

# Progression-Specific Genes Identified by Expression Profiling of Matched Ductal Carcinomas *In situ* and Invasive Breast Tumors, Combining Laser Capture Microdissection and Oligonucleotide Microarray Analysis

Christina S. Schuetz,<sup>1</sup> Michael Bonin,<sup>2</sup> Susan E. Clare,<sup>3</sup> Kay Nieselt,<sup>4</sup> Karl Sotlar,<sup>5</sup> Michael Walter,<sup>2</sup> Tanja Fehm,<sup>1</sup> Erich Solomayer,<sup>1</sup> Olaf Riess,<sup>2</sup> Diethelm Wallwiener,<sup>1</sup> Raffael Kurek,<sup>1</sup> and Hans J. Neubauer<sup>1</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, <sup>2</sup>Microarray Facility, Department of Medical Genetics, University of Tuebingen, Tuebingen, Germany; <sup>3</sup>Department of Surgery, Indiana University School of Medicine, Indianapolis, Indiana; and <sup>4</sup>Center of Bioinformatics, and <sup>5</sup>Department of Pathology, University of Tuebingen, Tuebingen, Germany

## Abstract

Becoming invasive is a crucial step in breast cancer oncogenesis. At this point, a lesion carries the potential for spreading and metastasis—a process, whose molecular characteristics still remain poorly understood. In this article, we describe a matched-pair analysis of ductal carcinoma *in situ* (DCIS) and invasive ductal carcinoma (IDC) of nine breast ductal carcinomas to identify novel molecular markers characterizing the transition from DCIS to IDC. The purpose of this study was to better understand the molecular biology of this transition and to identify candidate genes whose products might serve as prognostic markers and/or as molecular targets for treatment. To obtain cellular-based gene expression profiles from epithelial tumor cells, we combined laser capture microdissection with a T7-based two-round RNA amplification and Affymetrix oligonucleotide microarray analysis. Altogether, a set of 24 tumor samples was analyzed, comprised of nine matched DCIS/IDC and replicate DCIS/IDC preparations from three of the nine tumors. Cluster analysis on expression data shows the robustness and reproducibility of the techniques we established. Using multiple statistical methods, 546 significantly differentially expressed probe sets were identified. Eighteen candidate genes were evaluated by RT-PCR. Examples of genes already known to be associated with breast cancer invasion are *BPAG1*, *LRRC15*, *MMP11*, and *PLAU*. The expression of *BPAG1*, *DACT1*, *GREM1*, *MEF2C*, *SART2*, and *TNFAIP6* was localized to epithelial tumor cells by *in situ* hybridization and/or immunohistochemistry, confirming the accuracy of laser capture microdissection sampling and microarray analysis. (Cancer Res 2006; 66(10): 5278-86)

## Introduction

The prevailing multistep model of mammary carcinogenesis is a sequence of pathologically defined stages initiating as the noninvasive atypical ductal hyperplasia<sup>6</sup> and progressing to preinvasive ductal carcinoma *in situ* (DCIS), a nonobligate

precursor of the final stage of invasive ductal carcinoma (IDC; refs. 1, 2). Untreated, one group of DCIS will relatively rapidly progress to the invasive stage, whereas another group will change very little within 5 to 20 years (3). Classical histopathology has a limited ability to accurately predict the risk of progression to invasive disease (4, 5). Therefore, understanding the molecular biology of DCIS and its transition to IDC may provide insight into tumor initiation and progression with the aim of identifying biologically and clinically meaningful target genes and proteins. Gene expression profiling using microarray technology is a powerful approach for elucidating the transcriptional program and provides a requisite for the classification of a tumor according to its gene expression profile (6). Various studies suggest that a set of genes can show distinct patterns of expression in different types of breast cancer and that these profiles can be linked to prognosis (7, 8). Further investigations indicate the power of microarray analysis to characterize different progression stages in breast cancer (9–11). Morphologically, breast tumors are generally exceedingly heterogeneous (11). Hence, to obtain a rather more cellular-based than tissue-based resolution for gene expression profiles, the use of methods for cell separation such as laser capture microdissection (LCM) is essential (9–14).

The goal of this study was to identify genes that provide an understanding of the molecular biology of DCIS and to obtain genes marking the transition of stationary epithelial cells to migrating invasive cells. In order to reduce the false-positive rate caused by genetic variability, we used patient-matched DCIS/IDC tumor samples.

## Materials and Methods

**Tissue.** Primary breast cancer specimens were obtained with informed consent from patients who were treated at the Department of Gynecology and Obstetrics (University Hospital Tuebingen, Germany, a certified and multidisciplinary breast center). Tissue samples are cryopreserved by no later than 15 minutes postoperation in a liquid nitrogen tissue bank (ethical consent of the Medical Faculty Tuebingen, AZ.266/98). First selection of specimens was carried out using an Oracle-based tumor database (TumorAGENT),<sup>7</sup> based on clinical and histological information.

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org>).

R. Kurek and H.J. Neubauer share senior authorship.

**Requests for reprints:** Hans J. Neubauer, Tumor Progression Lab, Gynecological Hospital, Calwerstrasse 7/6, 72076 Tuebingen, Germany. Phone: 49-7071-298-5380; Fax: 49-7071-29-4947; E-mail: [hans.neubauer@med.uni-tuebingen.de](mailto:hans.neubauer@med.uni-tuebingen.de).

©2006 American Association for Cancer Research.

doi:10.1158/0008-5472.CAN-05-4610

<sup>6</sup> We use traditional terminology although it was replaced by ductal intraepithelial neoplasia classification because the histopathology for the tumors used in our experiments is based on the previous nomenclature.

<sup>7</sup> A. Endress, S. Clare, T. Fehm, H. Neubauer, E. Solomayer, S. Lessmann, C. Schuetz, H. Loeffler, K. Abt, D. Wallwiener, R. Kurek. TumorAGENT—an interdisciplinary tumor database for translational research in breast cancer. Submitted for publication.

Paraffin-embedded breast tissue was obtained from the Department of Pathology, University Hospital in Tuebingen.

**Laser capture microdissection.** Serial frozen tissue sections (10  $\mu\text{m}$ ) were fixed and rehydrated sequentially in decreasing concentrations of ethanol and RNase-free water followed by H&E staining. Finally, sections were dehydrated in 95% and 100% ethanol and 8 minutes in xylene. The PixCell Iie LCM System (Arcturus Engineering, Mountain View, CA) was used according to the manufacturer's instructions. For microarray experiments, ~20,000 cells (7,000 laser pulses, 30  $\mu\text{m}$  diameter laser) were dissected. LCM caps containing the isolated tumor cells were stored at  $-80^{\circ}\text{C}$  in 40  $\mu\text{L}$  RLT buffer (Qiagen, Hilden, Germany) + 1%  $\beta$ -mercaptoethanol until RNA extraction.

**RNA extraction and linear amplification.** Lysates of dissected tissue from several LCM caps were pooled into a final volume of 350  $\mu\text{L}$  RLT buffer (Qiagen) + 1%  $\beta$ -mercaptoethanol. Total RNA extraction was carried out using RNeasy Micro Kit (Qiagen) according to the manufacturer's instructions, combined with the protocol for LCM-derived small samples (15). Two-round linear amplification was carried out according to the GeneChip Eukaryotic Small Sample Target Labeling Assay version II protocol (Affymetrix, High Wycombe, United Kingdom), including 0.25  $\mu\text{L}$  T4gp32-protein in every reverse transcriptase reaction (16). The quality of total RNA of tumor sections, of microdissected tissue and of antisense RNA (aRNA) was monitored by Agilent 2100 Bioanalyzer using the RNA 6000 Nano LabChip Kit (Agilent Technologies, Boeblingen, Germany) following the manufacturer's instructions.

**Oligonucleotide microarrays and acquisition of data.** After the second round of amplification, 5  $\mu\text{g}$  of biotinylated fragmented aRNA was hybridized to Test3 arrays (Affymetrix). Affymetrix high-density oligonucleotide microarrays (GeneChip HG U133A and GeneChip HG U133 plus 2.0; both Affymetrix) were used for gene expression analysis. Hybridization experiments and evaluation were done by the Microarray Facility Tuebingen. The arrays were scanned at 3  $\mu\text{m}$  resolution using a GeneChip System confocal scanner (Agilent Technologies). Scanned images were subjected to visual inspection and analyzed using the Affymetrix's Microarray Suite version 5.0 (MAS 5.0) algorithms (Affymetrix) to generate report files for quality control. MAS 5.0 was used to compute detection  $P$  values. Transcripts with  $P < 0.04$  were classified as present, whereas they were classified as absent at  $P > 0.04$ . RMAExpress (17) was applied for normalization of raw data (CEL files) independently of the two different microarrays. The data discussed in this publication have been deposited in National Center for Biotechnology Information's Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) and are accessible through Gene Expression Omnibus Series accession number GSE3893.

**Hierarchical cluster analysis.** RMA-normalized expression values of the arrays of the HG U133A and the HG U133 plus 2.0 series were  $z$  score-normalized independently. A  $z$  score is computed by subtracting the probe's profile mean value from its expression value and subsequently dividing it by its SD. These normalized values were then combined into one data set, considering only genes represented on both Affymetrix arrays. Hierarchical clustering was done using the neighbor-joining method (18) applied to the Pearson correlation distance matrix. In addition, other hierarchical clustering methods, such as UPGMA, and other distance-generating functions, such as Euclidean, were used. All analyses were done within the Mayday Software Package.<sup>8</sup>

**Statistics.** Multiple statistical methods were conducted to obtain differentially expressed genes between DCIS and IDC. The HG U133A and HG U133 Plus 2.0 probe sets were combined, and only those probes for which at least one in all HG U133A samples was characterized as present were considered (as determined by the MAS 5.0 software using  $P < 0.04$ ). Thus, a total of 13,431 probes were assembled into a common expression matrix. Each test was applied to the RMA-normalized expression values of these probes. (a) Rank product (19) was conducted by using R library

package RankProd distributed by the Bioconductor project (<http://www.bioconductor.org>). The rank product is a new nonparametric type of statistical method suggested by Breitling et al. It detects genes that are consistently found among the most strongly up-regulated or down-regulated genes in a number of replicate experiments. First for each gene or transcript, the log ratio of the matched samples expression values were determined, and genes were then ranked with respect to decreasing log ratio. Then, for each gene, the geometric mean of its ranks was determined. From these means, two sorted lists for putative up-regulated and down-regulated genes and their rank product values were generated. To evaluate the significance of a rank product value, the percentage of false-positive predictions also known as the false discovery rate is calculated using a permutation-based procedure (100,000 permutations were conducted). Genes with a false-positive prediction  $< 0.01$  were regarded as significantly differentially expressed. (b) Paired  $t$  test was used to detect genes that are up-regulated and down-regulated in the IDC samples in comparison with the DCIS samples ( $P < 0.01$ ). (c) Paired Mann-Whitney test was used to detect genes that are up-regulated and down-regulated in the IDC samples in comparison to the DCIS samples ( $P < 0.01$ ). The respective lists of all methods (a-c) were combined, and the intersection was taken resulting into a "master gene list." This data set was analyzed through the use of Ingenuity Pathways Analysis (Ingenuity Systems, <http://www.ingenuity.com>). Functional Analysis identified the biological functions and/or diseases that were most significant to the data set. Genes from the data set that met the negative logarithmic significance cutoff of 5 or higher, and were associated with biological functions and/or diseases in the Ingenuity Pathways knowledge base were considered for the analysis. Fischer's exact test was used to calculate a  $P$  value determining the probability that each biological function and/or disease assigned to that data set is due to chance alone.

**RT-PCR.** A 1  $\mu\text{g}$  aliquot of aRNA after the first round of amplification was reverse-transcribed for single-stranded cDNA using random primers and Superscript II kit (Invitrogen, Karlsruhe, Germany). No separate microdissection of tissue was carried out for the RT-PCR experiments to enable direct validation of the microarray experiments, avoiding discrepancies due to technical variations. Classical PCR was done using Taq DNA polymerase kit (Invitrogen). Real-time PCR was done with the LightCycler System (Roche, Mannheim, Germany) and the Platinum SYBR Green qPCR Supermix UDG (Invitrogen) according to the manufacturer's instructions (primer sequences and specific annealing temperature; Supplementary data 1). For relative quantification of gene expression, triplicate reactions were set up. Expression of pyruvate dehydrogenase  $\beta$ -subunit (PDH) was used to determine the relative regulation of candidates employing the efficiency-corrected equation (20). The efficiency of every single PCR reaction was determined with LinRegPCR software (version 7.5, February 2004; ref. 21).

**In situ hybridization.** PCR products (primer sequences and specific annealing temperature; Supplementary data 1) were cloned into TOPO II vector (Invitrogen). Hybridization experiments on 4- $\mu\text{m}$  paraffin sections were carried out as described previously using digoxigenin-labeled RNA probes of ~300 bp in length targeting the 3'-end of the transcripts (RNA labeling kit, Roche; refs. 3, 22). The following improvements were accomplished: during the acetylation step, following proteinase K digestion, slides were rinsed in 0.5% acetic anhydride/0.1 mol/L triethanolamine for 10 minutes. Furthermore, an RNase One digest (10 U/mL; Promega, Mannheim, Germany) was carried out for 30 minutes during the washing steps after hybridization. The riboprobe (1 ng/ $\mu\text{L}$ ) was hybridized. Hybridizations were considered successful if the sense probe gave no significant signal.

**Immunohistochemistry.** Cryosections (10  $\mu\text{m}$ ) were fixed for 10 minutes in acetone, followed by incubation with the primary antibody for 30 to 60 minutes. Further steps were carried out using the respective Vectastain kit (Vector Laboratories, Burlingame, CA) according to the manufacturer's instructions. HistoGreen (Linaris, Wertheim-Bettingen, Germany) was used as the chromogen. Slides were counterstained with hematoxylin and mounted in pterex (Medite, Burgdorf, Germany). The following primary antibodies were used: polyclonal rabbit anti-human DACT1 antibody (Orbigen, San Diego, CA), polyclonal goat anti-human

<sup>8</sup> J. Dietzsch, N. Gehlenborg, and K. Nieselt, 2006. Mayday: a microarray data analysis workbench. *Bioinformatics Advance Access published online on February 24, 2006* *Bioinformatics*, doi:10.1093/bioinformatics/bt1070.

**Table 1.** Clinical and histopathologic data and 3'/M ratios

Tumor ID		Tumor characteristics				Hybridization characteristics		
No.	Specimen	Grade	ER-IRS*	PR-IRS*	HER2/neu-IRS*	Microarray	3'/M HUMGAPDH <sup>†</sup>	3'/M HSAC07 <sup>‡</sup>
1.1	T374 DCIS	G2-3	0	0	0	HG U133A	1.87	3.34
1.2	T374 IDC					HG U133A	1.44	2.65
2.1	T478 DCIS	G2-3	12	6	0	HG U133A	1.44	6.11
2.2	T478 IDC					HG U133A	1.25	3.35
3.1	T661 DCIS	G3	12	9	+3	HG U133A	1.19	3.35
3.2	T661 IDC					HG U133A	1.36	3.14
3.1 <sup>§</sup>	T661 DCIS					HG U133 Plus 2.0	1.14	1.78
3.2 <sup>§</sup>	T661 IDC					HG U133 Plus 2.0	1.39	2.45
4.1	T787 DCIS	G2	8	2	+2	HG U133A	2.79	8.50
4.2	T787 IDC					HG U133A	1.39	3.37
4.1 <sup>§</sup>	T787 DCIS					HG U133 Plus 2.0	1.6	2.42
4.2 <sup>§</sup>	T787 IDC					HG U133 Plus 2.0	1.37	2.94
5.1	T808 DCIS	G2	0	0	0	HG U133A	1.41	3.67
5.2	T808 IDC					HG U133A	1.40	5.12
5.1 <sup>§</sup>	T808 DCIS					HG U133 Plus 2.0	1.25	1.76
5.2 <sup>§</sup>	T808 IDC					HG U133 Plus 2.0	1.11	1.97
6.1	H191 DCIS	G2	8-12	0	0	HG U133 Plus 2.0	1.75	2.36
6.2	H191 IDC					HG U133 Plus 2.0	2.0	3.6
7.1	T796 DCIS	G2	12	12	0	HG U133 Plus 2.0	2.1	5.94
7.2	T796 IDC					HG U133 Plus 2.0	1.79	3.51
8.1	T396 DCIS	G2	12	0	0	HG U133 Plus 2.0	1.68	3.82
8.2	T396 IDC					HG U133 Plus 2.0	1.97	6.22
9.1	T706 DCIS	G2-3	0	0	+3	HG U133 Plus 2.0	1.81	3.56
9.2	T706 IDC					HG U133 Plus 2.0	2.22	4.39

\*Immunoreactive score.

† Glyceraldehyde 3-phosphate dehydrogenase.

‡ β-Actin.

§ Technical replicate.

GREM1 antibody and polyclonal goat anti-human MEF2C (both from Santa Cruz Biotechnology, Heidelberg, Germany). Polyclonal rabbit anti-human BPAG1 antibody was a kind gift from Dr. J.R. Stanley (Department of Dermatology, University of Pennsylvania, Philadelphia, PA).

## Results

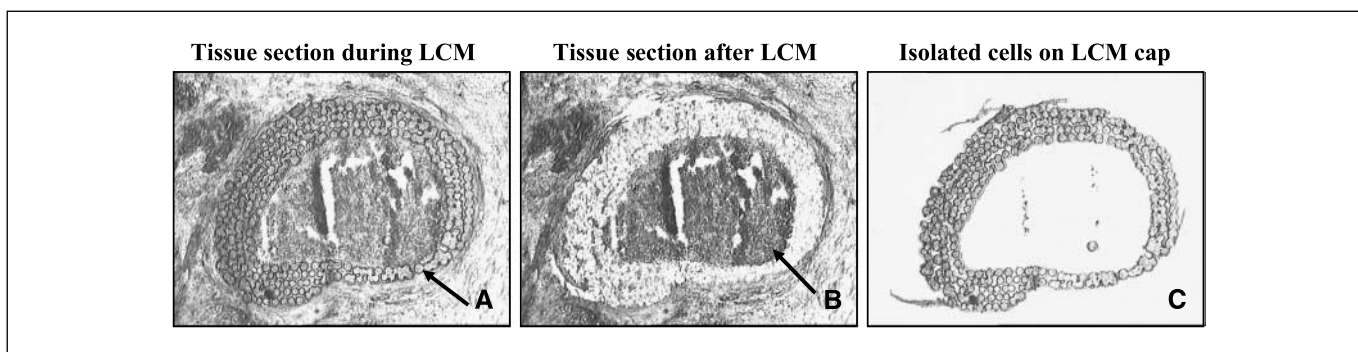
Protocols were established successfully, including extensive sequential quality controls, to enable the hybridization of small sample RNA from microdissected tissue, after a two-round linear amplification, to oligonucleotide microarrays.

**Tissue selection, LCM and RNA extraction.** Twenty breast tumors, which were diagnosed to carry both DCIS and IDC, of histological grades 2 and 3, were selected from the tissue bank of the Department of Gynecology and Obstetrics, Tuebingen, Germany. None of the patients included in this study had received preoperative systemic treatment. Finally, nine specimens with significant DCIS and IDC components were selected for further experiments (Table 1), based on repeated pathologist's evaluations of high-quality H&E-stained sections, and the integrity of its total RNA. All specimens passed through the RNA quality controls successfully, as is exemplified in Supplementary data 2. Twenty-four biological specimens were isolated from the nine tumors (nine matched DCIS/IDC and replicate DCIS/IDC preparations from three of nine tissues; Table 1). A representative example depicting different phases during

the microdissection procedure of DCIS cells is shown in Fig. 1. Only purified epithelial cells adhering to the LCM cap after LCM were processed further (Fig. 1). At least 125 ng of total RNA was retrieved after RNA extraction of all microdissected specimens and checked for high quality (see Supplementary data 2). Exclusively good quality RNA was used for linear RNA amplification.

**RNA amplification and microarray hybridization.** After T7-based two-round linear RNA amplification, typically 40 to 80 μg of amplified aRNA was generated from ~20,000 captured cells. Before hybridization, the quality of aRNA was verified as described in Materials and Methods (data not shown). Gene expression patterns were measured by hybridization of biotinylated aRNA to the Affymetrix GeneChip arrays HG U133A and HG U133 Plus 2.0, respectively. The mean 3'/M ratio retrieved from the high-density arrays for glyceraldehyde-3-phosphate dehydrogenase (HUM-GAPDH/M33197) was 1.61 (range, 1.11-2.79; Table 1). For β-actin (HSAC07/X00351), it was 3.72 (range, 1.76-8.5; Table 1).

**Hierarchical clustering for experiments.** Hierarchical clustering was done using the neighbor-joining method applied to the Pearson correlation distance matrix. In addition, other hierarchical clustering methods, such as UPGMA, and other distance-generating functions, such as Euclidean, were used. Only marginal deviations between the methods and the distance matrices used were observed. We found most stable results for the neighbor-joining



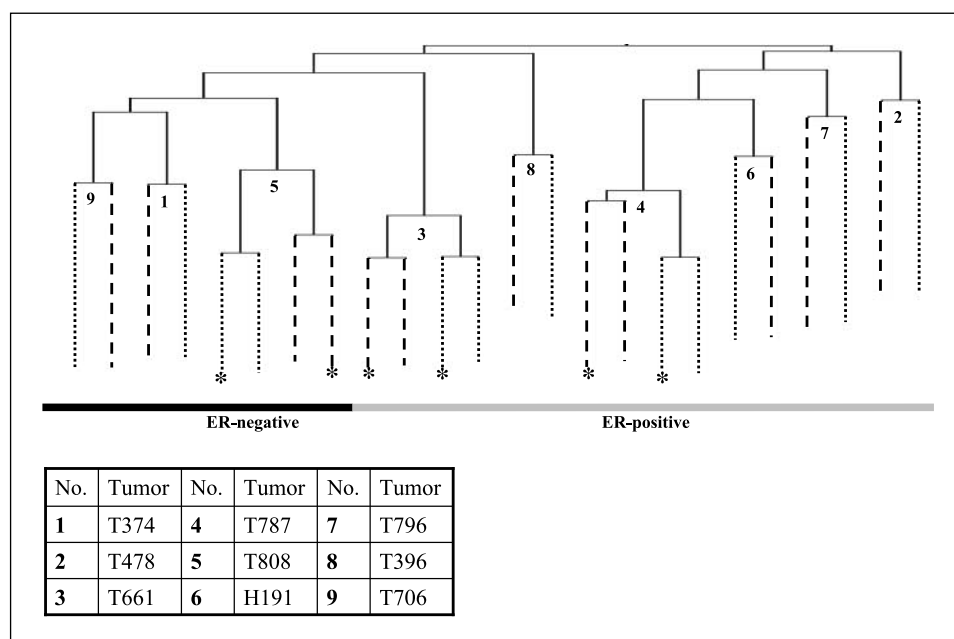
**Figure 1.** LCM from H&E-stained sections of DCIS epithelium. The epithelial cells of DCIS are selectively captured (A) and transferred to the film on the LCM cap (C). Surrounding stromal tissue and necrotic areas in the center of the DCIS remain on the tissue slide (B). A, arrow, laser imprint with captured cells; B, arrow, necrotic cells. LCM for tumor T374.

method applied to the Pearson correlation distance matrix. In Fig. 2, the clustering results for the combined data set is depicted, indicating that different tumor stages do not form distinct groups. Instead, the different synchronous stages of progression within an individual patient cluster more closely to one another than to their respective stage from different patients. Three matched tumor specimens (T661, T787, and T808), microdissected and hybridized independently as technical replicates, cluster to their matched tumor. In between these clusters, the respective tumor stages cluster together, corroborating our approach of combining the gene expression results retrieved from two different Affymetrix microarrays. Additionally, tumors can be stratified into a homogenous ER-negative tumor cluster, separated from distinct subgroups represented by ER-positive tumors.

**Identification of differentially expressed genes.** Independent statistical tests were done on the combined RMA-normalized expression raw data. First, genes that are consistently found most strongly up-regulated or down-regulated in the IDC samples were detected using the rank product method (19). Five hundred and seven probe sets were detected as significantly up-regulated in the IDC samples (false-positive prediction <0.01), and 198 probe sets

were detected as significantly down-regulated in the IDC samples (false-positive prediction <0.01). Next, a two-class paired *t* test was carried out ( $P < 0.01$ ). Differential expression was detected for 1,704 probe sets, of which 973 probe sets are up-regulated in the IDC samples, whereas 731 probe sets are down-regulated in the IDC samples ( $P < 0.01$ ). Third, the nonparametric Mann-Whitney test was applied, using the same parameters as for the *t* test. Differential expression was detected for 1,649 probe sets ( $P < 0.01$ ), of which 936 probe sets are up-regulated in the IDC samples; whereas 713 probe sets are down-regulated in the IDC samples. The respective gene lists of all three methods were combined and the intersection was taken, resulting in the so-called “master gene set,” consisting of 546 significantly regulated probe sets, characterizing the transition from DCIS to IDC. Four hundred and forty-five probe sets (360 genes) are up-regulated, 101 probe sets (85 genes) are down-regulated in IDC (see Supplementary data 1). Performing “Functional Analysis” with the “Ingenuity Pathway Analysis” software, 232 genes could be associated with special functions and diseases. The most significant subgroup of 100 genes (43%) acts in cell-to-cell signaling and interaction (Fischer’s exact test,  $P = 8.02 \times 10^{-11}$  to  $2.77 \times 10^{-7}$ ; see Supplementary data 2).

**Figure 2.** Dendrogram of breast tissue samples based on hierarchical clustering. Depicted is the dendrogram of all 24 breast cancer tissues based on the expression of genes represented on both Affymetrix arrays (HG U133A and HG U133 Plus). Numbers 1 to 9, tumors used (see Table 1). Altogether, nine DCIS, nine matched IDC, and three replicate DCIS/IDC preparations from three of nine breast tumors are shown. Replicate preparations for three tumors (\*), an IDC sample (\*\*\*\*\*), and a DCIS sample (— — —). In the dendrogram, the length of the branches between two elements reflect their degree of relatedness (the shorter, the more similar). Tumors cluster according to the expression of estrogen receptor in two subgroups. Different progression stages within the same tumor cluster together. Replicates (identical biological specimen) cluster to their equivalent.



**Table 2.** Expression validation by real-time PCR

Gene name, + Affymetrix ID / + gene symbol (acronyms) / + representative public ID	Average fold change in IDC vs. DCIS from microarray data ( $\pm$ SD)	Fold change in IDC vs. DCIS (real-time PCR results)			
		T796	T661	T787	T808
Biglycan, 213905_x_at / <i>BGN</i> / AA845258	2.63 $\pm$ 1.61	6.02	2.93	2.62	4.4
MADS box transcription enhancer, factor 2 209200_at / <i>MEF2C</i> / AL536517	2.03 $\pm$ 1.41	1.69	1.79	2.89	1.38
Bullous pemphigoid antigen 1, 204455_at / <i>BPAG1</i> / NM_001723	0.26 $\pm$ 2.92	-1.25	-2.07	-3.49	-12.75
Olfactomedin-like 2B, 213125_at / <i>OLEML2B</i> / AW007573	2.53 $\pm$ 1.87	5.61	2.55	1.44	3.43
Pleckstrin homology domain containing member 1, 209210_s_at / <i>PLEKHCI</i> / Z24725	2.5 $\pm$ 1.54	2.16	2.16	1.83	1.44
Periostin, 210809_s_at / <i>POSTN</i> / D13665	2.19 $\pm$ 1.73	8.89	4.21	1.78	2.29
Fibroblast activation protein $\alpha$ , 209955_s_at / <i>FAP</i> / U76833	2.96 $\pm$ 1.77	8.92	2.78	2.44	n.d.
*, Large tumor suppressor homologue 2, 223380_s_at / <i>LATS2</i> / AF207547	2.34 $\pm$ 1.37	1.12	1.65	1.25	n.d.
Ras-induced senescence 1, 213338_at / <i>RIS1</i> / BF062629	1.8 $\pm$ 1.68	3.65	3.31	3.17	n.d.
Tumor necrosis factor $\alpha$ -induced protein 6, 206026_s_at / <i>TNFAIP6</i> / NM_007115	2.61 $\pm$ 1.78	4.88	4.21	1.65	n.d.
213790_at / <sup>†</sup> / W46291	2.8 $\pm$ 1.90	6.21	2.39	2.54	n.d.
*, 226997_at / <sup>†</sup> / W74476	2.99 $\pm$ 1.59	12	2.92	3.85	n.d.
Dapper homologue, 1219179_at / <i>DACT1</i> / NM_016651	2.74 $\pm$ 1.86	6.75	2.24	n.d.	8.18
Gremlin 1 homologue, 218469_at / <i>GREM1</i> / NM_013372	3.35 $\pm$ 1.93	15.44	2.45	n.d.	2.74
Leucine-rich repeat containing 15, 213909_at / <i>LRRC15</i> / AU147799	3.61 $\pm$ 1.69	100% <sup>‡</sup>	5.01	n.d.	1.88
Squamous cell carcinoma antigen recognized by T cells 2, 218854_at / <i>SART2</i> / NM_fs013352	2.26 $\pm$ 1.51	3.61	1.68	n.d.	1.33

Abbreviation: n.d., "not determined."

\*Probe set only on HG U133 Plus 2.0 array.

<sup>†</sup>No gene symbol.

<sup>‡</sup>No expression in DCIS detected.

**Validation of candidate genes by RT-PCR.** Evaluation of gene expression for two invasion-related candidates, matrix metalloproteinase 11 (*MMP11*) and urokinase-type plasminogen activator (*PLAU*), by classical RT-PCR confirmed their up-regulation in IDC compared to matched DCIS, thus verifying microarray results (data not shown).

For 16 additional differentially regulated candidate genes, real-time PCR was applied in triplicate experiments. In Table 2, relative gene expression values (fold changes) between DCIS and corresponding invasive components are listed. In agreement with the microarray results, real-time PCR shows the up-regulation of pleckstrin homology domain containing member 1 (*PLEKHCI*), MADS box transcription enhancer factor 2 (*MEF2C*), biglycan (*BGN*), periostin (*POSTN*), olfactomedin-like 2B (*OLEML2B*), fibroblast activation protein  $\alpha$  (*FAP*), W46291 (Affymetrix-ID, 213790\_at), ras-induced senescence 1 (*RIS1*), tumor necrosis factor- $\alpha$ -induced protein 6 (*TNFAIP6*), leucine-rich repeat containing 15 (*LRRC15*), gremlin 1 homologue (*GREM1*), squamous cell carcinoma antigen recognized by T cell 2 (*SART2*), and dapper homologue 1 (*DACT1*) and down-regulation of bullous pemphigoid antigen 1 (*BPAG1*) in IDC (see Supplementary data 3). Up-regulation for large tumor suppressor homologue 2 (*LATS2*) and W74476 (Affymetrix-ID, 226997\_at), not included in the list derived from the statistical computations because they are only represented on the HG U133 2.0 Plus array, also corroborate microarray data.

#### Candidate genes are expressed in the epithelial tumor cells.

*In situ* hybridization was done for three genes (*MEF2C*, *SART2*, and *TNFAIP6*) to confirm their cellular specificity. Figure 3 shows a representative *in situ* hybridization example for *TNFAIP6*. The mRNA is present in the cytoplasm of DCIS and IDC, indicating its expression in epithelial tumor cells. *MEF2C* and *SART2* transcripts were also localized to the epithelial breast tumor cells (data not shown). No quantitative differences of gene expression could be observed using *in situ* hybridization.

**Protein expression of candidates resemble mRNA expression.** Protein expression of four candidates *BPAG1*, *DACT1*, *GREM1*, and *MEF2C* verified by RT-PCR, was evaluated by immunohistochemistry (Fig. 4). Again, as expected from the use of LCM and from the results obtained by *in situ* hybridization, signals localize to the epithelial cell compartment. *MEF2C* is expressed in the cytoplasm of both DCIS and IDC with stronger staining in the IDC component of the same section (Fig. 4A). *BPAG1* protein expression is increased in the DCIS component, confirming the RT-PCR and microarray results (Fig. 4B). *DACT1* transcript is up-regulated in IDC and this observation was also reiterated in the protein staining patterns. *DACT1* protein can be detected in the center of DCIS lesions (Fig. 4C). *GREM1* is expressed in the cytoplasm of epithelial cells in DCIS and IDC, respectively (Fig. 4D). Taken together, results from immunohistochemistry indicate the epithelial expression of candidate proteins.

## Discussion

The aim of this study was to gain insight into the molecular biology of DCIS to IDC transition and to provide new candidates which have the potential to become drug or prevention targets. To avoid expression differences based on the genetic background of individual patients, DCIS and IDC were compared in a matched-pair analysis. Additionally, LCM was combined with Affymetrix microarray technology to generate epithelial-specific gene expression profiles.

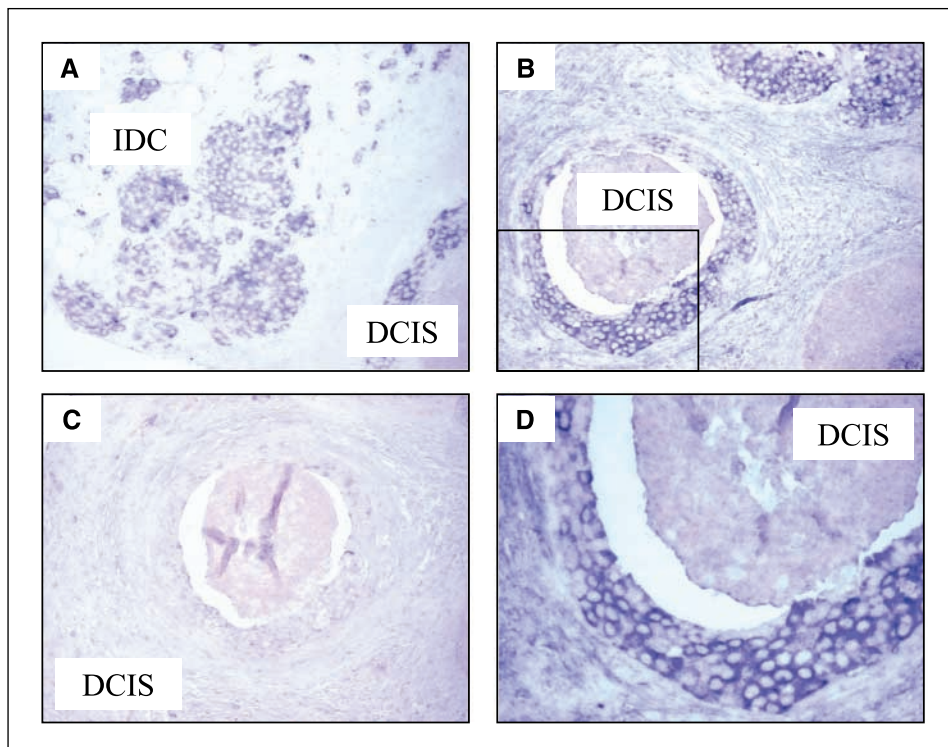
In the first part of the research, the major focus was to establish and standardize protocols to enable microarray analysis of small sample RNA from microdissected tissue. The involvement of the pathologist, documentation of LCM and stepwise RNA quality controls are fundamental attributes of this study. 3'/M ratios, a measure of quality for the labeled aRNAs after two-round amplification (GeneChip eukaryotic small sample target labeling assay, version II, Affymetrix), are in the recommended range for glyceraldehyde-3-phosphate dehydrogenase (3'/M  $\geq$  3). For  $\beta$ -actin, the ratios are slightly increased. Thus, the quality of the hybridized specimens, the comparability of matched-pair samples, and the robustness of the approach were finally determined by hierarchical cluster analysis. Here, different progression stages derived from the same patient cluster together, which was also previously observed (9). Additionally, replicate DCIS and IDC cluster to their equivalent from the same patient. These findings corroborate the robustness and stringency of our established experimental approach. The observation that different progression stages derived from the same patient cluster more closely to one another than the corresponding stages from different patients suggests, that "interpatient" genetic variability is overriding the differential gene expression between DCIS and IDC. These concordant findings emphasize the need to investigate pathological lesions from the same patient to reduce false-positive

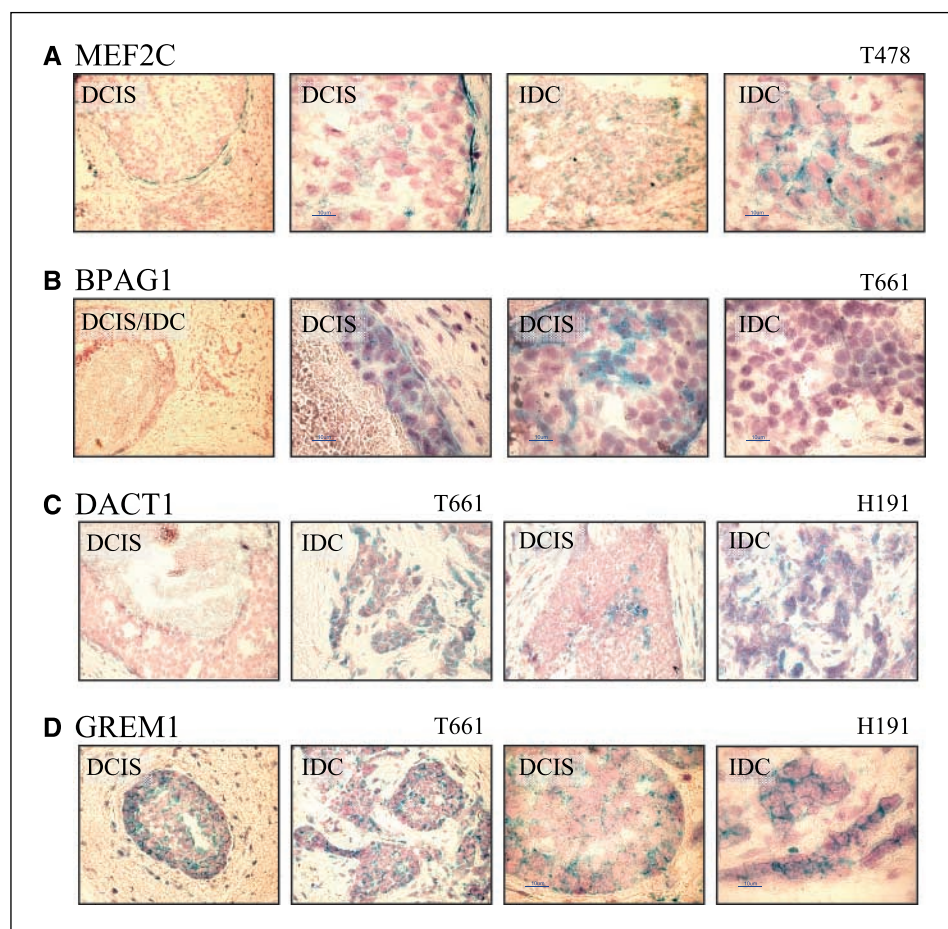
gene expression differences due to genetic variability between individual patients.

Recently, discussions have arisen regarding the cross-platform comparability of gene expression data and regarding reproducibility and the demand for standardization of microarray experiments (23, 24). Therefore, attention was focused on statistical analyses. A "master gene set" was composed by applying three independent statistical tests to increase the chance of identifying a robust gene set. Four hundred and forty-five probe sets (360 genes) were identified as significantly up-regulated in IDC (rank product, false-positive prediction  $<0.01$ ;  $t$  test and Mann-Whitney:  $P < 0.01$ ), making  $\sim 2\%$  of the total amount of probe sets analyzed (22,283 probe sets; HG U133A). In the comparable microarray- and LCM-based study done by Ma et al., 85 genes are significantly up-regulated in IDC ( $\sim 0.1\%$ ,  $P < 0.01$ ; ref. 9). An explanation for the increased number of differentially up-regulated genes in our experiment might be that Ma et al. used cDNA arrays that are more prone to cross-hybridization, and marginal differentially expressed genes might be lost in the background of signals (9, 23). However, these marginal differentially expressed genes may be the most significant or biologically causative (25). Many of the genes represented in our "master gene list" have relatively small but statistically significant fold changes which were validated for 18 candidate genes by RT-PCR.

It is known that comparison of gene expression data obtained from different array platforms results in a marginal overlap (24). Thus, it is not surprising that only four genes from our set are also represented in the gene list of Ma and colleagues: adipocyte enhancer-binding protein 1 (*AEBP1*), syndecan 2 (*SDC2*), chromosome 18 open reading frame 1 (*C18orf1*), and collagen type XV  $\alpha 1$  (*COL15A1*). *AEBP1* interacts with PTEN, a known tumor suppressor in breast cancer, whose inactivation is correlated with invasiveness and metastasis in breast cancer (26, 27).

**Figure 3.** Localization of *TNFAIP6* transcript by *in situ* hybridization. Antisense riboprobe for *TNFAIP6* specifically detects the mRNA in epithelial cells (A, B, and D). Specific signal is represented by a blue precipitate in the cytoplasm (original magnification,  $\times 20$ ; in A to C). C, sense riboprobe for *TNFAIP6*, used as negative control, detects no transcript. D,  $\times 40$  magnification of the insert in (B).





**Figure 4.** Analysis of protein expression of selected candidate progression markers. Expression was determined by immunohistochemistry using tissue sections from cryopreserved tumors used for microarray analysis. Specific signal is represented by a green precipitate. A, protein expression of MEF2C is increased in IDC compared to DCIS from the same section. Representative tumor specimen (*T478*). B, protein expression of BPAG1 is decreased in IDC compared to DCIS from the same section. Representative tumor specimen (*T661*). C, protein expression of DACT1 is increased in IDC compared to DCIS from the same section. Two representative tumor specimens (*T661* and *H191*). D, protein expression of GREM1 is localized to epithelial cells. Two representative tumor specimens (*T661* and *H191*).

We observe a disproportion of up-regulated versus down-regulated probe sets in IDC compared to DCIS (445:101). A possible explanation might be the clonal expansion of cell lineages during tumor progression, thereby increasing the number of mRNA species (28). Furthermore, some genes might be present in IDC samples because of co-isolation of nonepithelial cells in the invasive component during microdissection, although isolation of contaminating cells can be reduced to 0.6% using LCM (11). There might also be a biological explanation for the identification of genes supposed to be “stromal” or “fibroblast”: tumor progression includes processes like epithelial-mesenchymal transition. Therein epithelial cells lose epithelial polarity and acquire a fibroblastoid phenotype, making it a correlative for metastasis (29). Our “master gene set” contains *TWIST1*, which was recently published to be a master regulator promoting epithelial-mesenchymal transition contributing to metastasis in breast cancer (30), and several genes (e.g., *DCN*, *SDC2*, *SPARC*, *MMP13*, *PDGFRB*, *INHBA*, and *DAB2*)<sup>9</sup> represented in a gene list associated with epithelial-mesenchymal transition in mammary epithelial cells (31).

The master gene set also contains genes whose relative up-regulation in IDC was validated by RT-PCR and which are already correlated with invasion: *MMP11/STR-3* (32, 33), *PLAU/uPA* (34), *LRRIC15/Lib* (35), and *BPAG1* (36). In addition, *MMP11* is also a

predictor for poor prognosis (37). *PLAU* is a novel prognostic factor and is validated at the highest level of evidence regarding its clinical utility in breast cancer (34). *LRRIC15* was found to be highly and uniquely expressed in breast tumors compared to their matched normal tissue (38). It is present at the leading edge of migrating cells and siRNA-mediated mRNA suppression in highly invasive Hs467T breast carcinoma cells leads to abrogation of invasiveness in these cells (35). Being a transmembrane protein makes it an easily targetable protein. *BPAG1* is expressed in hemidesmosomes connecting epithelial cells to the basement membrane. Invasive breast cancer cells do not express hemidesmosomes or most of the component proteins (36).

The presence of *MMP11*, *PLAU*, *LRRIC15*, and *BPAG1* in our master gene set suggests that our approach might have detected more transcripts involved but not yet correlated with migration and invasion. Further support is provided by Functional Analysis using Ingenuity software, showing that a significant fraction of our master gene set is involved in cell-to-cell signaling and interaction. This emphasizes the importance of changing intercellular communication during the process of invasion in breast cancer. According to Nagaraja et al., a small set of nine transcripts playing a role in cell-to-cell signaling and interaction is sufficient to distinguish invasive breast cancer cell lines from noninvasive (39). Among this set, annexin A1 (*ANXA1*), cadherin 11 (*CDH11*), and claudin 3 (*CLDN3*) show the same pattern of expression in invasive and noninvasive cell lines as we observe for them in DCIS to IDC transition.

<sup>9</sup> *DCN*, decorin; *SDC2*, syndecan 2; *SPARC*, osteonectin; *MMP13*, matrix metalloproteinase 13; *PDGFRB*, platelet-derived growth factor receptor  $\beta$  polypeptide; *INHBA*, inhibin  $\beta$ A; *DAB2*, disabled homologue 2.

Using *in situ* hybridization and/or immunohistochemistry, several candidates which have thus far not been correlated with breast cancer invasion can be localized to the epithelial tumor cells, a result also corroborating the benefit of LCM for our experimental approach. These genes are *DACT1*, *GREM1*, *MEF2C*, *SART2*, and *TNFAIP6*. *DACT1* is not yet correlated with breast cancer. Its homologue, *DAPPER1* in *Xenopus*, modulates WNT signaling, which is important in the development of breast cancer (40, 41). *GREM1* is a secreted antagonist of bone morphogenic proteins, which are essential in initiating epithelial-mesenchymal signaling (42). Secreted proteins are potential interesting targets for diagnosis and therapy (43). *MEF2C* is a transcriptional enhancer whose biological function in human breast cancer is unknown. However, it was shown that its chromosomal localization is assigned to the mammary cancer susceptibility 1 locus (*Mcs1*) on chromosome 2q1 segregating with the sensitivity to mammary cancer development in rats (44). *TNFAIP6/TSG-6* is not yet correlated with cancer. Published data suggest that it might down-regulate inflammatory response, potentially enabling invasive cells to escape immune response (45). *SART2*, belongs to a protein family encoding antigenic peptides capable of inducing tumor-reactive CTLs in HLA-A24<sup>+</sup>/2<sup>+</sup> patients. Although *SART2* could not be detected in MCF-7 cells by Northern blotting (46), we were able to detect its expression in different breast cell lines and tumor tissue using RT-PCR methods and *in situ* hybridization.

Because *SART2*-derived peptides were used in vaccination studies in patients with lung, colon, and cervical cancers (47), it might also be considered to be applied in breast cancer, based on our results.

In conclusion, we identified progression-specific candidate genes using a quality-controlled approach, combining LCM and microarray analysis. We provided a deeper insight into the molecular biology of DCIS to IDC transition by characterizing the set of differentially regulated genes and single validated candidates. Among these candidates are gene products with potential clinical importance such as *GREM1*, *SART2*, or *LRRC15*.

## Acknowledgments

Received 12/27/2005; revised 2/22/2006; accepted 3/2/2006.

**Grant support:** Deutsche Forschungsgemeinschaft (Graduiertenkollege, 686-1 and 686-2) and fortune program of the University of Tuebingen (1234-0-0). K. Niesel is supported by the Deutsche Forschungsgemeinschaft (AZ BIZ 1/1-3). The Microarray Facility is sponsored by the Interdisciplinary Center of Clinical Research Tuebingen (IZKF) and the Federal Ministry of Education and Research (grant no. 01KS9602). M. Walter is supported by the European Integrated Project on Spinocerebellar Ataxias (EUROSCA, LSHM-CT-2004-503304).

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

The skillful technical assistance of U. Hilcher, B. Kootz, K. Abt, and Y. Wachtel is gratefully acknowledged. We thank Dr. U. Vogel for providing us with paraffin-embedded breast cancer tissue. The polyclonal rabbit anti-human BPAG1 (no. 1133) antibody was a kind gift from Dr. J.R. Stanley and his colleagues.

## References

- Allred DC, Mohsin SK, Fuqua SA. Histological and biological evolution of human premalignant breast disease. *Endocr Relat Cancer* 2001;8:47-61.
- O'Connell P, Pekkel V, Fuqua SA, Osborne CK, Clark GM, Allred DC. Analysis of loss of heterozygosity in 399 premalignant breast lesions at 15 genetic loci. *J Natl Cancer Inst* 1998;90:697-703.
- Porter D, Lahti-Domenici J, Keshaviah A, et al. Molecular markers in ductal carcinoma *in situ* of the breast. *Mol Cancer Res* 2003;1:362-75.
- Lakhani SR, Ashworth A. Microarray and histopathological analysis of tumours: the future and the past? *Nat Rev Cancer* 2001;1:151-7.
- Gradishar WJ. The future of breast cancer: the role of prognostic factors. *Breast Cancer Res Treat* 2005;89 Suppl 1:S17-26.
- Lockhart DJ, Dong H, Byrne MC, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 1996;14:1675-80.
- Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869-74.
- van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530-6.
- Ma XJ, Salunga R, Tuggle JT, et al. Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci U S A* 2003;100:5974-9.
- Luzzi V, Holtschlag V, Watson MA. Expression profiling of ductal carcinoma *in situ* by laser capture microdissection and high-density oligonucleotide arrays. *Am J Pathol* 2001;158:2005-10.
- Nishidate T, Katagiri T, Lin ML, et al. Genome-wide gene-expression profiles of breast-cancer cells purified with laser microbeam microdissection: identification of genes associated with progression and metastasis. *Int J Oncol* 2004;25:797-819.
- Tolosa E, Li W, Yasuda Y, et al. Cathepsin V is involved in the degradation of invariant chain in human thymus and is overexpressed in myasthenia gravis. *J Clin Invest* 2003;112:517-26.
- Weinschenk T, Gouttefangeas C, Schirle M, et al. Integrated functional genomics approach for the design of patient-individual antitumor vaccines. *Cancer Res* 2002;62:5818-27.
- Wulfkuehle JD, Sgroi DC, Krutzsch H, et al. Proteomics of human breast ductal carcinoma *in situ*. *Cancer Res* 2002;62:6740-9.
- DiLella Anthony G. Successful isolation and RT-PCR of RNA from small numbers of laser capture microdissected cells. *Qiagen News* 2001;4:3-5.
- Baugh LR, Hill AA, Brown EL, Hunter CP. Quantitative analysis of mRNA amplification by *in vitro* transcription. *Nucleic Acids Res* 2001;29:e29.
- Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4: 249-64.
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406-25.
- Breitling R, Armengaud P, Amtmann A, Herzyk P. Rank Products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett* 2004;573:83-92.
- Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 2001;29:e45.
- Ramakera C, Ruijter JM, Deprez RH, Moorman AF. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* 2003;339:62-6.
- Herrlinger U, Klingel K, Meyermann R, et al. Central nervous system Hodgkin's lymphoma without systemic manifestation: case report and review of the literature. *Acta Neuropathol (Berl)* 2000;99:709-14.
- Jarvinen AK, Hautaniemi S, Edgren H, et al. Are data from different gene expression microarray platforms comparable? *Genomics* 2004;83:1164-8.
- Ein-Dor L, Kela I, Getz G, Givol D, Domany E. Outcome signature genes in breast cancer: is there a unique set? [Epub 2004 Aug 12] *Bioinformatics* 2005;21: 171-8.
- Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 2003;34:267-73.
- Chung MJ, Jung SH, Lee BJ, Kang MJ, Lee DG. Inactivation of the PTEN gene protein product is associated with the invasiveness and metastasis, but not angiogenesis, of breast cancer. *Pathol Int* 2004;54: 10-5.
- Gorbenko O, Kuznetsov V, Kukharensko O, et al. Identification of a novel binding partners for tumor suppressor PTEN by a yeast two-hybrid approach. *Eksp Onkol* 2004;26:15-9.
- Lattimore BS, Crabbe MJ. Expression profiles in the progression of ductal carcinoma in the breast. *Comput Biol Chem* 2003;27:115-20.
- Grunert S, Jechlinger M, Beug H. Diverse cellular and molecular mechanisms contribute to epithelial plasticity and metastasis. *Nat Rev Mol Cell Biol* 2003;4:657-65.
- Yang J, Mani SA, Donaher JL, et al. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* 2004;117:927-39.
- Jechlinger M, Grunert S, Tamir IH, et al. Expression profiling of epithelial plasticity in tumor progression. *Oncogene* 2003;22:7155-69.
- Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer [Epub 2004 Dec 10]. *N Engl J Med* 2004; 351:2817-26.
- Andarawewa KL, Motrescu ER, Chenard MP, et al. Stromelysin-3 is a potent negative regulator of adipogenesis participating to cancer cell-adipocyte interaction/crosstalk at the tumor invasive front. *Cancer Res* 2005;65:10862-71.
- Harbeck N, Kates RE, Gauger K, et al. Urokinase-type plasminogen activator (uPA) and its inhibitor PAI-1: novel tumor-derived factors with a high prognostic and predictive impact in breast cancer. *Thromb Haemost* 2004;91:450-6.
- Reynolds PA, Smolen GA, Palmer RE, et al. Identification of a DNA-binding site and transcriptional target for the EWS-WT1(+KTS) oncoprotein. *Genes Dev* 2003; 17:2094-107.
- Bergstraesser LM, Srinivasan G, Jones JC, Stahl S, Weitzman SA. Expression of hemidesmosomes and component proteins is lost by invasive breast cancer cells. *Am J Pathol* 1995;147:1823-39.
- Tetu B, Brisson J, Wang CS, Lapointe H, Beaudry G, Blanchette C. Expression of cathepsin D, stromelysin-3,



- and urokinase by reactive stromal cells on breast carcinoma prognosis. *Cancer* 2001;92:2957-64.
38. Satoh K, Hata M, Yokota H. High lib mRNA expression in breast carcinomas. *DNA Res* 2004;11:199-203.
39. Nagaraja GM, Othman M, Fox BP, et al. Gene expression signatures and biomarkers of noninvasive and invasive breast cancer cells: comprehensive profiles by representational difference analysis, microarrays and proteomics. *Oncogene* 2005. Available from: <http://dx.doi.org/10.1038/sj.onc.1209265>.
40. Brott BK, Sokol SY. Frodo proteins: modulators of Wnt signaling in vertebrate development. *Differentiation* 2005;73:323-9.
41. Brown AM. Wnt signaling in breast cancer: have we come full circle? [Epub 2001 Sep 6] *Breast Cancer Res* 2001;3:351-5.
42. Hsu DR, Economides AN, Wang X, Eimon PM, Harland RM. The *Xenopus* dorsalizing factor Grem-1 identifies a novel family of secreted proteins that antagonize BMP activities. *Mol Cell* 1998;1:673-83.
43. Stitzel NO, Mar BG, Liang J, Westbrook CA. Membrane-associated and secreted genes in breast cancer. *Cancer Res* 2004;64:8682-7.
44. Laes JF, Quan X, Ravoet M, et al. Analysis of candidate genes included in the mammary cancer susceptibility 1 (Mcs1) region. *Mamm Genome* 2001;12:199-206.
45. Milner CM, Day AJ. TSG-6: a multifunctional protein associated with inflammation. *J Cell Sci* 2003;116:1863-73.
46. Nakao M, Shichijo S, Imaizumi T, et al. Identification of a gene coding for a new squamous cell carcinoma antigen recognized by the CTL. *J Immunol* 2000;164:2565-74.
47. Mochizuki K, Sato Y, Tsuda N, et al. Immunological evaluation of vaccination with pre-designated peptides frequently selected as vaccine candidates in an individualized peptide vaccination regimen. *Int J Oncol* 2004;25:121-31.