Oncogene (2007) 26, 7859–7871

www.nature.com/onc

ORIGINAL ARTICLE

Molecular signatures of metaplastic carcinoma of the breast by large-scale transcriptional profiling: identification of genes potentially related to epithelial–mesenchymal transition

HC Lien[1], YH Hsiao[2,3], YS Lin[4], YT Yao[1], HF Juan[5,6], WH Kuo[7], Mien-Chie Hung[8,9], KJ Chang[7,11] and FJ Hsieh[5,6,10,11]

[1]Department of Pathology, College of Medicine, National Taiwan University, Taipei, Taiwan; [2]Department of Obstetrics and Gynecology, Changhua Christian Hospital, Changhua, Taiwan; [3]Institute of Medical Research, Chang Jung Christian University, Tainan, Taiwan; [4]Welgene Biotech. Co. Ltd, Taipei, Taiwan; [5]Center for Systems Biology and Bioinformatics, National Taiwan University, Taipei, Taiwan; [6]Department of Life Science, National Taiwan University, Taipei, Taiwan; [7]Department of Surgery, College of Medicine, National Taiwan University, Taipei, Taiwan; [8]Center for Molecular Medicine, China Medical University Hospital, Taichung, Taiwan; [9]Department of Molecular and Cellular Oncology, University of Texas, MD Anderson Cancer Center, TX, USA and [10]Department of Obstetrics and Gynecology, College of Medicine, National Taiwan University, Taipei, Taiwan

Metaplastic carcinoma of the breast (MCB) is a poorly understood subtype of breast cancer. It is generally characterized by the coexistence of ductal carcinomatous and transdifferentiated sarcomatous components, but the underlying molecular alterations, possibly related to epithelial–mesenchymal transition (EMT), remain elusive. We performed transcriptional profiling using half-a-genome oligonucleotide microarrays to elucidate genetic profiles of MCBs and their differences to those of ductal carcinoma of breasts (DCBs) using discarded specimens of four MCBs and 34 DCBs. Unsupervised clustering disclosed distinctive expression profiles between MCBs and DCBs. Supervised analysis identified gene signatures discriminating MCBs from DCBs and between MCB subclasses. Notably, many of the discriminator genes were associated with downregulation of epithelial phenotypes and with synthesis, remodeling and adhesion of extracellular matrix, with some of them have known or inferred roles related to EMT. Importantly, several of the discriminator genes were upregulated in a mutant Snail-transfected MCF7 cell known to exhibit features of EMT, thereby indicating a crucial role for EMT in the pathogenesis of MCBs. Finally, the identification of SPARC and vimentin as poor prognostic factors reinforced the role of EMT in cancer progression. These data advance our understanding of MCB and offer clues to the molecular alterations underlying EMT.

Oncogene (2007) **26,** 7859–7871; doi:10.1038/sj.onc.1210593; published online 2 July 2007

Correspondence: Dr KJ Chang, Department of Surgery, College of Medicine, National Taiwan University, 1-1 Jen-Ai Road, Taipei, Taiwan or Dr FJ Hsieh, Department of Obstetrics and Gynecology, National Taiwan University, College of Medicine, 1-1 Jen-Ai Road, Taipei 104, Taiwan.
E-mail: kingjen@ha.mc.ntu.edu.tw or fjhsieh@ha.mc.ntu.edu.tw
[11]These authors have contributed equally to this work.
Received 7 February 2007; revised 8 May 2007; accepted 15 May 2007; published online 2 July 2007

Keywords: oligonucleotide microarray; metaplastic carcinoma of breast; epithelial–mesenchymal transition

Introduction

Metaplastic carcinoma of the breast (MCB) is a subtype of breast cancer. Although relatively rare, its peculiar histomorphology, classically characterized by the coexistence of carcinomatous and sarcomatous components, has long been intriguing. Several names, including carcinosarcoma, spindle carcinoma and sarcomatoid carcinoma, have been given to this cancer for its heterogeneous histologic features (Thiery, 2002). Carcinomas with these peculiar features are also encountered in other organs, including the female genital tract, lungs, prostate and urinary bladder (Thiery, 2002). Although several lines of evidence highly suggested a monoclonal origin (Zhuang *et al*., 1997; Lien *et al*., 2004), with the sarcomatous component converted from a carcinomatous component through a metaplastic (transdifferentiated) process (Thiery, 2002), it is currently not known whether this morphological subtype of breast cancer is genetically distinct from the conventional ductal carcinoma of the breast (DCB). Moreover, given the bona fide carcinomatous nature of MCB, the genetic programing underlying the metaplastic process remains elusive.

In addition to its histopathological features, MCB also has clinical behavior distinct from those of DCB, including relatively large size at diagnosis, infrequent expression of hormone receptors, lower incidence of axillary nodal involvement than typical breast carcinoma of similar size, and high rate of extranodal metastasis (Carter *et al*., 2006; Pezzi *et al*., 2007). Although presenting more commonly as node-negative disease, the outcome in MCB is generally poor with a

high risk of recurrence (Barnes *et al*., 2005; Carter *et al*., 2006; Pezzi *et al*., 2007). Patients with MCB usually do not benefit from conventional breast cancer chemotherapy or hormone therapy (Al Sayed *et al*., 2006). Despite these distinct clinicopathologic features, the genetic basis for the recognition of MCB as a discrete subtype of breast cancer distinct from DCB is still lacking and the molecular alterations of MCB, critical for understanding the pathology of this cancer, have not yet been explored.

Epithelial–mesenchymal transition (EMT) is the process of disaggregating structured polarized epithelial units into single-motile fibroblastoid cells to enable cell movement and morphogenesis, which was originally discovered from studies of embryonic development (Zavadil and Bottinger, 2005). The process of EMT gained wide recognition as a potential mechanism for the progression of malignancy, and was attributed to the loosening of epithelial characteristics and the acquisition of migratory and highly matrix invasive phenotypes (Thiery, 2002). EMT is characterized by loss of proteins associated with the polarized epithelial phenotype (for example, E-cadherin and cytokeratin) and *de novo* synthesis of proteins associated with mesenchymal and migratory morphology of transitioning cells (for example, vimentin) (Zavadil and Bottinger, 2005). Classically, MCB is composed of carcinomatous and sarcomatous components. Because the sarcomatous component is morphologically characterized by the less cohesive and more fibroblastoid growth patterns reminiscent of EMT, these morphological features would make metaplastic carcinoma an ideal *in vivo* model to study the genetic programing involved in EMT.

The purpose of this study was to investigate the molecular characteristics of MCBs and compare them with those of DCBs using microarray analysis of samples from four MCBs and 34 stage-matched DCBs. A more thorough understanding of the molecular profiles of MCBs and their distinctions from DCBs will contribute to our understanding of the pathology of this rare yet unique breast cancer and shed light on the genetic programing of EMT.

Results

Profiles of gene expression distinguish MCBs and DCBs

Representative histomorphological and immunophenotypic differences between MCBs and DCBs are shown in Figure 1a. A total of 44 breast tissue samples (4 MCBs, 34 DCBs and 6 normal breast samples) representing 44 different patients were profiled by oligonucleotide microarray technology. We initially analysed the microarray data in an unsupervised manner to determine whether there were any innate differences in the gene profiles between MCBs and DCBs. Unsupervised hierarchical clustering was performed using 4525 probes representing 4334 genes whose expression ratios varied by at least fourfold from the overall median abundance in at least one sample. We found distinctive gene expression profiles for MCBs and DCBs: the four MCBs clustered into one group and were separate from the 34 DCBs, exemplifying the inherent differences of gene expression between these two subtypes of breast cancer (Figure 1b). Not unexpectedly, the six normal breast samples clustered together and were separate from the cancer samples.

Identification of genes differentially expressed in MCBs relative to DCBs

To identify the most discriminating profiles between MCB and DCB samples, we performed supervised hierarchical clustering, using statistical analysis of microarrays (SAM) analysis (Figure 2a). With a false discovery rate (FDR) of 1%, a total of 213 probes representing 208 genes were found, with 87 genes overexpressed and 121 genes underexpressed in MCBs. The expression ratios and SAM scores for these 208 genes are shown in detail in Supplementary Table 1. Of the 208 genes, 124 had known biological functions, including 61 genes that were overexpressed and 63 genes that were underexpressed in the MCBs. Most of the 126 genes were categorized into five biological processes according to Gene Ontology annotations (Ashburner *et al*., 2000): cell adhesion/motility, transcription, development, signal transduction, and metabolism (Supplementary Table 2). Notably, several of the 87 MCB-overexpressed genes were functionally related to extracellular matrix (ECM), including genes associated with ECM synthesis (*PCOLCE2*, *PLOD*, *PLOD2*, *CSPG4*, *HAS*, *VIM* and *COL27A1*), remodeling (*CTSZ*), adhesion/motility/migration (*HAPLN1*, *VIM*, *COL9A3*, *COL9A1*, *CSPG4*, *KINDLIN 1* and *FLNA*), and genes associated with skeletal development and/or chondroossification (*BMP2*, *CSPG4* and *TNFRSF11* (*osteoprotegrin*)). *BMP2*, which also has documented function related to EMT (Ma *et al*., 2005), was identified among genes upregulated in MCBs. In contrast, genes encoding proteins related to maintaining epithelial phenotype, including cytoskeleton (*KRT18* and *KRT19*), cell–cell adhesion molecules (*CDH1*, *CEACAM6* and *TACSTD1* (*epithelial adhesion molecule*)) and tight junction (*EPPK1* (*epiplakin 1*), *CLDN7*, *CRB3* and *F11R* (*junctional adhesion molecule 1*)), were downregulated in MCBs (Liu *et al*., 2000; Fujiwara *et al*., 2001; Lemmers *et al*., 2004). *ESR1* and *ERBB2* were also downregulated in MCBs, consistent with universal lack of ERα and HER-2/neu expression in MCBs as compared to DCBs (Carter *et al*., 2006).

Identification of genes differentially expressed in MCB subclass 2

Among the four MCBs, BT-18 and BT-19, classified as MCB subclass 1, were histologically composed predominantly of carcinomatous components and focally (<10% of the whole-tumor component) of sarcomatous components, while BT-20 and L-5, classified as MCB subclass 2, were composed predominantly of sarcomatous components, with frequent ECM formation and osteochondroid change, and focally (<10% of the

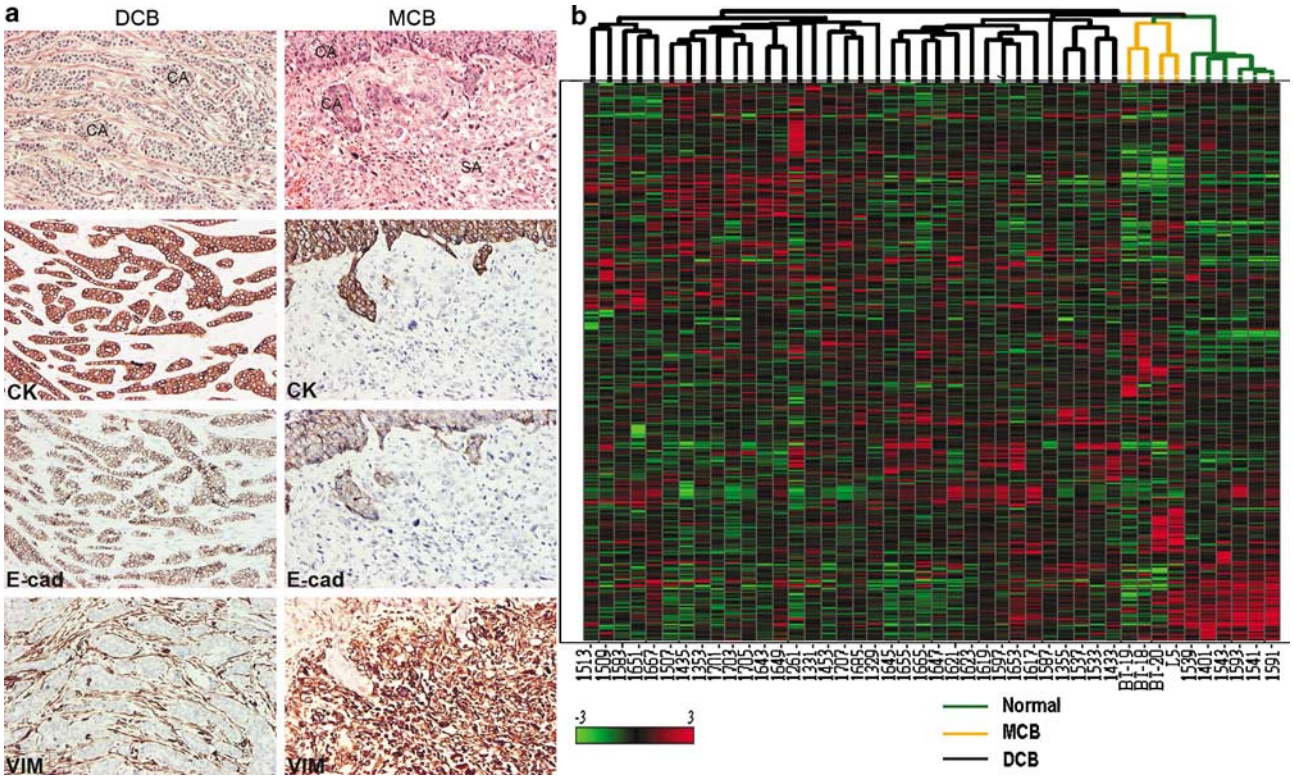

Figure 1 Histomorphology and gene expression profiling of metaplastic carcinoma of breasts (MCBs) and ductal carcinoma of breasts (DCBs). (**a**) Histology (stained with hematoxylin and eosin (H&E)) and immunophenotypes in a representative case of DCB (left) and MCB (right). DCB showed ductal arrangement of cancer cells that were strongly stained for cytokeratin (CK) and E-cadherin (E-cad), but negative for vimentin (VIM). MCB showed a mixture of carcinomatous (CA) and sarcomatous components (SA), with the latter displaying loss of epithelial phenotypes characterized by the loss of CK and E-cad, and the gain of mesenchymal phenotypes characterized by the fibroblastoid growth of cancer cells and the strong expression of VIM (original magnification, $\times 100$). (**b**) Unsupervised hierarchical clustering of 44 samples. Data are presented in a matrix format: each row represents a gene and each column represents a sample. For each sample, the $\log_2$ ratio of the abundance of each transcript to the abundance of each transcript in the reference sample was depicted according to a color scale (bottom left). The magnitude of deviation from the median was represented by the color saturation. The dendrogram of samples clearly revealed the distinctive gene expression profiles of MCBs and DCBs: the four MCBs clustered into one group and separated from the 34 DCBs. The six normal breast samples were clustered together and separated from cancer samples.

whole-tumor component) of carcinomatous components (Figure 3a). This morphological classification of the four MCBs into two subclasses was supported by the similar gene expression profiling between BT-20 and L-5 (MCB subclass 2) and between BT-18 and BT-19 (MCB subclass 1), as evidenced by correlation coefficients from scatter plotting among the four MCBs (Figure 3b), and by the clustering of BT-20 and L-5 (MCB subclass 2) and of BT-18 and BT-19 (MCB subclass 1) in the unsupervised hierarchical clustering analysis (Figure 1b). We hypothesized that the genetic difference between these two MCB subclasses might in someway account for the different morphological features seen between these two MCB subclasses, including the sarcomatous metaplasia and the prominent ECM formation with frequent osteochondroid change, which are features potentially related to EMT. Thus, we performed supervised hierarchical clustering using SAM analysis and disclosed 35 discriminator genes (FDR 5%). All of the 35 genes were overexpressed in MCB subclass 2 (Figure 3c). The expression ratios and SAM scores of the 35 genes are shown in Table 1. Twenty-seven of the 35 genes had known biological functions and most of the 27 genes were categorized into six biological processes according to Gene Ontology annotations: cell/cell matrix adhesion, development, signal transduction, metabolism, ion transport and transcription (Supplementary Table 3). Many of the genes were functionally related to ECM remodeling/synthesis (*ADAMTS5*, *HTRA3*, *MXRA8* and *TIPM3*), adhesion (*AGC1*, *EDIL3* and *PKD2*) and structural constituent/matricellular proteins (*COL16A1*, *COL18A1*, *LUM*, *P4H8*, *SPARC*, *THB1* and *THB2*). Several other genes had functions related to development (*HOXA7*, *MSX1*, *POSTN*, *AGC1*, *PRRX1*, *SFRP2*, *SFRP4* and *TBX2*), especially musculoskeletal/bone/cartilage development (*MSX1*, *POSTN*, *AGC1*, *PRRX1*, *SFRP2* and *SFRP4*). Genes involved in signal transduction (*PDGFA* and *PDGFRA*), ion transport (*KCNE4* and *PKD2*), cell division (*STAG2*) and muscle contraction (*PPP1R12A* and *TPM4*) were also among the genes identified by SAM analysis.

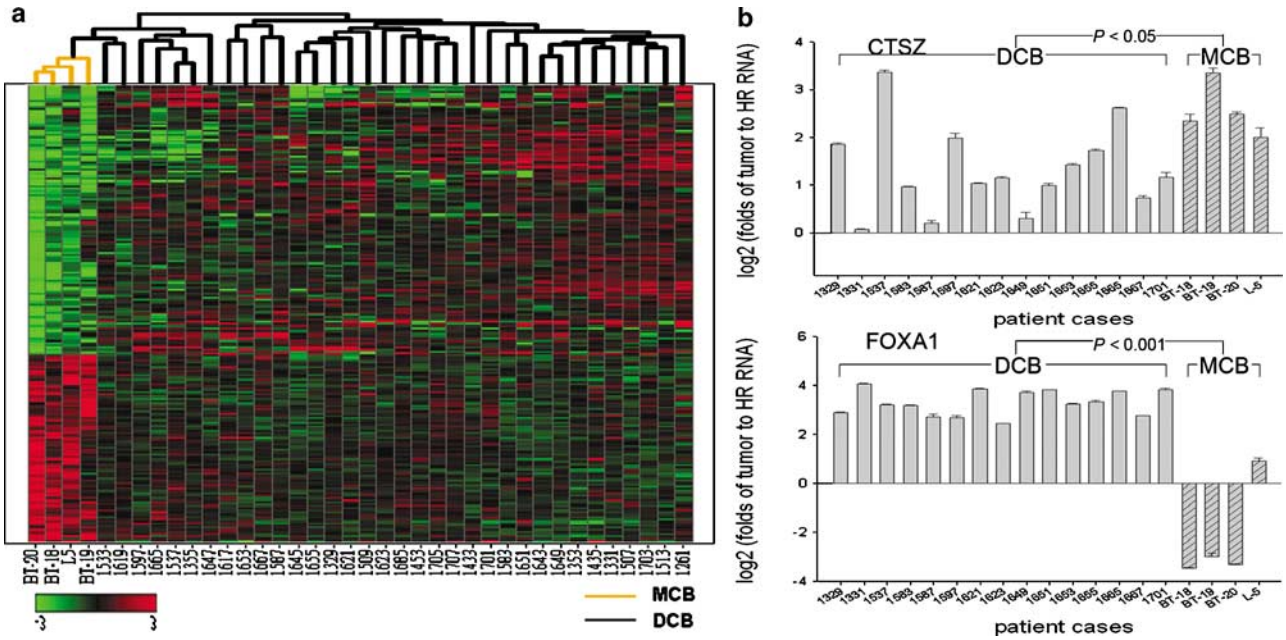

Figure 2 Identification of genes that distinguish MCBs and DCBs. (**a**) Supervised hierarchical clustering identified the 208 most discriminative genes, with 87 genes overexpressed and 121 genes underexpressed in MCBs. (**b**) Confirmation by quantitative reverse transcription–PCR (QRT–PCR) of an upregulated gene (*CTSZ*) and a downregulated gene (*FOXA1*) in tumor samples of MCBs and DCBs. Log_2-fold changes of tumor transcript level to reference transcript level and associated standard errors were plotted. Data are averages of triplicate QRT–PCR measurements. The patient case numbers of DCBs are as follows: 1329, 1331, 1583, 1651, 1655, 1665 and 1667 for stage II; and 1537, 1587, 1597, 1621, 1623, 1649, 1653 and 1701 for stage III. The independent samples *t*-test was used to analyse the differences in the expression levels between MCBs and DCBs. DCBs, ductal carcinoma of breasts; MCBs, metaplastic carcinoma of breasts.

Confirmation of microarray data

To validate the quality of the microarray data, we examined selected genes whose expression is particularly well coordinated in breast cancer cells. Specifically, we examined the correlation coefficients for *ERBB2* and *GRB7* expression, with *GRB7* being the neighboring gene of *ERBB2* and coamplified with it, and for distinctive coexpression pairs of cytokeratins (KRTs), including *KRT5*/*KRT14* and *KRT8*/*KRT18*. We found highly positive correlation for *ERBB2*/*GRB7* (0.833), *KRT8*/*KRT18* (0.788) and *KRT5*/*KRT14* (0.896) (Supplementary Figure 1), comparable to other data sets (Wilson and Dering, 2004). Gene expression profiles were further validated by comparing ERα, PR and HER-2/neu protein expression via immunohistochemical analyses to normalized gene expression data. We found statistically significant correlation between the expression of protein and gene expression level (ERα, $P = 0.01$ and HER-2/neu, $P = 0.0001$, *t*-test). To further confirm our microarray findings, we performed quantitative reverse transcription–PCR (QRT–PCR) analyses of several differentially expressed genes between MCBs and DCBs and between MCB subclasses. The QRT–PCR data correlated well with the expression levels, as revealed by microarray analysis; examples include those for *CTSZ*, *FOXA1*, *SPARC*, *PDGFRA*, *PDGFA*, *SFRP4*, *TIMP3* and *TSP1* (Figures 2b and 3d). These genes were next validated at the protein level in paraffin-embedded archival samples by IHC (immunohistochemistry) analyses, where the availability of appropriate antibodies made this possible. As shown in Figure 3e, SPARC and PDGFRA, which were encoded by the two genes upregulated in MCB subclass 2, were strongly immunostained in the sarcomatous components, which were the predominant tumor components in MCB subclass 2. This result also confirms that at least some of those genes upregulated in MCB subclass 2 are potentially related to the sarcomatous metaplasia of this carcinoma and highlights a potential role of these genes in EMT.

Hierarchical clustering of breast cancer samples using the 35 potentially EMT-related genes revealed clustering of MCBs and DCBs

Based on the distinct gene expression profiles of MCBs and DCBs and the EMT-like histologic features (sarcomatous change with ECM formation) seen in MCBs as compared to DCBs, we tested the possibility whether these 34 DCBs and 4 MCBs could be separated by hierarchical clustering using the 35 discriminator gene set derived from the MCB subclasses. Because these 35 genes were obtained by comparing the two subclasses of MCBs with relative abundance of morphological features of EMT, we hypothesized these genes might be more predictive of the genetic determinants underlying these EMT-related features. As shown in Figure 4, the 34 DCBs were clearly separated from the four MCBs, indicating a crucial role of these potentially EMT-related genes underlying the genetic distinction of MCBs and DCBs.

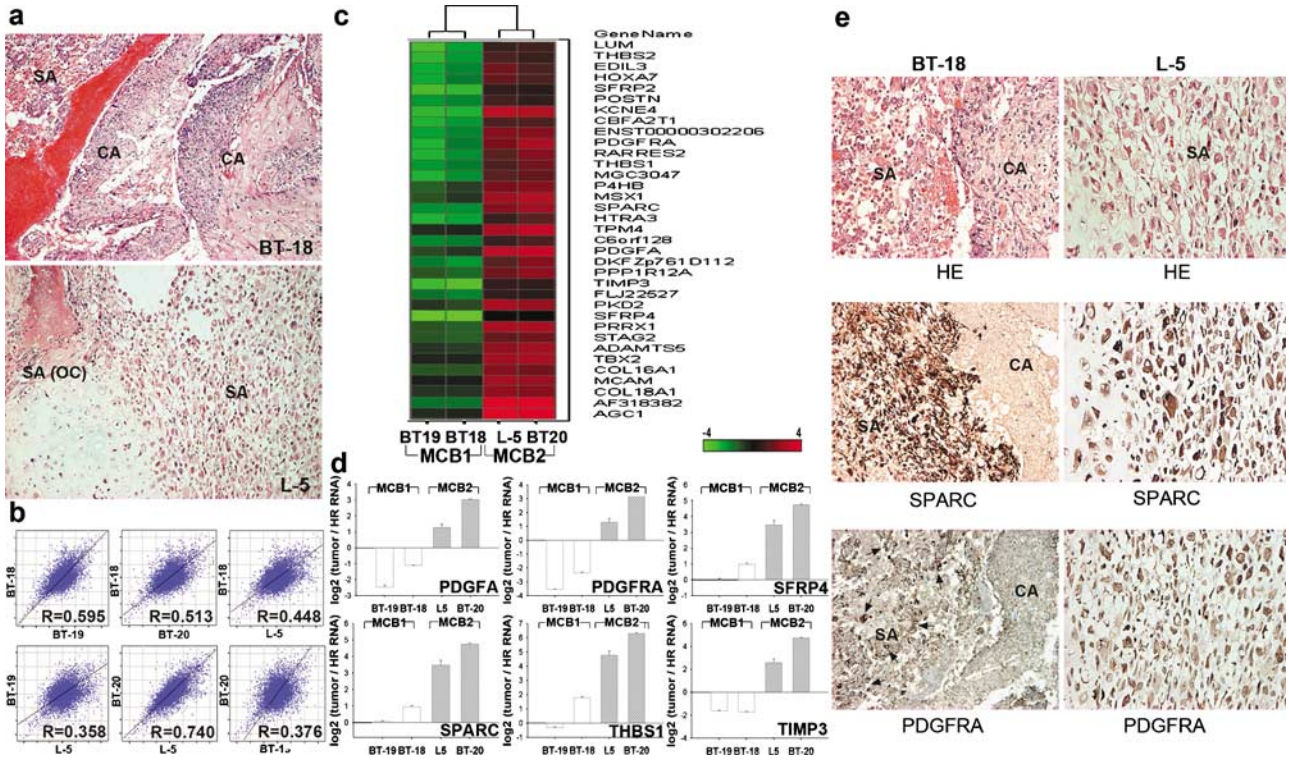

Figure 3 Histomorphology and gene expression profiles between subclasses of metaplastic carcinoma of breasts (MCBs). (**a**) Histomorphology in a representative case of MCB subclass 1 (BT-18) and 2 (L-5), respectively. BT-18 is composed predominantly of carcinomatous components (CA) and focally of sarcomatous components (SA). In contrast, L-5 is composed predominantly of sarcomatous components with frequent osteochondroid change (SA(OC)). (**b**) Scatter plots showing the substantial correlation between the gene expression profiling between the four MCBs. Among the four MCBs, the gene expression profiles were similar between L-5 and BT-20 ($R=0.74$) and between BT-18 and BT-19 ($R=0.59$). (**c**) Hierarchical clustering between MCB subclasses identified 35 discriminator genes. All 35 genes were upregulated in MCB subclass 2 and were annotated with gene symbol. (**d**) Confirmation by quantitative reverse transcription–PCR (QRT–PCR) of six genes (*PDGFA*, *PDGFRA*, *SFRP4*, *SPARC*, *THB1* and *TIMP3*) in tumor samples of MCBs. Log_2-fold changes in tumor transcript level to the reference transcript level and associated standard errors were plotted. The QRT–PCR data correlate well with the expression level as revealed by microarray analysis, that is overexpression of these genes in L-5 and BT-20 (MCB subclass 2) as compared to those in BT-18 and BT-19 (MCB subclass 1). (**e**) Immunohistochemistry (IHC) of SPARC and PDGFRA. SPARC and PDGFRA are strongly stained in the SA in both BT-18 and L-5, with no detection in the CA. Original magnifications, $\times 100$.

Validation of discriminator genes in a mutant Snail-transfected MCF7 breast cancer cell line exhibiting features of EMT

Snail, a zinc-finger transcription factor, plays a fundamental role in mediating EMT (Thiery, 2002; Zhou *et al*., 2004). Snail is highly unstable, but a variant of Snail (Snail-6SA) that abolishes two phosphorylation sites required, respectively, for nuclear to cytoplasmic translocation and subsequent ubiquitination, is much more stable and resides exclusively in the nucleus to induce EMT (Zhou *et al*., 2004). The Snail-6SA transfectant (Snail-6SA-MCF7) exhibits features of EMT, including the fibroblastoid and migratory phenotypes together with the loss of E-cadherin and the expression of vimentin, as compared to wild-type Snail-transfectant (Snail-WT-MCF7) or parental MCF7 cell line (Figures 5a–c) (Zhou *et al*., 2004). Using these cell lines as an *in vitro* model of EMT, we used QRT–PCR to examine whether genes differentially upregulated in the four MCBs and in MCB subclass 2, which showed prominent sarcomatous growth akin to EMT, were also upregulated in Snail-6SA-MCF7, as compared to Snail-WT-MCF7 or parental MCF7 cell lines. Based on our assumption that at least some of the 87 genes overexpressed in MCBs may be related to EMT, we selected two genes (*BMP2* and *VIM*) implicated in EMT (Lee *et al*., 2006) and two genes (*IMP2* and *CTSZ*) with no known association with EMT. Two of the highly ranked MCB-downregulated genes (*FOXA1* and *TFF3*) and *CDH1* (*E-cadherin*), which was critical in preventing EMT and was underexpressed in MCBs, were also included in the validation. As shown in Figure 5d, we found that *BMP2* and *VIM* were upregulated in Snail-6SA-MCF7 cells, consistent with the implicated roles of EMT in *BMP2* and *VIM*. In comparison, although having no implicated role in EMT, *IMP2* and *CTSZ* were also upregulated in Snail-6SA-MCF7 cells, suggesting a potential link between these two genes and EMT. Equally importantly, *FOXA1* and *TFF3*, which were downregulated in the four MCBs, were also downregulated in Snail-6SA-MCF7 cells. Similarly, we selected 11 MCB subclass-2-overexpressed genes including six (*PDGFRA*, *SFRP4*, *SPARC*, *TIMP3*, *PDGFA* and *HTRA3*) high-ranked genes and five (*EDIL3*,

Table 1 List of 35 genes differentially overexpressed in MCB subclass 2 as determined by SAM analysis with an FDR up to 0.05

Gene symbol	*Gene name*	*Reference sequence*	*SAM score*	*Fold difference (MCB subclass 2/1)*
AF318382	Pp9974 mRNA, complete cds	AF318382	6.50	71.30
KCNE4	Potassium voltage-gated channel, Isk-related family, member 4	NM_080671	8.78	56.55
PDGFRA	Platelet-derived growth factor receptor, α-polypeptide	NM_006206	3.93	22.42
AGC1	Aggrecan 1 (chondroitin sulfate proteoglycan 1, large aggregating proteoglycan, antigen identified by monoclonal antibody A0122)	NM_001135	5.35	22.35
SFRP4	Secreted frizzled-related protein 4	ENST00000223214	7.51	21.20
SPARC	Secreted protein, acidic, cysteine-rich (osteonectin)	NM_003118	5.51	20.84
ENST00000302206	Tropomyosin-1 (TM-β) mRNA, complete cds	ENST00000302206	5.32	18.86
RARRE2	Retinoic acid receptor responder (tazarotene-induced) 2	NM_002889	4.28	18.62
TIMP3	Tissue inhibitor of metalloproteinase 3 (Sorsby fundus dystrophy, pseudoinflammatory)	NM_000362	5.08	16.33
PRRX1	Paired-related homeobox 1	NM_006902	6.35	15.63
PDGFA	Platelet-derived growth factor α-polypeptide	NM_002607	4.93	15.52
MGC3047	Hypothetical protein MGC3047	NM_032348	4.71	15.08
DKFZp761D112	Homo sapiens mRNA; cDNA DKFZp761D112	AL136588	4.04	14.20
LUM	Lumican	NM_002345	4.00	14.10
SFRP2	Secreted frizzled-related protein 2	AF311912	5.19	14.09
HTRA3	Serine protease HTRA3	NM_053044	4.30	13.62
CBFA2T1	Core-binding factor, runt domain, α-subunit 2; translocated to, 1; cyclin D-related	NM_004349	5.02	12.83
EDIL3	EGF-like repeats and discoidin I-like domains 3	NM_005711	3.94	12.61
HOXA7	Homeo box A7	NM_006896	4.03	12.19
THB2	Thrombospondin 2	L12350	3.95	12.06
MSX1	Msh homeo box homolog 1 (*Drosophila*)	NM_002448	4.48	12.03
THB1	Thrombospondin 1	NM_003246	4.06	11.16
TPM4	Tropomyosin 4	NM_003290	4.18	10.06
PKD2	Polycystic kidney disease 2 (autosomal dominant)	NM_000297	4.46	9.97
PPP1R12A	Protein phosphatase 1, regulatory (inhibitor) subunit 12A	NM_002480	4.29	9.30
ADAMT5	A disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motif, 5 (aggrecanase-2)	NM_007038	4.73	8.24
COL16A1	Collagen, type XVI, α-1	NM_001856	4.69	8.23
POSTN	Periostin, osteoblast-specific factor	NM_006475	4.14	8.22
P4HB	Procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), β-polypeptide (protein disulfide isomerase; thyroid hormone binding protein p55)	NM_000918	3.91	8.03
C6orf128	Chromosome 6 open reading frame 128	NM_145316	4.25	7.71
TBX2	T-box 2	NM_005994	4.69	7.38
COL18A1	Collagen, type XVIII, α-1	NM_030582	4.46	6.93
STAG2	Stromal antigen 2	NM_006603	4.48	6.82
MCAM	Melanoma cell adhesion molecule	NM_006500	4.00	5.79
FLJ22527	Homo sapiens cDNA: FLJ22527 fis, clone HRC12820	AK026180	3.86	5.22

Abbreviations: FDR, false discovery rate; MCB, metaplastic carcinoma of the breast; SAM, significance analysis of microarrays.

THB2, *ADAMT5*, *STAG2* and *MCAM*) low-ranked known genes. We found nine (*SPARC*, *ADAMT5*, *MCAM*, *HTRA3*, *EDIL3*, *THB2*, *PDGFRA*, *TIMP3* and *SFRP4*) out of the 11 genes were upregulated in Snail-6SA-MCF7 cells (Figure 5d). Collectively, these data validate the microarray-derived results *in vitro* and further substantiate the potential relationship of these differentially expressed genes to EMT.

Validation of a previously diagnosed MCB (diagnostic test)

To validate the diagnostic utility of these two sets of discriminator genes, respectively, for MCBs and DCBs (Supplementary Table 1), and for MCB subclasses (Table 1), we performed supervised hierarchical clustering on a previously diagnosed MCB (MCB-MS) that was morphologically similar to MCB subclass 2, showing predominantly sarcomatous metaplasia with osteochondroid change (Figure 6a). In addition to the precise clustering of MCB-MS with MCBs on unsupervised clustering (Figure 6b), we found that these two gene sets correctly clustered MCB-MS with the four MCBs studied (Figure 6c), and clustered MCB-MS with MCB subclass 2, but not subclass 1 (Figure 6d). This validated the diagnostic utility of the results of our microarray experiments.

Evaluation of prognostic significance using tissue microarrays

Because EMT is implicated to play a key role in the invasiveness and metastasis of cancer progression and could be associated with a poor prognosis, we investigated whether these potential EMT-related genes demonstrated any correlation with prognosis in primary breast cancer tumor samples. Thus, we performed IHC analyses with antibodies against vimentin, SPARC and

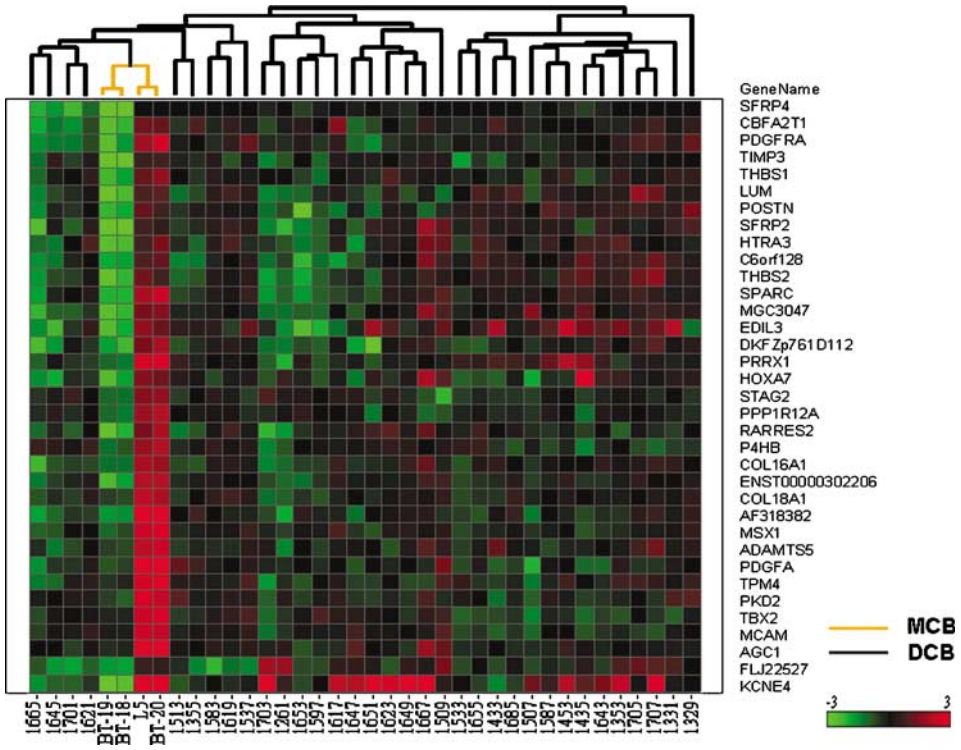

Figure 4 Hierarchical clustering of breast cancer samples using the 35 discriminator genes from the SAM (significance analysis of microarrays) analysis of metaplastic carcinoma of the breast (MCB) subclasses clearly separated the 34 ductal carcinoma of breasts (DCBs) from the four MCBs. Relative gene expression levels were depicted according to the color scale used in Figure 1.

PDGFRA on a tissue microarray consisting of primary breast tumor tissue samples from 144 patients with available clinicopathologic and overall survival data. Vimentin, SPARC and PDGFRA were positive in 7.0% (10/144), 16.6% (24/144) and 27.5% (39/142) of patients, respectively (Supplementary Table 4). We found statistically significant positive correlation between expression of EMT markers (vimentin and/or SPARC and/or PDGFRA) and tumor grade ($P=0.000$) and triple-negative status (ER, PR and HER2-negative) ($P=0.027$), suggesting a potential association between EMT and high-grade and triple-negative DCB. When Kaplan–Meier survival curves for overall survival were plotted, statistically significant poor prognoses were observed for vimentin-positive and SPARC-positive tumors ($P=0.0001$ and 0.0117, respectively, log-rank test) (Figure 7). By multivariate analysis, both vimentin and SPARC remained independent, poor prognostic indicators ($P=0.004$ and 0.005, respectively, Cox regression). PDGFRA, in contrast, was not associated with poor overall survival.

Discussion

We systemically surveyed gene expression of 4 MCBs and 34 DCBs on a genome-wide scale using oligonucleotide microarray techniques. Unsupervised hierarchical clustering clearly demonstrated distinctive gene expression profiles of MCBs and DCBs, exemplifying the inherent genetic differences between these two subtypes of breast cancer. Our data provided, for the first time, the genetic basis from a global approach for the classification of these two morphologically distinctive breast cancers.

Supervised hierarchical clustering disclosed 87 genes overexpressed and 121 genes underexpressed in MCBs. Most notably, several of the 87 MCB-overexpressed genes were functionally related to ECM, including genes associated with ECM synthesis, remodeling and adhesion/motility/migration, and genes associated with skeletal development and/or chondroossification. In contrast, genes encoding proteins related to maintaining epithelial phenotypes, including cytoskeleton, cell–cell adhesion molecules and tight junctions, were downregulated in MCBs. The upregulation of these ECM-related genes and the downregulation of epithelial-related genes in MCBs, especially mixed mesenchymal and epithelial MCBs, could account for the morphological features qof sarcomatous (fibroblastoid and motile) changes with ECM formation seen in MCBs, as compared to DCBs. Several other genes including *ADM*, *ENPP2*, *PLAGL1* and *BMP2* are also related to breast cancer. While ADM and ENPP2, potentially related to angiogenesis, are reportedly overexpressed in breast cancer (Oehler *et al*., 2003; Jansen *et al*., 2005), PLAGL1 and BMP2 exhibit tumor suppressor activity in breast cancer based on their growth inhibitory effects and their reduced expression in cancer tissue compared to normal

breast tissue (Soda *et al*., 1998; Bilanges *et al*., 1999; Reinholz *et al*., 2002). Despite their potential roles in the conventional type of breast cancer, how these genes participate in the pathophysiology of MCB remains to be clarified. BMP2, one of the TGF-β superfamily members, also plays a crucial role in EMT during cardiac cushion development (Ma *et al*., 2005). Since EMT is implicated in sarcomatous metaplasia of carcinoma cells in MCB (Thiery, 2002), the identification of *BMP2* as an upregulated gene in MCB and in Snail 6SA-MCF7 cells that display features of EMT further highlights a potential role for BMP2 underlying the features of EMT in MCBs and hinges on its role in EMT during cardiac cushion development. In addition, BMP2 also has a strong effect on the induction of bone formation (Ryoo *et al*., 2006), a feature commonly seen in MCBs. These multiple functions, taken together, strongly suggest a crucial role for BMP2 in the pathogenesis of MCB. CCAAT/enhancer binding protein beta (C/EBPbeta), a bZIP transcription factor involved in the regulation of genes involved in immune responses and in the stimulation of expression of the collagen type I gene, induces EMT of mammary epithelia (Bundy and Sealy, 2003). The identification of *C/EBPbeta* as an MCB-overexpressed gene could indicate a potential role for this transcription factor in the EMT in MCBs.

In contrast to conventional DCBs that show ER expression in more than two-thirds of cases, MCB generally lacks ER and PR expression (Barnes *et al*., 2005). Very little is known about the underlying biology of ER-negative breast cancer. Using microarray analysis, Doane *et al*. (2006) recently identified a subset of ER(−)/PR(−) breast cancers with paradoxical expression of genes known to be either direct targets of ER, responsive to ER or typically expressed in ER(+) breast cancer cases. They further suggested that these ER(−) cancers might be regulated by androgens, instead of estrogen hormones. Interestingly, several of those upregulated genes overlapped with genes downregulated in MCB in our study, including *AGR2*, *PIP*, *SPDEF*, *FOXA1*, *TFF3*, *AR*, and *XBP1*, suggesting that MCB, despite being ER(−)/PR(−), might have hormone regulatory pathways distinct from those in the reported subset of ER(−)/PR(−) breast cancer cases.

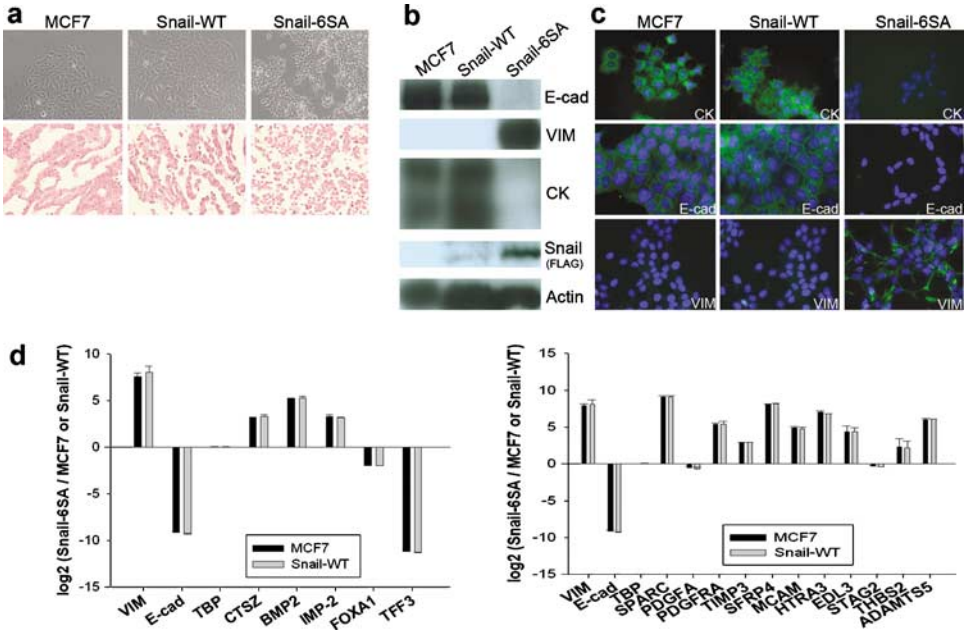

Figure 5 Phenotypes and differential expression of metaplastic carcinoma of the breast (MCB)-related genes in MCF7, Snail-WT-MCF7 and Snail-6SA-MCF7 cells. (**a**) Histomorphology. As compared to MCF7 and Snail-WT-MCF7 cells, Snail-6SA-MCF7 cells were more spindle shaped, less cohesive (phase-contrast microscopy, upper) and less duct-like (H&E, lower). (**b**) The expression of E-cadherin (E-cad), cytokeratin (CK), vimentin (VIM) and FLAG-tagged Snail (Snail) were analysed by western blotting. As compared to MCF7 and Snail-WT-MCF7 cells, Snail-6SA-MCF7 cells demonstrated the loss of epithelial markers of E-cadherin and cytokeratin and the gain of mesenchymal marker vimentin. The intense band of Snail in Snail-6SA-MCF7 cells is compatible with the more stable nature of Snail-6SA and the resultant features of EMT. (**c**) Immunofluorescent staining. Consistent with western blotting, Snail-6SA-MCF7 cells showed the loss of E-cadherin and cytokeratin, and the expression of vimentin. Together, these data demonstrated the features of epithelial–mesenchymal transition (EMT) in Snail-6SA-MCF7 cells, as compared to Snail-WE-MCF7 and MCF7 cells. (**d**) The differential expression of MCB-related genes in MCF7, Snail-WT-MCF7 and Snail-6SA-MCF7 cells were analysed by quantitative reverse transcription–PCR (QRT–PCR). Log_2-fold changes of Snail-6SA-MCF7 transcript level to Snail-WT-MCF7 or MCF7 transcript level and associated standard errors were plotted. Genes upregulated in MCBs relative to ductal carcinoma of breasts (DCBs), including *VIM*, *CTSZ*, *BMP2* and *IMP-2*, were also overexpressed in Snail-6SA-MCF7 cells (left). *E-cadherin*, *FOXA1* and *TFF3*, which were underexpressed in MCB, were downregulated in Snail-6SA-MCF7 cells. Nine of the 11 genes upregulated in MCB subclass 2 relative to MCB subclass 1, including *SPARC*, *ADAMT5*, *MCAM*, *HTRA3*, *EDIL3*, *THB2*, *PDGFRA*, *TIMP3* and *SFRP4*, were overexpressed in Snail-6SA-MCF7 cells (right). Consistent with western blotting and immunofluorescent staining, *E-cadherin* was downregulated in Snail-6SA-MCF7 cells. Equal expression levels of TATA box binding protein (TBP) were noted between these three cell lines.

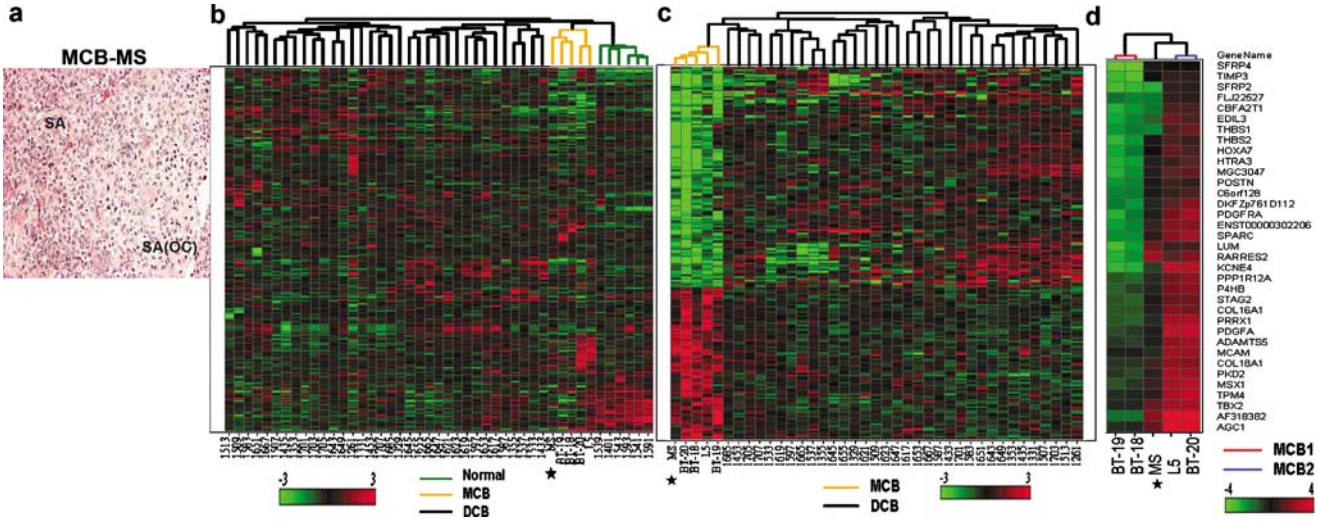

Figure 6 Hierarchical clustering analysis in the prediction of an additional case of metaplastic carcinoma of the breast (MCB) (MCB-MS). (**a**) The histology of MCB-MS revealed the predominance of sarcomatous components (SA) with frequent osteochondroid change (OC). (**b**) Unsupervised clustering clearly clusters MCB-MS (asterisk) with MCBs and separates it from ductal carcinoma of breast (DCBs). (**c**) Supervised clustering using the 208 discriminator genes between MCBs and DCBs correctly clusters MCB-MS (asterisk) with the four MCBs. (**d**) Supervised clustering using the 35 gene sets upregulated in MCB subclass 2 clearly clusters MCB-MS (asterisk) with MCB subclass 2.

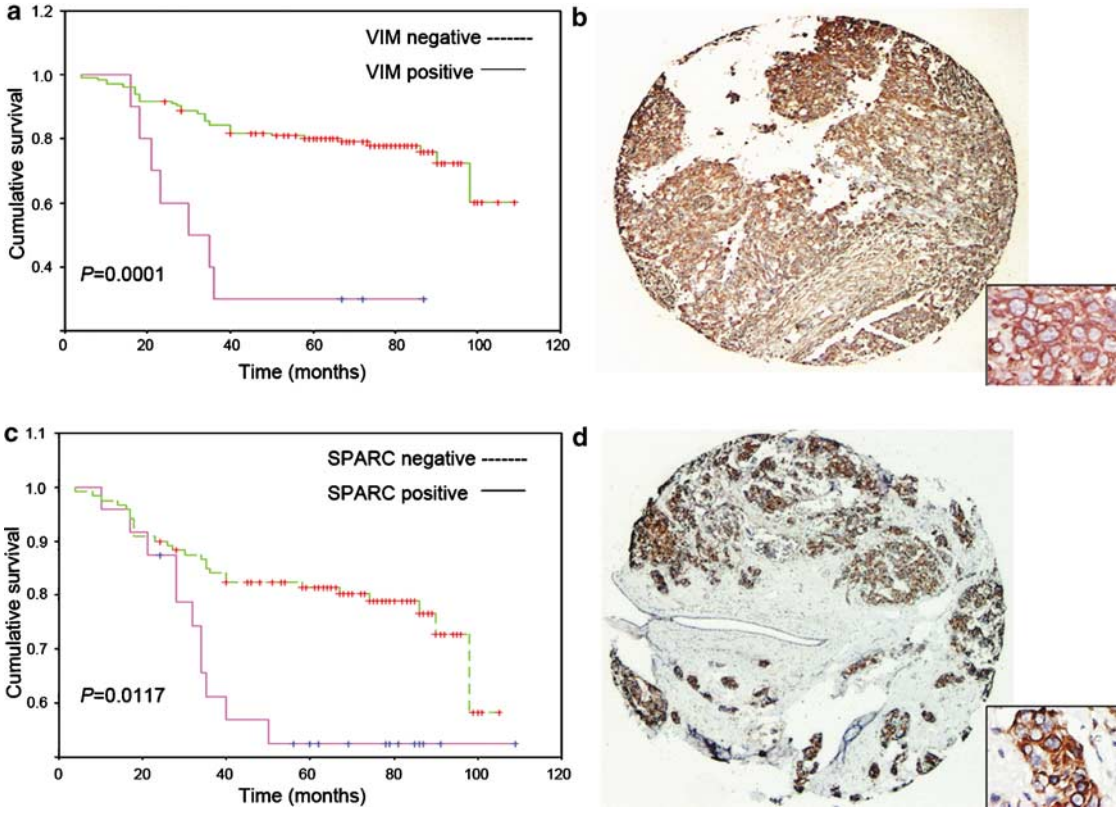

Figure 7 Kaplan–Meier survival curves from tissue microarray analysis. (**a** and **c**) Vimentin and SPARC expression were significantly associated with shorter overall survival in breast tumors ($P = 0.0001$ and 0.0117, respectively, log-rank test). (**b** and **d**) Expression of vimentin and SPARC in representative cases of invasive carcinoma. Original magnification, $\times 20$; $\times 60$ (inset).

In an effort to identify genes potentially related to the differences in morphological features seen between these two MCB subclasses, specifically the sarcomatous metaplasia with ECM formation and osteochondroid change, we performed supervised clustering and found 35 genes upregulated in MCB subclass 2. Expectedly, several of these genes are functionally related to synthesis, remodeling and adhesion of ECM. Notably,

some of these genes are related to the matrix component of cartilage (*AGC1*, *EDIL3*, *SPARC*, *MSX1* and *ADAMT5*) (Doege *et al*., 1991; Pfister *et al*., 2001; Framson and Sage, 2004; Gersch *et al*., 2005; Stanton *et al*., 2005) and bone (*SPARC*, *MSX1* and *POSTN*) (Framson and Sage, 2004; Gersch *et al*., 2005), which could account for the frequent osteochondroid change seen in MCB subclass 2 cancers. The identification of *SPARC* is of particular interest. SPARC, together with THB1 and THB2, belongs to a family of matricellular proteins functioning in mediating interactions between cells and their extracellular environments (Framson and Sage, 2004). The immunohistochemical demonstration of the intense expression of SPARC in sarcomatous tumor cells strongly suggests a potential role for SPARC in EMT. This notion is further supported *in vitro* by its deadhesive and motility-increasing effects on culture cells (Framson and Sage, 2004) and *in vivo* by its expression in pulmonary sarcomatoid carcinoma tumor cells, but not those of conventional lung cancer (Siddiq *et al*., 2004). In line with the critical role of EMT in cancer invasion and metastasis (Thiery, 2002), we found the expression of SPARC and vimentin to be independently associated with poor overall survival in breast cancer patients, findings also confirmed by other groups (Jones *et al*., 2004; Korsching *et al*., 2005). Similarly, the matrix proteoglycan lumican (*LUM*), which is also expressed in mesenchymal-like structures in pleomorphic adenomas of the salivary gland (Kusafuka *et al*., 2004), plays a role in EMT of lens epithelial cells in response to injury (Saika *et al*., 2003). Myoepithelial carcinoma or carcinoma with myoepithelial differentiation exhibits a partial or total spindle growth pattern (Rosai, 2004). MCB is suggested to have a myoepithelial origin and express myoepithelial markers (Reis-Filho *et al*., 2003). Interestingly, three of the MCB subclass-2-overexpressed genes, *SPARC*, *TIMP3* and *LUM*, are differentially overexpressed in myoepithelium (Jones *et al*., 2004; Kusafuka *et al*., 2004), and this might, in part, account for the sarcomatous (spindle) growth pattern seen in MCB. It is noteworthy that several of these potentially EMT-related genes have functions related to development, especially bone, cartilage and musculoskeletal development. How these genes participate in the process of EMT remains to be clarified in light of the pivotal role of EMT during embryonic development. Finally, the identification of both *PDGFA* and *PDGFRA* is of particular interest too. PDGF is a major growth factor for mesenchymal cells. A potential autocrine/paracrine loop of PDGFA and PDGFRA is implicated in epithelial–mesenchymal interaction during development (Xu *et al*., 2005) and is involved in the metastatic potential of TGF-β-induced EMT in breast cancer cells (Jechlinger *et al*., 2006). Taken together, these data substantiate the role of an autocrine loop of PDGFA and PDGFRA in EMT, which underlies the pathogenesis of MCB and might also imply a novel application of STI571, an inhibitor of PDGFRA, to therapeutically interfere with metastasis in this cancer.

Using Snail-6SA-MCF7 cells as an *in vitro* model of EMT, we found the majority of the tested discriminator genes between MCBs and DCBs and in MCB subclass 2 were also up-regulated in Snail-6SA-MCF7 cells as compared to Snail-WT-MCF7 or MCF7 parental cells, thereby suggesting a potential role of these MCB-related genes in EMT. Furthermore, the *in vivo* immunolocalization of SPARC and PDGFRA in the sarcomatous component of MCB and demonstrating that genes such as *VIM*, *SPARC*, *BMP2* and *PDGFRA* were upregulated in Snail-6SA-MCF7 cells indicates that some of those mesenchymal- or ECM-related genes, assumed to be restricted to mesenchymal or ECM, may also upregulate in carcinoma cells showing EMT. Collectively, these results validate our *in vivo* observation by showing that, in line with the features of EMT seen in MCBs, at least some of those genes differentially expressed in MCB are potentially related to EMT, and this in turn justifies the use of the mutant-Snail transfectant as a valid model system for the study of EMT.

Finally, the distinctive gene expression profile of MCB and the diagnostic utility of these two sets of discriminator genes, respectively, for MCB and DCB and for MCB subclasses, were validated by analysing the gene expression profile of a previously diagnosed additional case of MCB-MS that was morphologically similar to MCB subclass 2 cancer. In line with the purely morphological classification, MCB-MS was not only clearly clustered together with the MCBs and separated from the DCBs on unsupervised clustering, but also precisely clustered with the MCBs and MCB subclass 2 on supervised hierarchical clustering using these two sets of discriminator genes. These results demonstrate the robustness of these discriminator gene sets in the hierarchical clustering analysis by showing the consistence between the purely morphological classification and the gene expression profiling of the microarray experiments.

Although *SNAI1* (*Snail*) and *SNAI2* (*Slug*), which encode zinc-finger proteins crucial in the development of EMT, were identified as MCB-overexpressed genes on SAM analysis at an FDR of 0.05 (data not shown), some genes that are known to be activated during EMT (Zavadil and Bottinger, 2005) were absent in this genome-wide screen. It is possible that some of the molecular participants in EMT could appear only transiently in certain stages during the process of EMT, especially in the context of primary tumors when EMT takes a long time (Thiery, 2002; Zavadil and Bottinger, 2005). Alternatively, there might be potential differences in these molecular participants not only between embryonic development and cancer progression of different types, but also between *in vitro* and *in vivo* observations.

In conclusion, our study demonstrated distinctive gene profiling between MCBs and DCBs and validated the diagnostic utility of the discriminator gene sets. We elucidated and validated that certain genes differentially upregulated in MCB and MCB subclass 2 may have potential roles in EMT. The information derived from such global transcription profiling offers clues to the pathogenesis of MCB, and sheds light on the

mechanisms underlying EMT, and through such approaches, novel potential prognostic factors of cancer are found.

Materials and methods

Tumor tissue samples and cell lines
Surgical specimens of breast cancer tumor tissue were freshly collected and snap frozen from patients who underwent surgery at National Taiwan University Hospital (NTUH) between 1998 and 2005. Cancer samples containing relatively pure tumor, as defined by greater than 50% tumor cells per high-power field examined in a section adjacent to the tissue used, were included in this study. A total of four MCBs and 34 stage-matched DCBs were included. The microscopic slides from the patients' tumor specimens were reviewed by one of the authors (HCL) and the diagnosis and histological subclassification of cancers were confirmed in all cases. Normal breast tissues from six patients who underwent surgery for benign breast lesions were included for comparison. Linked clinicopathological data were obtained for all patients who contributed tumor specimens (Supplementary Table 5). One additional case of MCB (MCB-MS), provided by Min-Seng Su, MD, Ming-Seng Hospital, Taoyang, was used in the validation of microarray data. Studies using human tissues were approved by and conducted in accordance with the policies of the Institutional Review Board at NTUH. The wild-type MCF7 and Snail transfectants (Snail-WT-MCF7 and Snail-6SA-MCF7) were maintained in DMEM/F12 supplemented with 10% fetal calf serum.

Oligonucleotide microarray
Total RNA was extracted using Trizol Reagent (Invitrogen, Carlsbad, CA, USA), and followed by the RNeasy Mini Kit (Qiagen, Hilden, Germany). Extracted total RNA was amplified by a low RNA input fluor linear amp kit (Agilent Technologies, Foster City, CA, USA) during the *in vitro* transcription process (CyDye, PerkinElmer, Waltham, MA, USA). For the expression profiling of MCBs and DCBs, a human reference RNA pooled from 10 cell lines (Stratagene, La Jolla, CA, USA) served as the reference in the microarray comparison. Tumor RNA was labeled by Cy5 and RNA from human reference RNA was labeled by Cy3. Cy-labeled cRNA (2 μg) was fragmented to an average size of about 50–100 nucleotides by incubation with fragmentation buffer at 60°C for 30 min. Correspondingly, fragmented-labeled cRNA was then pooled and hybridized to the Human 1A (version 2) oligonucleotide microarray (Agilent Technologies) at 60°C for 17 h. After washing and drying in nitrogen, the microarrays were scanned with the Agilent microarray scanner at 535 nm for Cy3 and at 625 nm for Cy5. Scanned images were analysed using Feature Extraction software (Agilent Technologies). Only the features with signal noise ratios >2.6 in either the Cy3 or Cy5 channel were retrieved for further analysis.

Hierarchical clustering
A hierarchical clustering algorithm, UPGMA (Unweighted Pair-Group Method with Arithmetic mean), was applied to group genes and samples on the basis of their similarities in expression. The clustering was performed on all 44 samples by selecting transcripts that varied by at least fourfold from the median of the sample set in at least one of the samples. This resulted in a filtered gene list of 4525 probes representing 4334 genes (22.4% of 20 173 probes) for the unsupervised analysis. Results were displayed using the Spotfire program (Spotfire Inc., Somerville, MA, USA).

Statistical analysis of microarrays
The potentially significant changes in expression between DCBs and MCBs and between MCB subclasses were analysed using the SAM method (http://www-stat.stanford.edu/~tibs/SAM/). The filtered 4525 probes were performed with the SAM procedure, and a selection threshold was set at an FDR of 1 or 5%.

QRT–PCR
Total RNA from primary breast tumor tissues or cell lines was reverse transcribed into cDNA (Superscript II; Life Technologies Inc.). Primer sequences and the PCR program are described in Supplementary Table 6. Quantitative values were obtained from the threshold cycle number (C_t) and the fold-change in expression was calculated using the $\Delta\Delta C_t$ method. Target genes measurements in all samples were both normalized to the internal control gene TATA box binding protein and human common reference RNA.

IHC and immunofluorescent staining
Paraffin sections from primary breast cancer tissues were stained by IHC as described previously (Lien *et al*., 2004). Primary antibodies used in the IHC were as follows: monoclonal antibody against human SPARC (Biogenex, San Ramon, CA, USA), PDGFRA (Neomarker, Fremont, CA, USA), (DAKO, Glostrup, Denmark), E-cadherin and vimentin (Santa Cruz Biotechnology, Santa Cruz, CA, USA). For immunofluorescent staining, slides were incubated with primary mouse Ab (anti-vimentin, anti-E-cadherin or anti-CK), washed and incubated with FITC-conjugated secondary antibody (Ab), before counterstaining with DAPI. The slides were examined under a Zeiss Axioplan 2 fluorescence microscope fitted with an ApoTome slider.

Western blotting
Proteins from cell lysates were resolved on sodium dodecyl sulfate–polyacrylamide gels. Immunoblotting was performed with the addition of antibodies to E-cadherin, vimentin, cytokeratin, actin and FLAG-tagged Snail, according to the manufacturer's protocol.

Tissue array and survival analysis
One hundred and forty-four archives of breast cancer paraffin tissue blocks from the Department of Pathology at NTUH, with up to 11 years of clinical follow-up data, were arrayed in triplicate in a high-density tissue array, using an automated arrayer (Beecher Instruments, Sun Prairie, WI, USA). The clinicopathologic features are shown in Supplementary Table 7. Statistical analyses were carried out using SPSS 12.0 software for Windows. The cumulative overall survival was calculated using the Kaplan–Meier method, and the log-rank test was used to analyse differences in the survival times.

Acknowledgements

This study was supported by National Science Council Grant NSC 94-2320-B-002-057 (HCL), Grant NSC 95-2314-B002-012 (KJC) and Grant NSC 94-2314-B-002-118 (FJH).

References

Al Sayed AD, El Weshi AN, Tulbah AM, Rahal MM, Ezzat AA. (2006). Metaplastic carcinoma of the breast clinical presentation, treatment results and prognostic factors. *Acta Oncol* **45**: 188–195.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM *et al*. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–29.

Barnes PJ, Boutilier R, Chiasson D, Rayson D. (2005). Metaplastic breast carcinoma: clinical-pathologic characteristics and HER2/neu expression. *Breast Cancer Res Treat* **91**: 173–178.

Bilanges B, Varrault A, Basyuk E, Rodriguez C, Mazumdar A, Pantaloni C *et al*. (1999). Loss of expression of the candidate tumor suppressor gene ZAC in breast cancer cell lines and primary tumors. *Oncogene* **18**: 3979–3988.

Bundy LM, Sealy L. (2003). CCAAT/enhancer binding protein beta (C/EBPbeta)-2 transforms normal mammary epithelial cells and induces epithelial to mesenchymal transition in culture. *Oncogene* **22**: 869–883.

Carter MR, Hornick JL, Lester S, Fletcher CD. (2006). Spindle cell (sarcomatoid) carcinoma of the breast: a clinicopathologic and immunohistochemical analysis of 29 cases. *Am J Surg Pathol* **30**: 300–309.

Doane AS, Danso M, Lal P, Donaton M, Zhang L, Hudis C *et al*. (2006). An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. *Oncogene* **25**: 3994–4008.

Doege KJ, Sasaki M, Kimura T, Yamada Y. (1991). Complete coding sequence and deduced primary structure of the human cartilage large aggregating proteoglycan, aggrecan. Human-specific repeats, and additional alternatively spliced forms. *J Biol Chem* **266**: 894–902.

Framson PE, Sage EH. (2004). SPARC and tumor growth: where the seed meets the soil? *J Cell Biochem* **92**: 679–690.

Fujiwara S, Takeo N, Otani Y, Parry DA, Kunimatsu M, Lu R *et al*. (2001). Epiplakin, a novel member of the Plakin family originally identified as a 450-kDa human epidermal autoantigen. Structure and tissue localization. *J Biol Chem* **276**: 13340–13347.

Gersch RP, Lombardo F, McGovern SC, Hadjiargyrou M. (2005). Reactivation of Hox gene expression during bone regeneration. *J Orthop Res* **23**: 882–890.

Jansen S, Stefan C, Creemers JW, Waelkens E, Van Eynde A, Stalmans W *et al*. (2005). Proteolytic maturation and activation of autotaxin (NPP2), a secreted metastasis-enhancing lysophospholipase D. *J Cell Sci* **118**: 3081–3089.

Jechlinger M, Sommer A, Moriggl R, Seither P, Kraut N, Capodiecci P *et al*. (2006). Autocrine PDGFR signaling promotes mammary cancer metastasis. *J Clin Invest* **116**: 1561–1570.

Jones C, Mackay A, Grigoriadis A, Cossu A, Reis-Filho JS, Fulford L *et al*. (2004). Expression profiling of purified normal human luminal and myoepithelial breast cells: identification of novel prognostic markers for breast cancer. *Cancer Res* **64**: 3037–3045.

Korsching E, Packeisen J, Liedtke C, Hungermann D, Wulfing P, van Diest PJ *et al*. (2005). The origin of vimentin expression in invasive breast cancer: epithelial–mesenchymal transition, myoepithelial histogenesis or histogenesis from progenitor cells with bilinear differentiation potential? *J Pathol* **206**: 451–457.

Kusafuka K, Ishiwata T, Sugisaki Y, Takemura T, Kusafuka M, Hisha H *et al*. (2004). Lumican expression is associated with the formation of mesenchyme-like elements in salivary pleomorphic adenomas. *J Pathol* **203**: 953–960.

Lee JM, Dedhar S, Kalluri R, Thompson EW. (2006). The epithelial–mesenchymal transition: new insights in signaling, development, and disease. *J Cell Biol* **172**: 973–981.

Lemmers C, Michel D, Lane-Guermonprez L, Delgrossi MH, Medina E, Arsanto JP *et al*. (2004). CRB3 binds directly to Par6 and regulates the morphogenesis of the tight junctions in mammalian epithelial cells. *Mol Biol Cell* **15**: 1324–1333.

Lien HC, Lin CW, Mao TL, Kuo SH, Hsiao CH, Huang CS. (2004). p53 overexpression and mutation in metaplastic carcinoma of the breast: genetic evidence for a monoclonal origin of both the carcinomatous and the heterogeneous sarcomatous components. *J Pathol* **204**: 131–139.

Liu Y, Nusrat A, Schnell FJ, Reaves TA, Walsh S, Pochet M *et al*. (2000). Human junction adhesion molecule regulates tight junction resealing in epithelia. *J Cell Sci* **113**: 2363–2374.

Ma L, Lu MF, Schwartz RJ, Martin JF. (2005). Bmp2 is essential for cardiac cushion epithelial–mesenchymal transition and myocardial patterning. *Development* **132**: 5601–5611.

Oehler MK, Fischer DC, Orlowska-Volk M, Herrle F, Kieback DG, Rees MC *et al*. (2003). Tissue and plasma expression of the angiogenic peptide adrenomedullin in breast cancer. *Br J Cancer* **89**: 1927–1933.

Pezzi CM, Patel-Parekh L, Cole K, Franko J, Klimberg VS, Bland K. (2007). Characteristics and treatment of metaplastic breast cancer: analysis of 892 cases from the national cancer data base. *Ann Surg Oncol* **14**: 166–173.

Pfister BE, Aydelotte MB, Burkhart W, Kuettner KE, Schmid TM. (2001). Del1: a new protein in the superficial layer of articular cartilage. *Biochem Biophys Res Commun* **286**: 268–273.

Reinholz MM, Iturria SJ, Ingle JN, Roche PC. (2002). Differential gene expression of TGF-beta family members and osteopontin in breast tumor tissue: analysis by real-time quantitative PCR. *Breast Cancer Res Treat* **74**: 255–269.

Reis-Filho JS, Milanezi F, Paredes J, Silva P, Pereira EM, Maeda SA *et al*. (2003). Novel and classic myoepithelial/stem cell markers in metaplastic carcinomas of the breast. *Appl Immunohistochem Mol Morphol* **11**: 1–8.

Rosai J. (2004). *Rosai and Ackerman's Surgical Pathology*. St Louis: Mosby, pp 1828–1829.

Ryoo HM, Lee MH, Kim YJ. (2006). Critical molecular switches involved in BMP-2-induced osteogenic differentiation of mesenchymal cells. *Gene* **366**: 51–57.

Saika S, Miyamoto T, Tanaka S, Tanaka T, Ishida I, Ohnishi Y *et al*. (2003). Response of lens epithelial cells to injury: role of lumican in epithelial–mesenchymal transition. *Invest Ophthalmol Vis Sci* **44**: 2094–2102.

Siddiq F, Sarkar FH, Wali A, Pass HI, Lonardo F. (2004). Increased osteonectin expression is associated with malignant transformation and tumor associated fibrosis in the lung. *Lung Cancer* **45**: 197–205.

Soda H, Raymond E, Sharma S, Lawrence R, Cerna C, Gomez L *et al*. (1998). Antiproliferative effects of recombinant human bone morphogenetic protein-2 on human tumor colony-forming units. *Anticancer Drugs* **9**: 327–331.

Stanton H, Rogerson FM, East CJ, Golub SB, Lawlor KE, Meeker CT *et al*. (2005). ADAMTS5 is the major aggrecanase in mouse cartilage *in vivo* and *in vitro*. *Nature* **434**: 648–652.

Thiery JP. (2002). Epithelial–mesenchymal transitions in tumour progression. *Nat Rev Cancer* **2**: 442–454.

Wilson CA, Dering J. (2004). Recent translational research: microarray expression profiling of breast cancer – beyond classification and prognostic markers? *Breast Cancer Res* **6**: 192–200.

Xu X, Bringas Jr P, Soriano P, Chai Y. (2005). PDGFR-alpha signaling is critical for tooth cusp and palate morphogenesis. *Dev Dyn* **232**: 75–84.

Zavadil J, Bottinger EP. (2005). TGF-beta and epithelial-to-mesenchymal transitions. *Oncogene* **24**: 5764–5774.

Zhou BP, Deng J, Xia W, Xu J, Li YM, Gunduz M *et al*. (2004). Dual regulation of Snail by GSK-3beta-mediated phosphorylation in control of epithelial–mesenchymal transition. *Nat Cell Biol* **6**: 931–940.

Zhuang Z, Lininger RA, Man YG, Albuquerque A, Merino MJ, Tavassoli FA. (1997). Identical clonality of both components of mammary carcinosarcoma with differential loss of heterozygosity. *Mod Pathol* **10**: 354–362.

Supplementary Information accompanies the paper on the Oncogene website (http://www.nature.com/onc).