bcMCF* and *caMCF*), and, correspondently, showed largest copy number amplification (Fig. 1*B*). However, some exceptions were noteworthy. Chromosome 13 showed a high frequency of SNP amplification (50%) in bcMCF (Supplementary Fig. S1*A*), but was not enriched with up-regulated genes. In fact, there were only six genes up-regulated in this region. Moreover, this amplification was not maintained in caMCF (Fig. 1*B*), indicating that this aberration was not required for tumorigenesis. Also, chromosome 11 was overrepresented by down-regulated genes in trMCF (Fig. 2*B, trMCF*), although trMCF cells did not show any copy number deletion (Fig. 1*B*). This might imply that certain epigenetic modifications, instead of copy number loss, are related to the observed decrease in the number of genes expressed on chromosome 11 of trMCF cells.

**Biological processes enriched with the dysregulated genes.** GO analysis identified eight biological processes enriched with dysregulated genes in trMCF, bcMCF, and caMCF cells, suggesting that changes in these processes were involved in the *in vitro* transformation phenotype (Table 1). Two additional processes, DNA replication and nucleosome assembly, were found only in the trMCF cells. The DNA replication process in trMCF cells contained six genes. Each of these genes was up-regulated, including *replication factor C*, 2.1-fold; *SET translocation*, 1.8-fold; *DNA directed polymerase epsilon 2*, 2.7-fold; *ribonucleotide reductase M2 polypeptide*, 1.7-fold; *minichromosome maintenance deficient 4*, 1.7-fold; and *topoisomerase (DNA) IIα*, 1.8-fold. Moreover, 11 biological processes were found uniquely in the tumorigenic cells bcMCF and

[6] http://feinstone.memphis.edu/data/Estrogen

caMCF, suggesting that they were involved in the process of neoplastic transformation (Table 1). Most of the genes in the 11 GO categories showed decreased expression relative to MCF-10F. For example, the fraction of down-regulated genes was 24 of 28 in "apoptosis" (Supplementary Table S2), 13 of 16 in "positive regulation of IKB/NFKB cascade," and 19 of 21 in "lipid metabolism" (data not shown).

**Identification of significant canonical pathways.** Ingenuity Pathway Analysis revealed six canonical pathways significantly dysregulated in one or more groups (Fig. 2*C*). The genes associated with each of these pathways, integrin signaling, pyrimidine metabolism, transforming growth factor β (TGF-β) signaling, glutathione metabolism, extracellular signal-regulated kinase/mitogen-activated protein kinase signaling, and amyloid processing are listed in Table 2 and Supplementary Table S3.

Integrin signaling was the most significantly altered pathway in each of the three cell lines, indicating a continuum change within this pathway throughout the progressive, malignant cell transformation of MCF-10F cells. In this pathway, eight genes were down-regulated in trMCF cells (Table 2). Of interest, expression levels of several of these decreased genes, such as *ITGB6, LAMA3*, and *LAMC2*, were inhibited to even a greater extent in the tumorigenic cells. In contrast, *fibronectin 1*, which was completely suppressed (−18.9 fold, "Absent") in trMCF, was strongly induced in bcMCF (4.9-fold) and caMCF (2.8-fold; Fig. 3*A, FN1*). In comparison with the trMCF, the number of dysregulated genes associated with integrin signaling was much higher in the bcMCF and caMCF cells, showing both increased and decreased levels of expression (Table 2).

Levels of expression of several genes involved in glutathione metabolism were decreased in the bcMCF and caMCF

**Table 2.** Dysregulated genes involved in integrin signaling pathway

| Symbol | Gene name* | Fold change | | |
|---|---|---|---|---|
| | | trMCF | bcMCF | caMCF |
| *ARF3* | *ADP-ribosylation factor 3* | NS | −1.7 | NS |
| *ARF5* | *ADP-ribosylation factor 5* | NS | −2.5 | −2.4 |
| *ARF6* | *ADP-ribosylation factor 6* | NS | NS | 1.8 |
| *ARHGAP5* | *Rho GTPase activating protein 5* | NS | −1.8 | NS |
| *CAPN1* | *Calpain 1 (mu/I) large subunit* | NS | −2.6 | −2.4 |
| *CAV1* | *Caveolin 1* | NS | 1.8 | NS |
| *CDC42* | *Cell division cycle 42* | NS | 1.9 | 2.5 |
| *COL1A1* | *Collagen type I, α I* | −6.1 | −16.8 | NS |
| *DDEF1* | *Develop and differentiation enhancing factor 1* | NS | 2.2 | 2.2 |
| *FN1* | *Fibronectin 1* | −18.9 | 4.9 | NS |
| *FYN* | *FYN oncogene related to SRC, FGR, YES* | NS | 2.2 | NS |
| *GSK3B* | *Glycogen synthesis kinase 3 β* | NS | −2.2 | −1.9 |
| *HRAS* | *v-Ha-ras Harvey rat sarcoma viral oncogene* | NS | 6.0 | 7.4 |
| *ITGA6* | *Integrin $\alpha_6$* | NS | −1.8 | −1.8 |
| *ITGAV* | *Integrin $\alpha_V$* | −1.8 | NS | −2.5 |
| *ITGB1* | *Integrin $\beta_1$* | NS | NS | −2.0 |
| *ITGB4* | *Integrin $\beta_4$* | −1.9 | −3.9 | −3.7 |
| *ITGB6* | *Integrin $\beta_6$* | −2.6 | −50.4 | −58.5 |
| *KRAS* | *v-Ki-ras2 Kirsten rat sarcoma viral oncogene* | NS | −2.7 | NS |
| *LAMA3* | *Laminin α3* | −3.5 | −100.5 | −99.8 |
| *LAMA5* | *Laminin α5* | NS | NS | −1.7 |
| *LAMB3* | *Laminin β3* | −2.7 | −3.6 | −4.4 |
| *LAMC2* | *Laminin γ2* | −3.6 | −8.9 | −26.2 |
| *MYLK* | *Myosin, light polypetide kinase* | NS | NS | 4.4 |
| *NCK2* | *NCK adaptor protein 2* | NS | −2.3 | NS |
| *RAC1* | *ras-related C3 botulinum toxin substrate 1* | NS | −1.9 | NS |
| *RAC2* | *ras-related C3 botulinum toxin substrate 2* | NS | −1.9 | NS |
| *RALA* | *v-ral simian leukemia viral oncogene A* | NS | −2.0 | NS |
| *RALB* | *v-ral simian leukemia viral oncogene B* | NS | −1.7 | NS |
| *RHOB* | *ras homologue gene family, member B* | NS | 2.3 | NS |
| *RHOD* | *ras homologue gene family, member D* | NS | −3.2 | −4.2 |
| *RHOF* | *ras homologue gene family, member F* | NS | −3.8 | −4.3 |
| *TM4SF6* | *Tetraspanin 6* | NS | −2.5 | −1.8 |
| *TM4SF7* | *Tetraspanin 4* | NS | 1.7 | NS |
| *TM4SF8* | *Tetraspanin 3* | NS | NS | −1.9 |
| *TM4SF9* | *Tetraspanin 5* | NS | 6.5 | 7.1 |

Abbreviation: NS, nonsignificant gene.
*Genes were identified by Ingenuity Pathway Analysis (see Fig. 2*C*).

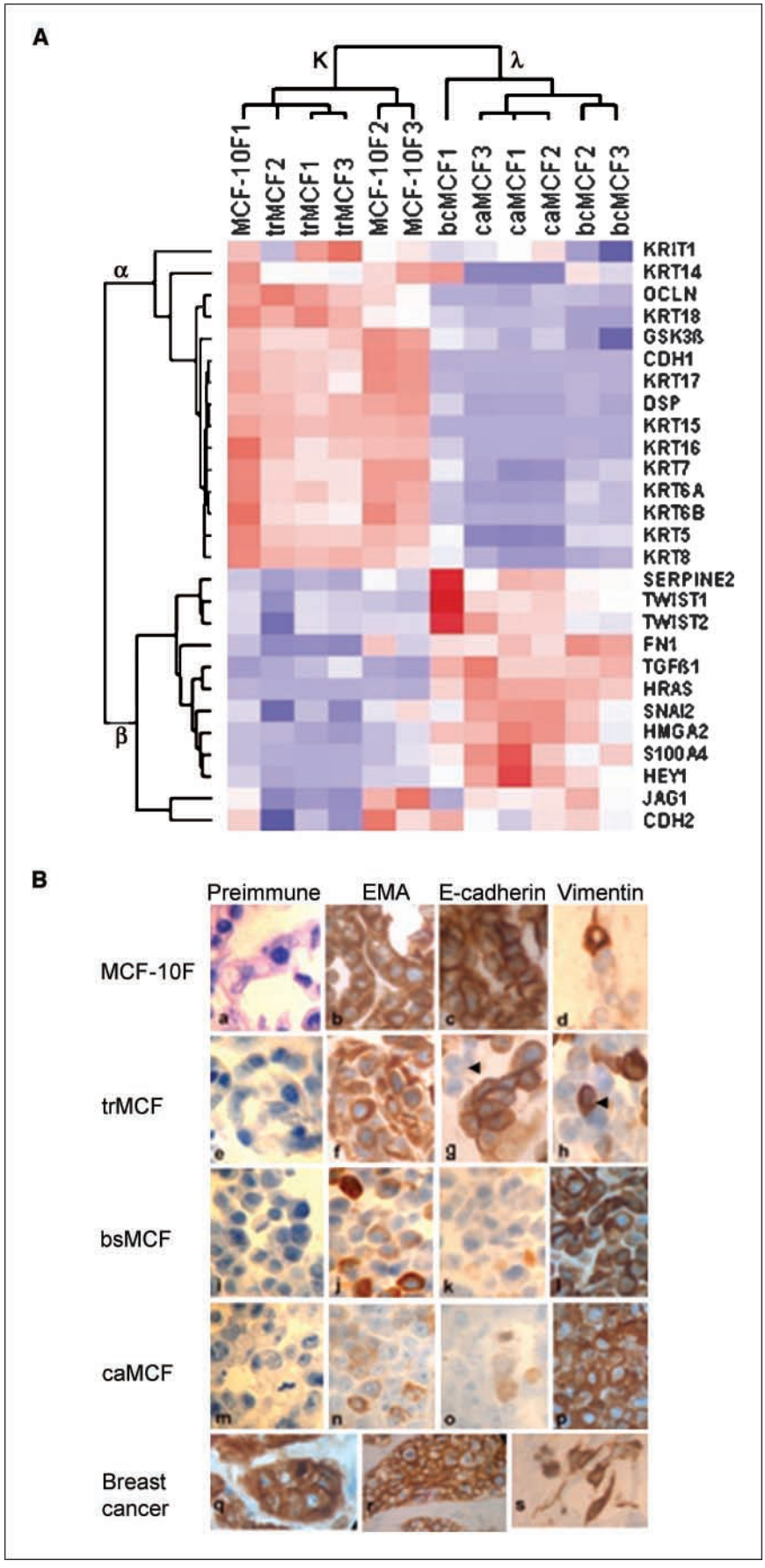


**Figure 3.** Expression profile of EMT markers and their regulators during malignant cell transformation. *A,* a list of EMT markers and promoting genes was generated a priori by literature search (Supplementary Table S5). Hierarchical clustering of cell lines and genes was performed using dChip software. Two sample clusters ($\kappa$ and $\lambda$) and two gene clusters ($\alpha$ and $\beta$) were identified. *Red, white*, and *blue,* level above, at, and below mean expression, respectively. *B,* detection of epithelial and mesenchymal markers by immunocytochemistry. *a,* histologic sections of MCF-10F cells, reacted with preimmune mouse serum, were used as the negative control (×100); *b,* MCF-10F reacted for EMA (×100); *c,* MCF-10F reacted for E-cadherin (×100); *d,* MCF-10F reacted for vimentin (×100); *e,* trMCF cells reacted with preimmune mouse serum used as negative control (×100); *f, g*, and *h,* trMCF cells reacted for EMA, E-cadherin, and vimentin, respectively (×100); *i,* bsMCF cells reacted with preimmune mouse serum as a negative control (×100); *j, k, l,* bsMCF cells reacted for EMA, E-cadherin, and vimentin, respectively (×100); *m,* caMCF tumor cell line cells reacted with preimmune mouse serum used as negative control (×100); *n, o, p,* caMCF tumor cell lines reacted for EMA, E-cadherin, and vimentin, respectively (×100); *q* and *r,* invasive ductal carcinoma of the breast as positive control and immunoreacted for EMA and E-cadherin, respectively (×100); *s,* histologic section of an invasive adenocarcinoma immunoreacted for vimentin (×100).

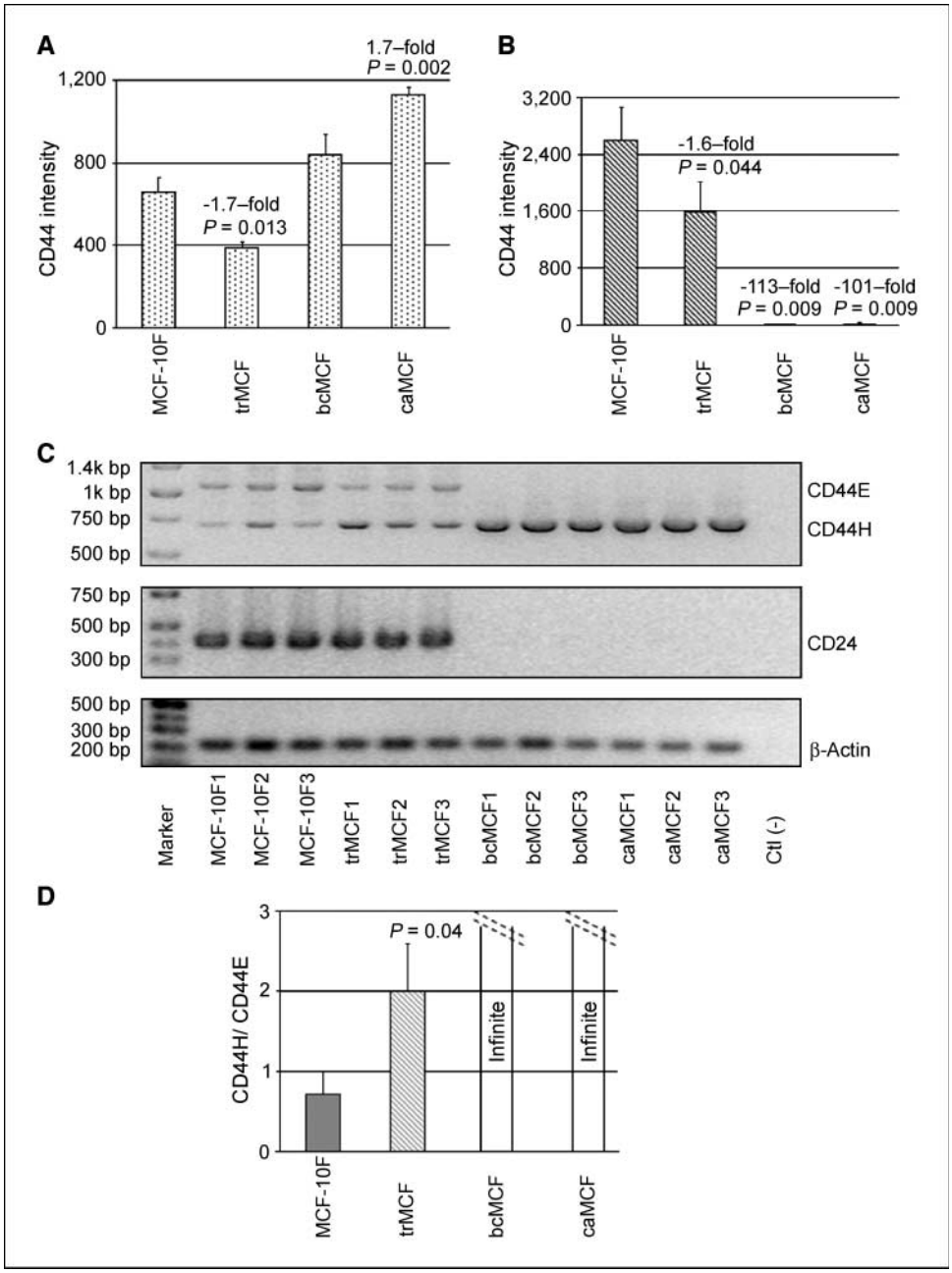


**Figure 4.** Characterization of the expression of *CD44* and *CD24*. *A* and *B*, microarray expression values of *CD44* and *CD24*, respectively. Significance: $P < 0.05$ in unpaired *t* test compared with MCF-10F. *C*, RT-PCR validation for *CD44* (*top*) and *CD24* (*middle*) expression levels. *D*, densitometry ratio of CD44H to CD44E detected in RT-PCR.

(Supplementary Table S3). Multiple glutathione *S*-transferase (*GST*) isoforms were down-regulated, including *GST κ1, ω2, pi, MGST 2*, and *MGST3*. Moreover, glutamate cysteine ligase, the enzyme catalyzing the first rate-limiting reaction in glutathione biosynthesis, was also reduced in the tumorigenic cells. Glutathione is a cellular antioxidant and a substrate for GST, which catalyzes the conjugation reactions with electrophiles. Repressed antioxidant activity is associated with the genomic damage and cancer incidence induced by exposure to the cytotoxic free radical and reactive oxygen species (27).

**EMT phenotype.** Analysis of the GO cellular component also identified categories distinct between the nontumorigenic and tumorigenic cells. The "intermediate filament" component was found in the bcMCF ($q = 2.0 \times 10^{-5}$, 12 genes) and caMCF cells ($q = 0.0$, 13 genes), but not in trMCF cells ($q = 0.4$, 1 gene; Supplementary Table S4). Numerous cytokeratins were suppressed or absent, whereas *vimentin* was strongly induced in bcMCF (7.0-fold) and caMCF (8.1-fold). Because of these findings, we generated a gene list from published literature for epithelial and mesenchymal markers and their regulators (Supplementary Table S5). The 52 genes in the list were filtered by low stringency criteria of combined coefficient of variation (CV) >0.3 and "Present calls" in >30% of the samples. The 27 genes passing these criteria were used for sample and gene clustering (Fig. 3*A*). Two sample groups and two gene groups were identified. The nontumorigenic MCF-10F and trMCF cells were grouped into sample cluster κ, whereas the tumorigenic bcMCF and caMCF cells were grouped into cluster λ. On the other side, the genes were grouped into

clusters α and β based on their expression pattern. The epithelial markers *E-cadherin* (*CDH1*), *occludin* (*OCLN*), *desmoplakin* (*DSP*), and *cytokertins* (*KRT*) were decreased, whereas the mesenchymal markers *fibronectin* (*FN1*), *vimentin* (*VIM*), and *N-cadherin* (*CDH2*) were increased in bcMCF and caMCF cells (Fig. 3). By real-time RT-PCR, it was confirmed that the expression of FN1, S100A4, SNAI2, HRAS, and TGFβ1 were increased, whereas CDH1 (E-cadherin) was decreased in bcMCF and caMCF cells (Supplementary Fig. S3). Immunocytochemical analysis using antibodies against EMA (also called MUC1) and E-cadherin displayed significant loss of these epithelial markers (Fig. 3*B, j, k, n, o*) and increased expression of the mesenchymal marker vimentin (Fig. 3*B, l, p*) in tumorigenic cells. These findings confirmed the EMT phenotype revealed by gene expression profile in Fig. 3*A*. Interestingly, whereas most trMCF cells were E-cadherin positive (+) and vimentin negative (−), a few cells were found to be E-cadherin (−) (Fig. 3*B, g, arrow*) and vimentin (+) (Fig. 3*B, h, arrow*).

**Characterization of molecular markers of tumorigenic cells.** The cell surface molecule $CD44^{+}/CD24^{-/low}$ phenotype has been identified as marker for tumor-initiating cells (28). Therefore, we examined the expression pattern of these markers in relation to the tumorigenic capacity of the cells in our model. Microarray data revealed that *CD44* was slightly decreased in trMCF but increased in bcMCF and caMCF cells, whereas *CD24* was completely lost in both bcMCF and caMCF (Figs. 4*A* and *B*). The loss of *CD24* in bcMCF and caMCF was independently shown by RT-PCR (Fig. 4*C, middle*). The alternative splicing variants of *CD44* were identified by the size of their PCR products and further validated by DNA sequence analysis (data not shown). Only the variants of *CD44H* and *CD44E* were expressed (Fig. 4*C, top*). *CD44H* was significantly increased in both bcMCF and caMCF, whereas *CD44E* was completely lost in these cells. The ratio of *CD44H/CD44E* was significantly increased in trMCF, bcMCF, and caMCF (Fig. 4*D*). The *CD44* and *CD24* gene expression in the different cell lines correlated well with the cell surface protein expression studied by FACS analysis (Supplementary Fig. S4). In addition, the *ESA* (epithelial-specific antigen, also termed as *Ep-CAM*, or tumor-associated calcium signal transducer 1), was completely lost in bcMCF and caMCF, but not changed in trMCF cells (see gene list online[6]).

## Discussion

This study integrates structural and functional genomic data analyses to elucidate the progressive molecular events in the $E_2$-mediated malignant transformation of ER (−) HBECs. Genomic aberrations progressively accumulated as the cells expressed more aggressive phenotypes (i.e., in the tumorigenic bcMCF and caMCF) in comparison with the nontumorigenic trMCF cells. Accordingly, the number of genes with altered levels of expression was greater in the tumorigenic cells, as where the chromosomes enriched with up- or down-regulated genes. Importantly, the 12 samples were correctly classified into tumorigenic or nontumorigenic groups based on the profile of copy number changes, indicating that in our model changes in copy number provide a genomic signature for the tumorigenic phenotype. Together, these findings revealed an intrinsic link between $E_2$-induced copy number changes, gene expression alterations, and tumorigenesis in ERα (−) HBECs.

The integrin signaling pathway was the most significantly altered pathway in the progression of the neoplastic transformation. Integrins function as heterodimeric receptors for extracellular matrix proteins, mediating cell anchorage. The capacity of cells to survive and proliferate in the absence of integrin-mediated adhesion *in vitro* strongly correlates with tumorigenesis *in vivo* (29). The integrin signaling pathway was enriched with dysregulated genes in trMCF, bcMCF, and caMCF, indicating that this pathway was affected in early stages of cell transformation. In addition, GO analysis revealed enrichment of dysregulated genes in the apoptotic process in tumorigenic cells (Supplementary Table S2) and resistance to apoptosis is a hallmark of tumorigenesis. Therefore, suppression of apoptosis in the tumorigenic cells might be a potential mechanism that confers survival onto these cells with disrupted integrin signaling and anchorage-independent growth (24).

EMT involves dedifferentiation of polarized epithelial cells to a migratory fibroblastoid phenotype, a phenomenon that is increasingly considered to be an important event during cancer progression and metastasis (30, 31). Cells selected from MCF-10F by Boyden chamber are nontumorigenic (24), whereas all bcMCF cells displayed consistent chromosomal aberrations and formed tumors in SCID mice in our model. Although it cannot be totally ruled out, it is very unlikely that bcMCF cells are derived from preexisting mesenchymal cells.

EMT is accompanied by a profoundly altered mesenchymal gene expression program, which is characterized by loss of epithelial keratins and induction of mesenchymal vimentin (32). Induction of *S100A4* is an important early event in the pathway toward EMT (33). The hallmark of EMT is the loss of expression of the cell adhesion molecule *E-cadherin* (34). E-cadherin is a cell-cell adhesion molecule that participates in homotypic, calcium-dependent interactions to form epithelial adherens junctions. This function is critical in the development and maintenance of a polar epithelium. GSK3β was down-regulated in bcMCF and caMCF (Fig. 3*A*). SNAI2 was down-regulated in trMCF but increased in bcMCF and caMCF (Fig. 3*A*). It was shown that inhibition of GSK3β resulted in the up-regulation of SNAI1 and down-regulation of E-cadherin *in vivo*. SNAI2 (or SLUG) and SNAI1 (or SNAIL) belong to the Snail family of proteins; both contain an $NH_2$-terminal repression domain and a COOH-terminal zinc-finger DNA-binding domain. Snail proteins repress the transcription of E-cadherin (35–39). E-cadherin loss is believed to contribute to both cancer development and progression (35). HRAS and TGFβ1 were up-regulated in bcMCF and caMCF (Supplementary Table S3). It has been shown that HRAS cooperates with TGFβ1 to cause EMT and also it interacts with CD44 directly by increasing its expression (40, 41). TGFβ1 via HMGA2, which was also up-regulated in bcMCF and caMCF, regulates the expression of TWIST, SNAI1, and SNAI2 (42). EMT is associated with higher tumor grade, high motility index, and ERα (−) status (43). Therefore, these findings revealed an intrinsic link between EMT and tumorigenic capacity in our model, which, in part, may explain the poor prognosis of ERα (−) human breast carcinomas.

$CD44^{+}/CD24^{-/low}$ is the cell surface marker of tumorigenic breast cancer cells in which the tumorigenic capacity is further increased by additional expression of *ESA* (28). These cells are characterized by a 186-gene "invasiveness" gene signature that is associated with risk of death and metastasis in breast cancer (44). *CD24* encodes a small, heavily glycosylated cell-surface adhesion protein. *CD44* undergoes extensive alternative splicing within its central region spanning exon 6a to 14, also termed as variable exon v1 to v10. The two variants expressed in our model are mRNA precursor variant 3 with exon 6a to 11 spliced out and variant 4 with exon 6a to 14 spliced out, corresponding to *CD44E* and

*CD44H*, respectively (45). The *CD44H* is mainly expressed on cells of lymphohematopoietic origin; it plays an important role in cell adhesion and its expression promotes tumor cell migration (46). *CD44E* is preferentially expressed on epithelial cells and it is involved in the recognition of a common determinant in *CD44H* and *CD44E* promoting homotypic cellular aggregation (47). Microarray and RT-PCR analysis of *CD44* and *CD24* revealed an expression pattern of $CD44H^{high}/CD44E^{-}/CD24^{-}$ in bcMCF and caMCF. The increased ratio of *CD44H/CD44E* in trMCF cells might represent an early marker for $E_2$-transformed HBECs. The significant increase of CD44H and complete loss of CD44E might be a novel phenotype associated with the tumorigenic capacity. In addition, the loss of *ESA* in bcMCF and caMCF indicated that *ESA* expression is not required for the tumorigenic capacity in our model.

Two recent studies, using a GeneChip containing 1495 SNPs (48) or comparative genomic hybridization (49), have shown that LOH and allelic loss in 4p and 5q occur more frequently in subtypes of breast cancer characterized as ERα (−). SNP mapping reveals that LOH in 4p14-15.3, 5q11-32, and 18q22-23 are significantly associated with a gene expression profile in basal-like subtype (48). In our model, the tumor cells caMCF also showed LOH in all these three regions. Moreover, it has been shown that the $CD44^{+}/CD24^{-}$ phenotype is associated with mesenchymal phenotype and invasion in breast cancer cell line, and may define breast cancers of basal/myoepithelial origin rather than luminal origin (50). These findings indicate a potential correlation of our model with the basal-like subtype. However, the bcMCF and caMCF cells displayed low or absent expression of both luminal and basal cytokeratins in microarray analysis and ERBB2 (−) in immunocytochemical staining (data not shown). Therefore, they cannot simply be classified into basal-like, ERBB2 (+), luminal A or B, or normal breast-like subtype (20). Carey et al. (51) has identified an "unclassified" subtype using immunohistochemical markers for the classification of breast cancer tissue. This subtype is characterized as ERα (−), ERBB2 (−), and progesterone receptor (−), which, for these markers, is the same phenotype as the basal-like subtype. Unlike the basal-like subtype, the unclassified subtype is cytokeratin 5 (−). It also displays a histologic grade and survival prognosis most close to that of basal-like subtype (51). Further comparison of the tumorigenic cells in our model to clinical basal-like or unclassified subtype is warranted.

Our results support the concept that $E_2$-induced breast cancer is a polygenic disease having a large range of genomic instabilities. $E_2$ and/or its metabolites can directly cause genomic aberrations without the mediation of ERα. The genomic aberrations lead to changes in gene expression, which result in disrupted integrin signaling and apoptotic pathways, and epithelial to mesenchymal transition. These functional changes lead to colony formation in agar-methocel, loss of ductulogenesis in collagen matrix, invasiveness *in vitro*, and tumor growth in SCID mice *in vivo* (24). However, MCF-10F is a spontaneously immortalized cell line harvested from a woman's breast that was free of malignancies, but had a diagnosis of benign fibrocystic disease (52). Hence, we cannot rule out its inherited susceptibility to estrogen and the disposition to tumorigenesis. Therefore, more normal primary human breast epithelial cell lines should be studied to validate and elaborate on the molecular mechanisms unveiled by our results.

## Acknowledgments

Received 4/12/2007; revised 9/14/2007; accepted 9/25/2007.

**Grant support:** Department of Defense grants DMAD17-00-1-0247 and DMAD17-03-1-0229 and the W. Harry Feinstone Center for Genomic Research.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Dr. Joseph Testa and the Research Cytogenetics and Genomics Facility of the Fox Chase Cancer Center for the performance and analysis of metaphase-CGH and karyotyping.

The microarray data have been deposited in National Center for Biotechnology Information Gene Expression Omnibus (series accession no. GSE5116) at http://www.ncbi.nlm.nih.gov/geo/.

## References

1. Beatson G. On the treatment of inoperable cases of carcinoma of the mammary. Suggestions for new method of treatment with illustrative cases. Lancet 1896;2:104–7.
2. Boyd S. An oophorectomy in cancer of the breast. Br Med J 1900;2:1161–7.
3. Henderson BE, Ross R, Bernstein L. Estrogens as a cause of human cancer: the Richard and Hinda Rosenthal Foundation award lecture. Cancer Res 1988;48:246–53.
4. Suga S, Kato K, Ohgami T, et al. An inhibitory effect on cell proliferation by blockage of the MAPK/estrogen receptor/MDM2 signal pathway in gynecologic cancer. Gynecol Oncol 2007;105:341–50.
5. Bocchinfuso WP, Hively WP, Couse JF, Varmus HE, Korach KS. A mouse mammary tumor virus-Wnt-1 transgene induces mammary gland hyperplasia and tumorigenesis in mice lacking estrogen receptor-α. Cancer Res 1999;59:1869–76.
6. Rogan EG, Badawi AF, Devanesan PD, et al. Relative imbalances in estrogen metabolism and conjugation in breast tissue of women with carcinoma: potential biomarkers of susceptibility to cancer. Carcinogenesis 2003;24:697–702.
7. Cavalieri EL, Stack DE, Devanesan PD, et al. Molecular origin of cancer: catechol estrogen-3,4-quinones as endogenous tumor initiators. Proc Natl Acad Sci U S A 1997;94:10937–42.
8. Li JJ, Li SA. Estrogen carcinogenesis in Syrian hamster tissues: role of metabolism. Fed Proc 1987;46:1858–63.
9. Newbold RR, Liehr JG. Induction of uterine adenocarcinoma in CD-1 mice by catechol estrogens. Cancer Res 2000;60:235–7.
10. Chakravarti D, Mailander PC, Li KM, et al. Evidence that a burst of DNA depurination in SENCAR mouse skin induces error-prone repair and forms mutations in the H-ras gene. Oncogene 2001;20:7945–53.
11. Fernandez SV, Russo IH, Lareef M, Balsara B, Russo J. Comparative genomic hybridization of human breast epithelial cells transformed by estrogen and its metabolites. Int J Oncol 2005;26:691–5.
12. Lareef MH, Garber J, Russo PA, et al. The estrogen antagonist ICI-182–780 does not inhibit the transformation phenotypes induced by 17-β-estradiol and 4-OH estradiol in human breast epithelial cells. Int J Oncol 2005;26:423–9.
13. Cavalieri E, Rogan E, Chakravarti D. The role of endogenous catechol quinones in the initiation of cancer and neurodegenerative diseases. Methods Enzymol 2004;382:293–319.
14. Russo J, Hasan Lareef M, Balogh G, Guo S, Russo IH. Estrogen and its metabolites are carcinogenic agents in human breast epithelial cells. J Steroid Biochem Mol Biol 2003;87:1–25.
15. Yager JD, Davidson NE. Estrogen carcinogenesis in breast cancer. N Engl J Med 2006;354:270–82.
16. Russo J, Gusterson BA, Rogers AE, et al. Comparative study of human and rat mammary tumorigenesis. Lab Invest 1990;62:244–78.
17. Russo IH, Russo J. *In vitro* models for human breast cancer. In: Molecular basis of breast cancer prevention and treatment. Heidelberg: Springer-Verlag; 2004. p. 227–80.
18. Pilat MJ, Christman JK, Brooks SC. Characterization of the estrogen receptor transfected MCF10A breast cell line 139B6. Breast Cancer Res Treat 1996;37:253–66.
19. Fernandez SV, Russo IH, Russo J. Estradiol and its metabolites 4-hydroxyestradiol and 2-hydroxyestradiol induce mutations in human breast epithelial cells. Int J Cancer 2006;118:1862–8.
20. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100:8418–23.
21. Christiansen JJ, Rajasekaran AK. Reassessing epithelial to mesenchymal transition as a prerequisite for carcinoma invasion and metastasis. Cancer Res 2006;66:8319–26.
22. van de Rijn M, Perou CM, Tibshirani R, et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. Am J Pathol 2002;161:1991–6.
23. Russo J, Lareef MH, Tahin Q, et al. 17β-Estradiol is carcinogenic in human breast epithelial cells. J Steroid Biochem Mol Biol 2002;80:149–62.

**24.** Russo J, Fernandez SV, Russo PA, et al. 17-β-Estradiol induces transformation and tumorigenesis in human breast epithelial cells. FASEB J 2006;20:1622–34.

**25.** Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74.

**26.** Consortium. The Gene Ontology (GO) project in 2006. Nucleic Acids Res 2006;34:D322–6.

**27.** Waris G, Ahsan H. Reactive oxygen species: role in the development of cancer and various chronic conditions. J Carcinog 2006;5:14.

**28.** Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A 2003;100: 3983–8.

**29.** Danen EH, Yamada KM. Fibronectin, integrins, and growth control. J Cell Physiol 2001;189:1–13.

**30.** Fuchs IB, Lichtenegger W, Buehler H, et al. The prognostic significance of epithelial-mesenchymal transition in breast cancer. Anticancer Res 2002;22:3415–9.

**31.** Thiery JP. Epithelial-mesenchymal transitions in tumour progression. Nat Rev Cancer 2002;2:442–54.

**32.** Sommers CL, Heckford SE, Skerker JM, et al. Loss of epithelial markers and acquisition of vimentin expression in Adriamycin- and vinblastine-resistant human breast cancer cell lines. Cancer Res 1992;52:5190–7.

**33.** Okada H, Danoff TM, Kalluri R, Neilson EG. Early role of Fsp1 in epithelial-mesenchymal transformation. Am J Phys 1997;273:F563–74.

**34.** Kang Y, Massague J. Epithelial-mesenchymal transitions: twist in development and metastasis. Cell 2004; 118:277–9.

**35.** Nollet F, Berx G, van Roy F. The role of the E-cadherin/catenin adhesion complex in the development and progression of cancer. Mol Cell Biol Res Commun 1999;2:77–85.

**36.** Cano A, Perez-Moreno MA, Rodrigo I, et al. The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. Nat Cell Biol 2000;2:76–83.

**37.** Hajra KM, Chen DY, Fearon ER. The SLUG zinc-finger protein represses E-cadherin in breast cancer. Cancer Res 2002;62:1613–8.

**38.** Hemavathy K, Ashraf SI, Ip YT. Snail/slug family of repressors: slowly going into the fast lane of development and cancer. Gene 2000;257:1–12.

**39.** Zhou BP, Deng J, Xia W, et al. Dual regulation of Snail by GSK-3β-mediated phosphorylation in control of epithelial-mesenchymal transition. Nat Cell Biol 2004;6: 931–40.

**40.** Risse-Hackl G, Adamkiewicz J, Wimmel A, Schuermann M. Transition from SCLC to NSCLC phenotype is accompanied by an increased TRE-binding activity and recruitment of specific AP-1 proteins. Oncogene 1998;16:3057–68.

**41.** Janda E, Lehmann K, Killisch I, et al. Ras and TGF[β] cooperatively regulate epithelial cell plasticity and metastasis: dissection of Ras signaling pathways. J Cell Biol 2002;156:299–313.

**42.** Thuault S, Valcourt U, Petersen M, et al. Transforming growth factor-β employs HMGA2 to elicit epithelial-mesenchymal transition. J Cell Biol 2006;174: 175–83.

**43.** Willipinski-Stapelfeldt B, Riethdorf S, Assmann V, et al. Changes in cytoskeletal protein composition indicative of an epithelial-mesenchymal transition in human micrometastatic and primary breast carcinoma cells. Clin Cancer Res 2005;11:8006–14.

**44.** Liu R, Wang X, Chen GY, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. N Engl J Med 2007;356:217–26.

**45.** Roca X, Mate JL, Ariza A, et al. CD44 isoform expression follows two alternative splicing pathways in breast tissue. Am J Pathol 1998;153:183–90.

**46.** Thomas L, Byers HR, Vink J, Stamenkovic I. CD44H regulates tumor cell migration on hyaluronate-coated substrate. J Cell Biol 1992;118:971–7.

**47.** Droll A, Dougherty ST, Chiu RK, et al. Adhesive interactions between alternatively spliced CD44 isoforms. J Biol Chem 1995;270:11567–73.

**48.** Wang ZC, Lin M, Wei LJ, et al. Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers. Cancer Res 2004;64:64–71.

**49.** Loo LW, Grove DI, Williams EM, et al. Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. Cancer Res 2004; 64:8541–9.

**50.** Sheridan C, Kishimoto H, Fuchs RK, et al. $CD44^+/CD24^-$ breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis. Breast Cancer Res 2006;8:R59.

**51.** Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 2006;295:2492–502.

**52.** Soule HD, Maloney TM, Wolman SR, et al. Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. Cancer Res 1990;50:6075–86.