Identification of dysregulated pathways associated with pancreatic cancer by survival analysis

QIONG-YING YUAN¹, YAN-PING GU¹, CONG-JUN WANG², HUI ZHANG² and XING-PENG WANG³,⁴

Departments of ¹Gastroenterology and ²Biliary and Pancreatic Surgery, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120; ³Department of Gastroenterology, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai 200072; ⁴Department of Gastroenterology, Shanghai First People's Hospital, JiaoTong University School of Medicine, Shanghai 200080, P.R. China

Received November 21, 2013; Accepted May 28, 2014

DOI: 10.3892/mmr.2014.2693

Abstract. In order to identify the dysregulated pathways associated with pancreatic cancer, the fourth leading cause of cancer mortality in the United States, tumor and non-tumor samples were systematically analyzed in the present study. Initially, dysregulated genes in pancreatic cancer were identified using paired t-test. Subsequently, dysregulated biological pathways involved in the development of pancreatic cancer were identified by enrichment analysis. Finally, individual survival analysis of the significantly dysregulated functions was conducted at the pathway level. Our results indicated that the pathway named 'Pathways in cancer' was significantly correlated with survival time. In addition, the mean survival time of individual and genetic variation demonstrated a significantly negative correlation, that is, the lower the genetic variation, the longer the survival time. Furthermore, detailed analysis of genes on the pathway named 'Pathways in cancer' denoted that this pathway involved multiple cancer hallmark signals and several dysregulated cancer genes, including tumor protein p53, myelocytomatosis, Kirsten rat sarcoma, phosphatidylinositol 3-kinase, v-raf murine sarcoma viral oncogene homolog B1 and cyclin-dependent kinase inhibitor 2A. According to the DrugBank database, certain oncogenes have been identified to be key targets of drugs, including Sorafenib, Trastuzumab, Imatinib and Paclitaxel or were under investigation. An improved understanding of the pathophysiology of pancreatic cancer has been achieved based on our results and the present study aimed to provide guidance for the development of drugs to treat pancreatic cancer.

Introduction

As the fourth leading cause of cancer mortality in the United States (1), pancreatic cancer has been widely investigated in the past decades and remains an unsolved health problem (2). Despite significant achievements in the understanding of pancreatic cancer pathogenesis, an effective treatment remains to be developed. The higher mortality rate (3) (97-98% in the past five years) and lower disease prognosis are attributable to the complexity of pancreatic cancer pathogenesis. Several studies have demonstrated that, with the exception of family history (4), the environment, cigarette smoking and age (5) are closely associated with pancreatic cancer.

Pancreatic cancer is considered as a complex process resulting from the accumulation of multiple genetic mutations (6). Considerable effort has been made to examine specific markers for early diagnosis and targeted therapy. Based on the accumulation of our understanding of the pathogenesis of pancreatic cancer, different types of genes and diverse pathways have been demonstrated to be associated with pancreatic cancer (7-9). Identifying mutations in K-ras, an important oncogene, has been used as an early detection strategy (10). Tumor suppressor gene inactivation, including p16 and TP53, are also widely detected in pancreatic cancers (11,12). The mutation of oncogenes and tumor suppressor genes mainly affect the metabolic pathways or signal transduction pathways and lead to the dysfunction of growth or differentiation (13). Furthermore, it has been systematically studied about the molecular mechanisms of the Hedgehog (Hh) pathway which mediated regulation of human pancreatic cancer (14). In addition, genome-wide pathway analysis has been applied to identify the important signaling pathways, for example, a core set of 12 cellular signaling pathways has been defined by global genomic analysis (15) and the susceptibility loci have been identified by a genome-wide association study (16). By contrast, mass-spectrometry assisted proteomic analysis is used to promote the identification of novel prognostic protein markers (17).

Microarray technology combined with statistical analysis has become an efficient approach for the identification of novel prognostic makers. Survival analysis, a frequently used method, has been applied to investigate the prognostic genetic variation, the longer the survival time. Furthermore, detailed analysis of genes on the pathway named 'Pathways in cancer' denoted that this pathway involved multiple cancer hallmark signals and several dysregulated cancer genes, including tumor protein p53, myelocytomatosis, Kirsten rat sarcoma, phosphatidylinositol 3-kinase, v-raf murine sarcoma viral oncogene homolog B1 and cyclin-dependent kinase inhibitor 2A. According to the DrugBank database, certain oncogenes have been identified to be key targets of drugs, including Sorafenib, Trastuzumab, Imatinib and Paclitaxel or were under investigation. An improved understanding of the pathophysiology of pancreatic cancer has been achieved based on our results and the present study aimed to provide guidance for the development of drugs to treat pancreatic cancer.

Correspondence to: Dr Xing-Peng Wang, Department of Gastroenterology, Shanghai Tenth People's Hospital, Tongji University School of Medicine, 301 Yanchang Middle Road, Shanghai 200072, P.R. China
E-mail: xingpengwang@hotmail.com

Key words: dysregulated pathways, pancreatic cancer, genetic, variation
pathways associated with breast cancer (18). The present study employed survival analysis at the pathway level to estimate survival distributions, assess the hypotheses of survival distributions and identify the risk or prognostic factors (19), so as to identify the differentially expressed genes in pancreatic tumor and non-tumor samples.

Materials and methods

Expression profiles of pancreatic cancer genes. The data of GSE28735 were downloaded from the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE28735) (20). A total of 90 samples were included in the present study consisting of 45 pancreatic cancer samples and 45 adjacent non-tumor samples. The data were gained by Affymetrix GeneChip Human Gene 1.0 ST arrays (Affymetrix, Santa Clara, CA, USA) and 33,297 probes were assessed. Subsequently, probe ID was converted to Entrez Gene ID. A total of 18,963 genes were identified and used for the analysis.

Data pre-processing. The processing of raw data, involving background correction and normalization by quantile algorithm, was conducted using the Robust Multi-Ary Array Average method included in R software (Bell Laboratories, Murray Hill, NJ, USA). The expression level of the genes, which correspond to more than one probe, was defined as the mean of those probes. Differentially expressed genes were identified by paired t-test. The false discovery rate (FDR; multiple testing) was set to 0.5 and subsequent analysis was based on the preprocessed expression profile.

Survival analysis at the pathway level. In order to examine the correlation between survival time and signaling pathways, the 45 pancreatic tumor samples and 45 adjacent non-tumors were subjected to the following analysis: Initially, the difference in gene expression between pancreatic tumor samples and non-tumor samples was calculated. Secondly, the difference in gene expression (upregulated or downregulated) was identified by paired t-test. The false discovery rate (FDR; multiple testing) was set to 0.5 and subsequent analysis was based on the preprocessed expression profile.

Survival analysis at the pathway level. In order to examine the correlation between survival time and signaling pathways, the 45 pancreatic tumor samples and 45 adjacent non-tumors were subjected to the following analysis: Initially, the difference in gene expression between pancreatic tumor samples and non-tumor samples was calculated. Secondly, the difference in gene expression (upregulated or downregulated) was identified according to the gene expression value. Finally, the difference in the pathway in a given individual was defined as the mean of the ranks of differentially expressed genes in this pathway. Three samples were excluded from the following analysis due to non-availability of survival information.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway data. The pathway data were downloaded from the KEGG database (http://www.genome.jp/kegg/pathway.html) on 26th April, 2011. The KEGG pathway data were grouped into the following six categories: Cellular processes, environmental information processing, genetic information processing, human diseases, metabolism and organismal systems. A total of 232 pathways were included. Subsequent analysis was based on these pathways.

Enrichment analysis of biological pathways. In order to identify the function of genes, KEGG pathways enrichment analysis was performed applying the hypergeometric distribution method (21). The theory can be described as below: suppose that KEGG pathways contain N genes and n and m represent the differentially expressed genes in a given pathway, then the P-value of overlapped genes more than k is:

$$P = 1 - \sum_{n}^{k} \frac{c_{k}^{m} c_{N-m}^{N-k}}{c_{N}^{N}}$$

Univariate Cox proportional hazards model. Univariate Cox proportional hazards model (22) was employed to assess the association between genes and prognosis to screen differentially expressed genes associated with survival. Univariate Cox proportional hazards models can be described as the following: $h(t,X) = h(0)\exp(\beta X)$. In the function, $h(t,X)$ represents the instantaneous hazard when the event occurs at time, t, in response to covariate X. X refers to the expression level of gene. $h(0)$ represents the hazard when X=0. The weight of covariate to survival is reflected by $\beta$. When the value of $\beta$ is positive ($\beta>0$), the hazard function $h(t,X)$ and mortality risk are positively correlated with the value of X and vice versa. When the value of $\beta$ is zero, the covariate X is not associated with prognosis. Patients were then divided into two groups, a high risk group and low risk group, based on the score. The log-rank test for prognosis differences of the two groups was employed considering the existence of censored data. The log-rank test can be divided into the following four steps: Firstly, the survival data was converted into a matrix with two columns and two lines based on a time point. Secondly, the total observation number (O) was obtained by adding the observation number together in every matrix. Thirdly, the total expectation (E) was calculated by adding the expectation of every matrix at any time point. Finally, a significance test was performed based on $\chi^2$ statistics.

Results

Differentially expressed genes screening and pathway enrichment analysis. Using the paired t-test, 7,593 differentially expressed genes were screened, including 4,099 upregulated and 3,944 downregulated genes. In order to minimize false positives, a stringent FDR control (FDR<0.05) was applied. Pathway enrichment was conducted with a hypergeometric distribution method and significantly enriched KEGG pathways were screened applying stringent FDR control (FDR<5%) and analyzed. The results demonstrated that enriched pathways were mainly involved in: Metastasis-related pathways, including focal adhesion and extracellular matrix-receptor interaction; important signaling pathways including the p53 signaling pathway and Wnt signaling pathways; the cancer pathway such as 'Pathways in cancer'; and pancreatic cancer-related pathways such as 'Pancreatic secretion'. 'Pathways in cancer' and 'Pancreatic secretion' are descriptions of the specific expressions of pathway names obtained from the KEGG pathway analysis. The details are shown in Table I.

Survival analysis at the pathway level. Individual significance of the seven enriched pathways was examined and the functions of differentially expressed genes were identified. Subsequently, the correlation between pathway and survival was assessed. The samples were divided into two groups by the mean of the difference of each pathway in the individual:
The higher expression group and the lower expression group. Survival package (within R environment) was applied for the survival analysis. The results demonstrated that only one pathway was significantly different (P<0.05) between the two groups in the seven enriched pathways. The mean individual survival time of the group with higher genetic differences was 11 months. The mean survival time of individuals in group 2 with less genetic differences (18 months) was clearly longer than group 1 (data not shown). The detailed analysis of results are shown in Table II. The pathway named 'Pathways in cancer' was the only one that significantly affected survival. Previous studies demonstrated that the variation of certain genes in this pathway have significant effects on cancer development (23,24). Furthermore, the genes in this pathway were analyzed in detail to examine its function and identify potential targets. Survival difference (formula = Surv(time, status) - 'Pathways in cancer', data = pancreatic cancer) was used to test survival curve differences in 'Pathways in cancer'. The formula is an expression as for other survival models, of the form Surv(time, status) - predictors. For a k-sample test, each unique combination of predictors defines a subgroup. Status is the number of subjects contained in each stratum and may be optionally designed. Data is an optional data frame in which to interpret the variables occurring in the formula. The detailed results are shown in Table III.

### Table I. KEGG pathways within which DEGs were enriched (FDR=5%).

<table>
<thead>
<tr>
<th>Index</th>
<th>Pathway Access</th>
<th>Pathway name</th>
<th>path Gene</th>
<th>DEGs (n)</th>
<th>P-value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>hsa04510</td>
<td>Focal adhesion</td>
<td>200</td>
<td>123</td>
<td>2.10E-08</td>
<td>4.87E-06</td>
</tr>
<tr>
<td>2</td>
<td>hsa03050</td>
<td>Proteasome</td>
<td>45</td>
<td>35</td>
<td>1.24E-06</td>
<td>9.56E-05</td>
</tr>
<tr>
<td>3</td>
<td>hsa04512</td>
<td>ECM-receptor interaction</td>
<td>85</td>
<td>58</td>
<td>1.04E-06</td>
<td>9.56E-05</td>
</tr>
<tr>
<td>4</td>
<td>hsa05200</td>
<td>'Pathways in cancer'</td>
<td>327</td>
<td>175</td>
<td>1.72E-05</td>
<td>0.000997</td>
</tr>
<tr>
<td>5</td>
<td>hsa04115</td>
<td>p53 signaling pathway</td>
<td>69</td>
<td>44</td>
<td>0.000244</td>
<td>0.011299</td>
</tr>
<tr>
<td>6</td>
<td>hsa04310</td>
<td>Wnt signaling pathway</td>
<td>151</td>
<td>85</td>
<td>0.000313</td>
<td>0.012093</td>
</tr>
<tr>
<td>7</td>
<td>hsa04972</td>
<td>'Pancreatic secretion'</td>
<td>103</td>
<td>59</td>
<td>0.001415</td>
<td>0.046909</td>
</tr>
</tbody>
</table>

Pathway names are from the KEGG database; path Gene is the gene number included in the pathway; DEGs are the number of differential genes in the pathway; P-values were obtained by hypergeometric enrichment; FDR was obtained by BH correction. FDR, false discovery rate; ECM, extracellular matrix; KEGG, Kyoto Encyclopedia of Genes and Genomes; DEG, differentially expressed gene.

### Table II. Individual survival analysis at the pathway level between the two groups.

<table>
<thead>
<tr>
<th>Index</th>
<th>Pathway Access</th>
<th>Pathway name</th>
<th>(\chi^2)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>hsa04510</td>
<td>Focal adhesion</td>
<td>2.8</td>
<td>0.092</td>
</tr>
<tr>
<td>2</td>
<td>hsa03050</td>
<td>Proteasome</td>
<td>1.9</td>
<td>0.172</td>
</tr>
<tr>
<td>3</td>
<td>hsa04512</td>
<td>ECM-receptor interaction</td>
<td>0.1</td>
<td>0.816</td>
</tr>
<tr>
<td>4</td>
<td>hsa05200</td>
<td>'Pathways in cancer'</td>
<td>5.2</td>
<td>0.023</td>
</tr>
<tr>
<td>5</td>
<td>hsa04115</td>
<td>p53 signaling pathway</td>
<td>0.1</td>
<td>0.751</td>
</tr>
<tr>
<td>6</td>
<td>hsa04310</td>
<td>Wnt signaling pathway</td>
<td>0.5</td>
<td>0.482</td>
</tr>
<tr>
<td>7</td>
<td>hsa04972</td>
<td>'Pancreatic secretion'</td>
<td>0.1</td>
<td>0.713</td>
</tr>
</tbody>
</table>

Pathway names are from the Kyoto Encyclopedia of Genes and Genomes database. The \(\chi^2\) statistic was obtained by survival analysis between two groups. The P-value was calculated by the \(\chi^2\) test. ECM, extracellular matrix.

The higher expression group and the lower expression group. Survival package (within R environment) was applied for the survival analysis. The results demonstrated that only one pathway was significantly different (P<0.05) between the two groups in the seven enriched pathways. The mean individual survival time of the group with higher genetic differences was 11 months. The mean survival time of individuals in group 2 with less genetic differences (18 months) was clearly longer than group 1 (data not shown). The detailed analysis of results are shown in Table II. The pathway named 'Pathways in cancer' was the only one that significantly affected survival. Previous studies demonstrated that the variation of certain genes in this pathway have significant effects on cancer development (23,24). Furthermore, the genes in this pathway were analyzed in detail to examine its function and identify potential targets. Survival difference (formula = Surv(time, status) - 'Pathways in cancer', data = pancreatic cancer) was used to test survival curve differences in 'Pathways in cancer'. The formula is an expression as for other survival models, of the form Surv(time, status) - predictors. For a k-sample test, each unique combination of predictors defines a subgroup. Status is the number of subjects contained in each stratum and may be optionally designed. Data is an optional data frame in which to interpret the variables occurring in the formula. The detailed results are shown in Table III.

**Analysis of the pathway named 'Pathways in cancer'**. According to the above analysis, the pathway named 'Pathways in cancer' was significantly associated with survival. The genes in this pathway, particularly the differentially expressed genes (175 genes), were investigated in detail. Initially, the differentially expressed genes were mapped to pathways applying the online tool KEGG Mapper. As shown in Fig. 1, the differentially expressed genes were mainly distributed in the hallmarks of cancer development, including self-sufficiency in growth signals, insensitivity to anti-growth signals, resisting cell death, limitless replicative potential, sustained vascular growth, tissue invasion, metastasis and genomic instability. In addition, the results confirmed that the development of cancer is a complex process that requires dysregulation of multiple genes. Oncogenes, which are closely associated with the development of cancer, in this pathway were then analyzed in detail. The oncogenes were downloaded from the Cancer Gene Census database (http://cancer.sanger.ac.uk/cosmic/census) containing 485 oncogenes. According to the Cancer Gene Census database, 42 oncogenes, including classical oncogenes such as tumor protein p53 (TP53), myelocytomatosis (MYC), Kirsten rat sarcoma (KRAS), phosphatidylinositol 3-kinase (PIK3CA), v-raf murine sarcoma viral oncogene homolog B1 (BRAF), a human gene encoding a protein called B-Raf,
and cyclin-dependent kinase inhibitor 2A (CDKN2A), were mapped on the pathway. According to the drugbank database, certain oncogenes have been validated to be the targets of drugs, including Sorafenib, Trastuzumab, Imatinib, Paclitaxel (approved by FDA) and are under investigation. Based on the above analysis, it is clear that regulation of dysregulated genes, particularly oncogenes, may aid in the prevention and treatment of pancreatic cancer. In addition, it is promising for drug development based on dysregulated genes.

Discussion

Based on the above analysis, dysregulated pathways in pancreatic cancer were identified using 90 samples by enrichment analysis. The present study initially identified that the pathway named ‘Pathways in cancer’ has significant correlation with survival rate by individual survival analysis. This result was also supported by functional analysis at the gene level. The pathway named ‘Pathways in cancer’ involves several signaling pathways, including the ErbB signaling pathway, Janus kinase-signal transducer and activator of transcription (Jak-STAT) signaling pathway, phosphoinositide 3-kinase-Akt signaling pathway, p53 signaling pathway and vascular endothelial growth factor signaling pathway. It has been demonstrated that the ErbB signaling pathway can mediate the tumor-stroma interaction in pancreatic cancer (25) and the ErbB family has been considered as the target for the treatment of human epidermal growth factor receptor in positive pancreatic cancer (26). Previous studies have indicated the abnormality of the Jak-STAT pathway in pancreatic cancer.

Table III. Survival difference (formula = Surv (time, stat) - ‘Pathways in cancer’, data = pancreatic cancer).

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Observed</th>
<th>Expected</th>
<th>(O-E)^2/E</th>
<th>(O-E)^2/V</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>11</td>
<td>16.7</td>
<td>1.94</td>
<td>5.17</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>18</td>
<td>12.3</td>
<td>2.62</td>
<td>5.17</td>
</tr>
</tbody>
</table>

\( \chi^2 = 5.2, 1 \) degrees of freedom, \( P=0.023 \). Observed, the weighted observed number of events in each group; Expected, the weighted expected number of events in each group. N, the number of subjects in each group; V, the variance matrix of the test; O, observation number; E, total expectation.
cancer (27,28). The critical p53 signaling pathway has also been identified to be important in pancreatic cancer (29). Furthermore, detailed analysis at the gene level identified the dysregulation of several classical oncogenes, including TP53, MYC, KRAS, PIK3CA, BRAF and CDKN2A. Several of these genes have been demonstrated to be associated with pancreatic cancer. Rozenblum et al demonstrated that all the 42 pancreatic cancer samples had a KRAS gene mutation, indicating that KRAS mutation frequency is positively associated with pancreatic cancer (30). The Hh pathway can be activated by oncogenic KRAS by RAF/MEK/mitogen activated protein kinase signaling in PDA cells (31). CDKN1A encodes two tumor suppressors (INK1A and alternative reading frame) (32) and CDKN1A allele mutations can give rise to a higher risk of pancreatic cancer (33,34). In addition, overexpression of CDKN1A associated with apoptosis can increase pancreatic invasive ductal carcinoma (35). Several studies have demonstrated that >50% of pancreatic cancers have TP53 tumor suppressor gene mutations (30) and p53 protein expression can be used as a prognostic factor in pancreatic cancer (36). The inactivation of the tumor suppressor gene TP53 is a late event in pancreatic cancer and mutations in TP53 may lead to constitutive activation of transcription factors (2). The MYC gene encoding transcription factor has been demonstrated to be overexpressed (~50%) in pancreatic cancer (37). The activation of the oncogenic MYC gene in pancreatic cancer can be mediated by the Ca2+/calcineurin signaling pathway and by ectopic activation of nuclear factor of activated T-cells cytoplasmic 1 (38). As a serine/threonine kinase, BRAF mutations can activate kinase effectors, which are important in pancreatic cancer (39).

According to the DrugBank database, certain oncogenes have been validated as targets of drugs, including TP53, BRAF, KRAS and PIK3CA, some of which are under investigation. With the aid of virtual screening and rational design, numerous small molecules have been designed which are able to bind to the tumor suppressor p53. PhiKan083 can bind to the drug-gable target, the oncogenic p53 mutant, with proper affinity, and the melting temperature of the mutant is raised (40). The KRAS gene has also been widely investigated as a target for the treatment of pancreatic cancer. In vitro experiments have demonstrated that silencing mutant KRAS altered the behavior of pancreatic cancer cells using RNA interference (41). Aoki et al also confirmed that silencing KRAS mutations significantly suppressed the growth of pancreatic cancer cells using liposome-mediated transfection (42).

Based on the above analysis, the development of drugs specific to those dysregulated genes associated with pancreatic cancer, is promising. An improved understanding of the pathophysiology of pancreatic cancer has been achieved based on our results and the present study aimed to provide guidance for the treatment of pancreatic cancer.

Acknowledgements

This study was supported by grants from the National Natural Science Foundation of China (nos. 30971359, 30872510, 81272534 and 81260349). The authors would like to thank the Department of Gastroenterology, Shanghai Tenth People's Hospital for their support.

References


