Cancer Cell
Article

Molecular Definition of Breast Tumor Heterogeneity

Michail Shipitsin,[1,3] Lauren L. Campbell,[1,5] Pedram Argani,[6] Stanislawa Weremowicz,[4,10] Noga Bloushtain-Qimron,[1] Jun Yao,[1,3] Tatiana Nikolskaya,[7,8] Tatiana Serebryiskaya,[7,8] Rameen Beroukhim,[1,3] Min Hu,[1,3] Marc K. Halushka,[6] Saraswati Sukumar,[6] Leroy M. Parker,[1,3] Karen S. Anderson,[1,3] Lyndsay N. Harris,[1,3] Judy E. Garber,[1,3] Andrea L. Richardson,[4,10] Stuart J. Schnitt,[4,11] Yuri Nikolsky,[7] Rebecca S. Gelman,[2,9] and Kornelia Polyak[1,3,5,*]
[1] Department of Medical Oncology
[2] Department of Biostatistics and Computational Biology
Dana-Farber Cancer Institute, Boston, MA 02115, USA
[3] Department of Medicine
[4] Department of Pathology
[5] Program in Biological and Biomedical Sciences
Harvard Medical School, Boston, MA 02115, USA
[6] Johns Hopkins University School of Medicine, Baltimore, MD 21231, USA
[7] GeneGo, Inc., St. Joseph, MI 49085, USA
[8] Vavilov Institute for General Genetics, Moscow B333, 117809, Russia
[9] Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA
[10] Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA
[11] Department of Pathology, Beth-Israel Deaconess Medical Center, Boston, MA 02115, USA
*Correspondence: kornelia_polyak@dfci.harvard.edu
DOI 10.1016/j.ccr.2007.01.013

SUMMARY

Cells with distinct phenotypes including stem-cell-like properties have been proposed to exist in normal human mammary epithelium and breast carcinomas, but their detailed molecular characteristics and clinical significance are unclear. We determined gene expression and genetic profiles of cells purified from cancerous and normal breast tissue using markers previously associated with stem-cell-like properties. CD24+ and CD44+ cells from individual tumors were clonally related but not always identical. CD44+ cell-specific genes included many known stem-cell markers and correlated with decreased patient survival. The TGF-β pathway was specifically active in CD44+ cancer cells, where its inhibition induced a more epithelial phenotype. Our data suggest prognostic relevance of CD44+ cells and therapeutic targeting of distinct tumor cell populations.

INTRODUCTION

Tumors originate from normal cells due to accumulated genetic and epigenetic alterations, but the identity of the tumor-initiating cells is largely unknown. Stem cells have been proposed as attractive targets since they share many characteristics with cancer cells, including the capacity to self-renew, give rise to heterogeneous progeny, and migrate and invade into surrounding tissue. Correlating with this, several pathways and genes required for normal stem-cell function are activated in cancer cells and play essential roles in tumorigenesis (Weissman, 2005). Cancer stem cells have been defined as a subset of tumor cells with stem-cell-like properties that are

SIGNIFICANCE

Clonal evolution is a prevailing concept of cancer biology explaining the initiation and progression of solid tumors. The cancer-stem-cell hypothesis is an alternative model that recently has received a lot of attention. Here, we determined molecular profiles of distinct breast cancer cell populations purified using markers previously associated with stem-cell-like properties. Although we found that CD24+ and CD44+ cells are more differentiated and progenitor-like, respectively, and that their gene expression differences may have prognostic relevance, our finding of genetic differences between these cells within individual tumors supports a clonal evolution hypothesis involving intratumoral heterogeneity.

thought to be responsible for the growth, progression, and recurrence of a tumor (Clarke and Fuller, 2006; Lynch et al., 2006; Polyak and Hahn, 2006; Weissman, 2005; Wicha et al., 2006). This hypothesis was proposed many years ago and revived in recent years with new experimental approaches. Putative cancer stem cells have been purified from various human tumor types, frequently using cell surface markers specific for the normal stem cells of the same organ (Clarke and Fuller, 2006; Lynch et al., 2006; Polyak and Hahn, 2006; Weissman, 2005; Wicha et al., 2006). The tumorigenicity and the "stemness" of these isolated cells have been demonstrated by performing in vitro clonogenicity and in vivo tumorigenicity studies.

The stages of human mammary epithelial cell differentiation and markers uniquely identifying differentiated and progenitor cells are not well defined. In vitro clonogenicity studies suggest the existence of bipotential progenitors that can give rise to both luminal epithelial and myoepithelial cells, lineage-restricted luminal epithelial and myoepithelial progenitors, and differentiated luminal epithelial and myoepithelial cells (Bocker et al., 2002; Clarke et al., 2005; Clayton et al., 2004; Dontu et al., 2003; Liu et al., 2006; Lynch et al., 2006; Stingl et al., 1998, 2005). MUC1 and CD10 (CALLA/MME) are thought to be expressed by luminal epithelial and myoepithelial cells, respectively (Clayton et al., 2004; Stingl et al., 1998, 2005). CD44 is present in progenitor cells, yet which epithelial cells express CD24 has not been determined in the normal human breast. In breast cancer, Al-Hajj et al. demonstrated that $lin^-/CD44^+/CD24^{-/low}$ (subsequently referred to as CD44+) cells from malignant pleural effusions of breast cancer patients were more tumorigenic in NOD/SCID mice than $CD44^-/CD24^+$ (subsequently referred to as CD24+) cells, and the resulting xenografts reproduced the heterogeneity of the original tumors, leading to the hypothesis that CD44+ cells are breast cancer stem cells (Al-Hajj et al., 2003). Subsequent studies in breast and prostate cancer confirmed that CD44+ cells are tumorigenic when injected into immunodeficient mice and have progenitor-like properties (Patrawala et al., 2005; Ponti et al., 2005). None of these studies have analyzed the comprehensive molecular profiles and clinical relevance of CD44+ and CD24+ cells or provided conclusive evidence that CD44+ tumor cells are stem cells and CD24+ tumor cells are their progeny. Furthermore, the relationship between the cancer-stem-cell and the genetic-clonal-evolution hypotheses of tumor progression has not been investigated.

To begin addressing these issues, we purified cells using CD44 and CD24 from breast carcinomas and determined their global gene expression and genetic profiles. For comparison, we also isolated and characterized CD44+ and CD24+ cells from normal human mammary epithelium. Gene signatures specific for CD44+ breast cancer cells were enriched for known stem-cell markers and correlated with clinical outcome and activity of signaling pathways, but cancer CD24+ and CD44+ cells were not always genetically identical.

RESULTS

Purification and Gene Expression Profiles of Distinct Breast Epithelial Cell Populations

We purified distinct subpopulations of normal mammary epithelial and breast cancer cells using cell surface markers that have been described to be specifically expressed in differentiated cells or in cells with stem-cell-like properties (Al-Hajj et al., 2003; Dontu et al., 2003; Stingl et al., 1998) (Figure 1A). We purified CD24+ and CD44+ cells from normal human breast tissue derived from reduction mammoplasty and from breast tumors of different stages. Based on the initial gene expression profiling of CD44+ cells from pleural effusion and ascites samples, we identified a CD44+ cell-specific gene (PROCR) that encodes a cell surface receptor and has expression more specific to CD44+ epithelial cells than CD44, which is also expressed in leukocytes and myofibroblasts. PROCR has been identified as a marker of hematopoietic, hair-follicle, neural, and embryonic stem cells (Blanpain et al., 2004; Ivanova et al., 2002; Ramalho-Santos et al., 2002). After confirming that 100% of CD44+ tumor cells are positive for PROCR (data not shown), we used this marker to isolate $CD44^+/PROCR^+$ (subsequently referred to as PROCR+) cells from primary invasive breast tumors (see Figure S1A in the Supplemental Data available with this article online). A detailed description of the purification procedure and the tissue samples is provided in the Supplemental Data.

We confirmed the purity and differentiation status of the cell fractions by semiquantitative RT-PCR using leukocyte, luminal epithelial, myoepithelial, and stem-cell markers (Dontu et al., 2003) (Figures 1B–1D and Figure S1B). The nearly mutually exclusive expression of known luminal epithelial and stem-cell markers in the CD24+ and CD44+ cells, respectively, suggested that they might indeed represent luminal epithelial and progenitor-like cells, respectively. The low abundance of estrogen receptor α (ESR1) in the CD44+ cells both in normal breast tissue and breast carcinomas implies that these cells are not responsive to estrogenic hormones (Figures 1B–1D and Table 1II). The lack of estrogen and progesterone receptors and ERBB2 in normal CD44+ cells further supports that these cells may represent progenitors since mouse mammary epithelial stem cells do not express these proteins (Asselin-Labat et al., 2006). We also analyzed the expression of genes associated with self-renewal, including BMI1 and hedgehog (Hh)-signaling pathway genes. Gli1 and Gli2 were more highly expressed in CD44+ cells than in CD24+ cells, reflecting activation of Hh signaling, while BMI1 expression was essentially the same in both cell populations (Figures S1C and S1D).

To determine the comprehensive gene expression profiles of the purified cells, we generated SAGE (serial analysis of gene expression) libraries from CD24+ and CD44+ cells purified from normal mammary epithelium and pleural effusion, ascites, and primary invasive tumor

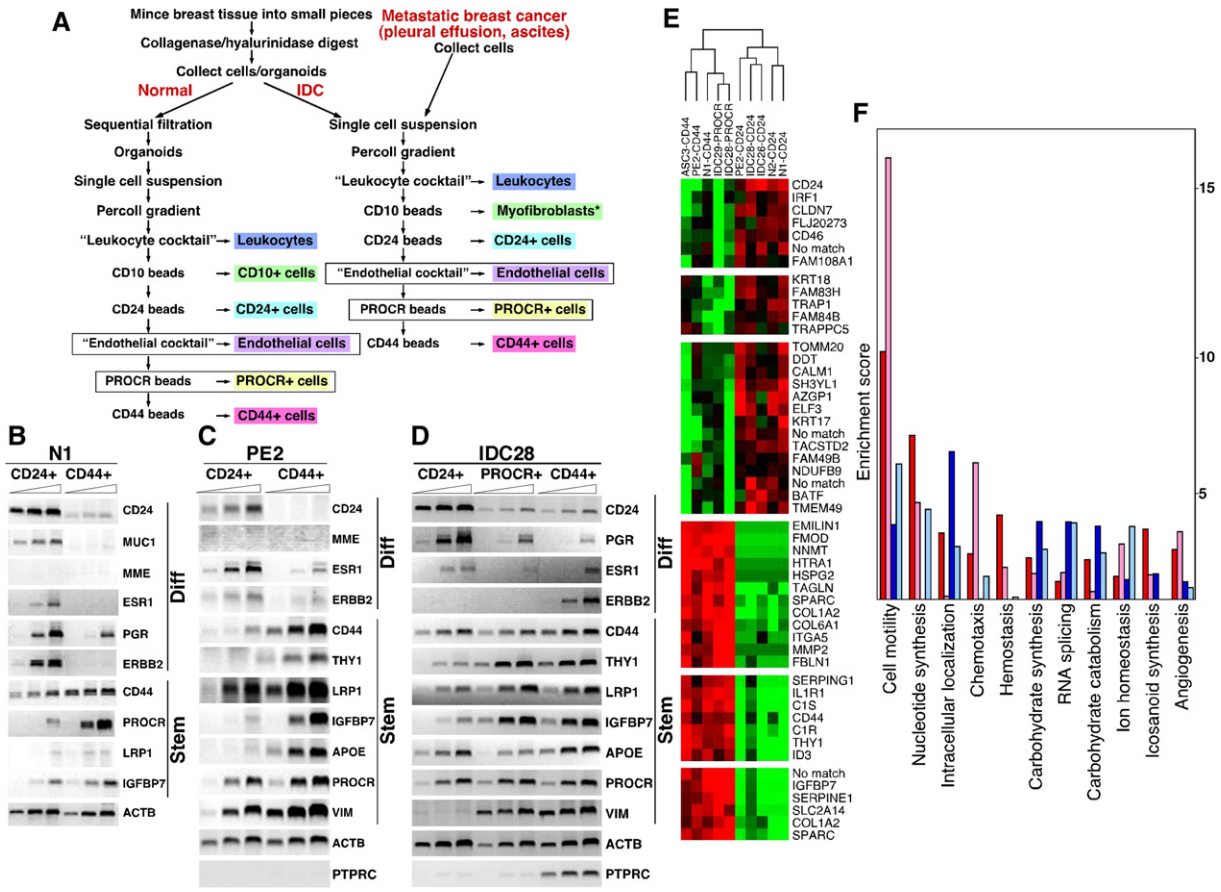

Figure 1. Purification and Gene Expression Analysis of Distinct Cell Subpopulations

(A) Schematic outline of purification of the various cells from normal breast tissue and invasive and metastatic breast carcinomas. Cells are captured using the indicated antibody-coupled magnetic beads specific for each cell type. Purification steps marked with rectangles were not always included in the procedure, while myofibroblasts, marked with an asterisk, were only present in invasive tumors. IDC, invasive ductal carcinoma. Semiquantitative RT-PCR analysis of purified cell fractions isolated from normal breast tissue (N1) (B), pleural effusion (PE2) (C), and primary invasive ductal carcinoma (IDC28) (D). RNA from CD24+, CD44+, and PROCR+ cells was tested for the expression of known differentiated (Diff) and stem (Stem)-cell-specific genes. CD44+ and PROCR+ cells lack differentiation markers and are positive for stem-cell markers. In the primary invasive tumor, the CD44+ fraction is contaminated by leukocytes, as demonstrated by high levels of CD45 leukocyte common antigen (PTPRC) expression. ACTB was used as loading control. Each triangle indicates an increasing number of PCR cycles (25, 30, 35). (E) Dendrogram depicting relatedness of SAGE libraries prepared from CD44+, PROCR+, and CD24+ cells. Hierarchical clustering was applied to SAGE data for the indicated libraries, and selected portions of the clustering heat map are shown here. Each row represents a tag and is labeled with the symbol of the gene that best matches that tag (or "no match" if no matching transcript was found). Red and green indicate high and low expression levels, respectively. The expression profiles of normal and cancer CD44+ and PROCR+ cells are more similar to each other than to those of CD24+ cells derived from the same tissues. ASC, ascites; PE, pleural effusion; N, normal; IDC, invasive ductal carcinoma. (F) Gene ontology biological process categories highly represented in pools of all SAGE libraries from different cell populations. Categories with an enrichment score >2 in at least one library pool using the DAVID Functional Annotation Tool are plotted. Cell populations represented include cancer CD44+ and PROCR+ (red), normal CD44+ and PROCR+ (pink), cancer CD24+ (dark blue), and normal CD24+ (light blue) cells.

samples collected from breast cancer patients (details in the Supplemental Data). We assigned each SAGE library a name based on the sample and cell surface marker used to purify the cells from which it was prepared. The SAGE data further supported the hypothesis that CD24+ and CD44+ cells represent more differentiated luminal epithelial and progenitor-like cells, respectively, since known markers of these cells were nearly mutually exclusively found in the respective SAGE libraries (Tables 1I and 1II and Tables S1–S6). Unsupervised hierarchical clustering of the SAGE libraries demonstrated that normal and cancer CD44+ and PROCR+ cells are more similar to each other than to CD24+ cells from the same tissue (Figure 1E). Functional classification of the genes expressed in the various SAGE libraries revealed that cancer and normal CD44+ and PROCR+ cells are more enriched for genes involved in cell motility, chemotaxis, hemostasis, and angiogenesis, while CD24+ cells more highly express genes implicated in carbohydrate metabolism and RNA splicing (Figure 1F).

Table 1. Selected Genes Differentially Expressed between CD44+ or PROCR+ and CD24+ Cells

	Tag	N1-CD44	PE2-CD44	ASC3-CD44	IDC28-PROCR	IDC29-PROCR	N1-CD24	N2-CD24	PE2-CD24	IDC26-CD24	IDC28-CD24	Gene Symbol	Gene Description
I	ATATGTATATTGCTGAG	78	30	27	11	19	34	44	8	27	22	*CD44*	CD44 molecule
	TGTCCTGGTTCCCGTTT	121	23	114	228	81	19	15	4	31	44	*CDKN1A*	Cyclin-dependent kinase inhibitor 1A
	CAGGCTGATGGGCCCCG	39	0	4	15	4	0	0	4	0	0	*CHRD*	Chordin
	AAAGAAATGGTGCTACC	0	84	421	0	11	0	0	4	0	0	*H19*	H19, imprinted maternally expressed untranslated mRNA
	AACCACTGCTACTCCCG	53	23	68	34	15	0	0	0	0	6	*ID3*	Inhibitor of DNA binding 3
	TATATATAGAGATGTTC	7	34	61	19	4	0	0	0	0	0	*ID4*	Inhibitor of DNA binding 4
	CATATCATTAAACAAAT	135	27	19	293	359	0	0	0	0	3	*IGFBP7*	Insulin-like growth factor binding protein 7
	CTCAACCCCCCTCCCAG	64	88	11	38	19	0	0	0	0	0	*LRP1*	Low density lipoprotein-related protein 1
	TCTCCTGCATAGCTTTT	57	0	0	46	30	0	0	0	4	0	*MSX1*	Msh homeobox homolog 1
	TTCCATAGCCTTGCTGG	89	0	0	27	15	5	10	8	0	0	*NOTCH3*	Notch homolog 3
	TCTTCTTCGAAGTGGCT	85	38	144	19	0	0	0	4	0	0	*PROCR*	Protein C receptor, endothelial
	GAGTGAGACCCAGGAGC	25	50	235	46	30	0	0	0	0	3	*THY1*	Thy-1 cell surface antigen
	TCCAAATCGATGTGGAT	1038	332	504	472	300	101	5	0	0	0	*VIM*	Vimentin
II	GGAACAAACAGATCGAA	18	0	0	15	0	168	102	48	354	166	*CD24*	CD24 molecule
	TTTTTCTATTAAAAATA	0	4	4	0	0	43	0	0	0	0	*CLDN1*	Claudin 1
	GGAGAGAAAACAAACCT	0	0	4	0	0	5	5	12	13	3	*CLDN12*	Claudin 12
	TATAGTCCTCTTGGGTT	11	4	0	4	0	29	29	40	13	53	*CLDN7*	Claudin 7
	ACAGCGGCAATCTTTTC	4	8	4	4	7	43	15	88	76	19	*DSP*	Desmoplakin
	AGGAAGGAACAGCAATG	0	4	11	8	0	10	5	24	4	3	*ERBB2*	V-erb-b2 erythroblastic leukemia viral oncogene homolog 2
	AGCAGGTGCCTGAGACA	0	0	0	0	4	0	0	12	13	3	*ESR1*	Estrogen receptor 1
	GTTTATTCTTTTTCTTT	0	0	0	0	7	5	0	32	67	13	*FOXA1*	Forkhead box A1
	CTTTGTGAACAAGTCCC	0	4	0	8	22	29	5	96	54	34	*GATA3*	GATA binding protein 3
	GTGTGGGGGGCTGGGGG	18	15	38	34	19	125	82	64	72	131	*JUP*	Junction plakoglobin
	CAAACCATCCAAAAGAC	7	38	80	4	26	168	44	119	54	147	*KRT18*	Keratin 18
	CCTCCAGCTACAAAACA	4	80	174	15	44	96	44	187	63	350	*KRT8*	Keratin 8
	CCTGGGAAGTGTTGTGG	0	27	8	4	0	29	5	231	0	6	*MUC1*	Mucin 1, cell surface associated
	CTGGCCCTCGGCACCCT	0	4	0	15	226	5	0	227	193	134	*TFF1*	Trefoil factor 1
	CTCCACCCGAGGACAGT	0	8	0	0	181	19	0	291	273	97	*TFF3*	Trefoil factor 3
	CAATTAAAAGGTACAAT	18	19	8	15	59	221	10	155	143	88	*XBP1*	X-box binding protein 1
III	ACCAAAAACCAAAAGTG	462	484	705	924	374	0	5	24	143	13	*COL1A1*	Collagen, type I, alpha 1
	TTTGCACCTTTCTAGTT	71	53	68	232	189	34	5	0	9	6	*CTGF*	Connective tissue growth factor
	ATCTTGTTACTGTGATA	7	46	30	91	26	0	0	0	0	0	*FN1*	Fibronectin 1
	TAAATCCCCACTGGGAC	78	15	0	4	4	0	0	0	0	0	*MMP9*	Matrix metallopeptidase 9
	TAAAAATGTTTCAAAAA	68	34	30	205	396	0	0	0	0	6	*SERPINE1*	Serpin peptidase inhibitor, clade E, member 1
	ATGTGAAGAGTTTCACA	281	160	300	848	715	0	0	8	13	19	*SPARC*	Secreted protein, acidic, cysteine-rich
	ACAGGCTACGGACGACC	270	23	186	137	67	0	5	0	0	0	*TAGLN*	Transgelin
	GGGGCTGTATTTAAGGA	7	23	19	27	52	0	0	0	0	34	*TGFB1*	Transforming growth factor, beta 1
	CACCTTCTGCCTGCGCC	18	11	19	19	11	0	5	4	0	3	*TGFB1I1*	Transforming growth factor beta 1 induced transcript 1
	TGCCACACAGTGACTTG	46	4	8	27	19	5	5	0	0	0	*TGFB3*	Transforming growth factor, beta 3
	GTGTGTTTGTAATAATA	92	65	19	19	4	0	0	0	0	0	*TGFBI*	Transforming growth factor, beta-induced, 68kDa
	TTTTGATACCCCTTTTT	7	0	0	8	7	0	5	0	0	0	*TGFBR1*	Transforming growth factor, beta receptor I
	TGATTTTTTTTTTCCTC	0	8	4	0	11	0	5	0	0	0	*TGIF*	TGFB-induced factor
	CTGCTGAATTCTGGTTG	25	0	8	11	7	5	10	0	0	22	*TGIF2*	TGFB-induced factor 2
	AGGTCTTCAATACTGTT	82	0	0	68	48	5	10	4	4	9	*THBS1*	Thrombospondin 1

Matching SAGE tag sequences, counts per 200,000 tags in each indicated SAGE library, and gene symbols and descriptions are listed for genes encoding known stem (I) or differentiated (II) cell markers or TGF-β pathway components and targets (III). N, normal; PE, pleural effusion; ASC, ascites; IDC, invasive ductal carcinoma.

Genetic Alterations in CD24+ and CD44+ Breast Cancer Cells

To determine the clonal relationship between CD24+ and CD44+ cells within individual tumors, genomic DNA isolated from these cells was analyzed on SNP arrays. The resulting data demonstrated that in some tumors the two cell fractions have the same copy-number alterations and appear to be genetically identical (Figure 2A). To analyze the genetic composition of the CD24+ and CD44+ cells at the single-cell level, we developed a protocol for their short-term in vitro culture that maintained their phenotypes based on the analysis of cell surface and cell-type-specific markers (Figures 2B and 4C). These primary cultures were used for metaphase FISH analyses using BAC probes corresponding to chromosomal areas frequently amplified in breast cancer and containing genes specifically and highly expressed in tumor CD24+ or CD44+ cells. This analysis identified a genetic alteration that was common to all cell fractions (8q24.3 rearrangement) as well as one only detected in CD24+ cells (1q21.3 gain) (Figure 2C). This distinct genetic change was detected in all CD24+ cells and was present in the original tumor (Figure 2D). Thus, we determined that CD44+ and CD24+ cells within individual tumors are

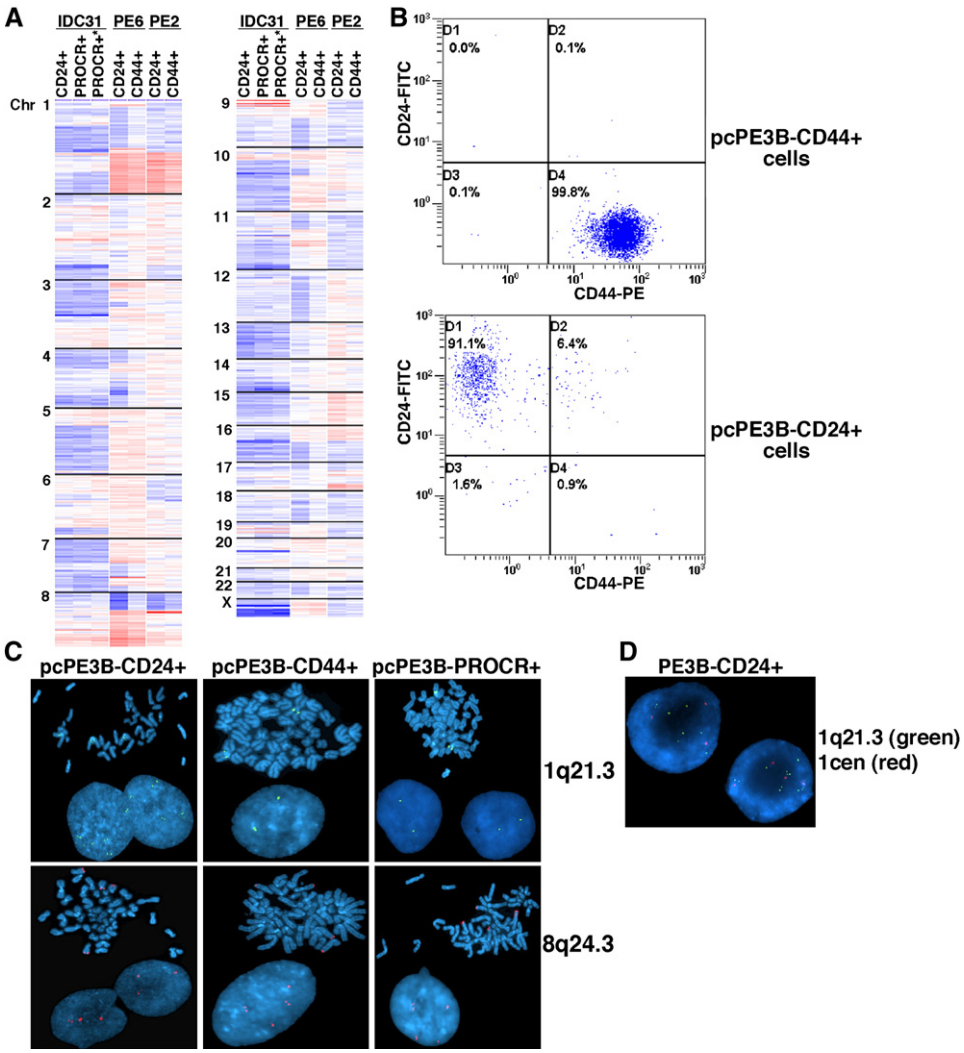

Figure 2. Analysis of the Genetic Profiles of Distinct Tumor Cell Subpopulations

(A) SNP array analysis of copy-number changes relative to normal leukocytes from the same patient in CD24+, CD44+, and PROCR+ cells isolated from pleural effusions (PE2 and PE6) and an invasive ductal carcinoma (IDC31). PROCR+* cells are PROCR+ cells further selected with CD44 antibodies. Red and blue indicate copy-number gain and loss, respectively. The pairs of distinct cell populations from each tumor overall appear to have identical copy number changes. Chr, chromosome.

(B) FACS analysis of primary cultured pleural effusion (pcPE) CD44+ and CD24+ cells using anti-CD24 and anti-CD44 antibodies conjugated with FITC and PE, respectively. CD44 and CD24 are mutually exclusively expressed in the CD44+ and CD24+ cells.

(C) FISH analysis of metaphase chromosomes and interphase nuclei prepared from primary cultured pleural effusion (pcPE) CD24+, CD44+, and PROCR+ cells using BAC RP11-157A11 mapped to 1q21.3 (green) and BAC CTD-2349A18 mapped to 8q24.3 (red). FISH with the 8q24.3-specific probe revealed a hybridization pattern characteristic of rearrangement at this locus (two brighter, larger and two weaker, smaller signals) in all three cell fractions. FISH with the 1q21.3-specific probe revealed multiple signals (four to seven) on metaphase chromosomes and interphase nuclei, characteristic of a gain of chromosomal material from the long (q) arm of chromosome 1.

(D) FISH analysis of uncultured pleural effusion (PE) CD24+ cells using BAC RP11-157A11 mapped to 1q21.3 and a chromosome 1 centromeric (cen) probe (D1Z5). Representative interphase nuclei are shown with multiple hybridization signals specific to the 1q21.3 locus (green) as well as three and four hybridization signals specific to the centromeric region (red), consistent with a gain of chromosome 1 material.

clonally related but that the CD24+ cells can have additional genetic changes not present in the CD44+ cells.

Signaling Pathways Active in CD44+ Cells

To identify signaling pathways based on our SAGE data that are specifically activated in CD24+ or CD44+ cells, we utilized the MetaCore data mining technology recently applied for functional analysis of high-throughput data (Ekins et al., 2006; Jarvinen et al., 2006; Lee et al., 2006; Nikolsky et al., 2005a, 2005b; Soreghan et al., 2005). First, we ranked gene ontology functional processes and canonical pathways according to how statistically significantly they fit to the lists of genes most highly overrepresented in normal or cancer CD44+ cell SAGE libraries

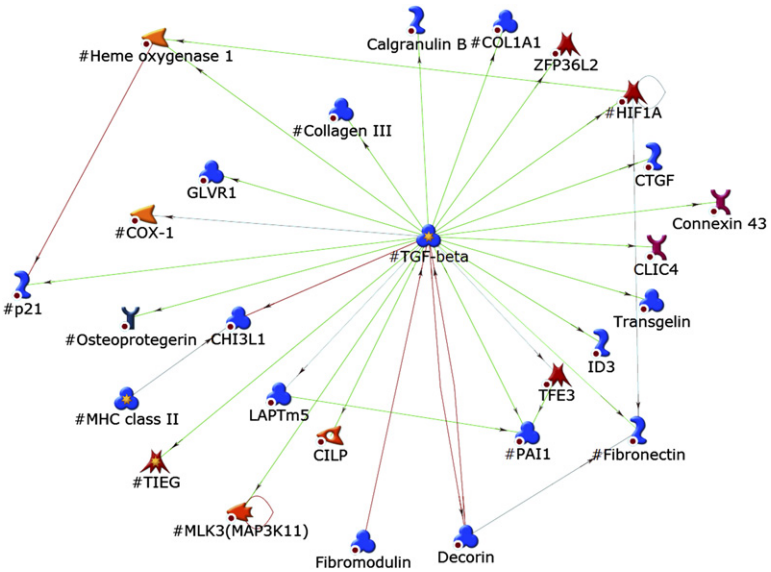

Figure 3. Network of TGF-β Signaling Pathway Upregulated in Cancer CD44+ Cells

Direct interaction network centered around TGF-β for genes upregulated in cancer and not normal CD44+ cells compared to corresponding CD24+ cells. Colors of the lines indicate inhibition (red), activation (green), and no clear link (gray).

compared to corresponding CD24+ cell SAGE libraries (Figures S2A and S2B and data not shown). Among gene ontology functional processes, cell motility, cell adhesion, protein biosynthesis, protein folding, and cell proliferation strongly fit genes upregulated in normal CD44+ cells and genes upregulated in cancer CD44+ cells. Cell motility and cell adhesion had high dispersion scores in this analysis, indicating that different sets of genes within each process are upregulated in cancer and normal CD44+ cells. Among canonical pathways, TGF-β and WNT signaling, cytoskeleton remodeling, integrin-mediated processes, reverse signaling by ephrin B, and chemokines and cell adhesion strongly fit genes upregulated in normal CD44+ cells and genes upregulated in cancer CD44+ cells. A map of the genes involved in TGF-β and WNT signaling and cytoskeleton remodeling and a more detailed view of the TGF-β pathway are depicted in Figures S3 and S4.

We also used the MetaCore data mining technology to perform a more detailed analysis of which pathways are differentially regulated in CD44+ and CD24+ cells by building direct interaction (DI) networks around the lists of genes most highly overrepresented in normal or cancer CD44+ cell SAGE libraries compared to corresponding CD24+ cell SAGE libraries. A DI network built around cancer CD44+ cell-specific genes is centered around TGF-β1 (Figure 3 and Figure S5). Modules around TGF-β1, dynamin, fibronectin, caveolin, and casein kinase II are upregulated only in cancer CD44+ cells, while modules around collagen 1 and transcription factors HIF1A and ETS1 are upregulated in cancer and normal CD44+ cells (Figures S6 and S7). DI networks built around normal CD44+ cell-specific genes are centered around VEGF-A, IL-1, NF-kβ, AP-1, Rac1, and SMAD3 and feature activation of the Notch pathway and TGF-β3 (Figures S8, S10, and S13). A subnetwork of the cancer CD44+ cell-specific genes DI network centered around TGF-β1 containing breast-cancer-relevant genes is shown in Figure S9, and additional networks built around genes upregulated in CD44+ cells from individual cancer samples are shown in Figures S11, S12, S14, and S15.

The TGF-β Pathway in Breast Cancer CD44+ Cells

Due to the known importance of TGF-β signaling in regulating the pluripotency of human embryonic stem cells (James et al., 2005); its roles in differentiation, tumorigenesis, and metastasis (Moses and Serra, 1996; Muraoka-Cook et al., 2005; Roberts and Wakefield, 2003; Siegel and Massague, 2003); and our SAGE (Table 1III) and MetaCore data implicating this pathway in CD44+ cells, we investigated its role in breast cancer CD44+ cells in further detail. We analyzed the expression of selected genes involved in TGF-β signaling by semiquantitative RT-PCR in CD44+ and CD24+ cells (Figure 4A). Correlating with our SAGE and MetaCore data, TGF-β1 was almost exclusively expressed in tumor CD44+ cells together with TGFBR2, one of the signaling receptors for TGF-β (Figure 4A). The dramatic difference in TGFBR2 mRNA levels between tumor CD24+ and CD44+ cells suggests potential underlying epigenetic regulation of this gene. Chromatin modification and promoter methylation have been described as potential mechanisms leading to silencing of TGFBR2 in some tumors (Osada et al., 2005; Zhang et al., 2004; Zhao et al., 2005). We analyzed the DNA methylation status of TGFBR2 in tumor CD44+ and CD24+ cells and determined that it is hypermethylated in CD24+ cells, potentially explaining its lack of expression in these cells (Figure 4B). This result also suggests that tumor CD24+ and CD44+ cells are epigenetically distinct, a finding confirmed by more extended analyses (N.B.-Q., J.Y., S.A. Mani, M.S., M.H., H. Chen, J.E. Antosiewicz, P.A., M.K.H., J.A. Thomson, P. Pharoah, A. Porgador, S.S., R. Parsons, A.L.R., R.S.G., K.P., unpublished data).

To test the functional relevance of the activation of TGF-β signaling in cancer CD44+ cells, we investigated the effect of a dual TGFBR1/TGFBR2 kinase inhibitor

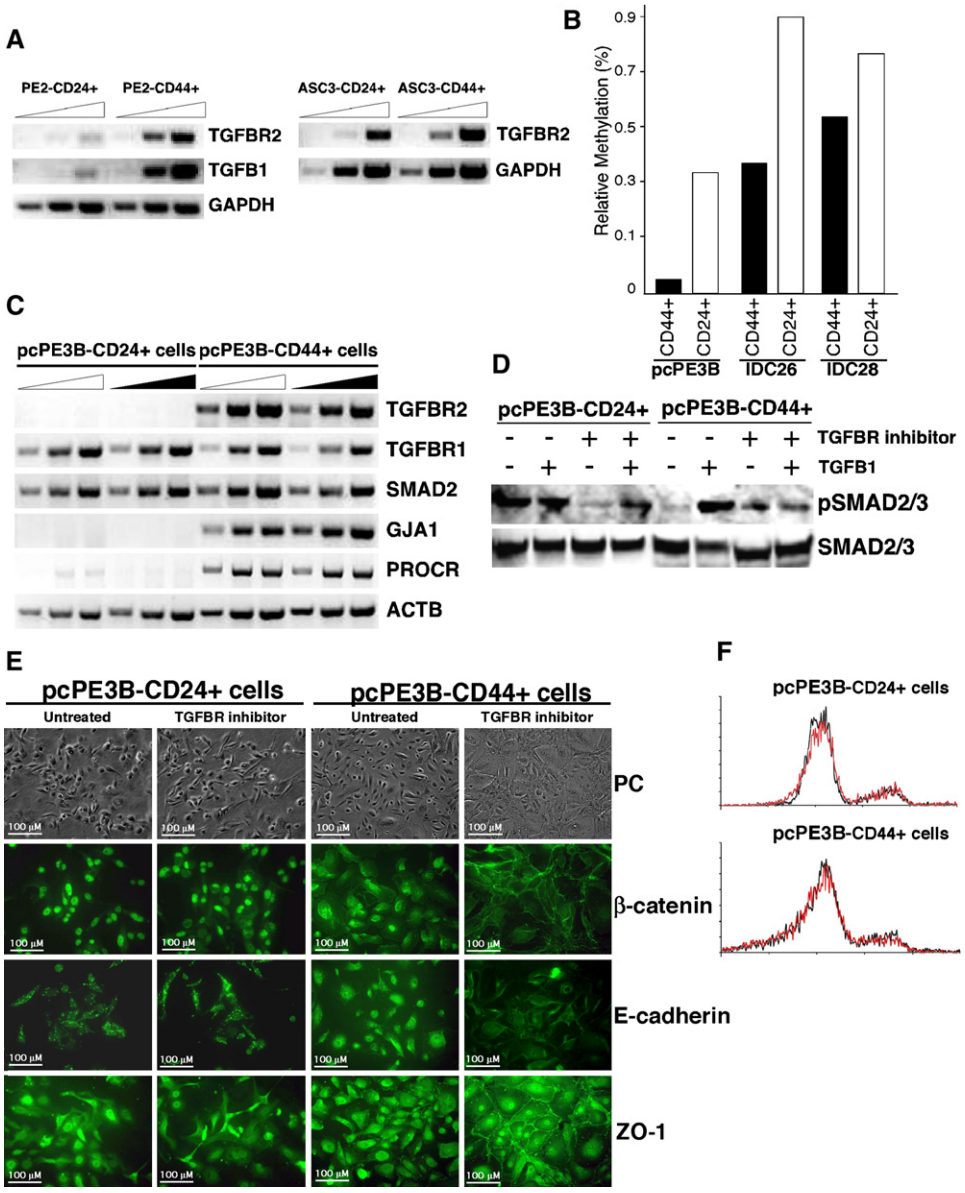

Figure 4. TGF-β Pathway in Tumor CD44+ and CD24+ Cells

(A) Semiquantitative RT-PCR analysis of the expression of genes involved in TGF-β signaling in CD24+ and CD44+ cells isolated from pleural effusion (PE2) and ascites (ASC3). TGFB1 and TGFBR2 are overexpressed in CD44+ cells. GAPDH was used as loading control. Each triangle indicates an increasing number of PCR cycles (25, 30, 35).

(B) Quantitative methylation-specific PCR (qMSP) analysis of TGFBR2 promoter in CD24+ (white bars) and CD44+ (black bars) cells from primary cultured pleural effusion (pcPE) or freshly isolated from the indicated invasive ductal carcinoma (IDC) tumor samples.

(C) Semiquantitative RT-PCR analysis of the expression of genes involved in TGF-β signaling and stem-cell function in primary cultured pleural effusion (pcPE) CD24+ and CD44+ cells. TGFBR2, GJA1, and PROCR are only detected in CD44+ cells, while the expression of TGFBR1 and SMAD2 are the same in the two cell populations. ACTB was used as loading control. Each triangle indicates an increasing number of PCR cycles (25, 30, 35), with color indicating cells cultured in the absence (white) or presence (black) of TGFBR inhibitor.

(D) Immunoblot analysis of phospho-SMAD2/3 and SMAD2/3 protein levels in primary cultured pleural effusion (pcPE) CD24+ and CD44+ cells in the presence or absence of TGF-β and the TGFBR inhibitor.

(E) Phase-contrast (PC) images and immunofluorescence analysis of β-catenin, E-cadherin, and ZO-1 expression and cellular localization before and after treatment with TGFBR inhibitor for primary cultured pleural effusion (pcPE) CD44+ and CD24+ cells. Following treatment with the inhibitor, CD44+ cells underwent dramatic morphologic changes and redistribution of β-catenin, E-cadherin, and ZO-1 to the cell membrane, while CD24+ cells demonstrated no response.

(F) Cell-cycle profiles of primary cultured pleural effusion (pcPE) CD24+ and CD44+ cells before (black line) and after (red line) treatment with TGFBR inhibitor. The inhibitor had no effect on the proliferation of either cell type.

(LY2109761) on the growth and differentiation of CD44+ and CD24+ cells cultured in vitro. First, we confirmed that the expression of TGF-β pathway components and cell-type-specific genes are the same in primary cultures as in fresh tumor samples and are not affected by treatment with the TGFBR inhibitor (Figure 4C). We also confirmed that TGF-β signaling is only activated in CD44+ cells in response to TGF-β treatment and that this is inhibited by the TGFBR inhibitor as determined by the phosphorylation of SMAD2/3 (Figure 4D). CD24+ cells exhibited high basal phopho-SMAD2/3 levels that could potentially be due to activin signaling or a mutation in TGFBR1. Next, we analyzed if treatment with the TGFBR inhibitor affects the phenotypes or proliferation of CD24+ and CD44+ cells. This treatment resulted in dramatic morphologic changes in CD44+ cells within 24 hr (Figure 4E). While untreated CD44+ cells were round-shaped and dispersed, CD44+ cells treated with the TGFBR inhibitor were more epithelial in appearance (Figure 4E). To determine if this morphologic change is the result of epithelial differentiation induced by the inhibitor, we analyzed β-catenin, E-cadherin, and ZO-1 by immunofluorescence (Figure 4E). TGFBR inhibitor treatment led to the localization of all three proteins to the cell membrane, consistent with the induction of an epithelial cell phenotype. These results prove that the TGF-β pathway is specifically activated in CD44+ breast cancer cells and that it regulates, at least in part, their more mesenchymal appearance. Consistent with the lack of expression of TGFBR2 in CD24+ cells (Figure 4C), we did not detect any changes in these cells following TGFBR inhibitor treatment. Interestingly, β-catenin and E-cadherin were abnormally localized in CD24+ cancer cells, suggesting that, although these cells lack stem-cell markers, they are not normal differentiated luminal epithelial cells. The TGFBR inhibitor had no significant effect on the proliferation and survival of CD24+ or CD44+ cells under the conditions tested (Figure 4F).

Prognostic Value of Gene Signatures Specific for Either CD44+ and PROCR+ or CD24+ Cancer Cells

To determine the clinical significance of the gene expression profiles of CD44+, PROCR+, and CD24+ breast cancer cells, we investigated whether expression in tumors of subsets of genes upregulated or downregulated in cancer CD44+ and PROCR+ cell SAGE libraries compared to cancer CD24+ cell SAGE libraries are correlated with breast cancer patient clinical outcome. Two groups of genes were identified that are statistically significantly associated with distant metastasis-free survival in three independent published data sets of patients with lymph-node-negative tumors who did not receive chemotherapy or hormonal therapy (Chang et al., 2005; Sotiriou et al., 2006; van de Vijver et al., 2002; Wang et al., 2005). Two data sets were used as training sets to select genes likely to be correlated with outcome, while the third data set was used only as a test set. Signatures "A" and "B" consist of genes upregulated and downregulated, respectively, in cancer CD44+ and PROCR+ cells compared to cancer CD24+ cells (Table 2). High expression of signature "A" genes is associated with shorter distant metastasis-free survival times, while high expression of signature "B" genes is associated with longer distant metastasis-free survival times (Figures 5A and 5B). In the third data set, signatures "A" and "B" are also associated with shorter and longer relapse-free and overall survival times, respectively (Supplemental Data). Neither signature is consistently statistically significantly associated with tumor estrogen receptor (ER) status or histologic grade, and each signature does correlate with survival time among patients with ER+, ER−, high-grade, and low-grade tumors (Supplemental Data); therefore, the correlation of signatures "A" and "B" with outcome is independent of ER expression and tumor grade.

Since our pathway analysis and TGFBR inhibitor treatment experiment implied that activation of the TGF-β pathway in CD44+ breast cancer cells might play an important role in their invasive phenotype, we also tested whether expression in tumors of a "TGF-β cassette" correlates with clinical outcome in the three data sets used above. The "TGF-β cassette" includes 15 genes more highly expressed in cancer CD44+ and PROCR+ cells than in cancer CD24+ cells (Table 1III). High expression of these genes is statistically significantly associated with shorter distant metastasis-free survival time in one data set (Figure 5C), suggesting that activation of the TGF-β pathway in CD44+ cancer cells may be relevant for disease progression in a subset of breast cancer patients and that these patients may benefit from therapy targeting this pathway.

Localization of CD24+ and CD44+ Cells in Normal and Tumor Tissue

Since the isolation of the CD24+ and CD44+ cells includes multiple steps, the possibility that the procedure altered the expression of some genes cannot be excluded. To verify the SAGE data by methods using intact tissue and to determine the number and location of CD24+ and CD44+ cells in normal breast tissue and breast carcinomas, we performed immunohistochemical analyses for selected genes specific for these two cell populations based on SAGE (Figure 6A). In normal breast tissue, cells positive for connexin 43 (Cx43), PROCR, and smooth muscle actin (SMA) were localized to the basal/myoepithelial layer. Virtually all epithelial (luminal and basal) cells were positive for CD44, while weak CD24 staining was detected in some luminal epithelial cells. Interestingly, in breast tissue of pregnant women, cells positive for Cx43, CD44, and PROCR were only detected in some ducts, suggesting a decrease in the expression of these genes or in the number of cells expressing them.

Next, we analyzed breast tumors of different stages, including 20 DCIS (ductal carcinoma in situ) and 250 primary invasive tumors with different lymph-node and distant-metastasis status. We observed varying expression of CD24, CD44, Cx43, CK19, CK17, FN1, and SMA and varying distribution of cells positive for them (Figure 6B and data not shown). Some tumors contained only CD44+ or

Table 2. Gene Signatures Specific for Either CD44+ and PROCR+ or CD24+ Cancer Cells with Prognostic Value

Signature	Tag	N1-CD44	PE2-CD44	ASC3-CD44	IDC28-PROCR	IDC29-PROCR	N1-CD24	N2-CD24	PE2-CD24	IDC26-CD24	IDC28-CD24	Gene Symbol	Gene Description
A	ATACTTTAATCAGAAGC	64	46	91	61	141	24	0	8	45	22	*ANXA5*	Annexin A5
	CTAGCCTCACGAAACTG	683	888	2298	278	415	418	1322	1039	264	422	*ACTG1*	Actin, gamma 1
	GTGCCAGCCCTCCTGGG	11	30	53	15	19	10	19	12	0	9	*ARF3*	ADP-ribosylation factor 3
	TAGGATGGGGGTCTTGT	36	15	0	65	81	14	24	4	40	28	*ATP1B3*	ATPase, Na+/K+ transporting, beta 3 polypeptide
	TCCCGAGGTCCACCTCC	68	61	87	34	26	5	58	52	4	22	*BAT3*	HLA-B associated transcript 3
	TTCTGTGAATCTGCCAT	11	4	8	30	41	38	0	0	0	0	*CALD1*	Caldesmon 1
	GGGTTCCCCGGCAGGGG	14	46	15	8	11	0	0	16	0	9	*CENTD2*	Centaurin, delta 2
	GTACTGTGGCTCAAGGG	164	80	311	61	78	101	145	44	4	75	*CLIC1*	Chloride intracellular channel 1
	GTACTGTGGCTCAAGGG	164	80	311	61	78	101	145	44	4	75	*CLIC1*	Chloride intracellular channel 1
	AATTTCTATTTCACAAG	18	23	23	0	30	14	5	0	18	9	*CTBS*	Chitobiase, di-N-acetyl-
	GGCTGCCCTGGGCAGCC	0	30	72	57	19	0	0	0	0	0	*DPYSL3*	Dihydropyrimidinase-like 3
	GGATTTGGAGTTAGGTG	4	15	38	19	7	14	24	4	0	13	*DVL3*	Dishevelled, dsh homolog 3
	CCGCTGATCCACTCTCA	7	0	30	15	11	0	5	8	0	0	*EXT1*	Exostoses 1
	GCTCAAACTACCACTCC	18	4	4	30	19	0	0	8	0	3	*FGFR1*	Fibroblast growth factor receptor 1
	CCCTGGGTTCTGCCCGC	228	1654	235	514	259	29	58	179	27	191	*FTL*	Ferritin, light polypeptide
	TTATGGGATCTCAACGA	167	57	102	141	74	144	271	48	31	44	*GNB2L1*	Guanine nucleotide binding protein, beta polypeptide 2-like 1
	CTGGGGCTGATGTGGGC	11	8	23	30	15	0	29	4	0	3	*GPRC5A*	G protein-coupled receptor, family C, group 5, member A
	GACGTGTGGGCGCGACT	156	53	133	57	48	19	121	48	40	25	*H2AFZ*	H2A histone family, member Z
	TAGATTTCAATAATTGA	50	30	53	38	15	19	19	12	0	13	*HIF1A*	Hypoxia-inducible factor 1, alpha subunit
	CAGTCTGGGAGTGGGGA	14	19	42	11	7	0	5	20	4	0	*IL13RA1*	Interleukin 13 receptor, alpha 1
	ATACCTTCCGAGTGGAG	11	15	46	11	11	5	68	0	0	0	*KDELR2*	KDEL endoplasmic reticulum protein retention receptor 2
	CCAACATAACCTCTGGG	0	8	15	11	0	0	10	4	0	0	*LARP1*	La ribonucleoprotein domain family, member 1
	CAGGCTGCCTCCGTTTT	0	15	8	11	4	5	0	0	0	0	*LPIN2*	Lipin 2
	CCCACGGTTAGTGCCAC	11	8	4	23	30	14	10	4	4	3	*MARS*	Methionine-tRNA synthetase
	TGCAATAGGTGAGAGAA	7	0	0	38	30	0	0	0	0	0	*MMP10*	Matrix metallopeptidase 10
	GTACCGGGGACTTGGGA	21	19	19	23	30	0	5	0	0	0	*MMP14*	Matrix metallopeptidase 14
	GATCCCAACTGCTCCTG	245	141	417	247	448	34	48	36	4	9	*MT2A*	Metallothionein 2A
	CAAAACTGTTTGTTGGC	7	4	0	38	30	24	10	0	9	9	*MYO10*	Myosin X
	ACCTGGCAAAGGAAGAA	4	19	0	8	4	0	0	0	0	0	*NUP62*	Nucleoporin 62kDa
	AACTTTTGGCGTCTACT	46	8	0	15	19	5	19	4	0	0	*PFKFB3*	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3
	TGTTAGAAAATTATAAA	7	11	8	4	19	5	10	4	0	3	*PLOD2*	Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2
	TGGCTAGTGTTTGCCGA	28	23	49	38	37	43	5	16	9	25	*PSMB7*	Proteasome subunit, beta type, 7
	GCCCCAGGTAGGGGGAC	4	15	27	8	33	5	5	4	18	3	*PSMD8*	Proteasome 26S subunit, non-ATPase, 8
	TGTGAAAATAATTTTTG	4	46	4	23	0	5	0	12	4	3	*RIN2*	Ras and Rab interactor 2
	CTAAAGTGTCAGTTGGT	4	0	0	30	4	0	5	4	0	0	*RYBP*	RING1 and YY1 binding protein
	AACGAGTACAACGCGCT	28	65	64	34	15	0	10	24	0	9	*SDF4*	Stromal cell derived factor 4
	TTGCCTGGGATGCTGGT	4	4	4	4	26	14	10	0	0	3	*SETD5*	SET domain containing 5
	AATAGAAATTTATGTAG	11	1223	4	15	0	5	0	52	0	0	*SPP1*	Secreted phosphoprotein 1
	GTGAAACGCTGTCTCTA	0	8	0	11	22	10	0	0	0	3	*STAU1*	Staufen, RNA binding protein, homolog 1
	AACGACCTGGTGTCCGA	7	8	11	11	4	0	0	0	0	3	*TUBB3*	Tubulin, beta 3
	CTGGCGAGCGCGATAAG	100	50	38	38	4	10	24	12	18	25	*UBE2S*	Ubiquitin-conjugating enzyme E2S
	AGACTATATTTATGATC	0	8	11	4	4	10	0	0	0	0	*XPNPEP1*	X-prolyl aminopeptidase 1, soluble
B	TGGTAACTGGCTGCTGA	7	0	4	4	4	0	0	0	31	6	*ABHD5*	Abhydrolase domain containing 5
	GTGGATTCATTTATACC	0	0	8	4	11	14	0	12	22	13	*ADNP*	Activity-dependent neuroprotector
	GGGTGCTTGGTTGTTTC	7	19	15	19	11	0	10	60	13	47	*ATP6AP1*	ATPase, H+ transporting, lysosomal accessory protein 1
	GCACAGAGCAAGGCGGG	7	11	0	0	7	14	19	4	58	31	*BATF*	Basic leucine zipper transcription factor, ATF-like
	TATATTGATTGTGGCAA	57	23	0	23	30	24	5	48	36	28	*BTG1*	B-cell translocation gene 1, anti-proliferative
	CTTAATCTTGTCTCTCT	18	15	8	8	11	5	10	44	63	3	*BTG2*	BTG family, member 2
	GAATAAAATAGCCAGGG	0	8	8	0	7	5	5	16	27	19	*CC2D1A*	Coiled-coil and C2 domain containing 1A
	ATTCCAATCTTGTGTGT	4	30	34	4	19	0	10	16	58	44	*CLTC*	Clathrin, heavy polypeptide
	GAACGGGCCCGTTTGTC	4	0	0	0	0	5	0	4	9	6	*DIMT1L*	DIM1 dimethyladenosine transferase 1-like
	GCATCTGTTTACATTTA	14	4	0	4	7	10	19	8	22	9	*ELOVL5*	ELOVL family member 5, elongation of long chain fatty acids
	TGTGAACACATAGGACG	14	8	0	11	0	38	19	8	13	34	*IRF1*	Interferon regulatory factor 1
	CATCCCGTGACTGCAAT	4	0	0	0	0	19	10	8	9	3	*LTA4H*	Leukotriene A4 hydrolase
	TTTGCACTACTTGGGGG	4	0	0	0	0	5	0	8	4	6	*PACAP*	Proapoptotic caspase adaptor protein
	GAATGTTTTTTCCTGAT	11	4	0	0	4	19	0	8	45	19	*PLXNA3*	Plexin A3
	CCAATGCTATGTCCACC	0	0	0	0	4	0	0	8	0	16	*RMND5B*	Required for meiotic nuclear division 5 homolog B
	ACATTGTTTTCCTTTTT	0	0	0	4	0	10	5	0	27	3	*SNX5*	Sorting nexin 5
	CTTTCAATGTTGGCTTC	0	11	4	8	15	19	15	8	81	25	*TBC1D8*	TBC1 domain family, member 8

Matching SAGE tag sequences, counts per 200,000 tags in each indicated SAGE library, and gene symbols and descriptions are listed for genes in signatures "A" and "B." These genes are statistically significantly up- and downregulated, respectively, in a pool of all cancer CD44+ and PROCR+ cell SAGE libraries compared to a pool of all cancer CD24+ cell SAGE libraries. N, normal; PE, pleural effusion; ASC, ascites; IDC, invasive ductal carcinoma.

only CD24+ cells, while others had a mix of the two cell types or lacked both of them. In some DCIS, we observed distinct subpopulations of cells mutually exclusively expressing CD44 and CD24. The expression of CD24 and CD44 was not statistically significantly correlated with any of the histopathologic characteristics of the tumors.

To determine if the number of CD44+ or CD24+ cells is different between primary tumors and distant metastases, we analyzed matched samples obtained from multiple independent patients. The frequency of CD24+ cells was dramatically higher in distant metastases regardless of their sites (all from solid organs such as liver, lung, bone, and adrenal gland) compared to primary tumors in all eight patients analyzed, while the numbers of cells positive for Cx43, CD44, SMA, CK19, and CK17 did not show consistent differences (Figure 6C and data not shown).

DISCUSSION

The hypothesis that tumors contain a subpopulation of cells with stem-cell characteristics has generated renewed interest and excitement in part due to the assumption that ineffective targeting of this cell population is responsible for therapeutic failures and recurrences (Dean et al., 2005; Eckfeldt et al., 2005). Unfortunately, there is very limited data available on the molecular identity and clinical relevance of breast "cancer stem cells." Analysis of the number of putative breast cancer stem cells

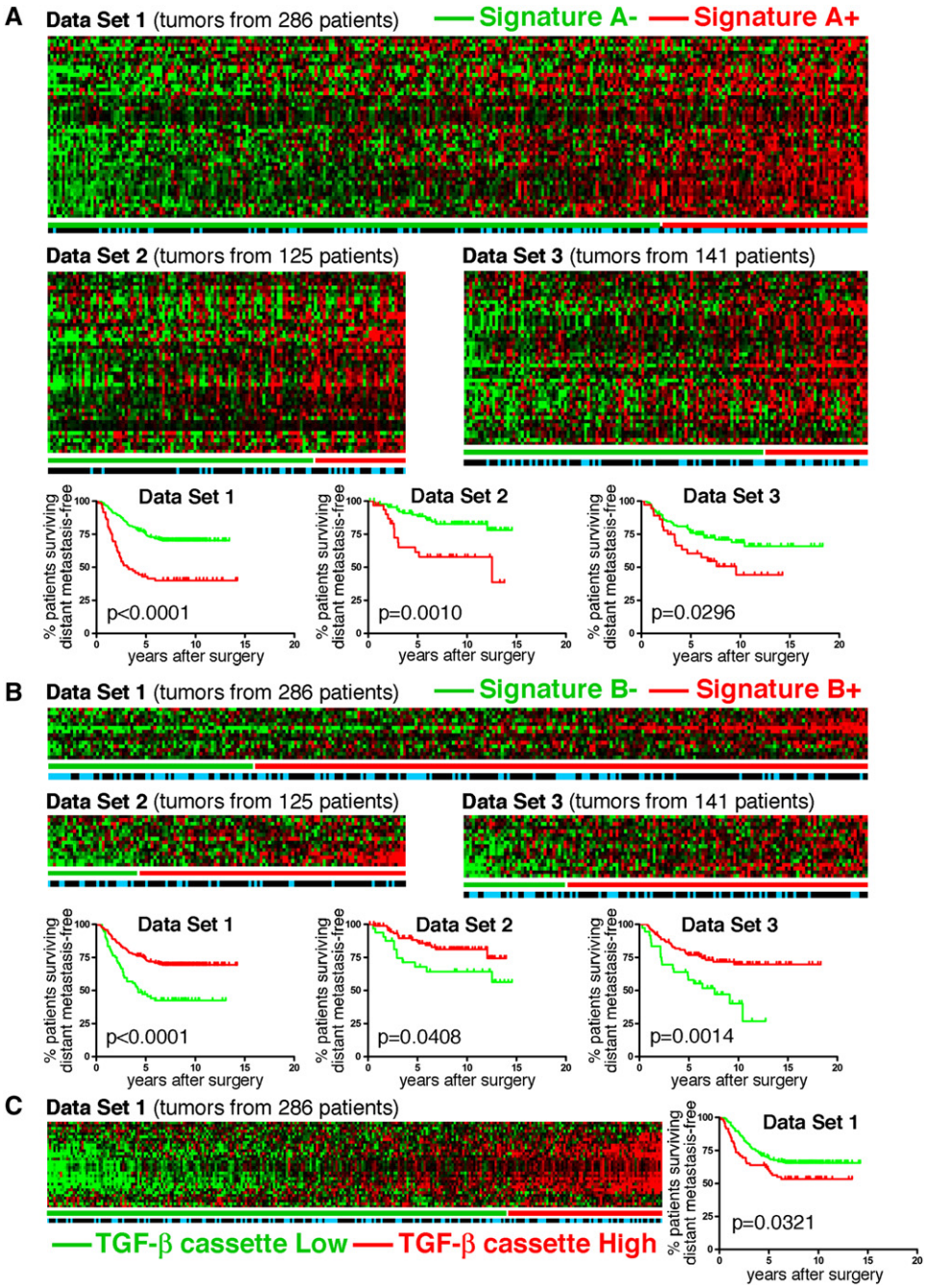

Figure 5. Clinical Relevance of Gene Expression Differences between CD24+, CD44+, and PROCR+ Breast Cancer Cells

(A) Tumor gene expression levels and distant metastasis-free survival for patients with and without signature "A," which consists of genes upregulated in pooled cancer CD44+ and PROCR+ cell SAGE libraries versus pooled cancer CD24+ cell SAGE libraries. In each heat map, rows represent microarray probes corresponding to the genes in signature "A," and columns represent tumors from data sets 1 (Wang et al., 2005), 2 (Sotiriou et al., 2006), and 3 (Chang et al., 2005; van de Vijver et al., 2002) ordered by average expression value for the signature. Patients with tumors with average expression values at or above the 75th percentile were called "signature A+"; all others were called "signature A−." Bottom bars indicate presence (blue) or absence (black) of distant metastases in the patients from which the tumors were obtained. Kaplan-Meier curves and log-rank test p values show that "signature A+" patients in all three data sets have statistically significantly ($p < 0.05$) shorter distant metastasis-free survival times than "signature A−" patients. (B) Tumor gene expression levels and distant metastasis-free survival for patients with and without signature "B," which consists of genes upregulated in pooled cancer CD24+ cell SAGE libraries versus pooled cancer CD44+ and PROCR+ cell SAGE libraries. Rows and columns of heat maps are as described above, except that probes correspond to the genes in signature "B." Patients with tumors with average expression values at or below the 25th percentile were called "signature B−"; all others were called "signature B+." Bottom bars are as described above. Kaplan-Meier curves and log-rank test p values show that "signature B−" patients in all three data sets have statistically significantly ($p < 0.05$) shorter distant metastasis-free survival times than "signature B+" patients.

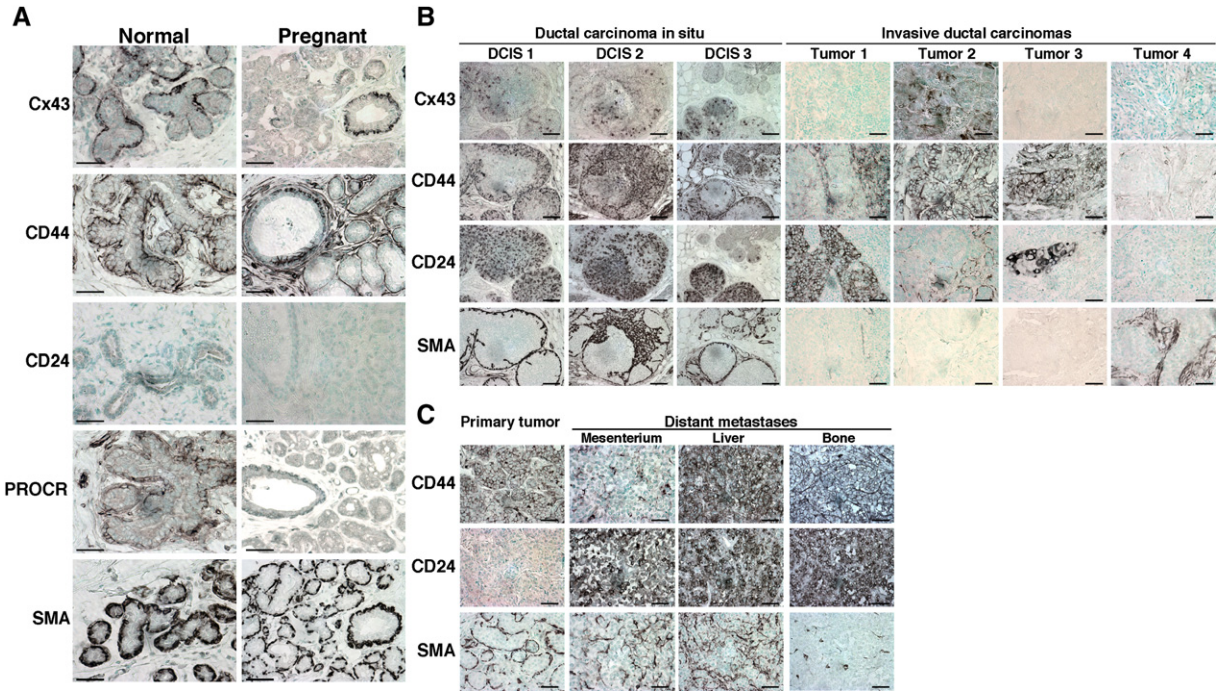

Figure 6. Immunohistochemical Analyses of Normal Breast Tissue and Breast Carcinomas

(A) Examples of immunohistochemical analyses of normal and pregnant breast tissue using the indicated genes. All CD44+ cell markers (Cx43, CD44, PROCR, and SMA) are positive in basal cells, and SMA is also positive in terminally differentiated myoepithelial cells. In pregnant breast tissue, cells positive for Cx43, CD44, and PROCR are only localized in terminal ducts. Scale bars, 100 μM.

(B) Representative examples of immunohistochemical staining patterns observed in DCIS (ductal carcinoma in situ) and invasive ductal breast carcinomas using antibodies against the indicated genes. Cx43, CD44, CD24, and SMA have varying expression and distribution in epithelial cells. SMA is also positive in DCIS myoepithelial cells and myofibroblasts. Scale bars, 200 μM.

(C) Immunohistochemical analyses of a primary breast tumor and matched distant metastases from the indicated organs for the expression of CD44, CD24, and SMA. The primary tumor was CD24 negative, but all distant metastases were strongly CD24 positive. Scale bars, 200 μM.

identified by immunohistochemistry as $CD44^+/CD24^-$ cells failed to identify statistically significant association between the frequency of these cells and clinical behavior, although tumors that developed distant metastases (mainly in the bone) had a higher fraction of $CD44^+/CD24^-$ cells (Abraham et al., 2005). Investigation of the prognostic significance of an 11-gene, BMI1-driven stem cell gene signature in a large collection of gene expression data from several different tumor types determined that this signature is a powerful predictor of short disease-free and overall survival and the risk of distant metastases (Glinsky et al., 2005).

To better understand the molecular differences between CD24+ and CD44+ cells in breast carcinomas as well as analogous cells from normal breast tissue, we determined their comprehensive gene expression and genetic profiles. Using this approach, we made several important conclusions. First, the gene expression profile of CD44+ cells resembles that of stem cells, and normal and tumor CD44+ cells are more similar to each other than to CD24+ cells from the same tissue. Second, tumor CD44+ and CD24+ cells are clonally related but not always identical since in some tumors CD24+ cells have additional genetic alterations besides the ones shared with CD44+ cells. Third, the distinct gene expression profiles of the cells reflect the activation of distinct signaling pathways, and some of these are specific for breast cancer CD44+ cells, suggesting approaches for their therapeutic targeting. Interestingly, genes involved in cell motility and angiogenesis were highly expressed in CD44+ cells, while genes involved in RNA splicing and carbohydrate metabolism were highly expressed in CD24+ cells. This is consistent with CD44+ cells demonstrating a more mesenchymal, motile, and less proliferative stem-cell-like profile. Fourth, CD44+ breast cancer cells are negative for ER even in some ER+ tumors (where ER is expressed

(C) Tumor gene expression levels and distant metastasis-free survival for patients with high and low expression of a CD44+ cell-specific "TGF-β cassette," which consists of the genes in Table 1III. Rows of the heat map represent microarray probes corresponding to genes in the "TGF-β cassette," and columns represent tumors from data set 1 ordered by average expression value for the cassette. Patients with tumors with average expression values at or above the 75th percentile were called "TGF-β cassette high"; all others were called "TGF-β cassette low." Bottom bars are as described above. Kaplan-Meier curves and a log-rank test p value show that "TGF-β cassette high" patients have statistically significantly ($p < 0.05$) shorter distant metastasis-free survival times than "TGF-β cassette low" patients.

in CD24+ cells). Fifth, breast cancer CD44+ and CD24+ cell gene expression signatures correlate with clinical outcome. Since these signatures were generated using expression data for bulk tumors, this result could mean that the numbers of cancer CD44+ and CD24+ cells within tumors are associated with clinical outcome. Specifically, tumors composed of mostly CD44+ cells may have worse clinical behavior than tumors mainly composed of CD24+ cells. Many genes in signature "A" (characteristic of breast cancer CD44+ cells) are involved in cell motility, invasion, apoptosis, and ECM remodeling, while signature "B" (characteristic of breast cancer CD24+ cells) contains several genes involved in inflammation and immune function.

One of the pathways we found specifically activated in CD44+ breast cancer cells is the TGF-β signaling pathway, known to play an important role in human embryonic stem cells as well as in tumorigenesis (James et al., 2005; Moses and Serra, 1996; Muraoka-Cook et al., 2005; Roberts and Wakefield, 2003; Siegel and Massague, 2003). TGF-β plays a dual role in tumor progression: it is one of the most potent inhibitors of cell proliferation, but it promotes invasion, angiogenesis, epithelial-mesenchymal transition (EMT), and metastasis (Bates and Mercurio, 2005; Roberts and Wakefield, 2003). Our results suggest that cells with different phenotypes, even within the same tumor and tissue type, respond differently to TGF-β activation. Intriguingly, we found that, in the tumors analyzed, the specific activation of TGF-β signaling in CD44+ breast cancer cells is due to the restricted expression of the TGFBR2 receptor in these cells and its epigenetic silencing in CD24+ cells. Correlating with this, treatment with a TGFBR kinase inhibitor specifically affected CD44+ tumor cells, leading to a mesenchymal-to-epithelial transition. Further emphasizing the importance of TGF-β signaling in CD44+ breast cancer cells, high expression of TGF-β pathway genes was associated with shorter distant metastasis-free survival in a set of breast cancer patients. These findings may have immediate therapeutic implications due to the current testing of TGF-β pathway inhibitors in clinical trials (Arteaga, 2006).

Our comprehensive gene expression profiling was limited to a few cases, so we analyzed the expression of selected genes by immunohistochemistry in a larger set of normal breast tissue and breast carcinomas of different stages. This demonstrated that cells expressing CD44+ cell-specific genes (CD44, Cx43, and PROCR) are localized to the basal cell layer of ducts and alveoli in normal breast tissue and that their number may dramatically decrease in late pregnancy. Pregnancy is thought to lead to the terminal differentiation of breast epithelial stem cells, and early full-term pregnancy reduces the risk of certain breast cancer (postmenopausal ER+ tumors), potentially by reducing the number of cells that could be targets of cellular transformation (Polyak, 2006; Russo et al., 2005; Schedin, 2006). Our results are consistent with this hypothesis, although proving it would require the analysis of normal breast tissue from women with differing parity and breast cancer history.

The immunohistochemical analysis of tumors revealed high heterogeneity for the expression of selected genes among samples. The expression of CD24 and CD44 (analyzed alone or in combination) in primary invasive breast carcinomas was not correlated with any tumor characteristics. The expression of SMA (in tumor cells) in primary tumors correlated with lymph node, distant metastasis, ER, and HER2 status, while Cx43 expression was associated with ER status. Surprisingly, we found that the number of CD24+ cells was dramatically and consistently increased in distant metastases irrespective of the type of the primary tumor and location of the distant metastasis. Although this observation seemingly contradicts the hypothesis that CD24+ cells represent more differentiated and less tumorigenic cancer cells, it is consistent with reports associating CD24 expression with tumor progression and metastatic behavior (Baumann et al., 2005; Bircan et al., 2006; Kristiansen et al., 2003). There are several possible explanations for this apparent paradox. The expression of one gene may not be sufficient to uniquely identify a cell with a particular phenotype. Tumorigenicity studies in experimental systems (e.g., injection of tumor cells into mammary fat pads of immunodeficient mice) may not reflect the behavior of the cells in patients. Tumor cells may change their phenotypes and gene expression profiles during the metastatic process. Finally, as also suggested by our FISH result demonstrating clonal genetic differences between CD24+ and CD44+ cells within a tumor, CD44+ and CD24+ cells may undergo independent clonal evolution.

In summary, our comprehensive molecular and phenotypic analysis of CD24+ and CD44+ cells from breast carcinomas revealed that they represent defined cell populations with distinct gene expression, epigenetic, and genetic profiles. Although CD44+ cells appear to express many stem-cell markers, the genetic difference between CD24+ and CD44+ cells within a tumor questions the validity of the cancer-stem-cell hypothesis in breast cancer and suggests clonal evolution involving intratumoral heterogeneity as an alternative explanation of our data and previously published data. Importantly, gene signatures associated with CD24+ and CD44+ tumor cells may have clinical relevance, and signaling pathways specifically activated in CD44+ cells could be used for their therapeutic targeting. Further studies, especially clinical trials, are necessary to validate our findings and to determine if targeting CD44+ tumor cells will have an impact on the clinical management of breast cancer patients.

EXPERIMENTAL PROCEDURES

Clinical Samples

Fresh, frozen, or formalin-fixed, paraffin-embedded tumor specimens were collected at Harvard-affiliated hospitals (Boston, MA) and Johns Hopkins University (Baltimore, MD). All human tissue was collected using protocols approved by the Institutional Review Boards; informed consent was obtained from each individual who provided tissues with linked clinical data. Fresh tissue samples were immediately processed for immunomagnetic purification as described in detail in the Supplemental Data. Tissue microarrays were purchased from Imgenex (CA),

obtained from Cooperative Breast Cancer Tissue Resource, or generated at Johns Hopkins University.

RT-PCR, SAGE, SNP Array, and qMSP Analysis

RT-PCR and SAGE analyses were performed essentially as previously described (Allinen et al., 2004). Clustering of SAGE libraries was performed using Cluster (Eisen et al., 1998) and the tags listed in Table S7, and results were visualized with MapleTree (developed by L. Simirenko). SNP array analysis was performed using Affymetrix 250K arrays, and data were analyzed using published protocols (LaFramboise et al., 2005). TGFBR2 promoter methylation analysis was performed as described previously (Hu et al., 2005) using published primers (Zhang et al., 2004; Zhao et al., 2005).

Fluorescence In Situ Hybridization

BAC probes RP11-157A11 (1q21.3), RP11-812I22 (17q21.2), and RP11-661N22 (6q21-q23.2) were labeled with digoxigenin (Roche, Indianapolis, IN) using an enzyme mix from the BioNick labeling kit (Invitrogen, Carlsbad, CA) as done previously (Zhao et al., 1995). RP11-167M14 (17q21.3-q22.1), CTD-2349A18 (8q24), RP11-606C3 (7p21.1), and RP11-697I2 (17q11-q12) were labeled with biotin using the same kit. CD24+, CD44+, and PROCR+ cultures were treated with colcemid, harvested, and used for metaphase chromosome spreads preparations according to standard protocols. Hybridization of metaphase chromosomes was performed as previously described (Ney et al., 1993). The probes were detected using reagents supplied by Cytocell Technologies, Ltd. according to the manufacturer's recommendation. Images were captured using the CytoVysion Imaging System (Applied Imaging, Pittsburgh, PA).

Functional Annotation, Network Analysis, and Correlation of Gene Expression with Outcome

Genes highly represented in the SAGE libraries were functionally classified by gene ontology biological processes using the DAVID Functional Annotation Tool (http://david.abcc.ncifcrf.gov/summary.jsp). For network analyses, sets of genes differentially expressed between CD44+ and CD24+ cells were uploaded into the MetaCore analytical suite version 2.0 (GeneGo, Inc., St. Joseph, MI), and analysis was conducted as described previously (Nikolsky et al., 2005a; Nikolsky et al., 2005b). Gene signatures that correlated with outcome were identified using Cox proportional hazards regression, hierarchical clustering, Kaplan-Meier analysis, and log-rank tests. TGF-β genes were tested for correlation with clinical outcome using Kaplan-Meier analysis and log-rank tests. Detailed descriptions of these methods are included in the Supplemental Data.

Primary Cell Culture, FACS, and Immunoblot Analyses

Purified cells were cultured using protocols described in detail in the Supplemental Data. FACS analysis was performed as described before (Polyak et al., 1994) using FITC- or PE-conjugated CD24, CD44, and PROCR antibodies from BD Biosciences (San Jose, CA). Immunocytochemistry and immunoblot analyses were performed using BCTN, phospho-SMAD2/3, and SMAD2/3 antibodies from Cell Signaling (Beverly, MA), CDH1 from BD Biosciences (San Jose, CA), and ZO1 from Chemicon (Temecula, CA) and protocols recommended by the provider. Cells were treated with the LY2109761 TGFBR inhibitor (Eli Lilly, IN) at 0.5 μM final concentration for 24 hr prior to analysis. Cell-cycle analysis was performed essentially as described (Polyak et al., 1994).

Immunohistochemistry

Immunohistochemistry was performed essentially as described earlier (Porter et al., 2003) using the following primary antibodies: CD24 and CD44 (LabVision, Fremont CA), Cx43 (Cell Signaling, Beverly, MA), PROCR (BD Biosciences, San Jose, CA), and SMA, CK17, and CK19 (DAKO, Carpinteria, CA). Antibody staining was scored by pathologist P.A. on a scale of 0–3 for intensity (0 = no staining, 1 = faint signal, 2 = moderate, and 3 = intense staining) and 0–3 for extent (0 = no, 1 < 30%, 2 = 30%–70%, and 3 > 70% positive cells). For statistical analyses, a cumulative score at or above 2 was considered positive.

Supplemental Data

The Supplemental Data include 15 supplemental figures, seven supplemental tables, and Supplemental Experimental Procedures and can be found with this article online at http://www.cancercell.org/cgi/content/full/11/3/259/DC1/.

ACKNOWLEDGMENTS

We greatly appreciate the help of Diana Calogrias and Dr. Myles Brown with the acquisition of human tissue samples and Natasha Pliss with immunohistochemical analyses. We thank Dr. Jonathan Yingling (Eli Lilly, Indianapolis, IN) for providing us with the TGFBR inhibitor; Drs. Ian Krop, Myles Brown, and Bert Vogelstein for their critical reading of the manuscript; and the Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, Canada for SAGE library sequencing. This work was supported by Novartis Pharmaceuticals, Inc., NIH (CA89393 and CA94074) and DOD (DAMD17-02-1-0692 and W8IXWH-04-1-0452) grants awarded to K.P., and a DOD (BCO30054) grant awarded to S.S. and P.A. K.P. receives research support from and is a consultant to Novartis Pharmaceuticals, Inc. K.P. also receives research support from Biogen-Idec and is a consultant to Aveo Pharmaceuticals, Inc.

Received: August 11, 2006
Revised: November 20, 2006
Accepted: January 16, 2007
Published: March 12, 2007

REFERENCES

Abraham, B.K., Fritz, P., McClellan, M., Hauptvogel, P., Athelogou, M., and Brauch, H. (2005). Prevalence of CD44+/CD24-/low cells in breast cancer may not be associated with clinical outcome but may favor distant metastasis. Clin. Cancer Res. *11*, 1154–1159.

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. USA *100*, 3983–3988.

Allinen, M., Beroukhim, R., Cai, L., Brennan, C., Lahti-Domenici, J., Huang, H., Porter, D., Hu, M., Chin, L., Richardson, A., et al. (2004). Molecular characterization of the tumor microenvironment in breast cancer. Cancer Cell *6*, 17–32.

Arteaga, C.L. (2006). Inhibition of TGFbeta signaling in cancer therapy. Curr. Opin. Genet. Dev. *16*, 30–37.

Asselin-Labat, M.L., Shackleton, M., Stingl, J., Vaillant, F., Forrest, N.C., Eaves, C.J., Visvader, J.E., and Lindeman, G.J. (2006). Steroid hormone receptor status of mouse mammary stem cells. J. Natl. Cancer Inst. *98*, 1011–1014.

Bates, R.C., and Mercurio, A.M. (2005). The epithelial-mesenchymal transition (EMT) and colorectal cancer progression. Cancer Biol. Ther. *4*, 365–370.

Baumann, P., Cremers, N., Kroese, F., Orend, G., Chiquet-Ehrismann, R., Uede, T., Yagita, H., and Sleeman, J.P. (2005). CD24 expression causes the acquisition of multiple cellular properties associated with tumor growth and metastasis. Cancer Res. *65*, 10783–10793.

Bircan, S., Kapucuoglu, N., Baspinar, S., Inan, G., and Candir, O. (2006). CD24 expression in ductal carcinoma in situ and invasive ductal carcinoma of breast: An immunohistochemistry-based pilot study. Pathol. Res. Pract. *202*, 569–576.

Blanpain, C., Lowry, W.E., Geoghegan, A., Polak, L., and Fuchs, E. (2004). Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. Cell *118*, 635–648.

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.

Chang, H.Y., Nuyten, D.S., Sneddon, J.B., Hastie, T., Tibshirani, R., Sorlie, T., Dai, H., He, Y.D., van't Veer, L.J., Bartelink, H., et al. (2005). Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc. Natl. Acad. Sci. USA *102*, 3738–3743.

Clarke, M.F., and Fuller, M. (2006). Stem cells and cancer: Two faces of eve. Cell *124*, 1111–1115.

Clarke, R.B., Spence, K., Anderson, E., Howell, A., Okano, H., and Potten, C.S. (2005). A putative human breast stem cell population is enriched for steroid receptor-positive cells. Dev. Biol. *277*, 443–456.

Clayton, H., Titley, I., and Vivanco, M. (2004). Growth and differentiation of progenitor/stem cells derived from the human mammary gland. Exp. Cell Res. *297*, 444–460.

Dean, M., Fojo, T., and Bates, S. (2005). Tumour stem cells and drug resistance. Nat. Rev. Cancer *5*, 275–284.

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.

Eckfeldt, C.E., Mendenhall, E.M., and Verfaillie, C.M. (2005). The molecular repertoire of the 'almighty' stem cell. Nat. Rev. Mol. Cell Biol. *6*, 726–737.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA *95*, 14863–14868.

Ekins, S., Andreyev, S., Ryabov, A., Kirillov, E., Rakhmatulin, E.A., Sorokina, S., Bugrim, A., and Nikolskaya, T. (2006). A combined approach to drug metabolism and toxicity assessment. Drug Metab. Dispos. *34*, 495–503.

Glinsky, G.V., Berezovska, O., and Glinskii, A.B. (2005). Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. J. Clin. Invest. *115*, 1503–1521.

Hu, M., Yao, J., Cai, L., Bachman, K.E., van den Brule, F., Velculescu, V., and Polyak, K. (2005). Distinct epigenetic changes in the stromal cells of breast cancers. Nat. Genet. *37*, 899–905.

Ivanova, N.B., Dimos, J.T., Schaniel, C., Hackney, J.A., Moore, K.A., and Lemischka, I.R. (2002). A stem cell molecular signature. Science *298*, 601–604.

James, D., Levine, A.J., Besser, D., and Hemmati-Brivanlou, A. (2005). TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development *132*, 1273–1282.

Jarvinen, A.K., Autio, R., Haapa-Paananen, S., Wolf, M., Saarela, M., Grenman, R., Leivo, I., Kallioniemi, O., Makitie, A.A., and Monni, O. (2006). Identification of target genes in laryngeal squamous cell carcinoma by high-resolution copy number and gene expression microarray analyses. Oncogene *25*, 6997–7008.

Kristiansen, G., Winzer, K.J., Mayordomo, E., Bellach, J., Schluns, K., Denkert, C., Dahl, E., Pilarsky, C., Altevogt, P., Guski, H., and Dietel, M. (2003). CD24 expression is a new prognostic marker in breast cancer. Clin. Cancer Res. *9*, 4906–4913.

LaFramboise, T., Weir, B.A., Zhao, X., Beroukhim, R., Li, C., Harrington, D., Sellers, W.R., and Meyerson, M. (2005). Allele-specific amplification in cancer revealed by SNP array analysis. PLoS Comput. Biol. *1*, e65. 10.1371/journal.pcbi.0010065.

Lee, T.L., Alba, D., Baxendale, V., Rennert, O.M., and Chan, W.Y. (2006). Application of transcriptional and biological network analyses in mouse germ-cell transcriptomes. Genomics *88*, 18–33.

Liu, S., Dontu, G., Mantle, I.D., Patel, S., Ahn, N.S., Jackson, K.W., Suri, P., and Wicha, M.S. (2006). Hedgehog signaling and Bmi-1 regulate self-renewal of normal and malignant human mammary stem cells. Cancer Res. *66*, 6063–6071.

Lynch, M.D., Cariati, M., and Purushotham, A.D. (2006). Breast cancer, stem cells and prospects for therapy. Breast Cancer Res. *8*, 211.

Moses, H.L., and Serra, R. (1996). Regulation of differentiation by TGF-beta. Curr. Opin. Genet. Dev. *6*, 581–586.

Muraoka-Cook, R.S., Dumont, N., and Arteaga, C.L. (2005). Dual role of transforming growth factor beta in mammary tumorigenesis and metastatic progression. Clin. Cancer Res. *11*, 937s–943s.

Ney, P.A., Andrews, N.C., Jane, S.M., Safer, B., Purucker, M.E., Weremowicz, S., Morton, C.C., Goff, S.C., Orkin, S.H., and Nienhuis, A.W. (1993). Purification of the human NF-E2 complex: cDNA cloning of the hematopoietic cell-specific subunit and evidence for an associated partner. Mol. Cell. Biol. *13*, 5604–5612.

Nikolsky, Y., Ekins, S., Nikolskaya, T., and Bugrim, A. (2005a). A novel method for generation of signature networks as biomarkers from complex high throughput data. Toxicol. Lett. *158*, 20–29.

Nikolsky, Y., Nikolskaya, T., and Bugrim, A. (2005b). Biological networks and analysis of experimental data in drug discovery. Drug Discov. Today *10*, 653–662.

Osada, H., Tatematsu, Y., Sugito, N., Horio, Y., and Takahashi, T. (2005). Histone modification in the TGFbetaRII gene promoter and its significance for responsiveness to HDAC inhibitor in lung cancer cell lines. Mol. Carcinog. *44*, 233–241.

Patrawala, L., Calhoun, T., Schneider-Broussard, R., Zhou, J., Claypool, K., and Tang, D.G. (2005). Side population is enriched in tumorigenic, stem-like cancer cells, whereas ABCG2+ and ABCG2− cancer cells are similarly tumorigenic. Cancer Res. *65*, 6207–6219.

Polyak, K. (2006). Pregnancy and breast cancer: The other side of the coin. Cancer Cell *9*, 151–153.

Polyak, K., and Hahn, W.C. (2006). Roots and stems: Stem cells in cancer. Nat. Med. *12*, 296–300.

Polyak, K., Lee, M.H., Erdjument-Bromage, H., Koff, A., Roberts, J.M., Tempst, P., and Massague, J. (1994). Cloning of p27Kip1, a cyclin-dependent kinase inhibitor and a potential mediator of extracellular antimitogenic signals. Cell *78*, 59–66.

Ponti, D., Costa, A., Zaffaroni, N., Pratesi, G., Petrangolini, G., Coradini, D., Pilotti, S., Pierotti, M.A., and Daidone, M.G. (2005). Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. Cancer Res. *65*, 5506–5511.

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. (2003). Molecular markers in ductal carcinoma in situ of the breast. Mol. Cancer Res. *1*, 362–375.

Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R.C., and Melton, D.A. (2002). "Stemness": Transcriptional profiling of embryonic and adult stem cells. Science *298*, 597–600.

Roberts, A.B., and Wakefield, L.M. (2003). The two faces of transforming growth factor beta in carcinogenesis. Proc. Natl. Acad. Sci. USA *100*, 8621–8623.

Russo, J., Moral, R., Balogh, G.A., Mailo, D., and Russo, I.H. (2005). The protective role of pregnancy in breast cancer. Breast Cancer Res. *7*, 131–142.

Schedin, P. (2006). Pregnancy-associated breast cancer and metastasis. Nat. Rev. Cancer *6*, 281–291.

Siegel, P.M., and Massague, J. (2003). Cytostatic and apoptotic actions of TGF-beta in homeostasis and cancer. Nat. Rev. Cancer *3*, 807–821.

Soreghan, B.A., Lu, B.W., Thomas, S.N., Duff, K., Rakhmatulin, E.A., Nikolskaya, T., Chen, T., and Yang, A.J. (2005). Using proteomics and network analysis to elucidate the consequences of synaptic protein oxidation in a PS1 + AbetaPP mouse model of Alzheimer's disease. J. Alzheimers Dis. *8*, 227–241.

Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B., et al. (2006). Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. J. Natl. Cancer Inst. *98*, 262–272.

Stingl, J., Eaves, C.J., Kuusk, U., and Emerman, J.T. (1998). Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast. Differentiation *63*, 201–213.

Stingl, J., Raouf, A., Emerman, J.T., and Eaves, C.J. (2005). Epithelial progenitors in the normal human mammary gland. J. Mammary Gland Biol. Neoplasia *10*, 49–59.

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.

Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J., et al. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet *365*, 671–679.

Weissman, I.L. (2005). Normal and neoplastic stem cells. Novartis Found. Symp. *265*, 35–50; discussion 50–34, 92–37.

Wicha, M.S., Liu, S., and Dontu, G. (2006). Cancer stem cells: An old idea—A paradigm shift. Cancer Res. *66*, 1883–1890.

Zhang, H.T., Chen, X.F., Wang, M.H., Wang, J.C., Qi, Q.Y., Zhang, R.M., Xu, W.Q., Fei, Q.Y., Wang, F., Cheng, Q.Q., et al. (2004). Defective expression of transforming growth factor beta receptor type II is associated with CpG methylated promoter in primary non-small cell lung cancer. Clin. Cancer Res. *10*, 2359–2367.

Zhao, Y., Bjorbaek, C., Weremowicz, S., Morton, C.C., and Moller, D.E. (1995). RSK3 encodes a novel pp90rsk isoform with a unique N-terminal sequence: Growth factor-stimulated kinase function and nuclear translocation. Mol. Cell. Biol. *15*, 4353–4363.

Zhao, H., Shiina, H., Greene, K.L., Li, L.C., Tanaka, Y., Kishi, H., Igawa, M., Kane, C.J., Carroll, P., and Dahiya, R. (2005). CpG methylation at promoter site -140 inactivates TGFbeta2 receptor gene in prostate cancer. Cancer *104*, 44–52.