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

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

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. G., 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 carcinoma 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 was used to compare the gene expression in intraepithelial neoplasia (15), some of which progress to pancreatic cancer tissue. A total of 294,920 tags in 10 pancreatic samples.

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 cell lines. 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 glmol Tris–HCl (pH 7.5), 500 glmol LiCl, 10 glmol EDTA, 1% LiDS, 5 glmol DTT] (Dynabead mRNA direct kit; Dynal, Oslo, Norway) by addition of 66 glmol of BSA. Total RNA (5 glmol) 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 as part of the SAGEmap database. 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.

1 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.

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.

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

3 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@jhmu.edu.

4 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.
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. 1A). A more informative means of displaying the key correlations is by scatter plot analysis. 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. 1B). 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. 1B). 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. 1B). 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. 1B). 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. 1B). 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. 1B). 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. 1B). 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. 1B). 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. 1B).
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).

![Fig. 1. Relationships of gene expression profiles.](image)

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.

### Table 2. Pair-wise comparison of SAGE profiles

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

* 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.

Downloaded from cancerres.aacrjournals.org on April 20, 2017. © 2002 American Association for Cancer Research.
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: Differential gene expression in pancreatic cancer as compared with normal ductal epithelium

<table>
<thead>
<tr>
<th>Tag</th>
<th>X²</th>
<th>Gene/ESTs</th>
<th>N</th>
<th>D1</th>
<th>D2</th>
<th>CL1</th>
<th>CL2</th>
<th>CL3</th>
<th>CL4</th>
<th>CL5</th>
<th>CL6</th>
<th>H1</th>
<th>I2</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCATGAGGT</td>
<td>Keratin 19</td>
<td>64</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>83</td>
<td>123</td>
<td>502</td>
<td>177</td>
<td>64</td>
<td>98</td>
<td>Cytoskeletal and microfibrillar</td>
</tr>
<tr>
<td>ATCCCTCAAACG</td>
<td>S100A4 (Mst1)</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>112</td>
<td>77</td>
<td>0</td>
<td>291</td>
<td>55</td>
<td>70</td>
<td>Calcium-binding protein</td>
<td></td>
</tr>
<tr>
<td>ATCCCTCAGG</td>
<td>Claudin 4</td>
<td>72</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>29</td>
<td>116</td>
<td>40</td>
<td>81</td>
<td>58</td>
<td>56</td>
<td>Tight junction barrier function</td>
<td></td>
</tr>
<tr>
<td>GCCCTGACCAC</td>
<td>Trop-2</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>49</td>
<td>99</td>
<td>121</td>
<td>0</td>
<td>55</td>
<td>11</td>
<td>Tumor-associated calcium signal transducer</td>
<td></td>
</tr>
<tr>
<td>CACACTCCGCC</td>
<td>Keratin 18</td>
<td>112</td>
<td>9</td>
<td>9</td>
<td>28</td>
<td>322</td>
<td>151</td>
<td>142</td>
<td>132</td>
<td>425</td>
<td>26</td>
<td>15</td>
<td>Cytoskeletal and microfibrillar</td>
<td></td>
</tr>
<tr>
<td>GCGGGCCTCC</td>
<td>FXVY3 (Mat-8)</td>
<td>75</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>62</td>
<td>47</td>
<td>40</td>
<td>19</td>
<td>81</td>
<td>11</td>
<td>FXVY domain-containing ion transport regulator</td>
<td></td>
</tr>
<tr>
<td>TTGTTTTTTA</td>
<td>Tissue-type plasminogen activator</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>107</td>
<td>39</td>
<td>9</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>Tissue remodeling and degradation/protease</td>
<td></td>
</tr>
<tr>
<td>CTCTCTGTC</td>
<td>HE4</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>17</td>
<td>93</td>
<td>6</td>
<td>12</td>
<td>5</td>
<td>Secreted protease inhibitor</td>
<td></td>
</tr>
<tr>
<td>GCGCCGACCTTT</td>
<td>Prostate stem cell antigen</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>73</td>
<td>12</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>GPI-anchored glycoprotein/prostate-specific</td>
<td></td>
</tr>
<tr>
<td>CGCTGGACAA</td>
<td>Basic transcription factor 3</td>
<td>26</td>
<td>1</td>
<td>1</td>
<td>47</td>
<td>20</td>
<td>0</td>
<td>44</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>15</td>
<td>General transcription factor</td>
<td></td>
</tr>
<tr>
<td>GAATATTGCT</td>
<td>Collagen, type I, 1a</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>57</td>
<td>17</td>
<td>9</td>
<td>34</td>
<td>15</td>
<td>2</td>
<td>may be an extracellular matrix protein</td>
<td></td>
</tr>
<tr>
<td>AGCTGCTGAC</td>
<td>CHD-2</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>8</td>
<td>30</td>
<td>71</td>
<td>19</td>
<td>35</td>
<td>5</td>
<td>Sequence selective DNA binding</td>
<td></td>
</tr>
<tr>
<td>GCAGTCTCTG</td>
<td>Tetraspan 1</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>4</td>
<td>31</td>
<td>34</td>
<td>16</td>
<td>50</td>
<td>26</td>
<td>14</td>
<td>Possible interconnecting cell surface molecules</td>
<td></td>
</tr>
<tr>
<td>TTCTGGCTCA</td>
<td>Cysteine-rich protein 1 (intestinal)</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>68</td>
<td>13</td>
<td>30</td>
<td>3</td>
<td>16</td>
<td>49</td>
<td>22</td>
<td>Potential intracellular zinc transport protein</td>
<td></td>
</tr>
<tr>
<td>GCGGCTCAGA</td>
<td>Major vault protein</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>10</td>
<td>30</td>
<td>19</td>
<td>62</td>
<td>17</td>
<td>30</td>
<td>Nucleo-cytoplasmic transport/Drug resistant-related protein</td>
</tr>
<tr>
<td>GCTGTGGAC</td>
<td>Ribosomal protein S20</td>
<td>38</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>132</td>
<td>81</td>
<td>125</td>
<td>28</td>
<td>34</td>
<td>26</td>
<td>14</td>
<td>Protein synthesis</td>
<td></td>
</tr>
<tr>
<td>ATTTTTCAAA</td>
<td>Anterior gradient 2 homolog</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>62</td>
<td>47</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>19</td>
<td>Secreted</td>
<td></td>
</tr>
<tr>
<td>AMGATTTGAA</td>
<td>CEACAM6</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>34</td>
<td>4</td>
<td>31</td>
<td>47</td>
<td>3</td>
<td>50</td>
<td>Carcinomembrane antigen-related cell adhesion molecule</td>
<td></td>
</tr>
<tr>
<td>AGGCTGACGN</td>
<td>Adenyl cyclase-associated protein</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>26</td>
<td>30</td>
<td>16</td>
<td>22</td>
<td>15</td>
<td>30</td>
<td>Membrane protein/regulatory bifunctional roles</td>
</tr>
<tr>
<td>ATCTTTTTAA</td>
<td>Hypothetical protein FLJ11068</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>47</td>
<td>17</td>
<td>0</td>
<td>43</td>
<td>15</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCCTCTGACG</td>
<td>X</td>
<td>Homo sapiens cDNA: FLJ23356</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>10</td>
<td>13</td>
<td>31</td>
<td>47</td>
<td>6</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCCTCTTCTG</td>
<td>Retinoid acid induced 3</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>13</td>
<td>53</td>
<td>25</td>
<td>12</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGCGGGCGCTT</td>
<td>Hypothetical protein FLJ10903</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>16</td>
<td>30</td>
<td>19</td>
<td>31</td>
<td>12</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGCCTGACTT</td>
<td>Secretory leukocyte protease inhibitor</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>42</td>
<td>43</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGCTGACCA</td>
<td>Endobrevin</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>5</td>
<td>34</td>
<td>34</td>
<td>16</td>
<td>15</td>
<td>16</td>
<td>Vesicle-associated membrane protein 8</td>
<td></td>
</tr>
<tr>
<td>GCAATTTCFGT</td>
<td>Coated vesicle membrane protein</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>28</td>
<td>23</td>
<td>30</td>
<td>12</td>
<td>6</td>
<td>17</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACACCTGCTG</td>
<td>Homo sapiens cDNA: FLJ22182</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>23</td>
<td>0</td>
<td>56</td>
<td>12</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTAAGCCTGC</td>
<td>No reliable matches</td>
<td>71</td>
<td>18</td>
<td>18</td>
<td>10</td>
<td>68</td>
<td>70</td>
<td>77</td>
<td>487</td>
<td>1054</td>
<td>131</td>
<td>252</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCGGCGCGCT</td>
<td>No reliable matches</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>43</td>
<td>22</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GACACTCTTT</td>
<td>No reliable matches</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>30</td>
<td>25</td>
<td>19</td>
<td>9</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGGTGCCAG</td>
<td>Claudin 3</td>
<td>318</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>16</td>
<td>9</td>
<td>9</td>
<td>40</td>
<td>3</td>
<td>8</td>
<td>13</td>
<td>Tight junction barrier function</td>
<td></td>
</tr>
<tr>
<td>SGCGCTGCAG</td>
<td>S100A10</td>
<td>112</td>
<td>15</td>
<td>9</td>
<td>58</td>
<td>380</td>
<td>135</td>
<td>133</td>
<td>140</td>
<td>264</td>
<td>197</td>
<td>246</td>
<td>Calcium-binding protein</td>
<td></td>
</tr>
<tr>
<td>BACCTTGCAG</td>
<td>Transglutaminase 2</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>9</td>
<td>43</td>
<td>6</td>
<td>9</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGCCCTACTT</td>
<td>Homo sapiens cDNA: FLJ21010</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>8</td>
<td>26</td>
<td>13</td>
<td>9</td>
<td>28</td>
<td>3</td>
<td>2</td>
<td>Ca(2+)-binding protein required for T cell receptor-, Fas-, &amp; glucocorticoid-induced cell death</td>
<td></td>
</tr>
<tr>
<td>GCCCTCTTCTT</td>
<td>RAMP1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>10</td>
<td>17</td>
<td>3</td>
<td>47</td>
<td>0</td>
<td>8</td>
<td>Receptor (calcitonin) activity modifying protein</td>
<td></td>
</tr>
<tr>
<td>TTCTCTGCTC</td>
<td>14-3-3 protein sigma (stratifin)</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>18</td>
<td>0</td>
<td>40</td>
<td>19</td>
<td>6</td>
<td>5</td>
<td>p53-regulated inhibitor of G2/M progression</td>
<td></td>
</tr>
<tr>
<td>ATCTGGCGCT</td>
<td>Zona occludens 3</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>22</td>
<td>16</td>
<td>34</td>
<td>0</td>
<td>5</td>
<td>Tight junction protein</td>
<td></td>
</tr>
</tbody>
</table>

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).

* ESTs, tag matches multiple expressed sequence tags.

* Tags representing multiple genes.
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
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 ductal 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 be-
cause of cell culture. Finally, we set a cutoff to draw the greatest
attention to genes of higher expression levels, requiring expression at
level
3). Third, the genes that failed to conform by RT-PCR did not
observe differential tags retained a spectrum of GC content (Table
3). The number of tags produced in each permutation. Second, the set 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 differen-
tially 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.

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 artificial 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 deter-
mnant 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 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, trans-
membrane, and tight junction protein coding genes. This could cor-
respond 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 sug-
gested 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,
deferred in this study was subsequently developed as a histological
marker of pancreatic malignancy (30) and is a secreted protein de-
lected in the serum of a set of prostate cancer patients evaluated
(29). Interestingly, a group of ion-homeostasis related proteins, espe-
sially those specific for the calcium ion (Ca$^{2+}$) such as S100A4,
S100A10, Trop-2, AIF-1, and ALG-2, were identified as overex-
pressed.

---

**Table 3 Continued**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Gene Description</th>
<th>Count difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAGCCGACCCT</td>
<td>Frequentin (Drosophila) homolog</td>
<td>1</td>
</tr>
<tr>
<td>CCAAGGAAATAA*</td>
<td>Homo sapiens cDNA FLJ12683 fis ESTs</td>
<td>1</td>
</tr>
<tr>
<td>CAAAGGATGATG</td>
<td>Early growth response 1</td>
<td>9</td>
</tr>
<tr>
<td>GAAAGAAAGA*</td>
<td>Tolloid-like protein</td>
<td>1</td>
</tr>
<tr>
<td>TCAAGGATGATG</td>
<td>KIAA0460 protein</td>
<td>7</td>
</tr>
<tr>
<td>CTGCTCTCTC</td>
<td>Nuclear receptor subfamily 1</td>
<td>2</td>
</tr>
<tr>
<td>GAAAGAAAGA*</td>
<td>A novel transmembrane protein</td>
<td>0</td>
</tr>
<tr>
<td>GCCGAGGAGG</td>
<td>KIAA1333 protein</td>
<td>1</td>
</tr>
<tr>
<td>AGCTAGCGG</td>
<td>Ribosomal protein L13a</td>
<td>1</td>
</tr>
<tr>
<td>CAGAGGAGTC</td>
<td>Translation initiation factor 2C</td>
<td>0</td>
</tr>
</tbody>
</table>

---

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 “nomord” 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, Mkp1-like protein tyrosine phosphatase, tumor necrosis factor α-inducible protein, catenin α1 (102 kDa), RAD51 (Saccharomyces cerevisiae) homologue C, guanine nucleotide-binding protein γ5, BCL2-associated athanogene 3, and 21 others were overexpressed in the “nomord” 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 cRso00 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. 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, M KK4, and BRCA2, were known within the cell lines studied by SAGE. We performed groupwise 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 M KK4 wild-type versus M KK4 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.

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

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.

Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/62/3/819

Cited articles
This article cites 34 articles, 15 of which you can access for free at:
http://cancerres.aacrjournals.org/content/62/3/819.full.html#ref-list-1

Citing articles
This article has been cited by 23 HighWire-hosted articles. Access the articles at:
/content/62/3/819.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.