Time series expression patterns reveal the molecular processes of pancreatic cancer progression

Yanmei Zhong1, Fan Cheng1, Mengmeng Yang2, Hao Song1, Ruixi Zhao1

1Department of Gastroenterology, Weifang People’s Hospital, Weifang 261000, P.R. China. 2Department of Thyroid and Breast Surgery, Weifang People’s Hospital, Weifang 261000, P.R. China.

Summary

Purpose: Pancreatic cancer is a fatal malignant tumor with no obvious characteristics in the early stage of onset and high metastatic ability which results in a low survival. Understanding the detailed process of pancreatic cancer contributes to new treatments to prolong patients’ survival.

Methods: We carried out an in-depth analysis by modularization while seeking for critical genes in the pathogenesis of pancreatic cancer so as to identify the molecular mechanisms of the condition using differential analysis, co-expression module analysis, enrichment analysis, and network connectivity analysis. In light of the hypergeometric test, ncRNA (non-coding RNA) and transcription factors that regulate the module would be predicted.

Results: Conclusively, seven co-expression modules were obtained, in which CPA2 and A1BG were significantly differentially expressed in patients who had pancreatic cancer with active regulation in dysfunction modules. The modular genes significantly participated in second-messenger-mediated signaling as well as cellular calcium homeostasis and also controlled the interactions of neuroactive ligand-receptor. Besides, we identified ncRNA pivot including FENDRR and miR-92a-3p as well as transcription factors pivot including SPI1, STAT5A which significantly regulated the dysfunction module.

Conclusion: This study can help reveal core dysfunction modules, potential regulatory factors, and driver genes for pancreatic cancer, enhancing the understanding of its pathogenesis and providing a reference for prediction with respect to the survival time of patients with this condition.

Key words: pancreatic cancer, co-expression network, enrichment analysis, regulatory factors

Introduction

Pancreatic cancer, a highly fatal gastrointestinal malignancy [1], currently is a deadliest cancer, with its incidence keeping increasing [2]. Since not many screening tools could be used to detect it at an early stage, 94% of patients diagnosed with advanced pancreatic cancer will die within 5 years while the main features are fibroproliferation and matrix reaction [3,4]. In patients who have locally advanced pancreatic cancer, the main complications of intraoperative radiation therapy are gastric fistula, abdominal infection, and hemorrhage, with the highest incidence of gastric fistula [5]. So far, although certain risk factors have been identified, the specific pathogenesis of pancreatic cancer remains unclear [6]. For example, in genetics, tumorigenesis in pancreatic cancer is associated with germline genetic and somatic mutations variants on BRCA1/2 [7]. Lee and Song analyzed the GWAS data of pancreatic cancer by using ICSN pathway and could identify 18 candidate single nucleotide polymorphisms (SNPs) 11 genes involving ADCY9, ADCY10, HNF1A and HNF4G, and 30 signaling pathways which made patients susceptible to pancreatic cancer [8]. Scientists have identi-
fied some potential target sites for the treatment of pancreatic cancer involving Liprin-α4 [9], CD44 [10], CBX7 [11], and TM4SF1 [12] and TUSC3 [13] with salient efforts. Coronet et al. have a deeper comprehension of the pancreatic cancer biology and of new technologies that have developed to enhance the detection, staging, and treatment of this condition [14]. However, due to characteristics like drug resistance, high metastatic rate, poor prognosis, and recurrence, pancreatic cancer still has a high mortality rate and surgical resection is the sole treatment that may cure this disease, while FOLFIRINOX (fluorouracil, irinotecan, folinic acid, oxaliplatin) and gemcitabine plus nanoparticle albumin bound paclitaxel (nab-paclitaxel) are not preferred treatments for patients not suitable for surgery [15]. Although researchers have conducted a series of studies on pancreatic cancer, the global effects of these outcomes remain unclear. For comprehensive and in-depth apprehension of the pathogenesis of this condition, we conducted a systematic modular analysis to determine the dysfunction modules and core molecules between them in order to further explore the driving genes of pancreatic cancer.

Methods

Data resource

Tumor Cancer Genome Atlas (TCGA) is a project jointly supervised by the National Cancer Institute and the National Human Genome Research Institute. High-throughput genomic analysis techniques would have a better understanding of cancer and develop skills for preventing disease, diagnosing disease and treating cancer. We first downloaded TCGA pancreatic cancer RNA-Seq data and screened ncRNA-mRNA interaction pairs with a score of ≥ 0.5 from the RAID v2.0 database [16] which has 451957 interaction pairs involving 5431 ncRNAs and enrolls over 5.27 million RNA-related interactions, comprising over 4 million RNA-RNA interactions and over 1.2 million RNA-protein interactions. Additionally, there are 130 000 RNA/protein symbols in 60 species, which can help comprehensively observe various RNA-related interactions. All human transcription factor target data were downloaded and used in the general database TRRUST v2 of transcriptional studies [17], comprising 2492 transcription factors and 9396 interaction pairs.

Differentially expressed gene

We performed the differential expression analysis of the gene expression profiling data using the R language limma package [18-20]. The correct background function was used to perform both background correction and data normalization. The normalization between arrays function quantize, the control probe, and the low expression probe would be filtered. Then, the differentially expressed genes of the data set were determined based on the lmFit and eBayes functions with default parameters.

Co-expression analysis

For exploring the molecular process of pancreatic cancer staging, we performed a differential analysis for each stage of pancreatic cancer and integrated differential gene expression profiles of pancreatic cancer staging in light of disease samples and normal samples, and used a weighted gene co-expression network analysis (WGCNA) [21] to analyze the differential expression matrix of pancreatic cancer to find out a gene module for synergistic expression, exploring differentially expressed genes in pancreatic cancer. The correlation coefficient weighting value was used, that is, the gene correlation coefficient which was taken to work out the correlation coefficient (Pearson’s coefficient) between any two genes which were calculated. The connections between genes were subject to scale-free networks, enabling the algorithm more significant in terms of biology. Thereupon, correlation coefficients established a hierarchical clustering tree between genes while diverse branches of it represented various gene modules as well as diverse colors represented varying modules. In each dysfunctional module, according to the magnitude of the regulatory power of the gene, the essential genes were found leading to the dysfunctional module, considered as critical genes that are responsible for the growth and proliferation of pancreatic cancer cells.

Enrichment analysis

Exploring the functions and signaling pathways involved in gene expression often helps study the molecular mechanisms of disease. The enrichment analysis of the functions and pathways of genes in dysfunctional modules is a useful means to explore the underlying mechanisms of pancreatic cancer. Therefore, regarding functional genes of pancreatic cancer, GO function and KEGG pathway enrichment analysis were performed with the R language Cluster profiler package [22]. Cluster profiler is a Bioconductor software package that provides both statistical analysis and visualization of functional clustering of gene sets or gene clusters.

Transcription factors and ncRNAs regulate dysfunctional modules

Non-coding RNA (ncRNA) and transcription factors (TF) often drive the transcriptional and post-transcriptional regulation of genes. We have scientifically predicted and tested the role of the pancreatic cancer dysfunction module. In the development and proliferation of the cells, pivot regulators would be defined as modulators with significant regulatory impact on modules, including ncRNA and TF. There were more than two organizational connections between each regulator and each module, and the significance of the enriched target in each module calculated based on the hypergeometric test is p value<0.01.

Patient and blood samples

All blood samples were obtained by experienced hematologists and informed consents were obtained from all patients. Human tissue samples were collected according to the International Ethical Guidelines for Bio-
medical Research involving Human and Subjects. This research was approved by the China-Japan Union Hospital of Jilin University and carried out in line with the regulations of the China-Japan Union Hospital of Jilin University.

Verification of key genes by qRT-PCR

Specifically, total RNA in the blood was extracted and transcribed into cDNA with a reverse transcription kit and qRT-PCR reaction was conducted with the SYBR qRT-PCR Detection Kit. The qRT-PCR program began the initial 3-min denaturation step at 95°C to stimulate the hot-start iTaq™ DNA polymerase, followed by 45 cycles of denaturation at 95°C for 10 s and annealing and extension at 60°C for 45 s. The internal reference genes were beta-actin and U6.

Results

Identification of disordered molecules in the staging of pancreatic cancer

Disordered molecules are determining the pancreatic cancer staging. Biologists have conducted many experiments and studies on the pathogenesis of pancreatic cancer, and thus they identified potential genes for the deterioration of pancreatic cancer. However, the complex molecular connections and overall effects of these genes remain unclear. According to pancreatic cancer in four different peri-

Table 1. Hub gene of modules

<table>
<thead>
<tr>
<th>Colour</th>
<th>Hub Genes</th>
<th>Module</th>
</tr>
</thead>
<tbody>
<tr>
<td>black</td>
<td>CPA2</td>
<td>m7</td>
</tr>
<tr>
<td>blue</td>
<td>MIR7-1</td>
<td>m2</td>
</tr>
<tr>
<td>brown</td>
<td>TMCC2</td>
<td>m3</td>
</tr>
<tr>
<td>green</td>
<td>A1BG</td>
<td>m5</td>
</tr>
<tr>
<td>red</td>
<td>CERS1</td>
<td>m6</td>
</tr>
<tr>
<td>turquoise</td>
<td>NCKAP1L</td>
<td>m1</td>
</tr>
<tr>
<td>yellow</td>
<td>DNAJC30</td>
<td>m4</td>
</tr>
</tbody>
</table>

Figure 1. Synergistic expression of four differentially expressed genes in pancreatic cancer in patient samples. A: The seven co-expression groups which were obtained by clustering were identified as modules, and seven colors represent seven co-expression modules. B: All genes were shown in expression heat map in the sample, and the expression behavior was clustered into seven co-expression modules. C: Each row represents a module, each column represents a phenotype, the color of each cell is mapped by the corresponding correlation coefficient, the value would be counted from -1 to 1, the color transitions from blue to white, and then transitions to red.
ods (clinical stage I, II, III, IV) from both disease samples and normal samples, we used differential microarray data for differential expression analysis to observe the molecular differences in the progression of pancreatic cancer, identify differential gene expression (DEG) of time series expression from pancreatic cancer and obtain genes that may lead to deterioration of pancreatic cancer by integrating differential genes in four periods. The results showed a total of 1339 differential genes (Figure 2 S1) and that disorder molecules exhibited in these differential genes when it comes to time for series expression from pancreatic cancer.

Identification of functional disorder module in pancreatic cancer

In light of 1339 differential genes, an expression profile matrix was created by patient samples and their interaction genes that were dysregulated by time-series expression from pancreatic cancer. According to the Weighted Gene Co-Expression Network Analysis (WGCNA), the genes exhibited markedly co-expression in the performance of the disease sample. Regarding expression behavior, we can learn which of the complex synergy between genes would be beneficial. We could cluster the expression behavior of pancreatic cancer from patient samples into modules, observing the relationship between genes. Since complex synergies could be beneficial, seven functional barrier modules for pancreatic cancer were obtained (Figure 1A, B) by determining the co-expression panel as a module. According to essential genes of each module from the functional disorder module, core genes such as CPA2, A1BG were found (Table 1). Linking the module to phenotypic data, we found that MEyellow was associated with clinical stage I of pancreatic cancer, MEblack was related to II, MEturquoise was related to III, and MEblue was associated with IV (Figure 1C). In addition, the downregulation of genes in these modules may have a certain impact on different stages while the modules related to the
four periods stood for different levels of expression with respective regulatory effort (Figure 2). This may also be a key factor affecting the staging of pancreatic cancer.

**Functions and pathways involved in the genes of interest**

Function and pathway are essential mediators of the physiological response of the disease. Exploring both the services and pathways involved in the dysfunctional module genes not only determines the upstream and downstream relationship between different genes in the same biological pathway in the module but also establishes a molecular bridge between the module and the disease in system biology. More importantly, it deepens the understanding of the underlying molecular mechanism of the disease. We carried out function and KEGG pathway enrichment analysis on seven modules and obtained 11861 biological processes, 1407 cells, 2208 molecular functions, and 675 KEGG pathways (Figure 2 S2, Figure 3A and B). These functions were mainly focused on second-messenger-mediated signaling, cellular

---

**Figure 3.** Functional and pathway enrichment analysis excerpts of the module gene. A: Module gene GO function enrichment analysis excerpt. The color increases from blue to purple, and the enrichment rises significantly. The larger the circle, the more significant the proportion of the gene in the module that accounts for the GO function. B: Pathway enrichment analysis excerpt of KEGG as the module gene. The color goes from blue to purple, and the enrichment increases significantly. The larger the circle, the more significant the proportion of the gene in the KEGG pathway entry.
calcium homeostasis, activation of protein kinase activity, regulation of fluid and electrolyte balance, and other biological processes.

**TF and ncRNA drive pancreatic cancer progression**

In terms of systems biology and systems genetics, transcription and post-transcriptional regulation of genes have long been recognized as crucial regulators of disease development, while transcription factors and ncRNAs as common regulators. Although ncRNA has been valued by many biologists with respect to the development of the condition, the regulation of both single TF, ncRNA and multiple TF, few studies focused on their overall global effects on dysfunctional mechanisms and the role of bridges in development. Thus, in this study, to the module genes, based on the targeted regulatory relationship of TF and ncRNA, the pivotal analysis for module genes was carried out in order to explore critical transcriptional regulators that regulate the progression of pancreatic cancer. The predicted results (Figure 2 S3,S4, Figure 4 A,B) showed that 263 ncRNAs got involved in 263 ncRNA-module regulatory pairs while 21 transcription factors in 21 TF-module target pairs. In addition, the number of pivot control modules was statistically analyzed, while only one dysfunctional module would be adjusted for both ncRNA and TF. By mediating dysfunctional modules, these transcription factors and ncRNAs may regulate the progression of pancreatic cancer. Thus, we identified these potential regulatory factors as dysfunctional molecules in the pathogenesis of pancreatic cancer. When the general observation was done to the staging mechanism of pancreatic cancer, five of the target genes of ncRNA including FENDRR and

**Figure 4.** Regulatory effect of the regulator on the dysfunction module. **A:** The blue circle represents the module, and the orange circle represents the ncRNA. **B:** Orange squares represent modules, and blue squares represent TF.
miR-92a-3p were found to be hub genes. Therefore, FENDRR is a crucial factor in regulating progression of pancreatic cancer as the expression level of key genes was verified by qRT-PCR (Figure 5), consistent with the previous results.

**Discussion**

Pancreatic cancer is typically difficult to be detected early, since most patients with this disease have no apparent symptoms until the disease enters advanced stage [23,24]. Even after surgical resection, pancreatic cancer will recur in most patients [25], and hence it is essential to understand the pathogenesis of the disease. Although the researchers studied pancreatic cancer from various aspects and they have done a summary report to the TCGA database, the detailed pathogenesis of the disease remains unclear. We collected RNA-seq data from the TCGA database of pancreatic cancer and conducted a series of bioinformatics analyses based on the differential gene expression profiles of different clinical stages of pancreatic cancer, aiming to better realize the molecular mechanism of the disease. At the module level, the module significantly participated in second-messenger-mediated signaling, cellular calcium homeostasis, activation of protein kinase activity, regulation of humoral levels and other biological processes. Besides, the module was also significantly involved in signaling pathways such as neuroactive ligand-receptor interactions. The Genome-wide association study found that pancreatic cancer risk was remarkably related to neuroactive ligand-receptor interaction and olfactory transduction [26]. In addition, the module gene determines JAK-STAT, Wnt, PI3K-AKT, AMPK, and other signaling pathways. For example, the JAK/STAT pathway up-regulates D-L1 expression in pancreatic cancer cell lines, while stemming the PD-1/PD-L1 pathway may improve chances of survival in patients with this condition [27]. In addition, CDGSH iron-sulfur domain 2 activated the proliferation and EMT of pancreatic cancer cells through the Wnt/-catenin pathway with prognostic value in human pancreatic cancer [28]. Targeted Akt/PI3K signaling pathway can be used as an underlying therapy for the condition. Abnormal activation of this pathway includes the regulation of metabolism, survival, cell cycle progression and apoptosis [29]. Besides, without AMPK activation invasion and metastasis of pancreatic cancer are enhanced through hsf1-dependent pathways [30]. At the molecular level, through co-expression modules, we have explored the core genes of dysfunction from seven regulatory modules like CPA2 and A1BG. These core genes were not only significantly differentially expressed, but also essential to the regulation in the dysfunction module. Among them, in the carboxypeptidase, the expression of four carboxypeptidase proteins such as A1 (CPA1), A2 (CPA2), B1 (CPB1) and chymotrypsin C (CTRC) were significantly downgrading in pancreatic ductal adenocarcinoma (PDAC) tissues.

**Figure 5.** The relative expression level of A1BG (A), CPA (B), FENDRR (C), and SPI1 (D). P<0.05.
Moreover, they may be novel biomarkers in patients with PDAC [30]. Recently, except for PNET, MMP-9, DJ-1 and α-1-B glycoprotein (A1BG), have been proved as prognostic markers for some malignancies [31]. The overexpression of miR-7-1 has promoted the efficacy of green tea polyphenols and induced the efficiency of apoptosis in malignant neuroblastoma SH-SY5Y and SK-N-DZ cells [32]. Studies with respect to miR-7-1, A1BG, TMCC2 and CERS1 have not found any effect on pancreatic cancer, and studies on NCKAP1L and DNAJC30 are even fewer. The effect of other driving genes in the pathogenesis of pancreatic cancer remains to be further explored. The above genes, the driving genes of the dysfunction modules, promote the occurrence and development of the disease, which can be considered as the potential driving genes for the deterioration of pancreatic cancer.

Finally, we predicted that 263 ncRNAs and 21 transcription factors were involved in pancreatic cancer progression through the mediating module and, according to the statistical analysis, we determined that all ncRNAs and transcription factors had significant effects on only one dysfunctional module. Among them, FENDRR suppresses proliferation and promotes apoptosis, which is related to the good prognosis of breast cancer and is important to the growth and development of breast cancer [33]. Meanwhile, FENDRR also inhibited the growth and invasiveness of NSCLC cells through sponge miR-761 [34]. Besides, signal transduction and transcriptional activators (STAT) are transcription factors participating in a variety of cellular functions [35]. Activation of JAK/STAT promotes the status of inflammatory cancer-associated fibroblasts (iCAF), which contributes to the growth of pancreatic PDAC in the fibroblast wall (PDAC) microenvironment [36]. JAK/STAT signal transduction pathway is the target of many regulatory factors to inhibit the invasion and migration of pancreatic cancer cells [37]. These ncRNA and regulatory factors that significantly regulate cancer may also take part in the basic process of tumor microenvironment, which can be used as candidate molecules for further molecular validation and therapeutic targets. In summary, this study introduced the association between multifactor-mediated dysfunction module genes and pancreatic cancer in detail, identifying potential therapeutic targets and associated biological processes. In the future this will provide abundant candidate resources for experimental verification and drug retargeting and theoretical direction for the biological research of pancreatic cancer.

In this study, we identified a number of coding and non-coding factors to regulate the progression of pancreatic cancer, which play a regulatory role through related biological processes and signaling pathways.

Conflict of interests

The authors declare no conflict of interests.

References

35. Wu HT, Liu J, Li GW, Shen JX, Huang YT. The transcriptional STAT3 is a potential target, whereas transcriptional STAT5A/5B/6 are new biomarkers for prognosis in human breast carcinoma. Oncotarget 2017;8:36279-88.