ejpmr, 2021,8(11), 527-533

SJIF Impact Factor 6.222

EUROPEAN JOURNAL OF PHARMACEUTICAL AND MEDICAL RESEARCH

www.ejpmr.com

Research Article

ISSN 2394-3211

EJPMR

BIOINFORMATICAL ANALYSIS OF MICROARRAY GENE CHIP DATA FOR THE SCREENING OF KEY GENES INVOLVED IN PANCREATIC DUCTAL ADENOCARCINOMA

Mubeen Hussein Arawker[1], Li Zhen Nan[1]*, Shantanu Baral[2], Mohammad Said Jalloh[2] and Kashif Ali[3]

[1]Department of Hepatobiliary, General Surgery, Northern Jiangsu People's Hospital (Subei), Affiliated Hospital of Yangzhou University, Yangzhou, China.

[2]Department of Gastrointestinal Surgery, General Surgery, Northern Jiangsu People's Hospital (Subei), Affiliated Hospital of Yangzhou University, Yangzhou, China.

[3]Department of Gastroenterology, Yangzhou University, China.

***Corresponding Author: Li Zhen Nan**

Department of Hepatobiliary, General Surgery, Northern Jiangsu People's Hospital (Subei), Affiliated Hospital of Yangzhou University, Yangzhou, China.

Article Received on 09/09/2021 **Article Revised on 29/09/2021** **Article Accepted on 19/10/2021**

ABSTRACT

Objective- The goal of this study is to use bioinformatics to investigate important genes and pathways linked to pancreatic ductal adenocarcinoma (PDAC) in order to better understand the underlying processes. Our discovery sheds fresh light on the pathophysiology of PDAC. **Methods**: GSE28735, GSE15471, and GSE101448 gene expression profiles were acquired from the Gene Expression Omnibus collection, which included 108 pancreatic ductal adenocarcinoma samples and 97 precancerous tissues. DEGs were evaluated with the help of the R programmes limma and impute. The gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were carried out using the online analytic tools DAVID. STRING was used to create a protein-protein interaction network, which was then displayed using Cytoscape software. The Molecular Complex Detection (MCODE) plugin was used to identify the hub genes. **Results**: A total of 161 DEGs with overlap were identified, including 54 up-regulated genes and 107 down-regulated genes (| log2 fold-change (FC) | > 2, P 0. 05). Extracellular exosome, extracellular space, and extracellular matrix structure were all substantially enriched in DEGs. In addition, three KEGG pathways, including pancreatic secretion, protein digestion and absorption, and ECM - receptor interaction, were substantially enriched. **Conclusion**: The hub genes ALB, COL1 1 A1, COL3 A1, FN1, EGF, COL1 A1, MMP9, COL5A2, ITGA2, and COL6A3 might be used as biomarkers or therapeutic targets for PDAC. Furthermore, protein digestion and absorption, as well as ECM-receptor interaction pathways, play important roles in PDAC development. Our discovery sheds fresh light on the pathophysiology of PDAC. **Summary**: To analyze pancreatic ductal adenocarcinoma utilizing bioinformatics approaches (PDAC) and screen important genes. Gene Expression Database from a Public Database was used as the approach (GEO) PDAC Gene Expression Profiling Chip (PDAC) may be downloaded here. GSE28735, GSE15471, GSE101448. 108 example PDAC Sample and 97 sample of tissue adjacent to cancer. R Code Language limma Package was used in screening of differentially expressed genes. DAVID Database and online analysis tools were used to carry out differential genes separately GO Functional enrichment analysis and KEGG Pathway enrichment analysis. We created a differential protein interaction network and further screened important genes using the STRING Database and Cytoscape tools. the outcome: 3 All gene expression profiling chips have this feature. 54 Up-regulated genes and 107 Down-regulated genes are among the 164 differentially expressed genes (| log2 fold - change (FC) | > 2,P 0.05). GO Functional enrichment analysis showed that the differential gene and extracellular exosome, extracellular space, extracellular matrix organization closely related. KEGG Pathway analysis shows that differential genes are mainly enriched in protein digestion and absorption, ECM — receptor interaction with focal adhesion. The protein interaction network diagram shows the most nodes 10 The pivot genes are ALB, COL11A1, COL3A1, FN1, EGF, COL1A1, MMP9, COL5A2, ITGA2, COL6A3. **Conclusion:** These 10 Key genes may be Play an important role in the occurrence and development, and is expected to become PDAC Biological targets for diagnosis and treatment, for further research PDAC The molecular mechanism of development provides a theoretical basis.

KEYWORDS: PDAC, Bioinformatics, Microarray, Gene, Gene Chip, Cancer, Prognosis.

INTRODUCTION

Pancreatic ductal adenocarcinoma (pancreatic ductal adenocarcinoma, PDAC) It is a malignant and aggressive tumor that occurs in the digestive system, with a global incidence of about 6.29/100 000.[1] PDAC is usually detected at an advanced stage when it is discovered, its malignancy is high, the disease progresses quickly, and the prognosis of the disease is poor. The overall survival rate is only 8%. The median survival time is only 4 ~ 6 Months, although in PC, progress has been made in related tumor biology research. However, molecular targeted therapy has not significantly improved the prognosis of this fatal disease.[2,3] In recent years, PDAC The morbidity and mortality are on the rise, and it is expected that by 2030, PDAC will become the second most common cause of cancer-related deaths.[4] Therefore, searching for specific tumor biomarkers is of great significance for the early detection of PDAC and the evaluation of the disease process.

As a new treatment strategy, gene therapy has made progress, and one of the most important aspects is to determine the correct target gene. Gene Expression Omnibus Database (GEO) provides for the study of tumor gene changes and pathway changes. Given the opportunity, bioinformatics methods combined with microarray technology can quickly and effectively monitor changes in the expression levels of tens of thousands of genes and has been widely used in molecular diagnosis, tumor classification, disease prognosis prediction, and the development of anti-tumor drugs. In this article, we include 3 PDAC Gene expression profiling chip includes 108 example PDAC with 97 samples of adjacent tissues, using bioinformatics to screen differentially expressed genes and analyzing differentially expressed genes through multiple databases, aiming to screen for clinical PDAC Related molecular markers and drug targets provide a theoretical basis.

1. MATERIALS AND METHODS

1.1 Gene expression profiling chip

The data was obtained from GEO database,[5] for PDAC Gene expression profiling chip GSE28735, GSE15471 and GSE101448. GSE28735 include 45 PDAC samples and 45 sample of adjacent tissues; GSE15471 include 39 PDAC samples and 39 samples of adjacent tissues; GSE101448 include 24 PDAC samples and 13 adjacent tissue samples, including a total of 3 expression profiling chips include 108 PDAC samples and 97 tissue samples adjacent tissue, the chip specific information is shown in Table 1.

Table 1: Features of the included data set.

	PDAC	Tumor	
ID	samples(n)	tissue(n)	Analysis platform
GSE28735	45	45	[HuGene - 1_0 - st] Affymetrix Human Gene 1. 0 ST Array
GSE15471	39	39	[HG - U133_Plus_2] Affymetrix Human Genome U133 Plus 2. 0 Array
GSE101448	24	13	Illumina Human HT - 12 V4. 0 expression bead chip

1.2 Screening for differentially expressed genes

Standardized analysis for the three respective expression profile chip was normalized by R Language limma Package.[6] PDAC The experimental group and the control group were compared and analyzed, log2 and the parameters used were fold - change (FC) |＞2 and P＜ 0.05. Vein diagram was constructed on this basis and the DEG's were obtained from the three datasets.

1.3 Differentially expressed genes GO function enrichment analysis and KEGG pathway enrichment analysis

Using DAVID (https://david.ncifcrf.gov/) Online database differentially expressed genes were obtained and distinguished into upregulated and down regulated genes. GO Functional enrichment analysis was done by Kabos 3.0 program.[7] This database was also used for KEGG Pathway enrichment analysis, and the criterion for selection was all based on P<0.05.

1.4 Protein interaction and module analysis

STRINGdb was used to evaluate protein interaction information. For conditional screening of differentially interacting proteins, use plug-ins CytoHubba Screen key genes based on node degree. In addition, the application Cytoscape Software builds protein interaction network and uses plug-ins Molecular Complex Detection (MCODE) to analyze the hub genes in the protein interaction network. MCODE Score ＞ 3, Number of nodes>4 was used as criterion for hub genes.

2. RESULTS

2. 1 Data preprocessing

The RMA method is used to preprocess the data and eliminate the interference factors in the chip experiment. The results show that the median expression of the three expression profile chip samples is at the same level, as shown in Figure 1, suggesting that the samples have been normalized.

2.2 Differentially expressed gene screening results

The differentially expressed genes were screened by R language. There were 475 differentially expressed genes in GSE28735, including 174 up-regulated differentially expressed genes and 301 down regulated differentially expressed genes; There were 1072 differentially expressed genes in GSE15471, including 225 up-regulated differentially expressed genes, 847 down regulated differentially expressed genes; There were 1579 differentially expressed genes in GSE101448,

including 750 up-regulated differentially expressed genes and 829 down regulated differentially expressed genes. There were 161 differentially expressed genes in the three expression microarrays, including 54 up-regulated differentially expressed genes (Fig. 2a) and 107 down-regulated differentially expressed genes (Fig. 2b).

2.3 GO Functional enrichment analysis and KEGG Pathway enrichment analysis

The up-regulated differentially expressed genes and down-regulated differentially expressed genes were analyzed by David database and kobas online database KEGG pathway enrichment analysis and GO analysis ($P < 0.05$ and $FDR < 0.05$) showed that differentially expressed genes were enriched in extracellular exosome, extracellular matrix organization and extracellular space (Fig. 3a-b). KEGG results ($P < 0.05$ and $FDR < 0.05$) showed that the differentially expressed genes were mainly enriched in pancreatic secret, protein digestion and absorption and ECM receiver interaction (Fig. 3c-d).

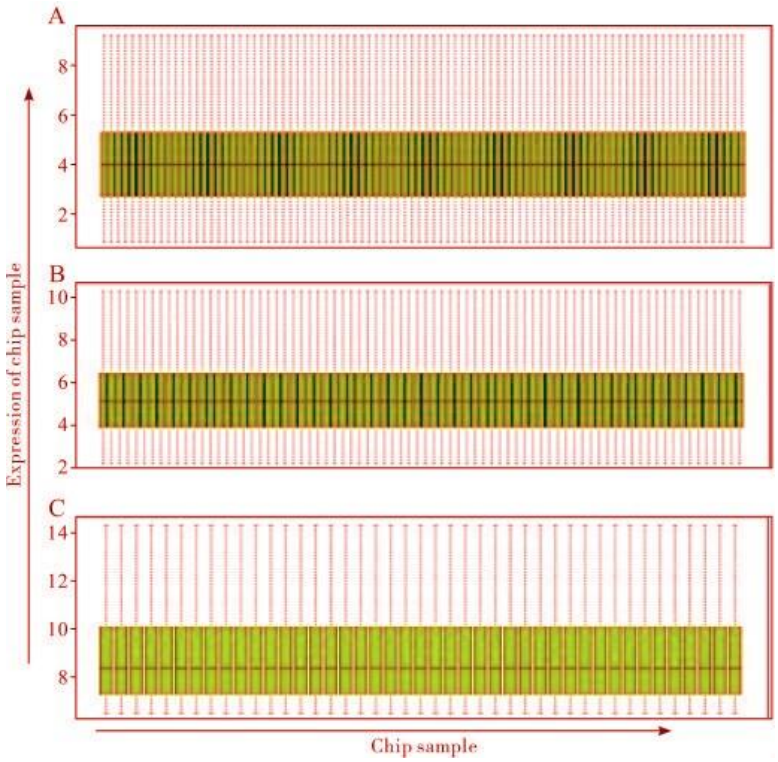

Fig. 1: Box plots of the expression spectrum data after normalization.

A: GSE28735. B: GSE15471. C: GSE101448. The thick line in the box represents median. All the thick lines are almost on the same line respectively, suggesting a good normalization.

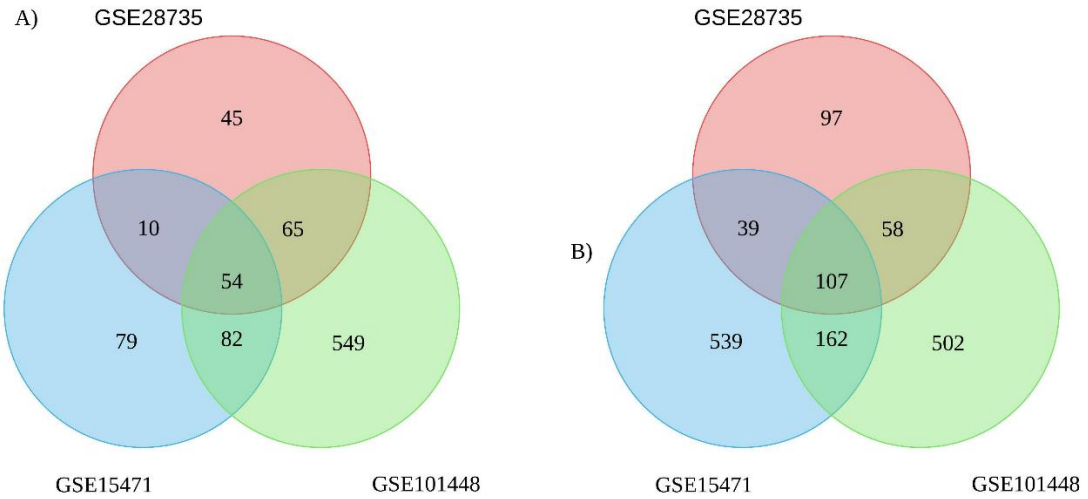

Fig. 2: Identification of DEGs.

A: Venn diagram of 54 overlapping up ─ regulated genes in GSE28735, GSE15471 and GSE101448.
B: Venn diagram of 107 overlapping down ─ regulated genes in GSE28735 , GSE15471 and GSE101448

2.4 PPI Network construction and module analysis

Online tools such as Stringdb were used to screen interaction scores greater than 0.4 which can be seen in protein construct protein-protein interaction network diagram (Figure 4). There are 111 nodes and 256 relationships that meet the conditions, of which the 10 genes with the highest node degree are ALB, COL11A1, COL3A1, FN1, EGF, COL1A1, MMP9, COL5A2, ITGA2, COL6A3. Cytoscape software plug-in MCODE, analyzed the whole protein interaction network and screened three sub modules, which were mainly enriched in protein digestion and absorption, ECM receiver interaction and focal adhesion (Fig. 5).

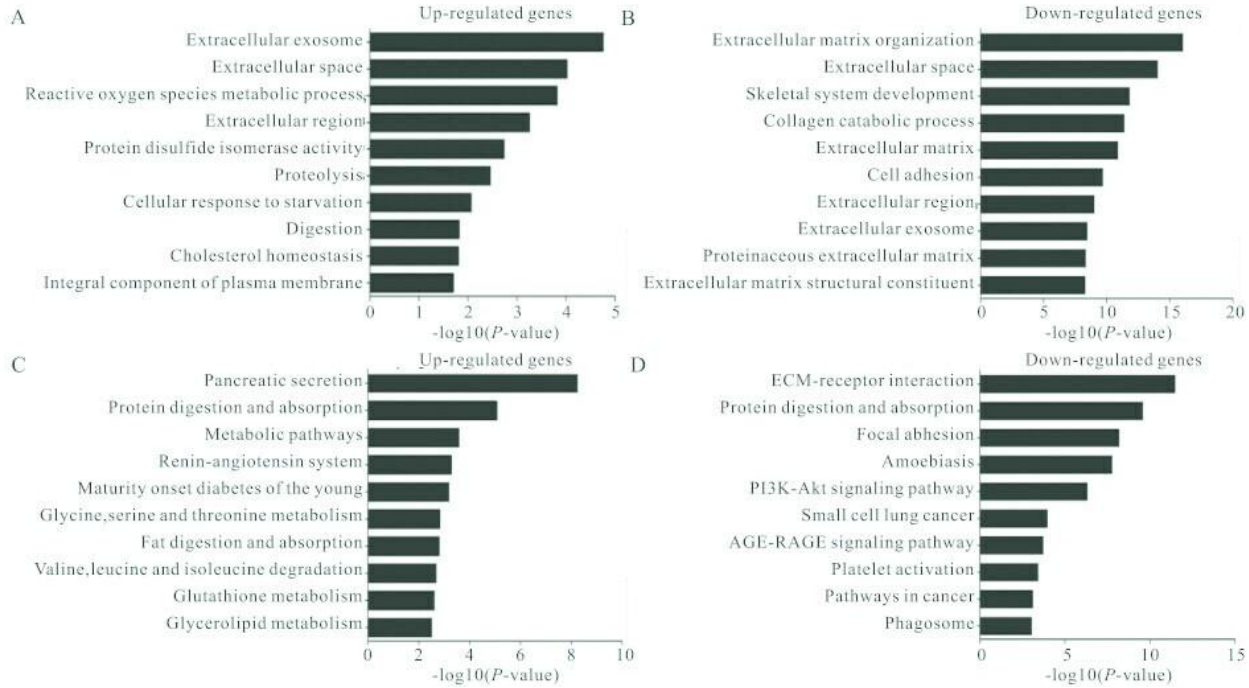

Fig. 3: GO analysis and KEGG pathway analysis of the overlapped DEGs.

A ─ B: GO analysis of the first ten up regulated DEGs and down regulated DEGs.

C ─ D: KEGG pathway analysis of the first ten up regulated DEGs and down regulated DEGs

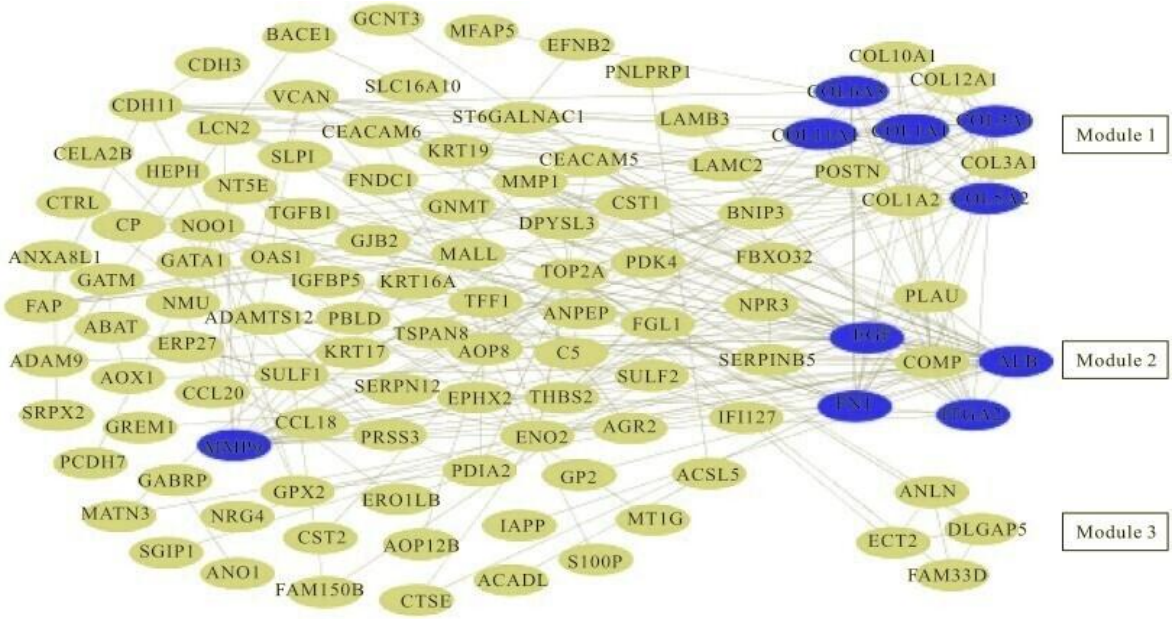

Fig. 4: PPI network construction and module analysis.

There are three modules were selected and the light blue nodes in the middle represent the DEGs, the red nodes represent the hub genes in the network, the lines represent the interaction between two nodes.

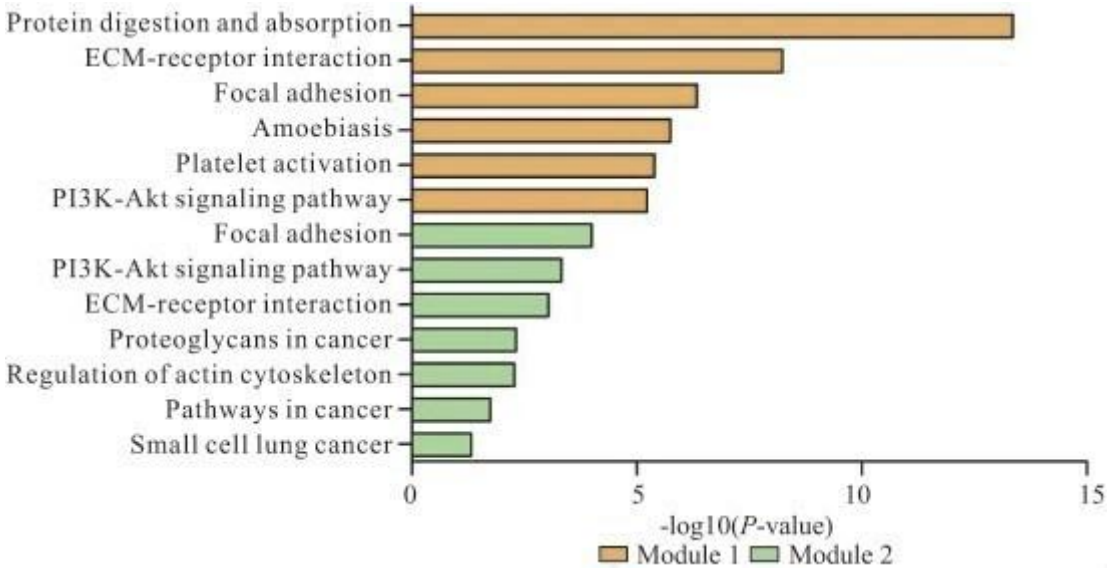

Fig. 5: KEGG pathway analysis of genes in modules

3. DISCUSSIONS

PDAC is a complex malignant tumor of digestive system. Surgery and chemotherapy were the main methods before treatment, but the effects were not ideal. Lack of screening and early specific biological target detection, coupled with the tendency of early metastasis and tolerance to systematic treatment, these are still the main obstacles to the treatment of PDAC patients. A few patients can accept surgical resection, but still face disease recurrence. The 5-year survival rate is only 12% to 27%.[8-10] Therefore, it is still urgent to study the early biological diagnostic markers of PDAC. In recent years, the rise of gene chip technology reveals that genetic changes in disease progression play an important role in tumor diagnosis, treatment and prognosis. In this paper, we analyzed the differentially expressed genes co expressed by multiple gene chips, excluded false positives to a certain extent, and provided a theoretical basis for studying the specific occurrence and development mechanism of PDAC. A total of 161 differentially expressed genes were screened, including 54 up-regulated differentially expressed genes and 107 down-regulated differentially expressed genes. Go functional enrichment analysis showed that the differentially expressed genes were closely related to extracellular exosome, extracellular matrix organization and extracellular space. Extracellular exosome is a vesicle structure with a diameter of 40 ~ 100 nm. It exists in almost all biological liquids and plays a role in the process of intercellular information transmission.[11] It has been found that in PDAC, exosomes such as glypican 1, miR-21 and erdj3 can manipulate the tumor microenvironment, promote immune escape to some extent, and mediate the formation, proliferation, invasion and metastasis of tumors.[12] Extracellular matrix organization mainly includes collagen, elastin, proteoglycan, etc. in the process of tumor development, it is often accompanied by extracellular matrix remodeling, which changes its interaction with cell surface receptors,[13] so as to accelerate the disease process. KEGG pathway analysis showed that the differentially expressed genes were mainly enriched in pancreatic secret, protein digestion and absorption and ECM receiver interaction. Pancreatic secretion is mainly affected by nerves and hormones. Sousa et al.[14] found that in PDAC, pancreatic stellate cells secrete alanine through autophagy, compete with glucose and glutamine for carbon atoms in TCA cycle, and finally synthesize non essential amino acids and lipids. This change in metabolic mode can accelerate the growth of tumor cells. The changes of ECM receptor interaction pathway are consistent with extracellular matrix remodeling, suggesting that PDAC is a process in which extracellular matrix is extremely involved.[15] The results of go functional enrichment analysis and KEGG pathway analysis suggest that the occurrence of PDAC is the result of multiple genes participating in multiple pathway changes.

In the protein-protein interaction network, we screened 10 genes with the highest number of nodes as key genes: FN1, COL6A3, ALB, COL3A1, EGF, MMP9, COl5A2, COL11A1, and ITGA2.

So far, no study has shown that the expression of albumin (ALB) is related to PDAC, but some studies have discussed the correlation between serum ALB and PDAC. Deng et al.[16] used nomogram to predict the prognosis of PDAC patients, and found that the decrease of serum ALB level suggested poor prognosis. Serum ALB is mainly affected by patients' nutritional status, liver function and other factors, and cannot reflect the expression of Alb in tissues. Therefore, further research is still needed. COL11A1, COL3A1, COL11A1, col5a2

and col6a3 all belong to the collagen gene family. The expression of COL11A1 in PDAC was significantly increased compared with normal tissues and chronic pancreatitis and the study also pointed out that COL11A1 may be a new marker for the diagnosis of PDAC.[17] When compared with adjacent tissues, COL6A3 protein level in PDAC was significantly up-regulated and related to tumor stage.[18]

FN1 is a glycoprotein, which is mainly involved in the process of cell adhesion and migration.
In nasopharyngeal carcinoma, FN1 promoted p65 nuclear entry and translocation by up regulating Bcl-2, and abnormally activated NF - κ B eventually leads to the proliferation and metastasis of tumor cells. It is speculated that FN1 mediated activation of the NF - κB/p65 pathway and inhibition of tumor cell apoptosis may be closely related to the occurrence of PDAC.[19]

A variety of growth factors and their receptors, notably the epidermal growth factor (EGF) family, have been found to play an essential role in pancreatic cancer. This ligand-receptor family is crucial in the development of pancreatic ductal carcinoma and contributes to its aggressiveness.[20] miR-21 and miR-211 synergistically up regulate the expression of MMP9 in PDAC and promote pancreatic epithelial mesenchymal transformation (EMT), which plays an important role in driving the growth and metastasis of tumor cells.[21] Itga2 is a member of the integrin family, like MMP9.

ITGA2 can also mediate EMT. In addition, ITGA2 demethylation in PDAC makes the upregulation of ITGA2 expression leads to poor prognosis of patients.[22]

Methylated drugs are of great significance to improve the survival rate of PDAC patients. In conclusion, we used bioinformatics methods to analyze PDAC gene expression.

The differentially expressed genes of microarray are comprehensively integrated and analyzed from functional enrichment and pathway enrichment. The selected 10 key genes may play an important role in the occurrence and development of PDAC, which is expected to become a biological target for PDAC diagnosis and treatment, and provide theoretical support and guidance for the further study of the molecular mechanism of PDAC occurrence and development.

REFERENCES

1. Carrato, A., et al., *A Systematic Review of the Burden of Pancreatic Cancer in Europe: Real-World Impact on Survival, Quality of Life and Costs.* J Gastrointest Cancer, 2015; 46(3): 201-11.
2. Dunne, R.F. and A.F. Hezel, *Genetics and Biology of Pancreatic Ductal Adenocarcinoma.* Hematol Oncol Clin North Am, 2015; 29(4): 595-608.
3. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer Statistics, 2017.* CA Cancer J Clin, 2017; 67(1): 7-30.
4. Akbani, R., et al., *A pan-cancer proteomic perspective on The Cancer Genome Atlas.* Nat Commun, 2014; 5: 3887.
5. Clough, E. and T. Barrett, *The Gene Expression Omnibus Database.* Methods Mol Biol, 2016; 1418: 93-110.
6. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015; 43(7): e47.
7. Bu, D., et al., *KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis.* Nucleic Acids Res, 2021; 49(W1): W317-W325.
8. Ferrone, C.R., et al., *Pancreatic ductal adenocarcinoma: long-term survival does not equal cure.* Surgery, 2012; 152(3 Suppl 1): S43-9.
9. He, J., et al., *2564 resected periampullary adenocarcinomas at a single institution: trends over three decades.* HPB (Oxford), 2014; 16(1): 83-90.
10. Katz, M.H., et al., *Long-term survival after multidisciplinary management of resected pancreatic adenocarcinoma.* Ann Surg Oncol, 2009; 16(4): 836-47.
11. Zhang, J., et al., *Exosome and exosomal microRNA: trafficking, sorting, and function.* Genomics Proteomics Bioinformatics, 2015; 13(1): 17-24.
12. Robinson, S.M., et al., *The role of exosomes in the pathogenesis of pancreatic ductal adenocarcinoma.* Int J Biochem Cell Biol, 2016; 75: 131-9.
13. Pickup, M.W., J.K. Mouw, and V.M. Weaver, *The extracellular matrix modulates the hallmarks of cancer.* EMBO Rep, 2014. 15(12): 1243-53.
14. Sousa, C.M., et al., *Erratum: Pancreatic stellate cells support tumour metabolism through autophagic alanine secretion.* Nature, 2016. 540(7631): 150.
15. Ansari, D., et al., *Pancreatic cancer stroma: controversies and current insights.* Scand J Gastroenterol, 2017; 52(6-7): 641-646.
16. Deng, Q.L., et al., *Development and Validation of a Nomogram for Predicting Survival in Patients with Advanced Pancreatic Ductal Adenocarcinoma.* Sci Rep, 2017; 7(1): 11524.
17. Garcia-Pravia, C., et al., *Overexpression of COL11A1 by cancer-associated fibroblasts: clinical relevance of a stromal marker in pancreatic cancer.* PLoS One, 2013; 8(10): e78327.
18. Arafat, H., et al., *Tumor-specific expression and alternative splicing of the COL6A3 gene in pancreatic cancer.* Surgery, 2011; 150(2): 306-15.
19. Wang, J., et al., *High expression of Fibronectin 1 suppresses apoptosis through the NF-kappaB pathway and is associated with migration in nasopharyngeal carcinoma.* Am J Transl Res, 2017; 9(10): 4502-4511.

20. Friess, H., et al., *Molecular aspects of pancreatic cancer and future perspectives.* Dig Surg, 1999; 16(4): 281-90.
21. Funel, N., *The role of miR-21 and miR-211 on MMP9 regulation in pancreatic ductal adenocarcinoma: cooperation in invasiveness behaviors?* Epigenomics, 2015; 7(3): 333-5.
22. Nones, K., et al., *Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signaling.* Int J Cancer, 2014; 135(5): 1110-8.