Preinvasive ductal carcinoma in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis.

Christina S. Schuetz, Michael Bonin, Susan E. Clare, Kay Nieselt, Karl Sotlar, Michael Walter, Tanja Fehm, Erich Solomayer, Olaf Riess, Diethelm Wallwiener, Rafael Kurek, and Hans J. Neubauer.

1Department of Obstetrics and Gynecology, 2Microarray Facility, Department of Medical Genetics, University of Tuebingen, Tuebingen, Germany; 3Department of Surgery, Indiana University School of Medicine, Indianapolis, Indiana; and 4Center of Bioinformatics, and 5Department 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 BPA1, LRRCL, MPPI1, and PLAU. The expression of BPA1, DACT1, GREMI, MEF2C, SART2, and TNFAP6 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 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 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), based on clinical and histological information. 6

---

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

Requests for reprints: Hans J. Neubauer, Tumor Progression Lab, Gynecological Hospital, Calverstraße 7/6, 72076 Tuebingen, Germany. Phone: 49-7071-298-5388; Fax: 49-7071-29-4947; E-mail: hans.neubauer@med.uni-tuebingen.de.

©2006 American Association for Cancer Research. doi:10.1158/0008-5472.CAN-05-4610

---

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

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

**Laser capture microdissection.** Serial frozen tissue sections (10 μm) 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 II 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 μm diameter laser) were dissected. LCM caps containing the isolated tumor cells were stored at −80°C in 40 μL RLT buffer (Qiagen, Hilden, Germany) + 1% β-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 μL RLT buffer (Qiagen) + 1% β-mercaptoethanol. Total RNA extraction was carried out using RNaseasy 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 μL 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 μg 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 μm 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 GSE8893.

**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.6

**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 μg 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 β-subunit (PDH) was selected 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, 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-μm 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 protease 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/μL) was hybridized. Hybridizations were considered successful if the sense probe gave no significant signal.

**Immunohistochemistry.** Cryosections (10 μm) 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 pertex (Medite, Burgdorf, Germany). The following primary antibodies were used: polyclonal rabbit anti-human DACT1 antibody (Orbigen, San Diego, CA), polyclonal goat anti-human
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 $\mu$g 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 (HUMGAPDH/M33197) was 1.61 (range, 1.11-2.79; Table 1). For $\beta$-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 approach.

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

<table>
<thead>
<tr>
<th>Tumor ID</th>
<th>Tumor characteristics</th>
<th>Hybridization characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>Specimen</td>
<td>Grade</td>
</tr>
<tr>
<td>1.1</td>
<td>T374 DCIS</td>
<td>G2-3</td>
</tr>
<tr>
<td>1.2</td>
<td>T374 IDC</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>T478 DCIS</td>
<td>G2-3</td>
</tr>
<tr>
<td>2.2</td>
<td>T478 IDC</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>T661 DCIS</td>
<td>G3</td>
</tr>
<tr>
<td>3.2</td>
<td>T661 IDC</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>T787 DCIS</td>
<td>G2</td>
</tr>
<tr>
<td>4.2</td>
<td>T787 IDC</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>T808 DCIS</td>
<td>G2</td>
</tr>
<tr>
<td>5.2</td>
<td>T808 IDC</td>
<td></td>
</tr>
<tr>
<td>6.1</td>
<td>H191 DCIS</td>
<td>G2</td>
</tr>
<tr>
<td>6.2</td>
<td>H191 IDC</td>
<td></td>
</tr>
<tr>
<td>7.1</td>
<td>T796 DCIS</td>
<td>G2</td>
</tr>
<tr>
<td>7.2</td>
<td>T796 IDC</td>
<td></td>
</tr>
<tr>
<td>8.1</td>
<td>T396 DCIS</td>
<td>G2</td>
</tr>
<tr>
<td>8.2</td>
<td>T396 IDC</td>
<td></td>
</tr>
<tr>
<td>9.1</td>
<td>T706 DCIS</td>
<td>G2-3</td>
</tr>
<tr>
<td>9.2</td>
<td>T706 IDC</td>
<td></td>
</tr>
</tbody>
</table>

*Immunoreactive score.
†Glyceraldehyde 3-phosphate dehydrogenase.
‡$\beta$-actin.
§Technical replicate.
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).
Table 2. Expression validation by real-time PCR

<table>
<thead>
<tr>
<th>Gene name, + Affymetrix ID / + gene symbol (acronyms) / + representative public ID</th>
<th>Average fold change in IDC vs. DCIS from microarray data (±SD)</th>
<th>Fold change in IDC vs. DCIS (real-time PCR results)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biglycan, 213905_x_at / BGN / AA845258</td>
<td>2.63 ± 1.61</td>
<td>T796</td>
</tr>
<tr>
<td>MADS box transcription enhancer, factor 2 / 209200_at / MEF2C / AL536517</td>
<td>2.03 ± 1.41</td>
<td>T661</td>
</tr>
<tr>
<td>Bullous pemphigoid antigen 1, 204455_at / BPAG1 / NM_001723</td>
<td>0.26 ± 2.92</td>
<td>T787</td>
</tr>
<tr>
<td>Olfactomedin-like 2B, 210809_s_at / OLFML2B / AW007573</td>
<td>2.53 ± 1.87</td>
<td>T808</td>
</tr>
<tr>
<td>Plekstrin homology domain containing member 1, 209210_s_at / PLEKHC1 / Z24725</td>
<td>2.5 ± 1.54</td>
<td>6.02</td>
</tr>
<tr>
<td>Periostin, 210809_s_at / POSTN / D13665</td>
<td>2.19 ± 1.73</td>
<td>2.93</td>
</tr>
<tr>
<td>Fibroblast activation protein α, 209955_s_at / FAP / U76833</td>
<td>2.96 ± 1.77</td>
<td>1.69</td>
</tr>
<tr>
<td>Large tumor suppressor homologue 2, 223380_s_at / RIS1 / AF207547</td>
<td>2.34 ± 1.37</td>
<td>1.49</td>
</tr>
<tr>
<td>Ras-induced senescence 1, 213338_at / DACT1 / NM_007115</td>
<td>1.8 ± 1.68</td>
<td>5.61</td>
</tr>
<tr>
<td>Tumor necrosis factor α-induced protein 6, 206026_s_at / TNFAIP6 / NM_007115</td>
<td>2.61 ± 1.78</td>
<td>5.11</td>
</tr>
<tr>
<td>213790_at / W46291</td>
<td>2.8 ± 1.90</td>
<td>2.16</td>
</tr>
<tr>
<td>Dapper homologue, 1219179_at / DACT1 / NM_016651</td>
<td>2.99 ± 1.59</td>
<td>2.16</td>
</tr>
<tr>
<td>Gremlin 1 homologue, 218854_at / SART2 / BF062629</td>
<td>1.78 ± 4.88</td>
<td>5.61</td>
</tr>
<tr>
<td>Squamous cell carcinoma antigen recognized by T cells 2, 218854_at / SART2 / NM_007115</td>
<td>2.26 ± 1.51</td>
<td>4.21</td>
</tr>
</tbody>
</table>

Abbreviation: n.d., “not determined.”
*Probe set only on HG U133 Plus 2.0 array.
†No gene symbol.
*No expression in DCIS detected.

Validation of candidate genes by RT-PCR. Evaluation of gene expression 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 plekstrin homology domain containing member 1 (PLEKHC1), MADS box transcription enhancer factor 2 (MEF2C), bullous pemphigoid (BGN), perinostin (POSTN), olfactomedin-like 2B (OLFML2B), fibroblast activation protein α (FAP), W46291 (Affymetrix-ID, 213790_at), ras-induced senescence 1 (RI1), tumor necrosis factor-α-induced protein 6 (TNFAIP6), leucine-rich repeat containing 15 (LRRC15), W46291 (Affymetrix-ID, 213790_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 ≥ 3). For β-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 ~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 (~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 α1 (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, ×20; in A to C). C, sense riboprobe for TNFAIP6, used as negative control, detects no transcript. D, ×40 magnification of the insert in (B).
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)\(^9\) 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), LRRC15/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). LRRC15 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, LRRC15, 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.

---

\(^9\) DCN, decorin; SDC2, syndecan 2; SPARC, osteonectin; MMP13, matrix metalloenzyme 13; PDGFRB, platelet-derived growth factor receptor β polypeptide; INHBA, inhibin βA; DAB2, disabled homologue 2.
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 (Graduiertenkolleg, 686-1 and 686-2) and fortune program of the University of Tuebingen (1234-0-0). K. Nieselt is supported by the Deutsche Forschungsgemeinschaft (AZ BIZ 1.1-3). The Microarray Facility is supported by the Interdisciplinary Center of Clinical Research Tuebingen (EUKF) and the Federal Ministry of Education and Research (grant no. 01K59602). M. Walter is supported by the European Integrated Project on Spinos cerebellar Ataxias (EUROSICA, LSHM-CT-2004-503349).

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. Koozt, 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


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, Michael Bonin, Susan E. Clare, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/66/10/5278

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2006/07/14/66.10.5278.DC1

Cited articles
This article cites 45 articles, 18 of which you can access for free at:
http://cancerres.aacrjournals.org/content/66/10/5278.full.html#ref-list-1

Citing articles
This article has been cited by 30 HighWire-hosted articles. Access the articles at:
/content/66/10/5278.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.