[CANCER RESEARCH 59, 5464–5470, November 1, 1999]

Advances in Brief

Combining Serial Analysis of Gene Expression and Array Technologies to Identify Genes Differentially Expressed in Breast Cancer[1]

Mariana Nacht, Anne T. Ferguson,[2] Wen Zhang, Joseph M. Petroziello, Brian P. Cook, Yu Hong Gao, Sharon Maguire, Deborah Riley, George Coppola, Gregory M. Landes, Stephen L. Madden, and Saraswati Sukumar[3]

Genzyme Molecular Oncology, Framingham, Massachusetts 01701 [M. N., W. Z., J. M. P., B. P. C., Y. H. G., G. C., G. M. L., S. L. M.]; Department of Oncology, Novartis Institute for Biomedical Research, Summit, New Jersey 07901 [S. M.]; Lion Bioscience, Cambridge, Massachusetts 02141 [D. R.]; and Breast Cancer Program, Johns Hopkins Oncology Center, Baltimore, Maryland 21205 [A. T. F., S. S.]

Abstract

Several methods have been used recently to determine gene expression profiles of cell populations. Here we demonstrate the strength of combining two approaches, serial analysis of gene expression (SAGE) and DNA arrays, to help elucidate pathways in breast cancer progression by finding genes consistently expressed at different levels in primary breast cancers, metastatic breast cancers, and normal mammary epithelial cells. SAGE profiles of 21PT and 21MT, two well-characterized breast tumor cell lines, were compared with SAGE profiles of normal breast epithelial cells to identify differentially expressed genes. A subset of these candidates was then placed on an array and screened with clinical breast tumor samples to find genes and expressed sequence tags that are consistently expressed at different levels in diseased and normal tissues. In addition to finding the predicted overexpression of known breast cancer markers *HER-2/neu* and *MUC-1*, the powerful coupling of SAGE and DNA arrays resulted in the identification of genes and potential pathways not implicated previously in breast cancer. Moreover, these techniques also generated information about the differences and similarities of expression profiles in primary and metastatic breast tumors. Thus, combining SAGE and custom array technology allowed for the rapid identification and validation of the clinical relevance of many genes potentially involved in breast cancer progression. These differentially expressed genes may be useful as tumor markers and prognostic indicators and may be suitable targets for various forms of therapeutic intervention.

Introduction

Gene expression patterns are crucial for maintaining and altering phenotypes of cells. Recent technological advances have resulted in several widely used methods for the large-scale study of gene expression, including comprehensive open systems, such as SAGE[4] (1), READS (2), AFLP (3), TOGA (Digital Gene Technologies, Inc., La Jolla, CA), and Gene Calling (Curagen Corp., New Haven, CT), and focused closed systems such as cDNA microarrays (4) and oligonucleotide chips (5). Each of these techniques alone is powerful in identifying genes that are differentially expressed in normal and diseased cells. However, only the open systems can identify expressed genes that have not yet been cloned or partially sequenced. Moreover, the open system platform, with SAGE as a prime example, can evaluate the expression patterns of tens of thousands of genes in a quantitative manner (1, 6–8). SAGE and other open systems are limited, however, to generating expression profiles of a restricted number of samples because of both labor considerations and the availability of sufficient RNA from experimental samples. Closed systems are excellent methods for rapidly screening large numbers of sequences, although the value of the information generated is limited to the choice, often arbitrary, of the known sequences screened. By combining open and closed system technologies in series, a limited number of samples can be used to identify sequences not previously implicated in a disease, and then custom arrays can rapidly verify these expression patterns in a large number of clinical samples.

A comparison of gene expression profiles in normal and tumor cells can yield critical information about the processes of transformation and metastasis. Although many studies have identified oncogenes and tumor suppressor genes that mark the transformation of cells from the colon, pancreas (9), and lung (8), comparable studies in breast cancer have met with limited success. This reflects both the difficulty in finding genetic and epigenetic alterations that are present in a significant proportion of breast cancers and the heterogeneity of breast cancer itself.

To help elucidate pathways that mediate breast cancer progression and to find targets for potential therapeutic intervention, we combined SAGE and custom array analysis. We identified differentially expressed genes by generating SAGE profiles of two well-characterized breast tumor cell lines, 21PT and 21MT, and two normal mammary epithelial cell cultures. To find genes and ESTs that are consistently expressed at different levels in diseased and normal tissues, a subset of the differentially expressed genes identified by SAGE was then spotted on an array and screened with complex cDNA probes extracted from 17 breast tumor samples and four normal breast epithelial cell cultures. SAGE analysis of 21PT, 21MT, and normal breast epithelial cells identified >200 transcripts that were differentially expressed at least 10-fold between cancer and normal cells. Custom array analysis verified the expression of some of these candidates in primary and metastatic tumors and identified several differences between the two cancerous states. Our data demonstrate that by coupling the advantages of SAGE and cDNA array technologies, we can identify genes and potential pathways that were not implicated previously in breast cancer. Moreover, this combination of techniques generated information about the similarities and differences in the expression profiles of primary and metastatic breast tumors.

Materials and Methods

Cell Lines and Primary Tissues. SAGE was performed on two independent primary cultures (HMECs) derived from normal mammary epithelial cells (Clonetics, San Diego, CA) and two breast cancer cell lines, 21PT and 21MT (a generous gift of Dr. Vimla Band, Tufts University School of Medicine, Boston, MA). HMEC cells were maintained in mammary epithelial cell basal

Received 8/6/99; accepted 9/17/99.

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.

[1] The research was funded by Genzyme Molecular Oncology and by USPHS Grants T32 CA09630 (to A. T. F.), CA48943, and P21 CA/ES 66204, a grant from the Susan G. Komen Foundation, and a grant from Johns Hopkins Oncology Research Funds (to S. S.).

[2] Present address: Calydon, Inc., 1324 Chesapeake Terrace, Sunnyvale, CA 94089.

[3] To whom requests for reprints should be addressed, at Johns Hopkins Oncology Center, 370 Ross, 720 Rutland Avenue, Baltimore MD 21205. Phone: (410) 614-2479; Fax: (410) 614-4073; E-mail: saras@pop.jhmi.edu.

[4] The abbreviations used are: SAGE, serial analysis of gene expression; EST, expressed sequence tag; HMEC, human mammary epithelial cell; poly(A)+, polyadenylated; HMG, high mobility group; MMP, matrix metalloproteinase; GST, glutathione *S*-transferase; IGF, insulin-like growth factor.

Table 1 *Summary of SAGE analysis*

	Normal breast 1	Normal breast 2	21PT	21MT
Total clones	3,222	2,813	2,003	3,035
Total tags	58,775	49,513	59,537	60,827
Unique tags	17,762	16,123	16,187	14,368
GenBank matches	13,614	12,483	12,387	11,229

medium supplemented with 52 μg/ml bovine pituitary extract, 10 μg/ml human epidermal growth factor, 5 μg/ml insulin, 0.5 μg/ml hydrocortisone, 50 μg/ml gentamicin, and 50 ng/ml amphotericin-B (Clonetics). Both normal mammary epithelial cell cultures used for the SAGE analysis were propagated to passage 10 and allowed to remain in culture for 4 days after reaching confluency, before preparing the samples for SAGE. 21PT (primary tumor) and 21MT (metastatic tumor) are two cell lines derived from the same patient (10). These cells were maintained in 90% α-MEM (Life Technologies, Inc., Gaithersburg, MD) supplemented with 10% FCS, 10 mM HEPES, 0.1 mM nonessential amino acids, 1 mM sodium pyruvate, 2 mM glutamine, 100 units/ml penicillin, 100 μg/ml streptomycin, 1 μg/ml insulin, 25 ng/ml epidermal growth factor, and 1 μg/ml hydrocortisone. 21PT and 21MT were not grown to confluence before preparing RNA for SAGE. Normal breast epithelial cells 184 were a generous gift of Dr. Martha Stampfer (University of California, Berkeley, CA) and were maintained according to her growth conditions described elsewhere.[5] Additional normal breast epithelial cells used to screen the custom arrays were also obtained from Clonetics and maintained as described above. The cultures used to screen the arrays were propagated to passage 7 and either used at confluency or 4 days after confluency. Primary breast tumor tissues and metastatic samples were obtained immediately after surgical resection at Johns Hopkins University. Microscopic examination of representative tissue sections from each tumor revealed that these samples contained >50% tumor cells.

SAGE Analysis. SAGE analysis was performed as described (1, 9). The sequence and abundance of each of the transcript tags were determined by SAGE software as described previously (8).

RNA Extraction. Total RNA was isolated from cultured cells using RNazol B (Biotecx Laboratories, Inc., Houston, TX), according to manufacturer's instructions. poly(A)+ RNA was extracted using the Oligotex mRNA Midi kit (Qiagen Inc., CA). Total RNA was isolated from tumor samples using the acid phenol extraction method (11) or Trizol reagent (Life Technologies, Inc.) according to manufacturer's specifications.

Custom Array Generation and Analysis. Plasmids containing cDNA sequences for genes of interest were obtained from Genome Systems (St. Louis, MO). cDNA target sequences were amplified by PCR, purified using QIAQuick PCR purification kit or QIAQuick Gel extraction kit (Qiagen Inc. CA), and arrays were prepared by spotting the targets on replicate nylon membranes (Biotrans; ICN, Costa Mesa, CA) at a concentration of 2 ng of DNA/spot, using a BIOMEK 2000 robot (Beckman Coulter, Inc., Fullerton, CA). Each target was spotted in quadruplicate. For probes, poly(A)+ RNA was isolated from cell lines and tissues as described above. poly(A)+ RNA (70 ng) was converted to cDNA and labeled with [α-^{32}P]dCTP by reverse transcription using Superscript II RT (Life Technologies, Inc.). Hybridizations were performed overnight at 42°C in 6× SSC/0.1% SDS/50% formamide/5× Denhardt's solution. Membranes were washed three times in 2× SSC/0.1% SDS at room temperature for 10 min each, followed by three high stringency washes in 0.1× SSC/0.1% SDS at 65°C for 30 min each. Hybridization intensities were quantitated on a STORM PhosphorImager (Molecular Dynamics, Sunnyvale, CA), and an average signal for each set of four targets was determined. The coefficient of variation for each set of four spots was ~30% (data not shown). For each membrane, the signal intensity of each cDNA was graphed, and a normalization factor based on the 95% of gene expression was applied to each data point to generate normalized values.

Results

Generation of SAGE Libraries and Tag Analysis. mRNA was isolated from two breast cancer cell lines and two normal mammary epithelial cell cultures and used to construct four SAGE libraries. 21PT and 21MT are well-characterized tumor lines derived from the primary and metastatic tumors, respectively, of the same patient with infiltrating and intraductal carcinoma (10). More than 2000 clones were isolated and sequenced for each library (Table 1). SAGE generates a 14- or 15-bp "tag" sequence from a defined position within the transcript, which is sufficient to uniquely identify individual transcripts (1). A total of 228,652 tags were sequenced, with 48,691 representing unique tags. Ten % (4,916) of the unique tags matched to known mRNA sequences in GenBank, and 73% of the unique tags appeared more than once in the libraries. Fifty-six % (27,415) of the unique tags matched to either known genes or EST clusters in the Unigene database. Although 21,276 (44%) of the unique tags did not match to any known sequences in the database, the majority of these tags (17,674 tags or 83%) appeared only once in the sequenced libraries. Previous SAGE studies found that sequencing errors produce a tag error rate of ~7% (9). Therefore, some of the tags that only appeared once may be due to sequencing errors (~1,237), whereas others may represent genes expressed at very low levels. 3,603 (7.4%) tags represent potentially novel genes that appeared more than once.

Identification of Transcripts Differentially Expressed in Breast Cancer Using SAGE. To identify genes that were differentially expressed in the breast cancer cell lines, for each unique tag, we compared the average expression in 21PT and 21MT to the average expression in the normal breast cells. The vast majority of tags did not show differential expression. However, 119 tags were expressed at least 10-fold greater in the tumor cells than in the normal cells, and 539 tags were expressed at least 5-fold higher in the cancer lines. Conversely, 94 tags were expressed at least 10-fold greater in the normal cells than in the cancer cells, and 381 tags were expressed at least 5-fold more highly in the normal cultures.

The SAGE data revealed differential expression of several genes that have been implicated previously in breast and other cancers (Table 2). Table 2A describes those known genes that were expressed at elevated levels in the tumor cell lines. The tags representing two well-known breast tumor-associated antigens, HER2/neu and Mucin, were highly overexpressed in the tumor cell lines. The *HER2/neu* oncogene encodes a tyrosine kinase receptor with homology to the epidermal growth factor receptor. Overexpression of HER2/neu has

Table 2 *Differentially expressed genes identified by SAGE*

TAG	T/N[a]	Identity
A.		
AGGAAGGAAC	51/1	*HER-2/neu*[b]
GGACTCTGGA	42/0	*Zinc-α-glycoprotein*[b]
GACATCAAGT	29/1	*Cytokeratin-19*
CAAACCATCC	29/1	*Cytokeratin-18*
CCTGGGAAGT	18/0	*Mucin*[b]
CCTCCAGCTA	15/1	*Cytokeratin-8*
ATTTGTCCCA	15/1	*HMG(Y)*[b]
CAGGGGAGTG	10/1	*N-Methyl purine DNA glycosylase*
GCCCTGAGCG	8/1	*Membrane-type MMP-15*
ACTGAGGTGC	7/1	*FGF-1*
B.		
GATCTCTTGG	1/152	*S100A2/CaN19*
AGGTCCTAGC	1/134	*GST-pi*[b]
CTTCCTTGCC	1/109	*Cytokeratin-17*
GCCTGTACAA	1/102	*IGF binding protein 2*[b]
TAATAAAGAA	1/26	*Cytokeratin-15*
CATTGTAAAT	1/20	*Maspin*
TAAACCTGCT	1/18	*Galectin-7*
AGGTCTTCAA	1/18	*Thrombospondin-1*[b]
GTAAATATGG	1/12	*Bullous pemphigoid antigen*[b]
GACCAGGCCC	1/12	*Tropomyosin-2*

[a] Ratios were calculated by comparing the average of 21PT and 21MT with the average of the two normal breast libraries.
[b] cDNA spotted on custom array.

[5] Internet address: http://www.lbl.gov/LBL-Programs/mrgs/review.html.

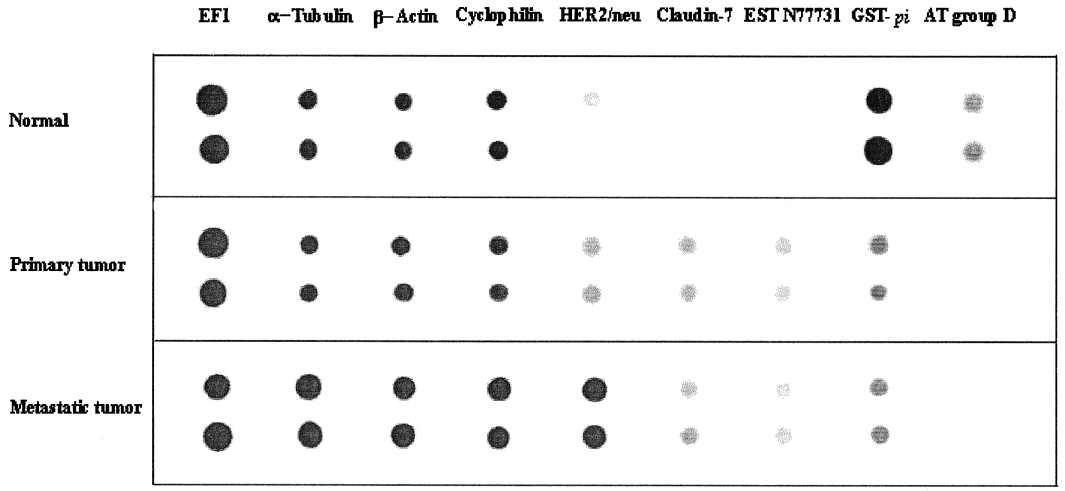

Fig. 1. Array analysis of genes identified by SAGE. Two ng of cDNA corresponding to tags identified by SAGE were spotted in quadruplicate on nylon membranes. Complex DNA probes from normal breast epithelial cells, primary breast tumors, and metastatic breast tumors were hybridized to replicate arrays. Duplicate spots from each type of probe are shown here. *EF1*, *α-tubulin*, *β-actin*, and *cyclophilin* were spotted as internal controls on each membrane. The signal intensities were quantitated on a STORM phosphorimager, and the membranes were exposed to film. *GTP*, GST-pi; *AT*, ataxia telangiectasia.

been observed in a variety of cancers, including ~30% of breast cancers (reviewed in Ref. 12). Similarly, epithelial mucin, encoded by the *MUC-1* gene, is highly expressed in many invasive tumors including carcinomas of the colon (13) and breast (reviewed in Ref. 14). Another secreted protein, zinc-α-2-glycoprotein, has been detected at elevated levels in the serum and breast fluids of breast cancer patients, particularly in those with well-differentiated tumors (15).

Maintaining the integrity of chromosomal structure and transcriptional regulation is critical to nonneoplastic cells. SAGE identified the transcription factor *HMGI(Y)* overexpressed in the breast tumor lines. HMG proteins are believed to be involved in chromatin organization, and *HMGI(Y)* has been found previously at elevated levels in transformed cells of the prostate (16), lung (17), thyroid (18), and breast (19). Interestingly, the gene encoding another protein that may contribute to alterations in chromosomal architecture in breast cancer cells, *N-methyl purine DNA glycosylase* (20), was also highly expressed in the breast tumor lines.

Changes in cytoskeletal structure also accompany neoplastic transitions. Accordingly, alterations in cytokeratin expression, especially *K8*, *K18*, and *K19*, have been correlated with breast cancer progression (21). Similarly, elevated levels of *FGF-1* were observed in the cancer cells and have been associated with hormone-independent breast tumor growth, vascularization, and increased metastatic potential (12). SAGE tags for each of these genes were found to be overrepresented in breast cancer cells compared with normal breast epithelial cells. Moreover, a member of the MMP family *(MMP-15)* was highly expressed in the tumor lines, consistent with previous observations for other MMPs (12). MMP proteins are able to degrade collagen and therefore may aide epithelial cell migration associated with metastatic progression (22).

SAGE analysis also identified genes that are known to be underexpressed in transformed cells (Table 2B). Several genes encoding proteins with tumor suppressor capabilities were detected. Maspin, a member of the serpin superfamily of protease inhibitors, was highly overexpressed in the normal breast cells with respect to the tumor cell lines, as described previously (23). Although the target protease of maspin remains unknown, the protein has been shown to inhibit mammary tumor cell invasion and metastasis. Thrombospondin, an inhibitor of both tumorigenic and angiogenic activity in breast and other tissues (12), was also overexpressed in the normal breast samples. Furthermore, *IGF binding protein-2*, a modulator of the mitogenic activities of IGFs, was very highly expressed in the normal breast cells. Many studies have implicated IGF pathways in breast cancer progression (24, 25).

Markers of normal mammary epithelium were also detected by SAGE. CaN19 is a member of the S100 family of Ca^{2+} binding proteins. Unlike other S100 proteins, *CaN19* expression is high in normal mammary epithelia and lost in tumor cells (26). Loss of *CaN19* expression in tumor cells may contribute to the calcification seen in mammograms as an early marker for breast cancer (26). As described above, tumor cells produce mainly keratins *K8*, *K18*, and *K19*, whereas nonhyperplastic cells express a different subset of keratins, including *K15* and *K17* (27), consistent with our results. The *GST-pi* gene was highly expressed in the normal breast cultures when compared with the tumor cell lines. Accordingly, *GST-pi* has been found to be hypermethylated and not expressed in tumors of the prostate, kidney, and breast (28, 29). Galectin-7, a protein involved in cell-cell and cell-matrix interactions, was abundant in the normal breast cells and has been observed previously in nontransformed cells but not in carcinoma cells (30). Interestingly, *Galectin-7* was found to be a *P53*-inducible gene in a SAGE analysis of colorectal cancer cells (7), although the tag sequence was different from the one detected here, suggesting that there are alternative splice forms of the mRNA. Tropomyosin 2 was also detected at elevated levels in the normal cells. High molecular weight tropomyosins, which regulate microfilament stability, are more highly expressed in noncancerous mammary cells than in breast tumors (31). Finally, hemidesmosomes, adhesion structures found in stationary epithelial cells but not in migratory cells, are readily detected in normal mammary epithelial cells but are lost in invasive carcinoma cells (32). Bullous pemphigoid antigen, a protein associated with hemidesmosomes, was found overexpressed by 12-fold in the normal breast cells compared with the breast tumor cell lines.

Confirmation of SAGE-identified, Differentially Expressed cDNAs with Clinical Tumors. Genes that are highly expressed, or repressed, in breast tumors may be useful as tumor markers and may also play a functional role in disease progression. To extend the SAGE results to a larger set of samples with more clinical relevance, a subset of the differentially expressed genes identified by SAGE was spotted on an array and further screened using clinical tumors as probes. Sixty-eight known genes and ESTs, corresponding to SAGE tags that were either up- or down-regulated at least 5-fold, on average, in 21PT

and 21MT with respect to the normal cells, were spotted on replicate nylon membranes. In addition to spotting genes already implicated in breast cancer, genes and ESTs not known previously to be involved in breast cancer were also included on the arrays for further study. SAGE analysis had detected little variation in the averaged expression of *EF-1*, *β-actin*, *α-tubulin*, and *cyclophilin* in the tumor and normal samples (data not shown), and these genes were included on the arrays as controls. mRNA was extracted from 7 primary breast tumors, 10 metastatic breast tumors, and 4 normal human mammary epithelial cell cultures. The transcripts were converted to ^{32}P-labeled cDNAs and were used to screen replicate membranes. The signals from each hybridization were quantitated and normalized to decrease variations attributable to differences in overall hybridization efficiency. For each spotted cDNA target, the average signal intensity was calculated for hybridizations with the primary tumors, metastatic tumors, and normal breast tissue. Array expression profiles for the primary and metastatic tumors were compared with the array expression profiles for the normal breast cells.

As expected from the SAGE results, *EF-1*, *β-actin*, *α-tubulin*, and *cyclophilin* showed relatively even expression by array analysis (Fig. 1). The average of the tumor *versus* normal ratios for these four control genes was 1.1.

Overall, the expression patterns of the chosen genes in the primary and metastatic breast tumors were similar (Table 3 and Fig. 1). However, only 6 of the top 10 differentially expressed targets were found in both primary and metastatic tumors (Table 3, A and B). Notably, 2 uncharacterized ESTs were among the 10 most differentially expressed genes in the primary tumor analysis, and 4 undefined ESTs were among the 10 most differentially expressed genes in the metastatic tumor analysis. As expected, *MUC-1, HER2/neu*, and *Zinc-α-2-glycoprotein* were overexpressed in the tumors. SAGE had detected the tag for the *Claudin-7* gene 12 times in the tumor cell lines, whereas no *Claudin-7* tags were found in the normal breast cultures. Accordingly, the array analysis showed that *Claudin-7* was nearly undetectable in the normal cells but was highly induced in both primary and metastatic tumors (Fig. 1 and Table 3). Claudin-7 is a new member of the multigene claudin family (33), which has not been implicated previously in cancers.

Notably, there was more consistency between primary and metastatic tumors with respect to underexpressed genes. Only one difference was noted in the 10 most underexpressed genes (Table 3, C and D). Among the transcripts overexpressed in the normal cells were genes involved in cell adhesion and migration [*Plakophilin-1* (34) and *Integrin α-6* (35)], a serine protease with an IGF-binding domain that is inhibited in tumors [*PRSS11* (36)], a known tumor suppressor [*MEN1* (37)], a DNA damage response gene [*Ataxia-telangiectasia* (12)] and a cell cycle regulator shown previously to be underexpressed in breast carcinoma [cyclin D2 or *CCND2* (38)]. Consistent with the SAGE analysis (Table 2), *GST-pi* was highly overexpressed in the normal breast cells (Fig. 1 and Table 3, C and D). Intriguingly, an uncharacterized EST was suppressed more than 6-fold in both primary and metastatic breast tumors. The custom array analysis detected the gene encoding a small proline-rich protein, SPR1, as the most highly differentially expressed transcript screened in the normal samples. This gene has been found in normal bronchial epithelial cells (39) and keratinocytes (40) but not in malignant counterparts.

Gene expression patterns in the primary and metastatic tumors were further analyzed by examining the individual expression levels of the most highly differential genes in each of the tumors used to screen the arrays (Table 4). Although *MUC-1* showed an average overexpression of 32-fold in the primary tumors and 23-fold in the metastatic tumors (Table 3, A and B), 5 of 7 of the primary tumors and 7 of 10 of the metastatic tumors actually expressed it >100-fold greater than normal cells. Similarly, *Claudin-7* was overexpressed >100-fold by 6 of 7 (85%) of the primary tumors and 6 of 10 (60%) of the metastatic tumors. The induction of *HER-2/neu* was <10-fold in 100% of the primary tumors and 70% of the metastatic tumors. Interestingly, *B94*, a truncated form of apolipoprotein B, showed a >3-fold induction in all but one of the primary tumors but was induced <3-fold in 60% of the metastatic tumors. Previous studies have found correlations between low expression of apolipoprotein B and both familial breast cancer (41) and breast cancer recurrence (42). An uncharacterized EST (GenBank accession number R08988) showed 3–10-fold overexpression in 43% of the primary tumors (data not shown) and 60% of the metastatic tumors (Table 4B). Finally, thrombospondin (*THBS1*), a gene known to be involved in preventing angiogenesis, was actually slightly overexpressed by each of the primary tumors (Table 4A and data not shown) but was underexpressed by 90% of the metastatic tumors (Table 4B and data not shown).

The expression patterns for genes overexpressed in the normal samples were also examined. Many of the genes showed similar repression patterns in the primary and metastatic tumors analyzed

Table 3 *SAGE-identified genes verified by array analysis*

Identity	TAG		Accession no.
A.		P/N[a]	
Mucin	CCTGGGAAGTG	32.1	N27731
Claudin-7	TATAGTCCTCT	18.3	W07258
B94	ACTCAGCCCGG	4.4	M92357
HER2/neu	AGGAAGGAACA	3.2	N22791
Neurosin	CACTCAATAAA	2.4	AF013988
Zn-α-2-GP	GGACTCTGGAG	2.3	D90427
Thrombospondin	GGCTGGGTTGC	1.9	R78085
NGAL/Lipocalin 2	TGCCCTCAGGA/C[b]	1.8	AA075896
EST	CAGGCTTCCAG	1.7	N77731
EST	ACCGCCTGTGG	1.7	W04477
B.		M/N[a]	
Mucin	CCTGGGAAGTG	23.4	N27731
HER2/neu	AGGAAGGAACA	17.5	N22791
Claudin-7	TATAGTCCTCT	14.8	W07258
EST	CAGGCTTCCAG	2.9	N77731
NGAL/Lipocalin 2	TGCCCTCAGGA/C[b]	2.8	AA076565
Cytochrome B561	TTAGCTGAGTC	2.7	U06715
B94	ACTCAGCCCGG	2.3	M92357
EST	CCCAGACTCCT	1.8	R08988
EST	GCAGTCGCTTG	1.6	W46969
EST	CGGCTGAATTC	1.4	W61355
C.		N/P[a]	
Spr1	CCCTTGAGGAG	55.57	AA447684
GST-pi	AGGTCCTAGCC	6.16	W21219
EST	GCATCTCCAGA	6.15	W30727
Ataxia-telangiectasia group D	TTGCATATCAG	5.13	W63617
Integrin α-6	GGAGGTCATCA	5.03	H13519
RIG-like 7-1	CTTTCTTTGAG	4.58	W45126
Heparin binding protein	GCCCACACAGC	3.34	AA936257
Cyclin D2	ATATAGTCAGC	3.10	N71047
MEN1 region epsilon/β	CCAGGGCAACA	3.02	N21656
Plakophilin-1	TTTGTAGAGGA	2.87	R54716
D.		N/M[a]	
Spr1	CCCTTGAGGAG	49.24	AA447684
EST	GCATCTCCAGA	6.59	W30727
GST-pi	AGGTCCTAGCC	6.52	W21219
Ataxia-telangiectasia group D	TTGCATATCAG	5.44	W63617
RIG-like 7-1	CTTTCTTTGAG	4.80	W45126
Integrin α-6	GGAGGTCATCA	4.43	H13519
Heparin binding protein	GCCCACACAGC	3.73	AA936257
Cyclin D2	ATATAGTCAGC	3.22	N71047
Serine protease PRSS11	TTTCCCTCAAA	2.91	W94696
Plakophilin-1	TTTGTAGAGGA	2.87	R54716

[a] P, primary tumors; M, metastatic tumors; N, normal breast cells. The average of the primary tumors or metastatic tumors was compared with the average of the normal samples.

[b] The 11th base of this tag sequence could not be determined unambiguously.

Table 4 *Expression of SAGE- and array-identified genes in individual tumors*

A. Identity	Accession no.	Primary *versus* Normal[a] ≤3-fold	>3- to <10-fold	10–100-fold	>100-fold
Mucin	N27731	1	0	1	5
Claudin-7	W07258	0	1	0	6
B94	M92357	1	5	1	0
HER2/neu	N22791	3	4	0	0
Neurosin	AF013988	7	0	0	0
Zn-α-2-GP	D90427	0	7	0	0
Thrombospondin	R78085	7	0	0	0
NGAL/Lipocalin 2	AA075896	5	2	0	0
EST	N77731	4	2	1	0
EST	W04477	4	3	0	0

B. Identity	Accession no.	Metastatic *versus* Normal[a] ≤3-fold	>3- to <10-fold	10–100-fold	>100-fold
Mucin	N27731	2	0	1	7
HER2/neu	N22791	4	3	2	1
Claudin-7	W07258	0	0	4	6
EST	N77731	8	0	1	1
NGAL/Lipocalin 2	AA075896	8	1	1	0
Cytochrome B561	U06715	1	8	1	0
B94	M92357	6	4	0	0
EST	R08988	3	6	1	0
EST	W46969	9	1	0	0
EST	W61355	8	2	0	0

C. Identity	Accession no.	Normal *versus* Primary[a] ≤3-fold	>3- to <10-fold	10–100-fold	>100-fold
Spr1	AA447684	0	0	6	1
GST-pi	W21219	0	6	1	0
EST	W30727	0	5	2	0
Ataxia telangiectasia group D	W63617	2	2	3	0
Integrin α-6	H13519	1	6	0	0
RIG like 7-1	W45126	3	1	3	0
Heparin binding protein	AA936257	6	1	0	0
Cyclin D2	N71047	4	2	1	0
MEN1 region epsilon/beta	N21656	7	0	0	0
Plakophilin	R54716	1	2	3	1

D. Identity	Accession no.	Normal *versus* Metastatic[a] ≤3-fold	>3- to <10-fold	10–100-fold	>100-fold
Spr1	AA447684	0	0	3	7
EST	W30727	1	4	5	0
GST-pi	W21219	2	1	5	2
Ataxia telangiectasia group D	W63617	1	2	6	1
RIG like 7-1	W45126	7	2	1	0
Integrin α-6	H13519	4	4	2	0
Heparin binding protein	AA936257	7	2	1	0
Cyclin D2	N71047	4	2	3	1
Serine protease PRSS11	W94696	7	2	1	0
Plakophilin	R54716	7	0	2	1

[a] Fold differences were determined by calculating the tumor:normal or normal:tumor ratio for each tumor (7 primary and 10 metastatic) and each normal (4 samples). Average fold-induction for each spotted target was calculated.

(Table 4, C and D). *SPR1* showed at least 10-fold repression in all of the tumors, and approximately half of all of the tumors showed an underexpression of ataxia-telangiectasia group D (*ATDC*) of at least 10-fold. Interestingly, *GST-pi* was repressed 3–10-fold in 6 of 7 primary tumors, but it was repressed more than 10-fold in 7 of 10 metastatic tumors. In contrast, the gene encoding cyclin D2 was repressed >10-fold in only 1 of 7 primary tumors but was repressed at least 10-fold in nearly half of the metastatic tumors. The uncharacterized EST (GenBank accession number R08988) was repressed >10-fold in 41% of the tumors.

Discussion

Gene expression profiles can provide powerful insights into the mechanisms regulating the transition from normal growth to cancerous proliferation. We have applied the combination of SAGE and array analysis to identify genes and pathways that may play a role in the biology of human breast tumors. The strength of combining these two methods lies in the ability to quantitatively analyze and identify an unlimited number of known and unknown genes by SAGE and then use arrays to rapidly and effectively evaluate the expression of a subset of sequences in fresh tumor tissues.

Our SAGE analysis of breast tumor cell lines and normal mammary epithelial cell cultures revealed >200 genes that were differentially expressed at least 10-fold in the normal and diseased cells. Of the 119 tags overexpressed in the breast tumors at least 10-fold, 50 corresponded to uncharacterized ESTs, and 12 tags did not match anything in either the known mRNA or EST databases, suggesting that they represent novel genes. Similarly, of the 94 genes overexpressed at least 10-fold in the normal cells, 21 are uncharacterized ESTs and 9 are potentially novel genes.

SAGE analysis identified several genes that have already been implicated in breast cancer, such as *HER2/neu*, *MUC-1*, and *FGF-1*. In the normal breast cells, SAGE detected genes thought to be involved in differentiation and growth control, such as cytokeratins *K15* and *K17*, *IGF-BP2*, and the gene encoding maspin, *PI5*. In addition, this analysis uncovered many genes and ESTs not implicated previously in breast cancer.

To validate the clinical relevance of the expression profiles seen in the cancer cell lines, a subset of the differentially expressed sequences identified by SAGE was further screened, in array format, against four normal breast epithelial cell cultures and 17 human primary and metastatic breast tumors. The 17 clinical tumors allowed us to find consistent patterns of gene expression in breast cancers. Of particular interest, *Claudin-7*, a gene cloned recently, was significantly overexpressed in all of the breast tumors examined (at least 100-fold in 85% of the primary tumors and >10-fold in all of the metastatic tumors). Claudins are proteins found in the specialized membrane domain of epithelial and endothelial cells known as tight junctions. In addition to facilitating cell-cell adhesion, tight junctions help maintain cell polarity and serve as a physical barrier to the passing of solutes and water from the paracellular space (43). The expression pattern of *Claudin-7* suggests that it plays a role in promoting or maintaining tumorigenesis, perhaps by altering adhesion properties of the cells and/or by allowing the flow of soluble growth factors or angiogenic molecules that are normally prevented from passing. Alternatively, Claudin-7 may create a barrier against angiogenesis inhibitors and/or proteins that hinder cell growth. Because *Claudin-7* is not expressed in normal breast epithelial cells but is readily detectable in all of the breast tumors examined, it may serve as a good marker of breast carcinogenesis.

Another protein thought to be involved in transport of lipophilic substances, neutrophil gelatinase-associated lipocalin (*NGAL/Lipocalin-2*), was heterogeneously overexpressed in the breast tumors examined (Tables 3 and 4), as was reported previously (44). This gene has also been detected at relatively high levels in cancers of the pancreas (45) and colon (46), and its tag was found to be overrepresented in colon cancers.[6]

Several of the genes expressed at relatively high levels in normal breast epithelial cells, including the heparin binding protein *HBp17*, *Ataxia telangiectasia group D*, and *GST-pi*, were also detected at higher levels in our analyses of normal melanocytes and nontransformed cells of the lung and colon.[7] This suggests that these genes may be involved in controlling the normal growth of multiple cell types. The tumor suppressor, maspin, however, was detected in nor-

[6] Internet address: www.ncbi.nlm.nih.gov/SAGE/, and unpublished data.
[7] Unpublished data.

mal breast, lung, and prostate epithelia but not in normal colon cells or melanocytes,[7] suggesting that this protein may have a more specialized function.

An uncharacterized EST (W30727) that was elevated >6-fold in the normal breast cells with respect to both primary and metastatic tumors may represent a breast-specific epithelial cell marker because it was not observed in the >674,000 tags analyzed by SAGE from normal and transformed cells of the lung, prostate, colon, and melanocytes.[7]

Overall, primary and metastatic breast tumors showed similar gene expression profiles. Both types of cancers showed relatively high levels of induction of the genes encoding *Mucin, Claudin-7, HER2/neu, B94, NGAL/Lipocalin 2,* and an uncharacterized EST (N77731). This EST was induced >10-fold in three of the breast tumors, with one of the metastatic tumors expressing it >100-fold higher than normal cells.

Interestingly, there were some genes more highly induced in either the primary or metastatic tumors, with several of them representing uncharacterized ESTs. These may be unknown genes involved in genome instability, cell motility, or invasion. Zinc-α-2-glycoprotein was more highly expressed in primary breast tumors than in metastatic tissues, suggesting that it may be involved in early events in the tumorigenesis pathway. Given its similarity to *MHC* genes and its proposed function in intercellular processing of antigens (47), it is interesting to speculate that this protein helps tumor cells evade the host immune system.

The profile of genes that were underexpressed in primary tumors was strikingly similar to those genes underexpressed in metastatic tumors. Of all of the genes screened, the small proline-rich protein, SPR1 (also called SPRR1b), was the most highly overexpressed in normal cells when compared with all of the tumor tissues examined. Interestingly, previous work has shown that SPR1 is a specific marker for squamous lung carcinoma (48). Accordingly, the tag corresponding to *SPR1* was detected in SAGE libraries of squamous lung tumors but not in the >674,000 tags analyzed in libraries generated from malignancies of the colon, pancreas, prostate, or melanocytes.[7] SPR1 is an envelope protein found commonly in epithelial cells and believed to be involved in cross-bridging. Therefore, SPR1 may contribute to the mechanical properties of cell envelopes (49), and loss of this protein may promote tumorigenesis by facilitating invasion or loss of contact inhibition.

Overall, there were more differences seen in genes induced when comparing primary and metastatic tumors than when comparing normal cells with either tumor type. This suggests that the transition from normal regulated growth to cancer, and specifically from a primary tumor to a metastatic state, may require a positive, growth-promoting transcriptional program to be initiated. Accordingly, genes involved in maintaining normal regulated growth would be more consistently expressed with respect to either tumor type.

Combining two powerful gene expression profiling technologies allowed the rapid and efficient identification of many clinically relevant genes potentially involved in preventing or promoting breast cancer. The data from our SAGE analysis of tumor cell lines and array analysis using clinical tumors indicated some differences in gene expression profiles. This underscores the importance of applying a level of screening that includes clinical tumor samples. This broad study has identified several genes previously unknown to be involved in breast carcinogenesis. Many of these genes may be useful as diagnostic markers or prognostic indicators. In addition, studying the function of some of these genes may uncover new pathways for therapeutic intervention.

Acknowledgments

We thank Dr. Walter Demkowicz for providing BC10TN, a primary breast tumor used to screen the custom arrays. We are grateful to Dana Barberio and Lesley Michalowsky for help in the custom array analysis.

References

1. Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. Serial analysis of gene expression. Science (Washington DC), *270:* 484–487, 1995.
2. Prashar, Y., and Weissman, S. M. READS: a method for display of 3′-end fragments of restriction enzyme- digested cDNAs for analysis of differential gene expression. Methods Enzymol., *303:* 258–272, 1999.
3. Bachem, C. W., van der Hoeven, R. S., de Bruijn, S. M., Vreugdenhil, D., Zabeau, M., and Visser, R. G. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J., *9:* 745–753, 1996.
4. Drmanac, S., Stavropoulos, N. A., Labat, I., Vonau, J., Hauser, B., Soares, M. B., and Drmanac, R. Gene-representing cDNA clusters defined by hybridization of 57,419 clones from infant brain libraries with short oligonucleotide probes. Genomics, *37:* 29–40, 1996.
5. Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E. L. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol., *14:* 1675–1680, 1996.
6. Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D., Jr., Hieter, P., Vogelstein, B., and Kinzler, K. W. Characterization of the yeast transcriptome. Cell, *88:* 243–251, 1997.
7. Polyak, K., Xia, Y., Zweier, J. L., Kinzler, K. W., and Vogelstein, B. A model for p53-induced apoptosis. Nature (Lond.), *389:* 300–305, 1997.
8. Hibi, K., Robinson, C. R., Booker, S., Wu, L., Hamilton, S. R., Sidransky, D., and Jen, J. Serial analysis of gene expression in non-small cell lung cancer. Cancer Res., *58:* 5690–5694, 1998.
9. Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B., and Kinzler, K. W. Gene expression profiles in normal and cancer cells. Science (Washington DC), *276:* 1268–1272, 1997.
10. Band, V., Zajchowski, D., Swisshelm, K., Trask, D., Kulesa, V., Cohen, C., Connolly, J., and Sager, R. Tumor progression in four mammary epithelial cell lines derived from the same patient. Cancer Res., *50:* 7351–7357, 1990.
11. Chomczynski, P., and Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate- phenol-chloroform extraction. Anal. Biochem., *162:* 156–159, 1987.
12. Welch, D. R., and Wei, L. L. Genetic and epigenetic regulation of human breast cancer progression and metastasis. Endocrine-Related Cancer, *5:* 155–197, 1998.
13. Aoki, R., Tanaka, S., Haruma, K., Yoshihara, M., Sumii, K., Kajiyama, G., Shimamoto, F., and Kohno, N. MUC-1 expression as a predictor of the curative endoscopic treatment of submucosally invasive colorectal carcinoma. Dis. Colon Rectum, *41:* 1262–1272, 1998.
14. Segal Eiras, A., and Croce, M. V. Breast cancer associated mucin: a review. Allergol. Immunopathol., *25:* 176–181, 1997.
15. Diez Itza, I., Sanchez, L. M., Allende, M. T., Vizoso, F., Ruibal, A., and Lopez-Otin, C. Zn-α2-glycoprotein levels in breast cancer cytosols and correlation with clinical, histological and biochemical parameters. Eur. J Cancer, *9:* 1256–1260, 1993.
16. Bussemakers, M. J., van de Ven, W. J., Debruyne, F. M., and Schalken, J. A. Identification of high mobility group protein I(Y) as potential progression marker for prostate cancer by differential hybridization analysis. Cancer Res., *51:* 606–611, 1991.
17. Giancotti, V., Buratti, E., Perissin, L., Zorzet, S., Balmain, A., Portella, G., Fusco, A., and Goodwin, G. H. Analysis of the HMGI nuclear proteins in mouse neoplastic cells induced by different procedures. Exp. Cell Res., *184:* 538–545, 1989.
18. Chiappetta, G., et al. The expression of the high mobility group HMGI (Y) proteins correlates with the malignant phenotype of human thyroid neoplasias. Oncogene. 1995. 10(7): p. 1307–14.
19. Ram, T. G., Reeves, R., and Hosick, H. L. Elevated high mobility group-I(Y) gene expression is associated with progressive transformation of mouse mammary epithelial cells. Cancer Res., *53:* 2655–2660, 1993.
20. Kaina, B., Fritz, G., and Coquerelle, T. Contribution of O^6-alkylguanine and *N*-alkylpurines to the formation of sister chromatid exchanges, chromosomal aberrations, and gene mutations: new insights gained from studies of genetically engineered mammalian cell lines. Environ. Mol. Mutagen., *22:* 283–292, 1993.
21. Brotherick, I., Robson, C. N., Bronell, D. A., Shentine, J., White, M. D., Cunliffe, W. J., Shenton, B. K., Egan, M., Webb, L. A., Lunt, L. J., Young, J. R., and Higgs, M. Cytokeratin expression in breast cancer: phenotypic changes associated with disease progression. Cytometry, *32:* 301–308, 1998.
22. Stetler Stevenson, W. G., Aznavoorian, S., and Liotta, L. A. Tumor cell interactions with the extracellular matrix during invasion and metastasis. Annu. Rev. Cell Biol., *9:* 541–573, 1993.
23. Sager, R., Sheng, S., Pemberton, P., and Hendrix, M. J. Maspin: a tumor suppressing serpin. Adv. Exp. Med. Biol., *425:* 77–88, 1997.
24. Manni, A., Badger, B., Wei, L., Zaenglein, A., Grove, R., Khin, S., Heitjan, D., Shimasaki, S., and Ling, N. Hormonal regulation of insulin-like growth factor II and insulin-like growth factor binding protein expression by breast cancer cells *in vivo*: evidence for stromal epithelial interactions. Cancer Res., *54:* 2934–2942, 1994.
25. Nickerson, T., Zhang, J., and Pollak, M. Regression of DMBA-induced breast carcinoma following ovariectomy is associated with increased expression of genes

encoding insulin-like growth factor binding proteins. Int. J. Oncol., *14:* 987–990, 1999.
26. Lee, S. W., Tomasetto, C., Swisshelm, K., Keyomarsi, K., and Sager, R. Down-regulation of a member of the S100 gene family in mammary carcinoma cells and reexpression by azadeoxycytidine treatment. Proc. Natl. Acad. Sci. USA, *89:* 2504–2508, 1992.
27. Franzen, B., Linder, S., Alaiya, A. A., Eriksson, E., Uruy, K., Hirano, T., Okuzawa, K., and Auer, G. Analysis of polypeptide expression in benign and malignant human breast lesions: down-regulation of cytokeratins. Br. J. Cancer, *74:* 1632–1638, 1996.
28. Lee, W. H., Morton, R. A., Epstein, J. I., Brooks, J. D., Campbell, P. A., Bova, G. S., Hsieh, W. S., Isaacs, W. B., and Nelson, W. G. Cytidine methylation of regulatory sequences near the pi-class glutathione *S*-transferase gene accompanies human prostatic carcinogenesis. Proc. Natl. Acad. Sci. USA, *91:* 11733–11737, 1994.
29. Esteller, M., Corn, P. G., Urena, J. M., Gabrielson, E., Baylin, S. B., and Herman, J. G. Inactivation of glutathione *S*-transferase P1 gene by promoter hypermethylation in human neoplasia. Cancer Res., *58:* 4515–4518, 1998.
30. Magnaldo, T., Fowlis, D., and Darmon, M. Galectin-7, a marker of all types of stratified epithelia. Differentiation, *63:* 159–168, 1998.
31. Franzen, B., Linder, S., Uryu, K., Alaiya, A. A., Hirano, T., Kato H., and Auer, G. Expression of tropomyosin isoforms in benign and malignant human breast lesions. Br. J. Cancer, *73:* 909–913, 1996.
32. Bergstraesser, L. M., Srinivasan, G., Jones, J. C., Stahl, S., and Weitzman, S. A. Expression of hemidesmosomes and component proteins is lost by invasive breast cancer cells. Am. J. Pathol., *147:* 1823–1839, 1995.
33. Morita, K., Furuse, M., Fujimoto, K., and Tsukita, S. Claudin multigene family encoding four-transmembrane domain protein components of tight junction strands. Proc. Natl. Acad. Sci. USA, *96:* 511–516, 1999.
34. Moll, I., Kurzen, H., Langbein, L., and Franke, W. W. The distribution of the desmosomal protein, plakophilin 1, in human skin and skin tumors. J. Investig. Dermatol., *108:* 139–146, 1997.
35. Shaw, L. M., Chao, C., Wewer, U. M., and Mercurio, A. M. Function of the integrin $\alpha 6\beta 1$ in metastatic breast carcinoma cells assessed by expression of a dominant-negative receptor. Cancer Res., *56:* 959–963, 1996.
36. Zumbrunn, J., and Trueb, B. Primary structure of a putative serine protease specific for IGF-binding proteins. FEBS Lett., *398:* 187–192, 1996.
37. Chandrasekharappa, S. C., Guru, S. C., Manickam, P., Olujemi, S. E., Collins, F. S., Emmert-Buck, M. R., Debelenko, I. V., Zhuang, Z., Lubensky, I. A., Liotta, L. A., Crabtree, J. S., Wang, Y., Roe, B. A., Weisemann, J., Boguski, M. S., Agarwal, S. K., Kester, M. B., Kim, Y. S., Heppner, C., Dong, Q., Spiegel, A. M., Burns, A. L., and Marx, S. J. Positional cloning of the gene for multiple endocrine neoplasia-type 1. Science (Washington DC), *276:* 404–407, 1997.
38. Zhou, Q., Stetler-Stevenson, M., and Steeg, P. S. Inhibition of cyclin D expression in human breast carcinoma cells by retinoids *in vitro*. Oncogene, *15:* 107–115, 1997.
39. DeMuth, J. P., Weaver, D. A., Crawford, E. L., Jackson, C. M., and Willey, J. C. Loss of spr1 expression measurable by quantitative RT-PCR in human bronchogenic carcinoma cell lines. Am. J. Respir. Cell Mol. Biol., *19:* 25–29, 1998.
40. Yaar, M., Gilani, A., DiBenedetto, P. J., Harkness, D. D., and Gilchrest, B. A. Gene modulation accompanying differentiation of normal *versus* malignant keratinocytes. Exp. Cell Res., *206:* 235–243, 1993.
41. Boyd, N. F., Connelly, P., Lynch, H., Knaus, M., Michal, S., Fili, M., Martin, L. J., Lockwood, G., and Tritchler, D. Plasma lipids, lipoproteins, and familial breast cancer. Cancer Epidemiol. Biomark. Prev., *4:* 117–122, 1995.
42. Lane, D. M., Boatman, K. K., and McConathy, W. J. Serum lipids and apolipoproteins in women with breast masses. Breast Cancer Res. Treat., *34:* 161–169, 1995.
43. Gumbiner, B. M. Breaking through the tight junction barrier. J. Cell Biol., *123* (6 Pt. 2)*:* 1631–1633, 1993.
44. Stoesz, S. P., Friedl, A., Haag, J. D., Lindstrom, M. J., Clark, G. M., and Gould, M. N. Heterogeneous expression of the lipocalin NGAL in primary breast cancers. Int. J. Cancer, *79:* 565–572, 1998.
45. Furutani, M., Arii, S., Mizumoto, M., Kato, M., and Imamura, M. Identification of a neutrophil gelatinase-associated lipocalin mRNA in human pancreatic cancers using a modified signal sequence trap method. Cancer Lett., *122:* 209–214, 1998.
46. Nielsen, B. S., Borregaard, N., Bundgaard, J. R., Timshel, S., Sehested, M., and Kjeldsen, L. Induction of NGAL synthesis in epithelial cells of human colorectal neoplasia and inflammatory bowel diseases. Gut, *38:* 414–420, 1996.
47. Freije, J. P., Fueyo, A., Uria, J. A., Velasco, G., Sanchez, L. M., Lopez-Boado, Y. S., and Lopez-Otin, C. Human Zn-α2-glycoprotein. complete genomic sequence, identification of a related pseudogene and relationship to class I major histocompatibility complex genes. Genomics, *18:* 575–587, 1993.
48. Hu, R., Wu, R., Deng, J., and Lau, D. A small proline-rich protein, spr1: specific marker for squamous lung carcinoma. Lung Cancer, *20:* 25–30, 1998.
49. Kartasova, T., Darwiche, N., Kohno, Y., Koizumi, H., Osada, S., Huh, N., Lichti, U., Steinert, P. M., and Kuroki, T. Sequence and expression patterns of mouse SPR1: correlation of expression with epithelial function. J. Investig. Dermatol., *106:* 294–304, 1996.