Breast Cancer Res Treat (2009) 114:47–62
DOI 10.1007/s10549-008-9982-8

PRECLINICAL STUDY

# Molecular signatures suggest a major role for stromal cells in development of invasive breast cancer

**Theresa Casey · Jeffrey Bond · Scott Tighe · Timothy Hunter · Laura Lintault · Osman Patel · Jonathan Eneman · Abigail Crocker · Jeffrey White · Joseph Tessitore · Mary Stanley · Seth Harlow · Donald Weaver · Hyman Muss · Karen Plaut**

Received: 8 January 2008 / Accepted: 17 March 2008 / Published online: 29 March 2008
© Springer Science+Business Media, LLC. 2008

**Abstract** *Background* Breast cancer invasion and metastasis involves both epithelial and stromal changes. Our objective was to delineate the pivotal role stroma plays in invasion by comparing transcriptomes among stromal and epithelial cells in normal tissue and invasive breast cancer. *Methods* Total RNA was isolated from epithelial and stromal cells that were laser captured from normal breast tissue ($n = 5$) and invasive breast cancer ($n = 28$). Gene expression was measured using Affymetrix U133A 2.0 GeneChips. Differential gene expression was evaluated and compared within a model that accounted for cell type (epithelial [E] versus stromal [S]), diagnosis (cancer [C] versus normal [N]) as well as cell type-diagnosis interactions. *Results* Compared to NE, the CE transcriptome was highly enriched with genes in proliferative, motility and ECM ontologies. Differences in CS and NS transcriptomes

**Electronic supplementary material** The online version of this article (doi:10.1007/s10549-008-9982-8) contains supplementary material, which is available to authorized users.

T. Casey · L. Lintault · J. White
Animal Science, University of Vermont, Burlington, VT 05405, USA

L. Lintault
e-mail: llintault@gmail.com

J. White
e-mail: Jeffrey.White@uvm.edu

T. Casey (✉) · O. Patel · K. Plaut
Animal Science, Michigan State University, East Lansing, MI 48824, USA
e-mail: ande1218@msu.edu

O. Patel
e-mail: patelo@msu.edu

J. Bond
Vermont Cancer Center Bioinformatics Core, University of Vermont, Burlington, VT 05405, USA
e-mail: Jeffrey.Bond@uvm.edu

S. Tighe · T. Hunter
University of Vermont Microarray Core Facility, University of Vermont, Burlington, VT 05405, USA
e-mail: Scott.Tighe@uvm.edu

T. Hunter
e-mail: Timothy.Hunter@uvm.edu

J. Eneman · A. Crocker · H. Muss
Hematology and Oncology, University of Vermont, Burlington, VT 05405, USA
e-mail: jeneman@yorkhospital.com

A. Crocker
e-mail: Abigail.Crocker@uvm.edu

H. Muss
e-mail: Hyman.Muss@uvm.edu

J. Tessitore · D. Weaver
Pathology, University of Vermont, Burlington, VT 05405, USA
e-mail: Joseph.Tessitore@vtmednet.org

D. Weaver
e-mail: Donald.Weaver@uvm.edu

M. Stanley · S. Harlow
Surgery, University of Vermont, Burlington, VT 05405, USA
e-mail: Mary.Stanley@uvm.edu

S. Harlow
e-mail: Seth.Harlow@uvm.edu

suggested that the ECM was being remodeled in invasive breast cancer, as genes were over-represented in ECM and proteolytic ontologies. Genes more highly expressed in CS compared to CE were primarily ECM components or were involved in the remodeling of ECM, suggesting that ECM biosynthesis and remodeling were initiated in the tumor stroma. *Conclusion* Based on identified molecular cross-talk between the two contiguous cell populations, a mechanistic model that spurs invasion is proposed, that shows breast cancer invasion proceeds through the acquisition of a motile phenotype in tumor epithelial cells and a reactive phenotype in cancer associated fibroblasts.

**Keywords** Epithelium · Invasion · Microarray analysis · Stroma · Transcriptome

**Abbreviations**

| | |
|---|---|
| NE | Normal epithelium |
| NS | Normal stroma |
| CE | Cancer epithelium |
| CS | Cancer stroma |
| LCM | Laser capture microdissection |
| EMT | Epithelial to mesenchyme transdifferentiation |

## Introduction

Most cancer deaths are preceded by local invasion and metastasis, with metastasis being the cause of 90% of deaths from solid tumors [1]. However, the molecular basis for invasion and metastasis is not well understood [2]. To progress to stage I breast cancer from DCIS, malignant epithelial cells must infiltrate through the basement membrane extracellular matrix (ECM) into the surrounding breast stroma [1]. Once invasion has occurred and cancer cells have infiltrated the stroma, they gain the potential to invade surrounding vascular structures and to metastasize [1, 3, 4].

Genetic and cell-biology studies have shown that tumor growth and progression is not only due to the transformed epithelial cells but also requires participation of the tumor stroma [1, 5]. Stromal tissue consists of fibroblasts, adipocytes, the ECM, and blood and lymph vessels, which all have been shown to influence tumor progression [5–9]. The stroma within the tumor microenvironment is often altered as cancer progresses, with alterations that include the activation of fibroblasts, remodeling of the ECM and angiogenesis [10, 11]. It has been hypothesized that these changes are instrumental in transforming the stroma into a metastasis supportive microenvironment [12–14].

Studies focused on describing an independent molecular signature of cancer associated stromal tissue are limited [15–17] because the transcriptome of each cellular compartment must be evaluated separately in order to ascribe effects to the stromal compartment. Recently, investigators used laser capture microdissection (LCM) to examine the gene expression signatures of tumor stroma versus epithelium in primary head-neck cancer and found that the metastasizing primary head-neck tumor is characterized by up-regulation of stroma specific genes, concomitant with predominant inactivation of tumor-epithelial specific genes [15]. LCM was also used to dissect normal epithelium and stroma derived from patients undergoing reduction mammoplasty or surgical treatment of breast cancer. Gene expression profiles revealed that normal stroma and epithelium from breast cancer patients are not statistically distinct from epithelium and stroma isolated from reduction mammoplasties and do not possess gene expression changes associated with standard clinical characteristics [18], suggesting that the changes in the stromal microenvironment are directly related to cancer progression, rather than being present prior to tumorigenesis.

These studies support the concept that the transcriptome of the tumor micro-environment is essential for understanding progression to metastatic disease. We hypothesize that while breast cancer is an epithelial cell phenomenon, it is likely that cancer associated fibroblasts (CAFs) act as mediators of tumor invasion and metastasis and are different from normal fibroblasts. As breast cancer progresses, gene expression within the tumor stroma is altered to allow for invasion and metastasis of epithelial tumor cells, and that epithelial cells and fibroblasts work together to promote tumor invasion and metastasis. We further hypothesize, that invasion occurs in part by remodeling the ECM to provide the scaffolding for epithelial tumor cell movement and angiogenesis. Our objective was to identify molecular signatures of the tumor microenvironment by analyzing differences in gene expression between stromal and epithelial cells in normal breast tissue and invasive breast cancer. To meet this objective, distinct epithelial and stromal cell populations were isolated from normal breast tissue and invasive breast cancer using LCM. Total RNA was extracted and gene expression was profiled using Affymetrix U133A 2.0 GeneChips. Differential gene expression was evaluated and compared within a model that accounted for cell type (epithelial [E] versus stromal [S]), diagnosis (cancer [C] versus normal [N]) as well as cell type-diagnosis interactions using gene ontology and pathway group analysis. We report that compared to NE, the CE transcriptome was highly enriched with genes in proliferative, motility and ECM ontologies. Differences in CS and NS transcriptomes suggested that the ECM was being remodeled in invasive breast cancer, as genes were over-represented in ECM and proteolytic ontologies. Genes more highly expressed in CS compared to CE were

primarily ECM components or were involved in the remodeling of ECM, suggesting that ECM biosynthesis and remodeling were initiated in the tumor stroma. Based on differences in trasnscriptomes we propose a model of breast cancer invasion that describes the acquisition of a motile phenotype in tumor epithelial cells and a reactive phenotype in CAFs.

## Materials and methods

### Study population

Breast tissue was collected from cancer patients and controls (reduction mammoplasty patients) that were identified through the Breast Care Center at Fletcher Allen Health Care; the University of Vermont affiliated hospital. Informed consent meeting all federal, state and institutional guidelines was obtained from all subjects. Tissue was collected from 28 patients with clinical stage I, II or III breast cancer (BC) and from 5 patients who underwent reduction mammoplasty (control). Final pathologic tumor stage was determined with the TNM staging system (AJCC Cancer Staging Manual, 6th edition, 2002) and graded using the Nottingham system [19]. In addition tumor grade, estrogen receptor (ER) and progesterone receptor (PR) status, presence or absence of lymphovascular invasion (LVI), and lymph node status were assessed in each tumor. Table 1 summarizes the clinical characteristics of this BC patient population. The mean, median and range of ages of control patients were 41.2, 39 and 21–55 years old, respectively.

**Table 1** Summary of breast cancer patient clinical data

| Age | Years |
|---|---|
| Mean | 55.4 |
| Median | 55 |
| Range | 40–78 |
| **Tumor size** | **No. (%)** |
| T1 (0.1–2cm) | 12 (42.9) |
| T2 (>2–5 cm) | 13 (46.4) |
| T3 (>5 cm) | 3 (10.7) |
| Node status | |
| Positive | 11 (39.3) |
| Negative | 17 (60.7) |
| Tumor grade[a] | |
| Well | 5 (17.9) |
| Moderate | 14 (50.0) |
| Poor | 9 (32.1) |
| Tumor type | |
| Ductal | 25 (89.3) |
| Lobular | 3 (10.7) |
| Estrogen receptor status[b] | |
| ER positive | 19 (73.1) |
| ER negative | 7 (26.9) |
| Progesterone receptor status[b] | |
| PR positive | 16 (61.5) |
| PR negative | 10 (38.5) |
| Lymphovascular invasion | |
| Positive | 10 (35.7) |
| Negative | 18 (64.3) |
| Stage | |
| I | 12 (42.9) |
| II | 10 (35.7) |
| III | 6 (21.4) |

All tumors were obtained with consent by surgical resection for primary diagnosis of breast cancer between 2004 and 2005. The ethnic composition of the patients was caucasian. Final pathologic tumor stage was determined using TNM staging system (AJCC staging manual)

[a] The Nottingham system was used to assess tumor grade (differentiation): grade 1 = well differentiated, grade 2 = moderately differentiated, grade 3 = poorly differentiated [103]

[b] Data missing from 2 patients

### Tissue collection and processing

Tissue was procured through surgical pathology and placed in Dulbecco's Modified Eagle's Medium (DMEM; Sigma Chemical Corporation, St. Louis, MO), pH 7.2–7.3, supplemented with 100,000 U/L penicillin G, 100 mg/l streptomycin (Gibco Invitrogen Corporation, Grand Isle, NY) and 10% fetal bovine serum (FBS, Gibco Invitrogen Corporation), and immediately transported on ice to the laboratory. The tissue was embedded in Tissue Tek OCT (optimal cutting temperature) cryostat embedding compound (Ames Co, Division of Miles Laboratory, Elkhart, ID), snap frozen in liquid nitrogen and stored at −80°C until sectioning. Serial tissue sections (7 μm) were prepared with a cryostat microtome using RNase-free techniques and stored at −80°C until use (less than 8 weeks). On day of LCM, tissue sections were fixed and stained in In Situ Hybridization Pap Jars (Evergreen Scientific, Los Angeles, CA) using the Histogene Kit (Arcturus Engineering, Mountain View, CA) following manufacturer's directions and left in xylene until LCM to ensure tissue was dehydrated.

### LCM and isolation of total RNA

LCM was used to isolate populations of normal epithelial cells and associated fibroblasts as well as epithelial cells from tumor and associated fibroblast cells. LCM was performed with an Arcturus PixCell IIe LCM system

(Arcturus Engineering) using CapSure LCM Transfer Film (Arcturus Engineering) according to manufacturer's protocol. While capturing cells, careful consideration of histopathology was taken into account. When microdissecting stromal tissue, we focused on capturing fibroblasts, while excluding lymphocytes and obvious vascular structures. Further, we captured stroma immediately associated with tumor epithelial cells, rather than stromal tissue distal to the tumor. Tissue captured onto transfer film was immediately immersed in 50 µl of isolation buffer from Acturus' Picopure kit (Arcturus Engineering).

Total RNA was ioslated and purified by pooling lysates from several slides of the same samples following manufacturer's protocol for laser captured frozen tissue. An optional DNase step was ommitted to prevent unnecessary losses of the limited RNA. Although this DNase step is critical for samples being amplified through a random priming technique, it is not necessary when doing synthesis reactions using Oligo (d) T. Recovered RNA was subsequently analyzed using a Bioanalyzer 2100 (Agilent Inc; Palo Alto, CA) to determine quality and quantity by use of the Picochip and multiple standards.

### Target preparation and microarray analysis

Target preparation was accomplished using NuGen Ribo-SPIA Ovation Technology with a 10 minute increase in the incubation time for the first strand cDNA synthesis reaction and SPIA amplification step in order to generate enough products for the Affymetrix HG U133A 2.0 GeneChip. Gene chips were hybridized and scanned using standard Affymetrix protocols. In accordance with MIAME convention microarray data were made available to the public through the Gene Expression Omnibus (GEO) website (http://www.ncbi.nlm.nih.gov/geo/) with the following accession number GSE10797.

### Analysis of microarray data

Calculations were performed using the R Programming Language and Environment for Statistical Computing [20] along with Bioconductor [21] packages. Normalized probe set statistics were calculated using the RMA statistic [22, 23]. For each pair-wise comparison (for example, CS vs NS) the difference between the mean RMA statistics was calculated as well as a *P*-value (using a two-sided Welch's *t*-test).

Rather than attempt to justify the assumptions required for the *t*-test (which are often violated in microarray studies) we interpreted the *P*-value simply as a second statistic and estimated the number of probe sets passing filters (based on fold change, *P*-value, or both) under the null hypothesis by permuting sample labels. When sample sizes are small (we have five normal samples) or the variance associated with differentially expressed genes is large use of a *P*-value filter can eliminate many true positives and remove the signature of important biological processes. For each probe set and each of 200 permutations of sample labels we calculated the same two statistics and estimated false discovery rates for a variety of filters (fold change-based filters are shown in Supplemental Figure 3). Although both each types of filters allowed identification of sets of genes having acceptable false discovery rates, the focus of this study is on highly differentially expressed genes and we adopted a two-fold differential expression threshold. The number of probe sets exceeding an eight-fold differential expression threshold was significant at $P < 0.005$, that is, none of the 200 permuted sets yielded as many hits as the natural data set. Inference was based on analysis of the association of sets of differentially expressed genes with GO annotation, performed as described by Falcon and Gentleman [24] using their Bioconductor package (GOstats) to calculate *P*-values.

GeneOntology, Pubmed, GeneSifter™ software (VizXLaboratories, Seattle, WA), as well as Ingenuity Pathways Analysis (Ingenuity® Systems, Inc., Redwood City, CA) were used to define function and categorize genes as well as biological significance of the differentially expressed genes. Highly enriched gene ontologies were defined as ontologies assigned z-scores >2 with at least 5 genes within the given category. Highly under-represented categories were assigned z-scores $< -2$.

## Results

### RNA suitable for microarray studies was isolated from homogeneous populations of cells

Although both normal and cancerous breast stromal tissue contain a mixed population of cells including fibroblasts, endothelial and immune cells, we decided to focus our study on the role that fibroblasts play in cancer invasion. Care was thus taken to enrich fibroblasts by avoiding blood vessels and obvious immune cells during LCM of the stromal tissue. Pictures were taken before and after LCM to document cell types and tissue captured (Fig. 1). RNA was isolated from samples and amplified using the Nugen Ovation SPIA system allowing for use of less than 20 ng of input RNA. The use of small quantities of RNA relative to other GeneChip experiments suggests that special care be taken in analyzing expression statistics. Although quality statistics (for example, percentage of probe sets called present) reflected the smaller quantities of RNA, it is noteworthy that, for certain probe sets, differences in signal intensity were very clear (an example is shown in supplemental Fig. 1) and inferences were robust with respect to the quality of the samples included in the analysis (an

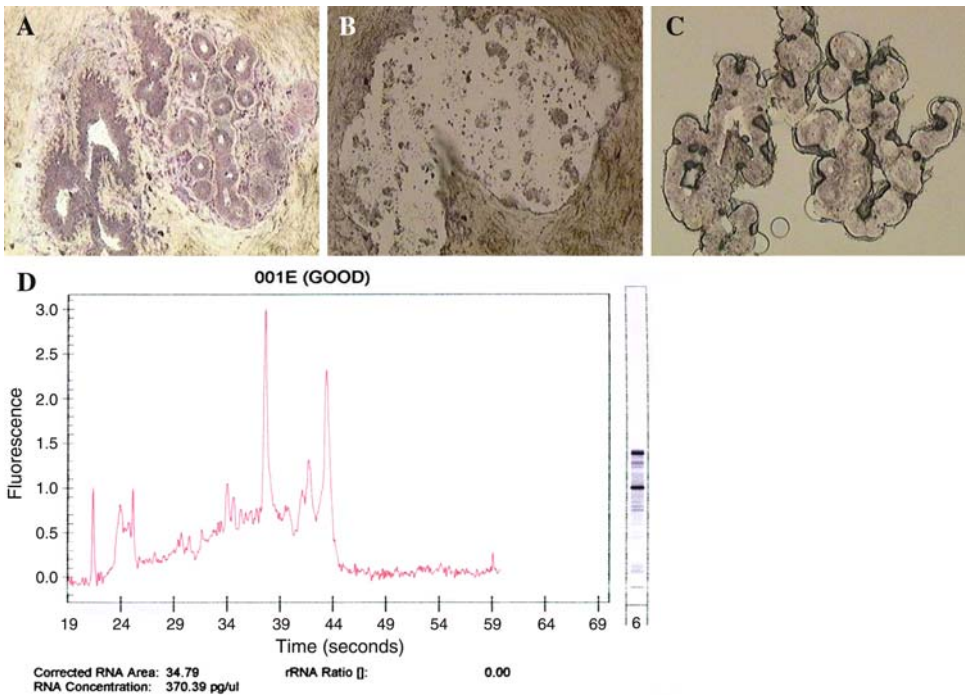


**Fig. 1** Laser capture microdissection of breast epithelial cells using Arcturus PixCell II; (**a**) area to be captured (200×); (**b**) remaining tissue after capture (200×); (**c**) cells on CapSure LCM transfer film. (**d**) Shows the quality of total RNA isolated from laser captured cells

example is shown in supplemental Fig. 2). Wilcoxon tests of the association of sample summary statistics (percentage of probe sets called present, scale factor, or the nine raw expression deciles) with sample group failed to reject the null hypotheses for either stromal or epithelial cells.

*Genes differentially expressed between normal breast stroma and epithelial tissues are categorized into unique ontologies typical of normal tissues*

When gene expression profiles were compared between normal stroma (NS) and normal epithelial (NE) tissue, using a 2-fold cut-off, 1,264 genes were found to be differentially expressed, with 266 of these genes up regulated in stroma relative to epithelial tissue. This subset of genes was then sorted by ontology and z-score using GeneSifter™ and revealed differences in molecular signatures between NS and NE (Fig. 2a). Ontologies assigned to NE with high negative z-scores were cell motility, locomotion and cell migration (−3.20, −3.22 and −2.80, respectively), while "cytoskeletal organization and development", "anti-apoptosis", and "programmed cell death" were assigned relatively high positive z-scores (2.17, 3.39, and 2.30). Within the "cellular component" ontology, 26 gene products were categorized as "extracellular matrix (ECM)" and/or "ECM sensu Metazoa research community" with z-scores of 4.53 for genes up-regulated in NS versus NE (Fig. 2a). KEGG pathways that were assigned high z-scores included "focal adhesion" pathway, "ECM receptor" and "cell communication" for genes up-regulated in NS versus NE (6.09, 7.86, 6.24, respectively). Many of these genes were among the same genes within ECM ontologies. The KEGG pathways assigned high z-scores for genes more highly expressed in NE versus NS included "oxidative phosphorylation" and "ATP synthesis" (6.40 and 5.02, respectively).

*Genes encompassing cell cycle regulation and the extracellular matrix ontologies predominate in epithelium harvested from breast cancer patients*

When gene expression profiles were compared between cancer epithelial (CE) and NE tissues, we found that 914 genes were differentially expressed at a 2-fold cut-off, with 235 of these genes up-regulated in cancer versus normal epithelial tissue. As expected, when these genes were sorted by ontology "M-phase mitotic cell cycle" was assigned a high z-score, 2.94, in CE, while genes within this ontology were underrepresented in NE, and were thus assigned a z-score of −2.78 (Fig. 2b). Accordingly, the ontologies "negative regulation of cell cycle" and "negative regulation of programmed cell death" were assigned 4.24 and 3.94, respectively for genes more highly expressed in NE. ECM related ontologies were assigned high z-scores for CE, while cytoskeletal related ontologies were assigned strong z-scores for genes down regulated in CE versus NE (Fig. 2b). KEGG pathways related to ECM ontologies "cell communication", "focal adhesion" and "ECM-receptor interaction" were assigned high z-scores for genes up-regulated in CE versus NE (Fig. 2b). The "GAP junction" pathway was assigned a high z-score

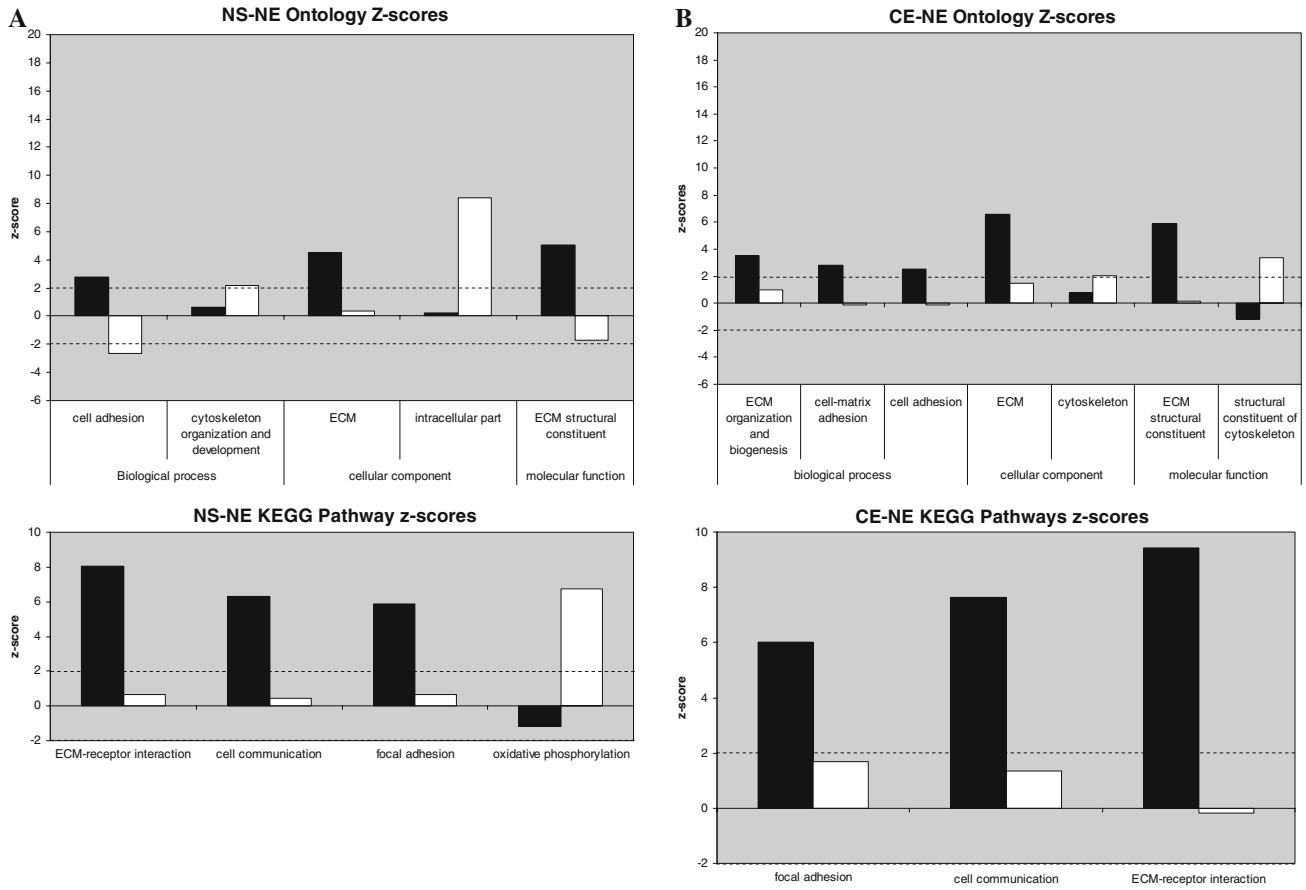


**Fig. 2** GeneSifter™ was used to sort and annotate genes differentially expressed between (**a**) normal stromal (NS) and normal epithelial (NE) breast tissue; (**b**) invasive breast cancer epithelial (CE) tissue and NE; (**c**) invasive breast cancer stromal (CS) tissue and NS; and (**d**) CS and CE transcriptomes using a 2-fold threshold. Graphs show ontologies within "biological process", "cellular component" and "molecular functions" and KEGG pathways that were highly enriched (z-score $> 2$) or highly under-represented (z-score $< -2$) with genes up-regulated (*black bars*) or down regulated (*white bars*) between comparisons

(2.84) for genes more highly expressed in normal versus cancer epithelial tissue. Importantly, when Ingenuity's® functional analysis tool was used to sort data "cell motility", "cell death" and "cell cycle" were found to be highly significant ($P < .00001$) pathways changed in CE versus NE.

*Ontological classification reveals coordinated alteration in genes associated with ECM and proteolysis with invasive breast cancer stroma*

Gene expression patterns associated with invasive breast stroma were significantly different from normal breast stroma. There were 1,151 genes differentially regulated with a 2-fold cut off, with 276 of these genes up-regulated in cancer stroma (CS) versus (NS). GeneSifter™ was used to sort genes within the three major ontologies (biological process, cellular component and molecular function) and assigned z-scores. Gene ontologies with significant z-scores for CS versus NS are presented in Fig. 2c. "ECM organization and biogenesis" as well as "proteolysis" were assigned strong z-scores for genes up-regulated in CS versus NS (4.44 and 2.37 respectively). When genes were sorted within the cellular component ontology, 17 and 25 genes were categorized as "ECM" and "ECM sensu Metazoa research community", respectively, with genes up-regulated in CS versus NS assigned strong z-scores (7.86 and 6.65, respectively). KEGG pathways with strong z-scores identified through GeneSifter™ were "ECM-receptor interaction" and "cell communication" with many of the same genes being highlighted in both pathways.

*Functional characterization of cancer stromal and epithelial transcriptomes suggests that ECM production and remodeling is primarily regulated through the stromal compartment*

There were 1,691 genes differentially expressed at a 2-fold cut off between CS and CE, with only 45 of these genes

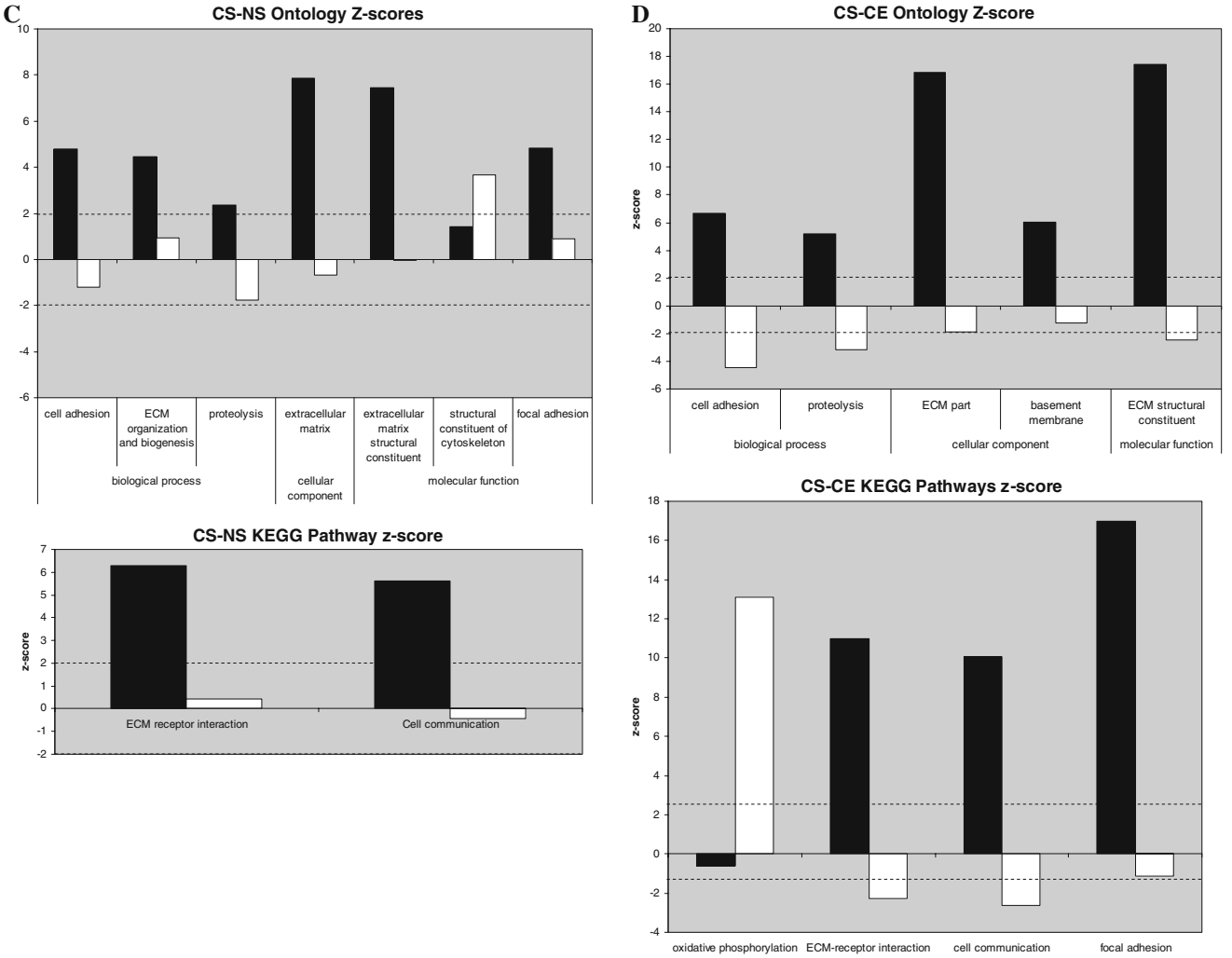


**Fig. 2** continued

up-regulated in CS. When GeneSifter$^{TM}$ was used to sort these genes into ontologies within biological process (Fig. 2d), "cell adhesion" and "proteolysis" were assigned high z-scores for genes up-regulated in CS (6.66 and 5.20, respectively), while CE was assigned highly negative z-scores for these same ontologies (−4.48 and −3.15). Genes up-regulated in CS were also sorted into ontologies that suggest that ECM biosynthesis and remodeling are initiated in these tissues as "ECM part", "basement membrane", "proteolysis" were assigned high z-scores. KEGG pathways that were assigned high z-scores included "focal adhesion" pathway, "ECM receptor" and "cell communication" pathways for up-regulated genes (9.48, 10.97, and 10.05 respectively) in CS; many of these genes were the same as the genes in the extracellular matrix ontology. KEGG pathways assigned high scores for genes more highly expressed in cancer epithelium relative to cancer stroma were several metabolic pathways including "oxidative phosphorylation".

*Detailed characterization of differentially expressed genes clustered into ECM ontological category*

Since ECM related ontologies were consistently assigned strong z-scores (>2) within all comparisons made, we decided to focus on the genes that were differentially expressed within these categories. There were 61 genes that were differentially expressed in at least one of the comparisons within "ECM" and "ECM sensu Metazoa research community" ontologies. The Venn diagram (Fig. 3) shows the relationship of these genes among the comparisons. Based on these relationships we were able to categorize changes in gene expression among the groups as cancer associated changes (occurring in both epithelial and stromal tissues), cancer stroma specific, cancer epithelium specific, or differential expression found in normal breast tissue (Table 2).

Six genes (*COL1A1, COL1A2, COL5A1, COL6A3, THBS2, DST*) were differentially expression in all

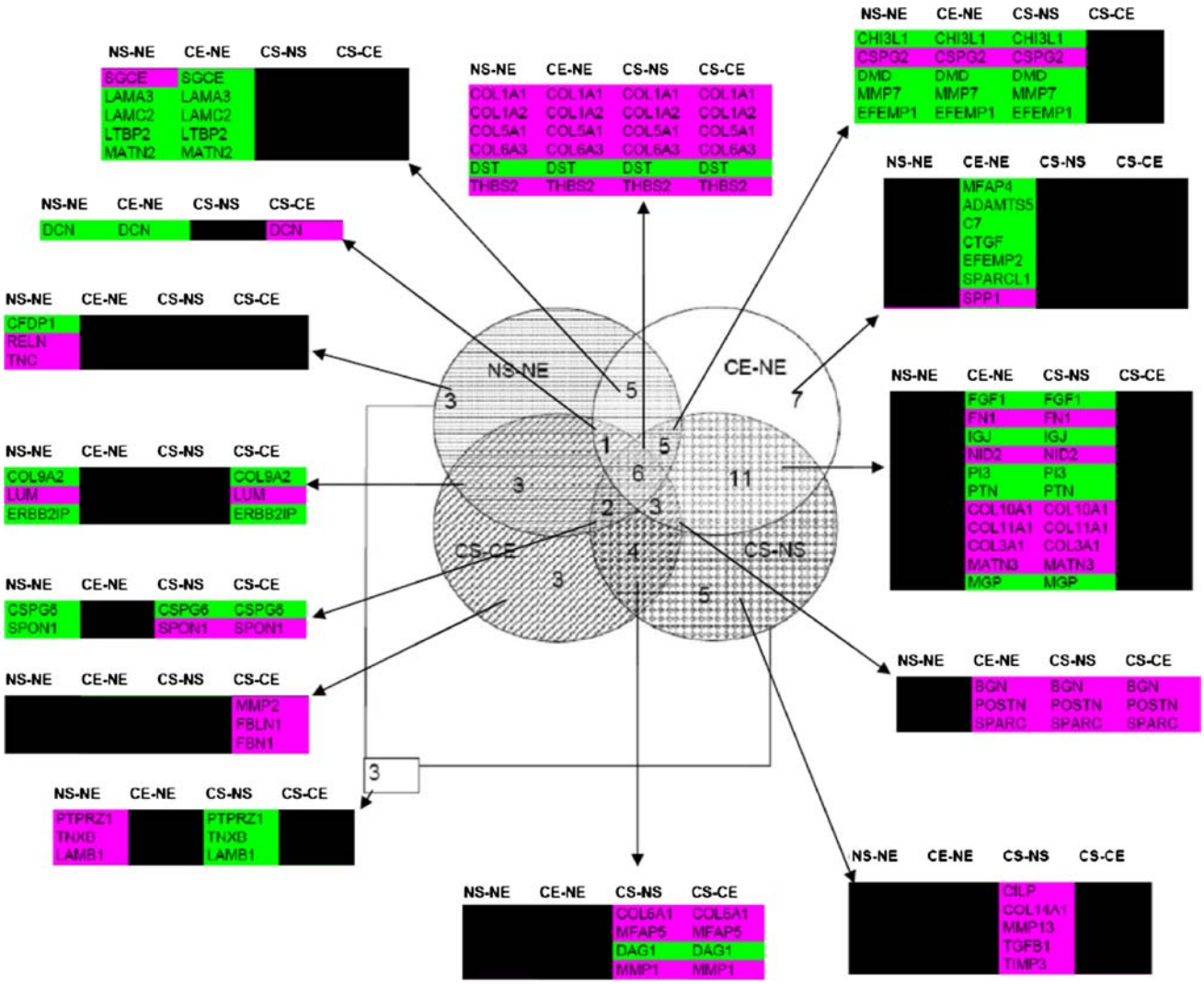


**Fig. 3** Venn diagram illustrates the relationship of the 61 genes in "ECM" and/or "ECM sensu Metazoa research community" ontologies differentially expressed among the 4 comparisons. Arrows point to "heat" maps indicating genes differentially expressed among the comparisons, with pink representing at least a 2-fold higher expression and green indicating at least a 2-fold lower expression of gene between comparison within a column (NS-NE; CE-NE; CS-NS; CS-CE), and black indicating no difference in expression between comparison

comparisons (Fig. 3; Table 2). Five of these genes were expressed at a greater level in malignant versus normal tissue and in stromal tissue versus epithelial. *DST*, dystonin, was greater in epithelial tissue and less in cancer. The expression of 10 genes (*FN1, NID2, COL10A, COL11A1, COL3A1, MATN3, BGN, POSTN, SPARC* and *CSPG2*) was increased in both cancer epithelium and stroma relative to normal epithelium and stroma, respectively. In contrast the expression of 9 genes was at least 2-fold less in cancer epithelial and stromal tissues versus normal tissues, and included, *FGF1, IGJ, PI3, PTN, MGP, CHI3L1, DMD, MMP7* and *EFEMP1*.

The ECM genes that were specific to stroma and had at least a 2-fold greater level of expression in cancer stroma versus normal breast stroma included *CILP, COL14A1, MMP13, TGFB1, TIMP3, COL6A1, MFAP5, MMP1*, and *SPON1*. The other 5 genes were expressed at a greater level in normal breast stroma versus stromal tissue isolated from breast cancer patients. Gene expression differences specific to tumor epithelium were for the most part down regulated in cancer epithelium versus normal epithelium except *SSP1* (osteopontin).

## Discussion

Our initial examination of microarray data using gene ontology and pathway analysis clustering of differential gene expression showed showed that trasncriptomes were distinctly different between normal and cancerous

**Table 2** Cell-specific gene signatures from breast cancer stromal and epithelial cells form enriched ECM gene ontology

| Gene symbol | Gene name |
|---|---|
| *Expression difference due to cancer; change evident in both tumor epithelium and stroma* | |
| BGN | Biglycan |
| CFDP1 | Cranofacial development protein 1 |
| CHI3L1 | Chitinase 3-like |
| COL10A1 | Collagen type X, alpha 1 |
| COL11A1 | Collagen type XI, alpha 1 |
| COL1A1 | Collagen type I, alpha 1 |
| COL1A2 | Collagen type I, alpha 2 |
| COL3A1 | Collagen type III, alpha 1 |
| COL5A1 | Collagen type V, alpha 1 |
| COL6A3 | Collagen type VI, alpha 3 |
| CSPG2 | Bamacan (chondroitin sulfate proteoglycan 2) |
| DMD | Dystrophin |
| DST | Dystonin |
| EFEMP1 | EGF-containing fibulin-like extracellular matrix protein 1 |
| FBLN1 | Fibulin 1 |
| FBN1 | Fibrillin 1 |
| FGF1 | Fibroblast growth factor 1 |
| FN1 | Fibronectin 1 |
| IGJ | Enamelin |
| MATN3 | Matrilin 3 |
| MGP | Matrix gla protein |
| MMP2 | Matrix metalloproteinase II |
| MMP7 | Matrix metalloproteinase VII |
| NID2 | Nidogen 2 (osteonidogen) |
| PI3 | Peptidase inhibitor 3 |
| POSTN | Periostin, osteoblast specific factor |
| PTN | Pleiotrophin (heparan binding growth factor 8) |
| RELN | Reelin |
| SPARC | Secreted protein, acidic, cysteine-rich (ostenectin) |
| THBS2 | Thrombospondin 2 |
| TNC | Tenascin C |
| *Gene expression difference specific to stroma* | |
| CILP | Cartilage intermediate layer protein |
| COL14A1 | Collagen type XIV, alpha 1 |
| COL6A1 | Collagen type VI, alpha 1 |
| CSPG6 | Chondroitin sulfate proteoglycan 6 |
| DAG1 | Dystroglycan 1 |
| LAMB1 | Laminin, beta 1 |
| MFAP5 | Microfibillar associated protein 5 |
| MMP1 | Matrix metalloproteinase I |
| MMP13 | Matrix metalloproteinase XIII |
| PTPRZ1 | Protein tyrosine phosphatase, receptor-type, Z polypeptide I |
| SPON1 | Spondin 1, extracellular matrix protein |
| TGFB1 | Transforming growth factor-beta 1 |
| TIMP3 | TIMP metallopeptidase inhibitor 3 |
| TNXB | Tenascin XB |
| *Gene expression difference specific to tumor epithelium* | |
| ADAMTS5 | Adam metallopeptidase with thrombospondin type1 motif |
| C7 | Complement component 7 |
| CTGF | Connective tissue growth factor |
| DCN | Decorin |
| EFEMP2 | EGF-containing fibulin-like extracellular matrix protein 2 |
| LAMA3 | Laminin, alpha3 |
| LAMC2 | Laminin, gamma 2 |
| LTBP2 | Latent TGF-beta binding protein 2 |
| MATN2 | Matrilin 2 |
| MFAP4 | Microfibillar asscociated protein 4 |
| SGCE | Sarcoglycan epsilon |
| SPARCL1 | Sparc-like 1 |
| SPP1 | Secreted phosphoprotein 1 (osteopontin) |
| *Gene expression difference is found in normal breast* | |
| COL9A2 | Collagen type IX, alpha 2 |
| ERBB2IP | ERBB2 interacting protein |
| LUM | Lumican |

epithelial and stromal tissues. When stromal and epithelial transcriptomes were compared within normal tissue, we found expression profiles concurred with a previous report that compared gene expression profiles between normal epithelial and stromal tissues isolated by LCM from reduction mammoplasty tissue and normal tissue distal from breast tumors. In their study 883 genes were differentially expressed between matched epithelial and stromal samples, with molecular profiles similar to ours including an enrichment of ECM-related genes in normal stromal tissue [18]. In our study, genes were highly under-represented in ontologies related to cell motility and migration (highly negative z-scores), while cytoskeletal organization and development as well as anti-apoptosis ontologies were enriched with genes more highly expressed in normal breast epithelial cells. The transcriptome of normal epithelial tissue also indicated that it is more highly metabolic than adjacent normal stromal tissue as several metabolic pathways, including oxidative phosphorylation, were assigned strong z-scores.

There were 28 ECM genes differentially expressed between normal stroma and epithelial tissues. The 13 genes that were up-regulated in stromal tissue versus epithelial were assigned high z-scores, and included collagens, tenascin C and XB, as well as three genes important to

cell-cell and cell-ECM interaction, reelin, thrombospondin 2 and sarcoglycan epsilon. Although 15 ECM ontology genes were more highly expressed in normal epithelial tissue compared to normal stromal tissue they were not assigned high z-scores. These 15 genes included two laminin chains, gamma 2 and alpha 3, as well as chitinase 3-like, bamacan (*CSGP6*), decorin, matrilin 2, and spondin 1 all of which encode extracellular matrix proteins, supporting evidence that both normal breast epithelial and stromal cells synthesize components of the extracellular matrix [25–29]. Other genes that were more highly expressed in epithelial tissue are important to signaling in epithelial cells and included ERBB2 interacting protein, dystrophin and dystonin. Specifically, protein products of these genes act as integrin signaling molecules, anchor cytokeratin intermediate filaments (IF) to hemidesmisomes and bridge the ECM to intracellular IF, respectively.

When molecular signatures were compared between cancer and normal epithelial tissues, we found that ontologies assigned high z-scores as well as pathways identified through Ingenuity's® functional analysis tool were typical of breast cancer and normal epithelial phenotypes, respectively. Our findings that "M-phase mitotic cell cycle" ontology genes were up-regulated in cancer epithelium versus normal epithelium, substantiated the EASE analysis of the "Intrinsic/UNC" breast cancer gene set [30] that showed the GO categories "mitotic cell cycle" and "M phase" were over-represented in the proliferation signature. We found that genes more highly expressed in normal epithelial tissue were over-represented in "negative regulation of cell cycle" and "negative regulation of cell death" ontologies relative to cancer epithelium, suggesting that there is a loss of function of cell cycle and cell death negative regulatory check points in transformed epithelial cells, and may shed light on the fact that breast tumors exhibit both high rates of proliferation and cell death [31–36]. Further, gene expression abnormalities detected at the in situ stage of breast cancer are related to specific genes in charge of regulating the proper homeostasis between cell death and cell proliferation [37]. Ten genes within "structural constituent of cytoskeleton" were down regulated in cancer epithelium relative to normal epithelium, including five cytokeratin genes, three genes involved in actin IF assembly, and gene products which bridge IF of the cytoskeleton to hemidesmisomes and ECM receptors. The KEGG GAP junction pathway was enriched with genes more highly expressed in normal epithelial tissue including EDG2, the LPA receptor, and tubulin, which is involved in polymerization of actin IF. These data support previous reports which showed loss of expression of normal epithelial markers, junctional proteins and proteins involved in cell communication in invasive breast cancer, including the down regulation of basal cytokeratins in the tumor epithelium [38–40]. It is important to note that since our study population is primarily composed of women with ER+ tumors, which most often show ER+/luminal molecular subtypes [41], the lower expression of basal cytokeratin (KRT) genes in CE compared to NE may be indicative of proliferation of malignant cells with a luminal phenotype rather than the down regulation of gene expression. Loss of gap junctions is associated with cancer invasion (for review [42]), with the loss of EDG2 expression being associated with increased breast cancer cell motility [43]. Cell motility involves cytoskeletal changes, cell-matrix interactions, localized proteolysis, actin-myosin contractions, and focal contact disassembly [44]. Thus the overall molecular signature of cancer epithelium versus normal epithelium suggests a cell motility phenotype for breast tumor cells. This is supported by our findings through Ingenuity's® Functional Analysis tool that showed an enrichment of cellular movement genes to be differentially expressed between the 2 populations ($P < 0.000001$).

ECM related genes were also differentially regulated between breast cancer epithelium and normal epithelium. Cancer epithelium expressed higher levels of collagens I, III, V and VI, fibronectin, *SPARC* (osteonectin), *SSP1* (osteopontin) and nidogen. Increased expression of fibronectin [45], *SPARC* and *SSP1* [46, 47] have been associated with the molecular signature of metastatic breast cancers. *POSTN*, periostin, was increased in cancer epithelium versus normal epithelium, which is in agreement with previous reports describing tumor epithelial specific transcriptomes [38]. When expression of POSTN was analyzed immunohistochemically with breast cancer tissue arrays, its expression correlated with poor outcome in a cohort of estrogen receptor-positive tumors. Further, our findings that laminins alpha 3 and gamma were among the genes that were down regulated in tumor versus normal epithelium concurs with previously described breast tumor epithelial transcriptomes [38].

Differences in molecular signatures of stromal tissue from invasive breast cancer patients and normal reduction mammoplasty patients were over-represented by genes primarily in ECM related ontologies as well as proteolysis. Of the 39 ECM-related genes ("ECM" and "ECM sensu Metazoa research community" combined) differentially regulated between cancer and normal stroma, the 24 that were up-regulated in cancer stroma included collagens I, III, V, VI, X, XI and XIV, biglycan, fibronectin, spondin 1 and nidogen. *SPARC* (osteonectin), *POSTN* (periostin) and *TGFB* were also more highly expressed in cancer versus normal breast stroma. SPARC is an anti-adhesion matricellular protein that mediates interactions between the extracellular matrix (ECM) and cells [48], and is involved in a variety of diverse biological processes including tissue remodeling, cell adhesion, proliferation, differentiation, matrix synthesis/turnover, angiogenesis and tumor cell

migration and invasion [48]. SPARC has been found to be over-expressed in a variety of cancers including in the stroma of invasive ductal carcinomas of the breast, and its anti-adhesive role is believed to facilitate tumor cell movement within the stromal compartment [49]. TGFB is a cytokine which inhibits epithelial cell growth, stimulates mesenchymal cell proliferation, regulates ECM deposition and degradation, and modulates immune function and wound repair [50–54] and is believed to both inhibit and promote progression of cancer [55, 56]. The dissociation of TGFB's growth inhibitory effects from its effects on ECM and induction of epithelial to mesenchyme transdifferentiation (EMT) is believed to be the turning point in its role as a tumor suppressor to its role as a breast cancer progression and metastatic factor [55].

There were 13 ECM proteases whose mRNA expression was up-regulated in cancer stroma versus normal stroma including, *MMP-1, MMP-13, FAP* and *TLL2*. MMP-1 and MMP-13 are interstitial collagenases and thus cleave collagen types I, II and III [57], and have been extensively implicated in tumor invasion of the stroma [58–62]. Mouse xenograph models, in which human breast cancer cells were implanted in mammary fat pads and bone, showed the specific induction of *MMP-13* mRNA expression in stromal tissue [60]. FAP (fibroblast activated protein alpha) is an integral membrane gelatinase belonging to the serine protease family, and is selectively expressed in reactive fibroblasts of epithelial cancers, granulation tissue of healing wounds, and malignant cells of bone and soft tissue sarcomas. FAP is thought to be involved in the control of fibroblast growth and epithelial-mesenchymal interactions during development, tissue repair, and epithelial carcinogenesis [63–67], and has thus been proposed as molecular anti-breast cancer invasion target [68]. TLL2 (tolloid-like 2) is a member of the BMP proteins which are metallopeptidases with high specificity for laminin 5, a major basement membrane component. TLL2 has been shown to be important to embryogenesis as well as implicated in solid tumor metastasis [69, 70]. The lysosomal cysteine proteases cathepsins B and K were among the genes up-regulated in cancer stroma relative to normal stroma within the proteolysis ontology. Cathepsin K is involved in bone remodeling and resorption and expressed in osteoclasts, as well as human breast cancer in which it has been associated with invasion [71, 72]. Similarly, cathepsin B has also been implicated in breast tumor progression and invasion (for review [73]). Cathepsin B mutants were shown to impair tumor formation and angiogenesis, while cathepsin B knockouts retarded cell proliferation and tumor growth [74].

The overall description of the breast cancer stromal transcriptome observed in this study was similar to the transcriptome of CAFs from basal cell carcinoma isolated by LCM [75], in that they both had an enrichment of up-regulated ECM genes. Thirteen CAF-associated genes were identified as being highly up-regulated in at least 2 of the 3 basal cell carcinoma patient transcriptomes studied [75]. Five of these thirteen genes were up-regulated in the stromal transcriptome of invasive breast cancer patients in our study and including *SPARC*, laminin-alpha 2, collagen type V, cathepsin K (*CTSK*), and cartilage intermediate layer protein (*CILP*). However 2 of the 13 genes up-regulated in basal cell carcinoma, matrix Gla-protein (*MGP*) and dermatopontin (*DPT*), were down regulated in our invasive breast cancer stromal transcriptome relative to normal breast stroma. The remaining genes identified in the basal cell carcinoma study were not differentially expressed at the 2-fold cut off. When CAF transcriptomes of basal cell carcinoma patients were compared to prostate and colon CAF transcriptomes, it was reported that basal cell carcinoma genes were not consistently up-regulated in CAFs of prostate or colon cancer, and in fact in colon cancer most of the analyzed genes were down-regulated [75].

When we compared the transcriptomes of breast cancer stroma and tumor epithelium we found that genes more highly expressed in tumor stroma were primarily either ECM components or involved in the remodeling of ECM (Fig. 2d). Seventy-five percent of the 22 ECM genes ("ECM" and "ECM sense Metazoa research community" combined ontologies) differentially expressed between cancer stroma and epithelium were up-regulated in the stroma. There were 3 differentially expressed ECM ontology genes unique to this comparison: *MMP-2*, *FBLN1*, and *FBN1* (Fig. 4). *MMP-2* encodes an enzyme, gelatinase A, which degrades type IV collagen, a major structural component of basement membranes [76]. *FBLN1* encodes fibulin 1, a secreted glycoprotein that becomes incorporated into a fibrillar extracellular matrix that binds to laminin, nidogen and fibronectin, regulates cell adhesion and has been shown to inhibit cell motility [77, 78]. *FBN1* encodes a member of the fibrillin family. Fibrillin 1 is a large, extracellular matrix glycoprotein that provides force bearing structural support in elastic and non-elastic connective tissue as well as modulates the activity of members of the TGF-$\beta$ family [79]. The protein products of 3 of the 5 genes more highly expressed in cancer epithelium versus stroma directly regulate interactions between epithelial cells and the ECM. Dystroglycan (DAG1) is a membrane spanning glycoprotein that connects the ECM to the cytoskeleton through binding to laminin and other basement membrane proteins in its extracellular domain and the actin cytoskeleton in its cytoplasmic domain [80, 81]. Dystonin (DST) is an adhesion junction plaque protein, which functions to anchor keratin-containing intermediate filaments to hemidesmosomes [82, 83]. ERBB2 interacting protein (ERBB2IP) binds integrins and initiates integrin

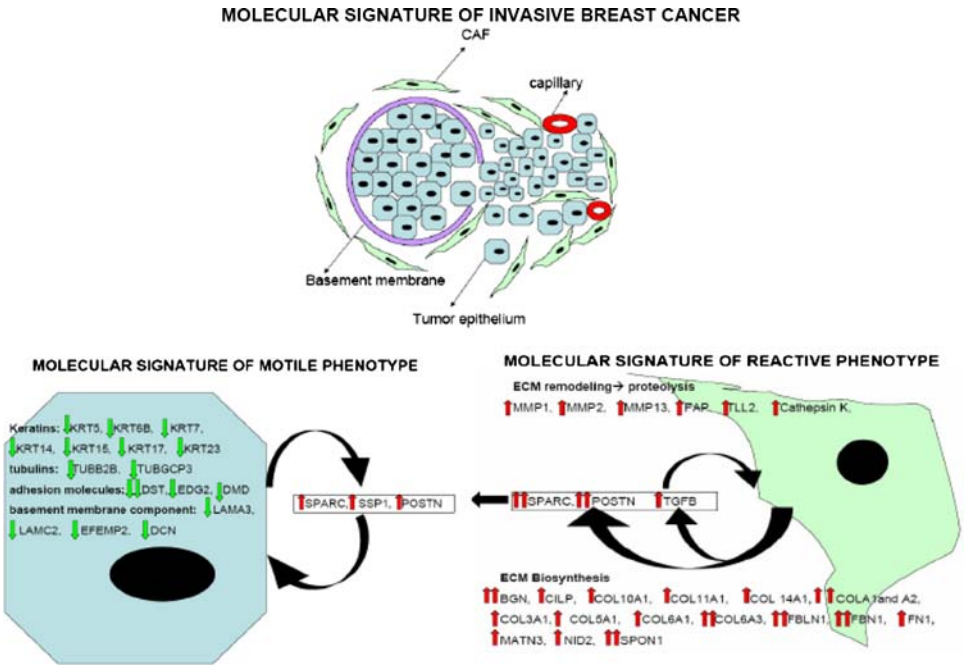


**Fig. 4** Molecular signature of invasive breast cancer: a model of the differential contributions of tumor epithelial cells and CAFs in the transition from DCIS to invasive breast cancer. During malignant transformation tumor epithelial cells develop a motile phenotype and the surrounding stroma becomes 'reactive'. Development of a motile phenotype is primarily through the loss of expression or down-regulation of genes that encode cytoskeletal proteins: keratins and tubulins, adhesion molecules as well as basement membrane components. Reactive stroma is marked by proliferation of activated CAFs that inturn express relatively high levels of ECM proteins and proteases, which is characteristic of the reactive phenotype. Increased levels of TGFB auto-induce the reactive phenotype of CAFs. Upregulation of SPARC, POSTN, SSP1, and TGFB in tumor epithelial cells and or CAFs stimulate migration of tumor epithelial cells. Red arrows indicate 2-fold increase and green arrows indicate 2-fold decrease in expression of gene within cancer tissue versus normal tissue. Double red arrows indicate gene expression was at least 2-fold greater in CS versus CE

signaling (Ras-Raf signaling), as well as binds to the unphosphorylated form of the ERBB2 protein and regulates ERBB2 function and localization [84, 85]. However, in our study we found that the differential expression of *ERBB2IP* in epithelial versus stromal tissue was also seen in normal reduction mammoplasty patients, so this difference in transcriptomes is not a cancer associated change (Table 2).

ECM ontologies genes were consistently highly enriched in stromal tissue versus epithelial and in invasive breast cancer versus normal reduction patients in all comparisons made among the transcriptomes. Using these data we were able to describe expression signatures which characterized the roles that tumor stroma and epithelium play in ECM biosynthesis and remodeling during invasion (Table 2). ECM related genes were over-represented in genes up-regulated in cancerous epithelial and stromal tissues versus normal tissue and in cancer stroma versus cancer epithelium. This is very interesting in light of observations of Roepman et al. [15] that showed in primary head and neck squamous cell carcinoma, transcriptomes of tumor and cancer associated stroma were characterized by up-regulation of stroma specific genes and inactivation of tumor cell specific genes [15].

The differential contribution of tumor stroma and tumor epithelium and the cross talk between the tissues needed for tumor cell invasion of the stroma has been demonstrated in cell biology and animal models as well as in clinical data [6, 11, 60, 86–99]. Using transcriptome analysis we developed a model of breast cancer invasion based on the molecular signatures of cancer stroma and epithelium (Fig. 4). Invasion occurs through the acquistion of a motile phenotype in tumor epithelial cells and a reactive phenotype in CAFs. Malignant epithelial cells acquire the molecular signature of a motile phenotype prior to invasion, with the down regulation of basement membrane proteins likely contributing to their ability to transverse this barrier during invasion (Fig. 4). Reactive stroma may form in response to mailgnant transformation of epithelial cells [100]. CAFs within reactive stroma are highly proliferative and express higher levels of ECM proteases and proteins [5, 100]. ECM proteases hydrolyze basement membrane proteins and thus completely dissolve the barrier giving motile tumor epithelial cells access to the stroma. Biosynthesis of ECM proteins result in an altered ECM scaffolding that tumor epithelial cells can track along for high speed movement as well as add structural support for angiogenesis. Changes in gene expression leading to or enhancing the motile phenotype of tumor epithelial are stimulated by the upregulation of *SPARC, POSTN, SSP1* and *TGFB* (Fig. 4). These protein products may stimulate tumor

cell motility directly or indirectly through induction of EMT, as with TGFB. The upregulation of TGFB in CAFs autoinduces their activation, and thus stimulates proliferation of cells and synthesis of ECM proteins and proteases. Thus the molecular signature of the tumor stroma shows that it plays a major role in the invasion of breast cancer.

Our model supports the literature that malignant cells in breast cancers move much faster than similarly transformed cells on tissue-culture plastic or in reconstituted extracellular matrix [101]. The faster mobility in vivo appears to involve the ability of malignant cancer cells to rapidly track along collagenous extracellular matrix fibers [101]. Further, our data supports the findings that as breast cancer progresses there is a loss of epithelial specific gene expression with the exception of increased expression of genes that encode ECM proteins and remodel the ECM. This is consistent with the literature in which in situ lesions were compared to fully invasive lesions and found that extracellular matrix remodeling, invasion and cell motility functions [37] were altered leading to an induction of a host stromal response in the transition of DCIS to IDC [102]. Future work related to these data will examine genes that did not fit into Gene Ontology categories and/or pathway groups as these genes may be important to progression and invasion of breast cancer.

## Conclusion

Transcriptome analysis of enriched populations of normal and cancerous stromal and epithelial tissues allowed us to begin to define the differential roles of tumor epithelium and stroma in ECM biosynthesis and remodeling and to develop a model for invasion of tumor epithelial cells into the surrounding stroma. Our data demonstrated that tumor cell invasion of the stroma likely proceeds through the down regulation of the laminin basement membrane proteins and cell-cell/cell-ECM adhesion molecules as well as cytoskeletal proteins; these changes in the transcritpome of the tumor epithelial results in a motile phenotype. Changes in the transcriptome of cancer stroma result in a reactive phenotype, characterized by changes in expression of genes responsible for ECM remodeling and the hydrolysis of the basement membrane as well as biosynthesis of an 'invasive' ECM scaffolding. Further we suggest that invasion of tumor epithelial cells into stroma following basement membrane hydrolysis proceeds through the induction of movement or tracking along ECM scaffolding upon SSP1, SPARC, POSTN and/or TGFB stimulation.

**Acknowledgements** This work was supported by a grant from the Breast Cancer Research Foundation; New York, N.Y. Microarray analysis was performed in the University of Vermont Microarray Facility and was supported in part by grant P30CA22435 from the NCI.

## References

1. Gupta GP, Massague J (2006) Cancer metastasis: building a framework. Cell 127:679–695
2. Janda E, Lehmann K, Killisch I, Jechlinger M, Herzig M, Downward J, Beug H, Grunert S (2002) Ras and TGF[beta] cooperatively regulate epithelial cell plasticity and metastasis: dissection of Ras signaling pathways. J Cell Biol 156:299–313
3. Chambers AF, Groom AC, MacDonald IC (2002) Dissemination and growth of cancer cells in metastatic sites. Nat Rev Cancer 2:563–572
4. Fidler IJ (2003) The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. Nat Rev Cancer 3:453–458
5. Kalluri R, Zeisberg M (2006) Fibroblasts in cancer. Nat Rev Cancer 6:392–401
6. Tlsty TD, Hein PW (2001) Know thy neighbor: stromal cells can contribute oncogenic signals. Curr Opin Genet Dev 11:54–59
7. de Visser K, Korets L, Coussens L (2005) De novo carcinogenesis promoted by chronic inflammation is B lymphocyte dependent. Cancer Res 7:411–423
8. Kim K, Li B, Winer J, Armanini M, Gillett N, Phillips H, Ferrara N (1993) Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumour growth in vivo. Nature 362:841–844
9. Condeelis J, Pollard JW (2006) Macrophages: obligate partners for tumor cell migration, invasion, and metastasis. Cell 124:263–266
10. Noel A, Foidart J-M (1998) The role of stroma in breast carcinoma growth in vivo. J Mammary Gland Biol Neoplasia V3:215–225
11. Tuxhorn JA, McAlhany SJ, Dang TD, Ayala GE, Rowley DR (2002) Stromal cells promote angiogenesis and growth of human prostate tumors in a differential reactive stroma (DRS) xenograft model. Cancer Res 62:3298–3307
12. Mueller MM, Fusenig NE (2004) friends or foes- bipolar effects of the tumour stroma in cancer. Nat Rev Cancer 4:839–849
13. Bissell MJ, Radisky D (2001) Putting tumours in context. Nat Rev Cancer 1:46–54
14. Liotta LA, Kohn EC (2001) The microenvironment of the tumour-host interface. Nature 411:375–379
15. Roepman P, de Koning E, van Leenen D, de Weger R, Kummer JA, Slootweg P, Holstege F (2006) Dissection of a metastatic gene expression signature into distinct components. Genome Biol 7:R117
16. Yamabuki T, Daigo Y, Kato T, Hayama S, Tsunoda T, Miyamoto M, Ito T, Fujita M, Hosokawa M, Kondo S, Nakamura Y (2006) Genome-wide gene expression profile analysis of esophageal squamous cell carcinomas. Int J Oncol 28(6):1375–1384
17. Nakamura N, Iijima T, Mase K, Furuya S, Kano J, Morishita Y, Noguchi M (2004) Phenotypic differences of proliferating fibroblasts in the stroma of lung adenocarcinoma and normal bronchus tissue. Cancer Sci 95:226–232
18. Finak G, Sadekova S, Pepin F, Hallett M, Meterissian S, Halwani F, Khetani K, Souleimanova M, Zabolotny B, Omeroglu A, Park M (2006) Gene expression signatures of morphologically normal breast tissue identify basal-like tumors. Breast Cancer Res 8:R58
19. Elston CW, Ellis IO (1991) Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 19:403–410

20. R Development Core Team (2006) R: a language and environment for statistical computing. http://www.R-project.org
21. Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini A, Sawitzki G, Smith C, Smyth G, Tierney L, Yang J, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80
22. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP (2003) Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 31:e15
23. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19:185–193
24. Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. Bioinformatics 23:257–258
25. Marisa MF, Jérôme T, Marie-Ange D, Ilaria Taddei-De La H, Jean Paul T, Marina AG (2005) Myoepithelial cells in the control of mammary development and tumorigenesis: data from genetically modified mice. J Mammary Gland Biol Neoplasia 10:211–219
26. Woodward TL, Mienaltowski AS, Modi RR, Bennett JM, Haslam SZ (2001) Fibronectin and the {{alpha}}5{beta}1 integrin are under developmental and ovarian steroid regulation in the normal mouse mammary gland. Endocrinology 142:3214–3222
27. Pujuguet P, Simian M, Liaw J, Timpl R, Werb Z, Bissell MJ (2000) Nidogen-1 regulates laminin-1-dependent mammary-specific gene expression. J Cell Sci 113:849–858
28. Nakano S, Iyama K, Ogawa M, Yoshioka H, Sado Y, Oohashi T, Ninomiya Y, (1999) Differential tissular expression and localization of type IV collagen alpha1(IV), alpha2(IV), alpha5(IV), and alpha6(IV) chains and their mRNA in normal breast and in benign and malignant breast tumors. Lab Invest 79:281–292
29. Silberstein GB, Daniel CW (1987) Reversible inhibition of mammary gland growth by transforming growth factor-beta. Science 237:291–293
30. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L, Nobel A, Parker J, Ewend MG, Sawyer LR, Wu J, Liu Y, Nanda R, Tretiakova M, Ruiz Orrico A, Dreher D, Palazzo JP, Perreard L, Nelson E, Mone M, Hansen H, Mullins M, Quackenbush JF, Ellis MJ, Olopade OI, Bernard PS, Perou CM (2006) The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7:96
31. de Jong JS, van Diest PJ, Baak JP (2000) Hot spot microvessel density and the mitotic activity index are strong additional prognostic indicators in invasive breast cancer. Histopathology 36:306–312
32. de Jong JS, van Diest PJ, Baak JP (2000) Number of apoptotic cells as a prognostic marker in invasive breast cancer. Br J Cancer 82:368–373
33. Liu S, Edgerton SM, Moore DH 2nd, Thor AD (2001) Measures of cell turnover (proliferation and apoptosis) and their association with survival in breast cancer. Clin Cancer Res 7:1716–1723
34. Mori I, Yang Q, Kakudo K (2002) Predictive and prognostic markers for invasive breast cancer. Pathol Int 52:186–194
35. Srinivas P, Abraham E, Ahamed I, Madhavan M, Vijayalakshmi NR, Nair MK, Balaram P (2002) Apoptotic index: use in predicting recurrence in breast cancer patients. J Exp Clin Cancer Res 21:233–238
36. Vakkala M, Lahteenmaki K, Raunio H, Paakko P, Soini Y (1999) Apoptosis during breast carcinoma progression. Clin Cancer Res 5:319–324
37. Abba M, Drake J, Hawkins K, Hu Y, Sun H, Notcovich C, Gaddis S, Sahin A, Baggerly K, Aldaz C (2004) Transcriptomic changes in human breast cancer progression as determined by serial analysis of gene expression. Breast Cancer Res 6:R499–R519
38. Grigoriadis A, Mackay A, Reis-Filho J, Steele D, Iseli C, Stevenson B, Jongeneel CV, Valgeirsson H, Fenwick K, Iravani M, Leao M, Simpson A, Strausberg R, Jat P, Ashworth A, Neville AM, O'Hare M (2006) Establishment of the epithelial-specific transcriptome of normal and malignant human breast cells based on MPSS and array expression data. Breast Cancer Res 8:R56
39. Turashvili G, Bouchal J, Baumforth K, Wei W, Dziechciarkova M, Ehrmann J, Klein J, Fridman E, Skarda J, Srovnal J, Hajduch M, Murray P, Kolar Z (2007) Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis. BMC Cancer 7:55
40. Folgueira M, Brentani H, Katayama M, Patrão D, Carraro D, Mourão Netto M, Barbosa E, Caldeira J, Abreu A, Lyra E, Kaiano J, Mota L, Campos A, Maciel M, Dellamano M, Caballero O, Brentani M (2006) Gene expression profiling of clinical stages II and III breast cancer. Braz J Med Biol Res 39:1101–1113
41. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale A-L, Brown PO, Botstein D (2000) Molecular portraits of human breast tumours. Nature 406:747–752
42. Mesnil M, Crespin S, Avanzo J-L, Zaidan-Dagli M-L (2005) Defective gap junctional intercellular communication in the carcinogenic process. Biochim Biophys Acta 1719:125–145
43. Stadler CR, Knyazev P, Bange J, Ullrich A (2006) FGFR4 GLY388 isotype suppresses motility of MDA-MB-231 breast cancer cells by EDG-2 gene repression. Cell Signal 18:783–794
44. Friedl P, Wolf K (2003) Proteolytic and non-proteolytic migration of tumour cells and leucocytes. Biochem Soc Symp 70:277–285
45. Clark EA, Goiub TR, Lander ES, Hynes RO (2000) Genomic analysis of metastasis reveals an essential role for RhoC. Nature 406:532–535
46. Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordon-Cardo C, Guise TA, Massague J (2003) A multigenic program mediating breast cancer metastasis to bone. Cancer Cell 3:537–549
47. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, Viale A, Olshen AB, Gerald WL, Massague J (2005) Genes that mediate breast cancer metastasis to lung. Nature 436:518–524
48. Bradshaw A, Sage E (2001) SPARC, a matricellular protein that functions in cellular differentiation and tissue response to injury. J Clin Invest 107:1049–1054
49. Barth PJ, Moll R, Ramaswamy A (2005) Stromal remodeling and SPARC (secreted protein acid rich in cysteine) expression in invasive ductal carcinomas of the breast. Virchows Arch 446:532–536
50. Robinson S, Silberstein G, Daniel CW (1992) Evidence supporting a role for TGF-beta isoforms in growth regulation and functional differentiation of the mouse mammary gland. In: Mechanisms regulating lactation and infant nutrient utilization. Wiley-Liss, Inc, New York, pp 43–52
51. Faler B, Macsata R, Plummer D, Mishra L, Sidawy A (2006) Transforming growth factor-beta and wound healing. Perspect Vasc Surg Endovasc Ther 18:55–62
52. O'Kane S, Ferguson MWJ (1997) Transforming growth factor [beta]s and wound healing. Int J Biochem Cell Biol 29:63–78
53. Wahl SM (2007) Transforming growth factor-[beta]: innately bipolar. Curr Opin Immunol 19:55–62

54. Fleisch MC, Maxwell CA, Barcellos-Hoff M-H (2006) The pleiotropic roles of transforming growth factor beta in homeostasis and carcinogenesis of endocrine organs. Endocrinology 13:379–400
55. Muraoka-Cook RS, Dumont N, Arteaga CL (2005) Dual role of transforming growth factor beta in mammary tumorigenesis and metastatic progression. Clin Cancer Res 11:937s–943s
56. Roberts A, Wakefield L (2003) The two faces of transforming growth factor beta in carcinogenesis. Proc Natl Acad Sci USA 100:8621–8623
57. Minond D, Lauer-Fields JL, Cudic M, Overall CM, Pei D, Brew K, Moss ML, Fields GB (2007) Differentiation of secreted and membrane-type matrix metalloproteinase activities based on substitutions and interruptions of triple-helical sequences. Biochemistry 46(12):3724–3733
58. Merrell MA, Ilvesaro JM, Lehtonen N, Sorsa T, Gehrs B, Rosenthal E, Chen D, Shackley B, Harris KW, Selander KS (2006) Toll-like receptor 9 agonists promote cellular invasion by increasing matrix metalloproteinase activity. Mol Cancer Res 4:437–447
59. Leivonen SK, Ala-aho R, Koli K, Grenman R, Peltonen J, Kahari VM (2006) Activation of Smad signaling enhances collagenase-3 (MMP-13) expression and invasion of head and neck squamous carcinoma cells. Oncogene 25:2588–2600
60. Lafleur M, Drew A, de Sousa E, Blick T, Bills M, Walker E, Williams E, Waltham M, Thompson E (2005) Upregulation of matrix metalloproteinases (MMPs) in breast cancer xenografts: a major induction of stromal MMP-13. Int J Cancer 114:544–554
61. Andreas MS, Bartosz A, Rolf M, Toshiyuki Y, Mehdorn HM, Janka H-F (2007) Differential expression of matrix metalloproteinases in brain- and bone-seeking clones of metastatic MDA-MB-231 breast cancer cells. J Neurooncol 81:39–48
62. Luukkaa M, Vihinen P, Kronqvist P, Vahlberg T, Pyrhönen S, Kähäri V, Grénman R (2006) Association between high collagenase-3 expression levels and poor prognosis in patients with head and neck cancer. Head Neck 28:225–234
63. Garin-Chesa P, Old LJ, Rettig WJ (1990) Cell surface glycoprotein of reactive stromal fibroblasts as a potential antibody target in human epithelial cancers. PNAS 87:7235–7239
64. Niedermeyer J, Enenkel B, Park JE, Lenter M, Rettig WJ, Damm K, Schnapp A (1998) Mouse fibroblast-activation protein – conserved Fap gene organization and biochemical function as a serine protease. Eur J Biochem 254:650–654
65. Scanlan MJ, Raj BKM, Calvo B, Garin-Chesa P, Sanz-Moncasi MP, Healey JH, Old LJ, Rettig WJ (1994) Molecular cloning of fibroblast activation protein {alpha}, a member of the serine protease family selectively expressed in stromal fibroblasts of epithelial cancers. Proc Natl Acad Sci USA 91:5657–5661
66. Park JE, Lenter MC, Zimmermann RN, Garin-Chesa P, Old LJ, Rettig WJ (1999) Fibroblast activation protein, a dual specificity serine protease expressed in reactive human tumor stromal fibroblasts. J Biol Chem 274:36505–36512
67. Rettig WJ, Garin-Chesa P, Beresford HR, Oettgen HF, Melamed MR, Old LJ (1988) Cell-surface glycoproteins of human sarcomas: differential expression in normal and malignant tissues and cultured cells. Proc Natl Acad Sci USA 85:3110–3114
68. Tahtis K, Lee F-T, Wheatley JM, Garin-Chesa P, Park JE, Smyth FE, Obata Y, Stockert E, Hall CM, Old LJ, Rettig WJ, Scott AM (2003) Expression and targeting of human fibroblast activation protein in a human skin/severe combined immunodeficient mouse breast cancer xenograft model. Mol Cancer Res 2:729–737
69. Scott IC, Blitz IL, Pappano WN, Imamura Y, Clark TG, Steiglitz BM, Thomas CL, Maas SA, Takahara K, Cho KWY, Greenspan DS (1999) Mammalian BMP-1/tolloid-related metalloproteinases, including novel family member mammalian tolloid-like 2, have differential enzymatic activities and distributions of expression relevant to patterning and skeletogenesis. Dev Biol 213:283–300
70. Veitch DP, Nokelainen P, McGowan KA, Nguyen T-T, Nguyen NE, Stephenson R, Pappano WN, Keene DR, Spong SM, Greenspan DS, Findell PR, Marinkovich MP (2003) Mammalian tolloid metalloproteinase, and not matrix metalloprotease 2 or membrane type 1 metalloprotease, processes laminin-5 in keratinocytes and skin. J Biol Chem 278:15661–15668
71. Ishikawa T, Kamiyama M, Tani-Ishii N, Suzuki H, Ichikawa Y, Hamaguchi Y, Momiyama N, Shimada H (2001) Inhibition of osteoclast differentiation and bone resorption by cathepsin K antisense oligonucleotides. Mol Carcinog 32:84–91
72. Lipton A (2005) New therapeutic agents for the treatment of bone diseases. Expert Opin Biol Ther 5:817–832
73. Turk V, Turk B, Guncar G, Turk D, Kos J (2002) Lysosomal cathepsins: structure, role in antigen processing and presentation, and cancer. Adv Enzyme Regul 42:285–303
74. Gocheva V, Zeng W, Ke D, Klimstra D, Reinheckel T, Peters C, Hanahan D, Joyce JA (2006) Distinct roles for cysteine cathepsin genes in multistage tumorigenesis. Genes Dev 20:543–556
75. Micke P, Kappert K, Ohshima M, Sundquist C, Scheidl S, Lindahl P, Heldin C-H, Botling J, Ponten F, Ostman A (2007) In situ identification of genes regulated specifically in fibroblasts of human basal cell carcinoma. J Invest Dermatol 127(6):1516–1523
76. Cockett M, Murphy G, Birch M, O'Connell J, Crabbe T, Millican A, Hart I, Docherty A (1998) Matrix metalloproteinases and metastatic cancer. Biochem Soc Symp 63:295–313
77. Twal WO, Czirok A, Hegedus B, Knaak C, Chintalapudi MR, Okagawa H, Sugi Y, Argraves WS (2001) Fibulin-1 suppression of fibronectin-regulated cell adhesion and motility. J Cell Sci 114:4587–4598
78. Hayashido Y, Lucas A et al (1998) Estradiol and fibulin-1 inhibit motility of human ovarian- and breast-cancer cells induced by fibronectin. Int J Cancer 75:654–658
79. Hubmacher D, Tiedemann K, Reinhardt DP, Gerald PS (2006) Fibrillins: from biogenesis of microfibrils to signaling functions. In: Current topics in developmental biology. Academic Press, pp 93–123
80. Weir M, Oppizzi M, Henry M, Onishi A, Campbell K, Bissell M, Muschler J (2006) Dystroglycan loss disrupts polarity and beta-casein induction in mammary epithelial cells by perturbing laminin anchoring. J Cell Sci 119:4047–4058
81. Barresi R, Campbell K (2006) Dystroglycan: from biosynthesis to pathogenesis of human disease. J Cell Sci 119:199–207
82. Young KG, Pinheiro B, Kothary R (2006) A Bpag1 isoform involved in cytoskeletal organization surrounding the nucleus. Exp Cell Res 312:121–134
83. Roper K, Gregory SL, Brown NH (2002) The 'Spectraplakins': cytoskeletal giants with characteristics of both spectrin and plakin families. J Cell Sci 115:4215–4225
84. Carraway KL, Sweeney C (2006) Co-opted integrin signaling in ErbB2-induced mammary tumor progression. Cancer Cell 10:93–95
85. Kolch W (2003) Erbin: sorting out ErbB2 receptors or giving Ras a break? Science pe37
86. Singer C, Kronsteiner N, Marton E, Kubista M, Cullen K, Hirtenlehner K, Seifert M, Kubista E (2002) MMP-2 and MMP-9 expression in breast cancer-derived human fibroblasts is differentially regulated by stromal-epithelial interactions. Br Cancer Res Treat 72:69–77
87. Basset P, Wolf C, Chambon P (1993) Expression of the stromelysin-3 gene in fibroblastic cells of invasive carcinomas of the breast and other human tissues: a review. Br Cancer Res Treat 24:185–193

88. Okada A, Bellocq J, Rouyer N, Chenard M, Rio M, Chambon P, Basset P (1995) Membrane-type matrix metalloproteinase (MT-MMP) gene is expressed in stromal cells of human colon, breast, and head and neck carcinomas. Proc Natl Acad Sci USA 92:2730–2734
89. Polette M, Gilles C, Marchand V, Lorenzato M, Toole B, Tournier J-M, Zucker S, Birembaut P (1997) Tumor collagenase stimulatory factor (TCSF) expression and localization in human lung and breast cancers. J Histochem Cytochem 45:703–710
90. Tang Y, Kesavan P, Nakada MT, Yan L (2004) Tumor-stroma interaction: positive feedback regulation of extracellular matrix metalloproteinase inducer (EMMPRIN) expression and matrix metalloproteinase-dependent generation of soluble EMMPRIN. Mol Cancer Res 2:73–80
91. Bhowmick NA, Chytil A, Plieth D, Gorska AE, Dumont N, Shappell S, Washington MK, Neilson EG, Moses HL (2004) TGF-{beta} signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia. Science 303:848–851
92. Joesting MS, Perrin S, Elenbaas B, Fawell SE, Rubin JS, Franco OE, Hayward SW, Cunha GR, Marker PC (2005) Identification of SFRP1 as a candidate mediator of stromal-to-epithelial signaling in prostate cancer. Cancer Res 65:10423–10430
93. Ronnov-Jessen L, Petersen OW, Bissell MJ (1996) Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. Physiol Rev 76:69–125
94. Barcellos-Hoff MH, Ravani SA (2000) Irradiated mammary gland stroma promotes the expression of tumorigenic potential by unirradiated epithelial cells. Cancer Res 60:1254–1260
95. Barcellos-Hoff MH (1993) Radiation-induced transforming growth factor beta and subsequent extracellular matrix reorganization in murine mammary gland. Cancer Res 53:3880–3886
96. Cheng N, Bhowmick NA, Chytil A, Gorksa AE, Brown KA, Muraoka R, Arteaga CL, Neilson EG, Hayward SW, Moses HL (2005) Loss of TGF-beta type II receptor in fibroblasts promotes mammary carcinoma growth and invasion through upregulation of TGF-alpha-, MSP- and HGF-mediated signaling networks. Oncogene 24:5053–5068
97. Moinfar F, Man YG, Arnould L, Bratthauer GL, Ratschek M, Tavassoli FA (2000) Concurrent and independent genetic alterations in the stromal and epithelial cells of mammary carcinoma: implications for tumorigenesis. Cancer Res 60: 2562–2566
98. Parrott JA, Nilsson E, Mosher R, Magrane G, Albertson D, Pinkel D, Gray JW, Skinner MK (2001) Stromal-epithelial interactions in the progression of ovarian cancer: influence and source of tumor stromal cells. Mol Cell Endocrinol 175:29–39
99. Fukumura D, Xavier R, Sugiura T, Chen Y, Park E-C, Lu N, Selig M, Nielsen G, Taksir T, Jain RK, Seed B (1998) Tumor induction of VEGF promoter activity in stromal cells. Cell 94:715–725
100. Radisky E, Radisky D (2007) Stromal induction of breast cancer: inflammation and invasion. Rev Endocr Metab Disord 8(3):279–287
101. Condeelis J, Segall JE (2003) Intravital imaging of cell movement in tumours. Nat Rev Cancer 3:921–930
102. Porter D, Krop I, Nasser S, Sgroi D, Kaelin C, Marks J, Riggins G, Polyak K (2001) A SAGE (Serial Analysis of Gene Expression) view of breast tumor progression. Cancer Res 61:5697–5702
103. Elston CW, Ellis IO (2002) Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. C. W. Elston & I. O. Ellis. Histopathology 1991; 19:403–410. Histopathology 41:151