# Screening of genes specifically activated in the pancreatic juice ductal cells from the patients with pancreatic ductal carcinoma

**Koji Yoshida,[1,6] Shuichi Ueno,[1,2] Toshiyasu Iwao,[6] Souichirou Yamasaki,[7] Akira Tsuchida,[7] Ken Ohmine,[1,3] Ruri Ohki,[1,2] Young Lim Choi,[1] Koji Koinuma,[1,4] Tomoaki Wada,[1,5] Jun Ota,[1] Yoshihiro Yamashita,[1] Kazuaki Chayama,[7] Kazuhiro Sato[6] and Hiroyuki Mano[1,8]**

Divisions of [1]Functional Genomics, [2]Cardiology and [3]Hematology, Departments of [4]Surgery and [5]Gynecology, Jichi Medical School, 3311-1 Yakushiji, Kawachi-gun, Tochigi 329-0498, [6]Gastroenterological Center, Aizu Central Hospital, 1-1 Tsurugamachi, Aizuwakamatsu-shi, Fukushima 956-8611 and [7]First Department of Internal Medicine, Faculty of Medicine, Hiroshima University, 1-2-3 Kasumi, Hiroshima-shi, Hiroshima 734-8551

(Received November 8, 2002/Revised January 7, 2003/Accepted January 14, 2003)

**Pancreatic ductal carcinoma (PDC) is one of the most intractable human malignancies. Surgical resection of PDC at curable stages is hampered by a lack of sensitive and reliable detection methods. Given that DNA microarray analysis allows the expression of thousands of genes to be monitored simultaneously, it offers a potentially suitable approach to the identification of molecular markers for the clinical diagnosis of PDC. However, a simple comparison between the transcriptomes of normal and cancerous pancreatic tissue is likely to yield misleading pseudopositive data that reflect mainly the different cellular compositions of the specimens. Indeed, a microarray comparison of normal and cancerous tissue identified the *INSULIN* gene as one of the genes whose expression was most specific to normal tissue. To eliminate such a "population-shift" effect, the pancreatic ductal epithelial cells were purified by MUC1-based affinity chromatography from pancreatic juice isolated from both healthy individuals and PDC patients. Analysis of these background-matched samples with DNA microarrays representing 3456 human genes resulted in the identification of candidate genes for PDC-specific markers, including those for *AC133* and carcinoembryonic antigen-related cell adhesion molecule 7 (*CEACAM7*). Specific expression of these genes in the ductal cells of the patients with PDC was confirmed by quantitative real-time polymerase chain reaction analysis. Microarray analysis with purified pancreatic ductal cells has thus provided a basis for the development of a sensitive method for the detection of PDC that relies on pancreatic juice, which is routinely obtained in the clinical setting. (Cancer Sci 2003; 94: 263–270)**

Pancreatic carcinoma remains the most intractable disorder among gastroenterological malignancies, with a 5-year survival rate of <5%.[1,2] More than 90% of pancreatic carcinomas are adenocarcinomas of ductal cell origin. In part because of the lack of disease-specific symptoms, individuals at an early stage of pancreatic carcinoma are rarely detected, and the probability of tumors being suitable for surgical resection at the time of discovery is low (10 to 20%). Several improvements in imaging analysis of pancreatic structure have recently been achieved, including endoscopic retrograde cholangiopancreatography (ERCP), magnetic resonance cholangiopancreatography (MRCP), and endoscopic ultrasound examination.[3] However, even with these procedures, it often remains difficult to distinguish pancreatic carcinoma from other disorders such as chronic pancreatitis. Furthermore, these methods usually detect only those pancreatic tumors with a diameter of >5 mm. Given the low 5-year survival rate (20 to 30%) even of individuals with small, resectable tumors, the sensitivity of current technologies is not sufficient to allow detection of pancreatic carcinoma at curable early stages. A cure for this disorder will thus depend on development of an approach that is able to detect tumors at an early stage of carcinogenesis.

Pancreatic ductal carcinoma (PDC) arises from epithelial cells of the pancreatic duct. Carcinoma cells of individuals with this condition are thus shed into pancreatic juice. Analysis of these cells appears a promising approach to the development of a sensitive method for the diagnosis of pancreatic carcinoma. Indeed, molecular biological analysis of these tumor cells has revealed a variety of genetic alterations associated with the pathogenesis of pancreatic carcinoma. Activating point mutations of the K-*RAS* proto-oncogene have thus been identified in >80% of individuals with pancreatic carcinoma,[4] and inactivation of the *TP53* tumor suppressor gene has been detected at a similar frequency.[5] Other mutations have been identified in the genes for p16, DPC4, and DCC.[6–8] However, K-*RAS* mutations are also evident at a relatively high frequency in nonmalignant pancreatic disorders.[9] To date, no molecular markers proven to be specific to carcinoma cells of pancreatic ductal origin have been identified.

DNA microarray analysis allows the simultaneous monitoring of the expression of thousands of genes[10,11] and is therefore a potentially suitable approach to identify PDC-specific genes. The high throughput of this methodology also may be disadvantageous, however. Without careful selection of samples for analysis or data normalization procedures, DNA microarray experiments yield large numbers of pseudopositive and pseudonegative results. In the case of PDC, a simple comparison of pancreatic tissue obtained from individuals with nonmalignant or cancerous conditions would likely not prove informative. Most normal pancreatic tissue comprises exocrine and endocrine cells, with ductal structures constituting only a small proportion of the total volume of the normal pancreas. In contrast, cancerous pancreatic tissue consists mostly of tumor cells that arise from ductal epithelial cells. A comparison between nonmalignant and cancerous tissue would thus likely identify differences between the gene expression profiles of exocrine and endocrine cells and that of tumor cells of ductal cell origin, rather than differences between those of normal and transformed cells of the same origin.

We now show that such a tissue comparison for PDC is indeed uninformative with regard to the identification of tumor-specific genes. To avoid this pitfall, we therefore adopted the strategy of "background-matched population (BAMP) screening,"[12] in which the sample characteristics are matched as closely as possible, with the exception of the feature of interest (in this case, transformation), before microarray analysis. To achieve this goal, we purified pancreatic carcinoma cells and normal ductal cells from pancreatic juice with the use of affinity chromatography based on the shared surface marker MUC1.

[8]To whom correspondence and reprint requests should be addressed.
E-mail: hmano@jichi.ac.jp

Comparison of these two cell preparations by DNA microarray analysis revealed a group of genes that are potential molecular markers specific to PDC.

## Materials and Methods

**Preparation of pancreatic ductal cells.** The study subjects comprised individuals who were subjected to ERCP and to the collection of pancreatic juice for cytological examination and who gave informed consent. The study was approved by the institutional review boards of Jichi Medical School, Aizu Central Hospital and Hiroshima University. Diagnosis of patients was confirmed on the basis both of the combination of results obtained by ERCP, cytological examination of pancreatic juice, abdominal computed tomography, and measurement of the serum concentration of CA19-9, as well as of follow-up observations. About one-third of each pancreatic juice specimen was used to purify MUC1$^+$ ductal cells. Cells were collected from the pancreatic juice by centrifugation and resuspended in 1 ml of MACS binding buffer [150 m*M* NaCl, 20 m*M* sodium phosphate (pH 7.4), 3% fetal bovine serum, 2 m*M* EDTA]. The cells were then incubated for 30 min at 4°C with 0.5 $\mu$g of mouse monoclonal antibodies to MUC1 (Novocastra Laboratories, Newcastle upon Tyne, UK), washed with MACS binding buffer, and mixed with MACS MicroBeads conjugated with antibodies to mouse immunoglobulin G (Miltenyi Biotec, Auburn, CA). The resulting mixture was subjected to chromatography on miniMACS magnetic cell separation columns (Miltenyi Biotec). The eluted MUC1$^+$ cells were divided into aliquots and stored at −80°C. Portions of the unfractionated cells as well as of the isolated MUC1$^+$ cells of each individual were stained with Wright-Giemsa solution to examine the purity of the ductal cell-enriched fractions.

**Isolation of RNA and microarray analysis.** Total RNA was extracted from the MUC1$^+$ cell preparations with the use of RNAzol B (Tel-Test, Friendswood, TX), and portions (20 $\mu$g) of the resulting preparations were subjected to amplification of mRNA with T7 RNA polymerase as described.[13] Biotin-labeled cRNA was synthesized from the amplified RNA (2 $\mu$g) with the use of the ExpressChip labeling system (Mergen, San Leandro, CA) and was then subjected to hybridization with microarrays (HO-1 to -3, Mergen) that contain oligonucleotides corresponding to a total of 3456 human genes (for a list of the genes, see http://www.mergen-ltd.com). The microarrays were then incubated consecutively with streptavidin, antibodies to streptavidin, and Cy3-conjugated secondary antibodies (Mergen). Detection and digitization of hybridization signals were performed with a GMS 418 array scanner (Affymetrix, Santa Clara, CA). The fluorescence intensity for each gene was normalized relative to the median fluorescence value for all genes in each array hybridization. Statistical analysis of the data was performed with GeneSpring 5.0 software (Silicon Genetics, Redwood, CA).

**Real-time polymerase chain reaction (PCR) analysis.** Portions of unamplified cDNA were subjected to the PCR with SYBR Green PCR Core Reagents (PE Applied Biosystems, Foster City, CA). Incorporation of the SYBR Green dye into the PCR products was monitored in real time with an ABI PRISM 7700 sequence detection system (PE Applied Biosystems), thereby allowing determination of the threshold cycle ($C_T$) at which exponential amplification of PCR products begins. The $C_T$ values for cDNAs corresponding to the *β-actin* gene and target genes were used to calculate the abundance of the target transcripts relative to that of β-actin mRNA. The oligonucleotide primers for PCR were as follows: 5′-CCATCATGAAGTGTGACGTGG-3′ and 5′-GTCCGCCTAGAAGCATTTGCG-3′ for β-actin cDNA, 5′-TCCTGGGACTGTGACTTTCA-3′ and 5′-CTTTTGGTCCAGACCCTCAA-3′ for small ubiquitin-like modifier (SUMO) 1 cDNA, 5′-CCATCATGAAGTGTGACGTGG-3′ and 5′-GTCCGCCTAGAAGCATTTGCG-3′ for carcinoembryonic antigen-related cell adhesion molecule (CEACAM) 7 cDNA, and 5′-GAGACTCAGAACACAACCTACCTG-3′ and 5′-AGCCAGTACTCCAATCATGATGCT-3′ for AC133 cDNA.

## Results

**Purification of ductal cells from pancreatic juice.** Pancreatic juice contains various types of cells, including pancreatic ductal cells, erythrocytes, neutrophils, and lymphocytes (Fig. 1A). Given that the proportions of these cellular components of pancreatic juice vary markedly among individuals, the purification of ductal cells is required for reliable comparison of gene expression profiles. Normal and cancerous pancreatic ductal cells express various mucins. Among those, MUC1 is known to be expressed in both normal and cancerous ductal cells, whereas others, such as MUC3 and MUC5, are differentially expressed in a disease-dependent manner.[14, 15] We therefore developed an affinity purification approach for pancreatic ductal cells based on MUC1 as a common surface marker. Cells specifically eluted from a magnetic bead separation column exhibited an epithelial cell-like morphology (Fig. 1B).

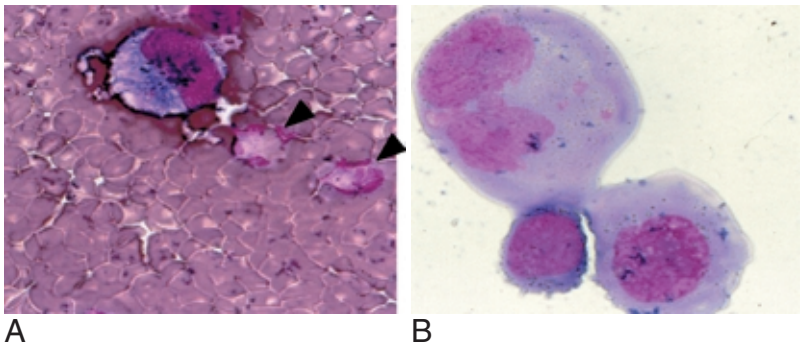
Previous attempts to identify genes whose expression is specific to PDC have often compared the gene expression profiles of normal and cancerous pancreatic tissues.[16] However, such an approach may result in the identification of genes that are differentially expressed between exocrine-endocrine cells and ductal cells. To directly examine if this is the case, we first compared the transcriptomes of surgically resected normal ($n$=1) and cancerous ($n$=2) pancreatic tissues by oligonucleotide microarray analysis. The digitized expression intensities for the 3456 human genes examined were normalized relative to the median expression level of all genes in each hybridization; in the case of the cancer tissue, the average expression value for each gene in the two specimens was further calculated. The expression level of every gene was then compared between the normal and cancerous tissues. One of the genes whose expression was most specific for the normal pancreatic tissue was that for insulin; its expression level in normal tissue was 6.869 arbitrary units (U) whereas the averaged value in the cancerous tissues was 1.22 U. Given that insulin is expressed only in islets of Langerhans, this result likely reflects the difference in the proportion of endocrine cells between the samples, not a difference in the number of *INSULIN* gene transcripts per cell between normal and cancer cells.

We next prepared MUC1$^+$ ductal cells from two individuals who were diagnosed as negative for PDC. Microarray analysis of these cells and comparison of the resulting data with those obtained with normal pancreatic tissue also identified the *INSULIN* gene as one of the most differentially expressed genes between the two types of sample; the averaged *INSULIN* expression level in the ductal specimens was 0.495 U, while that in the normal tissue section was 6.869 U.

Given that the proportion of cells of ductal origin would be expected to be markedly increased in cancerous pancreatic tissue compared with that in normal pancreatic tissue, these data support our expectation that a simple comparison of surgically resected specimens of normal and cancerous tissues from the pancreas is not a suitable approach to identify transformation-related genes of the ductal cell lineage.

**Gene expression profiles of ductal cells obtained from pancreatic juice.** An ideal strategy to identify potential molecular markers specific to PDC would be to compare the transcriptomes of ductal cells isolated from the pancreatic juice of healthy individuals and cancer patients. Any difference identified between the transcriptomes by such screening would thus likely reflect the transformation process, given that both of the samples would be of the same cellular origin. Furthermore, from the

**Fig. 1.** Purification of pancreatic duct cells from pancreatic juice. (A) Cells isolated by centrifugation from the pancreatic juice of an individual with PDC were subjected to Wright-Giemsa staining (magnification, 100×). In addition to cells of epithelial origin, both red blood cells and neutrophils (arrowheads) are apparent. (B) Cells separated from the pancreatic juice of the same individual with PDC were subjected to chromatography on a MUC1-based affinity column. Cells specifically eluted from the column were then subjected to Wright-Giemsa staining (magnification, 200×). Some of the eluted cells exhibited a cancer-specific aberrant phenotype (large nuclei with fine chromatin structure).



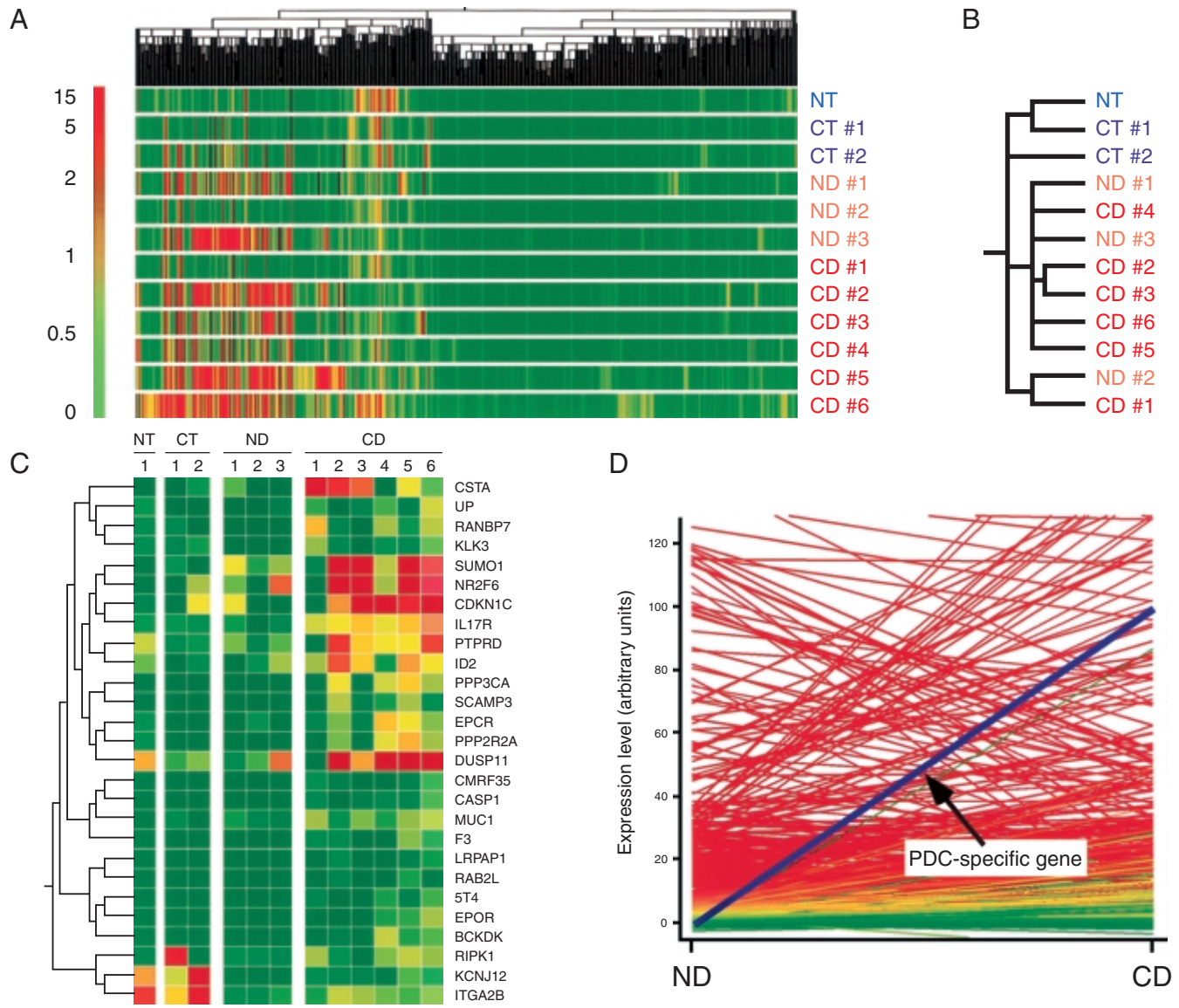




**Fig. 2.** (A) Hierarchical clustering of 3456 genes based on their expression profiles in pancreatic tissue specimens from one normal individual (NT) and two PDC patients (CT #1 and #2) as well as in MUC1$^+$ ductal cells obtained from three normal individuals (ND #1–3) and six cancer patients (CD #1–6). Each column represents a single gene on the microarray, and each row corresponds to a different subject. The normalized fluorescence intensity for each gene is shown color-coded as indicated at the left. (B) Two-way clustering analysis of the transcriptomes shown in (A) was performed to assess statistically the similarity among the samples from the different subjects and to generate a subject dendrogram. (C) Hierarchical clustering of the "disease-dependent" genes. Expression intensities are shown color-coded according to the scale in (A). Gene symbols are indicated at the right. (D) Comparison of the expression levels of 3456 human genes between normal and cancerous ductal cells. The normalized value for the expression level of each gene was averaged for three normal ductal cell specimens and was compared with the corresponding value obtained with six cancerous ductal cell samples. Each line corresponds to a single gene on the array and is presented color-coded according to the expression level in the normal tissue according to the scale shown in (A). The line for a hypothetical "PDC-specific gene" is indicated in blue.

point of view of clinical application, this BAMP screening approach also appears desirable. The identification of *bona fide* cancer-specific genes would thus allow development of a sensitive method for the diagnosis of PDC based on reverse transcription and PCR (RT-PCR) analysis of cells isolated from pancreatic juice, which can be obtained during the ERCP procedure.

In an attempt to realize this goal, we compared the expression profiles of 3456 genes among one specimen of normal pancreatic tissue (NT), two specimens of cancerous pancreatic tissue (CT #1 and #2), three normal ductal cell preparations (ND #1 to #3), and six ductal cell preparations obtained from PDC patients (CD #1 to #6). The clinical information is summarized in Table 1 for the PDC patients who provided pancreatic juice. All of the ductal cell preparations of the CD patients were cytologically diagnosed to contain "class IV" cells, the proportion of which is also shown in the table. Since all CD patients already had tumor invasion into either the splenic artery or the portal vein as judged by angiography, none of them was suitable for surgical operation. Therefore, we do not have any pathological data of pancreatic tissues for any of the PDC patients in Table 1. All CD patients died within 12 months after diagnostic procedures.

**Table 1. Clinical characteristics of the patients with PDC**

| Patient ID | Sex | Age (yr) | Liver metastasis | SA or PV invasion | Proportion of class IV cells (%) |
|---|---|---|---|---|---|
| CD #1 | M | 71 | – | + | 6.4 |
| CD #2 | F | 61 | – | + | 45.3 |
| CD #3 | F | 82 | – | + | 4.6 |
| CD #4 | F | 68 | + | + | 4.2 |
| CD #5 | F | 73 | + | + | 12.6 |
| CD #6 | F | 71 | – | + | 33.4 |

M, male; F, female; yr, year; SA, splenic artery; PV, portal vein.

The ND #1–3 individuals were subjected to ERCP procedure due to a slight elevation in blood amylase level or to the echographic finding of dilation of the pancreatic duct. However, ERCP examination could detect no anomaly in their ductal structure. These individuals were also negative for PDC in cytological analysis of pancreatic juice, and are still healthy after >12 months of observation.

The gene expression profiles of each sample were subjected to clustering analysis in order to generate a dendrogram, or "gene tree," in which genes with similar expression profiles are clustered together (Fig. 2A). Such analysis revealed that the patterns of gene expression of ND #1 and #3 were similar to those of CD #2 to #6. However, despite this overall similarity, significant differences between these two types of sample were apparent, some of which might reflect the carcinogenic process.

To statistically analyze the similarity of transcriptomes among the samples, we performed two-way clustering analysis[17] to generate a "subject tree," in which samples with similar transcriptomes are grouped together (Fig. 2B). All ductal cell samples (ND and CD) were clustered in two major branches, separated from the tissue samples, which indicates that the transcriptomes of the cancerous ductal cells were more similar to those of the normal ductal cells than they were to those of the cancer tissue specimens. The transcriptomes of ductal cell samples from cancer patients #2 and #3 exhibited the greatest similarity.

**Potential molecular markers for PDC.** Our data suggest that a direct comparison between normal and cancerous ductal cells would be a suitable means to efficiently identify the PDC-specific

**Table 2. Expression level of the disease-dependent genes**

| Gene symbol | GenBank # | NT | CT #1 | CT #2 | ND #1 | ND #2 | ND #3 | CD #1 | CD #2 | CD #3 | CD #4 | CD #5 | CD #6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *DUSP11* | AF023917 | 3.833 | 0.924 | 1.387 | 0.358 | 1.056 | 5.102 | −0.056 | 22.841 | 3.826 | 22.211 | 30.331 | 15.227 |
| *KCNJ12* | L36069 | 4.157 | 2.096 | 10.149 | −0.166 | −0.422 | −0.284 | 0.356 | 0.169 | 0.087 | −0.164 | 1.106 | 0.168 |
| *ITGA2B* | J02764 | 7.498 | 3.165 | 9.311 | 0.322 | 0.261 | 0.773 | 0.730 | 2.072 | 1.668 | 1.284 | 1.000 | 1.340 |
| *CSTA* | X05978 | −0.962 | −0.905 | 0.638 | 1.100 | −1.487 | −0.366 | 12.865 | 7.687 | 5.928 | −0.164 | 2.528 | 0.809 |
| *UP* | X90858 | 0.637 | −0.182 | −0.378 | −0.447 | −0.092 | 0.062 | 0.977 | 0.021 | 0.209 | 1.000 | 0.056 | 2.205 |
| *SUMO1* | U61397 | −1.043 | 0.557 | −0.933 | 2.703 | 0.841 | 1.618 | 0.196 | 10.385 | 11.219 | 1.952 | 16.324 | 8.469 |
| *PPP3CA* | M29550 | −0.511 | −0.926 | −0.938 | −0.747 | −1.191 | −0.214 | −1.587 | 2.656 | −0.427 | 2.267 | 3.231 | 1.701 |
| *PTPRD* | L38929 | 2.149 | 0.333 | 0.694 | 1.310 | 0.149 | 1.140 | −0.104 | 7.369 | 3.223 | 2.997 | 2.793 | 6.908 |
| *LRPAP1* | M63959 | −0.387 | −0.790 | −0.541 | −0.805 | −0.595 | −0.265 | −0.339 | 0.399 | −0.352 | −0.291 | 0.062 | 0.759 |
| *RANBP7* | AF098799 | 0.351 | 0.256 | −0.741 | −0.389 | −0.303 | 0.018 | 3.466 | 0.142 | −0.395 | 1.834 | 0.673 | 1.968 |
| *EPCR* | L35545 | 0.440 | −0.558 | −0.393 | −0.483 | 0.581 | −0.050 | 0.194 | 1.277 | 0.125 | 3.373 | 2.891 | 1.384 |
| *RAB2L* | U68142 | −0.638 | −1.144 | −0.829 | −0.856 | −1.214 | −1.341 | −1.507 | 0.226 | −0.848 | −0.014 | 0.803 | −0.100 |
| *5T4* | Z29083 | −0.412 | −0.913 | −0.338 | −0.710 | −0.724 | −0.104 | 0.221 | −0.136 | −0.492 | 0.897 | 0.102 | 1.044 |
| *PPP2R2A* | M64929 | 0.501 | −0.858 | −0.442 | −0.501 | −0.376 | −0.194 | −0.448 | 1.272 | −0.380 | 2.474 | 3.584 | 1.677 |
| *EPOR* | M34986 | −0.008 | −1.112 | −0.744 | −0.692 | −0.671 | −0.143 | −0.093 | −0.246 | −0.438 | 0.857 | 0.898 | 1.727 |
| *BCKDK* | AF026548 | −0.202 | −1.266 | −0.713 | −1.002 | −0.900 | −0.260 | −0.876 | 0.129 | −0.256 | 2.188 | 0.852 | 1.089 |
| *ID2* | M97796 | 1.247 | −0.076 | 0.389 | −0.531 | 0.572 | 1.695 | 1.721 | 6.419 | 3.258 | 0.659 | 4.139 | 2.867 |
| *NR2F6* | X12794 | −1.279 | −0.502 | 1.768 | 1.270 | −0.713 | 5.963 | −0.921 | 20.658 | 30.415 | 1.672 | 28.746 | 23.043 |
| *RIPK1* | U50062 | 0.095 | 8.471 | −0.001 | 0.422 | 0.367 | 0.190 | 1.819 | 0.450 | 0.032 | 1.404 | 2.050 | 1.607 |
| *KLK3* | M26663 | 0.384 | 0.192 | 0.780 | −0.038 | −0.095 | −0.066 | 1.653 | 0.276 | 0.041 | 0.556 | 0.277 | 0.937 |
| *CMRF35* | X66171 | −1.406 | −0.541 | −0.824 | −0.171 | −0.327 | −0.129 | 0.521 | 0.076 | −0.024 | −0.190 | 0.076 | 0.579 |
| *SCAMP3* | AF005039 | −0.362 | −1.002 | −0.628 | −0.493 | −1.060 | −0.270 | −0.734 | 1.936 | 0.099 | 0.781 | 1.827 | 0.410 |
| *CASP1* | U13698 | −0.808 | −0.132 | −1.027 | 0.025 | −0.113 | −0.070 | 0.300 | −0.009 | 0.123 | 0.279 | 0.187 | 0.726 |
| *F3* | J02931 | −1.096 | −0.562 | −0.286 | −0.261 | −0.607 | −0.325 | 0.364 | 0.349 | 0.038 | −0.276 | 1.530 | 0.388 |
| *MUC1* | J05581 | −2.454 | 0.398 | −0.759 | 0.771 | 0.330 | 0.070 | 1.775 | 0.534 | 1.387 | 0.863 | 1.642 | 1.925 |
| *CDKN1C* | U22398 | 0.057 | −0.434 | 3.034 | 2.572 | −1.003 | −0.375 | −0.138 | 4.131 | 10.992 | 9.788 | 8.131 | 11.543 |
| *IL17R* | U58917 | 0.707 | 0.647 | 0.432 | 0.765 | −0.138 | 0.263 | 2.322 | 2.692 | 3.360 | 2.452 | 3.319 | 5.122 |

Expression level of the "disease-dependent" genes is shown in arbitrary units (U). Gene symbol as well as GenBank accession number (#) is indicated for each gene.

transcriptome changes while keeping pseudo-positive data minimum. To identify *bona fide* PDC-specific genes from the array data, we here took two approaches.

First, expression levels of 3456 genes were compared between ND and CD sample types by Welch ANOVA test. Twenty-seven genes were thus identified, whose expression levels were statistically significantly different in the two types ($P<0.05$). A dendrogram of such disease-dependent genes is shown in Fig. 2C. Many genes in the list, including those for *SUMO1* (GenBank accession no. U61397) and dual specificity phosphatase (*DUSP*) *11* (GenBank accession no. AF023917), were inducibly expressed in PDC cells. Like ubiquitin, SUMO1 functions as a protein "tag," transfer of which is mediated by a SUMO $E_3$ ligase. In contrast to ubiquitin, however, modification with SUMO1 not only drives the substrates into a proteasome pathway, but has a pleiotropic effect on the substrates, such as protection against proteolysis, induction of apoptosis, and regulation of substrate function.[18, 19] The *in vivo* role of SUMO1 is thus likely to be context-dependent, and it is an interesting question whether increased SUMO-tagging has a transforming or anti-apoptotic activity in PDC cells. The array data for these "disease-dependent" genes are shown in detail in Table 2. These genes would be good candidates to be included in custom-made DNA microarrays specialized for the diagnosis of PDC.

However, there is a caveat that this type of comparison may isolate genes whose absolute expression levels may be negligibly low. Actually, fifteen out of twenty-seven genes in Table 2 did not have expression levels of more than 3.0 U in any ductal cell preparation.

Therefore, we also tried another approach to select PDC-specific genes. The mean expression value of each gene was calculated for the ND or CD sample type, and the differences in the resulting values are represented in Fig. 2D. To identify genes whose mean expression values were induced only in the cancerous ductal cells, with the use of GeneSpring software, we searched for genes whose expression profiles were statistically similar, with a minimum correlation of 0.99, to that of a hypothetical "PDC-specific gene" (blue line in Fig. 2D) that exhibits a mean expression level of 0.0 U in the ND group and 100.0 U in the CD group. Taking the 188 genes thus identified, we then applied the criteria that the gene expression value should be (1) $<3.0$ U in all NT/ND samples and (2) $\geq 19.0$ U in at least one of the CD samples. Thirty-one genes were finally identified to be "PDC-specific" (Table 3). Through this approach, we tried to extract genes whose expression levels were negligible in all normal pancreatic specimens, but significantly high in at least a part of the cancerous ones. They may be good candidates for molecular markers to develop PCR-based diagnostic tests for PDC.

These potential PDC-specific markers include the genes for FYN protein tyrosine kinase (*FYN*; GenBank accession no. M14676/M14333); insulin-like growth factor binding protein 1 (*IGFBP1*; Y00856); collagen, type I, alpha 1 (*COL1A1*; Z74615); calpain, large polypeptide L2 (*CAPN2*; M23254); eukaryotic translation elongation factor 1 beta 2 (*EEF1B2*; X60489); *AC133* (AF027208) and *CEACAM7* (X98311).

**Table 3. Expression level of the PDC-specific genes**

| Gene symbol | GenBank # | NT | CT #1 | CT #2 | ND #1 | ND #2 | ND #3 | CD #1 | CD #2 | CD #3 | CD #4 | CD #5 | CD #6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *FYN* | M14676 | −0.317 | −0.982 | −0.671 | 2.198 | −1.203 | −0.470 | 3.000 | 1.327 | −0.029 | 1.435 | 27.266 | 13.246 |
| *FYN* | M14333 | −0.642 | −0.874 | −1.010 | 1.131 | −1.229 | 0.721 | 2.771 | 1.340 | −0.032 | 1.796 | 28.936 | 12.357 |
| *RGR* | U14910 | 0.174 | −1.073 | −0.249 | −0.795 | −0.439 | 2.719 | 6.824 | 1.531 | 2.929 | 2.021 | 1.863 | 27.514 |
| *IGFBP1* | Y00856 | −0.504 | −0.880 | −0.671 | −0.486 | −1.098 | −0.461 | −1.244 | 77.812 | −0.820 | −0.772 | 52.414 | 8.442 |
| *DUSP1* | X68277 | 0.062 | 1.488 | 0.701 | 1.412 | −0.682 | 1.444 | 0.782 | 15.259 | 0.723 | 2.374 | 21.301 | 2.082 |
| *IL1RN* | X52015 | −0.288 | −0.879 | −0.489 | 2.410 | −0.874 | 3.110 | −1.102 | 75.070 | 1.968 | 4.507 | 10.914 | 5.436 |
| *HSJ2* | L08069 | −0.882 | −1.352 | −0.752 | 0.909 | −1.269 | 2.420 | −0.992 | −0.165 | −0.712 | 0.438 | 27.059 | 2.441 |
| *APCS* | X04608 | −0.740 | −0.471 | −0.212 | 1.062 | −1.140 | 0.038 | −1.127 | 22.176 | 6.088 | 19.942 | 0.012 | −0.482 |
| *GTF2A1* | U21242 | −0.181 | −0.754 | −0.090 | 2.395 | −1.085 | 2.376 | −0.930 | 12.423 | 1.016 | −0.140 | 28.966 | 7.717 |
| *GTF2F2* | X16901 | −0.697 | −1.348 | −0.356 | −0.392 | −0.231 | 0.523 | 2.121 | 3.955 | 0.329 | 0.655 | 20.209 | 3.319 |
| *IRF4* | U52682 | 0.269 | −1.213 | −0.509 | −0.835 | 0.141 | −0.073 | −0.561 | 0.202 | −0.103 | −0.236 | 25.817 | 0.570 |
| *POU2AF1* | Z49194 | −0.698 | −1.264 | −0.461 | −1.069 | −0.643 | 0.090 | −0.642 | 2.758 | −0.109 | −0.623 | 47.368 | 1.189 |
| *SNRPG* | X85373 | −0.374 | −1.027 | −0.827 | 2.082 | −1.095 | 3.342 | −0.053 | 11.652 | −0.214 | −0.355 | 33.614 | 7.384 |
| *SLC16A3* | U81800 | 0.463 | −0.588 | 0.296 | 0.712 | −0.913 | 0.332 | −0.841 | 4.999 | 0.359 | −0.222 | 21.756 | 0.495 |
| *H1F5* | X83509 | −0.092 | −0.239 | 0.118 | 1.197 | −1.118 | 0.418 | −0.886 | 1.573 | 0.166 | −0.593 | 23.560 | 0.481 |
| *GTF2B* | M76766 | −0.893 | −0.670 | −0.816 | 0.824 | −1.126 | 0.467 | −1.093 | 32.156 | 0.339 | 4.622 | 34.587 | 23.964 |
| *SNRPC* | M18465 | −0.149 | −1.184 | −0.542 | −0.835 | −0.403 | 0.547 | −0.610 | 11.491 | 0.819 | 1.282 | 22.521 | 1.158 |
| *ECM1* | U68186 | −1.969 | −0.882 | −0.971 | 0.048 | −0.921 | −0.454 | −1.218 | 15.501 | 0.389 | −0.425 | 50.772 | 0.072 |
| *KLK6* | AF013988 | −4.069 | −1.028 | −3.372 | −0.121 | −1.441 | −0.372 | −1.324 | 26.647 | 0.122 | −0.715 | 60.203 | 3.603 |
| *COL1A1* | Z74615 | −2.193 | 0.018 | 98.459 | 1.133 | −1.197 | −0.466 | −1.134 | 10.098 | 13.086 | 3.584 | 131.260 | 10.451 |
| *CAPN2* | M23254 | −0.996 | −1.063 | 0.483 | 1.178 | −1.387 | −0.030 | −1.320 | 12.394 | 6.932 | −0.419 | 20.623 | 0.570 |
| *RGS5* | AB008109 | 0.026 | −0.950 | −0.386 | −0.837 | −1.315 | −0.458 | −1.093 | 0.814 | 0.140 | −0.675 | 0.000 | 52.133 |
| *EEF1B2* | X60489 | −0.287 | −0.713 | 0.037 | 1.154 | −1.509 | 2.050 | −1.269 | 9.485 | 20.314 | −0.133 | 30.121 | 0.971 |
| *F7* | M13232 | −1.686 | −1.055 | −0.909 | −0.500 | −1.476 | −0.512 | −1.363 | −0.254 | −0.331 | −0.770 | −0.353 | 22.485 |
| *CEACAM7* | X98311 | −0.065 | −0.802 | −0.728 | −0.247 | −1.036 | −0.085 | −0.900 | 10.468 | 22.096 | 0.021 | −0.244 | −0.011 |
| *CAMLG* | U18242 | −0.703 | −0.916 | 0.285 | 0.582 | −1.435 | 0.092 | −1.238 | 1.829 | 1.801 | −0.431 | 22.154 | 0.461 |
| *APOA4* | X13629 | −3.473 | −1.048 | −0.974 | 0.105 | −0.835 | −0.514 | −1.278 | −0.240 | −0.357 | −0.704 | −0.324 | 37.780 |
| *GAPDH* | M33197 | 0.142 | −0.765 | 1.590 | 2.756 | −0.920 | 2.505 | −0.854 | 3.372 | 3.908 | 2.241 | 4.070 | 21.745 |
| *MYBPC3* | X84075 | −1.825 | −1.043 | −1.217 | −0.475 | −1.472 | −0.496 | −1.283 | −0.102 | −0.307 | −0.772 | −0.355 | 521.712 |
| *AC133* | AF027208 | −0.741 | −0.970 | −0.666 | 0.071 | −1.527 | −0.264 | −1.398 | 19.820 | 8.152 | −0.460 | 0.546 | −0.364 |
| *APOBEC1* | L25877 | −0.740 | 1.771 | −1.215 | 2.591 | 1.060 | 0.982 | 0.979 | 9.583 | 2.584 | 2.151 | 12.211 | 24.971 |
| *EIF3S6* | U62962 | −0.703 | 0.244 | −0.123 | 2.402 | 0.053 | 2.350 | −0.626 | 11.403 | 28.308 | 0.607 | 11.966 | 6.648 |

Expression intensities of the "PDC-specific" genes are shown in arbitrary units (U). Gene symbol as well as GenBank accession number (#) is indicated for each gene. Two distinct oligonucleotides were spotted on the array for the *FYN* gene.

**Quantitation of mRNA for potential PDC marker genes.** Finally, we confirmed the expression of three of the potential PDC marker genes by real-time PCR. Unamplified cDNA was prepared from MUC1$^+$ ductal cells obtained from 8 normal individuals and 10 patients with PDC and was subjected to PCR with primers specific for *β-actin*, *SUMO1*, *AC133*, or *CEACAM7* genes. The amount of each PCR product was monitored in real time, thereby allowing determination of the corresponding $C_T$ values. The abundance of SUMO1, AC133 and CEACAM7 mRNAs was then calculated relative to that of β-actin mRNA.

Consistent with the microarray data, expression of *SUMO1*, *AC133* and *CEACAM7* genes was highly specific to PDC; in particular, the latter two genes were almost silent in normal ductal cells (Fig. 3). These genes are thus candidates for PDC-specific markers. The expression levels of the *SUMO1*, *AC133* and *CEACAM7* genes varied among the cancer specimens, as might be expected from nonuniformity of the transformation process in pancreatic ductal cells.

## Discussion

We have demonstrated that a simple comparison of transcriptomes between normal and cancerous tissue of the pancreas is not a suitable approach for characterization of the transformation process. In contrast, through screening with isolated ductal cells derived from normal and carcinoma tissue, we were able to identify a group of genes that may prove helpful in the diagnosis of PDC.

In addition to the purification of PDC cells from pancreatic juice, there is another way to isolate PDC cells, i.e., the laser capture microdissection (LCM) method.[20] Although, with LCM, it is theoretically possible to purify any cell type in a given tissue, fixation and staining procedures of the specimens prior to LCM may severely impair the quality of mRNA in the samples. Furthermore, it would be a demanding task to pick up $10^5$–$10^6$ cells by LCM. Small number of cells obtained by LCM often requires multiple rounds of mRNA amplification before microarray experiments, making the data evaluation more difficult. Therefore, purification of intact and live PDC cells through pancreatic juice would be advantageous for obtaining high-quality mRNA and good reproducibility in transcriptome analysis.

Moreover, as with our CD cases (see Table 1), it is rare to find patients with PDC at early stages competent for surgical resection. Therefore, it may be difficult to complete a large-scale clinical screening of PDC tissue sections. In contrast, screening of hundreds of "pancreatic juice" samples is a realistic project.

For the improvement of PDC treatment, it is essential to detect PDC at the stage of curable carcinoma *in situ*. We assume that the direct analysis of PDC cell-containing specimens would be the most sensitive way to detect PDC, and, in a routine clinical setting, pancreatic juice is the only source to obtain PDC cells. These are the reasons why we attempted to develop a novel PDC diagnosis procedure based on pancreatic juice.

As expected, pancreatic juice contained various amounts of non-ductal cells (mainly blood cells). Therefore, we had first to enrich pancreatic ductal cells from the juice by means of an affinity column directed toward MUC1. It was interesting to find *MUC1* in the "disease-dependent" gene list (Fig. 2C and Table 2). In our analysis, *MUC1* expression was induced in cancerous ductal cells (1.35 U±0.547; mean value±SD) compared to normal ductal cells (0.390 U±0.354). An increase in mRNA[21] or protein[22] level of MUC1 in PDC cells has been also reported. Low yet significant expression of MUC1 in our ductal cell specimens also argues that the MUC1-column eluents did contain pancreatic ductal cells, since MUC1 is expressed only by epithelial cells, not by blood cells.

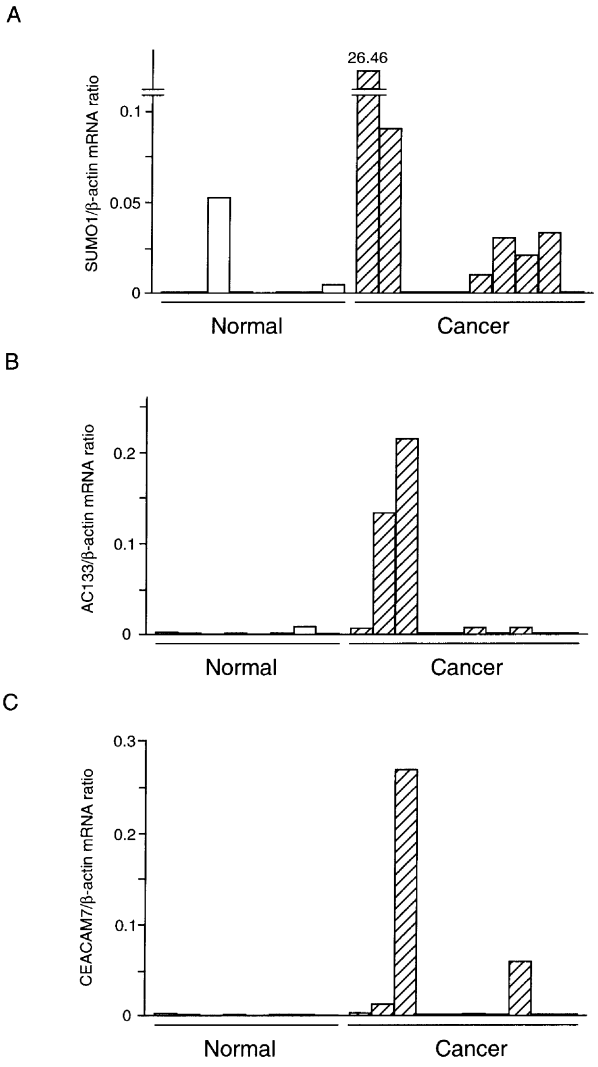



**Fig. 3.** Quantitation of *SUMO1*, *AC133* and *CEACAM7* gene transcripts in MUC1$^+$ ductal cells. Complementary DNA prepared from pancreatic ductal cells of 8 normal individuals and 10 PDC patients was subjected to real-time PCR with primers specific for *SUMO1* (A), *AC133* (B), *CEACAM7* (C), or *β-actin* genes. The ratio of the abundance of the target transcripts to that of β-actin mRNA was calculated as $2^n$, where $n$ is the $C_T$ value for β-actin cDNA minus the $C_T$ value of the target cDNA.

Our MUC1-based purification system does not discriminate normal ductal cells from malignant ones. Therefore, ductal cells isolated from PDC patients (such as CD #1–6) should be a mixture of normal ductal cells and PDC ones. Since there are no cell membrane proteins known to be specifically expressed in PDC, it is currently impossible to directly purify PDC cells from pancreatic juice. Rather, we here aimed to develop a sensitive method to detect a trace amounts of PDC cells shed into pancreatic juice.

For this purpose, there may be two distinct types of molecular markers. One type is useful in statistically distinguishing normal and cancerous ductal cell types. Such analyses choose genes whose expression level has a small deviation, and, therefore, may be suitable to construct custom-made DNA microarrays. Genes of the other type would be active only in cancerous ductal cells, but strictly absent in normal ones. These genes would be good candidates for the target transcripts used in RT-PCR-based detection systems. Expression levels of such genes in cancerous cells may have a relatively large SD, and such genes may not be expressed in all cancerous cells. However, if

their transcription is completely silent in normal cells, an RT-PCR-based detection kit would be of practical value. The genes in Tables 2 and 3 are our first results from the approaches above, and expression profiles of some of them were confirmed by real-time PCR (Fig. 3).

Among the genes listed in Tables 2 and 3, several were already known to be highly expressed in carcinoma cells. For example, ID2 drives cell cycle progression by interacting with, and suppressing the activity of, a tumor suppressor, Rb.[23] ID2 can also suppress another growth suppressor, p16.

CEACAM7 belongs to the CEA family of proteins. In contrast to the high level of expression of CEA apparent in colorectal carcinomas, CEACAM7 is abundant in normal colonic epithelium and its expression is down-regulated during malignant transformation.[24, 25] Although its expression in pancreas has not been well characterized, previous data indicate that CEACAM7 is expressed in normal pancreatic ductal cells.[24] However, our observation that the *CEACAM7* gene is preferentially expressed in ductal cells of PDC patients suggests that this gene is a potential marker for cancer diagnosis with either ductal cell- or serum-based assays.

AC133 was initially identified as a cell surface marker specific to a hematopoietic stem cell-enriched fraction with a $CD34^{high}$, $CD38^{low}$ or $CD38^{-}$, and c-$Kit^{+}$ phenotype.[26] This protein is also expressed on the precursor of endothelial cells,[27] indicating that it may be a marker for immature hemangioblasts, which are common precursors for blood cells and blood vessels. Although expression of AC133 in tissues other than bone marrow and the retina has not been previously demonstrated, we have now shown that the *AC133* gene is expressed in the pancreatic ductal cells of PDC patients. Given the abundance of AC133 in normal hemangioblasts, the expression of the *AC133* gene in carcinoma ductal cells may suggest that AC133 is also a marker of the precursor for ductal cells. The increased expression of the *AC133* gene in PDC may thus reflect the immature nature of the cancer cells with regard to the differentiation program of ductal cells.

It should be noted, however, that none of the single genes listed in Tables 2 and 3 was able to distinguish all PDC samples from normal ductal cells. In addition to such single gene-based prediction systems, it may be possible to use the expression profiles of a combination of "class predictor" genes[28] for PDC diagnosis. We have indeed examined the feasibility of this approach with the statistically "disease-dependent" genes listed in Table 2. Prediction of PDC diagnosis was tried with the k-nearest neighbor algorithm by using the GeneSpring software (http://www.silicongenetics.com/Support/GeneSpring/GSnotes/class_prediction.pdf). In a "cross-validation" trial, all three ND samples were correctly diagnosed by the expression profiles of the disease-dependent genes (data not shown). With regard to the CD samples, four out of six samples were correctly predicted, and the other two were called "unpredictable." Therefore, among the nine ductal cell specimens, seven (77.8%) were correctly diagnosed. Selection of stronger "predictor" genes through large-scale microarray studies may make it possible to construct reliable "PDC diagnosis arrays" harboring a small number of such predictor genes.

In conclusion, we have shown that DNA microarray analysis with purified ductal cell fractions is a promising approach to the identification of PDC-specific genes, being greatly superior to a mere comparison of tissue specimens. Our data thus provide a basis for the possible development of an ERCP-dependent sensitive and specific test for the detection of pancreatic cancer.

This work was supported in part by a Grant-in-Aid for Scientific Research (C) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

1. Bornman PC, Beckingham IJ. Pancreatic tumours. *Br Med J* 2001; **322**: 721–3.
2. Rosewicz S, Wiedenmann B. Pancreatic carcinoma. *Lancet* 1997; **349**: 485–9.
3. Adamek HE, Albert J, Breer H, Weitz M, Schilling D, Riemann JF. Pancreatic cancer detection with magnetic resonance cholangiopancreatography and endoscopic retrograde cholangiopancreatography: a prospective controlled study. *Lancet* 2000; **356**: 190–3.
4. Kondo H, Sugano K, Fukayama N, Kyogoku A, Nose H, Shimada K, Ohkura H, Ohtsu A, Yoshida S, Shimosato Y. Detection of point mutations in the K-ras oncogene at codon 12 in pure pancreatic juice for diagnosis of pancreatic carcinoma. *Cancer* 1994; **73**: 1589–94.
5. Sugano K, Nakashima Y, Yamaguchi K, Fukayama N, Maekawa M, Ohkura H, Kakizoe T, Sekiya T. Sensitive detection of loss of heterozygosity in the TP53 gene in pancreatic adenocarcinoma by fluorescence-based single-strand conformation polymorphism analysis using blunt-end DNA fragments. *Genes Chromosom Cancer* 1996; **15**: 157–64.
6. Caldas C, Hahn SA, da Costa LT, Redston MS, Schutte M, Seymour AB, Weinstein CL, Hruban RH, Yeo CJ, Kern SE. Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. *Nat Genet* 1994; **8**: 27–32.
7. Hahn SA, Schutte M, Hoque AT, Moskaluk CA, da Costa LT, Rozenblum E, Weinstein CL, Fischer A, Yeo CJ, Hruban RH, Kern SE. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* 1996; **271**: 350–3.
8. Hohne MW, Halatsch ME, Kahl GF, Weinel RJ. Frequent loss of expression of the potential tumor suppressor gene DCC in ductal pancreatic adenocarcinoma. *Cancer Res* 1992; **52**: 2616–9.
9. Furuya N, Kawa S, Akamatsu T, Furihata K. Long-term follow-up of patients with chronic pancreatitis and K-ras gene mutation detected in pancreatic juice. *Gastroenterology* 1997; **113**: 593–8.
10. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nat Genet* 1999; **21**: 10–4.
11. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; **270**: 467–70.
12. Miyazato A, Ueno S, Ohmine K, Ueda M, Yoshida K, Yamashita Y, Kaneko T, Mori M, Kirito K, Toshima M, Nakamura Y, Saito K, Kano Y, Furusawa S, Ozawa K, Mano H. Identification of myelodysplastic syndrome-specific genes by DNA microarray analysis with purified hematopoietic stem cell fraction. *Blood* 2001; **98**: 422–7.
13. Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA* 1990; **87**: 1663–7.
14. Balague C, Audie JP, Porchet N, Real FX. *In situ* hybridization shows distinct patterns of mucin gene expression in normal, benign, and malignant pancreas tissues. *Gastroenterology* 1995; **109**: 953–64.
15. Terada T, Ohta T, Sasaki M, Nakamura Y, Kim YS. Expression of MUC apomucins in normal pancreas and pancreatic tumours. *J Pathol* 1996; **180**: 160–5.
16. Lacobuzio-Donahue CA, Maitra A, Shen-Ong GL, van Heek T, Ashfaq R, Meyer R, Walter K, Berg K, Hooingsworth MA, Cameron JL, Yeo CJ, Kern SE, Goggins M, Hruban RH. Discovery of novel tumor markers of pancreatic cancer using global gene expression technology. *Am J Pathol* 2002; **160**: 1239–49.
17. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, Levine AJ. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci USA* 1999; **96**: 6745–50.
18. Hay RT. Protein modification by SUMO. *Trends Biochem Sci* 2001; **26**: 332–3.
19. Johnson ES, Gupta AA. An E3-like factor that promotes SUMO conjugation to the yeast septins. *Cell* 2001; **106**: 735–44.
20. Luo L, Salunga RC, Guo H, Bittner A, Joy KC, Galindo JE, Xiao H, Rogers KE, Wan JS, Jackson MR, Erlander MG. Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nat Med* 1999; **5**: 117–22.
21. Ho SB, Niehans GA, Lyftogt C, Yan PS, Cherwitz DL, Gum ET, Dahiya R, Kim YS. Heterogeneity of mucin gene expression in normal and neoplastic tissues. *Cancer Res* 1993; **53**: 641–51.
22. Osako M, Yonezawa S, Siddiki B, Huang J, Ho JJL, Kim YS, Sato E. Immunohistochemical study of mucin carbohydrates and core proteins in human pancreatic tumors. *Cancer* 1993; **71**: 2191–9.
23. Lavarone A, Garg P, Lasorella A, Hsu J, Israel MA. The helix-loop-helix protein Id-2 enhances cell proliferation and binds to the retinoblastoma protein. *Genes Dev* 1994; **8**: 1270–84.
24. Scholzel S, Zimmermann W, Schwarzkopf G, Grunert F, Rogaczewski B, Thompson J. Carcinoembryonic antigen family members CEACAM6 and

CEACAM7 are differentially expressed in normal tissues and oppositely deregulated in hyperplastic colorectal polyps and early adenomas. *Am J Pathol* 2000; **156**: 595–605.
25. Thompson J, Seitz M, Chastre E, Ditter M, Aldrian C, Gespach C, Zimmermann W. Down-regulation of carcinoembryonic antigen family member 2 expression is an early event in colorectal tumorigenesis. *Cancer Res* 1997; **57**: 1776–84.
26. Hin AH, Miraglia S, Zanjani ED, Almeida-Porada G, Ogawa M, Leary AG, Olweus J, Kearney J, Buck DW. AC133, a novel marker for human hematopoietic stem and progenitor cells. *Blood* 1997; **90**: 5002–12.
27. Gallacher L, Murdoch B, Wu DM, Karanu FN, Keeney M, Bhatia M. Isolation and characterization of human CD34(−)Lin(−) and CD34(+)Lin(−) hematopoietic stem cells using cell surface markers AC133 and CD7. *Blood* 2000; **95**: 2813–20.
28. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; **286**: 531–7.