[CANCER RESEARCH 62, 819–826, February 1, 2002]

Relationships and Differentially Expressed Genes among Pancreatic Cancers Examined by Large-scale Serial Analysis of Gene Expression[1]

Byungwoo Ryu, Jessa Jones, Natalie J. Blades, Giovanni Parmigiani, Michael A. Hollingsworth, Ralph H. Hruban, and Scott E. Kern[2]

Departments of Oncology [B. R., J. J., G. P., R. H. H., S. E. K.], Biostatistics [N. J. B., G. P.], and Pathology [R. H. H., G. P., S. E. K.], The Johns Hopkins Medical Institutions, Baltimore, Maryland 21231, and The Eppley Institute, The University of Nebraska Medical Center, Omaha, Nebraska 68198 [M. A. H.]

ABSTRACT

Pancreatic adenocarcinoma is among the most fatal of cancers, in part because of late diagnosis and a lack of effective therapies. Comprehensive studies are needed to better understand and address the cellular mechanisms and pathways of tumorigenesis. Serial analysis of gene expression was used to analyze gene expression profiles of pancreatic cancer cell lines, short-term cultures of normal pancreatic ductal epithelium, and primary pancreatic cancer tissue. A total of 294,920 tags representing 77,746 genes in 10 serial analysis of gene expression libraries were analyzed. A pancreatic cancer cell line (Hs766T) that exhibited a "normoid" profile of gene expression was identified. Several genes that may be involved in the fundamental nature of malignant changes in pancreatic ductal epithelium were suggested from those differentially and highly expressed in pancreatic cancer cells as compared with normal epithelium. Some overexpressed genes, such as S100A4, prostate stem cell antigen, carcinoembryonic antigen-related cell adhesion molecule 6, and mesothelin, suggest potential use as diagnostic markers. Others suggest potential novel therapeutic targets.

INTRODUCTION

Pancreatic cancer is the fifth leading cause of cancer death in the United States. Annually, ~28,200 people are diagnosed and die of pancreatic cancer (1). The mortality rate is the highest among cancer types, in part because of the asymptomatic nature of the disease in early stages, a lack of sensitive and specific diagnostic tools, and limited progress in development of effective therapeutics. Better knowledge of changes of gene expression that accompany pancreatic cancer may suggest new screening tools and therapeutic strategies. Several genes that are overexpressed in pancreatic cancer have been identified by subtractive and comparative methods (2–10). Initial surveys of these tumors and cell lines by RNA-based gene expression analysis have been reported (11–13).

A series of genetic changes within pancreatic ductal epithelium accompanies the development of precursor lesions, termed pancreatic intraepithelial neoplasia (15), some of which progress to pancreatic adenocarcinoma. SAGE[3] was used to compare the gene expression in short-term cultures of normal pancreatic epithelium to cultures of pancreatic carcinoma cells using data from primary tumors to filter the data from cultured cells. SAGE technology, developed by Velculescu *et al.* (14) provides a simultaneous and comprehensive enumeration of gene transcripts of a given sample in a quantitative manner, whereas it also provides sequence information that is used to identify the differentially expressed genes. Here, we report differentially expressed genes identified from the comparative SAGE profiling of 294,920 tags in 10 pancreatic samples.

MATERIALS AND METHODS

Samples. Pancreatic cancer cell lines AsPc1, CAPAN1, CAPAN2, Hs766T, and Panc-1 were obtained from the American Type Culture Collection (Manassas, VA). Cell line PL45 was developed from a primary pancreatic ductal adenocarcinoma (16). Short-term cultures of pancreatic ductal epithelial cells (HX, H48, H110, and H126) were prepared as described and validated as having the characteristics of ductal epithelium (17). Profiles of genetic alterations of the pancreatic cancer cell lines used in this study are known in considerable depth (18). Within these lines, the *CDKN2A* gene is inactivated by genetic or epigenetic changes in all; the *KRAS2* gene is mutated in all; the *AKT2* amplicon is present in AsPc1 and Panc-1, *BRCA2* is genetically inactivated in CAPAN1; *TP53* is genetically inactivated in AsPc1, CAPAN1, Panc-1, and PL45; *MKK4* is genetically inactivated in AsPc1 and CAPAN1; and *MADH4* is genetically inactivated in AsPc1, CAPAN1, and Hs766T.

Panc-1 and Hs766T cell lines were cultured in DMEM supplemented with 10% fetal bovine serum and antibiotics (100 units/ml penicillin and 100 μg/ml streptomycin). CAPAN1 and CAPAN2 cell lines were cultured in RPMI 1640 and McCoy's growth medium (Life Technologies, Inc., Gaithersburg, MD) supplemented with 10% fetal bovine serum and antibiotics (100 units/ml penicillin and 100 μg/ml streptomycin), respectively. Use of different medium was required to minimize the variance in growth rates that would otherwise be exaggerated with the use of a single medium.

The cell lines samples AsPc1 and PL45, and the two primary pancreatic ductal adenocarcinomas were described (19). The isolation and validation of normal pancreatic ductal short-term cultures HX, H126, H48, and H110 were described (17).

SAGE. Total cellular RNA was obtained from pancreatic cancer cells, ~90% confluent CAPAN1, CAPAN2, Hs766T, and Panc-1 cell line cultures, using TRIZOL reagent (Life Technologies, Inc.). Polyadenylated mRNA was purified from total RNA (Messagemaker; Life Technologies, Inc.), and cDNA was generated (cDNA Synthesis System; Life Technologies, Inc.). SAGE was performed as described by Velculescu *et al.* (14) for all of the pancreatic cancer libraries. For the generation of the two normal pancreatic ductal epithelial cell line (HX and H126) libraries, MicroSAGE, a SAGE technique modified for limited sample sizes (20), was used with a slight modification. In brief, total RNA was prepared from HX and H126 using TRIZOL (Life Technologies, Inc.) instead of direct mRNA isolation from cells as described originally in the MicroSAGE protocol (20). A modified lysis/binding buffer was prepared from 1 ml of lysis/binding buffer [100 mM Tris-HCl (pH 7.5), 500 mM LiCl, 10 mM EDTA, 1% LiDS, 5 mM DTT] (Dynabead mRNA direct kit; Dynal, Oslo, Norway) by addition of 66 μg tRNA and 10 μg of BSA. Total RNA (5 μg) was dissolved in 1-ml modified lysis/binding buffer. mRNA was purified using Oligo$(dT)_{25}$ Dynabeads, and SAGE libraries were constructed. As part of the CGAP (NIH) SAGE consortium, all six of the SAGE libraries were arrayed at the Lawrence Livermore National Laboratories and Washington University Human Genome Center. The SAGE library data were posted at the CGAP website[4] as part of the SAGEmap database (21, 22). SAGE data of other tumor types were obtained from this database for comparison. SAGE libraries of two pancreatic cancer cell lines (AsPc1 and PL45) and from primary pancreatic cancer tissues (91-16113 and 96-6252) were obtained from earlier efforts (19).

Received 4/25/01; accepted 11/20/01.

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] Supported by Grants CA62924 and CA72712 from the NIH Specialized Programs of Research Excellence in Gastrointestinal Cancer.

[2] To whom requests for reprints should be addressed, at Department of Oncology, 451 Cancer Research Building, 1650 Orleans Street, The Johns Hopkins School of Medicine, Baltimore, MD 21231. Phone: (410) 614-3316; Fax: (410) 614-9705; E-mail: sk@jhmi.edu.

[3] The abbreviations used are: SAGE, serial analysis of gene expression; PSCA, prostate stem cell antigen; CEACAM6, carcinoembryonic antigen-related cell adhesion molecule 6; CGAP, Cancer Genome Anatomy Project; EST, expressed sequence tag; PCA, principal component analysis; RT-PCR, reverse transcription-PCR.

[4] Internet address: http://www.ncbi.nlm.nih.gov/SAGE/.

Table 1 *Overall descriptions of SAGE libraries*[a]

Library[b]	D1	D2	CL1	CL2	CL3	CL4	CL5	CL6	I1	I2	Total
Tags	32157	32420	10467	24879	37926	23222	31864	32300	33942	35743	294920
Unique tags[c]	8529	6872	4127	7815	11711	7824	8635	8286	11355	11504	86658
Genes[d]	8411	6845	4112	7783	11660	7788	8583	8228	11297	11450	77746
>500											
Unique tags[c]	0	0	0	1 (5)	0	0	1 (2)	1 (5)	1 (5)	1 (7)	5 (3)
Genes[d]	0	0	0	1	0	0	1	1	1	1	5
>50 and ≤500											
Unique tags[c]	69 (29)	71 (31)	14 (11)	35 (13)	76 (21)	46 (17)	65 (23)	62 (22)	38 (12)	38 (11)	514 (20)
Genes[d]	68	70	14	35	76	46	65	62	38	38	512
>3 and ≤50											
Unique tags[c]	1112 (35)	1196 (39)	345 (38)	1033 (41)	1389 (37)	818 (39)	1104 (37)	1075 (37)	1211 (38)	1376 (39)	10659 (38)
Genes[d]	1111	1193	344	1028	1386	7817	1095	1060	1203	1368	10605
≤3											
Unique tags[c]	7348 (36)	5605 (30)	3768 (51)	6746 (41)	10246 (42)	6960 (45)	7465 (38)	7148 (35)	10105 (45)	10089 (43)	75480 (39)
Genes[d]	7219	5572	3754	6719	6799	6925	7422	7105	10055	10043	74832

[a] The libraries are: H126 and HX, human pancreas duct epithelium from short-term cultures; Hs766T, Panc-1, CAPAN1, CAPAN2, PL45, and AsPc1, pancreatic cancer cell lines; 91-16113 and 96-6252, primary pancreatic adenocarcinomas.

[b] D1 and D2 are normal pancreatic ductal epithelium, HX and H126, respectively. CL1 to CL6 are pancreatic cancer cell lines, Hs766T, Panc-1, CAPAN1, CAPAN2, PL45, and AsPcl. I1 and I2 are primary pancreatic cancer, 91-16113 and 96-6252, respectively.

[c] Number of different transcripts represented by the total tags analyzed. Numbers in parentheses indicates the mass fraction represented by the indicated abundance class (the sum of tag counts in a category divided by total tags for the sample, in percentage).

[d] Number of genes, corrected by the estimated SAGE sequence error rate (6.8%) (14). These numbers are obtained by elimination of ESTs, mitochondrial DNA sequences, and repetitive DNA sequences from a dataset of total genes that matched to an entry of SAGEtag to UniGene Mapping database (http://ncbi.nlm.nih.gov/pub/sage/map/). A database, SAGEmap_tag_ug.rel.NlaIII.hs, representing those tag to gene assignments deemed reliable by CGAP, was used.

Statistical Analysis. Partek Pro2000 (Partek Inc., St. Louis, MO), Cluster and TreeView,[5] SigmaStat 2.03 (SPSS Science, Chicago, IL), Access, and Excel (Microsoft, Seattle, WA) and R[6] programs were used. For all of the statistical analysis beyond the initial description of datasets, SAGE data were normalized to tags per 100,000. A subset of the 10 SAGE libraries was obtained using Cluster and TreeView programs by filtering to require each tag to have one observation in > 3 of the 10 libraries, which produced a total of 6,245 unique tags. This filtered subset of data was used for all of the additional analysis. PCA analysis (23) was carried out on the filtered subset of data using Partek Pro2000 software to graphically plot the three major components and Cluter/TreeView to list the genes of each component.

RT-PCR. Constant amounts (1.0 μg) of total RNA from pancreatic cancer cell lines (CAPAN1, CAPAN2, PL45, and AsPc1) and two normal pancreatic ductal epithelial cells (H48 and H110, which were derived from pancreas of a 16-year-old male and a 17-year-old male, respectively, and cultured short-term; Ref. 17), were used. Normal duct samples HX and H126, which were used for SAGE, were not available because of the small quantities of such cultures. Reverse-transcription was performed in a total reaction volume of 20 μl using Oligo(dT)$_{25}$ primers and the SuperScript First-Strand Synthesis System for RT-PCR kit (Life Technologies, Inc.) according to the manufacturer's protocol. The products were serially diluted and used for subsequent PCRs. Primers for corresponding genes were selected from mRNA sequences obtained from GenBank.[7] Optimal PCR cycle numbers for each gene were determined by empirical identification of a range of cycles that produced exponential accumulation of amplified DNA on examination of a series of reactions comprising 25–40 PCR cycles. Aliquots (10 μl) of RT-PCR products were separated by electrophoresis in 1.5% or 2% agarose gels, depending on the product size, and were stained with ethidium bromide. Primer sequences and numbers of temperature cycles were: *S100A4* (32 cycles): sense 5′-CCCCTCTCTACAACCCTCTC, antisense 5′-AGCACGTGTCTGAAGGAGCC; *TSPAN-1* (36 cycles); sense 5′-ACTGTCGTCCAGTGCCATGC, antisense 5′-TAGCCCCAAGTCTGGAGCAG; *CEACAM6* (32 cycles); sense 5′-CCTGCAGATTGCATGTCCCC, antisense 5′-GTCCTATTGAGGCCAGTGCC; *ALG-2* (35 cycles): sense 5′-GACACCGAGCTTCAGCAAGC, antisense 5′-CACCTGTGCTCCATTCCCTC; and *glyceraldehyde-3-phosphate dehydrogenase* (29 cycles): sense 5′-GGCACCGTCAAGGCTGAGAA, antisense 5′-GAGACCACCTGGTGCTCAGT.

RESULTS AND DISCUSSION

Analysis of Group Data. Descriptive statistics of the 10 pancreas SAGE libraries are presented in Table 1. A total of 294,920 tags were sequenced, and 86,658 (28.5%) were identified as unique tags. The average number of unique tags per library was 8,666, with a range of 4,127–11,711. Transcripts found three times or less in each library comprised ~87.1% of the unique tags identified in this study, but this low-abundance class represented only 39% of total mRNA mass as judged by numbers of total tags. The number of unique tags that matched genes was 77,746 (90%) after elimination of mitochondrial DNA sequence, repetitive DNA sequences, and correction for the estimated SAGE tag sequence error rate (6.8%, attributable to sequencing errors; see Ref. 14). These parameters were similar to those reported in similar SAGE datasets of other tissue types (19, 24).

A dendrogram created by hierarchical clustering analysis (Spearman rank correlation; Ref. 25) suggested a close relationship between the gene expression profiles of Hs766T, a pancreatic cancer cell line, and the two normal ductal epithelial samples (Fig. 1*A*). A more informative means of displaying the key correlations is by scatter plot (Fig. 1*B*). Scatter plot analysis visualizes gene expression profiles in two dimensions, and the Pearson correlation coefficients yield quantitative comparisons between the gene expression profiles. The relationships suggested by hierarchical clustering were in broad accordance with those shown by the pairwise comparisons of the Pearson correlation coefficients (Table 2). As expected, the normal ductal epithelia HX and H126 were very similar in gene expression profiles (Fig. 1*B*, D1 *versus* D2; Table 2, $r = 0.981$). Such a resemblance between samples of independent derivation provided a measure of the reliability of the SAGE technique. Cell line Hs766T had the highest degree of correlation to the normal pancreatic ductal epithelium ($r = 0.935$ for HX and 0.930 for H126) followed in degree of similarity by the cell lines CAPAN2, CAPAN1, and Panc-1 ($r = 0.797$, 0.748, and 0.515, comparison to HX). This relationship was also seen in the scatter plot analysis (Fig. 1*B*, D1–2 *versus* CL1). The two primary pancreatic cancer tissues, 91–16113 and 96–6252, were highly correlated with each other ($r = 0.980$), as expected. Pancreatic cancer cell line AsPc1 had the highest similarity of expression profile to the primary cancer tissues ($r = 0.770$ and 0.778 for 91-16113 and 96-6252, respectively). Scatter plots confirm the larger degrees of difference among the latter comparisons (Fig. 1*B*, D1–2 *versus* CL2–6 and D1–2 *versus* I1–2).

To display a high fraction of the available data in three dimensions, we used PCA. PCA can provide a global overview of the relatedness of gene expression profiles among samples while better avoiding the

[5] Internet address: http://www.microarrays.org/software.html.
[6] Internet address: www.stat.auckland.ac.nz/rproj.html.
[7] Internet address: http://www.ncbi.nlm.nih.gov/GenBank/index.html.

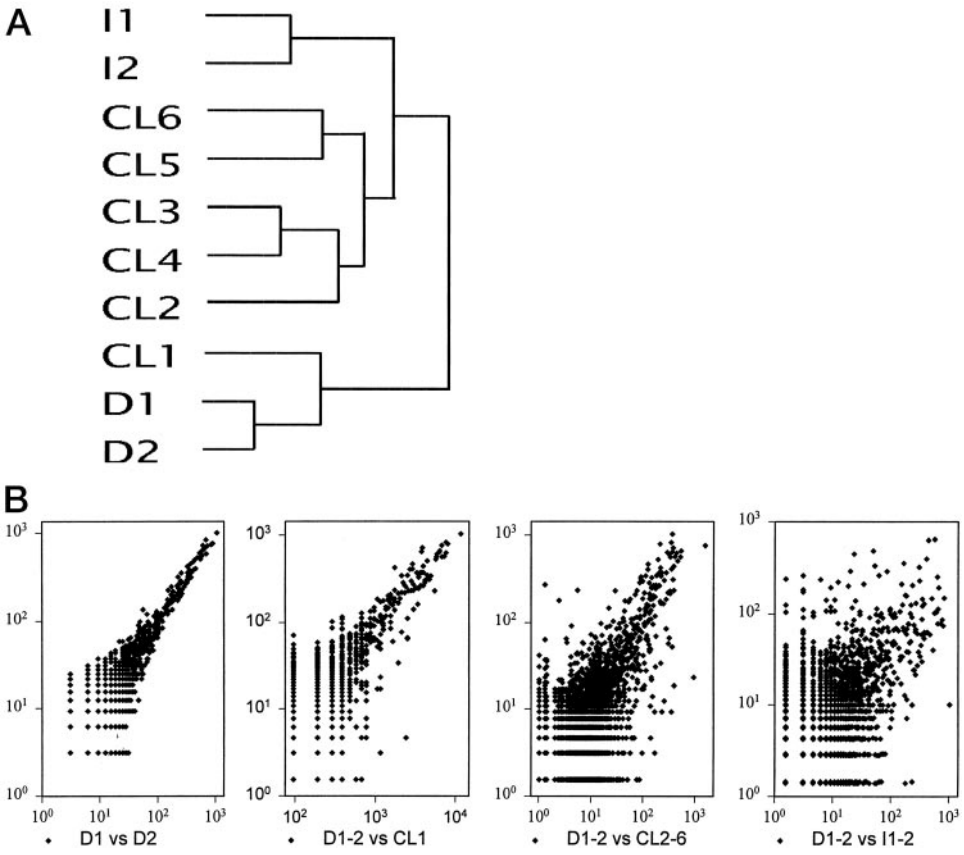

Fig. 1. Relationships of gene expression profiles. *A*, dendrogram produced by hierarchical cluster analysis of the 10 pancreatic SAGE libraries. D1 and D2 are the normal pancreatic ductal epithelial samples HX and H126, respectively. CL1 to CL6 are the pancreatic cancer cell lines Hs766T, Panc-1 CAPAN1, CAPAN2, PL45, and AsPc1, respectively. I1 and I2 are the invasive primary pancreatic carcinomas 92-16113 and 96-6252, respectively. *B*, scatter plots. The initial specimen in the label "y *versus* x" refers to the ordinate. The scale is in counts for unique tags per 100,000 total tags. D1–2 refers to the average normalized values of duct epithelial samples HX and H126. CL2–6 refers to the average of the normalized values of Panc-1, CAPAN1, CAPAN2, PL45, and AsPc1. I1–2 refers to the average of the normalized values of the invasive tumors 92-16113 and 96-6252.

deterministic and rather arbitrary nature of hierarchical clustering. The filtered and normalized dataset, which is described in "Materials and Methods," was log-transformed, and PCA was carried out using Partek Pro2000 bioinformatic software. PCA also distinguished the three original sample groups, *i.e.*, normal pancreatic ductal epithelium, pancreatic adenocarcinoma cells, and primary pancreatic tumors. PCA confirmed the very close similarity of the samples of normal duct epithelium, the close resemblance of the normal duct profiles to cell line Hs766T, the more "distant" profiles of the other five pancreatic cancer lines, and the distinctive nature of the invasive lesions (data not shown).

Validation of SAGE Using RT-PCR. To validate the differential expression of candidate genes identified by mapping of gene identities to the SAGEtag results, we performed semiquantitative RT-PCR analysis on two short-term normal pancreatic ductal cultures (H48 and H116) and four pancreatic cancer cell lines (AsPc1, CAPAN1, CAPAN2, and PL45). These two additional pancreatic ductal epithelial samples, which were derived from different individuals, were used because of the limited quantity of HX and H126 primary ductal epithelial samples. This validation thereby also served as a test of the generality of the results. Eight genes from among those up-regulated (Table 3) were selected and assayed by RT-PCR. Differential expression was confirmed for four genes (*S100A4*, *TSPAN-1*, *CEACAM6*, and *ALG-2*) as shown in Fig. 2. Interestingly, the degree of differential expression between normal ductal cells and pancreatic adenocarcinoma cells detected by RT-PCR approximated the differences observed in their respective SAGE tag counts. Lack of validation of some genes (*keratin 19*, *claudin 4*, *basic transcription factor 3*, and *adenylyl cyclase-associated protein*) could in part be attributable to the statistical false discovery rate (~18%, see below), the incomplete current state of tag-to-gene mapping, and a lack of uniformity of gene expression among samples of different patients (a test of generality).

Table 2 *Pair-wise comparison of SAGE profiles*[a]

	D1	D2	CL1	CL2	CL3	CL4	CL5	CL6	I1	I2
D1	1	0.981	0.935	0.515	0.748	0.797	0.291	0.286	0.148	0.132
D2		1	0.930	0.552	0.759	0.799	0.293	0.288	0.152	0.134
CL1			1	0.514	0.761	0.802	0.289	0.277	0.152	0.134
CL2				1	0.478	0.490	0.204	0.189	0.123	0.093
CL3					1	0.927	0.411	0.369	0.225	0.194
CL4						1	0.408	0.372	0.218	0.190
CL5							1	0.569	0.351	0.396
CL6								1	0.860	0.880
I1									1	0.978
I2										1

[a] Pearson correlation coefficients (r). A subset of SAGE data set was obtained by filtering with the criterion that each tag have at least one observation more than twice among the 10 libraries. This subset was used for calculation of r values. The generated r values were similar using different filtration criteria such as no filtration, or requiring each tag to have >5, 10, and 15 observation in at least one sample. D1 and D2 are normal pancreatic ductal epithelium, HX and H126, respectively. CL1 to CL6 are pancreatic cancer cell lines, Hs766T, Panc-1, CAPAN1, CAPAN2, PL45, and AsPc1, respectively. I1 and I2 are primary pancreatic cancer, 91-16113 and 96-6252, respectively.

Table 3 *Differential gene expression in pancreatic cancer as compared with normal ductal epithelium*[a]

Tag	X[b]	Gene/ESTs	N	D1	D2	CL1	CL2	CL3	CL4	CL5	CL6	I1	I2	Function
Up-regulated in cancer														
GACATCAAGT		Keratin 19	64	0	0	0	20	83	125	502	177	64	98	Cytoskeletal and microfibrillar
ATGTGTAACG	X	S100A4 (Mst1)	9	0	0	0	40	112	77	0	291	55	70	Calcium-binding protein
ATCGTGGCGG		Claudin 4	72	0	0	0	28	29	116	40	81	58	56	Tight junction barrior function
GCCTACCCGA		Trop-2	23	0	0	0	4	49	99	121	0	55	11	Tumor-associated calcium signal transducer
CAAACCATCC		Keratin 18	112	9	9	28	132	151	142	1132	425	26	75	Cytoskeletal and microfibrillar
GCAGGGCCTC		FXYD3 (Mat-8)	75	0	0	0	0	62	47	40	19	81	11	FXYD domain-containing ion transport regulator
TTAGTTTTTA	X	Tissue-type plasminogen activator	15	0	0	0	0	107	39	9	9	3	2	Tissue remodeling and degradation/protease
CCTGCTTGTC		HE4	12	0	0	0	0	26	17	93	6	12	5	Secreted ptotease inhibitor
GCCCAGCATT	X	Prostate stem cell antigen	1	0	0	0	0	49	73	12	6	3	2	GPI-anchored glycoprotein/prostate-specific
CTGAGACAAA		Basic transcription factor 3	26	1	1	47	20	0	44	40	40	15	8	General transcription factor
GAATGATTTC		Collagen, type I, alpha 1	17	0	0	0	20	57	17	9	34	15	2	may be an extracellular matrix protein
GAGTTCGACC*		CHD-2	5	0	0	0	8	8	30	71	19	35	8	Sequence selective DNA binding
GGAACTGTGA		AIF-1 Tetraspan 1	39	0	0	10	4	31	34	16	50	26	14	Ionized calcium binding adapter Possible interconnecting cell surface molecules
TTTGGGCCTA	X	Cysteine-rich protein 1 (intestinal)	10	0	0	0	68	13	30	3	16	49	22	Potential intracellular zinc transport protein
GCCGTGGAGA		Major vault protein	15	0	0	0	8	10	30	19	62	17	30	Nucleo-cytoplasmic transport/Drug resistant-related protein
GCTGTTGCGC		Ribosomal protein S20	38	3	3	0	132	81	125	28	34	26	30	Protein synthesis
ATTTTCTAAA	X	Anterior gradient 2 homolog	11	0	0	0	8	62	47	3	3	3	19	Secreted
AAGGATAAAA	X	CEACAM6	5	0	0	0	0	34	4	31	47	3	56	Carcinoembryonic antigen-related cell adhesion molecule
AAGGTAGCAG		Adenylyl cyclase-associated protein	13	0	0	0	20	26	30	16	22	15	30	Membrane protein/regulatory bifunctional roles
ATGTAAAAAA*	X	Hypothetical protein FLJ11068 Leukocyte ig-like receptor Lysozyme	11	0	0	0	4	47	17	0	43	15	19	
CCCCCTGCAG	X	Homo sapiens cDNA: FLJ23356 fis Mesothelin	2	0	0	0	4	10	13	31	47	6	33	GPI-anchored/cell adhesion/mesothelioma and ovarian cancer ag
GTCTCCTAAT		Retinoic acid induced 3	5	0	0	0	0	13	13	53	25	12	2	Putative G protein-coupled receptor
GAGGGCCGGT		Hypothetical protein FLJ10903	16	0	0	0	8	16	30	19	31	12	5	
TGTGGGAAAT		Secretory leukocyte protease inhibitor	6	0	0	0	0	42	43	9	9	0	2	Secreted serine protease inhibitor
TGGCTGGGAA		Endobrevin	13	0	0	0	12	5	34	34	16	15	16	Vesicle-associated membrane protein 8
CCAAGTTTTT		Coated vesicle membrane protein	6	0	0	10	28	23	30	12	6	17	19	Could bind cargo molecules to collect them into vesicles
ACTGTATTTT		Homo sapiens cDNA: FLJ22182 fis	2	0	0	0	8	23	0	56	12	3	2	Maybe secreted serine protease
CTAAGACTTC		No reliable matches	71	18	18	10	68	70	77	487	1054	151	252	
GCGGCGCTGC		No reliable matches	11	0	0	0	0	23	43	22	6	6	8	
GGCACCGTGC		No reliable matches	16	0	0	0	0	18	30	25	19	9	30	
CTCGCGCTGG		Claudin 3	318	0	0	0	12	16	9	9	40	3	8	Tight junction barrier function
AGCAGATCAG		S100A10	112	15	9	58	380	135	133	140	264	197	246	Calcium-binding protein
TATATTTTCT*		Transglutaminase 2 Homo sapiens cDNA: FLJ21010 fis sapiens cDNA: FLJ21680 fis	7	0	0	0	0	26	9	43	6	9	14	
TGCCTTACTT		ALG-2	9	0	0	10	8	26	13	9	28	3	2	Ca(2+)-binding protein required for T cell receptor-, Fas-, & glucocorticoid-induced cell death
GCCGCTACTT		RAMP1	2	0	0	0	4	10	17	3	47	0	8	Receptor (calcitonin) activity modifying protein
TTTCCTCTCA		14-3-3 protein sigma (stratifin)	6	0	0	0	4	18	0	40	19	6	5	p53-regulated inhibitor of G2/M progression
ATGCGGAGTC		Zona occludens 3	6	0	0	0	0	8	22	16	34	0	5	Tight junction protein

[a] All numbers have been converted to tags per 100,000. Data for primary pancreatic cancer tissues, 91-16113 and 96-6252 were from Zhang *et al.* (19). Data for other pancreatic samples are from SAGE libraries that were constructed for this study as described in "Materials and Methods." *N*, bulk tissues of normal colorectal mucosae (NC1 and NC2), prostate (PR317, Chen Normal Pr), breast (BrN), ovary (HOSE 4 and IOSE 92-11), brain (BB542 whitematter), and vascular (Duke HMVEC). Data of these normals were obtained from the SAGEmap site as of March 26, 2001, and average numbers of nine samples were used.

[b] X indicates a differentially expressed tag which is identified by X-profiler as being among the 50 genes most likely represent at least 10-fold difference between sample groups of pancreatic cancer cell lines CL2 to CL4 *vs* normal samples from ovary (IOSE29-11), prostate (PR317 and TSU), and pancreas (HX and H126).

[c] ESTs, tag matches multiple expressed sequence tags.

* Tags representing multiple genes.

Differences between Pancreatic Cell Lines and Normal Epithelium: Analysis of Individual Genes. The identification of genes differentially expressed in a disease state as compared with normal tissue is one major goal of gene expression profile analyses. In this study, we focused primarily on a comparison of pancreatic cancer cell lines to normal pancreatic ductal epithelial cells. The majority of pancreatic cancers (>90%) originate from ductal epithelial cells; however, primary pancreatic cancer tissue

Table 3 *Continued*

TGAAAAAAAA*		Homo sapiens cDNA FLJ14267 fis Homo sapiens cDNA FLJ21963 fis Ankyrin repeat-containing ASB-2 Gastric protein ZA43P DEAD/H box polypeptide 11	21	0	0	10	12	26	13	9	16	6	8	
AATGGAATGG		Homo sapiens cDNA FLJ12145 fis	7	0	0	0	0	29	34	6	6	0	5	
GAAATTTAAA		HMG-1	15	0	0	0	4	29	30	3	9	12	0	Bind single-stranded DNA & unwinds double-stranded DNA
TCATAGAAAC		Sec61 homolog	10	0	0	20	8	26	13	19	9	3	2	Possible ribosome receptor in ER
CTGAGAAACT		Ferritin, heavy polypeptide 1	4	0	0	0	4	5	9	19	37	6	16	Iron binding protein
AAATAAAAGC*		Thyrotropin receptor Villin 2	11	0	0	10	4	10	4	31	19	26	8	
CCCTGATTTT		NAT-1	13	0	0	0	24	3	4	25	12	23	22	Translation repressor
GGGGCAGCCG		ESTs[c]	4	0	0	0	24	5	17	0	22	3	2	
CCCTCAATCC		Interleukin 18	4	0	0	0	0	5	9	9	43	3	2	Stimulate interferon gamma production
AAAAAGCAGA		Superoxide dismutase 1, soluble	4	0	0	10	12	8	17	12	16	0	2	Radical metabolism
GACCACGAAT	X	Cathepsin H	6	0	0	0	0	29	9	6	19	0	16	Cysteine protease
AATTTTATTT		PCBP1	8	1	1	4	1	0	18	16	22	6	3	Major cellular poly(rC)-binding proteins
Down-regulated in cancer														
GCTGCTGCGC	X	Aldo-keto reductase family 1 B1	1	226	319	298	4	0	0	0	0	0	0	Reduces glucose and other carbonyl-containing substrates
TAGGCATTCA	X	Ribosomal protein L26	0	118	115	115	0	0	0	0	0	0	0	Protein synthesis
GTGCCCGTGC		Triosephosphate isomerase 1	1	59	74	48	0	3	0	0	0	0	0	Glycolysis
TCAGGGATCT		No reliable matches	0	40	53	38	0	0	0	0	0	0	0	
GGTGACCACC	X	Nuclear hormone receptor MB67	11	61	58	29	0	0	0	0	0	0	0	Orphan member of the nuclear hormone receptor superfamily
CTGAGACGAA		Basic transcription factor 3	0	37	43	58	0	0	0	0	0	0	0	General transcription factor
GGGCTCGGGG		SCGF-alpha	2	37	31	38	0	0	0	0	0	6	2	Stem cell growth factor; lymphocyte secreted C-type lectin
GCAAAAGCCG		No reliable matches	0	34	31	29	0	0	0	0	0	0	0	
GGTTATTTTG		PAI, type I	7	31	31	38	0	0	0	0	0	0	8	Serine (or cysteine) proteinase inhibitor
CAAACTGGTC		Stanniocalcin 1	1	47	53	10	0	0	0	0	0	3	2	Stimulates renal phosphate reabsorption
GTGGCGCACT*		Selenoprotein W,1 E3 ubiqutin ligase SMURF2 Omo sapiens cDNA FLJ11697 fis	2	43	24	28	0	3	0	0	0	3	2	
GGGGCACCCG		Lamin, beta 1	2	96	87	96	0	0	13	3	0	0	0	Member of a family of nuclear envelope proteins
ACATTTCCAA		G0S2	18	49	34	67	0	3	0	0	3	12	2	Putative lymphocyte G0/G1 switch protein
CTAACGCAGC		AP-1 (proto-oncogene c-Jun)	4	15	24	0	0	0	0	0	0	12	2	Transcription factor
GGCTCGTCTG		No reliable matches	0	19	22	29	0	0	0	0	0	0	0	
CCGAAGTCGA		ESTs[c]	4	52	55	77	0	5	0	6	0	0	2	
CCTGCGGTCT		No reliable matches	0	25	25	38	0	3	0	0	0	0	0	
TGGCTGCAGA		EST	0	19	43	10	0	0	4	0	0	0	0	
GAGAAGGGCA*		Matrilin 1 Sphingosine kinase 1	2	37	37	29	4	0	0	3	0	3	14	Major component of extracellular matrix kinase
ATCCGGACCC		GADD34	2	21	24	29	0	3	0	0	0	6	0	Apoptosis associated /growth arrest & DNA damage induced
ACATCCCAGA		Coronin 1C	6	24	34	19	4	0	0	3	0	3	5	Actin-binding protein
TGGCTCCTCC		L-plastin	3	34	37	0	0	5	4	0	0	3	2	Actin-binding protein
ACATTGGGTT		ATP1B3	2	12	16	0	0	0	0	0	0	0	0	ATPase, Na+/K+ transporting, beta 3
TGAGGAGGTT		Hypothetical protein	1	12	16	0	0	0	0	0	0	0	0	
CAAGAGAGTA		GFA-2	1	21	15	10	4	0	0	0	0	0	5	Hexosamine pathway/high glucose levels induce TGF-beta1 production
GCGGCAGCGG		SMAP	3	21	15	10	0	0	0	3	0	3	0	Smg GDS-associated /G protein GDP dissociation stimulator
GCTGCGCAGA		ESTs[c]	3	31	18	19	0	3	4	0	0	0	2	

removed from a patient contains only a small fraction of cancer cells (usually <30%; Ref. 26). Cell-type heterogeneity of bulk tumors was explored in a previous study (11).

We used a four-step data reduction algorithm to identify genes differentially expressed between pancreatic cancer cells and normal ductal epithelium. First, we tested for tags exhibiting consistent differences (Student's *t* test) between two groups of samples, the two normal pancreas ductal epithelial cells (HX and H126) *versus* the five "non-normoid" pancreatic cancer cells (AsPc1, CAPAN1, CAPAN2, Panc-1, and PL45). A total of 669 tags that had a $P < 0.02$ were

Table 3 *Continued*

AGGCCACCTC	Frequenin (Drosophila) homolog	1	6	27	0	0	0	0	0	3	0	5	Neuronal Ca^{2+} sensor-1/calcium-binding protein
CCTGGAATGA*	Homo sapiens cDNA FLJ12683 fis ESTs[c]	2	15	18	10	0	3	0	3	0	0	2	
TCAAGCCATC	Early growth response 1	9	15	12	0	0	0	0	0	3	32	2	Transcriptional regulator
GGAGGAGGAG*	Tolloid-like protein	1	18	18	0	0	8	0	0	0	9	5	BMP-1 and Tolloid-related metalloproteases
TGCAGCGCCT	KIAA0460 protein Uridine phosphorylase	7	229	235	259	56	5	4	12	28	12	8	Catalysis the reversible phosphorolysis of uridine
CTTCCTGGCC	Nuclear receptor subfamily 1	2	19	12	0	0	0	0	0	0	3	0	Heterodimer with the retinoid-X receptor
GAAAGTGGCT	A novel transmembrane protein	0	12	9	0	0	0	0	0	0	0	0	Have two follistatin modules and an EGF domain
GACAGATGGA	KIAA1533 protein	1	9	12	0	0	10	0	0	0	3	3	
AGGCTACGGG	Ribosomal protein L13a	1	9	12	0	0	10	0	0	0	3	3	Protein synthesis
CAGAGGCGTC	Translation initiation factor 2C	0	25	16	0	4	19	3	0	0	0	0	Protein synthesis

identified. Among the 669 tags, 376 tags were up-regulated, and 293 tags were down-regulated in pancreatic cancer cells as compared with the normal duct cells. Second, we retained only the tags that expressed differences of ≥10-fold. To calculate fold difference for each unique tag, the average of normalized tags from the cancer cell lines was divided by the average of normalized tags from the two normal duct cultures. For the tags that had no occurrences in a sample set, the arbitrary number 1 was assigned for this calculation. Third, we filtered out any tags not expressed in at least one primary tumor sample at a level >two tags/100,000 (this was performed only in the identification of overexpressed tags). The latter criterion was applied to reduce the possibility of identifying genes that had experienced induction because of cell culture. Finally, we set a cutoff to draw the greatest attention to genes of higher expression levels, requiring expression at an average of ≥12 tags/100,000 in the five cancer cell lines (for the genes up-regulated in cancer) or in the two normal duct samples (for the genes down-regulated). This algorithm identified a group of 86 tags that exhibited a robust overexpression or underexpression in pancreatic cancer cells (Table 3). The genes, *tissue-type plasminogen activator*, *cathepsin H*, and *CEACAM6*, which are known to be up-regulated in pancreatic cancer (10, 12, 13), were confirmed in this study (Table 3), validating our current approach to identify differentially expressed genes.

A permutation procedure was performed to estimate the false discovery rate. There are 21 possible permutations for which two libraries could be considered the comparison set (corresponding in form to the two ductal libraries of the original analysis). We evaluated the four-step process for each of these assignment choices. A total of 330 tags emerged from these pseudo-trials, which included the 86 of the original permutation. Thus, the average number of tags produced by these trials was 16, a conservative estimate of the number of false-positive tags to be expected under the null hypothesis of no real difference between pancreatic cancer and normal ductal expression profiles. The false discovery tag estimate of 16 tags is 18% of the observed tag count for the original analysis.

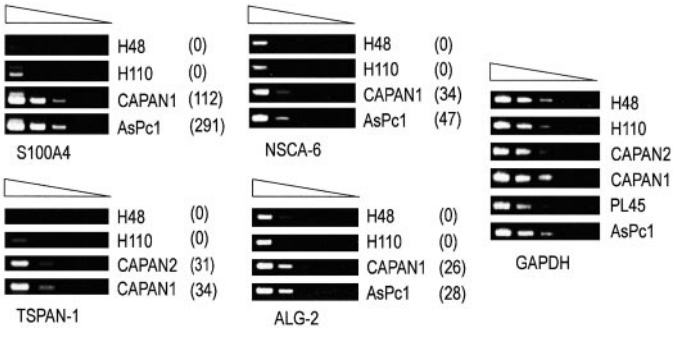

Fig. 2. RT-PCR analysis of genes overexpressed in pancreatic adenocarcinoma cells as compared with normal ductal epithelial cells. Total cellular RNA from pancreatic cancer cell lines (AsPc1CAPAN1 and CAPAN2) and normal ductal epithelial cells (H48 and H110) was analyzed. *Triangles* represent concentration gradients of cDNA templates: reverse transcribed samples were diluted by factors of 100, 1,000, 10,000, and 100,000 to serve as templates for PCR. The amplified products were separated electrophoretically in agarose gels and visualized by ethidium bromide fluorescence. *Numbers in parentheses* represent the results of SAGE analysis in tags per 100,000. NCSA-6 is CEACAM6.

A potential source of bias could be the known variable GC content bias present in most SAGE libraries (27). The following approaches excluded this bias as a significant source of artifactual results in our study. First, the GC content of the two comparison libraries for each of our 21 permutations (above) was found not to correspond to the number of tags produced in each permutation. Second, the set of 86 observed differential tags retained a spectrum of GC content (Table 3). Third, the genes that failed to conform by RT-PCR did not represent a skewed GC content as compared with those that were confirmed. Thus, whereas one must consider variable ditag melting and the resultant enrichment of GC content as an uncontrolled determinant in analysis of SAGE data, this did not appear to exert a major effect in the assessment of differential gene expression.

The NCBI SAGEmap website[8] provides the X-profiler program to reduce SAGE data (21, 22). Using this additional tool, we also identified genes differentially expressed in pancreatic cancer cells as compared with normal control tissues or cultures of pancreas and other organs that were available in the SAGEmap database as of March 26, 2001. Several genes in Table 3 were consistently identified as differentially expressed using X-profiler (Table 3).

Biological and Clinical Implications for Pancreatic Cancer. A notable feature of the overexpressed genes identified in this study is that nearly half of the genes comprise secretory, cell-surface, transmembrane, and tight junction protein coding genes. This could correspond to altered cellular attachments and cell surface architecture, resulting in aberrant cell-cell interactions that are a reproducible characteristic of cancer cells. One reason to explore such alterations would be to develop new therapeutic strategies. Another use is suggested by the detection of secretory proteins such as HE4 (a putative ovarian cancer marker; Refs. 24, 28), PSCA (a putative prostate cancer marker; Ref. 29), and CEACAM6, which suggest a use as potential diagnostic markers. Indeed, one of the markers, PSCA, identified in this study was subsequently developed as a histological marker of pancreatic malignancy (30) and is a secreted protein detectable in the serum of a set of prostate cancer patients evaluated (29). Interestingly, a group of ion-homeostasis related proteins, especially those specific for the calcium ion (Ca^{2+}) such as S100A4, S100A10, Trop-2, AIF-1, and ALG-2, were identified as overex-

[8] Internet address: http://www.ncbi.nlm.nih.gov/SAGE/sagexpsetup.cgi.

pressed. Another example is the major vault protein, which functions to produce multidrug resistance in a cancer cell (31); overexpression of this gene in cancer is reported to predict the response to chemotherapy in several tumor types (32). A group of genes down-regulated in pancreatic cancer was also identified (Table 3).

Differences between Pancreatic Cell Lines and Normal Epithelium: The Question of the Minimal Deviation Malignant Profile. The Pearson correlation coefficient (Table 2), hierarchical clustering, and scatter plot analysis (Fig. 1, *A* and *B*) classified Hs766T as closely resembling normal ductal epithelium in its expression profile: a "normoid" cancer cell line. However, differences in gene expression between Hs766T and normal ductal epithelium might suggest that this cell line contains a minimal set of changes responsible for key features of pancreatic neoplasia. On such a comparison, SAGE tags mapped to genes encoding protein translocation complex β, regulator of G protein 5, nuclear phosphoprotein B23, MKP-1-like protein tyrosine phosphatase, tumor necrosis factor α-inducible protein, catenin α 1 (102 kDa), RAD51 (*Saccharomyces cerevisiae*) homologue C, guanine nucleotide-binding protein γ 5, BCL2-associated athanogene 3, and 21 others were overexpressed in the "normoid" cancer cell Hs766T by >10-fold.

PCA detected groups of genes that could represent cell line-specific expression deficits, that is, genes that were not expressed in one cell line but were expressed in the others. For example, 116 known genes lacked expression in Hs766T but were expressed in all of the other cell lines. Examination of the chromosomal locations of these genes revealed that the cytogenetic distributions of these genes were concentrated in several regions and did not have a random distribution. For an example, the cytogenetic locations of 13 of the 116 genes deficient in Hs766T cells reside between 214 and 263 cR_{3000} on chromosome 11 (33), which is closely associated with the fragile site of chromosomal band 11q13. This "regional dropout" of gene expression in a single cell line raised the possibility of a homozygous deletion, but none of the genes (10 were tested) were absent from the genomic DNA, and these down-regulated genes were physically interspersed with expressed genes.[9] Possible explanations of this regional gene dropout in gene expression include regional gene silencing by methylation (34, 35) and regional chromatin structural changes (36, 37).

Gene Expression *versus* Known Genetic Mutations. Profiles of genetic changes are well established in pancreatic cancer cell lines (18). For example, homozygous deletions and mutations of tumor suppressor genes, such as *p53*, *MADH4*, *MKK4*, and *BRCA2*, were known within the cell lines studied by SAGE. We performed group-wise comparisons of SAGE data, including *p53* wild-type *versus p53* mutated, *MADH4* wild-type *versus MADH4* homozygous deleted/mutated, *BRCA2* wild-type *versus BRCA2* mutated, and *MKK4* wild-type *versus MKK4* homozygous deleted/mutated, but no distinct patterns emerged. In the Student *t* test, the number of expressed genes that achieved a given *P* cutoff level appeared to depend primarily on the power of the comparison (*i.e.*, the number of cell lines in each arm of the comparison). No other pattern could be discerned with the dataset, as might be expected from the small numbers of cell lines available for comparison.

We may offer the following summary and perspective. The progression from normal cell to cancer cell undoubtedly involves stochastic alterations in genetic composition and gene expression; however, selective pressures related to the process of tumorigenesis and metastasis result in the accumulation of common sets of defects that contribute to survival and spread of tumors. An unbiased survey by SAGE analysis identified a candidate list of differentially expressed genes (Table 3). This gene set likely includes genes of which the deregulation contributes to tumorigenesis in the pancreas. Such genes may be robust markers of pancreatic neoplasia and suggest new targets for directed diagnostic and therapeutic approaches.

[9] Unpublished observations.

ACKNOWLEDGMENTS

We thank Dr. Greg J. Riggins and CGAP for helpful suggestions and for the sequencing of SAGE libraries from pancreatic cancer cell cultures.

REFERENCES

1. Greenlee, R. T., Murray, T., Bolden, S., and Wingo, P. A. Cancer statistics, 2000. CA Cancer J. Clin., *50:* 7–33, 2000.
2. Kang, S., Xu, H., Duan, X., Liu, J. J., He, Z., Yu, F., Zhou, S., Meng, X. Q., Cao, M., and Kennedy, G. C. *PCD1*, a novel gene containing PDZ and LIM domains, is overexpressed in several human cancers. Cancer Res., *60:* 5296–5302, 2000.
3. Schweinfest, C. W., Graber, M. W., Chapman, J. M., Papas, T. S., Baron, P. L., and Watson, D. K. CaSm, a Sm-like protein that contributes to the transformed state in cancer cells. Cancer Res., *57:* 2961–2965, 1997.
4. Wallrapp, C., Hahnel, S., Muller-Pillasch, F., Burghardt, B., Iwamura, T., Ruthenburger, M., Lerch, M. M., Adler, G., and Gress, T. M. A novel transmembrane serine protease (TMPRSS3) overexpressed in pancreatic cancer. Cancer Res., *60:* 2602–2606, 2000.
5. Wang, W., Abbruzzese, J. L., Evans, D. B., and Chiao, P. J. Overexpression of urokinase-type plasminogen activator in pancreatic adenocarcinoma is regulated by constitutively activated RelA. Oncogene, *18:* 4554–4563, 1999.
6. Kleeff, J., Ishiwata, T., Kumbasar, A., Friess, H., Buchler, M. W., Lander, A. D., and Korc, M. The cell-surface heparan sulfate proteoglycan glypican-1 regulates growth factor action in pancreatic carcinoma cells and is overexpressed in human pancreatic cancer. J. Clin. Investig., *102:* 1662–1673, 1998.
7. Lutz, M. P., Esser, I. B., Flossmann-Kast, B. B., Vogelmann, R., Luhrs, H., Friess, H., Buchler, M. W., and Adler, G. Overexpression and activation of the tyrosine kinase Src in human pancreatic carcinoma. Biochem. Biophys. Res. Commun., *243:* 503–508, 1998.
8. Muller-Pillasch, F., Zimmerhackl, F., Lacher, U., Schultz, N., Hameister, H., Varga, G., Friess, H., Buchler, M., Adler, G., and Gress, T. M. Cloning of novel transcripts of the human guanine-nucleotide-exchange factor Mss4: *in situ* chromosomal mapping and expression in pancreatic cancer. Genomics, *46:* 389–396, 1997.
9. Muller-Pillasch, F., Wallrapp, C., Bartels, K., Varga, G., Friess, H., Buchler, M., Adler, G., and Gress, T. M. Cloning of a new Kunitz-type protease inhibitor with a putative transmembrane domain overexpressed in pancreatic cancer. Biochim. Biophys. Acta, *1395:* 88–95, 1998.
10. Paciucci, R., Berrozpe, G., Tora, M., Navarro, E., Garcia de Herreros, A., and Real, F. X. Isolation of tissue-type plasminogen activator, cathepsin H, and non-specific cross-reacting antigen from SK-PC-1 pancreas cancer cells using subtractive hybridization. FEBS Lett., *385:* 72–76, 1996.
11. 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.
12. Gress, T. M., Muller-Pillasch, F., Geng, M., Zimmerhackl, F., Zehetner, G., Friess, H., Buchler, M., Adler, G., and Lehrach, H. A pancreatic cancer-specific expression profile. Oncogene, *13:* 1819–1830, 1996.
13. Gress, T. M., Wallrapp, C., Frohme, M., Muller-Pillasch, F., Lacher, U., Friess, H., Buchler, M., Adler, G., and Hoheisel, J. D. Identification of genes with specific expression in pancreatic cancer by cDNA representational difference analysis. Genes Chromosomes Cancer, *19:* 97–103, 1997.
14. Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. Serial analysis of gene expression. Science (Wash. DC), *270:* 484–487, 1995.
15. Hruban, R. H., Wilentz, R. E., and Kern, S. E. Genetic progression in the pancreatic ducts. Am. J. Pathol., *156:* 1821–1825, 2000.
16. Caldas, C., Hahn, S. A., da Costa, L. T., Redston, M. S., Schutte, M., Seymour, A. B., Weinstein, C. L., Hruban, R. H., Yeo, C. J., and Kern, S. E. Frequent somatic mutations and homozygous deletions of the *p16* (*MTS1*) gene in pancreatic adenocarcinoma. Nat. Genet., *8:* 27–31, 1994.
17. Kolar, C., Caffrey, T. C., Hollingsworth, M. A., Sheetz, M., Sutherlin, M., Weide, L., and Lawson, T. Duct epithelial cells cultured from human pancreas processed for transplantation retain differentiated ductal functions. Pancreas, *15:* 265–271, 1997.
18. Kern, S. E. Molecular genetic alterations in ductal pancreatic adenocarcinomas. Med. Clin. North Am., *84:* 691–695, xi, 2000.
19. 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 (Wash. DC), *276:* 1268–1272, 1997.
20. St. Croix, B., Velculescu, V. E., Zhang, L., Zhou, W., Traverso, G., Vogelstein, B., and Kinzler, K. W. MicroSAGE. Detailed protocol, Version 1.0e. (http://www.sagenet.org/sage_protocol.htm), 2000.
21. Lal, A., Lash, A. E., Altschul, S. F., Velculescu, V., Zhang, L., McLendon, R. E., Marra, M. A., Prange, C., Morin, P. J., Polyak, K., Papadopoulos, N., Vogelstein, B., Kinzler, K. W., Strausberg, R. L., and Riggins, G. J. A public database for gene expression in human cancers. Cancer Res., *59:* 5403–5407, 1999.

22. Lash, A. E., Tolstoshev, C. M., Wagner, L., Schuler, G. D., Strausberg, R. L., Riggins, G. J., and Altschul, S. F. SAGEmap: a public gene expression resource. Genome Res., *10:* 1051–1060, 2000.
23. Jackson, J. E. A User's Guide to Principal Components. New York: Wiley, 1991.
24. Hough, C. D., Sherman-Baust, C. A., Pizer, E. S., Montz, F. J., Im, D. D., Rosenshein, N. B., Cho, K. R., Riggins, G. J., and Morin, P. J. Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer. Cancer Res., *60:* 6281–6287, 2000.
25. 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.
26. Seymour, A. B., Hruban, R. H., Redston, M., Caldas, C., Powell, S. M., Kinzler, K. W., Yee, C. J., and Kern, S. E. Allelotype of pancreatic adenocarcinoma. Cancer Res., *54:* 2761–2764, 1994.
27. Margulies, E. H., Kardia, S. L., and Innis, J. W. Identification and prevention of a GC content bias in SAGE libraries. Nucleic Acids Res., *29:* E60, 2001.
28. Schummer, M., Ng, W. V., Bumgarner, R. E., Nelson, P. S., Schummer, B., Bednarski, D. W., Hassell, L., Baldwin, R. L., Karlan, B. Y., and Hood, L. Comparative hybridization of an array of 21,500 ovarian cDNAs for the discovery of genes overexpressed in ovarian carcinomas. Gene (Amst.), *238:* 375–385, 1999.
29. Reiter, R. E., Gu, Z., Watabe, T., Thomas, G., Szigeti, K., Davis, E., Wahl, M., Nisitani, S., Yamashiro, J., Le Beau, M. M., Loda, M., and Witte, O. N. Prostate stem cell antigen: a cell surface marker overexpressed in prostate cancer. Proc. Natl. Acad. Sci. USA, *95:* 1735–1740, 1998.
30. Argani, P., Rosty, C., Reiter, R. E., Wilentz, R. E., Murugesan, S. R., Leach, S. D., Ryu, B., Skinner, H. G., Goggins, M., Jaffee, E. M., Yeo, C. J., Cameron, J. L., Kern, S. E., and Hruban, R. H. Discovery of new markers of cancer through serial analysis of gene expression: prostate stem cell antigen is overexpressed in pancreatic adenocarcinoma. Cancer Res., *61:* 4320–4324, 2001.
31. Scheffer, G. L., Wijngaard, P. L., Flens, M. J., Izquierdo, M. A., Slovak, M. L., Pinedo, H. M., Meijer, C. J., Clevers, H. C., and Scheper, R. J. The drug resistance-related protein LRP is the human major vault protein. Nat. Med., *1:* 578–582, 1995.
32. Scheffer, G. L., Schroeijers, A. B., Izquierdo, M. A., Wiemer, E. A., and Scheper, R. J. Lung resistance-related protein/major vault protein and vaults in multidrug-resistant cancer. Curr. Opin. Oncol., *12:* 550–556, 2000.
33. Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega-Czarny, N., Spillett, D., Muselet, D., Prud'Homme, J. F., Dib, C., Auffray, C., Morissette, J., Weissenbach, J., and Goodfellow, P. N. A radiation hybrid map of the human genome. Hum. Mol. Genet., *5:* 339–346, 1996.
34. Prokhorchuk, A. V., and Ruzov, A. S. Genome methylation and its role in functioning of the eukaryotic organism. Genetika, *36:* 1475–1486, 2000.
35. Feinberg, A. P. DNA methylation, genomic imprinting and cancer. Curr. Top. Microbiol. Immunol., *249:* 87–99, 2000.
36. Vogelauer, M., Wu, J., Suka, N., and Grunstein, M. Global histone acetylation and deacetylation in yeast. Nature (Lond.), *408:* 495–498, 2000.
37. Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M. C., van Asperen, R., Boon, K., Voute, P. A., Heisterkamp, S., van Kampen, A., and Versteeg, R. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. Science (Wash. DC), *291:* 1289–1292, 2001.

Relationships and Differentially Expressed Genes among Pancreatic Cancers Examined by Large-scale Serial Analysis of Gene Expression

Byungwoo Ryu, Jessa Jones, Natalie J. Blades, et al.

Cancer Res 2002;62:819-826.