

# The Desmoplastic Response to Infiltrating Breast Carcinoma: Gene Expression at the Site of Primary Invasion and Implications for Comparisons between Tumor Types<sup>1</sup>

Christine A. Iacobuzio-Donahue, Pedram Argani, Paula M. Hemen, Jessa Jones, and Scott E. Kern<sup>2</sup>

Departments of Pathology [C. A. I-D., P. A.] and Oncology [P. M. H., J. J., S. E. K.], The Johns Hopkins Medical Institutions, Baltimore, Maryland 21231

## ABSTRACT

The gene expression patterns of desmoplasia are becoming exposed through the application of global gene expression technologies such as cDNA microarrays or serial analysis of gene expression (SAGE). These patterns represent the sum of the many cellular components of the host stromal response to an infiltrating carcinoma. In studies of human neoplasms, it would be useful to identify those prototypical genes that characteristically indicate the recognizable forms of the responses to individual tumor types. Such genes may offer clues to better understand the process of invasion itself, the interactions between tumor and host cells, and tumor-specific differences in invasion. We used SAGE-defined genes and *in situ* transcript labeling to characterize the desmoplastic stroma induced by infiltrating ductal carcinomas of the breast. Principal component analysis identified 103 SAGE tags as specific for invasive breast carcinomas, in comparison with *in situ* duct carcinomas or normal breast epithelium. Of these, 68 tags corresponded to known genes. Six of the 68 genes from this breast cancer “invasion-specific” cluster were further characterized by *in situ* hybridization to breast cancer tissues. Results of *in situ* hybridization demonstrated that each gene was expressed within one of five distinct regions of the invasive tumors (neoplastic epithelium; angioendothelium; inflammatory, panstromal, and juxtatumoral stroma), reflecting a defined architectural structure to the transcriptome of invasive breast cancers. Two of these 6 genes were specifically expressed by the stromal cells within the invasive carcinoma; however, 1 (*collagen 1 $\alpha$ 1*) was expressed throughout the stromal response (panstromal expression), whereas the second (*osteonectin*) was specifically expressed within the juxtatumoral stromal cells, indicating a critical “regionality” of gene expression within the stromal response itself. A comparison of the gene expression profiles of the juxtatumoral stroma in breast and pancreatic carcinomas indicated important differences between the two, suggesting tumor-specific or organ-specific differences in the desmoplastic responses. Some of the genes presented are novel markers of the invasive process, imply communication at the host/tumor interface, and suggest potential therapeutic targets.

## INTRODUCTION

The ability of malignant neoplasms to invade adjacent normal tissues is fundamental to the neoplastic process. For many epithelial neoplasms, this process includes the ability to induce a desmoplastic response of the host tissues at the site of primary invasion. This host stromal response is a result of a complex interaction between the host and invading neoplasm, comprising fibroblasts, various inflammatory cells, proliferating vascular structures, as well as normal parenchymal cells undergoing atrophy at the invasive edge.

Recent investigations into the host desmoplastic response to infil-

trating pancreatic adenocarcinoma have identified or refined an architectural organization of gene expression within this host response (1, 2). Ryu *et al.* (1) identified an invasion-specific cluster of genes when comparing SAGE<sup>3</sup> libraries of primary carcinomas to those derived from passaged cancer cell lines. Many of the genes identified were found to be markers of the exuberant host stromal response present in infiltrating pancreatic cancers, representing distinct classes of genes with differing cellular functions. *In situ* hybridization, using 12 of these invasion-specific genes as probes, illustrated how gene expression patterns are partitioned into spatial compartments within the desmoplastic response to the tumor cells, including a distinct “juxtatumoral stroma,” a region of the host response thought to be important for tumor-host interactions (2). These genes, and their associated architectural compartments of gene expression, represent potential new targets for diagnostic screening or for therapeutic development.

We proposed that similar study of the desmoplastic response to an infiltrating carcinoma might be useful in understanding the molecular biology of other tumor types such as infiltrating duct carcinomas of the breast, which also characteristically can produce a prominent host stromal response (3). We applied principal component analysis to a set of SAGE libraries generated from normal and neoplastic breast cancer tissues and cell lines and characterized those genes identified by *in situ* hybridization: (a) we determined whether the gene expression characteristic of the desmoplastic response to breast cancer is similar to that found in response to other epithelial neoplasms; (b) we examined whether spatially defined regions of gene expression exist among the desmoplastic responses to breast cancers, with specific attention to the juxtatumoral stroma; and (c) we sought to determine whether specific genes potentially important in the desmoplastic response to one tumor type may play a similar role in other epithelial neoplasms as well.

## MATERIALS AND METHODS

**SAGE Data.** SAGE data of 11 breast-derived libraries were acquired from the CGAP database available in the NCBI SAGEmap database.<sup>4</sup> The SAGE libraries of the two normal tissues, two ductal carcinomas *in situ*, and two invasive cancers and their matched lymph node metastases were prepared and sequenced as described in detail by Porter *et al.* (4). Breast cancer cell line libraries LacZ, MCF70 h, and MDA453 were included in the analyses to aid in determination of invasion-specific gene expression. A total of 88,178 unique SAGE tags, which were identified among 467,742 total tags sequenced from 11 breast SAGE libraries, were used for all subsequent analyses.

**Biocomputational Tools.** The Cluster and TreeView computer programs were obtained from the online resource<sup>5</sup> and used for PCA and visualization of tree diagrams (5). SAGE data were filtered as follows. Exclusion was applied to tags when fewer than two samples contained at least 5 tags in the raw data and when minimum and maximum values among all samples differed by <4 tags. This produced a dataset of 2,575 tags from an original 88,178 unique tags.

Received 4/17/02; accepted 7/19/02.

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.

<sup>1</sup> Supported by the NIH Specialized Programs of Research Excellence in Gastrointestinal Cancer Grant CA 62924 (to S. E. K.) and the NIH Specialized Programs of Research Excellence in Breast Cancer Grant CA88843.

<sup>2</sup> To whom requests for reprints should be addressed, at Department of Oncology, Room 461, Cancer Research Building, 1650 Orleans Street, The Johns Hopkins University School of Medicine, Baltimore, MD 21231. Phone: (410) 614-3316; Fax: (410) 614-9705; E-mail: sk@jhmi.edu.

<sup>3</sup> The abbreviations used are: SAGE, serial analysis of gene expression; NCBI, National Center for Biotechnology Information; PCA, principal component analysis; EST, expressed sequence tag; DCIS, ductal carcinoma *in situ*; IGF1BP, insulin-like growth factor binding protein.

<sup>4</sup> Internet address: <http://www.ncbi.nlm.nih.gov/SAGE/>.

<sup>5</sup> Internet address: <http://www.microarrays.org/software.html>.

The data were imported into the Cluster program and log-transformed, and PCA was performed. The names of genes and ESTs that matched the tag sequences were obtained using an online resource from NCBI.<sup>6</sup>

**Tissues.** Paraffin-embedded tissues of four samples of infiltrating duct carcinoma of the breast were obtained from the files of The Johns Hopkins Hospital. For each case, one representative section was chosen that contained invasive carcinoma and normal duct and lobule structures on the same slide. Three of the four cases also contained high-grade DCIS within the same paraffin section. All four carcinomas were Elston grade II/III (6).

**Nonradioactive *in Situ* Hybridization of Paraffin Sections.** To generate riboprobes for use in *in situ* hybridization of genes of interest, DNA templates were generated by PCR with incorporation of a T7 promoter into the antisense or sense primer (7). After phenol:chloroform purification of amplified DNA, 200 ng of the DNA templates were used to generate either antisense or sense riboprobes by *in vitro* transcription with digoxigenin labeling reagents and T7 polymerase according to the manufacturer's protocol (Roche Diagnostics, Indianapolis, IN).

*In situ* hybridization of paraffin-embedded tissues were performed following methods modified from Kadkol *et al.* (8). Five- $\mu$ -thick sections were cut from the paraffin blocks, deparaffinized in xylene, and hydrated in graded concentrations of ethanol for 5 min each. Sections were incubated with 1% hydrogen peroxide, followed by digestion in 10  $\mu$ g/ml of proteinase K at 37°C for 30 min. Sections were hybridized overnight at 15–25°C below the  $T_m$  calculated for each individual riboprobe with a 200 ng/ml dilution of either antisense or sense riboprobes in mRNA hybridization buffer (DAKO, Carpinteria, CA). The following day, sections were washed in 2 $\times$  SSC (0.3 M sodium chloride and 0.03 M sodium citrate) and incubated with a 1:35 dilution of RNase A cocktail (Ambion, Austin, TX) in 2 $\times$  SSC for 37°C. Next, sections were stringently washed in 2 $\times$  SSC/50% formamide twice, followed by one wash at 0.08 $\times$  SSC at 5–8°C below the calculated  $T_m$ . For signal amplification, a horseradish peroxidase-conjugated rabbit anti-digoxigenin antibody (DAKO) was used to catalyze the deposition of biotinyl-tyramide, followed by secondary streptavidin complex (GenPoint kit; DAKO). The final signal was developed with 3,3'-diaminobenzidine chromagen (GenPoint kit; DAKO), and the tissue was counterstained in hematoxylin for 15 s.

**Histological Evaluation of Tissue Sections.** *In situ* hybridization labeling of mRNA expression in samples of paraffin-embedded pancreatic carcinoma was evaluated by three of the authors (C. A. I-D., P. A., and S. E. K.) with agreement in all cases examined. For each case, the labeling pattern obtained following *in situ* hybridization was evaluated for the presence or absence of gene expression individually within the normal duct epithelium, duct carcinoma *in situ* (if present), and infiltrating duct carcinoma. Expression was also evaluated within the desmoplastic stroma of the neoplasm and the vasculature within the normal tissues and tumor mass. In those cases having positive expression noted within the tumor stroma, gene expression was scored as occurring within the entire stromal region of the tumor or in the stroma immediately adjacent to tumor epithelium (juxtatumoral stroma).

## RESULTS AND DISCUSSION

**Hierarchical Clustering.** One-way hierarchical clustering was used to examine the relationships among the 11 SAGE libraries samples based on their global gene expression profiles (Fig. 1). A dendrogram created by this analysis indicated that breast cancer tissues and their matched metastases were more similar to each other than to other breast cancer tissues or cell lines. Samples of normal breast epithelium also clustered together on a terminal branch of the dendrogram, and samples of DCIS (DCIS and DCIS2) were arranged between these two groups. The hierarchical clustering analysis also confirmed the expected dissimilarity of gene expression between the breast tissue libraries and breast cancer cell lines. Two of the three cancer cell lines clustered together on an independent arm of the dendrogram. However, one cell line, MCF70h, clustered more closely to the tissue-derived breast samples than to the other cell lines analyzed. These findings are in agreement with similar reports (4, 9) but

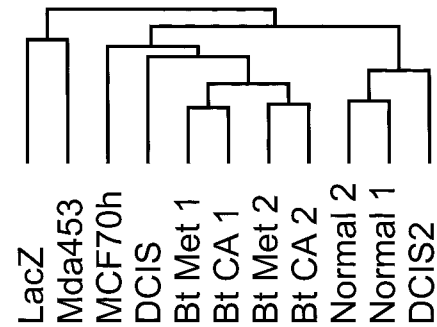


Fig. 1. Relationships among samples shown by clustering of gene expression. *LacZ*, *Mda453*, and *MCF70h* are breast cancer cell lines; *Normal 1* and *Normal 2* are two samples of normal breast duct epithelium; *DCIS* and *DCIS2* are samples of DCIS; *Bt CA 1* and *Bt Met 1*, *Bt CA 2* and *Bt Met 2* are two sets of matched infiltrating duct carcinoma and lymph node metastasis. Spearman rank correlation complete linkage clustering was used to produce tree diagrams.

also formed the basis for additional analyses to identify invasion-specific gene expression in breast cancers.

**Principal Component Analysis.** PCA can provide a global overview of the relatedness of gene expression profiles among samples while better avoiding the deterministic and rather arbitrary nature of hierarchical clustering. We therefore used PCA to delineate the gene cluster that distinguished the invasive breast cancer specimens from all others in this SAGE library dataset (Fig. 2). A cluster of genes specific for, and highly expressed in, invasive breast cancer tissues (primary carcinomas and matched metastases) was identified. This cluster thus identified the "invasion-specific genes" of breast carcinomas, as defined by comparison with breast cancer cell lines or samples of DCIS. This gene cluster is not to be confused with tumor-specific genes (present in both invasive breast cancer tissues and cell lines) but instead includes the gene expression associated with the presence of the host stromal response present within samples of infiltrating duct carcinoma. These genes may represent stromal gene expression or the expression of certain genes within the neoplastic epithelium as a function of tumor-stromal interactions (10, 11).

A smaller yet distinct cluster of genes was also identified to correspond to normal duct epithelium, similar to that described previously in a normal breast duct epithelium gene cluster (4). Breast cancer cell lines and DCIS samples were less well delineated from the other samples and did not show distinct gene clusters by PCA. However, because our goal was to investigate the host response to breast cancer, we directed our attention solely to the invasion-specific gene cluster.

**Genes Characteristic of Invasive Breast Cancer.** Table 1 contains the identities of the SAGE tags and their frequency of appearance in the invasion-specific gene cluster of breast duct carcinoma. Among 103 tags, 68 matched to known transcripts, and 35 might include novel genes. A comparison of the genes identified in our invasion-specific cluster to that of Porter *et al.* (4) revealed several similarities, with 15 of the genes identified in their analysis also being identified by our methods (Table 1). However, several genes were identified in our cluster that were not reported (*e.g.*, *thymosin  $\beta$ 4*, *apolipoprotein E*, *laminin receptor 1*, and *IGFBP7*), suggesting the contrasting of primary tumors to breast cancer cell lines and analysis by PCA may be more appropriate for the identification of invasion-specific gene expression.

Infiltrating carcinomas are often associated with a dense fibrous host stromal reaction to the neoplasm, known as desmoplasia. At the advancing edge of the infiltrating carcinoma, there is often entrapment of normal duct and lobule structures, as well as foci of residual duct carcinoma *in situ* present within the original site of neoplastic forma-

<sup>6</sup> Internet address: <http://www.ncbi.nlm.nih.gov/SAGE/SAGetag.cgi>.

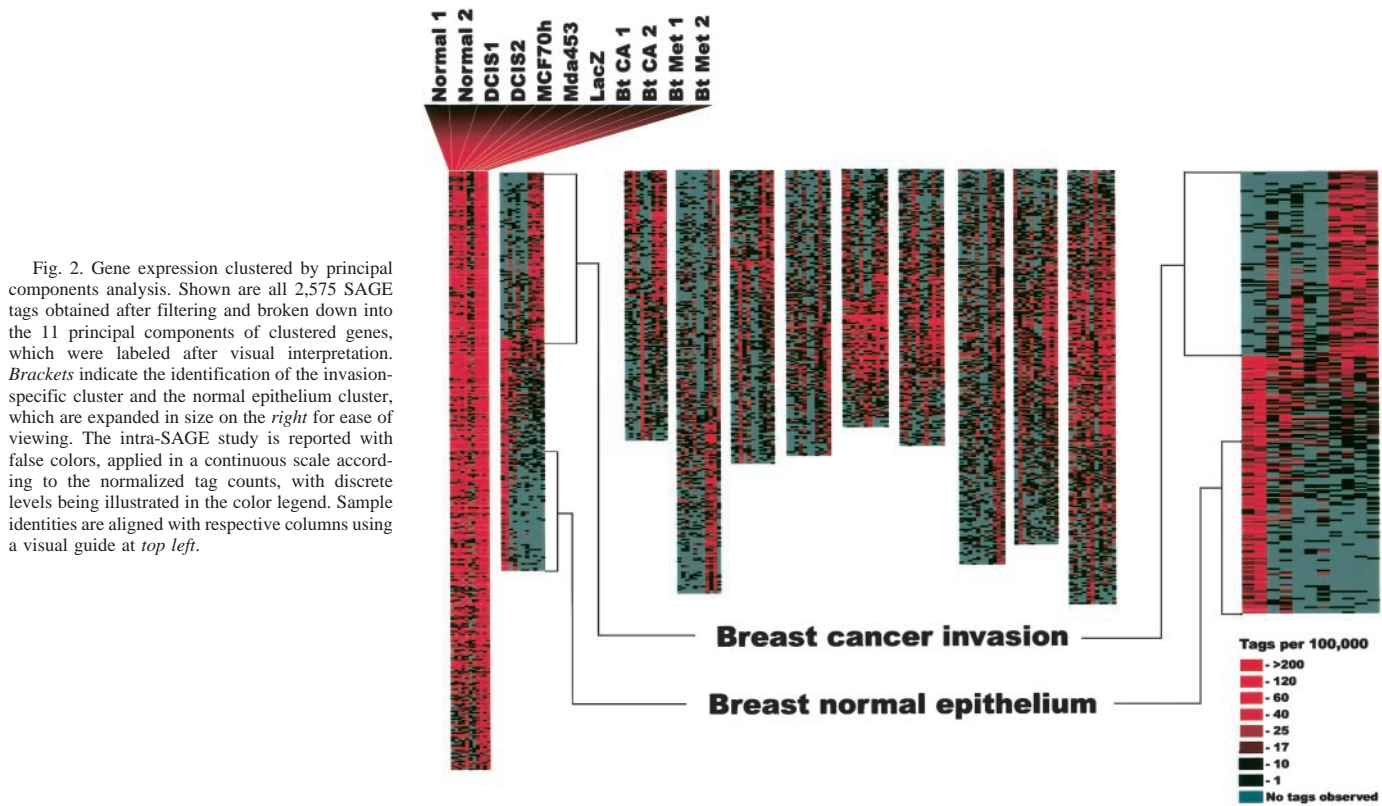


Fig. 2. Gene expression clustered by principal components analysis. Shown are all 2,575 SAGE tags obtained after filtering and broken down into the 11 principal components of clustered genes, which were labeled after visual interpretation. Brackets indicate the identification of the invasion-specific cluster and the normal epithelium cluster, which are expanded in size on the right for ease of viewing. The intra-SAGE study is reported with false colors, applied in a continuous scale according to the normalized tag counts, with discrete levels being illustrated in the color legend. Sample identities are aligned with respective columns using a visual guide at top left.

tion. Inflammatory cells may also represent a proportion of the cellularity of the mass and are usually found at the advancing edges of the neoplasm as it invades through normal structures. The genes identified within this invasion-specific cluster are therefore best categorized with an understanding of the variety of cell types that constitute the primary site of invasion within breast carcinomas.

Genes identified within this cluster reflected the presence of various components of the host stromal response, including extracellular matrix remodeling (e.g., *collagen 1 $\alpha$ 1*; Ref. 12), angiogenesis (e.g., *IGFBP7* and *osteonectin*; Ref. 13), the immune response (e.g., *immunoglobulin heavy chain  $\gamma$ 3*; Refs. 9, 14), increased proliferation (*cdk inhibitor 3* and *SMC4-like 1*; Refs. 9, 14), or elevated transcriptional demands (*ribosomal proteins*; Ref. 15). Relatively few genes identified within this breast invasion-specific gene cluster, however, were also present within the invasion-specific gene cluster characteristic of pancreatic cancer (1). Genes that were found in both invasion-specific gene clusters included *apolipoprotein C-1*, *osteonectin*, and *collagen 1 $\alpha$ 1*, suggesting that some genes may play a universal role in the host stromal response to infiltrating cancer. However, most genes identified in the breast invasion-specific gene cluster were not identified in the invasion-specific cluster of the pancreas, and *vice versa*. It was thus possible that invasion-specific gene expression might relate to the primary organ in which the host stromal response occurs. Alternatively, because the identification of invasion-specific genes by PCA predominantly reflects quantitative changes in expression, the differences in desmoplastic gene expression between these two tumor types might primarily reflect the relatively more exuberant host stromal response to pancreas cancer, with a more cellular host response perhaps being represented in SAGE libraries from those tumors. These possibilities were addressed by *in situ* studies.

***In situ* Hybridization of Selected Invasion-specific Genes.** Because invasive breast carcinomas represent an aggregate of diverse cell types, the precise cellular origin of these transcripts cannot be

determined without additional study. To define the cellular origin and patterns of expression of these genes associated with the host stromal response to breast cancer, 6 genes were selected for further study of their expression in invasive breast carcinoma tissues by *in situ* hybridization (Fig. 3). These gene expression markers were selected for their presumed role in the host stromal response, such as new vessel formation (*IGFBP7* and *osteonectin*; Ref. 13), fibroblastic proliferation (*collagen 1 $\alpha$ 1* and *apolipoprotein C-1*; Ref. 1), extracellular matrix remodeling (*collagen 1 $\alpha$ 1* and *laminin receptor 1*; Refs. 1, 16), or the inflammatory response (*fusin*; Ref. 17).

*In situ* hybridization was performed for each of the 6 invasion-specific genes on four paraffin-embedded tissue samples obtained from mastectomy specimens removed for infiltrating duct carcinoma of the breast. Detectable expression of all 6 genes was observed in all four neoplasms. For each gene, detectable expression was found to localize to one or more of five distinct architectural regions, or "gene expression compartments," of the invasive tumors: (a) neoplastic epithelium; (b) angioendothelium and/or vascular smooth muscle; (c) juxtatumoral stroma (i.e., only those stromal cells immediately adjacent to the invasive neoplastic epithelium); (d) panstromal tissue (i.e., all areas of stromal tissue of the invasive tumor); or (e) inflammatory cells within the invasive focus.

Four of the 6 genes were expressed within a single architectural compartment in the four samples of invasive cancer (Fig. 3). Expression of *laminin receptor 1* was localized to neoplastic epithelium, with no additional expression noted in the surrounding stromal or angioendothelial compartments. In contrast, *collagen 1 $\alpha$ 1* gene expression was observed throughout the stromal response (panstromal), whereas the neoplastic epithelial and angioendothelial compartments were negative for expression of this gene. Finally, the gene expression of *fusin* and *apolipoprotein C-1* was observed within leukocytic (inflammatory) cells infiltrating within the invasive carcinomas. *Fusin* gene expression was predominantly within small lymphocytes, whereas

Table 1 Breast cancer invasion-specific cluster: Genes and frequencies among breast tissue libraries<sup>a</sup>

SAGE tag	Gene	Individual SAGE data										
		N1 <sup>b</sup>	N2	Dcis1	Dcis2	CL1	CL2	CL3	I1	M1	I2	M2
<b>Known genes</b>												
tcactcaaa	<i>Similar to Zn finger protein 83</i>	0	20	24	0	16	52	0	118	132	177	221
gcaccctcag	<i>Reg. of G-protein signalling 3</i>	0	40	0	34	0	0	0	59	99	101	110
gaagcaataa	<i>MHC II DQ α1/sialyltransferase 6</i>	26	0	0	0	16	0	0	74	82	101	177
ttttactcac	<i>Adducin 1 (α)</i>	26	0	24	0	16	0	0	89	66	101	132
gaccacaccg	<i>β1,4-Galactosyltransferase</i>	26	20	24	0	16	0	53	237	99	304	376
aagcacaata	<i>EST/tyrosine kinase bp</i>	0	0	24	0	0	0	0	133	82	228	154
agctgtttct	<i>Ribosomal protein L13a</i>	53	20	48	69	49	0	107	133	115	228	221
ttacgatgaa	<i>Phosphatidylinositol-4-phosphate 5-kinase, II, β</i>	0	20	0	69	16	0	0	461	99	582	796
tgagcaagcc	<i>Zinc finger protein 278</i>	0	0	72	0	32	0	0	118	181	202	154
gttgacactg	<i>Mitochondrial Ts translation elongation factor</i>	0	0	24	0	49	0	0	74	115	177	154
gcttatagtc	<i>Histidine triad nucleotide-bp</i>	26	0	48	69	49	52	53	104	99	228	265
tgccccagg	<i>Apolipoprotein C-I<sup>c</sup></i>	26	20	97	0	32	0	0	654	314	988	907
aatagggcca	<i>Amyloid β (A4) P1P2</i>	0	0	72	34	32	52	107	163	99	126	199
gagtttaaaa	<i>MHCII, DR β</i>	0	0	0	0	16	0	0	59	82	101	332
cagataacat	<i>Translocase of outer mitochondrial membrane 20 (Yeast) homologue</i>	0	0	0	0	49	52	53	89	82	152	88
actaatcggt	<i>Ribosomal protein L37</i>	26	0	24	0	16	0	0	59	82	329	221
gttgtctttg	<i>FLJ20003/complement C3</i>	53	0	97	0	32	52	53	297	348	253	221
ttcccttctt	<i>MHC II, DP β</i>	0	20	72	0	0	0	0	237	82	126	110
gtgtgtttgt	<b>TGF-β induced, 68 kDa</b>	0	0	97	0	0	0	0	118	115	152	265
caaaactaac	<i>Ig heavy constant mu</i>	0	0	121	0	0	0	0	788	148	1950	3607
cgacccacag	<i>Apolipoprotein E</i>	26	40	145	0	0	0	0	416	314	329	354
gcttgatctc	<i>FK506 binding protein 12</i>	26	40	72	69	0	0	0	89	214	101	88
catatcatta	<i>IGFBP7</i>	0	0	121	69	16	0	0	713	132	532	619
gtgaaacatt	<i>Ribosomal protein S6 kinase, polypeptide 1</i>	53	20	24	69	49	0	53	104	82	76	88
ggaagtgtac	<i>ATPase, H+ transporting, lysosomal, subunit 1</i>	0	0	24	69	81	52	0	252	264	126	132
caagtctctt	<i>Thymosin, β4, X chromosome</i>	0	20	121	34	0	0	0	208	264	202	132
ccttatattt	<i>Tetrapeptide repeat domain 3</i>	0	0	24	0	65	105	53	118	66	228	243
gaaataaacg	<b>Ig heavy constant γ 3</b>	0	0	630	0	0	0	0	7450	1190	8157	16091
aaacccaat	<i>Ig λ locus</i>	0	0	169	0	0	0	0	981	115	1849	2877
taaagtgtct	<i>Similar to nuclear localization signals binding protein 1</i>	0	0	72	34	0	0	0	59	66	278	132
tgtagattt	<i>Chromobox homologue 1</i>	0	40	24	0	49	0	0	118	115	76	66
taaacccctc	<i>Ribonuclease 1</i>	0	20	291	0	0	0	0	654	512	329	199
gacagctgag	<i>Adenylate kinase 1</i>	26	20	48	34	65	105	53	74	115	126	66
tcacatcag	<i>Ribosomal protein P1/arylsulfatase D</i>	26	0	97	0	49	52	0	133	148	76	199
gcagccggct	<i>Ribosomal protein S3</i>	53	40	266	34	16	0	0	282	595	228	199
gaagcaagaa	<i>Phosphatidylinositol 4-kinase II/DKFZP586B1621</i>	0	72	0	0	16	105	0	104	49	76	154
ttaaacttaa	<b>Fusin</b>	53	20	194	0	0	0	0	312	99	202	509
ctgtatttga	<i>Transformer-2α</i>	53	0	24	34	130	52	53	106	332	118	202
tcgttgttta	<i>Leucine carboxyl methyltransferase/CGI-68</i>	0	20	34	0	114	52	53	89	165	152	110
cccactgtta	<i>ATPase, Ca<sup>2+</sup> transporting, plasma membrane 4</i>	0	0	169	0	0	52	53	148	198	152	110
tggaatgac	<b>Collagen, type I, α1</b>	26	0	2231	0	16	0	0	1026	975	2026	2833
atgtgaagag	<b>Osteonectin (secreted)</b>	0	20	1358	34	0	0	0	1085	347	1216	1837
acctataagt	<i>Putative heme-binding protein</i>	53	0	0	0	130	52	53	163	66	228	199
agtgggtgct	<b>Fibromodulin</b>	0	0	194	0	0	105	0	401	66	380	243
ttgttctgct	<i>TSPY-like/EST</i>	53	40	97	0	163	105	107	163	132	278	287
tggaagggct	<i>Putative mitochondrial outer membrane protein import receptor</i>	26	40	48	34	147	105	107	74	115	228	110
atggctaagc	<b>Regulator of G-protein signaling 10</b>	53	20	242	34	49	0	0	163	231	126	154
gggcagaata	<b>Peroxisomal farnesylated protein</b>	0	0	121	0	65	0	53	133	115	101	110
agggccgggg	<i>Aldehyde dehydrogenase 4 family, A</i>	0	0	48	34	130	105	53	118	82	152	88
tttctaaag	<i>RNA binding protein</i>	26	40	24	34	294	105	53	223	181	126	154
cattccagag	<i>Metaxin 1</i>	26	0	0	34	147	105	53	74	82	126	154
ccactcctcc	<i>Defender against cell death 1</i>	0	40	0	34	130	0	53	59	132	253	88
tggttcaag	<i>Ribosomal protein L27a</i>	0	0	218	0	114	52	0	89	247	278	398
gactttttaa	<i>SMC4-like 1</i>	0	0	97	34	130	0	53	148	115	101	177
atcaagaatc	<i>IF γ-inducible protein 30</i>	0	20	24	0	130	0	161	148	148	202	243
caggaaccac	<i>Ribosomal protein L37a</i>	26	0	0	34	65	0	0	59	99	126	177
gagagtaaca	<i>Syncoilin</i>	26	40	145	69	98	52	0	118	115	126	132
tatcactctg	<i>Male-enhanced antigen</i>	0	0	0	34	245	0	107	89	115	101	199
tgcatattt	<i>cdk inhibitor 3</i>	0	0	0	0	98	0	107	89	49	101	88
ttaagaggga	<i>Transducer of ERBB2, 1</i>	0	0	121	34	376	105	53	208	132	202	332
ctgacccctc	<i>Ubiquitin-conjugating enzyme E2L 3/glucuronosyltransferase 1</i>	53	20	0	103	81	52	53	163	165	152	132
ttgtacaaca	<i>CGI-27 protein</i>	26	0	97	0	98	0	107	59	82	126	110
tggtttttaa	<b>Poly(rC)-binding protein 2</b>	26	20	72	34	147	105	107	74	132	76	132
aagaatctga	<b>NADH dehydrogenase 1β subcomplex, 1/FLJ12549</b>	0	0	169	0	376	0	161	223	512	380	796
ctgcttaaga	<i>Laminin receptor 1</i>	0	72	0	0	98	105	0	74	49	101	132
cgctgttttt	<b>Flap structure-specific endonuclease 1</b>	26	20	194	0	81	105	161	133	281	329	154
gttgtaaat	<i>Lysophosphatidic acid acyltransferase-delta</i>	53	0	97	0	229	52	53	89	247	76	265
aggaaagaa	<i>v-erb-b2</i>	0	0	97	103	32	211	53	475	132	684	1128

Table 1 Continued

SAGE tag	Gene	Individual SAGE data										
		N1 <sup>b</sup>	N2	Dcis1	Dcis2	CL1	CL2	CL3	I1	M1	I2	M2
Novel genes												
ctctgagga	EST <sup>d</sup>	0	20	0	0	0	52	0	104	82	101	110
gtgacaccgc	?	0	20	48	0	0	105	0	59	165	126	132
ttaatgcaa	?	26	20	24	34	32	0	53	118	66	101	132
taattaactc	?	0	0	24	0	32	0	0	74	99	76	132
ggccttgtag	FLJ11088	0	20	24	69	49	0	107	104	66	101	154
tgcccttagta	Similar to mouse Dnaj11	26	40	0	0	49	0	0	118	165	152	110
ggcatcaggg	Hypothetical protein, estradiol-induced	26	20	0	34	0	52	53	74	429	202	88
tcaccagggg	?	0	20	97	0	32	52	0	133	214	278	243
tttctgtaa	?	0	20	72	69	16	52	53	163	82	278	553
taactttat	Similar to murine leucine-rich repeat protein	26	0	0	0	49	0	0	118	82	76	110
tgccctaaaa	?	53	0	24	0	49	0	0	59	181	177	88
gattccacag	?	0	20	48	34	32	0	0	178	49	76	221
gtaaaacaat	EST	26	0	24	34	65	105	0	59	82	101	199
gaaccacaag	?	0	0	121	69	0	0	53	89	181	304	110
actgatgcaa	EST	0	0	121	0	16	0	53	148	132	101	132
gatctcatct	EST	0	0	0	0	81	52	53	74	82	126	132
tcaaatgcaa	<b>KIAA0156</b>	0	20	97	0	0	0	53	74	115	76	154
ctttattcca	EST	0	0	218	0	0	0	0	252	165	278	221
gtagacacct	?	0	0	266	0	0	105	0	178	231	354	199
acaaggtgcg	?	53	40	97	0	114	52	0	74	165	481	332
tttctggagg	<b>KIAA0545</b>	0	20	169	34	32	52	53	133	66	177	221
ggagggaaca	?	26	0	97	0	65	0	0	89	99	76	110
ttcccca	MGC2477	0	0	121	34	49	0	0	59	132	101	132
ggaatatgca	?	26	20	121	0	130	105	0	74	165	202	243
gcagaccac	FLJ22940	26	20	72	34	98	0	0	74	198	76	110
gctggaataa	FLJ20211	0	20	97	34	81	105	53	59	82	126	177
agctgggttg	FLJ11099	0	40	97	69	147	52	53	89	165	177	110
aggacacaa	?	53	0	242	0	81	158	53	118	247	253	376
gagtgagtga	MGC3234	0	40	97	69	163	105	53	252	99	152	88
gcataactgc	DKFZP434D1335	26	20	121	34	49	52	107	118	66	101	66
tgataattc	?	26	20	121	69	65	105	107	104	82	101	177
tgcccttta	?	0	0	169	69	81	105	0	118	264	152	132
gtgctgtcc	Similar to RIKEN cDNA 1110021J02 gene	0	20	121	69	16	0	107	74	132	177	177
tcgtaacgag	Homo sapiens, clone IMAGE:3343149, mRNA, partial cds	26	20	121	0	196	0	53	89	99	278	243
tgtggtggtg	MLN51	26	61	72	34	0	105	0	282	198	988	1040

<sup>a</sup> All values have been converted to tags per million. All samples are from the SAGEmap site as of 7/30/01. Complete breast data are shown for all SAGE tags identified in the breast cancer invasion-specific cluster by PCA (Fig. 2).

<sup>b</sup> Samples listed are N1 and N2, two independent normal duct epithelium samples (BrN and mammary epithelium); Dcis1 and Dcis2, two independent samples of Dcis (Dcis and Dcis2); CL1, MCF70h; CL2, MDA453; CL3, LacZ; I1 and M1, 95-347 and 95-348; I2 and M2, 95-259 and 95-260.

<sup>c</sup> Names listed in bold are those genes also identified in the invasion-specific cluster of breast cancer by Porter *et al.* (4).

<sup>d</sup> EST, a match to the Unigene cluster defined by expressed sequence tags.

*apolipoprotein C-1* gene expression was within macrophages infiltrating the tumor or within necrotic debris associated with DCIS. *Osteonectin* and *IGFBP7* were expressed within two architectural compartments in all four carcinomas studied. *Osteonectin* gene expression was observed within angioendothelial cells and the juxtatumoral stroma in all four cases. *IGFBP7* was predominantly expressed within the angioendothelium, although two of four cases also showed weak labeling of tumor epithelium. Thus, although the identification of *IGFBP7* within the invasion-specific cluster of breast cancer can largely be attributed to endothelial expression, our results are fully consistent with the published reports of *IGFBP7* gene expression within breast tumor epithelium (18, 19). Three genes (*osteonectin*, *collagen 1α1*, and *IGFBP7*) were specifically expressed within the invasive tumor as compared with adjacent normal breast tissue and serve as markers of the desmoplastic response in infiltrating breast carcinomas (20, 21). *Osteonectin* is a phosphorylated, acidic, glycine-rich glycoprotein of  $M_r$  43,000 with multiple  $Ca^{2+}$ -binding domains. The function of osteonectin is not primarily known, but it is thought to be involved in angiogenesis and remodeling of the extracellular matrix in keeping with its elevated expression in the host stromal response (22, 23). *Collagen 1α1* expression by the stroma likely reflects the transcriptional activity of proliferating fibroblastic tissue within the host response. *IGFBP7*, which was strongly expressed by endothelial cells, has not been described as an endothelial-specific marker in human tumors, although other members of this gene family have been so implicated (13). For each invasive carcinoma analyzed,

samples of normal breast terminal duct lobular unit epithelium were present within the same tissue section. Expression of three genes, *fusin*, *apolipoprotein C-1*, and *laminin receptor 1*, were also noted in tissues of the normal breast. *Fusin* expression was noted within small lymphocytes within the intralobular stroma of normal lobules, and *apolipoprotein C-1* was expressed in macrophages present in benign ducts. *Laminin receptor 1*, although most strongly expressed by the neoplastic epithelium, was also weakly expressed in atrophic ducts, as well as in areas of DCIS both within and outside the mass.

Invasion-specific genes were thus spatially localized to distinct compartments of gene expression in the host stromal response to breast cancer. This extends our prior observations in the pancreas and supports the existence of a highly structured organization of gene expression within the host desmoplastic response to infiltrating carcinoma. Specifically, the finding of *osteonectin* gene expression localized to the juxtatumoral stroma validates this newly defined architectural region of the host stromal transcriptional response. Expression of *osteonectin* by the juxtatumoral stroma may thus be intimately involved with the invasive process and highlights this region as a potential site of tumor-host interactions to be targeted for therapeutic intervention.

**Comparison of Juxtatumoral Gene Expression in Breast and Pancreas Carcinomas.** We have noted previously that *apolipoprotein C-1*, *apolipoprotein D*, and *MMP11* are each gene expression markers of the juxtatumoral stromal compartment in adenocarcinomas of the pancreas (2). In an effort to better discern the gene expression

Fig. 3. *In situ* detection of invasive tissue-specific genes in breast cancer. Invasive tissue-specific gene expression was found to be located within angioendothelial tissue (A), tumor epithelium (B), juxtatumoral stroma (C), panstromal tissue (D), or inflammatory cells within the invasive carcinoma (E and F). A, *IGFBP7* mRNA expression detected within angioendothelium. Adjacent tumor epithelium and stroma are negative in expression in this field. B, *laminin receptor 1* mRNA expression detected within tumor epithelium but not within adjacent stroma. C, *osteonectin* mRNA expression detected within stromal cells immediately adjacent to tumor epithelium (juxtatumoral stroma) but not in stromal cells away from the invasive tumor glands. *Osteonectin* mRNA expression was also seen in endothelial cells (not shown). D, *collagen 1 $\alpha$ 1* mRNA expression detected throughout the stromal response. Tumor epithelium is negative. E, *apolipoprotein C-1* mRNA expression detected within macrophages scattered throughout the stroma, as well as within areas of comedonecrosis (bottom right and inset). F, *fusin* mRNA expression detected within mature lymphocytes at the infiltrating edges of the tumor. Tumor epithelium, angioendothelium, and stroma were all negative.

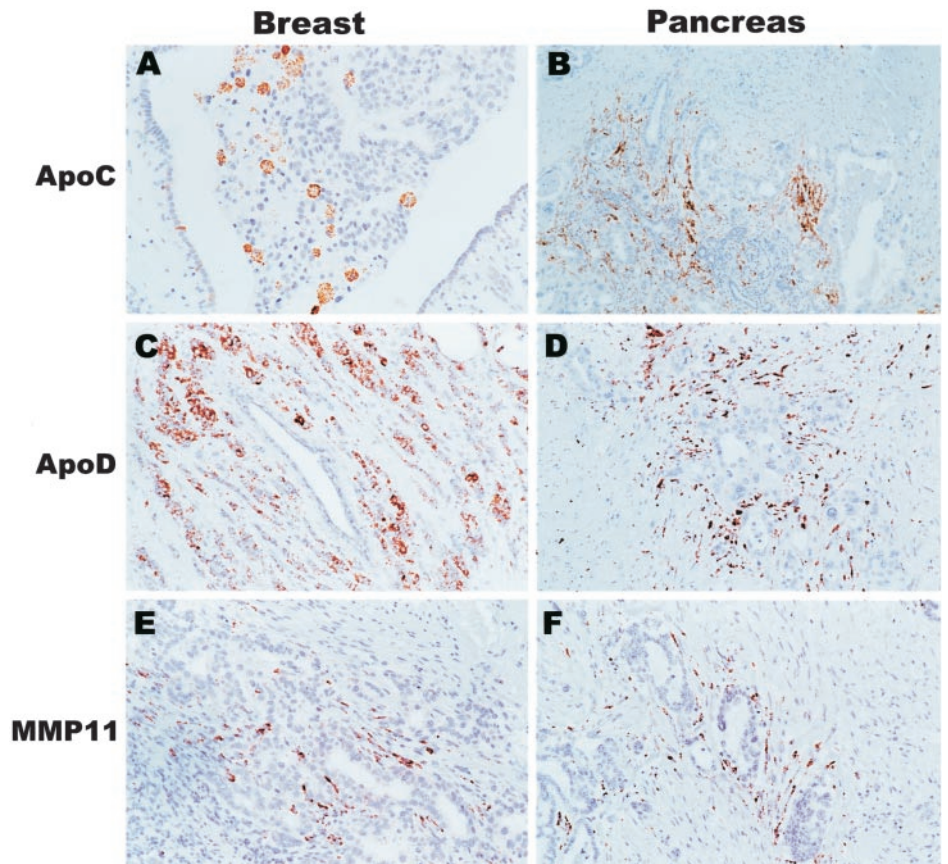
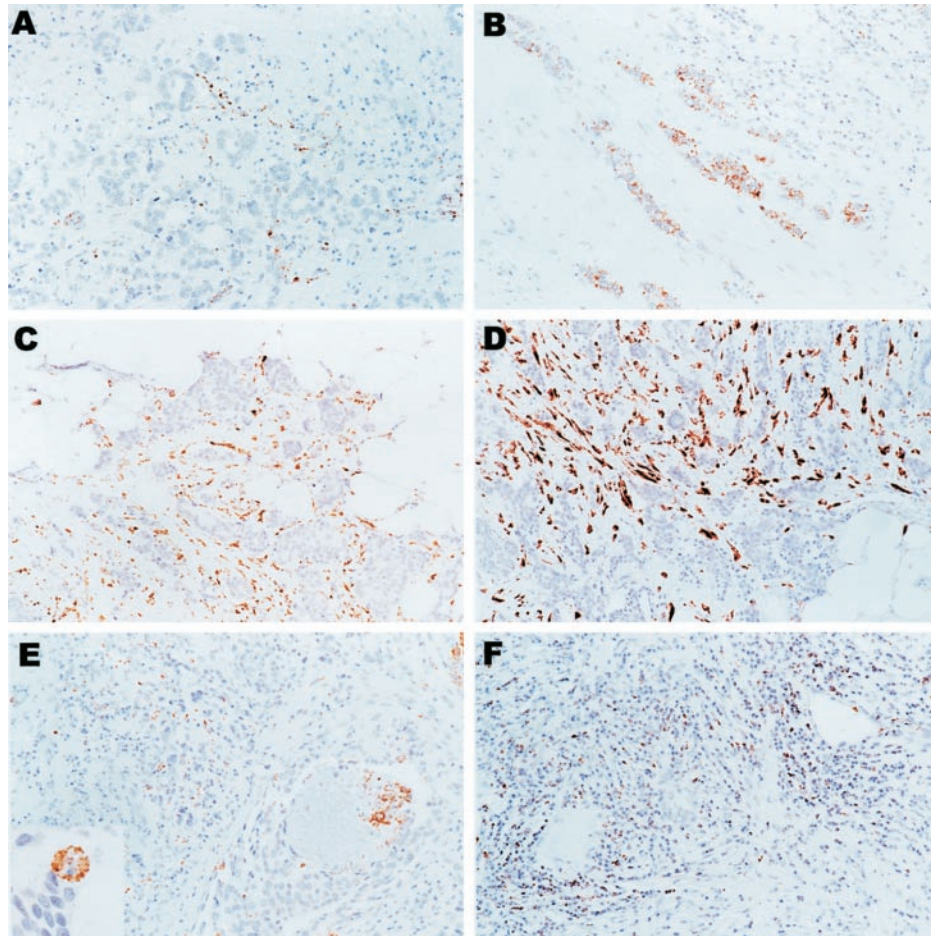


Fig. 4. Comparison of juxtatumoral gene expression in invasive breast and pancreas cancers. *In situ* detection of *apolipoprotein C-1* (A and B), *apolipoprotein D* (C and D), and *MMP11* (E and F) in paraffin-embedded sections of breast and pancreas cancers ( $\times 100$ ). A, C, and E are sections of infiltrating breast cancer, and B, D, and F are sections of infiltrating pancreas cancer. *Apolipoprotein C-1* (*ApoC*) localizes to macrophages within areas of tissue invasion and necrosis in breast cancer specimens (A), in contrast to the juxtatumoral stromal expression seen for *apolipoprotein C* in pancreas cancers (B). *Apolipoprotein D* (*ApoD*) expression localizes to tumor epithelium of breast cancers (C), also in contrast to the juxtatumoral stroma expression seen for *apolipoprotein D* in pancreas cancers (D). Only *MMP11* gene expression localizes to the juxtatumoral stroma of invasive breast (E) and pancreas (F) cancers.

patterns of this distinct region of the host stromal response, we performed *in situ* hybridization of each of these 3 genes in the four samples of invasive breast carcinoma, with comparison to the gene expression patterns seen previously in samples of paraffin-embedded pancreas cancers for these genes (Fig. 4).

Gene expression of the juxtatumoral stroma was found to differ among these two tumor types. As noted previously, *apolipoprotein C-1* was one of only 3 genes common to the invasion-specific clusters of both the breast and pancreas. Surprisingly, although this gene was expressed in both tumor types, its cellular distribution within breast or pancreas cancers was dissimilar. *Apolipoprotein C-1* gene expression localized to tumor-infiltrating macrophages in the four samples of breast carcinoma (Fig. 4A) but was clearly expressed by stromal fibroblasts within the juxtatumoral stroma in the four samples of pancreatic carcinoma (Fig. 4B). *Apolipoprotein D* expression also differed among breast and pancreas cancers. *Apolipoprotein D* gene expression was localized to tumor epithelium in the four breast carcinomas (Fig. 4C), in contrast to the juxtatumoral stromal pattern seen in all four pancreas cancer tumor tissues (Fig. 4D). Only MMP11, of the 3 invasion-specific genes studied, was found to localize to the juxtatumoral stroma in all four breast cancers and in all four pancreatic cancers studied (Fig. 4, E and F). Thus, although the juxtatumoral stroma appears to be a defined component of the host stromal response, the gene expression profile of this compartment must depend in part upon the site of tumor origin.

Surprisingly, although these genes are associated with the process of tissue invasion in both breast and pancreas cancers, their role in tissue invasion differs between these two tumor types. These observations in turn raise several other questions regarding the desmoplastic response to human tumors, *i.e.*, is the gene expression of the host stromal response to primary tumors similar to or different from the stromal response present in metastatic tumors? Do histologically different tumors that are derived from the same organ type produce similar or different host stromal responses? Our current data indicate that, with respect to the desmoplastic response, the robust patterns of gene expression in one tumor type are different from the robust patterns seen in other tumor types. Clearly, additional work needs to be done to determine how static or variable these gene expression patterns are.

In summary, the patterns of spatially organized compartments of gene expression in the host response to breast cancer and the comparisons among various cancer types provide new insights into the biology of desmoplasia. These similarities in the host stromal response to different tumor types may suggest some universal targets for therapeutic intervention. Additional studies to understand the desmoplastic response to invasive neoplasms may aid in identifying new targets for clinical imaging, serological diagnosis, drug development, and delivery.

## ACKNOWLEDGMENTS

We thank Dr. Sandra Rempel for the generous gift of the *osteonectin* cDNA used for preparation of the *osteonectin*-specific riboprobes for *in situ* hybridization.

## REFERENCES

- Ryu, B., Jones, J., Hollingsworth, M. A., Hruban, R. H., and Kern, S. E. Invasion-specific genes in malignancy: serial analysis of gene expression comparisons of primary and passaged cancers. *Cancer Res.*, *61*: 1833–1838, 2001.
- Iacobuzio-Donahue, C. A., Ryu, B., Hruban, R. H., and Kern, S. E. Exploring the host desmoplastic response to pancreatic carcinoma: gene expression of stromal and neoplastic cells at the site of primary invasion. *Am. J. Pathol.*, *160*: 91–99, 2002.
- Walker, R. A. The complexities of breast cancer desmoplasia. *Breast Cancer Res.*, *3*: 143–145, 2001.
- Porter, D. A., Krop, I. E., Nasser, S., Sgroi, D., Kaelin, C. M., Marks, J. R., Riggins, G., and Polyak, K. A SAGE (serial analysis of gene expression) view of breast tumor progression. *Cancer Res.*, *61*: 5697–5702, 2001.
- Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*, *95*: 14863–14868, 1998.
- Elston, C. W., and Ellis, I. O. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, *19*: 403–410, 1991.
- Baklanov, M. M., Golikova, L. N., and Malygin, E. G. Effect on DNA transcription of nucleotide sequences upstream to T7 promoter. *Nucleic Acids Res.*, *24*: 3659–3660, 1996.
- Kadkol, S. S., Gage, W. R., and Pasternack, G. R. *In situ* hybridization-theory and practice. *Mol. Diagn.*, *4*: 169–183, 1999.
- Perou, C. M., Sorlie, T., Eisen, M. B., Van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O., and Botstein, D. Molecular portraits of human breast tumours. *Nature (Lond.)*, *406*: 747–752, 2000.
- Tlsty, T. D. Stromal cells can contribute oncogenic signals. *Semin. Cancer Biol.*, *11*: 97–104, 2001.
- Tlsty, T. D., and Hein, P. W. Know thy neighbor: stromal cells can contribute oncogenic signals. *Curr. Opin. Genet. Dev.*, *11*: 54–59, 2001.
- Gress, T. M., Muller-Pillasch, F., Lerch, M. M., Friess, H., Buchler, M., Beger, H. G., and Adler, G. Balance of expression of genes coding for extracellular matrix proteins and extracellular matrix degrading proteases in chronic pancreatitis. *Gastroenterology*, *32*: 221–225, 1994.
- St Croix, B., Rago, C., Velculescu, V., Traverso, G., Romans, K. E., Montgomery, E., Lal, A., Riggins, G. J., Lengauer, C., Vogelstein, B., and Kinzler, K. W. Genes expressed in human tumor endothelium. *Science (Wash. DC)*, *289*: 1197–1202, 2000.
- Perou, C. M., Jeffrey, S. S., van de R. M., Rees, C. A., Eisen, M. B., Ross, D. T., Pergamenschikov, A., Williams, C. F., Zhu, S. X., Lee, J. C., Lashkari, D., Shalon, D., Brown, P. O., and Botstein, D. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl. Acad. Sci. USA*, *96*: 9212–9217, 1999.
- Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C., Trent, J. M., Staudt, L. M., Hudson, J., Jr., Boguski, M. S., Lashkari, D., Shalon, D., Botstein, D., and Brown, P. O. The transcriptional program in the response of human fibroblasts to serum. *Science (Wash. DC)*, *283*: 83–87, 1999.
- Kikkawa, Y., Sanzen, N., Fujiwara, H., Sonnenberg, A., and Sekiguchi, K. Integrin binding specificity of laminin-10/11: laminin-10/11 are recognized by  $\alpha 3\beta 1$ ,  $\alpha 6\beta 1$  and  $\alpha 6\beta 4$  integrins. *J. Cell Sci.*, *113* (Pt. 5): 869–876, 2000.
- Egawa, T., Kawabata, K., Kawamoto, H., Amada, K., Okamoto, R., Fujii, N., Kishimoto, T., Katsura, Y., and Nagasawa, T. The earliest stages of B cell development require a chemokine stromal cell-derived factor/pre-B cell growth-stimulating factor. *Immunity*, *15*: 323–334, 2001.
- Degeorges, A., Wang, F., Frierson, H. F., Jr., Seth, A., and Sikes, R. A. Distribution of IGFBP-rP1 in normal human tissues. *J. Histochem. Cytochem.*, *48*: 747–754, 2000.
- Landberg, G., Ostlund, H., Nielsen, N. H., Roos, G., Emdin, S., Burger, A. M., and Seth, A. Downregulation of the potential suppressor gene *IGFBP-rP1* in human breast cancer is associated with inactivation of the retinoblastoma protein, cyclin E overexpression and increased proliferation in estrogen receptor negative tumors. *Oncogene*, *20*: 3497–3505, 2001.
- Gilles, C., Bassuk, J. A., Pulyaeva, H., Sage, E. H., Foidart, J. M., and Thompson, E. W. SPARC/osteonectin induces matrix metalloproteinase 2 activation in human breast cancer cell lines. *Cancer Res.*, *58*: 5529–5536, 1998.
- Bellahcene, A., and Castronovo, V. Increased expression of osteonectin and osteopontin, two bone matrix proteins, in human breast cancer. *Am. J. Pathol.*, *146*: 95–100, 1995.
- Bradshaw, A. D., Reed, M. J., and Sage, E. H. SPARC-null mice exhibit accelerated cutaneous wound closure. *J. Histochem. Cytochem.*, *50*: 1–10, 2002.
- Kato, Y., Lewalle, J. M., Baba, Y., Tsukuda, M., Sakai, N., Baba, M., Kobayashi, K., Koshika, S., Nagashima, Y., Frankenne, F., Noel, A., Foidart, J. M., and Hata, R. I. Induction of SPARC by VEGF in human vascular endothelial cells. *Biochem. Biophys. Res. Commun.*, *287*: 422–426, 2001.