Molecular characterization of the tumor microenvironment in breast cancer

Minna Allinen,[1,3] Rameen Beroukhim,[1,3,10] Li Cai,[2,10] Cameron Brennan,[1,10] Jaana Lahti-Domenici,[1] Haiyan Huang,[2,6,11] Dale Porter,[1,3] Min Hu,[1,3] Lynda Chin,[1,4] Andrea Richardson,[5,7] Stuart Schnitt,[5,8] William R. Sellers,[1,3,9] and Kornelia Polyak[1,3]

[1]Department of Medical Oncology
[2]Department of Research Computing
Dana-Farber Cancer Institute, Boston, Massachusetts 02115
[3]Department of Medicine
[4]Deparment of Dermatology
[5]Department of Pathology
Harvard Medical School, Boston, Massachusetts 02115
[6]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115
[7]Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115
[8]Department of Pathology, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02115
[9]Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142
[10]These authors contributed equally to this work.
[11]Present address: Department of Statistics, University of California, Berkeley, Berkeley, California 94720.
*Correspondence: kornelia_polyak@dfci.harvard.edu

Summary

Here we describe the comprehensive gene expression profiles of each cell type composing normal breast tissue and in situ and invasive breast carcinomas using serial analysis of gene expression. Based on these data, we determined that extensive gene expression changes occur in all cell types during cancer progression and that a significant fraction of altered genes encode secreted proteins and receptors. Despite the dramatic gene expression changes in all cell types, genetic alterations were detected only in cancer epithelial cells. The CXCL14 and CXCL12 chemokines overexpressed in tumor myoepithelial cells and myofibroblasts, respectively, bind to receptors on epithelial cells and enhance their proliferation, migration, and invasion. Thus, chemokines may play a role in breast tumorigenesis by acting as paracrine factors.

Introduction

Breast cancer is the most commonly identified and one of the deadliest neoplasms in women in Western countries. The recent trend toward improvement in breast cancer mortality rate is largely due to increased diagnosis of early stage disease, while our therapeutic options for advanced stage breast carcinomas are still fairly limited. Thus, there is a need to better understand the molecular basis of breast cancer initiation and progression and to use this knowledge for the design of targeted, molecular-based therapies. In the past few years, newly developed technologies such as microarrays and SAGE (serial analysis of gene expression) have enabled us to analyze molecular differences between normal and cancer cells at a genome-wide level in comprehensive and unbiased ways (Schena et al., 1995; Velculescu et al., 1995). Using these approaches, the molecular-based classification of breast cancer has become a reality, and molecular signatures correlating with metastatic behavior and clinical outcome have been identified (Ramaswamy et al., 2003; Sorlie et al., 2001; van 't Veer et al., 2002; van de Vijver et al., 2002). However, since most of these analyses were performed using bulk tissue samples that are composed of multiple cell types or purified tumor epithelial cells, the specific contribution of epithelial and stromal cells to these tumor classifiers and prognostic signatures is unknown. Similarly, in the past decades the major focus of cancer research has been the transformed tumor cell itself, while the role of the cellular microenvironment in tumorigenesis has not been widely explored. Early studies demonstrated the ability of stromal tissues to regulate the growth and differentiation state of breast cancer cells (DeCosse

SIGNIFICANCE

Despite compelling cell biological studies and histopathological observations incriminating myoepithelial and stromal cells in tumorigenesis, our knowledge of the genes that mediate changes in the tumor microenvironment and interactions among various cell types in breast cancer and their role in tumorigenesis is limited. Similarly, the occurrence and role of genetic changes in stromal cells are undefined. Here, we describe a comprehensive molecular characterization of each cell type composing normal breast tissue and in situ and invasive breast carcinomas. We identified several genes as potential mediators of epithelial-stromal/myoepithelial cell interactions, including the CXCL12 and CXCL14 chemokines. These data should therefore provide a valuable resource for future basic and clinical studies addressing the role of epithelial-stromal/myoepithelial cell interactions in breast cancer.

et al., 1973, 1975), and several recent in vivo and in vitro studies have demonstrated that the growth, differentiation, invasive behavior, and polarity of normal mammary epithelial cells and breast carcinomas are influenced by surrounding stromal cells including fibroblasts, myofibroblasts, leukocytes, and myoepithelial cells (Bissell and Radisky, 2001; Radisky et al., 2001; Tlsty, 2001). In addition, certain histopathological features of breast tumors, including lymphocytic infiltration, fibrosis, and angio- and lymphangiogenesis, have proven prognostic significance. Despite these convincing data implicating a role for the tumor microenvironment in breast tumorigenesis, our understanding of the genes mediating cellular interactions and paracrine regulatory circuits among various cell types in normal and cancerous breast tissue and their role in breast tumorigenesis is limited.

In the past few years, the role of the cellular microenvironment in tumorigenesis has become an intense area of research. This is in part due to studies demonstrating that genetic abnormalities, such as loss of heterozygosity (LOH), occur not only in cancer cells, but in stromal cells as well (Kurose et al., 2001, 2002; Lakhani et al., 1998; Moinfar et al., 2000). However, no genes presumably targeted by these genetic events in stromal cells have been identified; thus, their role in breast tumorigenesis is still unknown.

As a consequence of studies focusing almost exclusively on cancer cells, nearly all of the currently used cancer therapeutic agents target the cancer cells that, due to their inherent genomic instability, frequently acquire therapeutic resistance (Rajagopalan et al., 2003). In part due to frequent therapeutic failures during the course of treatment of advanced stage tumors, increasing emphasis has been placed on targeting various stromal cells, particularly endothelial cells, via therapeutic interventions. Since these cells are thought to be normal and genetically stable, they are less likely to develop acquired resistance to cancer therapy. Thus, isolating and characterizing each cell type (epithelial, myoepithelial, and various stromal cells) comprising nonmalignant and cancerous breast tissue would not only help us to understand the role these cells play in breast tumorigenesis, but would likely give us new molecular targets for cancer intervention and treatment.

Results

Purification of all cell types present in breast tissue

To determine the molecular profile of each cell type that, together, compose the breast tissue and to identify autocrine and paracrine interactions that may play a role in breast tumor progression, we developed a purification procedure that allows the isolation of pure cell populations from normal breast tissue and from in situ (ductal carcinoma in situ, DCIS) and invasive breast carcinomas (Figure 1A). We utilized cell type-specific cell surface markers and magnetic beads for the rapid sequential isolation of the various cell types. We used the BerEP4 antigen restricted to epithelial cells, the CD45 panleukocyte marker, and the P1H12 antibody that specifically recognizes endothelial cells. The CD10 antigen is present in myoepithelial cells and myofibroblasts, but also in some leukocytes. Thus, to minimize the crosscontamination of these different cell types, in the case of normal (N-MYOEP-1) and DCIS breast tissue, myoepithelial cells were isolated from organoids (breast ducts), while in invasive tumors we first removed the leukocytes prior to capturing the myofibroblasts using CD10 beads. Several recent studies reported that some morphologically distinct myoepithelial cells lack CD10 and other myoepithelial cell markers (Zhang et al., 2003). Thus, due to the use of CD10 beads for the isolation of myoepithelial cells, a subset of myoepithelial cells may have been excluded from our study. We were not able to identify an antibody that would specifically recognize fibroblasts and allow their purification; thus, we used the unbound fraction following the removal of all other cell types as a fibroblast-enriched "stromal" fraction. A detailed description of the purification method is described in the Supplemental Data (http://www.cancercell.org/cgi/content/full/6/1/17/DC1). Since this protocol includes sequential enzymatic digestion of the tissue, the possibility that the expression of some genes could be altered due to the procedure cannot be excluded. However, since we were able to verify the SAGE data by alternative methods using unprocessed tissue (Figure 3), these changes (if any) are likely to be minimal. The success of the purification method and the purity of each cell fraction were confirmed by performing RT-PCR on a small fraction of the isolated cells using cell type-specific genes (Figure 1B). The remaining portion of the cells ($\sim$10,000–100,000 cells, depending on the sample) was used for the generation of micro-SAGE libraries following previously described protocols (Porter et al., 2001, 2003a) and for the isolation of genomic DNA to be used for array comparative genomic hybridization (aCGH) and single nucleotide polymorphism (SNP) array studies. We have generated SAGE libraries from epithelial and myoepithelial cells (myofibroblasts from invasive tumors), infiltrating lymphocytes, endothelial cells, and fibroblasts (stroma) from one normal breast reduction tissue, two different DCIS, and three invasive breast tumors. Not all libraries were generated from all cases due to our inability to obtain sufficient amounts of purified cells. In addition, we also included a fibroadenoma and a phyllodes tumor in our SAGE analyses. Fibroadenomas are the most common benign breast tumors that are not considered to progress to malignancy despite genetic changes detected in the stromal (but not epithelial) cells (Amiel et al., 2003). Phyllodes tumors, on the other hand, are rare fibroepithelial tumors that are usually benign but can recur and progress to malignant sarcomas. Initially, phyllodes tumors were considered stromal neoplasms, but recent molecular studies demonstrating (frequently discordant) genetic alterations in both epithelial and stromal cells suggests that phyllodes tumors may represent a true clonal co-evolution of malignant epithelial and stromal cells (Sawyer et al., 2000, 2002). A detailed description of the tissue samples and the SAGE libraries is included in the Supplemental Data online. Analysis of the SAGE data confirmed that the cell purification procedure worked well, since several genes known to be specific for a particular cell type were present in the appropriate SAGE libraries. For example, cytokeratins 8 and 19, E-cadherin, HIN-1, and CD24 were highly specific for epithelial cells (HIN-1 only for normal epithelial cells); myofibroblast and myoepithelial cells demonstrated high levels of smooth muscle actin, various extracellular matrix proteins including collagens, and matrix metalloproteinases; and leukocyte libraries had the highest levels of several chemokines and lysozyme (Table 1 and Supplemental Table S1). In general, SAGE libraries prepared from the same cell type purified from different tissue samples were highly similar to each other, although there were differences as well, likely due to variability among patients and also slight variability

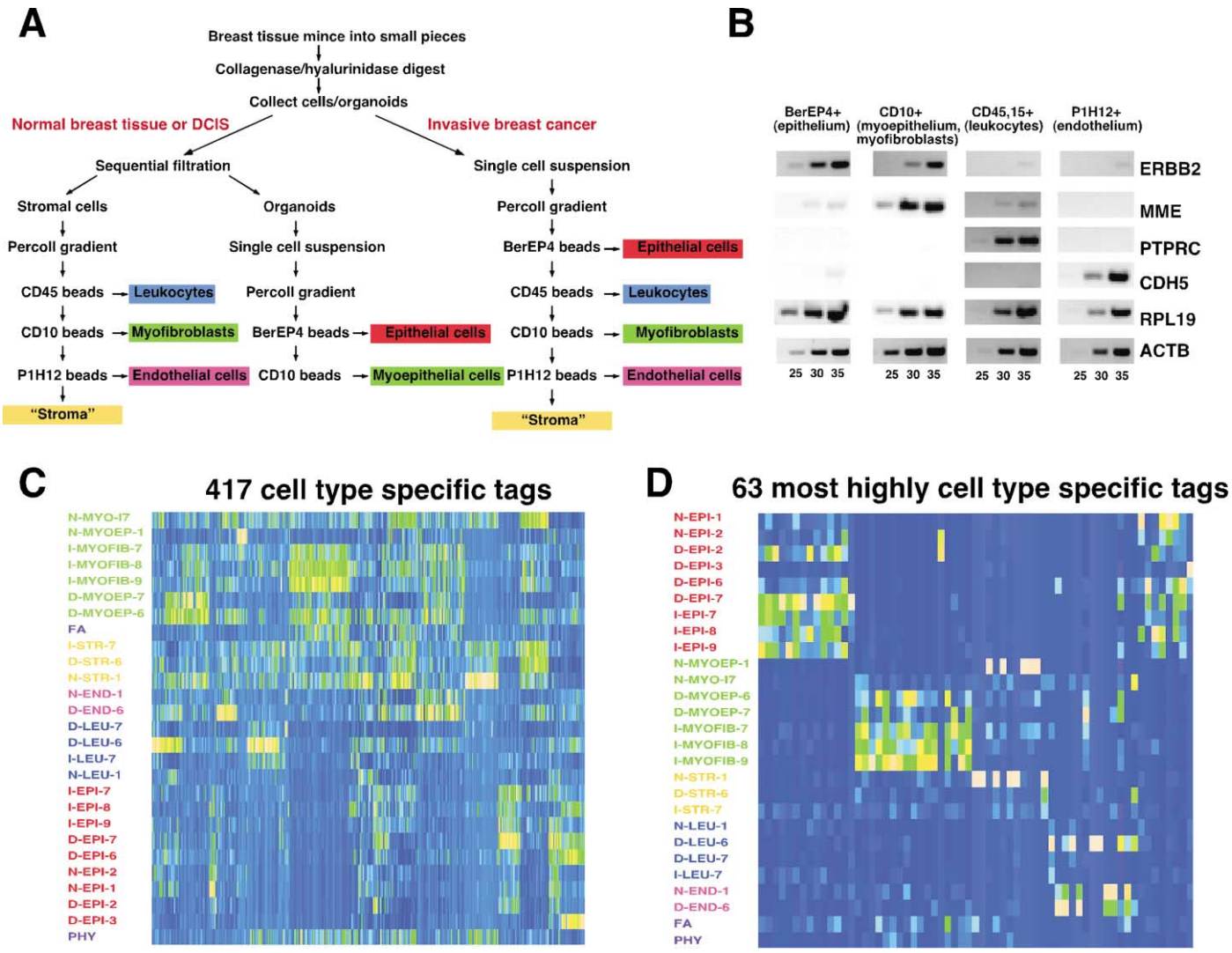

Figure 1. Isolation and characterization of each cell type comprising normal and cancerous breast tissue

A: Schematic outline of tissue fractionation and sequential purification of the various cell types from normal breast tissue and in situ and invasive breast carcinomas. The procedure is described in detail in the online Supplemental Data.

B: RT-PCR analysis of each cell fraction isolated from DCIS-7 using known cell type-specific genes to confirm the purity of the cells and integrity of the mRNA. MME (CD10) is highly specifically expressed in $CD10^+$ myoepithelial cells and myofibroblasts, PTPRC (CD45) in leukocytes, and CDH5 (endothelial cadherin) in endothelial cells. Although ERBB2 is not an absolutely epithelial cell-specific gene, its abundance is highest in luminal epithelial cells. PCR was performed at 25, 30, and 35 cycles. Genes expressed at equal levels in all cell types; β-actin (ACTB) and ribosomal protein L19 (RPL19) were used as controls.

C: Heat map depicting the relatedness of the different SAGE libraries based on 417 cell type-specific tags. Color scheme: blue, downregulated (low tag counts); green, mean tag counts; yellow, upregulated (high tag counts). The names of the SAGE libraries prepared from epithelial cells are in red, myoepithelial cells and myofibroblasts in green, stroma in yellow, leukocytes in blue, endothelial cells in pink, and fibroadenoma and phyllodes tumor (stroma fraction) in purple. A detailed description of the SAGE libraries and tissue samples is included in Supplemental Data.

D: Heat map depicting the relatedness of the different SAGE libraries based on the 63 most highly cell type-specific tags. Color scheme and SAGE library names are described as above.

in the purification procedure itself (see Supplemental Data for more details).

Comprehensive gene expression profile of each cell type

Based on statistical methods developed for the analysis of SAGE data (see Experimental Procedures, Supplemental Data; Cai et al., 2004), we identified genes that are specifically expressed in a particular cell type and tumor progression stage (Tables 1 and 2 and Supplemental Tables S1–S15). Genes were defined as specific for a particular cell type if the average tag number in all the SAGE libraries generated from the selected cell type was statistically significantly ($p < 0.02$) different from all other cell types. For the purpose of these comparisons, we considered myoepithelial cells and myofibroblasts as one group due to their high degree of similarity, although there are genes that are specific for myoepithelial cells and myofibroblasts, respectively (Ronnov-Jessen et al., 1996). Using these criteria,

Table 1. List of 63 most highly cell type-specific tags and corresponding genes

SAGE tag	M-myo-17	M-myop-1	I-myofib-7	I-myofib-8	I-myofib-9	D-myoep-7	D-myoep-6	I-str-7	D-str-6	M-str-1	M-end-1	D-end-6	D-leu-7	D-leu-6	I-leu-7	M-leu-1	I-epi-8	I-epi-9	D-epi-7	D-epi-6	M-epi-2	D-epi-2	D-epi-3	Pdhy	PA	Unigene	Gene description
CCTCCAGCTA	0	5	9	6	0	0	2	28	0	10	8	0	2	4	31	11	118	72	124	159	32	28	62	43	14	3	25	356123	KRT8 keratin 8
GACATCAAGT	0	0	5	0	0	0	15	0	0	4	9	0	5	9	26	11	73	64	59	153	48	15	18	55	2	0	5	309517	KRT19 keratin 19
TGTGGGTGCT	0	5	2	0	0	0	3	3	0	2	0	0	0	0	4	0	11	17	25	49	83	14	15	14	5	0	5	194657	CDH1 cadherin 1, type 1, E-cadherin
AGGAAGGAAC	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	18	0	2	24	90	0	0	3	7	0	3	446352	ERBB2
CTGGCCCTCG	0	0	0	9	0	0	0	0	0	0	4	0	0	2	3	4	33	149	74	74	10	62	163	39	6	0	5	350470	TFF1 trefoil factor 1
CTCCACCCGA	2	0	0	19	0	0	0	5	0	4	8	0	0	8	12	38	43	297	51	30	25	3	19	284	11	3	50	82961	TFF3 trefoil factor 3
AAGCCTCGCG	0	0	2	2	0	0	0	0	0	3	2	0	0	3	0	7	0	24	0	0	7	19	69	0	0	0	0	62492	SCGB3A1 secretoglobin, family 3A, member 1 (HIN-1)
CTTCCTGTGA	0	0	0	0	0	0	0	0	0	5	0	0	0	6	0	0	2	5	0	5	67	98	272	7	10	0	0	348419	LOC118430 small breast epithelial mucin
AAGAAAACCT	0	0	0	0	0	0	0	2	0	0	0	0	0	0	3	0	22	10	19	22	2	2	8	16	0	0	3	100686	BCMP11 breast cancer membrane protein 11
ATTTTCTAAA	0	0	0	8	0	0	0	2	0	0	0	0	0	0	3	0	8	68	13	5	6	3	2	25	0	0	3	226391	AGR2 anterior gradient 2 homolog (Xenopus laevis)
CGGACTCACT	0	2	3	2	2	0	0	0	2	4	3	2	0	0	0	9	23	13	89	12	0	3	11	3	0	3	300446	SPARD10 START domain containing 10
GGAACAAACA	0	0	3	0	0	0	11	0	8	11	0	6	6	9	14	62	7	129	94	122	62	30	57	3	0	13	375108	CD24 CD24 antigen
AATATGTGGG	13	9	7	17	9	2	0	29	6	6	3	6	0	0	14	0	89	56	80	112	2	2	6	235	4	8	12	98664	BPA-1 mRNA for brain peptide A1
GCACTCTGGA	0	0	0	4	0	0	0	2	0	0	6	3	0	0	5	7	5	25	39	23	31	14	56	11	7	0	0	439027	BDNF brain-derived neurotrophic factor
GCAGTGCCCT	2	0	0	0	0	0	0	0	0	3	5	0	0	4	0	2	8	26	14	25	11	3	3	32	5	0	11	396783	SLC9A3R1 solute carrier family 9, isoform 3 regulatory factor 1
TGGGGTTCTT	0	0	0	0	0	0	0	0	0	0	4	0	0	2	0	0	8	5	0	84	0	0	0	0	0	0	0	272499	DHRS2 dehydrogenase/reductase (SDR family) member 2
ATGCTCAGCC	0	0	4	2	0	0	0	0	0	0	0	0	0	0	3	0	9	12	86	48	9	2	4	6	0	0	0	98306	KIAA1862 KIAA1862 protein
TTGCGTTGCG	0	0	0	0	0	0	0	0	0	0	4	0	0	2	0	3	0	0	0	4	16	0	0	89	0	0	0	no match	
TCTCCATACC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	137	0	183	0	0	0	0	no match	
GATGTGCACG	0	339	41	5	0	0	5	19	2	0	0	0	0	0	3	0	0	0	0	0	0	6	4	0	6	0	2	355214	KRT14 keratin 14
GACCAGCAGA	0	0	40	15	24	18	44	0	0	3	13	3	4	6	0	0	0	0	0	0	0	0	0	0	0	0	4	137569	TP73L tumor protein p73-like
TTAAATAGCA	6	0	57	80	181	2	14	11	2	6	8	0	0	0	0	3	0	2	10	4	0	0	0	0	0	18	19	172928	COL1A1 collagen, type I, alpha 1
CCGGGGGAGC	3	0	43	52	104	46	55	6	0	7	17	0	0	0	4	0	0	6	2	2	2	0	0	0	0	18	10	172928	COL1A1 collagen, type I, alpha 1
GACTTTGGAA	8	0	18	33	53	15	100	21	6	7	0	0	2	0	0	0	2	0	0	2	0	0	0	0	0	18	27	172928	COL1A1 collagen, type I, alpha 1
TGGAAATGAA	4	0	11	16	18	3	24	4	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	15	172928	COL1A1 collagen, type I, alpha 1
CGGGGTGGCC	8	0	22	11	9	81	22	5	0	2	0	0	4	2	2	0	0	0	3	0	3	4	0	0	3	0	0	1584	COMP cartilage oligomeric matrix protein
TGGAAGCAGA	0	0	42	9	0	28	0	4	0	0	2	0	0	0	0	0	0	0	0	4	0	0	0	0	0	9	0	1584	COMP cartilage oligomeric matrix protein
CTGTCAGCGT	5	0	70	34	107	12	29	5	2	3	0	0	3	6	0	0	2	3	4	0	0	0	0	0	9	0	0	283713	CTHRC1 collagen triple helix repeat containing 1
CAGGAGACCC	0	0	33	51	102	0	8	8	2	0	0	0	3	2	0	0	3	2	0	0	0	0	0	0	0	0	0	143791	MMP11 matrix metalloproteinase 11 (stromelysin 3)
TCCCTACCGA	0	0	10	15	22	0	4	0	0	2	0	3	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	367877	MMP2 matrix metalloproteinase 2
TGGAAGCAGA	0	0	42	9	0	28	0	4	0	0	2	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	415041	THBS4 thrombospondin 4
CGGAAGGGAC	0	52	6	0	0	0	0	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13144	OSMDL2 OSM1-like 2
CTTCCTTGCC	8	785	179	34	4	7	63	19	7	27	0	2	0	3	8	0	6	0	17	6	3	5	15	2	33	0	4	449630	HBA2 hemoglobin, alpha 2
GGGGAAATCG	96	22	57	103	112	228	177	19	59	75	71	120	30	330	32	34	59	88	149	188	151	41	90	38	22	45	96	446574	TMSB10 thymosin, beta 10
TATTTTCACA	3	0	21	19	31	4	5	2	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1321331	Transcribed sequences
CCACGGGATT	20	0	63	164	78	23	168	17	10	0	19	16	2	14	0	4	0	0	8	5	0	0	0	0	0	0	68	no match	
GGTCTTCAAG	0	0	5	23	27	0	7	2	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	no match	
GTGCGCCGGA	0	40	7	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	no match	
GAGCTCGGAA	0	0	0	0	0	0	0	0	0	2	0	0	0	0	33	0	0	0	0	0	0	0	0	0	0	0	0	73875	FAH fumarylacetoacetate hydrolase
GAGCTCGAAA	0	0	0	0	0	0	0	0	0	2	0	0	0	0	33	0	0	0	0	0	0	0	0	0	0	0	0	20950	LHPP phospholysine phosphohistidine inorganic pyrophosphate phosphatase
GAGAAATCGT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	33	0	0	0	0	0	0	0	0	0	0	0	0	234734	LYZ lysozyme
AACGGGGCCC	2	0	2	0	0	0	0	0	0	0	0	0	0	0	17	4	2	0	0	0	0	0	0	0	0	0	0	80420	CX3CL1 chemokine
ATTCCTGAGC	2	0	0	0	0	0	0	0	0	0	0	0	0	2	24	0	2	0	0	0	0	0	0	0	0	0	0	no match	
ATACAGAATA	2	0	0	0	0	2	0	0	0	64	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	15	169238	MLK1 delta-like 1 homolog
CAGGAGAAGG	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24049	GOLGA2 golgi autoantigen, golgin subfamily a, 2
CAGGAGAAGG	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	366	MGC27165 hypothetical protein MGC27165
GCGGAGGTGG	2	0	0	0	0	0	2	2	0	283	4	11	10	6	2	4	0	0	0	0	0	0	0	0	0	0	0	366	MGC27165 hypothetical protein MGC27165
GCCGTTCTTA	41	0	0	0	2	0	0	0	42	27	277	0	11	0	0	0	2	0	3	0	0	5	5	3	0	32	0	no match	
TCAACACGCAG	2	0	0	0	0	0	0	0	4	4	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	no match	
GAGTTTATTC	3	0	0	0	3	0	0	0	4	2	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	no match	
AATGAATTAT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	9	0	0	293257	ECT2 epithelial cell transforming sequence 2 oncogene
TAGGTCAGGA	0	0	0	0	0	0	0	0	0	0	7	4	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	43666	PTP4A3 protein tyrosine phosphatase type IVA, member 3
CGAGAGTGTG	0	0	0	2	0	0	0	0	0	0	4	15	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	175804	CDNA FL742395 fis, clone ASTRO2001076
GCCCCTCCCG	0	0	0	0	0	0	0	0	0	0	0	11	5	0	4	31	2	91	0	2	0	0	0	12	3	0	0	435800	VIM vimentin
TGTTGAAAAA	104	0	9	0	0	0	3	15	82	0	4	31	2	91	0	2	0	0	0	0	12	3	0	0	0	0	0	89546	SELE selectin E
AAGTTTGTTG	0	0	0	0	0	0	0	0	0	3	12	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	66729	KCNJ10 potassium inwardly-rectifying channel, subfamily J, member 10
GGCCGCGAGG	0	0	0	0	0	0	0	2	0	0	18	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	78344	MYH11 myosin, heavy polypeptide 11, smooth muscle

SAGE tag numbers in the various libraries are indicated. Coloring reflects tag abundance in the different cell types.

Table 2. List of genes encoding secreted proteins and receptors overexpressed in DCIS myoepithelial cells compared to normal myoepithelium

SAGE Tag	N-MYOEP-1	D-MYOEP-7	D-MYOEP-6	Ratio D/N	Unigene	Gene description
ACCAAAAACC	2	274	849	244	172928	COL1A1 collagen, type I, alpha 1
GATCAGGCCA	0	191	181	124	443625	COL3A1 collagen, type III, alpha 1
TGGAAATGAC	0	50	228	93	172928	COL1A1 collagen, type I, alpha 1
CGGGGTGGCC	0	193	24	73	1584	COMP cartilage oligomeric matrix protein
CTAACGGGGC	0	169	20	63	513022	ISLR immunoglobulin superfamily containing leucine-rich repeat
CAGATAAGTT	0	72	101	58	222171	KIAA0182 KIAA0182 protein
CCGGGGGAGC	0	110	61	57	172928	COL1A1 collagen, type I, alpha 1
GTCAAAATTT	0	110	47	52	458354	THBS2 thrombospondin 2
GTGCTAAGCG	3	308	141	49	420269	COL6A2 collagen, type VI, alpha 2
GACTTTGGAA	0	36	110	49	172928	COL1A1 collagen, type I, alpha 1
CGCCGACGAT	0	100	32	44	287721	G1P3 interferon, alpha-inducible protein (clone IFI-6-16)
TTGGGATGGG	0	103	29	44	296941	HFL1 H factor (complement)-like 1
CATATCATTA	**0**	**21**	**94**	**38**	**435795**	**IGFBP7 insulin-like growth factor binding protein 7**
TCCAGGAAAC	0	72	39	37	11590	CTSF cathepsin F
GGCCCCTCAC	0	74	22	32	274313	IGFBP6 insulin-like growth factor binding protein 6
ACATTCCAAG	**0**	**50**	**42**	**31**	**245188**	**TIMP3 tissue inhibitor of metalloproteinase 3**
ATAAAAAGAA	**0**	**19**	**73**	**31**	**83942**	**CTSK cathepsin K**
GACCAGCAGA	0	43	48	30	172928	COL1A1 collagen, type I, alpha 1
ACTTATTATG	2	107	30	30	156316	DCN decorin
GTGCGCTGAG	0	33	52	28	274485	HLA-C major histocompatibility complex, class I, C
TGCGCTGGCC	0	67	18	28	289019	LTBP3 latent transforming growth factor beta binding protein 3
AGGCTCCTGG	**3**	**217**	**31**	**27**	**24395**	**CXCL14 chemokine**
CTCAACCCCC	2	105	19	27	162757	LRP1 low density lipoprotein-related protein 1
CAGCGGCGGG	0	57	13	23	2420	SOD3 superoxide dismutase 3, extracellular
GGCACCTCAG	2	36	65	22	512234	IL6 interleukin 6
GCCTGTCCCT	0	50	13	21	821	BGN biglycan
ATTTCTTCAA	0	19	44	21	31386	SFRP2 secreted frizzled-related protein 2
TCGAAGAACC	2	60	34	21	445570	CD63 CD63 antigen
ACATTCTTTT	0	17	44	20	389964	GPNMB glycoprotein (transmembrane)
CTGTCAGCGT	**0**	**29**	**32**	**20**	**283713**	**CTHRC1 collagen triple helix repeat containing 1**
CAGCTGGCCA	0	36	22	19	445240	FBLN1 fibulin 1
ACTGAAAGAA	3	124	50	19	458355	C1S complement component 1, s subcomponent
TTCTGTGCTG	3	105	40	16	376414	C1R complement component 1, r subcomponent
GGATGTGAAA	0	19	26	15	283477	CD99 CD99 antigen
ACTCAGCCCG	2	36	28	14	101382	TNFAIP2 tumor necrosis factor, alpha-induced protein 2
TTTCCCTCAA	2	21	42	14	75111	PRSS11 protease, serine, 11 (IGF binding)
CTAAAAAAAA	0	26	15	14	54457	CD81 CD81 antigen (target of antiproliferative antibody 1)
GGCCACGTAG	0	26	15	14	155597	DF D component of complement
AAGAAAGGAG	0	21	20	14	202097	PCOLCE procollagen C-endopeptidase enhancer
GGAGGAATTC	0	21	20	14	418123	CTSL cathepsin L
AGCCACCGCG	2	43	19	14	355874	RABL2B RAB, member of RAS oncogene family-like 2B
TGTAAACAAT	**0**	**19**	**22**	**14**	**170040**	**PDGFRL platelet-derived growth factor receptor-like**
ACCTTGAAGT	2	36	19	12	407546	TNFAIP6 tumor necrosis factor, alpha-induced protein 6
CATAAATGCG	**0**	**21**	**13**	**12**	**436042**	**CXCL12 chemokine (stromal cell-derived factor 1)**
TTGCTGACTT	12	122	279	11	415997	COL6A1 collagen, type VI, alpha 1
ATGGCAACAG	0	17	17	11	149609	ITGA5 integrin, alpha 5
CTCTCCAAAC	2	26	20	10	384598	SERPING1 serine proteinase inhibitor, clade G, member 1
TGCCTGCACC	**5**	**76**	**46**	**9**	**304682**	**CST3 cystatin C**
GGAAATGTCA	18	93	325	8	367877	MMP2 matrix metalloproteinase 2
CAGGTTTCAT	**12**	**124**	**117**	**7**	**24395**	**CXCL14 chemokine**
CCGTGACTCT	12	112	70	5	433622	FSTL1 follistatin-like 1

SAGE tag numbers reflect tag numbers normalized to the SAGE library with the highest tag number. Ratio was calculated as a ratio of the average tag numbers in the two DCIS myoepithelial libraries divided by the tag numbers in the normal myoepithelial library. Genes highlighted in red were selected for follow-up studies.

we identified 357 tags that differentiate epithelial cells from other cell types, 572 tags specifying myoepithelial cells and myofibroblasts, 502 tags discriminating leukocytes, 604 tags selecting stroma, and 124 tags discerning endothelial cells from other cells. To further define SAGE tags specific for each cell type, within each group of tags we selected the ones that were not only statistically significantly different, but also more abundant in the specific cell type. This led to the identification of 70 tags that were most abundant in epithelial cells, 117 tags present at highest levels in myoepithelial cells and myofibroblasts, 70 tags highly expressed in leukocytes, and 117 stroma- and 78 endothelium-specific tags (Supplemental Tables S3, S5, S7, S9, and S11). Several of these genes have previously been described as being specific for a particular cell type, such as keratins 8 and 19 for epithelial cells, keratins 14 and 17 for myoepithelial cells, and chemokines and chemokine receptors for leukocytes (Page et al., 1999), but the cell type-specific expression of the majority of the genes has not been documented. The majority of the transcripts corresponding to these cell type-specific SAGE tags encode known genes, but a significant fraction are uncharacterized ESTs or currently have no cDNA match ($\sim$10% of the tags on average belong to each of these last two groups). The only exceptions were tags most abundant in stroma, since in this group 25/117 tags (21%) had no database match, suggesting that they correspond to previously unidentified transcripts.

Next, using the SAGE tags most abundant in (417 tags) or most highly specific for (63 tags) each of the five cell types, we performed clustering analysis of all 27 SAGE libraries using a new Poisson model-based K-means algorithm (PK algorithm, Supplemental Data; Cai et al., 2004) to delineate similarities and differences among the samples (Figures 1C and 1D). In addition, we also performed clustering analysis of the SAGE libraries using each of the cell type-specific gene sets (Supplemental Figures S1 and S2). The PK clustering method orders the samples according to their relatedness. For example, using the 63 most highly cell type-specific SAGE tags, we obtained a division of the 27 SAGE libraries according to cell types, and within each cell type subgroup, the DCIS samples were located between normal breast tissue and invasive breast cancer SAGE libraries (Figure 1D). This result indicates that not only tumor epithelial cells, but also other cell types in the tumor, are different from their corresponding normal counterparts. Since these differences are already pronounced at a pre-invasive (DCIS) tumor stage, they suggest a role for stromal changes not only in tumor invasion and metastasis, but also in the earlier steps of breast tumorigenesis.

Based on our SAGE data, we found that the most consistent and dramatic gene expression changes occur in myoepithelial cells. More than 300 genes were differentially expressed at $p < 0.002$ in both DCIS myoepithelial libraries, and interestingly, a significant fraction of these genes (89 out of 245 characterized genes) encode secreted or cell surface proteins, suggesting extensive abnormal paracrine interactions between myoepithelial and other cell types (Supplemental Table S5). Myoepithelial cells are thought to be derived from bipotential stem cells that also give rise to luminal epithelial cells, although recently another progenitor has been identified that can differentiate only into myoepithelial cells (Bocker et al., 2002; Dontu et al., 2003). The function of myoepithelial cells and their role in breast cancer are not well understood, but myoepithelial cells have been shown to be able to suppress breast cancer cell growth, invasion, and angiogenesis (Deugnier et al., 2002; Sternlicht and Barsky, 1997). The main distinguishing feature of in situ and invasive carcinomas, which is also used as a diagnostic criteria, is that in DCIS, the cancer epithelial cells are separated from the stroma by a nearly continuous layer of myoepithelial cells and basement membrane, while in invasive and metastatic tumors, cancer cells are admixed with stroma. Due to our SAGE and previously published data suggesting a role for these cells in breast tumor progression, we focused our follow-up studies on myoepithelial cells with special emphasis on secreted proteins and receptors abnormally expressed in these cells. Several proteases (cathepsins F, K, and L, MMP2, and PRSS11), protease inhibitors (thrombospondin 2, SERPING1, cystatin C, and TIMP3), and many different collagens were highly upregulated in DCIS myoepithelial cells, suggesting a role for these cells in extracellular matrix remodeling (Table 2).

Analysis of the genotype of epithelial, myoepithelial, and stromal cells

To determine if the dramatic gene expression changes observed in tumor myoepithelial and stromal cell types could be due to underlying genetic alterations, we first performed aCGH analysis of epithelial and myoepithelial cells and of myofibroblasts from two DCIS (DCIS-6 and -7) and one invasive breast carcinoma (IDC7) used for SAGE. As expected, we detected numerous chromosomal gains and losses in the tumor epithelial cells, while no changes were detected in myoepithelial cells and myofibroblasts (Figures 2A and 2B). Similarly, no genetic changes were detected in epithelial and myoepithelial cells isolated from normal tissue adjacent to the tumors (Figure 2A). These data suggest that although nonepithelial cells in breast tumors are phenotypically distinct from their normal counterparts, genetic changes detectable by aCGH appear to be limited to cancer epithelial cells. However, since array CGH is thought to be more sensitive for the detection of copy number gains than losses and previous studies demonstrated LOH in stromal cells, we applied another technology, SNP arrays, for the analysis of isolated epithelial and stromal cells from a set of breast tumors. As expected, cancer epithelial cells from all but one invasive breast tumor demonstrated numerous LOH on nearly all chromosomes, while myofibroblasts and other stromal cells from the same tumors appeared to be mostly normal (Figure 2C and Supplemental Figure S3). Clustering analysis based on the inferred LOH data clearly divided the samples into two major groups, the tumor epithelial and stromal cells from different cases demonstrating more similarity to each other than to their corresponding other cell type (Figure 2C). The only exception was epithelial cells from IDC10 (a low-grade estrogen receptor-positive tumor) that did not appear to have major genetic changes (the purity of the tumor epithelial cells was confirmed by RT-PCR, data not shown), while in the phyllodes tumor, the stroma had numerous genetic alterations with much fewer LOH events detected in the epithelial cells. We did not detect significant LOH in the three fibroadenomas analyzed or in the one LCIS (lobular carcinoma in situ) case. Two nonepithelial samples (I-MYOFIB-8 and I-STR-13) had a few areas where 2–5 adjacent SNPs exhibited LOH (Figure 1C), but careful examination of these SNPs individually suggested that these LOH calls are likely due to poor hybridization results. In order to resolve this issue, we amplified and sequenced eight of these ambiguous

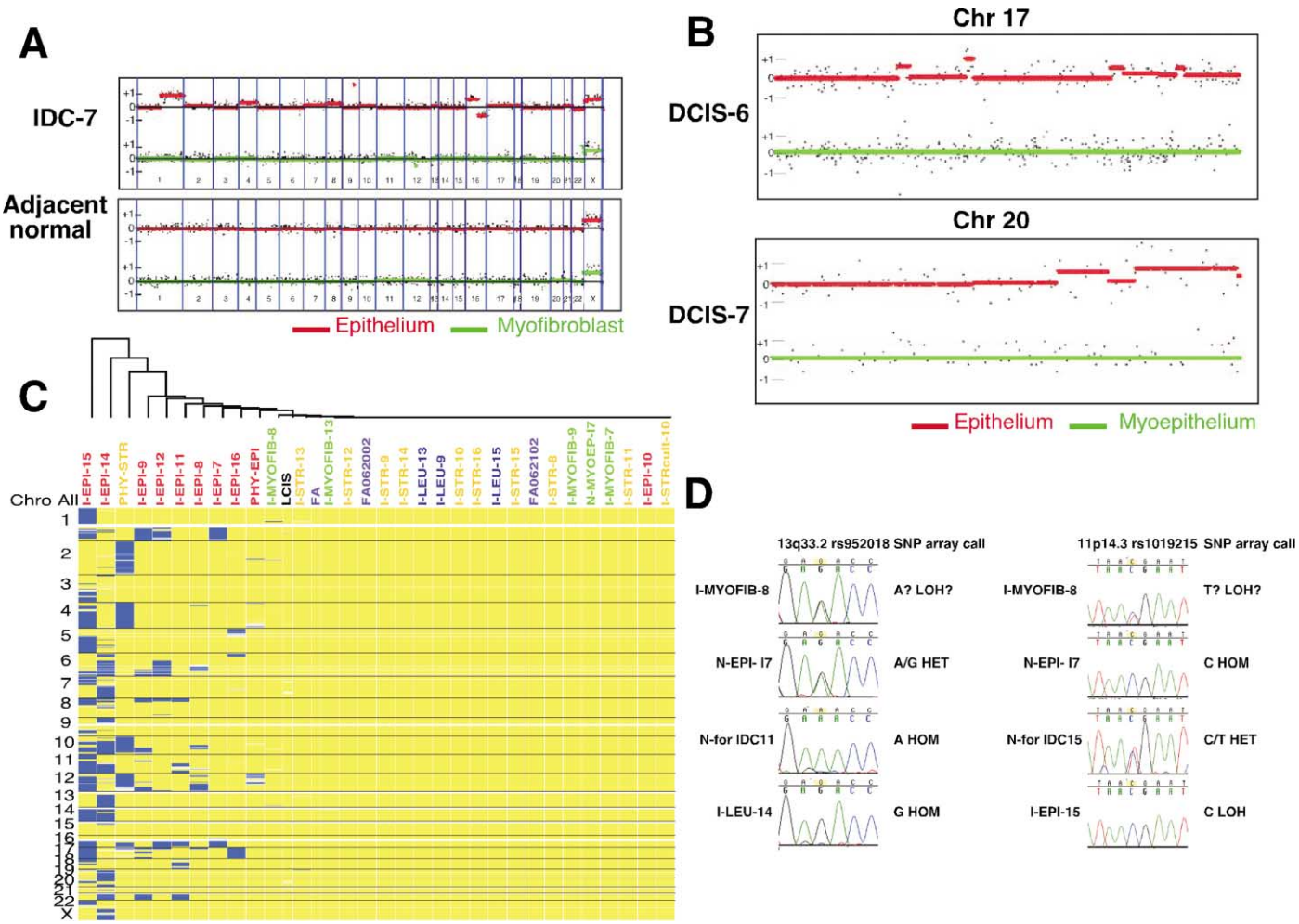

Figure 2. Genotype analysis of fractionated normal and tumor breast tissue

A: Array CGH analysis of luminal epithelial (red line) and myofibroblasts (green line) cells isolated from IDC-7 invasive breast tumor used for SAGE and from adjacent normal tissue. Mode centered segmented data, significant gains and losses defined as Log2 signal ratio of greater than or equal to +0.13 or −0.13, respectively, are depicted.

B: Array CGH analysis of luminal epithelial (red line) and myoepithelial (green line) cells from DCIS-6 and DCIS-7. Areas with statistically significant gains in the epithelial cells (chromosome 17 in the case of DCIS-6 and chromosome 20 for DCIS-7) are depicted, indicating that myoepithelial cells do not share these changes with the epithelial cells. No significant gains and losses were detected in any other areas of the genome in the myoepithelial cells (data not shown).

C: SNP array analysis of purified epithelial and stromal cells from invasive breast carcinomas, phyllodes tumor, fibroadenomas, and LCIS. Samples are clustered based on inferred loss of heterozygosity (LOH). All but one tumor epithelial DNA sample are clustered together to the left, while all stromal samples, regardless of their origin, are clustered together to the right. Inferred loss of heterozygosity (LOH) is indicated in blue, yellow indicates regions retaining heterozygosity, and white regions are indeterminate (noninformative). The names of DNA samples obtained from epithelial cells are depicted in red, myofibroblasts in green, stroma in yellow, leukocytes in blue, endothelial cells in pink, fibroadenoma in purple, and LCIS in black. A detailed description of the samples is included in the online Supplemental Data.

D: Sequence analysis of two ambiguous SNP cells present in I-MYOFIB-8 and in several controls. For all cases, the chromatograms of the sequence reads and the SNP array calls are indicated. One of the SNPs (rs952018) is on chromosome 13q33.2, while the other one (rs1019215) is on chromosome 11p14.3. As depicted in the figure in the case of SNP rs952018, the I-MYOFIB-8 sample had both "A" and "G" peaks just like the N-EPI-I7 sample, proving the retention of both alleles, while the I-LEU-14 sample was homozygous for the "G" allele and the N-for IDC11 (normal DNA corresponding to tumor IDC11) was homozygous for the "A" allele. Similarly in the case of SNP rs952018, the I-MYOFIB-8 sample had both T and C peaks just like the N-for IDC15 (normal DNA corresponding to tumor IDC15), while the N-EPI-I7 and I-EPI-15 were both homozygous for the "C" allele.

SNPs from these two stromal samples (I-MYOFIB-8 and I-STR-13) together with several controls, where the SNP results clearly depicted heterozygous or homozygous alleles. In all seven cases in which high-quality sequencing results were obtained, we found no evidence of LOH in either of these two ambiguous stromal samples (Figure 2D).

Evaluation of gene expression by immunohistochemistry and mRNA in situ hybridization

The generation of the SAGE libraries involved the in vitro purification of the cells that could potentially alter the in vivo gene expression patterns, although prior SAGE data from several

laboratories suggest that these changes are likely to be minimal (Porter et al., 2003a, 2003b; St Croix et al., 2000). However, in order to further investigate the expression of selected genes at the cellular level in vivo, we performed immunohistochemical analyses and mRNA in situ hybridization in a panel of DCIS and invasive breast tumors (including tumors used for SAGE as well as additional cases). In addition, the cell type specificity of some genes was verified by RT-PCR in the samples used for SAGE (data not shown). Immunohistochemical analysis confirmed that two genes, IL-1β and CCL3 (MIP1α), are highly expressed in leukocytes infiltrating DCIS, but not normal breast tissue, whereas the PTPRC (CD45) panleukocyte marker was expressed in both cases (Figure 3A). Despite the similar number of total leukocytes in invasive tumors, the frequency of IL-1β and CCL3-positive leukocytes was much lower than in DCIS, suggesting that in situ and invasive breast carcinomas may be immunologically dissimilar. mRNA in situ hybridization determined that in DCIS tumors, the expression of PDGF receptor β-like (PDGFRBL), cathepsin K (CTSK), and CXCL12 was localized to myofibroblasts as determined by smooth muscle actin (ACTA2) staining, CXCL14 was expressed only in myoepithelial cells, while TIMP3, cystatin C (CST3), and collagen triple helix repeat containing 1 (CTHRC1) were expressed in both myoepithelial cells and myofibroblasts. In invasive tumors, all seven genes were expressed in myofibroblasts. No signal was detected in normal breast tissue nor with the sense probes (Figure 3B, Supplemental Figure S4, and data not shown). Interestingly, although in DCIS tumors we detected CXCL14 expression only in myoepithelial cells, in some (4/9) invasive breast carcinomas, the expression of CXCL14 was restricted to the tumor epithelial cells (Figures 3B and 4A). Similarly, some breast cancer cell lines expressed high levels of CXCL12 or CXCL14 in vitro, suggesting that during tumor progression a paracrine factor may be converted into an autocrine one due to its upregulation in the tumor epithelial cells (Figure 4B). Interestingly, all CXCL14-positive invasive ductal carcinomas and even the CXCL14 expressing breast cancer cell line (UACC812) were obtained from young, premenopausal patients (average age of onset 39 years), suggesting a possible association of CXCL14 expression with hormone levels or clinico-pathologic characteristics of the tumors, the analysis of which requires the examination of larger tumor sets.

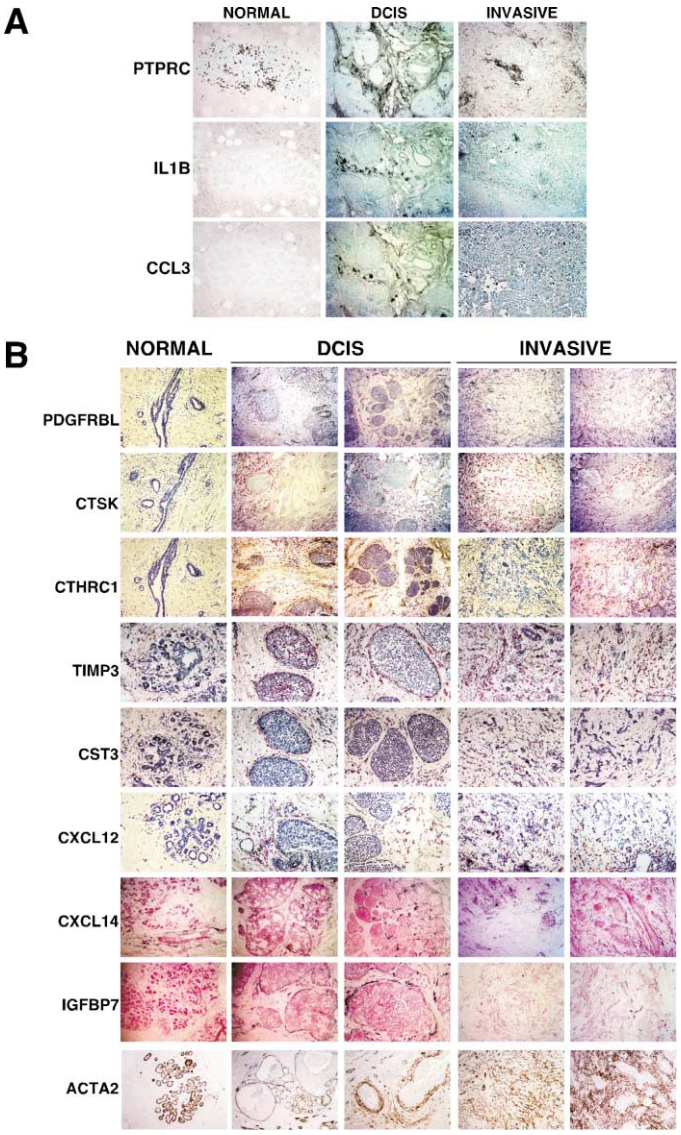

Figure 3. Validation of SAGE data using immunohistochemistry and mRNA in situ hybridization and Northern blot analysis

A: Immunohistochemical analysis of PTPRC (CD45), IL1β, and CCL3 expression in normal, DCIS, and invasive cancer breast tissue. Black signal indicates expression of the indicated proteins in leukocytes. Methyl green was used to stain the nuclei to visualize tissue histology. Magnification is 100×.

B: mRNA in situ hybridization analysis of the indicated genes using antisense ribo-probes in a panel of normal, DCIS, and invasive breast cancer tissue. Red (PDGFRL, CTSK, CTHRC1, TIMP3, CST3, and CXCL12) or black (CXCL14 and IGFBP7) staining indicates the presence of the mRNA depending on the hybridization protocol used. Paraffin sections were analyzed for ACTA2 (smooth muscle actin) expression by immunohistochemistry to confirm the identity of myoepithelial cells and myofibroblasts. Brown staining indicates the expression of SMA in myoepithelial cells and myofibroblasts. Magnification is 100×. More detailed images with higher (200×) magnification are included in Supplemental Data (Supplemental Figure S4).

The effect of CXCL12 and CXCL14 chemokines on breast cancer cells

The high level of expression of two chemokines, CXCL12 and CXCL14, in myoepithelial cells and myofibroblasts both in DCIS and invasive breast carcinomas was particularly interesting due to the known function of chemokines as regulators of cell proliferation, differentiation, migration, and invasion (Gerard and Rollins, 2001; Muller et al., 2001; Rossi and Zlotnik, 2000). To determine if CXCL12 and CXCL14 may act as autocrine and/or paracrine factors in breast tumors, we investigated which cell types appear to have receptors for these chemokines in vivo in primary breast tissue. The signaling receptor for CXCL12 is CXCR4, which is known to be widely expressed in various lymphoid as well as a variety of epithelial cells (Gerard and Rollins, 2001). We confirmed the expression of CXCR4 in lymphoid and breast epithelial cells using immunohistochemistry, while SAGE data indicated that its expression is increased in invasive tumors compared to DCIS and normal breast tissue (data not shown). The signaling receptor of CXCL14 is unknown, but cell surface ligand binding experiments have suggested the presence of a putative CXCL14 receptor on monocytes and B cells, suggesting that its receptor is not likely to be CXCR4 (Kurth et al., 2001; Sleeman et al., 2000). To determine if a

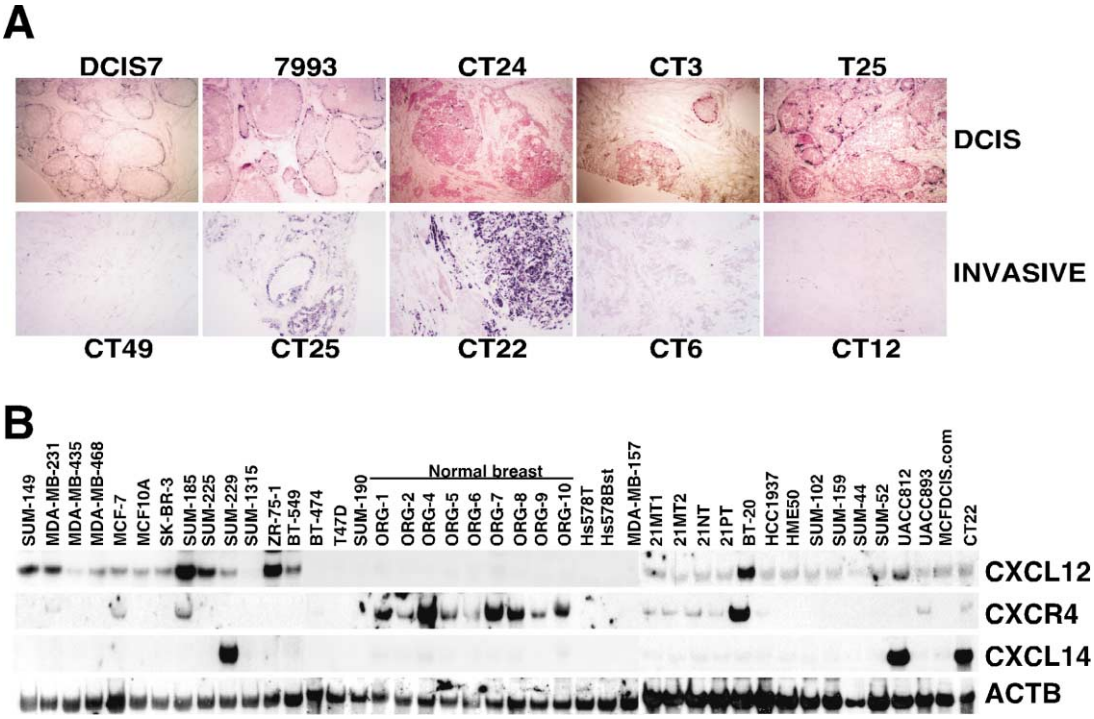

Figure 4. CXCL14 expression in primary breast tumors and breast cancer cell lines

A: mRNA in situ hybridization using CXCL14 antisense ribo-probe in multiple DCIS and invasive breast carcinomas including the tumors used for SAGE (DCIS-7). Black/purple staining indicates the presence of the CXCL14 mRNA, while nuclei were stained with nuclear FastRed to visualize tissue histology. The names of the tumor samples are indicated above/below of the pictures. In DCIS cases, CXCL14 is expressed only in myoepithelial cells, while in some invasive breast carcinomas (CT22 and CT25), strong expression is observed in tumor epithelial cells.

B: Northern blot analysis of CXCL12, CXCL14, and CXCR4 expression in the indicated breast cancer cell lines, breast organoids (ORG1–10, uncultured breast ducts from normal breast tissue), and primary breast tumor CT22. Hybridization with β-actin (ACTB) was used as a control for loading. Confirming the mRNA in situ hybridization data, strong CXCL14 expression is detected in tumor CT22, similarly in SUM-229 and UACC812 breast cancer cell lines.

CXCL14 binding cell surface protein(s) is also present on breast cancer cells, we generated an alkaline phosphatase-CXCL14 (AP-CXCL14) fusion protein to be used as a ligand in receptor binding assays. Conditioned media of AP-CXCL14 or control AP expressing cells was then used as an affinity reagent to stain normal and cancerous mammary tissue sections including the DCIS tumors used for SAGE. Blue staining indicated the presence of a CXCL14 binding protein in certain leukocytes and breast epithelial cells (Figure 5A). These results suggest the presence of a cell surface CXCL14 binding protein(s) in cancerous and normal mammary epithelial cells and are consistent with a paracrine mechanism of CXCL14 action in the breast. To test further the binding characteristics of AP-CXCL14, we performed in vitro ligand binding assays using various cell lines. Low-level AP-CXCL14 binding was detected in all cell lines tested, including MDA-MB-231 and MDA-MB-435 breast cancer and MCF10A immortalized mammary epithelial cells (data not shown). To further characterize the AP-CXCL14-putative CXCL14 receptor interaction, we performed more detailed binding assays in MDA-MB-231 breast cancer cells. Scatchard plot analysis showed two binding slopes in MDA-MB-231 cells indicating the presence of high-affinity ($K_d = 6.1 \times 10^{-8}$ M) and low-affinity ($K_d = 56.7 \times 10^{-8}$ M) binding sites (Figure 5B).

In previous studies, CXCL12 was demonstrated to enhance breast cancer cell growth, migration, and invasion (Hall and Korach, 2003; Muller et al., 2001). In order to determine if CXCL14 has similar effects, we tested the effect of conditioned medium containing AP-CXCL14 on the growth of MDA-MB-231 and MCF10A cells, while its effect on cell migration and invasion was investigated in MDA-MB-231 cells. Conditioned medias of cells transfected with AP alone and CXCL12 were used as negative and positive controls, respectively. Similar to CXCL12, CXCL14 enhanced the proliferation of MDA-MB-231 and MCF10A cells and the migration and invasion of MDA-MB-231 cells (Figures 5C and 5D, and data not shown). The concentration of AP-CXCL14 was 2–30 nM in these experiments, which is similar to the concentration required by several chemokines, including CXCL12, to exert biological effects. The same results were obtained in cell migration and invasion assays using CXCL14-AP (C-terminal AP-tag) and CXCL14-HA (C-terminal HA-tag) fusion proteins (Figure 5D and data not shown); thus, the observed effects are not likely to be due to the position or identity of the epitope tag. Preliminary results using recombinant CXCL14 protein and CXCL14 expressing adenovirus demonstrated possible induction of calcium flux in MDA-MB-231 and activation of AKT in MCF10A cells, respectively (data not shown), further suggesting that mammary epithelial cells have a functional CXCL14 receptor.

To determine if paracrine factors, including CXCL14, secreted by DCIS myoepithelial cells may influence the proliferation of tumor epithelial cells in vivo, we analyzed the expression of Ki67, a marker of cell proliferation, in the two DCIS samples

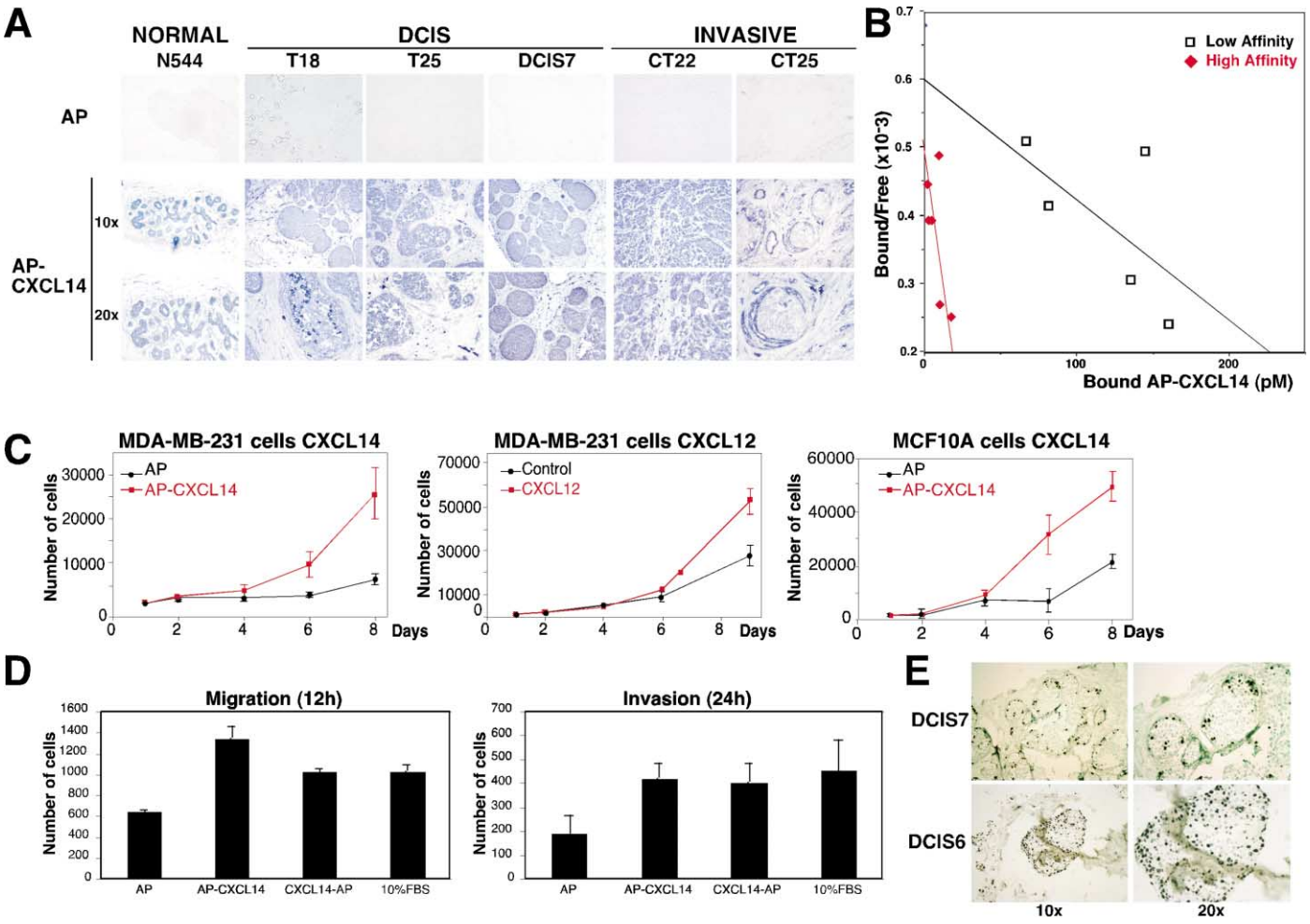

Figure 5. Analysis of CXCL14 ligand binding characteristics and function

A: Identification of a putative CXCL14 receptor in breast epithelial cells using an AP (alkaline phosphatase)-CXCL14 fusion protein as ligand. Blue staining reflecting AP activity indicates binding of AP-CXCL14 to breast epithelial cells and some stromal leukocytes, while no staining is detected with the AP alone negative control. All these tumor samples were also analyzed for the expression of CXCL14 by mRNA in situ hybridization (Figures 3B and 4A) and were expressing CXCL14 in tumor epithelial cells (CT22 and CT25) and DCIS myoepithelial cells (T18 and T25). Images were taken with 10× and 20× objectives (100× and 200× magnification).

B: Scatchard transformation of AP-CXCL14 binding assays in MDA-MB-231 cells. Red and black colored lines indicate high (K_d = 6.1 × 10^{-8} M) and low (K_d = 56.7 × 10^{-8} M) -affinity binding slopes, respectively.

C: The effect of CXCL12 and AP-CXCL14 on the growth of MDA-MB-231 breast cancer and MCF10A immortalized breast epithelial cells (red lines) compared to AP and control media (black lines). Representative result of experiments performed in triplicate.

D: The effect of CXCL14 and 10% fetal bovine serum (FBS) on the migration and invasion of MDA-MB-231 breast cancer cells. The number of cells that crossed the uncoated (migration) or Matrigel-coated membranes (invasion) is indicated on the *y* axis. Representative result of experiments performed in triplicate.

E: Immunohistochemical analysis of Ki67 expression in DCIS-6 and DCIS-7 samples to identify proliferating cells. Images were taken with 10× and 20× objectives (100× and 200× magnification). Ki67 is expressed in all phases of the cell cycle except in noncycling (G_0) cells. Tumor epithelial cells adjacent to the myoepithelial cell layer are more frequently positive than their more centrally located counterparts.

used for SAGE (Figure 5E). In both cases, epithelial cells adjacent to the myoepithelial cell layer were more frequently positive for Ki67 than tumor epithelial cells in other parts of the ducts. This result suggests that tumor epithelial cells may receive paracrine signals from adjacent myoepithelial cells that enhance their proliferation, although other reasons for this intraductal location-dependent proliferation difference cannot be excluded. Correlating with this, a recent study described that the gene expression profile of tumor epithelial cells located at the periphery and the center of DCIS ducts is significantly different (Zhu et al., 2003).

Discussion

Epithelial-mesenchymal interactions are known to be important for the normal development of the mammary gland and to play a role in breast tumorigenesis (Bissell et al., 2002; Coussens and Werb, 2002; Kenny and Bissell, 2003; Radisky et al., 2001; Shekhar et al., 2003; Tlsty, 2001; Tlsty and Hein, 2001; Wiseman and Werb, 2002). Early studies demonstrated that the normal mammary microenvironment is capable of "reverting" the neoplastic phenotype of breast cancer cells by inducing cellular differentiation (DeCosse et al., 1973, 1975), suggesting that

cancer cells can thrive only in a distorted environment or have to become independent of extracellular signals. The contribution of genetic host factors to tumor initiation, progression, and angiogenesis also support a role for nonepithelial cells in carcinogenesis (Hunter, 2004; Rohan et al., 2000). This was dramatically illustrated by the finding that inactivation of TGF-β type II receptor in stromal fibroblasts led to prostate and gastric epithelial neoplasia (Bhowmick et al., 2004). Similarly, a recent finding demonstrating that mammary tumors only formed in cleared mammary fat pads of rats treated with carcinogen, regardless of whether the injected epithelial cells were treated with carcinogen in vitro, also emphasizes the importance of stromal alterations in the initiating steps of breast cancer (Maffini et al., 2004). Numerous in vitro and in vivo studies using diverse experimental systems have demonstrated that the growth, survival, polarity, and invasive behavior of breast cancer cells can be modulated by myoepithelial and various stromal cells, and several genes have been implicated to play an important role in this process (Bissell et al., 2002; Coussens and Werb, 2002; Deugnier et al., 2002; Elenbaas and Weinberg, 2001; Gudjonsson et al., 2002; Kenny and Bissell, 2003; Radisky et al., 2001; Shekhar et al., 2003; Sternlicht and Barsky, 1997; Tlsty, 2001; Tlsty and Hein, 2001; Wiseman and Werb, 2002). However, comprehensive molecular analysis of all cell types that compose normal human mammary breast tissue and breast carcinomas has not been performed.

With the aim of delineating epithelial-stromal/myoepithelial cell interactions at the molecular level, we determined the comprehensive gene expression and genomic profiles of epithelial, myoepithelial, and stromal cells in normal breast tissue and in situ and invasive breast carcinomas. Our results confirm at the molecular level that the cellular microenvironment is dramatically different between normal breast tissue and breast carcinomas and that this is already evident at the in situ carcinoma stage. Based on our gene expression data, we determined that the most dramatic and consistent changes occur in myoepithelial cells and myofibroblasts and the majority of the differentially expressed genes encode secreted and cell surface proteins (Tables 1 and 2 and Supplemental Tables S2 and S5). Since previous data also implicated these two cell types in breast tumor progression, particularly in the transition of in situ to invasive carcinomas (Alpaugh et al., 2000; Barsky, 2003; Chauhan et al., 2003; Nguyen et al., 2000; Shao et al., 1998; Sternlicht and Barsky, 1997; Sternlicht et al., 1997; Walter-Yohrling et al., 2003), we mainly focused on tumor myoepithelial cells and myofibroblasts and the genes expressed by them.

Myoepithelial cells play a major role in the formation of the basement membrane and lactation due to their expression of type IV collagen, laminin, smooth muscle actin, and oxytocin receptor (Gudjonsson et al., 2002; Murrell, 1995). They also have been suggested to suppress breast cancer cell growth, invasion, and angiogenesis via shedding of CD44 and expression of protease inhibitors (Alpaugh et al., 2000; Barsky, 2003; Xiao et al., 1999). On the other hand, myoepithelial cells are also important for the survival, differentiation, and polarity of normal luminal epithelial cells (Gomm et al., 1997a, 1997b). Proteomic and mRNA expression profiling of short-term cultured myoepithelial cells and myoepithelial cell lines, respectively, gave a glimpse of the molecular basis for the tumor and invasion suppressor role of normal myoepithelium (Barsky, 2003; Page et al., 1999). Our SAGE-based profiling of freshly isolated, uncultured myoepithelial cells from normal breast tissue also demonstrated the high expression of laminin, tenascin, thrombospondin, and PAI-1 binding protein. However, the expression of these genes was downregulated in DCIS myoepithelial cells similar to that of cytokeratins 7, 14, and 17, oxytocin receptor, and tropomyosin, suggesting that DCIS myoepithelial cells are phenotypically altered and less differentiated than normal myoepithelial cells. Keeping with this, several recent studies described a lack of commonly used myoepithelial markers (including CD10 and SMA) in a subset of morphologically distinct myoepithelial cells, suggesting that myoepithelial cells may also be subject to pathological alterations (Zhang et al., 2003). Moreover, in support of a role for myoepithelial cells in breast tumor progression, it was recently reported that DCIS tumor epithelial cells adjacent to a disrupted myoepithelial cell layer are molecularly and genetically different from their more distant counterparts (Man et al., 2003).

Myofibroblasts are stromal fibroblasts with features of both myoblasts (e.g., expression of smooth muscle actin) and fibroblasts that have been implicated in breast cancer invasion, extracellular matrix remodeling, wound healing, and chronic inflammation (De Wever and Mareel, 2003; Gabbiani, 1999; Schurch, 1999). The cell type of origin of myofibroblasts is not conclusively established. Certain cytokines can induce (TGF-β) or inhibit (IFN-γ) the transformation of fibroblasts into myofibroblasts in vitro (De Wever and Mareel, 2003; Tanaka et al., 2003), while PDGF-B stimulates the proliferation of fibrocytes and their conversion into myofibroblasts in vivo (Oh et al., 1998). Isolation of various stromal and epithelial cells from breast tumors and their coculturing in vitro demonstrated that cancer epithelial cells can induce the expression of myofibroblast markers in a subset of fibroblasts (Ronnov-Jessen et al., 1995). However, the finding that only a small fraction of fibroblasts were transformed into myofibroblasts (Ronnov-Jessen et al., 1995) raises the question of whether myofibroblasts could be derived from specific stem cells that are normally present in the breast or in the bone marrow and are growth stimulated or recruited by adjacent cancer epithelial cells. Recent data both in animal models and human breast tumors support the hypothesis that at least a subset of cancer-associated myofibroblasts is derived from circulating bone-marrow derived cells (Chauhan et al., 2003; Ishii et al., 2003). Our finding that the gene expression profiles of myofibroblasts isolated from different invasive breast carcinomas are highly similar also suggest their common cell type of origin.

Two genes highly expressed in tumor myoepithelial cells and myofibroblasts encoding chemokines CXCL12 and CXCL14 were particularly interesting due to the recently demonstrated role of chemokines and chemokine receptors in cancer cell growth, invasion, and metastasis (Barbero et al., 2002; Chen et al., 2003; Hall and Korach, 2003; Kang et al., 2003; Muller et al., 2001; Scotton et al., 2002). CXCL12 has been previously implicated in breast cancer metastasis (Kang et al., 2003; Muller et al., 2001), but its high expression in DCIS (a pre-invasive tumor) myofibroblasts suggests that it might have additional roles in the earlier stages of breast tumorigenesis. Correlating with this hypothesis, CXCL12 was recently identified as a transcriptional target of the estrogen receptor that mediates estrogen-induced proliferation of breast cancer cells (Hall and Korach, 2003). Relatively little is known about the CXCL14 chemokine despite the fact that it was independently identified

by multiple labs using different approaches. The high expression of CXCL14 in inflammatory cells in multiple cancer types and its selectivity to monoctyes may suggest a role in macrophage development (Frederick et al., 2000; Hromas et al., 1999; Kurth et al., 2001; Sleeman et al., 2000). Although the receptor for CXCL14 has not been identified, the induction of calcium mobilization by recombinant CXCL14 in monocytes suggests that similar to other chemokines, it is also likely to signal via a G protein-coupled receptor. Our preliminary results demonstrating intracellular calcium flux in MDA-MB-231 breast cancer cells also support this hypothesis.

In addition to phenotypic alterations, several recent studies described genetic changes (including LOH and mutations in tumor suppressor genes) in stromal cells adjacent to breast cancer cells (Kurose et al., 2001, 2002; Lakhani et al., 1998; Moinfar et al., 2000; Wernert et al., 2001). Loss of heterozygosity at several loci has also been demonstrated in normal-appearing epithelial cells adjacent to breast carcinomas and short-term cultured luminal and myoepithelial cells (Deng et al., 1996; Forsti et al., 2001; Lakhani et al., 1999; Moinfar et al., 2000). In several cases, the tumor epithelial and stromal cells had discordant genetic changes, suggesting a clonal co-evolution for these two cell types. Moreover, due to the low probability of two adjacent cells simultaneously acquiring genetic changes, this would also suggest that breast cancer epithelial and stromal cells may be derived from a common stem cell and then undergo a divergent genetic selection process.

In order to determine if in the same population of tumor epithelial, myoepithelial, and stromal cells in which we detected dramatic gene expression changes there are also underlying genetic alterations, we analyzed the genotype of these cell types using different technologies in 2 DCIS and 12 invasive breast carcinomas. All but one of the tumor epithelial cells had numerous LOH involving almost all chromosome arms. The most frequent genetic changes we identified in the tumor epithelial cells (1q, 8q, 17q, and 20q gain, and 6q, 8p, 10q, 12q, and 17p LOH), both in DCIS and invasive tumors, were in agreement with that of prior studies (Nishizaki et al., 1997; Waldman et al., 2000). The one tumor DNA sample (IDC10) that appeared to be devoid of significant LOH was obtained from a low-grade estrogen and progesterone receptor-positive, HER2-negative breast tumor. The lack of gross chromosomal changes in this tumor is unlikely to be due to technical issues, but potentially reflects a special pathway of breast tumorigenesis. Correlating with this, an independent study using BAC array CGH analysis of a large set of breast tumors also found that a subset of breast tumors (9/146) have minimal ($<$1.5% of the genome) chromosomal changes (Dr. J. Gray, Lawrence Berkeley National Laboratory, personal communication). Using three different methods (aCGH, SNP arrays, and direct sequencing of specific SNPs), we detected genetic changes only in cancer epithelial cells. However, in a malignant phyllodes tumor that is thought to be composed of malignant stroma and epithelium, we detected LOH in both components. These results suggest that using the technologies we applied, genetic changes can be detected both in epithelial and stromal cells, but only if there is a mono- or oligoclonal proliferation of neoplastic epithelial or stromal cells. Our inability to find conclusive genetic alterations in stromal cells from invasive ductal breast carcinomas is seemingly in disagreement with the findings of several of the above referenced studies. However, we believe that the reason for the different results could be due to the use of different technologies and approaches. All the studies that described LOH in cancer stroma analyzed a few polymorphic markers and a fairly small population of stromal cells isolated by microdissection from the same area adjacent to tumor epithelial cells, while we analyzed all the stromal cells from the tumor and used comprehensive genome-wide SNP arrays. Thus, if the stromal cells are highly heterogeneous with respect to genetic alterations, these changes can be detected only if relatively few cells from the same area of the tumor are analyzed. However, in our view this argues against the hypothesis that the genetic changes in the stroma are selected for and thus play a major role in tumorigenesis.

In summary, this study provides a comprehensive molecular characterization of each cell type composing normal breast tissue and in situ and invasive breast carcinomas. The genes described here should therefore provide a valuable resource for future basic and clinical studies addressing the role of epithelial-stromal cell interactions in breast and other cancer types. The availability of specific chemokine receptor inhibitors and preclinical studies demonstrating dramatic tumor and metastasis suppressive effects using CXCR4 inhibitors in brain and breast tumors (Rubin et al., 2003; Tamamura et al., 2003) provide a proof of principle that therapeutic targeting of chemokines is a promising new opportunity for the treatment of breast carcinomas.

Experimental procedures

Cell lines and tissue specimens

Breast cancer cell lines were obtained from American Type Culture Collection (Manassas, VA) or were generously provided by Drs. Steve Ethier (University of Michigan) and Arthur Pardee (Dana-Farber Cancer Institute). Cells were grown in media recommended by the provider. Tumor specimens were obtained from Brigham and Women's and Massachusetts General Hospitals (Boston, MA), Duke University (Durham, NC), University Hospital Zagreb (Zagreb, Croatia), and the National Disease Research Interchange, snap frozen on dry ice, and stored at -80°C until use or were processed for purification as described below. All human tissue was collected using protocols approved by the Institutional Review Boards. We purified all the cell types from 2 different normal reduction mammoplasty tissues, 2 different DCIS, and 13 different invasive ductal carcinomas. Due to technical difficulties (insufficient number of cells), we were not able to generate SAGE libraries from each cell type of each tissue used for purification. In addition, selected cell types were isolated from a few additional normal and DCIS samples. The detailed protocol used for the purification of all cell types is included in the Supplemental Data. The estimated number of cells obtained from each fraction varied from 10,000 to 100,000.

Generation and analysis of SAGE libraries, mRNA in situ hybridization, and immunohistochemistry

All SAGE libraries were generated using a modified micro-SAGE protocol and the I-SAGE (libraries prepared in 2002) or long I-SAGE (I-epi-7, I-epi-8, I-epi-9, I-leu-7, I-str-7, I-myofib-7, I-myofib-8, I-myofib-9, D-str-6, FA, PHY) kits from Invitrogen. The samples were collected and SAGE libraries were generated during 2002–2004, and the long-SAGE kit became available only in 2003. SAGE libraries were sequenced by Agencourt (Beverly, MA) as part of the NCI-CGAP SAGE project, and all data will be deposited to the SAGEGenie website (http://cgap.nci.nih.gov/SAGE). Approximately 50,000 tags (average tag number 56,647 $\pm$ 4,383) were obtained from each library, and the preliminary analysis of the SAGE data was performed essentially as described (Porter et al., 2001). Briefly, genes significantly ($p \leq 0.002$) differentially expressed between normal and cancerous cells were identified by performing pair-wise comparisons using the SAGE2000 software and Monte Carlo analysis. Significance calculation among groups of SAGE libraries and clustering analyses were performed using a new Poisson model-based K-means algorithm (PK algorithm, Cai et al., 2004). A detailed descrip-

tion of the methodology used for the analysis and clustering of the SAGE data is provided in the Supplemental Data. Probes for the selected genes to be used for mRNA in situ hybridization were generated by PCR amplification of a 300–500 bp region of the 3′UTR and subcloning the fragments into pZERO1.0 (Invitrogen). The identity of the subcloned PCR products was confirmed by sequencing, and the resulting plasmids were used for the generation of digitonin-labeled riboprobes essentially as described (Porter et al., 2003a). mRNA in situ hybridizations and immunohistochemistry were performed as described or as recommended by the antibody supplier (Porter et al., 2003a). Mouse monoclonal antibodies for IL1β and CCL3 were purchased from R&D, while anti-CD45 and anti-Ki67 mouse monoclonal antibodies were obtained from DAKO.

Array comparative genomic hybridization

cDNA array comparative genomic hybridization using Agilent (Palo Alto, California) arrays were performed by the Belfer Genome Center at the Dana-Farber Cancer Institute. Genomic DNA was digested with DpnII and random prime labeled according to standard protocols with slight modifications (Pollack et al., 1999). (For a detailed protocol, see http://genomic.dfci.harvard.edu/array_cgh.htm.) Labeled DNAs were hybridized to human cDNA microarrays containing 12,814 unique cDNA clones (Agilent Technologies, Human 1 clone set). Among these clones, approximately 9,420 unique map positions were found for 12,020 unique GenBank sequences. The median interval between cDNAs is 100.1 kilobase, 92.8% of intervals are less than 1 megabase, and 98.6% are less than 3 megabases. The density of coverage is closely correlated with gene density. Following extensive QA analysis, fluorescence ratios of scanned images of the arrays were calculated and the raw array CGH profiles were processed to identify statistically significant transitions in copy number using a segmentation algorithm that employs permutation to determine the significance of change points in the raw data. By mode centering this segmented data set, we defined gains and losses as Log2 signal ratio of greater than or equal to $+0.13$ or -0.13, respectively, and amplification and deletion as a ratio greater than 0.52 or less than -0.58, respectively (e.g., 97% or 3% quantiles). Statistical analysis of the aCGH data will be described in detail elsewhere (Brennan et al., 2004). Segmentation of aCGH profiles was performed by changepoint identification algorithm provided by Adam Olshen and E.S. Venkatraman (Lucito et al., 2003).

Single nucleotide polymorphism array analysis

SNP array hybridizations were performed by the Dana-Farber Microarray Core using Affymetrix 11K XbaI SNP arrays and protocols recommended by Affymetrix (Santa Clara, CA). These arrays contain probes for both alleles at 11,565 SNP loci, with mean and median intermarker distances of 209 kb and 104 kb, respectively, with probe density closely correlating with gene density (Matsuzaki et al., 2004). Arrays were scanned using a confocal laser scanner (Agilent, Palo Alto, California), and Affymetrix genotyping software (Affymetrix GeneChip 5.0) was used to make allele calls for all loci. These data were then analyzed using dChipSNP (Lieberfarb et al., 2003; Lin et al., 2004). Loci that were heterozygous (AB call) in normal epithelium or leukocytes but homozygous (AA or BB call) in the test tissue were identified as potentially having undergone LOH. Some of these potential LOH events reflect genotyping error, but serial events among neighboring loci along a chromosome likely reflect true regional LOH. Regions of statistically likely LOH were delineated according to a Hidden Markov Model analysis; the detailed method will be described elsewhere. Two samples (LCIS and FA) had no normal reference counterparts. For these, regions of LOH were inferred when a stretch of consecutive homozygous loci exceeded what would be expected by chance alone. Again, a Hidden Markov Model analysis was used, assigning a marginal probability of heterozygosity of 0.37 to correspond to the actual rate found in these SNPs (Matsuzaki et al., 2004) and a transition probability between consecutive SNPs proportional to the genetic distance between them (Lander and Green, 1987). A detailed description of the method will be presented in a forthcoming study (M. Lin, R.B., X. Zhao, M. Meyerson, C. Li, and W.R.S., unpublished data). Samples were clustered hierarchically as previously described (Lin et al., 2004), based upon LOH calls in all statistically likely regions of LOH in all chromosomes.

Ligand binding, cell growth, migration, and invasion assays

We generated N-terminal or C-terminal alkaline phosphatase (AP) CXCL14 fusion proteins using the AP-TAG-5 expression vector (GenHunter, Nashville, TN). We transfected mammalian cells with Fugene6 (Roche, Indianapolis, IN), Lipofectamine, or Lipofectamine 2000 (LifeTechnologies, Rockville, MD) reagents. We performed in vivo and in vitro ligand binding assays on primary tissues and cell lines using AP-CXCL14 essentially as described (Flanagan and Leder, 1990; Porter et al., 2003b). Briefly, we fixed frozen sections of various human specimens, incubated with either AP-CXCL14 fusion protein or AP control conditioned medium, rinsed, and then incubated with AP substrate forming a blue/purple precipitate. For in vitro assays we incubated cells in suspension with conditioned media containing either AP alone or AP-CXCL14 fusion protein, rinsed, and then assayed for bound AP activity. To determine the effect of CXCL14 on cell growth, we plated MDA-MB-231 and MCF10A cells (4000 cells/well in a 24-well plate) and grew them in conditioned media containing AP or AP-CXCL14. Conditioned media were generated by transfecting 293 cells with pAP-tag5 or pAP-CXCL14 plasmids and growing them in McCoy's medium supplemented with 10% FBS (to be used for MDA-MB-231 cells) or in MCF10A media (American Type Culture Collection, Manassas, VA, to be used in MCF10A cells). Cells were counted (3 wells/time point) on days 1, 2, 4, 6, and 8 after plating. We used 10 nM CXCL12 in MDA-MB-231 cells as positive control. The experiment was repeated three times. In order to determine if CXCL14 binding to breast cancer cells has an effect on cell migration and invasion, we tested the ability of conditioned medium containing AP-CXCL14 or pCDNA3.1 expressing HA-tagged CXCL14 to induce the migration and invasion of MDA-MB-231 cells using BIOCOAT Matrigel invasion chambers essentially as described (Muller et al., 2001). For invasion assays, we plated 2.5×10^4 cells/well and assayed 24 hr later, while for migration assays we used 1.25×10^4 cells/well and determined cell numbers 12 hr later. Conditioned medium of cells transfected with pAP-Tag5 or pCDNA 3.1 empty vectors were used as negative control.

Acknowledgments

We thank Gabriela Lodeiro and Ruth Gomes for their help with the acquisition of tumor samples and Drs. Barrett Rollins and Myles Brown for critically reading the manuscript. This work is funded in part by NIH (NCI-CGAP SAGE and SPORE grant P50 CA89393-01 to K.P. and NRSA Fellowship 5F32CA94788-02 to D.P.) and DOD (grant DAMD 17 01 1 0221 to M.A.). W.R.S. and K.P. receive research support from and are consultants to the Novartis Institutes for Biomedical Research. K.P. is a consultant to GenPath Pharmaceuticals, Inc.

Received: November 12, 2003
Revised: March 16, 2004
Accepted: May 18, 2004
Published: July 19, 2004

References

Alpaugh, M.L., Lee, M.C., Nguyen, M., Deato, M., Dishakjian, L., and Barsky, S.H. (2000). Myoepithelial-specific CD44 shedding contributes to the anti-invasive and antiangiogenic phenotype of myoepithelial cells. Exp. Cell Res. *261*, 150–158.

Amiel, A., Kaufman, Z., Goldstein, E., Bruchim, R.B., Kidron, D., Gaber, E., and Fejgin, M.D. (2003). Application of comparative genomic hybridization in search for genetic aberrations in fibroadenomas of the breast. Cancer Genet. Cytogenet. *142*, 145–148.

Barbero, S., Bajetto, A., Bonavia, R., Porcile, C., Piccioli, P., Pirani, P., Ravetti, J.L., Zona, G., Spaziante, R., Florio, T., and Schettini, G. (2002). Expression of the chemokine receptor CXCR4 and its ligand stromal cell-derived factor 1 in human brain tumors and their involvement in glial proliferation in vitro. Ann. N Y Acad. Sci. *973*, 60–69.

Barsky, S.H. (2003). Myoepithelial mRNA expression profiling reveals a common tumor-suppressor phenotype. Exp. Mol. Pathol. *74*, 113–122.

Bhowmick, N.A., Chytil, A., Plieth, D., Gorska, A.E., Dumont, N., Shappell, S., Washington, M.K., Neilson, E.G., and Moses, H.L. (2004). TGF-beta signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia. Science *303*, 848–851.

Bissell, M.J., and Radisky, D. (2001). Putting tumours in context. Nat. Rev. Cancer *1*, 46–54.

Bissell, M.J., Radisky, D.C., Rizki, A., Weaver, V.M., and Petersen, O.W. (2002). The organizing principle: microenvironmental influences in the normal and malignant breast. Differentiation *70*, 537–546.

Bocker, W., Moll, R., Poremba, C., Holland, R., Van Diest, P.J., Dervan, P., Burger, H., Wai, D., Ina Diallo, R., Brandt, B., et al. (2002). Common adult stem cells in the human breast give rise to glandular and myoepithelial cell lineages: a new cell biological concept. Lab. Invest. *82*, 737–746.

Brennan, C., Zhang, Y., Leo, C., Feng, B., Cauwels, C., Aguirre, A.J., Kim, M., Protopopov, A., and Chin, L. High-resolution global profiling of genomic alterations with long oligonucleotide microarray. Cancer Res., in press.

Cai, L., Huang, H., Blackshaw, S., Liu, J.S., Cepko, C., and Wong, W.H. (2004). Clustering analysis of SAGE data using a Poisson approach. Genome Biol., in press.

Chauhan, H., Abraham, A., Phillips, J.R., Pringle, J.H., Walker, R.A., and Jones, J.L. (2003). There is more than one kind of myofibroblast: analysis of CD34 expression in benign, in situ, and invasive breast lesions. J. Clin. Pathol. *56*, 271–276.

Chen, Y., Stamatoyannopoulos, G., and Song, C.Z. (2003). Down-regulation of CXCR4 by inducible small interfering RNA inhibits breast cancer cell invasion in vitro. Cancer Res. *63*, 4801–4804.

Coussens, L.M., and Werb, Z. (2002). Inflammation and cancer. Nature *420*, 860–867.

DeCosse, J.J., Gossens, C.L., Kuzma, J.F., and Unsworth, B.R. (1973). Breast cancer: induction of differentiation by embryonic tissue. Science *181*, 1057–1058.

DeCosse, J.J., Gossens, C., Kuzma, J.F., and Unsworth, B.R. (1975). Embryonic inductive tissues that cause histologic differentiation of murine mammary carcinoma in vitro. J. Natl. Cancer Inst. *54*, 913–922.

Deng, G., Lu, Y., Zlotnikov, G., Thor, A.D., and Smith, H.S. (1996). Loss of heterozygosity in normal tissue adjacent to breast carcinomas. Science *274*, 2057–2059.

Deugnier, M.A., Teuliere, J., Faraldo, M.M., Thiery, J.P., and Glukhova, M.A. (2002). The importance of being a myoepithelial cell. Breast Cancer Res. *4*, 224–230.

De Wever, O., and Mareel, M. (2003). Role of tissue stroma in cancer cell invasion. J. Pathol. *200*, 429–447.

Dontu, G., Abdallah, W.M., Foley, J.M., Jackson, K.W., Clarke, M.F., Kawamura, M.J., and Wicha, M.S. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev. *17*, 1253–1270.

Elenbaas, B., and Weinberg, R.A. (2001). Heterotypic signaling between epithelial tumor cells and fibroblasts in carcinoma formation. Exp. Cell Res. *264*, 169–184.

Flanagan, J.G., and Leder, P. (1990). The kit ligand: a cell surface molecule altered in steel mutant fibroblasts. Cell *63*, 185–194.

Forsti, A., Louhelainen, J., Soderberg, M., Wijkstrom, H., and Hemminki, K. (2001). Loss of heterozygosity in tumour-adjacent normal tissue of breast and bladder cancer. Eur. J. Cancer *37*, 1372–1380.

Frederick, M.J., Henderson, Y., Xu, X., Deavers, M.T., Sahin, A.A., Wu, H., Lewis, D.E., El-Naggar, A.K., and Clayman, G.L. (2000). In vivo expression of the novel CXC chemokine BRAK in normal and cancerous human tissue. Am. J. Pathol. *156*, 1937–1950.

Gabbiani, G. (1999). Some historical and philosophical reflections on the myofibroblast concept. Curr. Top. Pathol. *93*, 1–5.

Gerard, C., and Rollins, B.J. (2001). Chemokines and disease. Nat. Immunol. *2*, 108–115.

Gomm, J.J., Browne, P.J., Coope, R.C., Bansal, G.S., Yiangou, C., Johnston, C.L., Mason, R., and Coombes, R.C. (1997a). A paracrine role for myoepithelial cell-derived FGF2 in the normal human breast. Exp. Cell Res. *234*, 165–173.

Gomm, J.J., Coope, R.C., Browne, P.J., and Coombes, R.C. (1997b). Separated human breast epithelial and myoepithelial cells have different growth factor requirements in vitro but can reconstitute normal breast lobuloalveolar structure. J. Cell. Physiol. *171*, 11–19.

Gudjonsson, T., Ronnov-Jessen, L., Villadsen, R., Rank, F., Bissell, M.J., and Petersen, O.W. (2002). Normal and tumor-derived myoepithelial cells differ in their ability to interact with luminal breast epithelial cells for polarity and basement membrane deposition. J. Cell Sci. *115*, 39–50.

Hall, J.M., and Korach, K.S. (2003). Stromal cell-derived factor *1*, a novel target of estrogen receptor action, mediates the mitogenic effects of estradiol in ovarian and breast cancer cells. Mol. Endocrinol. *17*, 792–803.

Hromas, R., Broxmeyer, H.E., Kim, C., Nakshatri, H., Christopherson, K., 2nd, Azam, M., and Hou, Y.H. (1999). Cloning of BRAK, a novel divergent CXC chemokine preferentially expressed in normal versus malignant cells. Biochem. Biophys. Res. Commun. *255*, 703–706.

Hunter, K.W. (2004). Host genetics and tumour metastasis. Br. J. Cancer *90*, 752–755.

Ishii, G., Sangai, T., Oda, T., Aoyagi, Y., Hasebe, T., Kanomata, N., Endoh, Y., Okumura, C., Okuhara, Y., Magae, J., et al. (2003). Bone-marrow-derived myofibroblasts contribute to the cancer-induced stromal reaction. Biochem. Biophys. Res. Commun. *309*, 232–240.

Kang, Y., Siegel, P.M., Shu, W., Drobnjak, M., Kakonen, S.M., Cordon-Cardo, C., Guise, T.A., and Massague, J. (2003). A multigenic program mediating breast cancer metastasis to bone. Cancer Cell *3*, 537–549.

Kenny, P.A., and Bissell, M.J. (2003). Tumor reversion: correction of malignant behavior by microenvironmental cues. Int. J. Cancer *107*, 688–695.

Kurose, K., Hoshaw-Woodard, S., Adeyinka, A., Lemeshow, S., Watson, P.H., and Eng, C. (2001). Genetic model of multi-step breast carcinogenesis involving the epithelium and stroma: clues to tumour-microenvironment interactions. Hum. Mol. Genet. *10*, 1907–1913.

Kurose, K., Gilley, K., Matsumoto, S., Watson, P.H., Zhou, X.P., and Eng, C. (2002). Frequent somatic mutations in PTEN and TP53 are mutually exclusive in the stroma of breast carcinomas. Nat. Genet. *32*, 355–357.

Kurth, I., Willimann, K., Schaerli, P., Hunziker, T., Clark-Lewis, I., and Moser, B. (2001). Monocyte selectivity and tissue localization suggests a role for breast and kidney-expressed chemokine (BRAK) in macrophage development. J. Exp. Med. *194*, 855–861.

Lakhani, S.R., Jacquemier, J., Sloane, J.P., Gusterson, B.A., Anderson, T.J., van de Vijver, M.J., Farid, L.M., Venter, D., Antoniou, A., Storfer-Isser, A., et al. (1998). Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. J. Natl. Cancer Inst. *90*, 1138–1145.

Lakhani, S.R., Chaggar, R., Davies, S., Jones, C., Collins, N., Odel, C., Stratton, M.R., and O'Hare, M.J. (1999). Genetic alterations in 'normal' luminal and myoepithelial cells of the breast. J. Pathol. *189*, 496–503.

Lander, E.S., and Green, P. (1987). Construction of multilocus genetic linkage maps in humans. Proc. Natl. Acad. Sci. USA *84*, 2363–2367.

Lieberfarb, M.E., Lin, M., Lechpammer, M., Li, C., Tanenbaum, D.M., Febbo, P.G., Wright, R.L., Shim, J., Kantoff, P.W., Loda, M., et al. (2003). Genome-wide loss of heterozygosity analysis from laser capture microdissected prostate cancer using single nucleotide polymorphic allele (SNP) arrays and a novel bioinformatics platform dChipSNP. Cancer Res. *63*, 4781–4785.

Lin, M., Wei, L.J., Sellers, W.R., Lieberfarb, M., Wong, W.H., and Li, C. (2004). dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics *20*, 1233–1240.

Lucito, R., Healy, J., Alexander, J., Reiner, A., Esposito, D., Chi, M., Rodgers, L., Brady, A., Sebat, J., Troge, J., et al. (2003). Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. Genome Res. *13*, 2291–2305.

Maffini, M.V., Soto, A.M., Calabro, J.M., Ucci, A.A., and Sonnenschein, C. (2004). The stroma as a crucial target in rat mammary gland carcinogenesis. J. Cell Sci. *117*, 1495–1502.

Man, Y.G., Tai, L., Barner, R., Vang, R., Saenger, J.S., Shekitka, K.M., Bratt-

hauer, G.L., Wheeler, D.T., Liang, C.Y., Vinh, T.N., and Strauss, B.L. (2003). Cell clusters overlying focally disrupted mammary myoepithelial cell layers and adjacent cells within the same duct display different immunohistochemical and genetic features: implications for tumor progression and invasion. Breast Cancer Res. *5*, R231–R241.

Matsuzaki, H., Loi, H., Dong, S., Tsai, Y.Y., Fang, J., Law, J., Di, X., Liu, W.M., Yang, G., Liu, G., et al. (2004). Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. Genome Res. *14*, 414–425.

Moinfar, F., Man, Y.G., Arnould, L., Bratthauer, G.L., Ratschek, M., and Tavassoli, F.A. (2000). Concurrent and independent genetic alterations in the stromal and epithelial cells of mammary carcinoma: implications for tumorigenesis. Cancer Res. *60*, 2562–2566.

Muller, A., Homey, B., Soto, H., Ge, N., Catron, D., Buchanan, M.E., McClanahan, T., Murphy, E., Yuan, W., Wagner, S.N., et al. (2001). Involvement of chemokine receptors in breast cancer metastasis. Nature *410*, 50–56.

Murrell, T.G. (1995). The potential for oxytocin (OT) to prevent breast cancer: a hypothesis. Breast Cancer Res. Treat. *35*, 225–229.

Nguyen, M., Lee, M.C., Wang, J.L., Tomlinson, J.S., Shao, Z.M., Alpaugh, M.L., and Barsky, S.H. (2000). The human myoepithelial cell displays a multifaceted anti-angiogenic phenotype. Oncogene *19*, 3449–3459.

Nishizaki, T., DeVries, S., Chew, K., Goodson, W.H., 3rd, Ljung, B.M., Thor, A., and Waldman, F.M. (1997). Genetic alterations in primary breast cancers and their metastases: direct comparison using modified comparative genomic hybridization. Genes Chromosomes Cancer *19*, 267–272.

Oh, S.J., Kurz, H., Christ, B., and Wilting, J. (1998). Platelet-derived growth factor-B induces transformation of fibrocytes into spindle-shaped myofibroblasts in vivo. Histochem. Cell Biol. *109*, 349–357.

Page, M.J., Amess, B., Townsend, R.R., Parekh, R., Herath, A., Brusten, L., Zvelebil, M.J., Stein, R.C., Waterfield, M.D., Davies, S.C., and O'Hare, M.J. (1999). Proteomic definition of normal human luminal and myoepithelial breast cells purified from reduction mammoplasties. Proc. Natl. Acad. Sci. USA *96*, 12589–12594.

Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D., and Brown, P.O. (1999). Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat. Genet. *23*, 41–46.

Porter, D.A., Krop, I.E., Nasser, S., Sgroi, D., Kaelin, C.M., Marks, J.R., Riggins, G., and Polyak, K. (2001). A SAGE (serial analysis of gene expression) view of breast tumor progression. Cancer Res. *61*, 5697–5702.

Porter, D., Lahti-Domenici, J., Keshaviah, A., Bae, Y.K., Argani, P., Marks, J., Richardson, A., Cooper, A., Strausberg, R., Riggins, G.J., et al. (2003a). Molecular markers in ductal carcinoma in situ of the breast. Mol. Cancer Res. *1*, 362–375.

Porter, D., Weremowicz, S., Chin, K., Seth, P., Keshaviah, A., Lahti-Domenici, J., Bae, Y.K., Monitto, C.L., Merlos-Suarez, A., Chan, J., et al. (2003b). A neural survival factor is a candidate oncogene in breast cancer. Proc. Natl. Acad. Sci. USA *100*, 10931–10936.

Radisky, D., Hagios, C., and Bissell, M.J. (2001). Tumors are unique organs defined by abnormal signaling and context. Semin. Cancer Biol. *11*, 87–95.

Rajagopalan, H., Nowak, M.A., Vogelstein, B., and Lengauer, C. (2003). The significance of unstable chromosomes in colorectal cancer. Nat. Rev. Cancer *3*, 695–701.

Ramaswamy, S., Ross, K.N., Lander, E.S., and Golub, T.R. (2003). A molecular signature of metastasis in primary solid tumors. Nat. Genet. *33*, 49–54.

Rohan, R.M., Fernandez, A., Udagawa, T., Yuan, J., and D'Amato, R.J. (2000). Genetic heterogeneity of angiogenesis in mice. FASEB J. *14*, 871–876.

Ronnov-Jessen, L., Petersen, O.W., Koteliansky, V.E., and Bissell, M.J. (1995). The origin of the myofibroblasts in breast cancer. Recapitulation of tumor environment in culture unravels diversity and implicates converted fibroblasts and recruited smooth muscle cells. J. Clin. Invest. *95*, 859–873.

Ronnov-Jessen, L., Petersen, O.W., and Bissell, M.J. (1996). Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. Physiol. Rev. *76*, 69–125.

Rossi, D., and Zlotnik, A. (2000). The biology of chemokines and their receptors. Annu. Rev. Immunol. *18*, 217–242.

Rubin, J.B., Kung, A.L., Klein, R.S., Chan, J.A., Sun, Y., Schmidt, K., Kieran, M.W., Luster, A.D., and Segal, R.A. (2003). A small-molecule antagonist of CXCR4 inhibits intracranial growth of primary brain tumors. Proc. Natl. Acad. Sci. USA *100*, 13513–13518.

Sawyer, E.J., Hanby, A.M., Ellis, P., Lakhani, S.R., Ellis, I.O., Boyle, S., and Tomlinson, I.P. (2000). Molecular analysis of phyllodes tumors reveals distinct changes in the epithelial and stromal components. Am. J. Pathol. *156*, 1093–1098.

Sawyer, E.J., Hanby, A.M., Rowan, A.J., Gillett, C.E., Thomas, R.E., Poulsom, R., Lakhani, S.R., Ellis, I.O., Ellis, P., and Tomlinson, I.P. (2002). The Wnt pathway, epithelial-stromal interactions, and malignant progression in phyllodes tumours. J. Pathol. *196*, 437–444.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467–470.

Schurch, W. (1999). The myofibroblast in neoplasia. Curr. Top. Pathol. *93*, 135–148.

Scotton, C.J., Wilson, J.L., Scott, K., Stamp, G., Wilbanks, G.D., Fricker, S., Bridger, G., and Balkwill, F.R. (2002). Multiple actions of the chemokine CXCL12 on epithelial tumor cells in human ovarian cancer. Cancer Res. *62*, 5930–5938.

Shao, Z.M., Nguyen, M., Alpaugh, M.L., O'Connell, J.T., and Barsky, S.H. (1998). The human myoepithelial cell exerts antiproliferative effects on breast carcinoma cells characterized by p21WAF1/CIP1 induction, G2/M arrest, and apoptosis. Exp. Cell Res. *241*, 394–403.

Shekhar, M.P., Pauley, R., and Heppner, G. (2003). Host microenvironment in breast cancer development: extracellular matrix-stromal cell contribution to neoplastic phenotype of epithelial cells in the breast. Breast Cancer Res. *5*, 130–135.

Sleeman, M.A., Fraser, J.K., Murison, J.G., Kelly, S.L., Prestidge, R.L., Palmer, D.J., Watson, J.D., and Kumble, K.D. (2000). B cell- and monocyte-activating chemokine (BMAC), a novel non-ELR alpha-chemokine. Int. Immunol. *12*, 677–689.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA *98*, 10869–10874.

St Croix, B., Rago, C., Velculescu, V., Traverso, G., Romans, K.E., Montgomery, E., Lal, A., Riggins, G.J., Lengauer, C., Vogelstein, B., and Kinzler, K.W. (2000). Genes expressed in human tumor endothelium. Science *289*, 1197–1202.

Sternlicht, M.D., and Barsky, S.H. (1997). The myoepithelial defense: a host defense against cancer. Med. Hypotheses *48*, 37–46.

Sternlicht, M.D., Kedeshian, P., Shao, Z.M., Safarians, S., and Barsky, S.H. (1997). The human myoepithelial cell is a natural tumor suppressor. Clin. Cancer Res. *3*, 1949–1958.

Tamamura, H., Hori, A., Kanzaki, N., Hiramatsu, K., Mizumoto, M., Nakashima, H., Yamamoto, N., Otaka, A., and Fujii, N. (2003). T140 analogs as CXCR4 antagonists identified as anti-metastatic agents in the treatment of breast cancer. FEBS Lett. *550*, 79–83.

Tanaka, K., Sano, K., Yuba, K., Katsumura, K., Nakano, T., Kobayashi, M., Ikeda, T., and Abe, M. (2003). Inhibition of induction of myofibroblasts by interferon gamma in a human fibroblast cell line. Int. Immunopharmacol. *3*, 1273–1280.

Tlsty, T.D. (2001). Stromal cells can contribute oncogenic signals. Semin. Cancer Biol. *11*, 97–104.

Tlsty, T.D., and Hein, P.W. (2001). Know thy neighbor: stromal cells can contribute oncogenic signals. Curr. Opin. Genet. Dev. *11*, 54–59.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M.,

Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536.

van de Vijver, M.J., He, Y.D., van 't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. *347*, 1999–2009.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. Science *270*, 484–487.

Waldman, F.M., DeVries, S., Chew, K.L., Moore, D.H., 2nd, Kerlikowske, K., and Ljung, B.M. (2000). Chromosomal alterations in ductal carcinomas in situ and their in situ recurrences. J. Natl. Cancer Inst. *92*, 313–320.

Walter-Yohrling, J., Pratt, B.M., Ledbetter, S., and Teicher, B.A. (2003). Myofibroblasts enable invasion of endothelial cells into three-dimensional tumor cell clusters: a novel in vitro tumor model. Cancer Chemother. Pharmacol. *52*, 263–269.

Wernert, N., Locherbach, C., Wellmann, A., Behrens, P., and Hugel, A. (2001). Presence of genetic alterations in microdissected stroma of human colon and breast cancers. Anticancer Res. *21*, 2259–2264.

Wiseman, B.S., and Werb, Z. (2002). Stromal effects on mammary gland development and breast cancer. Science *296*, 1046–1049.

Xiao, G., Liu, Y.E., Gentz, R., Sang, Q.A., Ni, J., Goldberg, I.D., and Shi, Y.E. (1999). Suppression of breast cancer growth and metastasis by a serpin myoepithelium-derived serine proteinase inhibitor expressed in the mammary myoepithelial cells. Proc. Natl. Acad. Sci. USA *96*, 3700–3705.

Zhang, R.R., Man, Y.G., Vang, R., Saenger, J.S., Barner, R., Wheeler, D.T., Liang, C.Y., Vinh, T.N., and Bratthauer, G.L. (2003). A subset of morphologically distinct mammary myoepithelial cells lacks corresponding immunophenotypic markers. Breast Cancer Res. *5*, R151–R156.

Zhu, G., Reynolds, L., Crnogorac-Jurcevic, T., Gillett, C.E., Dublin, E.A., Marshall, J.F., Barnes, D., D'Arrigo, C., Van Trappen, P.O., Lemoine, N.R., and Hart, I.R. (2003). Combination of microdissection and microarray analysis to identify gene expression changes between differentially located tumour cells in breast cancer. Oncogene *22*, 3742–3748.