



Published in final edited form as:

*Mol Cancer Res.* 2007 September ; 5(9): 881–890. doi:10.1158/1541-7786.MCR-07-0055.

## Breast Cancer Molecular Signatures as Determined by SAGE: Correlation with Lymph Node Status

Martín C. Abba<sup>1</sup>, Hongxia Sun<sup>1</sup>, Kathleen A. Hawkins<sup>1</sup>, Jeffrey A. Drake<sup>1</sup>, Yuhui Hu<sup>1</sup>, Maria I. Nunez<sup>1</sup>, Sally Gaddis<sup>1</sup>, Tao Shi<sup>2</sup>, Steve Horvath<sup>3</sup>, Aysegul Sahin<sup>4</sup>, and C. Marcelo Aldaz<sup>1</sup>

<sup>1</sup>Department of Carcinogenesis, The University of Texas M. D. Anderson Cancer Center, Science Park-Research Division, Smithville, Texas

<sup>2</sup>Ortho-Clinical Diagnostics, San Diego, California

<sup>3</sup>Human Genetics and Biostatistics, David Geffen School of Medicine, University of California, Los Angeles, California

<sup>4</sup>Department of Pathology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas

### Abstract

Global gene expression measured by DNA microarray platforms have been extensively used to classify breast carcinomas correlating with clinical characteristics, including outcome. We generated a breast cancer Serial Analysis of Gene Expression (SAGE) high-resolution database of ~2.7 million tags to perform unsupervised statistical analyses to obtain the molecular classification of breast-invasive ductal carcinomas in correlation with clinicopathologic features. Unsupervised statistical analysis by means of a random forest approach identified two main clusters of breast carcinomas, which differed in their lymph node status ( $P = 0.01$ ); this suggested that lymph node status leads to globally distinct expression profiles. A total of 245 (55 up-modulated and 190 down-modulated) transcripts were differentially expressed between lymph node (+) and lymph node (–) primary breast tumors (fold change,  $\geq 2$ ;  $P < 0.05$ ). Various lymph node (+) up-modulated transcripts were validated in independent sets of human breast tumors by means of real-time reverse transcription-PCR (RT-PCR). We validated significant overexpression of transcripts for *HOXC10* ( $P = 0.001$ ), *TPD52L1* ( $P = 0.007$ ), *ZFP36L1* ( $P = 0.011$ ), *PLINP1* ( $P = 0.013$ ), *DCTN3* ( $P = 0.025$ ), *DEK* ( $P = 0.031$ ), and *CSNK1D* ( $P = 0.04$ ) in lymph node (+) breast carcinomas. Moreover, the *DCTN3* ( $P = 0.022$ ) and *RHBDD2* ( $P = 0.002$ ) transcripts were confirmed to be overexpressed in tumors that recurred within 6 years of follow-up by real-time RT-PCR. In addition, meta-analysis was used to compare SAGE data associated with lymph node (+) status with publicly available breast cancer DNA microarray data sets. We have generated evidence indicating that the pattern of gene expression in primary breast cancers at the time of surgical

Copyright © 2007 American Association for Cancer Research.

**Requests for reprints:** C. Marcelo Aldaz, Department of Carcinogenesis, The University of Texas M. D. Anderson Cancer Center, Science Park-Research Division, P.O. Box 389, Smithville, TX 78957. Phone: 512-237-2403; Fax: 512-237-2475. maaldaz@mdanderson.org.

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

removal could discriminate those tumors with lymph node metastatic involvement using SAGE to identify specific transcripts that behave as predictors of recurrence as well.

---

## Introduction

Although breast cancer is the most common malignancy in women, the biology of breast cancer remains poorly understood mainly due to the characteristic cellular and molecular heterogeneity of breast tumors. Global gene expression profiling is providing novel information of biological and clinical relevance for the classification of breast cancers.

By means of DNA microarray analyses, various laboratories identified gene expression patterns that correlated with breast cancer patient prognosis (1–9). In spite of the described progress in molecular oncology, invasion into axillary lymph nodes and steroid hormone receptors status still remain as the most reliable prognostic factor for breast cancer patients (10).

Development of metastases (local and distant) requires that a cancer cell must complete a series of steps involving complex interactions with the host microenvironment. This process involves the dysregulation of multiple genes and transcriptional programs. The primary goal of this study was to identify gene expression signatures of relevance for breast cancer subclassification and prognosis. We analyzed a high-resolution Serial Analysis of Gene Expression (SAGE) database obtained from 27 breast-invasive ductal carcinomas (IDCA). A random forest (RF) clustering approach was used for SAGE data analysis (11, 12). This unsupervised analysis of gene expression profiles grouped breast carcinomas predominantly according to their lymph node status. This suggests that lymph node status leads to globally distinct breast cancer gene expression profiles.

The identification of gene expression profiles, individual biomarkers, and biological pathways that contribute to the development of lymph node metastases will be of significant benefit to improve tumor classification and may, in the future, influence clinical decision making and the development of targeted therapies.

## Results and Discussion

### Generation and Analysis of SAGE Libraries

Breast cancer phenotypic and genetic heterogeneity corresponds to heterogeneity of gene expression profiles. SAGE data were obtained from a set of invasive breast carcinomas at a resolution of 100,000 tags per library. Thus, a breast cancer SAGE high-resolution database of almost 2.7 million tags was generated and analyzed, monitoring the expression behavior of more than 30,000 transcripts.

An unsupervised clustering method (RF clustering) allowed us to group the invasive breast carcinomas on the basis of their gene expression pattern. Two dominant clusters were identified (Fig. 1A). To further elucidate the reasons driving the separation of breast carcinomas in two major groups, we analyzed the identified clusters in the light of available histopathologic data (see Table 1). Interestingly, the variable that correlated with the RF

clustering results was the lymph node status of tumors ( $P = 0.01$ ). A total of 7 of 9 breast cancers (78%) in the cluster A are lymph node (+), and 14 of 18 breast tumors (87%) in cluster B are lymph node (-) IDCA (Fig. 1A). Nonstatistically significant differences were detected for ER status, histologic grade, and tumor size ( $P > 0.05$ ). In contrast with results from previous gene expression studies, in which ER was the major discriminator between breast cancer groups, in our case, we interpret that the lack of spontaneous association between clusters and ER status in this subset of samples is likely due to that ~75% of the SAGE libraries generated derived from ER $\alpha$ (+) stages I and II primary breast carcinomas.

To identify the most representative differentially expressed transcripts between tumor groups, we employed a statistically supervised method previously described by us as a modified  $t$  test (13). This analysis revealed 245 genes differentially expressed ( $P < 0.05$ ) between lymph node (+) and lymph node (-) IDCA (Fig. 1B; Supplementary Data File 1). Among the 245 transcripts, 55 were up-modulated, and 190 were down-modulated transcripts in lymph node (+) tumors.

We used the Expression Analysis Systematic Explorer software (EASE) to annotate the 245 deregulated genes according to the information provided by the GO Consortium (14, 15). We observed that 32% of the transcripts are involved in biological processes related to metabolism, 22% are related to cellular physiologic process, and 14% are related to cell communication. Approximately 25% of these dysregulated genes are related to molecular functions associated with nucleic acid/protein binding, 15% are related to hydrolase/transferase activity, and 4% are related to metal ion-binding functions.

### Cross-Platform Gene Expression Profile Comparison

Comparing data sets generated on different gene expression platforms increases the confidence of specific gene expression classifier data sets (16). By performing a meta-analysis from publicly available breast cancer microarray studies, we provide a robust cross-platform validation of 55 up-regulated and 55 down-regulated (fold change,  $>3$ ) lymph node (+)-associated transcripts (Fig. 1C and D). Meta-analysis showed that 42% of the transcripts identified by SAGE (46 out of 110) were confirmed as having statistically significant up- or down-modulation in relation to lymph node (+) status (9 genes), distal metastasis (26 genes), and relapse (29 genes; Table 2, Supplementary Data Files 2 and 3). The lack of 100% overlap of findings between the various studies including ours is not surprising when it is considered that these studies have been done with different technologies (cDNA or various oligonucleotide microarrays), different number of genes in the various fixed platforms, different and heterogeneous patient populations (with regard to age, tumor staging, hormone receptor status, and treatment). Nevertheless, we show that a significant proportion of lymph node (+)-associated transcripts detected by our SAGE study behave as poor prognostic markers. More importantly, SAGE, an open gene expression platform, also identified novel sets of genes as highly expressed in lymph node (+) primary breast carcinomas not previously reported by others.

## Real-time Reverse Transcription-PCR Validation of Lymph Node (+)-Associated Transcripts

The most commonly dysregulated transcripts between lymph node (+) and lymph node (-) breast IDCA as determined by SAGE are represented in Table 3 (fold change, >3;  $P < 0.01$ ). To validate our findings, an independent set of 40 breast IDCA was analyzed by means of real-time reverse transcription-PCR (RT-PCR). In agreement with the SAGE data, we detected statistical differences in the overexpression of seven out of eight evaluated transcripts in lymph node (+) breast tumors, including *homeobox protein hox-c10 (HOXC10)*;  $P = 0.001$ , *tumor protein D52 like-1 (TPD52L1)*;  $P = 0.007$ , *zinc finger protein 36 like-1 (ZFP36L1)*;  $P = 0.011$ , *p53-responsive gene 6 (PLINP1)*;  $P = 0.013$ , *dynactin 3 (DCTN3)*;  $P = 0.025$ , *dek oncogene (DEK)*;  $P = 0.031$ , *casein kinase 1 $\delta$  (CSNK1D)*;  $P = 0.04$ ; Fig. 2A). A trend of borderline significance was detected for the *rhomboid domain containing 2 (RHBDD2)*;  $P = 0.069$ ; Fig. 2A). Hierarchical clustering analysis of the validated transcripts successfully classified tumors according to patient's lymph node status ( $P < 0.05$ ), distinguishing the lymph node (+) from the lymph node (-) breast carcinomas with an accuracy of 89.5% (2 out of 19 lymph node-positive samples misclassified; Fig. 3A). Nonstatistically significant associations were detected between the expressions of these transcripts and ER $\alpha$  status ( $P > 0.05$ ).

## Tumor Protein- and Transcription Factor-Related Genes

The *TPD52L1* gene encodes a member of the tumor protein D52 family. This protein contains a coiled-coil motif required for homo- and heteromeric interactions with other D52-like proteins (17). The *TPD52* gene was first identified as overexpressed in human breast carcinomas (17, 18). Subsequent studies also indicated that these genes are overexpressed in multiple human cancers such as lung, prostate, ovarian, endometrial, and hepatocellular carcinomas (18). *TPD52L1* was reported to be involved in cell proliferation and calcium signaling. It also interacts with the mitogen-activated protein kinase 5 (19).

*DEK* was originally described as a proto-oncogene and has been implicated in multiple cellular processes, including transcriptional regulation and chromatin remodeling (20). Transcriptional up-regulation of wild-type DEK was discovered in various tumor types, including myeloid leukemia, brain tumors, and hepatocellular carcinoma (21, 22). In addition, DEK overexpression was associated with a number of clinical autoimmune conditions (23, 24). Recently, it has been suggested that DEK up-regulation may be a common event in human carcinogenesis and may reflect its senescence inhibitory function (25). Despite these associations with several human disorders, little is known about how DEK could functionally be involved in these diseases (24).

*HOXC10* is one of the highly conserved HOXC family members of transcription factors that play an important role in morphogenesis, cell differentiation, and proliferation (26–28). The HOXC protein levels are controlled during cell differentiation and proliferation. Dysregulation of a variety of HOXC genes has been implicated in several human cancers, including leukemias, colorectal, breast, and renal carcinomas, melanomas, and squamous cell carcinomas of the skin (26, 27). Recently, it was shown that the overexpression of *HOXC4*, *HOXC5*, *HOXC6*, and *HOXC8* genes in malignant cell lines and prostate

carcinomas with lymph node metastases (29). In agreement with these data, we validated the overexpression of *HOXC10* gene in primary lymph node (+) breast carcinomas by real-time RT-PCR ( $P = 0.001$ ; Fig. 2A).

*ZFP36L1* also known as *C3H type-like 1* is a member of the 12-*O*-tetradecanoylphorbol-13-acetate (TPA)-inducible sequence 11 (TIS11) family of early-response genes. The encoded protein contains a zinc finger domain with a repeating cys-his motif (30). TIS11 gene expression is induced rapidly and transiently in response to extracellular hormone and growth factor signals. The potential role of this gene in breast carcinogenesis remains unknown.

### DCTN3 and RHBDD2 as Predictors of Recurrence

As mentioned, the quantitative RT-PCR analysis validated significant differences between lymph node (+) versus lymph node (-) primary breast carcinoma groups for *DCTN3* ( $P = 0.025$ ), and a trend was detected for *RHBDD2* ( $P = 0.069$ ; Fig. 2A). However, meta-analysis comparisons further confirmed our findings showing statistically significant over-expression of *DCTN3* ( $P = 0.003$ ) and *RHBDD2* ( $P = 0.042$ ) in lymph node (+) compared with lymph node (-) breast IDCA (Fig. 1C, Supplementary Data File 3). More importantly, the *DCTN3* ( $P = 0.022$ ) and *RHBDD2* ( $P = 0.002$ ) transcripts were also observed to be markedly up-modulated in tumors that recurred within 6 years of follow-up (Fig. 2B). Unsupervised hierarchical clustering analysis of these transcripts successfully classified tumors according to recurrence status ( $P < 0.05$ ; Fig. 3B). These data suggest that overexpression of *DCTN3* and *RHBDD2* genes could play a role in breast cancer progression.

The *DCTN3* gene (also known as *DCTN22*) encodes the smallest (p22/24) subunit of dynactin, a cytoplasmic motor protein complex involved in organelle trafficking, cytokinesis, spindle formation, chromosome movement, and nuclear positioning (31). Overexpression in mammalian cells of one dynactin subunit (dynamitin) disrupts the complex, resulting in the perturbation of mitosis (32). In addition, *DCTN2* over-expression disrupt the dynein-dynactin motor, shifting cellular movement and mitosis with predisposition to mitotic block and polyploidy (33). *DCTN3* localizes to the centrosomes during interphase and to kinetochores and spindle poles throughout mitosis. It was also proposed that the dynein-dynactin complex is involved in cytoplasmic/nuclear transport of p53 (34). The correct balance of dynactin subunits is important for adequate centrosome integrity before centrosome duplication, ultimately governing the G<sub>1</sub>-S transition.

The *RHBDD2* gene (*rhomboid domain containing 2*) encodes a protein that spans seven-transmembrane domains and is a member of the rhomboid veinlet-like family of genes. Several rhomboid protein members in *Drosophila* have been implicated in the processing of transforming growth factor- $\alpha$  (TGF- $\alpha$ )-like ligands, and consequent epidermal growth factor (EGF) receptor activation (35). Genetic and molecular studies have revealed that the production of an activated EGF ligand by the signal-sending cell is a key regulatory step in receptor activation (36). Thus, the RHBDD2 protein very likely functions in regulating the response to growth factors. However, the potential role of this protein in breast carcinogenesis remains to be elucidated.

## Tissue Microarray Immunohistochemical Analysis of DCTN3 Protein Expression

Because DCTN3 was identified by real-time RT-PCR as distinctively overexpressed in lymph node (+) primary breast carcinomas and in IDCA that recurred within 6 years, we decided to investigate further this gene at protein expression level using a breast cancer progression tissue microarray (Fig. 4).

DCTN3 tissue microarray immunohistochemical (TMA-IHC) analysis showed undetectable expression in 72% (13 out of 18) of the normal breast epithelial samples analyzed, whereas strong immunoreactivity for DCTN3 protein was detected in 60% (6 out of 10) of invasive ductal (IDC) and metastatic breast carcinoma tissues analyzed ( $P$  trend = 0.001; Table 4). In all positive cases, the DCTN3 immunostaining had a homogeneous and diffuse pattern that was localized to the cytoplasm. When DCTN3 expression was correlated with lymph node status, we determined that 75% (6 out of 8) of lymph node (+) carcinoma showed strong DCTN3 staining, whereas 67% (12 out of 18) of lymph node (–) breast carcinomas showed negative immunostaining ( $P$  = 0.027; Table 4). These data plus the aforementioned evidence strongly suggest a putative role for DCTN3 mRNA/protein expression and axillary lymph node metastasis and breast cancer progression.

## Conclusions

The genes that we identified and validated seem to be involved in signaling pathways related to invasion into axillary lymph nodes. Interestingly, deregulated transcripts that correlate with the presence of lymph node metastases at the time of surgery conform a gene expression signature distinguishable to that observed for the lymph node–negative counterparts, suggesting different molecular programs related to the meta-static process.

Gene expression profiling will not necessarily replace classic approaches to predict the outcome; however, it will likely add substantial information that may help in better defining breast cancer outcome classes. The identification of individual proteins is also of high relevance not only for the potential value as prognostic biomarkers but also may provide insight into mechanisms and pathways of relevance in breast cancer progression. Nevertheless, given the molecular heterogeneity of breast cancer, further global and individual gene expression studies are needed to reliably discriminate breast cancer subgroups of value for determining outcome. Results of this study will provide novel insights into the molecular biology underlying breast cancer lymph node metastasis and recurrence.

## Materials and Methods

### SAGE Libraries

We did a comparative analysis of the gene expression profiles of 27 IDCA using SAGE. Libraries were generated at our laboratory (~100,000 SAGE tags per library). Table 1 shows histopathologic characteristics of the specimens analyzed. For the generation of SAGE libraries, snap-frozen samples were obtained from the M.D. Anderson Breast Cancer Tumor Bank, and SAGE analysis was done as previously described (37, 38).



## Data Processing and Statistical Analysis of SAGE Libraries

SAGE tag extraction from sequencing files was done using the SAGE2000 software version 4.0 (kindly provided by Dr. K. Kinzler, John Hopkins University). SAGE data management, tag to gene matching, as well as additional gene annotations and links to publicly available resources such as GO, UniGene, RefSeq, were done using a suite of Web-based SAGE library tools developed by us.<sup>5</sup>

Our analysis of data involved the following steps: (a) use of unsupervised RF clustering to group the patients based on their SAGE expression profiles; (b) investigate potential associations with multiple histopathologic variables; (c) identification of differentially expressed transcripts between clusters; (d) gene ontology analysis of the resulting transcripts.

We propose to use the RF clustering for SAGE data analysis because it has several relevant theoretical advantages. First, the RF dissimilarity approach handles mixed covariate types well, i.e., it can handle ordinal and continuous covariates in an unbiased way: the more related the covariate is to other covariates, the more it will affect the definition of the RF dissimilarity. Second, the clustering results do not change when one or more covariates are monotonically transformed because the dissimilarity only depends on the feature ranks. Third, the RF dissimilarity does not require the user to specify threshold values for dichotomizing tumor expressions. For the detailed description of RF clustering algorithm, consult Breiman (11) and Shi and Horvath (12). Briefly, the RF clustering procedure is carried out as follows. The RF dissimilarity is used to represent each patient as a point in a two-dimensional space with the aid of multidimensional scaling (39, 40). The distances between the points are used in partitioning around medoids clustering. The number of clusters is chosen by visually inspecting multidimensional scaling plots.

We tested whether variables differed across groups using the Fisher's exact test. All  $P$  values were two sided, and  $P < 0.05$  was considered significant. RF clustering and the analyses described above were carried out with the freely available software R (41).

To identify differentially expressed transcripts between clusters, we used a modified  $t$  test. This test is based on a beta binomial sampling model that takes into account both the intra-library and the inter-library variability, thus identifying common patterns of SAGE transcript tag changes systematically occurring across samples (13).<sup>6</sup>

For automated functional annotation and classification of genes of interest based on GO terms, we used the EASE Web-based software resource (14).<sup>7</sup>

## Meta-analysis of Breast Cancer Microarray Data Sets

To identify and validate the most reliable set of genes able to discriminate primary breast carcinomas based on their lymph node status, we did a cross-platform comparison between

---

<sup>5</sup>[http://spi.mdacc.tmc.edu/bitools/about/sage\\_lib\\_tool.html](http://spi.mdacc.tmc.edu/bitools/about/sage_lib_tool.html)

<sup>6</sup>All raw SAGE data reported as Supplementary Tables in this manuscript is publicly available at [http://sciencepark.mdanderson.org/labs/ggeg/SAGE\\_Proj\\_11.htm](http://sciencepark.mdanderson.org/labs/ggeg/SAGE_Proj_11.htm).

<sup>7</sup>Available at the Database for Annotation, Visualization and Integrated Discovery (DAVID) at <http://david.niaid.nih.gov/david> (15).

the described SAGE data set with previously reported breast cancer studies based on DNA microarray methods (1–3, 5–8, 42–45). The Oncomine cancer microarray database was employed for data collection and to investigate histopathologic associations (46). The Oncomine database is an integrated bioinformatic resource providing data collection, processing, and storage of all publicly available cancer microarray studies. All data are log transformed, median centered per array, and SD normalized to one per array. Gene module application lists all differential expression analyses in which the target genes were included and allows the user to select studies of interest, providing comparative statistical analyses. Selected comparisons of interest for meta-analysis included lymph node (–) versus lymph node (+) status, non-metastasis versus metastasis (5 years of follow-up), non-disease versus relapse (5 years of follow-up). The 55 up-modulated and 55 most down-regulated genes in lymph node (+) primary breast carcinomas were included for meta-analysis comparison. Data processing was carried out using comprehensive meta-analysis software v2 (Biostat, 2006). Standardized mean difference measures as scale-free indices and fixed effects analyses were employed for statistical integration. To enable visualization of meta-analysis results, we used The Institute for Genomic Research MultiExperiment Viewer (MeV 3.0) software. This tool was employed for average clustering of the *P* values obtained from each gene analyzed. When statistically significant coincidence among studies (i.e., SAGE and microarray studies) was observed on the behavior of specific transcripts, this was represented by colored boxes (red or green). Other progression parameters such as metastasis and disease-free survival (DFS) were also compared with the SAGE lymph node status findings. Statistically significant *P* values ( $P < 0.05$ ) associated with gene overexpression in lymph node (+), metastasis, and relapse (DFS) are represented in red; statistically significant down-modulated expression is represented in green color.

### Real-time RT-PCR Analysis

Template cDNAs were synthesized on mRNAs isolated from snap-frozen samples from an independent set of 40 stages I to II human breast carcinomas [21 lymph node (–) and 19 lymph node (+) IDCA samples]. Primers and probes were obtained from TaqMan Assays-on-Demand Gene Expression Products (Applied Biosystems). All the PCR reactions were done using the TaqMan PCR Core Reagents kit and the ABI Prism 7700 Sequence Detection System (Applied Biosystems). Experiments were done in triplicate, and each data point and 18S rRNA were used as control. Results were expressed as mean  $\pm$  2 SE based on  $\log_2$  transformation of normalized real-time RT-PCR values of the assayed genes. We used the *t* test to compare the gene expression levels of validated genes between lymph node (+) and lymph node (–) breast tumors ( $P < 0.05$ ).

### DCTN3 Antibodies Production

Polyclonal antibody against DCTN3 (a kind gift of Dr. Kevin Pfister, Department of Cell Biology, University of Virginia, Charlottesville, VA) was generated according to standard procedures. Briefly, we obtained rabbit serum from animals previously immunized with DCTN3 peptides as antigen. After generation of GST-DCTN3 fusion protein, we did an antibody affinity purification of such serum. The antibodies obtained, which were known to work in Western blots, were optimized for immunohistochemical analysis on paraffin sections (47).



## Tissue Microarray and Immunohistochemical Analyses

A breast cancer progression TMA was obtained from the M. D. Anderson Cancer Center (Houston, TX), and we were able to analyze a total of 87 cases representative of normal breast epithelium, ductal carcinoma *in situ*, invasive breast carcinoma, and metastatic tissues. Before immunostaining, endogenous peroxidase activity was blocked with 3% H<sub>2</sub>O<sub>2</sub> in water for 10 min. Heat-induced epitope retrieval was done with 1.0 mmol/L EDTA buffer (pH 8.0) for 10 min in a microwave oven followed by a 20-min cool down. To block nonspecific antibody binding, the slides were incubated with 10% goat serum in PBS for 30 min. DCTN3 protein was detected using primary anti-DCTN3 polyclonal antibody (1:100 dilution), and horseradish peroxidase–conjugated anti-rabbit secondary antibody. Staining development was done with 3,3'-diaminobenzidine (DAB), and the slides were then counterstained with hematoxylin. DCTN3 protein expression were measured using a Chromavision Automated Cellular Imaging System (ACIS) by means of the generic DAB software application. The software determines brown intensity regardless of the area covered by the positive cells.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

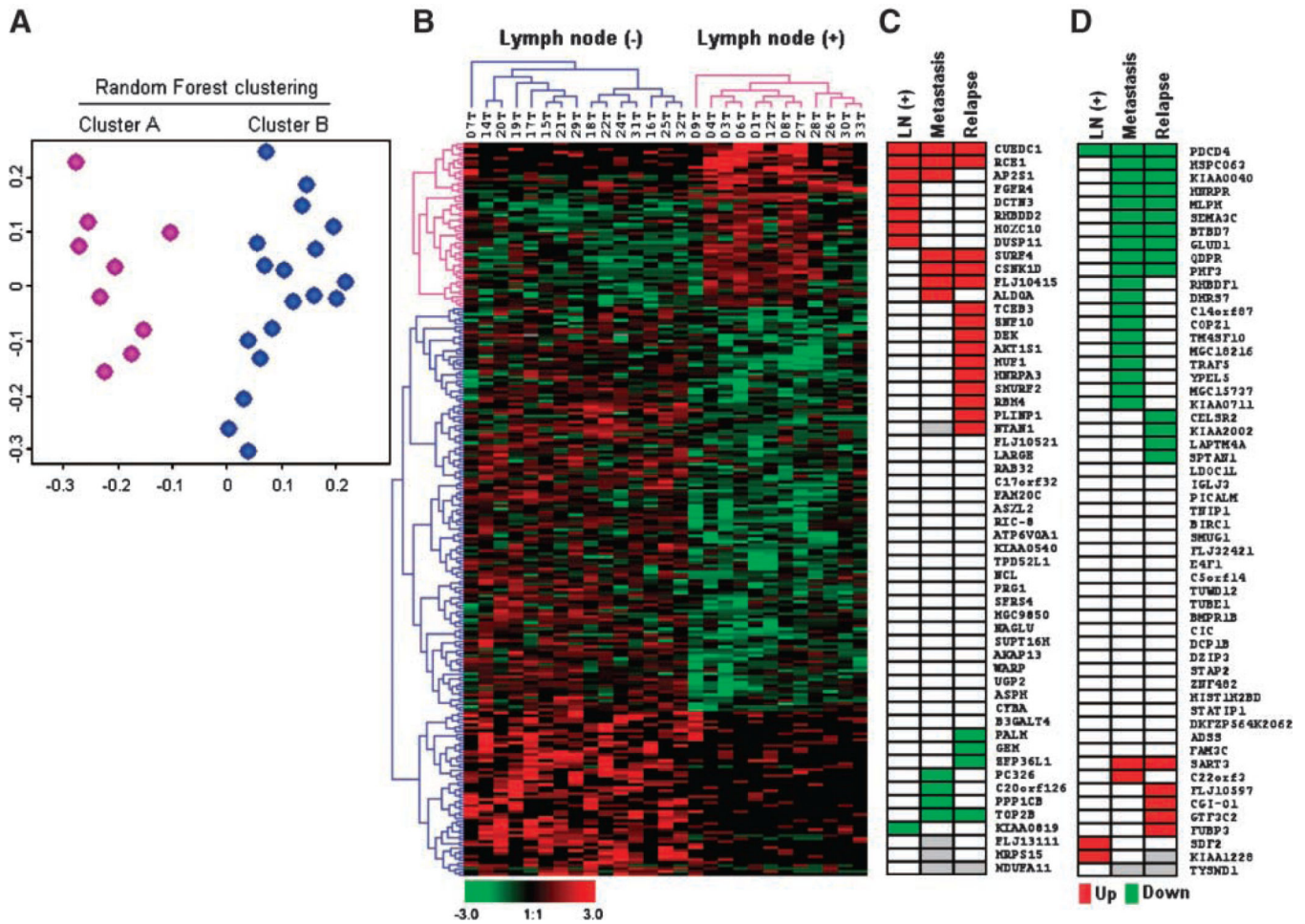
**Grant support:** NIH-National Cancer Institute grant 1U19 CA84978-1A1 (C.M. Aldaz), center grant ES-07784, and by the University of California at Los Angeles Integrative Graduate Education and Research Training Bioinformatics Program funded by NSF DGE 9987641 (T. Shi).

## References

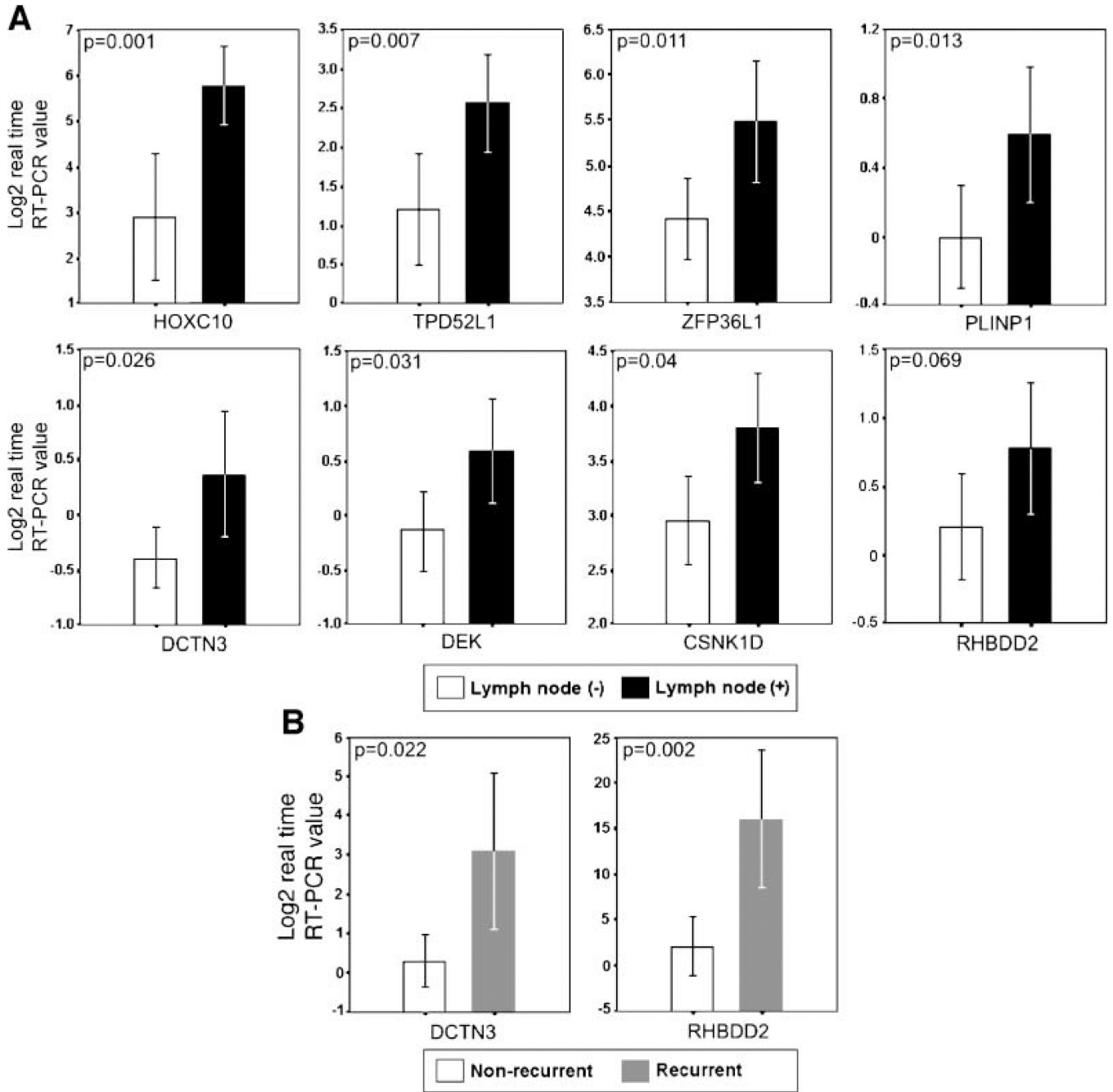
1. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001; 98:10869–10874. [PubMed: 11553815]
2. Van't Veer LJ, Dai Hongyue, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002; 415:530–536. [PubMed: 11823860]
3. Van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002; 347:1999–2009. [PubMed: 12490681]
4. Ahr A, Karn T, Solbach C, et al. Identification of high risk breast-cancer patients by gene expression profiling. *Lancet*. 2002; 359:131–132. [PubMed: 11809257]
5. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A*. 2003; 100:8418–8423. [PubMed: 12829800]
6. Sotiriou C, Neo S, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A*. 2003; 18:10393–10398. [PubMed: 12917485]
7. Huang E, Cheng SH, Dressman H, et al. Gene expression predictors of breast cancer outcomes. *Lancet*. 2003; 361:1590–1596. [PubMed: 12747878]
8. Wang Y, Klijn JGM, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*. 2005; 365:671–679. [PubMed: 15721472]
9. Chang HY, Nuyten DSA, Sneddon JB, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A*. 2005; 102:3738–3743. [PubMed: 15701700]

10. Krag D, Weaver D, Ashikaga T, et al. The sentinel node in breast cancer—a multicenter validation study. *N Engl J Med*. 1998; 339:941–946. [PubMed: 9753708]
11. Breiman L. Random forests. *Machine Learning*. 2001; 45:5–32.
12. Shi T, Horvath S. Unsupervised learning with random forest predictors. *J Comput Graph Stat*. 2006; 15:118–138.
13. Baggerly KA, Deng L, Morris JS, Aldaz CM. Differential expression in SAGE: accounting for normal between-library variation. *Bioinformatics*. 2003; 19:1477–1483. [PubMed: 12912827]
14. Hosack DA, Dennis G, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. *Genome Biol*. 2003; 4:R70. [PubMed: 14519205]
15. Dennis G, Sherman BT, Hosack DA, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol*. 2003; 4:r60.
16. Detours V, Dumont JE, Bersini H, Maenhaut C. Integration and cross-validation of high-throughput gene expression data: comparing heterogeneous data sets. *FEBS Lett*. 2003; 546:98–102. [PubMed: 12829243]
17. Byrne JA, Mattei MG, Basset P. Definition of the tumor protein D52 (TPD52) gene family through cloning of D52 homologues in human (hD53) and mouse (mD52). *Genomics*. 1996; 35:523–532. [PubMed: 8812487]
18. Boutros R, Fanayan S, Shehata M, Byrne JA. The tumor protein D52 family: many pieces, many puzzles. *Biochem Biophys Res Commun*. 2004; 325:1115–1121. [PubMed: 15555543]
19. Boutros R, Byrne JA. D53 (TPD52L1) is a cell cycle-regulated protein maximally expressed at the G2-M transition in breast cancer cells. *Exp Cell Res*. 2005; 310:152–165. [PubMed: 16112108]
20. Waldmann T, Scholten I, Kappes F, Hu HG, Knippers R. The DEK protein: an abundant and ubiquitous constituent of mammalian chromatin. *Gene*. 2004; 343:1–9. [PubMed: 15563827]
21. Kondoh N, Wakatsuki T, Ryo A, et al. Identification and characterization of genes associated with human hepatocellular carcinogenesis. *Cancer Res*. 1999; 59:4990–4996. [PubMed: 10519413]
22. Kroes RA, Jastrow A, Mclone MG, et al. The identification of novel therapeutic targets for the treatment of malignant brain tumors. *Cancer Lett*. 2000; 156:191–198. [PubMed: 10880769]
23. Dong X, Wang J, Kabir FN, et al. Autoantibodies to DEK oncoprotein in human inflammatory disease. *Arthritis Rheum*. 2000; 43:85–93. [PubMed: 10643703]
24. Kappes F, Scholten I, Richter N, Gruss C, Waldmann T. Functional domains of the ubiquitous chromatin protein DEK. *Mol Cell Biol*. 2004; 24:6000–6010. [PubMed: 15199153]
25. Wise-Draper TM, Allen HV, Thobe MN, et al. The human DEK proto-oncogene is a senescence inhibitor and an upregulated target of high-risk human papillomavirus E7. *J Virol*. 2005; 79:14309–14317. [PubMed: 16254365]
26. Cillo C, Cantile M, Faiella A, Boncinelli E. Homeobox genes in normal and malignant cells. *J Cell Physiol*. 2001; 188:161–169. [PubMed: 11424082]
27. Abate-Shen C. Deregulated Homeobox gene expression in cancer: cause or consequence? *Nat Rev Cancer*. 2002; 2:777–785. [PubMed: 12360280]
28. Gabellini D, Colaluca IN, Vodermaier HC, et al. Early mitotic degradation of the homeoprotein *HOXC10* is potentially linked to cell cycle progression. *EMBO J*. 2003; 22:3715–3724. [PubMed: 12853486]
29. Miller GJ, Miller HL, van Bokhoven A, et al. Aberrant *hoxc* expression accompanies the malignant phenotype in human prostate. *Cancer Res*. 2003; 63:5879–5888. [PubMed: 14522913]
30. Varnum BC, Ma QF, Chi TH, Fletcher B, Herschman HR. The TIS11 primary response gene is a member of gene family that encodes proteins with a highly conserved sequence containing an unusual cys-his repeat. *Mol Cell Biol*. 1991; 11:1754–1758. [PubMed: 1996120]
31. Karki S, LaMonte B, Holzbaun ELF. Characterization of p22 subunit of dynactin reveals the localization of cytoplasmic dynein and dynactin to the midbody of dividing cells. *Cell Biol*. 1998; 142:1023–1034.
32. Burkhardt JK, Echeverri CJ, Nisson T, Vallee RB. Overexpression of the dynamitin (p50) subunit of the dynactin complex disrupts dynein-dependent maintenance of membrane organelle disruption. *J Cell Biol*. 1997; 139:469–484. [PubMed: 9334349]

33. Bransfield KL, Askham JM, Leek JP, Robinson PA, Miqhell AJ. Phenotypic changes associated with dynactin-2 (DCTN2) over expression characterize SJS-1 osteosarcoma cells. *Mol Carcinog.* 2006; 45:157–163. [PubMed: 16369996]
34. Galigniana MD, Harrell JM, O'Hagen HM, Ljungman M, Pratt WB. HSP90- binding immunophilins link p53 to Dynein during p53 transport to the nucleus. *J Biol Chem.* 2004; 279:22483–22489. [PubMed: 15004035]
35. Pascall JC, Luck JE, Brown KD. Expression in mammalian cell cultures reveals interdependent, but distinct, functions for star and rhomboid proteins in the processing of the *Drosophila* transforming-growth-factor-a homologue Spitz. *Biochem J.* 2002; 363:347–352. [PubMed: 11931664]
36. Urban S, Lee JR, Freeman M. A family of rhomboid intramembrane proteases activates all *Drosophila* membrane-tethered EGF ligands. *EMBO J.* 2002; 21:4277–4286. [PubMed: 12169630]
37. Charpentier AH, Bednarek AK, Daniel RL, et al. Effects of estrogen on global gene expression: identification of novel targets of estrogen action. *Cancer Res.* 2000; 60:5977–5983. [PubMed: 11085516]
38. Hu Y, Sun H, Drake J, et al. From mice to human: identification of commonly deregulated genes in mammary cancer via comparative SAGE studies. *Cancer Res.* 2004; 64:7748–7755. [PubMed: 15520179]
39. Venables, WN.; Ripley, BD. *Modern applied statistic with S-Plus.* New York: Springer; 1999.
40. Cox, TF.; Cox, MAA. *Multidimensional scaling.* United Kingdom: CRC Press; 2001.
41. R Development Core Team. *R: a language and environment for statistical computing.* Vienna, Austria: R foundation for statistical computing; ISBN 3 – 900051 – 07 – 0. Available from: <http://www.r-project.org/>.
42. Zhao H, Langerod A, Ji Y, et al. Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. *Mol Biol Cell.* 2004; 15:2523–2536. [PubMed: 15034139]
43. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature.* 2000; 406:742–752. [PubMed: 10963601]
44. West M, Blanchette C, Dressman H, et al. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci U S A.* 2001; 98:11462–11467. [PubMed: 11562467]
45. Ma X, Salunga R, Tuggle JT, et al. Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci U S A.* 2003; 100:5974–5979. [PubMed: 12714683]
46. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia.* 2004; 6:1–6. [PubMed: 15068665]
47. Pfister KK, Benashski SE, Dillman JF, Patel-King RS, King SM. Identification and molecular characterization of the p24 dynactin ligh chain. *Cell Motil Cytoskeleton.* 1998; 41:154–167. [PubMed: 9786090]

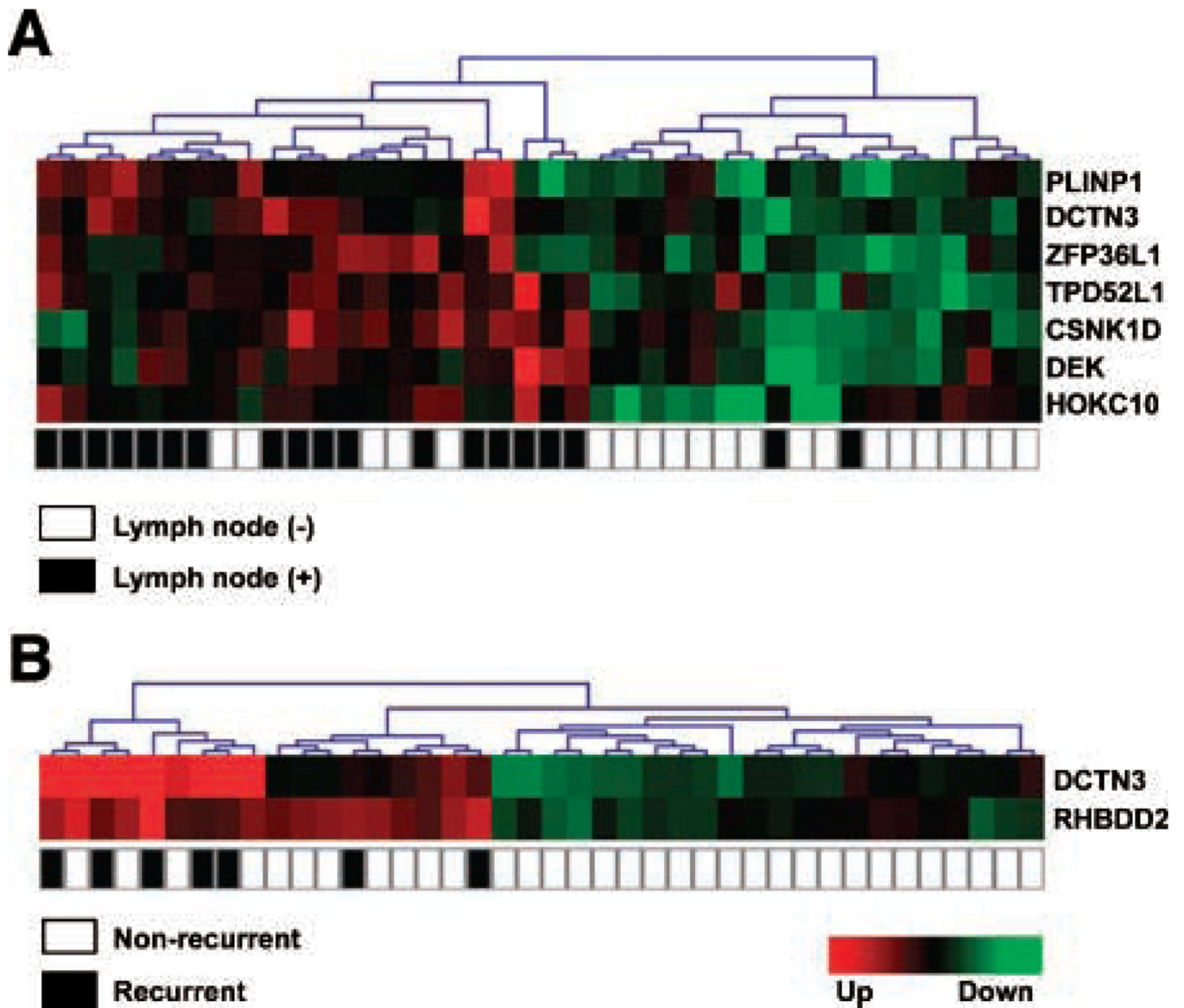


**FIGURE 1.** SAGE profiles of 27 primary invasive breast carcinomas. **A.** The SAGE profiles of 27 breast carcinomas are visualized in a two-dimensional multidimensional scaling plot where each dot represents one sample and the relative distances between samples are correlated with their RF dissimilarities. Breast carcinomas are colored by their RF clustering memberships: cluster A (*fuchsia*) composed by 78% of lymph node (+) carcinomas and cluster B (*blue*) composed of 87% of lymph node (-) breast carcinomas. **B.** Hierarchical clustering of 245 differentially expressed genes (55 up-modulated transcripts and 190 down-modulated transcripts) according to patient’s lymph node status based on pathologic diagnosis. Color scale at the bottom of the picture is used to represent expression level: low expression is represented by green, and high expression is represented by red. Results of meta-analysis (from publicly available gene expression microarray data sets) of 55 up-modulated (**C**) and 55 down-modulated transcripts (**D**) identified by SAGE. Red or green boxes, represent statistically significant agreement between our study and previously published studies not only on lymph node status, but also in association with other progression parameters such as metastasis or relapse. Red, statistically significant *P* values ( $P < 0.05$ ) associated with gene overexpression in lymph node (+), metastasis, and relapse (DFS); green, statistically significant down-modulated expression. Gray boxes, Unavailable data.



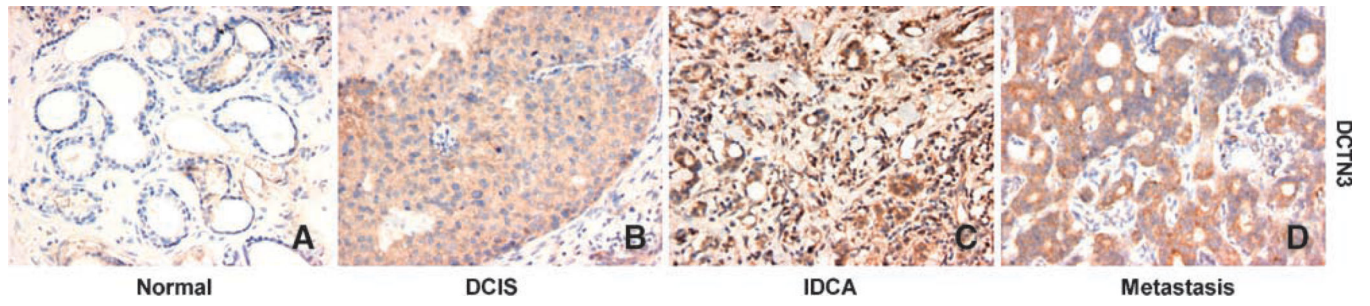
**FIGURE 2.** Validation assays of SAGE expression profiles in an independent set of primary invasive breast carcinomas ( $n = 40$ ). **A.** Real-time RT-PCR of seven up-modulated transcripts (*HOXC10*, *TPD52L1*, *ZFP36L1*, *PLINP1*, *DCTN3*, *DEK*, *CSNK1D*, *RHBDD2*) in LN(+) carcinomas. **B.** Real-time RT-PCR of two up-modulated transcripts (*DCTN3*, *RHBDD2*) in recurrent breast carcinomas. Mean  $\pm$  2 SE based on  $\log_2$  transformation of real-time RT-PCR values of the assayed gene relative to 18S rRNA used as normalizing control.





**FIGURE 3.** Hierarchical clustering of primary breast carcinomas based on real-time RT-PCR validation data. **A.** Cluster showing nodes in the basis of lymph node distribution ( $P = 0.0001$ ). **B.** Cluster showing nodes in the basis of recurrence status distribution ( $P = 0.001$ ).





**FIGURE 4.**  
DCTN3 immunohistochemical staining in normal (adjacent tumor), ductal carcinoma in situ (DCIS), invasive ductal carcinoma (IDCA), and metastatic breast samples.

**Table 1**

## Histopathologic Characteristics of Primary Breast Carcinomas Analyzed by SAGE

	Number (%)
Histology type	
Invasive ductal carcinoma	27 (100)
Tumor size	
1-2 cm	11 (41)
2-5 cm	13 (48)
>5 cm	3 (11)
Nodal status	
N0	16 (59)
N1	11 (41)
Nuclear grade	
Grade I	4 (15)
Grade II	14 (52)
Grade III	9 (33)
ER $\alpha$ status	
Negative	7 (26)
Positive	20 (74)

**Table 2**

Meta-analysis, Cross-Validated SAGE Transcripts as Poor-Prognosis Breast Cancer Biomarkers

Gene Name	Description	Lymph Node (+)	Metastasis (Yes)	Relapse (Yes)
<b>Up-modulated genes positively associated with the variable analyzed</b>				
<i>CUEDC1</i>	<i>CUE domain containing 1</i>	0.040	0.033	0.018
<i>RCE1</i>	<i>RCE1 homologue. prenyl protein protease</i>	0.038	0.0001	0.0001
<i>AP2S1</i>	<i>Adaptor-related protein complex 2</i>	0.015	0.038	0.508
<i>FGFR4</i>	<i>Fibroblast growth factor receptor 4</i>	0.035	0.279	0.913
<i>DCTN3</i>	<i>Dynactin 3 (p22)</i>	0.003	0.868	0.682
<i>RHBDD2</i>	<i>Rhomboid domain containing 2</i>	0.042	0.300	0.259
<i>HOXC10</i>	<i>Homeobox C10</i>	0.013	0.520	0.471
<i>DUSP11</i>	<i>Dual specificity phosphatase 11</i>	0.049	0.158	0.940
<i>SURF4</i>	<i>Surfeit 4</i>	0.393	0.024	0.009
<i>CSNK1D</i>	<i>Casein kinase 1 <math>\delta</math></i>	0.417	0.0001	0.001
<i>FLJ10415</i>	<i>Hypothetical protein</i>	0.029	0.0001	0.0001
<i>ALDOA</i>	<i>Aldolase A</i>	0.212	0.0001	0.109
<i>TCEB3</i>	<i>Transcription elongation factor B</i>	0.335	0.322	0.020
<i>ZNF10</i>	<i>Zinc finger protein 10 (KOK1)</i>	0.818	0.722	0.0001
<i>DEK</i>	<i>DEK oncogene</i>	0.419	0.673	0.005
<i>AKT1S1</i>	<i>AKT1 substrate 1 (proline-rich)</i>	0.533	0.360	0.037
<i>MUF1</i>	<i>MUF1 protein</i>	0.688	0.322	0.028
<i>HNRPA3</i>	<i>Heterogeneous nuclear ribonucleoprotein A3</i>	0.977	0.140	0.004
<i>SMURF2</i>	<i>SMAD specific E3 ubiquitin protein ligase 2</i>	0.704	0.058	0.0001
<i>RBM4</i>	<i>RNA binding motif protein 4</i>	0.882	0.783	0.032
<i>PLINP1</i>	<i>Growth arrest and DNA-damage-inducible</i>	0.424	0.208	0.039
<i>NTAN1</i>	<i>NH<sub>2</sub>-terminal asparagines amidase</i>	0.925	—	0.005
<b>Down-modulated genes negatively associated with the variables analyzed</b>				
<i>PDCD4</i>	<i>Programmed cell death 4</i>	0.007	0.001	0.0001
<i>HSPC063</i>	<i>HSPC063 protein</i>	0.185	0.0001	0.0001
<i>HNRPR</i>	<i>Heterogeneous nuclear ribonucleoprotein R</i>	0.803	0.0001	0.0001
<i>KIAA0040</i>	<i>KIAA0040 protein</i>	0.624	0.003	0.005
<i>MLPH</i>	<i>Melanophilin</i>	0.780	0.007	0.016
<i>SEMA3C</i>	<i>Sema domain, immunoglobulin domain (Ig)</i>	0.285	0.001	0.0001
<i>BTBD7</i>	<i>BTB (POZ) domain containing 7</i>	0.629	0.003	0.003
<i>GLUD1</i>	<i>Glutamate dehydrogenase 1</i>	0.220	0.002	0.0001
<i>QDPR</i>	<i>Quinoid dihydropteridine reductase</i>	0.217	0.0001	0.0001
<i>PHF3</i>	<i>PHD finger protein 3</i>	0.342	0.0001	0.001
<i>RHBDF1</i>	<i>Rhomboid family 1 (Drosophila)</i>	0.491	0.039	0.426
<i>DHRS7</i>	<i>Dehydrogenase/reductase (SDR family) member 7</i>	0.584	0.047	0.167
<i>C14orf87</i>	<i>Chromosome 14 open reading frame 87</i>	0.346	0.011	0.968
<i>COPZ1</i>	<i>Coatomer protein complex subunit <math>\zeta</math>1</i>	0.272	0.031	0.704
<i>TM4SF10</i>	<i>Transmembrane 4 superfamily member 10</i>	0.570	0.026	0.293

Gene Name	Description	Lymph Node (+)	Metastasis (Yes)	Relapse (Yes)
<i>MGC18216</i>	<i>Hypothetical protein MGC18216</i>	0.731	0.002	0.508
<i>TRAF5</i>	<i>Tumor necrosis factor receptor-associated factor 5</i>	0.188	0.016	0.673
<i>YPEL5</i>	<i>Yippee-like 5 (Drosophila)</i>	0.188	0.001	0.212
<i>MGC15737</i>	<i>Hypothetical protein MGC15737</i>	0.841	0.019	0.365
<i>KIAA0711</i>	<i>KIAA0711 gene product</i>	0.936	0.003	0.841
<i>CELSR2</i>	<i>Cadherin, EGF LAG seven-pass G-type receptor 2</i>	0.494	0.384	0.009
<i>KIAA2002</i>	<i>KIAA2002 protein</i>	0.600	0.172	0.017
<i>LAPTM4A</i>	<i>Lysosomal-associated protein transmembrane 4</i>	0.297	0.164	0.004
<i>SPTAN1</i>	<i>Spectrin <math>\alpha</math>, non-erythrocytic 1</i>	0.994	0.463	0.0001

NOTE: Studies included in the meta-analysis: Sorlie et al. (5); van de Vijver et al. (3); van't Veer et al. (2); Huang et al. (7); Ma et al. (45); Sotiriou et al. (6); Zhao et al. (42); Wang et al. (8).

**Table 3**

Most Highly Deregulated Transcripts in Lymph Node (+) Breast Carcinomas Identified by SAGE (Fold Change, >3;  $P < 0.01$ )

Gene Name	Description	Tag	Entrez Gene ID	Expression
Regulation of cell proliferation				
<i>DEK</i>	<i>DEK oncogene</i>	ACAAAAGTGA	7913	▲
<i>TPD52L1</i>	<i>Tumor protein D52-like 1</i>	ACTGTCTCCA	7164	▲
<i>GEM</i>	<i>GTP binding protein</i>	GAGCCATCAT	2669	▲
<i>AKAP13</i>	<i>A kinase (PRKA) anchor protein 13</i>	GGATGCGCAG	11214	▲
<i>CCRK</i>	<i>Cell cycle related kinase</i>	GGATGATGTC	23552	▼
Regulation of transcription related				
<i>MGC9850</i>	<i>Hypothetical protein</i>	TGCTTGACAA	219404	▲
<i>ZFP36L1</i>	<i>Zinc finger protein 36, C3H type-like 1</i>	CTTCTCTCCC	677	▲
<i>TP53BP1</i>	<i>Tumor protein p53 binding protein 1</i>	ACAGTGCTTG	7158	▼
<i>ATF2</i>	<i>Activating transcription factor 2</i>	GTGGATTCAT	1386	▼
<i>CBX4</i>	<i>Chromobox homologue 4</i>	AAAGTCTAGA	8535	▼
Signal transduction related				
<i>CSNK1D</i>	<i>Casein kinase 1, <math>\delta</math></i>	GCTGATCTAC	1453	▲
<i>PPP1CB</i>	<i>Protein phosphatase 1</i>	AAGATTTTAG	5500	▲
<i>IGFBP4</i>	<i>Insulin-like growth factor binding prot. 4</i>	TTTGAATGT	3487	▼
<i>ARHGAP1</i>	<i>Rho GTPase activating protein 1</i>	TGTCTGTGGT	392	▼
<i>FYCO1</i>	<i>FYVE and coiled-coil domain containing</i>	TTAAATGCAA	79443	▼
<i>P2RY2</i>	<i>Purinergic receptor P2Y, G-protein</i>	AGTAAACCAT	5029	▼
Cytoskeleton				
<i>DCTN3</i>	<i>Dynactin 3 (p22)</i>	CTGCCGCCT	11258	▲
<i>MYH3</i>	<i>Myosin, heavy polypeptide 3</i>	GTCTCATTTT	4621	▼
Protein transport/targeting and biosynthesis				
<i>AP2S1</i>	<i>Adaptor-related protein complex 2</i>	CCGTGGTCAC	1175	▲
<i>SUPT16H</i>	<i>Suppressor of Ty 16 homologue</i>	CCTTGGGCCT	11198	▲
<i>HSPS4</i>	<i>Hermansky-Pudlak syndrome 4</i>	TTTGACTG	89781	▼
<i>TOMM20</i>	<i>Translocase of outer mitochondrial memb.</i>	TGTGAGCCCT	9804	▼
Metabolism and miscellaneous				
<i>ATP6V0A1</i>	<i>ATPase, H<sup>+</sup> transporting</i>	TGGCTGTGAG	535	▲
<i>NTAN1</i>	<i>NH<sub>2</sub>-terminal asparagine amidase</i>	AATTACCAA	123803	▲
<i>NAGLU</i>	<i>N-acetylglucosaminidase, <math>\alpha</math></i>	GCTGAGCTGG	4669	▲
<i>SMURF2</i>	<i>SMAD specific E3 ubiquitin protein</i>	ATCTGAACA	64750	▲
<i>PRG1</i>	<i>Proteoglycan 1, secretory granule</i>	GCCATAAAAT	5552	▲
<i>CNKSR1</i>	<i>Connector enhancer of kinase suppres.</i>	TACAGTTCCC	10256	▼
<i>DAP13</i>	<i>13 kDa differentiation-associated protein</i>	TGTTATTA	55967	▼
<i>MANBAL</i>	<i>Mannosidase, <math>\beta</math>A, lysosomal-like</i>	CAACTAATTC	63905	▼
<i>ADSS</i>	<i>Adenylosuccinate synthase</i>	GACTACCTTT	159	▼
Function unknown				

Gene Name	Description	Tag	Entrez Gene ID	Expression
<i>FAM20C</i>	<i>Family with sequence similarity 20, C</i>	CGCCCGTCGT	56975	▲
<i>C20orf126</i>	<i>Chromosome 20 ORF 1</i>	GGTGTTGCT	81572	▲
<i>AKT1S1</i>	<i>AKT1 substrate 1 (proline-rich)</i>	CGCGCGCTGG	84335	▲
<i>RIC-8</i>	<i>Likely ortholog of mouse synembryn</i>	ATTTGCCTCT	60626	▲
<i>MUF1</i>	<i>MUF1 protein</i>	GGCTGCCAG	10489	▲
<i>WARP</i>	<i>Von Willebrand factor A domain</i>	CCCAGGACAC	64856	▲
<i>PKD1-like</i>	<i>Polycystic kidney disease 1-like</i>	TTGACACTTT	79932	▼
<i>MESDC1</i>	<i>Mesoderm development candidate 1</i>	ACAAGAATTG	59274	▼
<i>TTC15</i>	<i>Tetratricopeptide repeat domain 15</i>	TTTTACTCAC	51112	▼
<i>BCMP11</i>	<i>Breast cancer membrane protein 11</i>	CGGCAGAGCT	155465	▼
<i>MGC10067</i>	<i>Hypothetical protein</i>	GATGTCTTGT	134510	▼
<i>AMIGO2</i>	<i>Amphoterin induced gene 2</i>	CCCCATACTA	347902	▼

NOTE: ▲, up-regulated gene in lymph node (+) primary breast carcinomas. ▼, down-regulated gene in lymph node (+) primary breast carcinomas.



**Table 4**

## DCTN3 Protein Expression According to Histopathologic Characteristics

	Protein Expression <i>n</i> (%)			Statistic
	Absent	Moderate	Strong	
Normal epithelium	13 (72)	18 (30.5)	1 (10)	$\chi^2 = 15.3, P = 0.004, P \text{ trend} = 0.001$
DCIS	1 (5.5)	20 (34)	3 (30)	
IDCA/Metastasis	4 (22)	21 (35.5)	6 (60)	
LN(-)	12 (67)	3 (23)	2 (25)	$\chi^2 = 7.25, P = 0.027, P \text{ trend} = 0.007$
LN(+)	6 (33)	10 (77)	6 (75)	