Open Access Full Text Article

ORIGINAL RESEARCH

Exploration and Validation of Pancreatic Cancer Hub Genes Based on Weighted Gene Co-Expression Network Analysis and Immune Infiltration Score Analysis

Xiao-Xi Li[1,]*, Hong Li[2,]*, Li-Quan Jin[3], Yun-Bo Tan[1,3]

[1]Dali University of Clinical Medicine School, Dali, Yunnan, 671000, People's Republic of China; [2]Department of Radiology, Affiliated Renhe Hospital of China Three Gorges University, Hubei, 443001, People's Republic of China; [3]Department of General Surgery, The First of Affiliated Hospital of Dali University, Dali, Yunnan, 671000, People's Republic of China

*These authors contributed equally to this work

Correspondence: Yun-Bo Tan, Department of General Surgery, the First of Affiliated Hospital of Dali University, No. 32 of Carlsberg Road, Dali, Yunnan, 671000, People's Republic of China, Tel +86 15517162885, Email tanyunbo6365@163.com

Objective: To find pancreatic cancer (PC)-related hub genes based on weighted gene co-expression network analysis (WGCNA) construction and immune infiltration score analysis and validate them immunohistochemically by clinical cases, to generate new concepts or therapeutic targets for the early diagnosis and treatment of PC.

Material and Methods: In this study, WGCNA and immune infiltration score were utilized to identify the relevant core modules of PC and the hub genes within these core modules.

Results: Using WGCNA analysis, data from PC and normal pancreas integrated with TCGA and GTEX were analyzed and brown modules were chosen from the six modules. Five hub genes, including DPYD, FXYD6, MAP6, FAM110B, and ANK2, were discovered to have differential survival significance via validation tests utilizing survival analysis curves and the GEPIA database. The DPYD gene was the only gene associated with PC survival side effects. Validation of the Human Protein Atlas (HPA) database and immunohistochemical testing of clinical samples showed positive results for DPYD expression in PC.

Conclusion: In this study, we identified DPYD, FXYD6, MAP6, FAM110B, and ANK2, as immune-related candidate markers for PC. Only the DPYD gene had a negative impact on the survival of PC patients. Through validation of the HPA database and immunohistochemical testing of clinical cases, we believe that the DPYD gene brings novel ideas and therapeutic targets in the diagnosis and treatment of PC.

Keywords: DPYD, pancreatic cancer, tumor immunity, WGCNA

Introduction

Pancreatic cancer (PC) is a malignant tumor with a very high mortality rate for which there is no effective treatment other than surgery.[1,2] PC is prone to metastasis and invasion in the early stage, and there is no effective means for early detection. Most patients often have cancer metastasis at the time of diagnosis and lose the opportunity for surgical resection.[3] For more than 40 years, with the advancement of technology and increasing research on pancreatic cancer, some treatments such as targeted therapy, neoadjuvant therapy, and translational therapy have gradually emerged in people's vision, but their therapeutic effects on pancreatic cancer are often limited, and the survival rate of pancreatic cancer patients remains very low.[4] The unique microenvironment of PC promotes cancer cell proliferation and immune escape. Two main features characterize the immune microenvironment of PC: the dense fibrous tissue proliferation and extensive immunosuppression. Both of these features can promote immune surveillance escape of PC cells and lead to their proliferation by directly suppressing anti-tumor immunity.[5] Over the last decade, substantial progression in

molecular subtyping for PDAC has facilitated the understanding of molecular pathogenesis and provided clues for advanced therapy designing.[6–11] However, to date, there is still lacking an immune feature- based molecular subtyping for better understand the molecular mechanism of pancreatic cancer invasion and progression in order to facilitate early diagnosis and effective treatment. PC treatment should emphasize on early diagnosis, prevention of metastasis, and prevention of immune evasion. Moreover, the current understanding of PC is incomplete; consequently, it is essential to comprehend the molecular mechanisms of PC and to find novel prognostic biomarkers and potential therapeutic targets associated with immunity in PC.

Weighted gene co-expression network analysis (WGCNA) is a promising emerging bioinformatics analysis method.[12] It is primarily a systems biology approach to describe gene association patterns between different samples, to identify highly synergistic sets of genes, and ultimately screen essential hub genes for a specific trait as alternate biomarker genes or therapeutic targets based on the inline nature of the gene set and its association with clinical performance.

In this paper, we screened for immune signature-related hub genes in PC using the WGCNA method to select target genes from among them. This approach provides potential biomarkers and new therapeutic targets for PC treatment.

Materials and Methods
Data Source and Pre-Processing

The patient RNA sequencing and clinical information data were downloaded from the PC cohort in the TCGA database (https://www.cancer.gov/aboutnci/organization/ccg/research/structural-genomics/tcga) on July 7, 2022. The expression data was 182. The clinical data was 196. At the same time, the RNA sequencing data and clinical information of normal pancreatic tissues were downloaded from the GTEX database (https://www.g texportal.org/home/), with 167 gene expression data and 204 clinical data. Combining the two data using R language, the PC tissues that could not be distinguished between the two were then screened out from normal tissues using principal component analysis (PCA). We used PCA, a multifactorial statistical method, to reduce the dimensionality of the data. The DPYD gene immunohistochemical test results for the clinical case were obtained from the First Affiliated Hospital of Dali University, Yunnan, China.

The specific workflow is as follows (Figure 1).

Identification and Enrichment Analysis of Differentially Expressed Genes in PC

Limma[13] (linear models for microarray) data is a differential expression screening method based on generalized linear models. In this study, we utilized the R package "Limma (version 3.40.6)" to differentially analyze the differentially expressed genes (DEG) of the combined gene expression profiles[14] to identify the differentially up-regulated and down-regulated genes between normal pancreatic tissues and PC tissues. We analyzed the functional enrichment of differentially expressed genes (DEGs) in R using Gene Ontology (GO) for up-regulated and down-regulated genes.

WGCNA Co-Expression Network Construction and Co-Expression Module

A total of 22,515 differential genes were evaluated and a gene co-expression network was constructed using the "WGCNA" R language package for these differential genes. Co-expression modules are groups of genes that share a high degree of topological overlap similarity and genes within the same module typically co-express to a greater extent. We constructed a hierarchical clustering tree by clustering genes according to a dissimilarity matrix representing gene linkage, and used a dynamic tree-cutting algorithm to identify gene co-expression modules. The goodSamplesGenes method of the R package WGCNA[15,16] was used to remove outlier genes and samples. A scale-free co-expression network was subsequently constructed using WGCNA to obtain six co-expression modules, with the gray module representing the set of genes that could not be assigned to a module.

Tumor Infiltration Scoring by MCPCounter Method and Correlation with Modules to Identify Hub Genes

Immune scores were assigned to CD3+ T cells, CD8+ T cells, cytotoxic lymphocytes, NK cells, B lymphocytes, monocytes, bone marrow dendritic cells, neutrophils and endothelial cells, and fibroblasts in each of the six co-
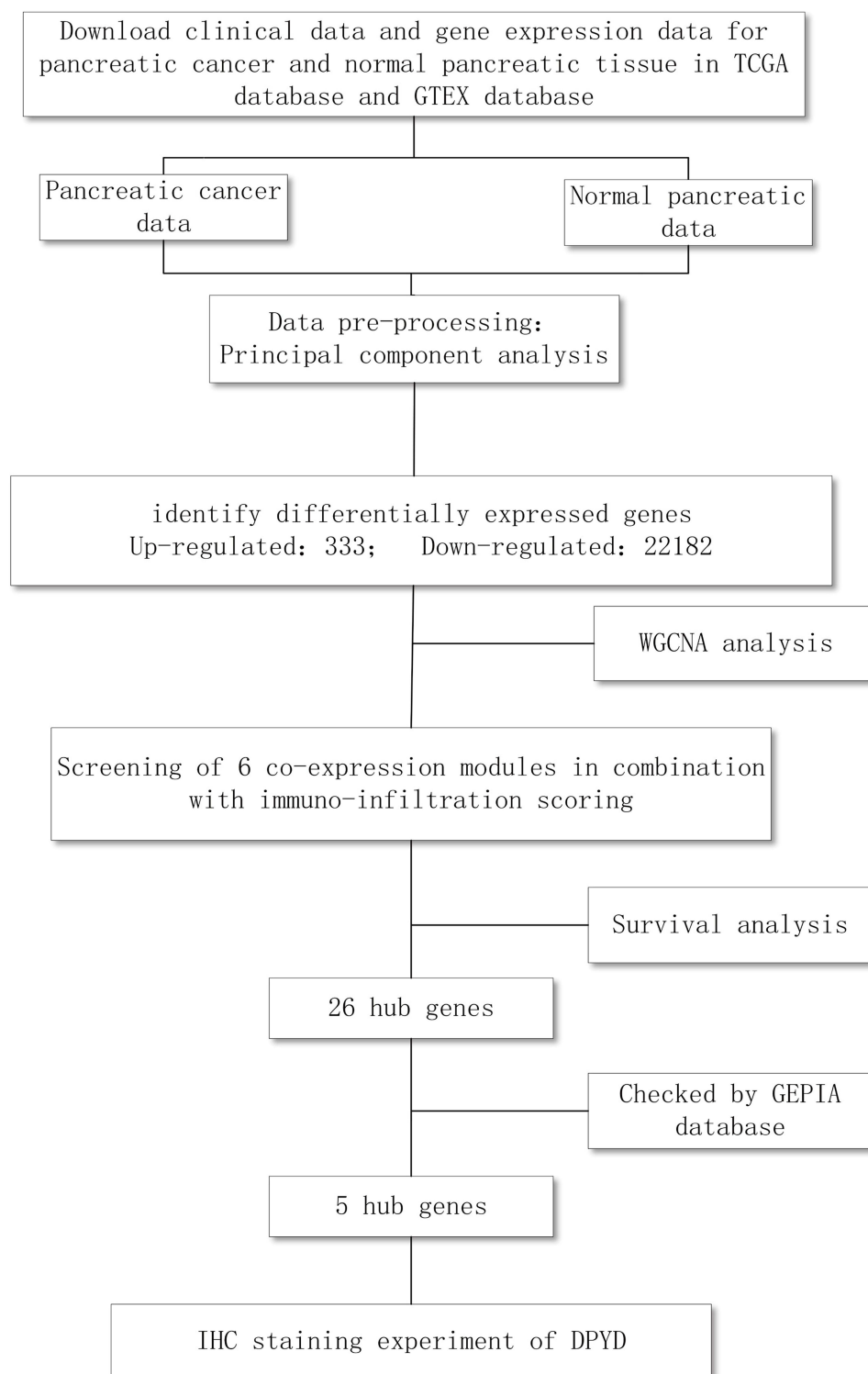
Download clinical data and gene expression data for pancreatic cancer and normal pancreatic tissue in TCGA database and GTEX database

Pancreatic cancer data

Normal pancreatic data

Data pre-processing: Principal component analysis

identify differentially expressed genes
Up-regulated: 333;   Down-regulated: 22182

WGCNA analysis

Screening of 6 co-expression modules in combination with immuno-infiltration scoring

Survival analysis

26 hub genes

Checked by GEPIA database

5 hub genes

IHC staining experiment of DPYD

Figure 1 Workflow for searching hub genes in PC.

expression modules. Using the R language package immune-oncology biological research (IOBR),[17] this score was used to quantify the degree of immune infiltration in PC and its response to therapy. This score quantifies the degree of immune infiltration in PC and represents the immune properties of the tumor.[18] The immune score was used as the

immune signature of PC to correlate with the modules screened. The hub genes with the highest association and variability with the immune signature and modules were identified.

Identification of Hub Genes

Gene Module Membership (MM) indicates the relationship between genes and modules. GeneTrait Significance (GS) refers to the relationship between the clinical characteristics or immune information of the gene and the gene itself. The criteria for hub genes in the immune-related gene module network are "cor.geneTraitSignificance" > 0.6 and "cor.gene ModuleMembership" > 0.8. Screening and filtering under conditions satisfying this criterion yields hub genes with high relevance for the desired immune profile.[15]

Survival Analysis of Final Hub Genes

Using the R package SURVIVAL, Kaplan-Meier survival curves and Log rank tests for differences in overall survival between high and low-expression groups of hub genes were utilized to identify hub genes having prognostic importance for survival. Log < 0.01 and $P < 0.05$ were judged statistically significant.

Bioinformatic Validation of Hub Genes

The GEPIA online database (http://gepia.cancer-pku.cn/) is a website that incorporates data from the TCGA and GTEX databases, and has interactive and customizable features.[19] The screened hub genes were then uploaded to the GEPIA website so that their expression and survival curves could be analyzed. The verification of the dependability of the hub genes was investigated in this study.

Human Protein Atlas Database for Examination

We verified the hub genes that were selected at the protein level. The Human Protein Atlas (HPA) database employs transcriptomics and proteomics technologies to examine protein expression levels at the RNA and protein levels in various human tissues and organs.[20] It has nearly 26,000 antibodies, and all results have immunohistochemical (IHC) staining in various tissues. In normal pancreatic and PC tissue, we found IHC markers for the expressed proteins of hub genes in HPA and downloaded them for comparison and localization.

Immunohistochemical Validation of the Hub Gene DPYD Gene

In this study, IHC staining was applied to the cancer tissues of three PC patients from the First Affiliated Hospital of Dali University.[21,22] The cancer tissues were formalin-fixed and embedded in paraffin wax and then three different PC tissue sections of the patients were taken. The tissue sections were sequentially treated with alcohol in the order of xylene, 100%X2, 95%, 90%, 80%, and 70%. Tissue antigen was repaired by boiling with sodium citrate buffer and adding endogenous peroxidase inhibitors. It was then blocked in 5% goat serum for 1 hour to prevent non-specific binding. Subsequently, anti-DPYD antibodies (dilution ratio: 1:300; Proteintech:27662-1-AP) were dropped onto the tissue sections separately and incubated overnight at 4 °C. Secondary antibodies were then dropped and incubated. Then, the sections were blocked and photographed after adding diaminobenzidine for color development and then hematoxylin re-staining. Finally, the staining results were evaluated by experienced researchers.

Results

Principal Component Analysis Extraction of Data from the TCGA and GTEX Databases—344 Samples Met the Inclusion Requirements

There were 182 samples from the TCGA database with 4 normal or para-cancerous tissue samples and 178 (182–4) PC samples. Then, 167 normal pancreatic tissue samples from the TCGA and GTEX database and 4 normal tissue samples were enrolled, for a total of 171 (167+4) samples (Figure 2A). Five normal pancreatic tissue samples were excluded after filtering by PCA, which can better distinguish tumor samples from normal ones. The first component, represented in the
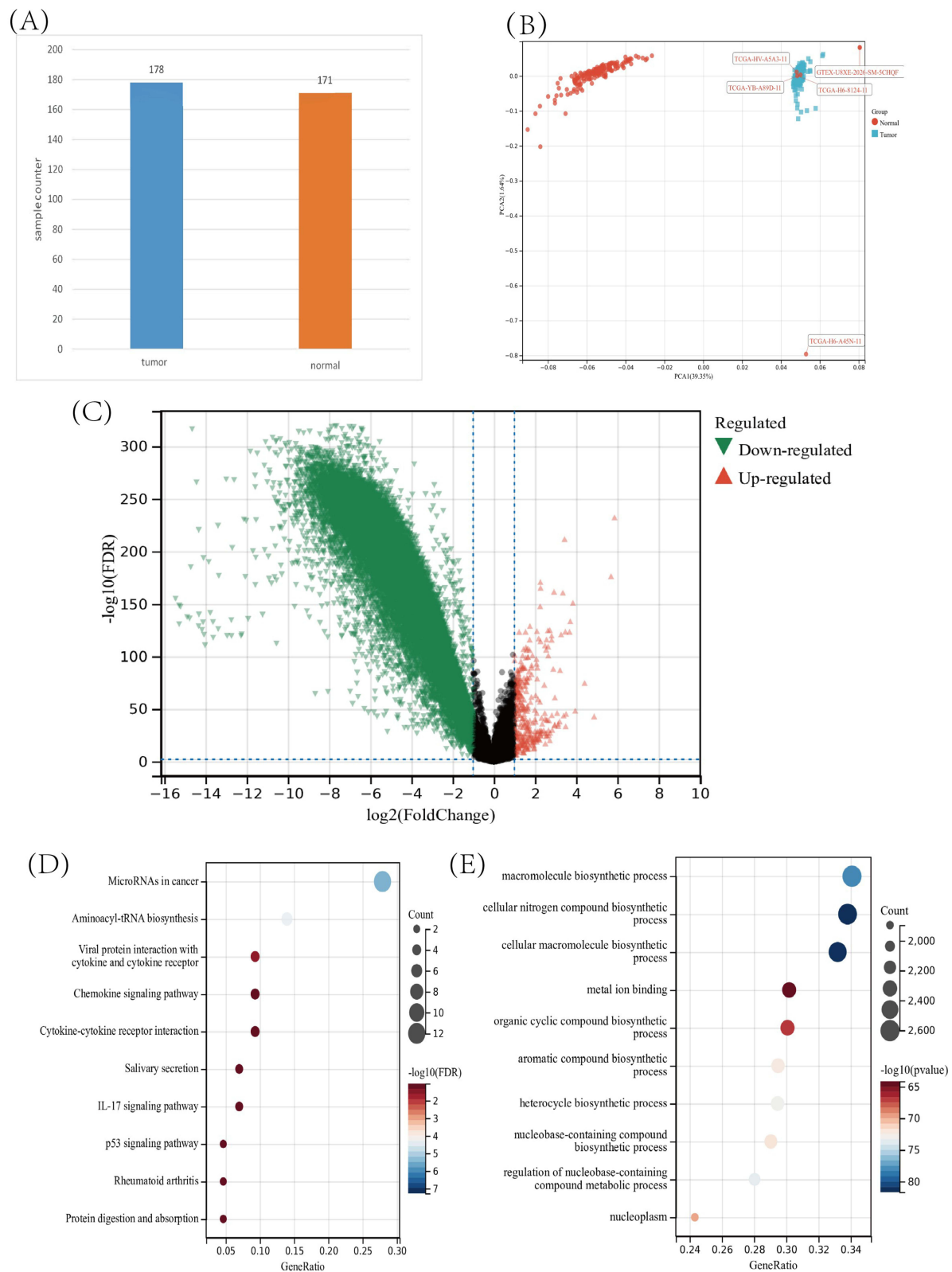
Figure 2 (**A**) Data of normal pancreas and PC. (**B**) Principal Component Analysis. (**C**) Volcano plot. Green dots represent genes down-regulated in PC; red dots represent genes up-regulated in PC. Black dots represent genes not differing significantly between PC and normal tissues. (**D**) Bubble chart: GO analysis of functional enrichment of up-regulated genes. The dot size reflects the number of genes enriched under the given ontology term, and the color indicates the significance of enrichment. (**E**) Bubble chart: GO analysis of functional enrichment of down-regulated genes.

red color was significantly distinguished from the second component, represented in the blue color (Figure 2B), and further analysis and screening was conducted on the remaining 344 (178+171-5) samples.

Differential Genes were Found by DEG Analysis of PC and GO Enrichment Analysis was Performed to Comprehend the Biology of the Differential Genes

Using the "Limma (version 3.40.6)" function of the R program, samples of 173 normal pancreatic tissues and 171 PC tissues were detected, and DEGs were defined as those with |log2 (fold change) | > 1 and $P < 0.01$. In all, 22,515 differential genes were identified, with 333 up-regulated and 22,182 down-regulated genes (Figure 2C). It is assumed that the abnormally high number of down-regulated genes is due to the integration of the normal pancreas gene expression profile into the GTEX data. At the same time, DEGs were also analyzed by GO enrichment to understand their potential biological functions. As can be seen from Figure 2D and E, the up-regulated DEGs in cancer are mainly involved in small molecule ribonucleic acid, aminoacyl tRNA biosynthesis, interactions between viral proteins and cytokines and cytokine receptors, chemokine signaling pathway, IL-17 signaling pathway and P53 signaling pathway (Figure 2D). Most down-regulated DEGs were involved in the macromolecular synthesis, cytosolic nitrogenation biosynthesis, metal ion binding, and biosynthesis of base-containing chemicals (Figure 2E). The biofunctional annotations of these genes with significant alterations, such as the chemokine signaling route and the IL-17 signaling pathway, are comparable with some of the recognized dysfunctions of PC. On the other hand, this demonstrates the excellent reliability of our findings.

Calculation of the PC Immune Cell Infiltration Score Using MCPcounter

By running the MCP-counter algorithm, we brought 178 PC samples with gene expression information into the R package, thus quantifying the transcriptome data of the samples into an absolute abundance of 8 immune cells and 2 stromal cell populations. This algorithm has been used in several gene expression studies to quantify the abundance of immune cells in different samples.[23,24] We used the scores of individual immune cells of PC calculated by MCPcounter as the immune signature of PC.

Expression Matrix of WGCNA and Co-Expression Network of the Immune Infiltrating Cell Score of MCPcounter Validated the Screening of 6 Modules

WGCNA was used to construct an expression matrix of 22,515 DEGs and a co-expression network from the immune infiltrating cell score data calculated by MCPcount. We clustered the samples of DEGs using Pearson correlation coefficients, and after removing outliers, drew a clustering tree of the samples. First, we calculated the soft threshold power and raised the co-expression similarity to calculate the neighborhood. We used the pickSoftThreshold function in WGCNA to analyze its network topology. The figure shows that the average connectivity is best when the plotted scalar independence reaches 0.86 and the soft threshold power is set to 8. Therefore, the soft threshold β is set to 8 (Figure 3A and B).

After establishing a soft threshold value of 8, the created gene network conforms to the scale-free distribution, allowing the hub modules and genes to be searched. We used the WGCNA R program to create the gene network further and identify the modules. We set the minimum module size for its clustering analysis to 30 and the sensitivity to 3, and after clustering DEGs with similar expression patterns into the same modules, modules set to show a difference in cut height < 0.25 were merged (Figure 3C). We built the connection matrix and the topological overlap matrix as an extension. Six modules were discovered based on the average hierarchical clustering and dynamic tree cropping, whose distribution spans from black, blue, brown, gray, red, and yellow modules (Figure 3D and E). Among them, the gray modules were genes that could not be integrated into other modules.

The modules screened in this study were correlated with the immune features represented by the immune infiltration score of PC calculated by the MCPcounter. This analysis showed (Figure 4A) that the brown modules showed a high correlation with the immune features (cor=0.89, p=2.2x10-105) and the immune features represented by T cell and Monocytic-lineage. The highest correlation was found with the immune features represented by T cell and Monocytic-lineage. The findings clearly show that genes within the brown module contribute to pancreatic carcinogenesis and development and are closely associated with immunity.
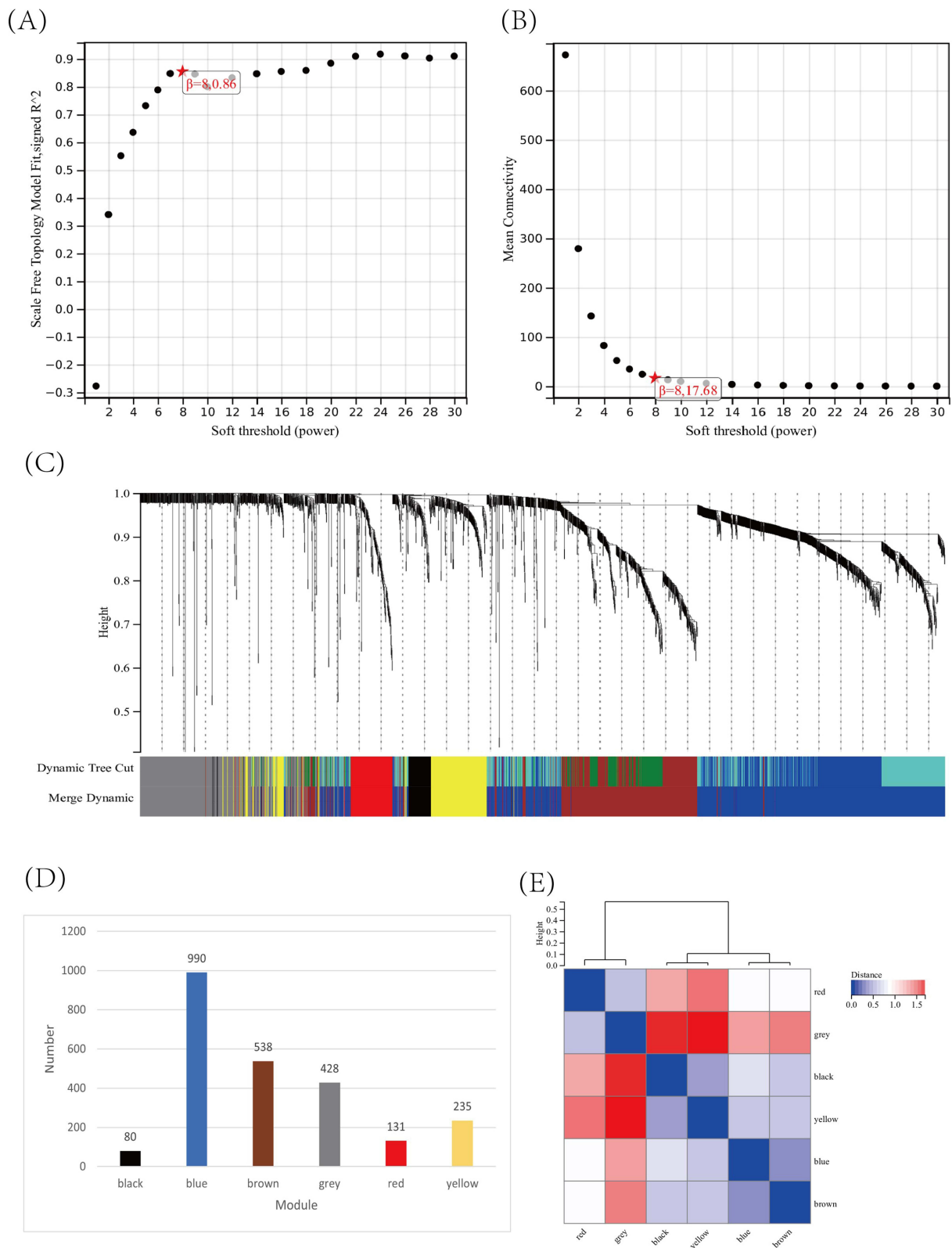
(A)

(B)

(C)

(D)

(E)

Figure 3 (**A**) Analysis of the scale-free index for various soft-threshold powers (β). (**B**) Analysis of the mean connectivity for various soft-threshold powers. (**C**) Dendrogram of 22,515 DEGs depending on the dissimilarity measure 1-TOM (see 'Material and methods section). Each branch represents a gene, and each color represents a co-expression module. (**D**) Numbers of genes in the six modules; Module heatmap of eigengene adjacency. (**E**) Heatmap of eigengene adjacency.
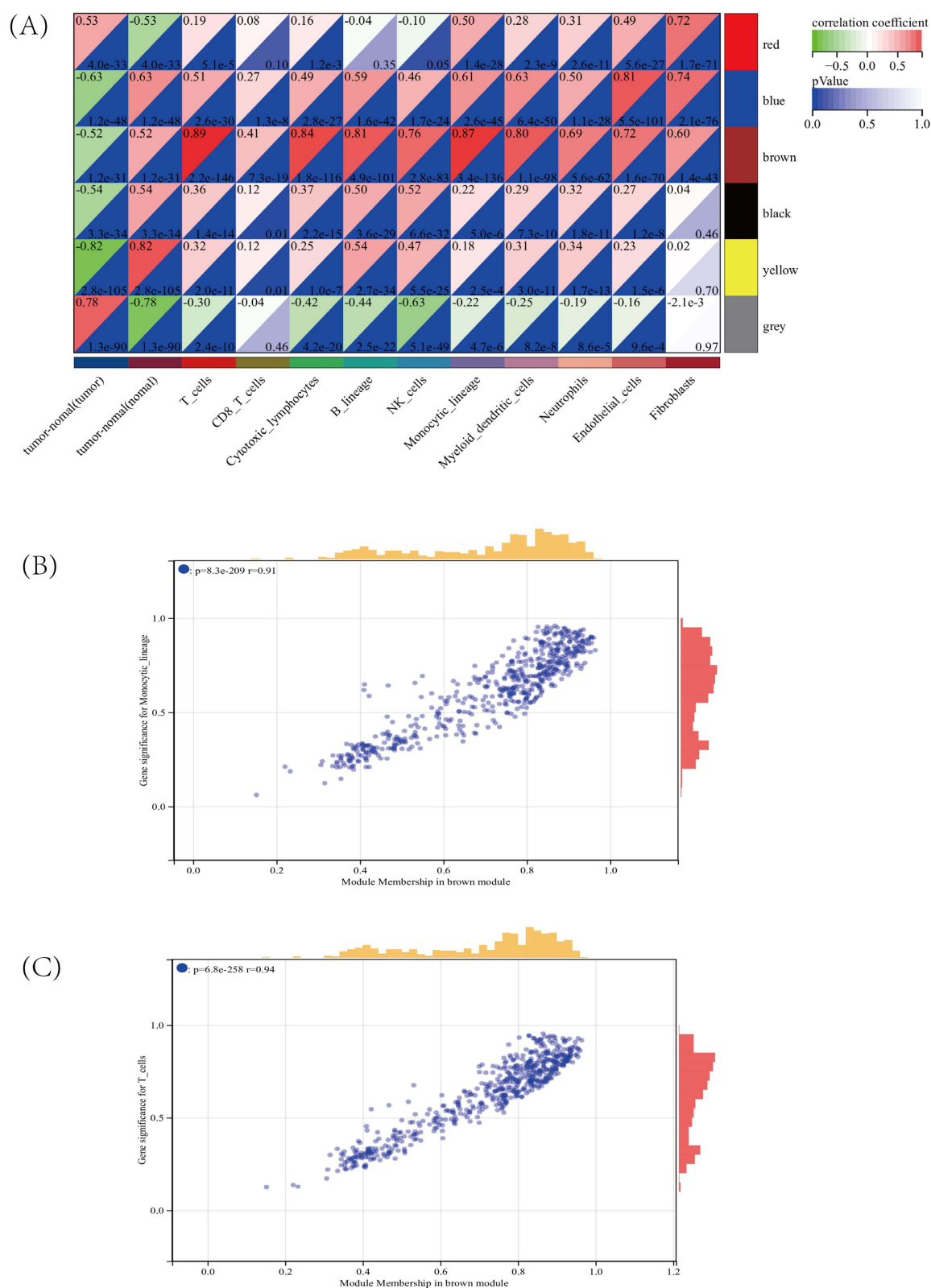
Figure 4 (**A**) Heat map of modules associated with immune features. (**B** and **C**) Correlation coefficients between MM and GS in T cells and monocyte lines.

Identification of Candidate Central Genes in the Brown Module

The analysis shows that MM and GS of the two modules representing T cell and Monocytic-lineage in the brown module are strongly positively correlated with each other, with correlation coefficients of r=0.94 and r=0.91, respectively (Figure 4B and C).

Under the condition that "cor.geneModuleMembership" > 0.8 threshold and satisfying "cor.geneTraitSignificance" > 0.6 thresholds, we screened 330 hub genes under the more strongly correlated T cell module.

Three Hundred and Thirty Hub Genes Were Screened for Survival Analysis to Identify 26 Hub Genes with Prognostic Relevance for Survival, Which Were Then Validated by the GEPIA Website

Based on the expression data and clinical information of PC samples in TCGA, we examined the potential association between the expression of 330 genes screened in the brown module and patient survival. After univariate Cox proportional risk regression analysis of the 330 genes by the survival package in R, 26 genes with prognostic relevance for survival were screened (log < 0.01, P < 0.05) (Figure 5A). For the 26 screened genes, validation using the GEPIA website resulted in survival differences for only 5 genes, including DPYD, FXYD6, MAP6, FAM110B, and ANK2. We defined these 5 genes as the final hub genes. As shown by the survival analysis curves (Figure 5B), only the DPYD gene has a negative impact on the survival of PC patients. That is, DPYD may be able to stimulate the proliferation and differentiation of PC cells, hence shortening patient survival time.

DPYD Gene is Highly Expressed in PC Cells as Verified by the HPA Database

To get additional insight into the protein expression level of the hub gene DPYD, we compared the IHC profiles of PC and normal pancreatic tissues from the HPA database. The comparison revealed that DPYD was substantially expressed in the cytoplasm of PC cells, but it was nearly absent in normal pancreatic tissue cells (Figure 6). Therefore, we propose that the DPYD gene has the potential to become a novel PC-specific marker.

Heterogeneous Expression of DPYD Gene in Cancer Cells Verified by IHC in Clinical PC Cases

We took pathological sections from surgically confirmed PC patients from the First Affiliated Hospital of Dali University in Yunnan, China. The heterogeneous expression of the DPYD gene in PC cells was detected clinically by IHC analysis in all cases (Figure 7), thus further validating the clinical application of the DPYD gene as an immune-related hub gene. We did not use normal pancreatic tissues for comparison due to the harsh conditions of obtaining normal pancreatic tissues in the clinic, which is a shortcoming of this study.

Discussion

The mortality rate of PC has been high and the cure rate and prognosis are unsatisfactory. The main reason is that PC is prone to metastasis and immune escape due to its unique immune microenvironment. In addition, early PC metastasis is associated with Epithelial-Mesenchymal Transition (EMT). The immune infiltrating cells within the EMT play a key role.[25,26] There is an inextricable link between immunity and PC. The molecular mechanisms involved in the immune microenvironment of PC are still unclear, so it is crucial to explore the genes associated with immunity in patients with PC.

Potentially Significant Differences in the Biology of Up-Regulated DEG in PC as Understood by GO Enrichment Analysis

According to GO enrichment analysis, the 333 up-regulated DEGs in our study were connected to small molecule RNAs, amyl tRNA production, viral protein-cytokine interactions, chemokine signaling pathways, IL-17 signaling pathways, and P53 signaling pathways in cancer. Pancreatic cancer is also linked to a chemokine signaling pathway with substantial heterogeneity in enrichment analysis. Chemokines are members of a family of tiny molecules of cytokines whose primary
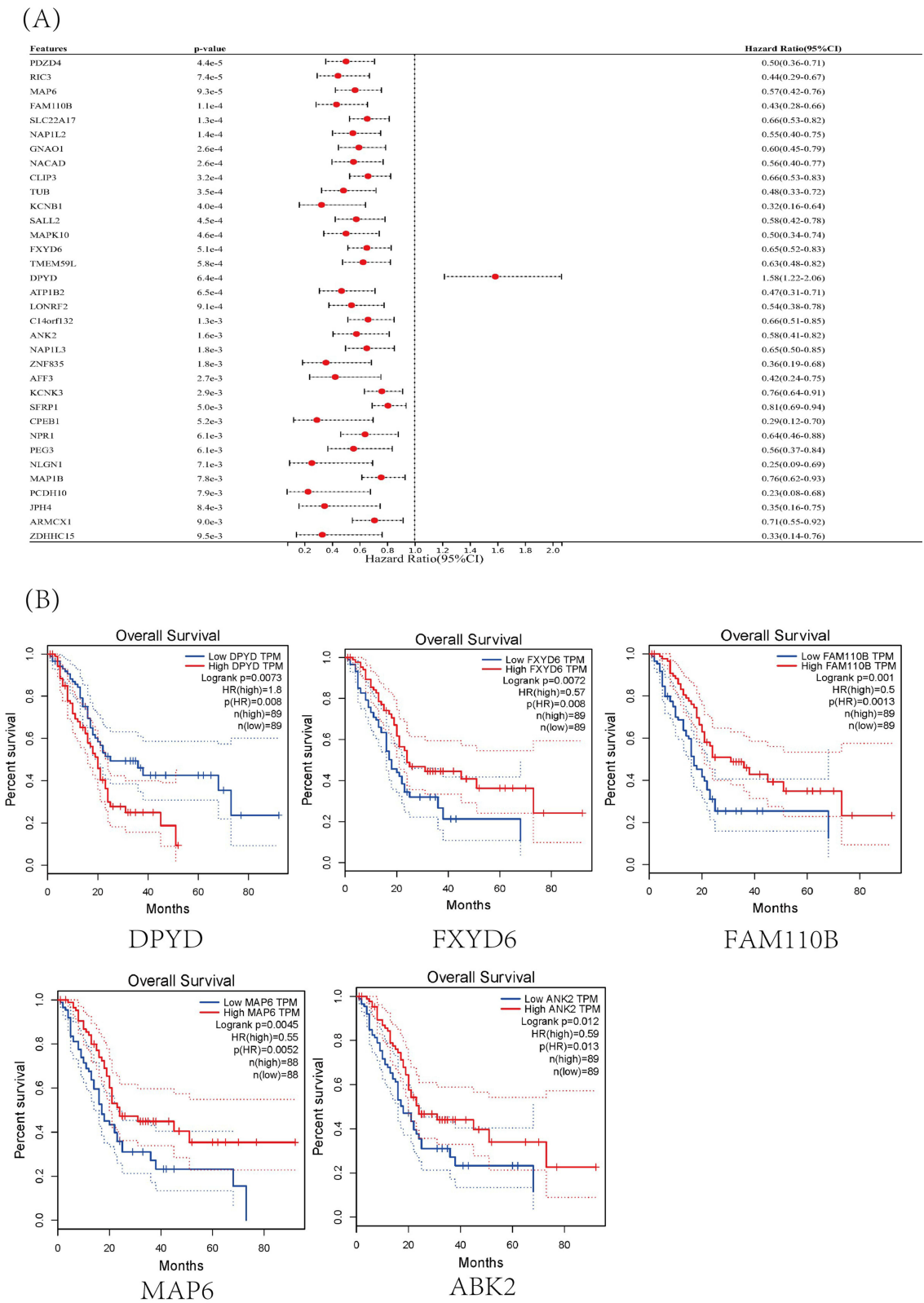
(A)

Features	p-value		Hazard Ratio(95%CI)
PDZD4	4.4e-5		0.50(0.36-0.71)
RIC3	7.4e-5		0.44(0.29-0.67)
MAP6	9.3e-5		0.57(0.42-0.76)
FAM110B	1.1e-4		0.43(0.28-0.66)
SLC22A17	1.3e-4		0.66(0.53-0.82)
NAP1L2	1.4e-4		0.55(0.40-0.75)
GNAO1	2.6e-4		0.60(0.45-0.79)
NACAD	2.6e-4		0.56(0.40-0.77)
CLIP3	3.2e-4		0.66(0.53-0.83)
TUB	3.5e-4		0.48(0.33-0.72)
KCNB1	4.0e-4		0.32(0.16-0.64)
SALL2	4.5e-4		0.58(0.42-0.78)
MAPK10	4.6e-4		0.50(0.34-0.74)
FXYD6	5.1e-4		0.65(0.52-0.83)
TMEM59L	5.8e-4		0.63(0.48-0.82)
DPYD	6.4e-4		1.58(1.22-2.06)
ATP1B2	6.5e-4		0.47(0.31-0.71)
LONRF2	9.1e-4		0.54(0.38-0.78)
C14orf132	1.3e-3		0.66(0.51-0.85)
ANK2	1.6e-3		0.58(0.41-0.82)
NAP1L3	1.8e-3		0.65(0.50-0.85)
ZNF835	1.8e-3		0.36(0.19-0.68)
AFF3	2.7e-3		0.42(0.24-0.75)
KCNK3	2.9e-3		0.76(0.64-0.91)
SFRP1	5.0e-3		0.81(0.69-0.94)
CPEB1	5.2e-3		0.29(0.12-0.70)
NPR1	6.1e-3		0.64(0.46-0.88)
PEG3	6.1e-3		0.56(0.37-0.84)
NLGN1	7.1e-3		0.25(0.09-0.69)
MAP1B	7.8e-3		0.76(0.62-0.93)
PCDH10	7.9e-3		0.23(0.08-0.68)
JPH4	8.4e-3		0.35(0.16-0.75)
ARMCX1	9.0e-3		0.71(0.55-0.92)
ZDHHC15	9.5e-3		0.33(0.14-0.76)

Hazard Ratio(95%CI)

(B)

DPYD

Overall Survival
Low DPYD TPM
High DPYD TPM
Logrank p=0.0073
HR(high)=1.8
p(HR)=0.008
n(high)=89
n(low)=89

FXYD6

Overall Survival
Low FXYD6 TPM
High FXYD6 TPM
Logrank p=0.0072
HR(high)=0.57
p(HR)=0.008
n(high)=89
n(low)=89

FAM110B

Overall Survival
Low FAM110B TPM
High FAM110B TPM
Logrank p=0.001
HR(high)=0.5
p(HR)=0.0013
n(high)=89
n(low)=89

MAP6

Overall Survival
Low MAP6 TPM
High MAP6 TPM
Logrank p=0.0045
HR(high)=0.55
p(HR)=0.0052
n(high)=88
n(low)=88

ABK2

Overall Survival
Low ANK2 TPM
High ANK2 TPM
Logrank p=0.012
HR(high)=0.59
p(HR)=0.013
n(high)=89
n(low)=89

Figure 5 (**A**) Survival forest map of 26 hub genes. (**B**) Kaplan-Meier survival curves of PC patients stratified by low or high expression of the five hub genes.
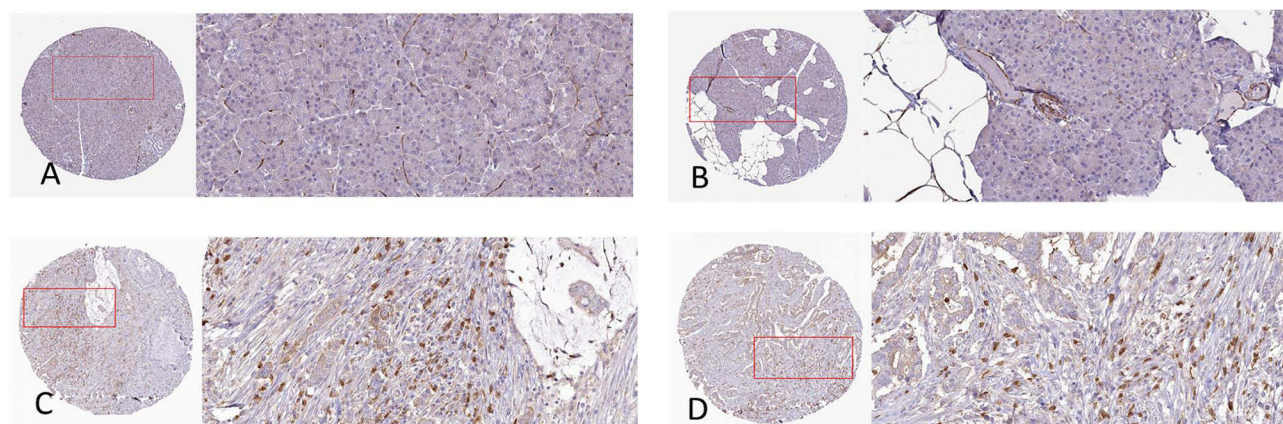
Figure 6 (**A** and **B**) indicates the expression of DPYD in normal pancreatic tissues; (**C** and **D**) indicates the expression of DPYD in PC tissues; DPYD in (**C** and **D**) was highly expressed, while DPYD in (**A** and **B**) was almost not expressed.
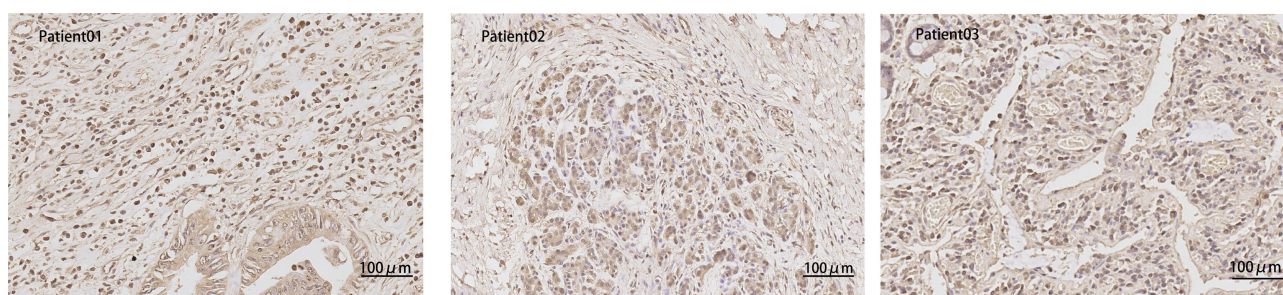

Figure 7 Differential expression of DPYD in PC tissues.

function is to draw immune cells to the site of inflammation. Their downstream signaling pathways, JAK/STAT, Ras, ERK, and Akt, are also linked to PC, encouraging cancer proliferation, differentiation, and metastasis,[27–30] and primarily activate these pathways.[31] According to a recent study, the inflammatory chemokine CXCL3 stimulates the spread of PC via a brand-new cancer escape mechanism hijacked by myofibroblasts.[32] Additionally, research has revealed that the IL-17 and p53 pathways are involved in the initiation, progression, and prognosis of PC and that the action of IL-17 promotes PC metastasis and invasion as well as predicts PC prognosis and gemcitabine therapy effectiveness.[33,34] The high P53 mutation level in the pancreas also demonstrates that our projections generally match the facts.[35]

Identification of Immune and Survival-Related Hub Genes in PC Based on WGCNA Construction and Immune Infiltration Score Analysis

By WGCNA construction, we identified five crucial genes—DPYD, FXYD6, MAP6, FAM110B, and ANK2—linked to immunity and survival. In prior investigations and research, FXYD6, MAP6, FAM110B, and ANK2 were not significantly related to PC. FXYD6 is significantly altered in hepatocellular carcinoma. Its inhibition reduces the growth of mouse hepatocellular carcinoma cells, suggesting that it may serve as a novel therapeutic target for hepatocellular carcinoma.[36] MAP6 is an intraluminal protein that stimulates the coiling of neuronal microtubules.[37] Recent research indicates that MAP6 may influence the malignant course of osteosarcoma via the bax/bcl and Wnt/-catenin signaling pathways.[38] FAM110B is a member of the gene family FAM110. This family protein is localized to the centrosome and accumulates at the center of microtubule organization in interphase and at the mitotic spindle pole; FAM110B expression inhibits cell cycle progression in the G1 phase.[39] FAM110B has also been demonstrated to decrease the proliferation and differentiation of lung small-cell carcinoma.[40] Ank2 is mainly expressed in genes associated with human breast cancer and autism risk.[41]

The DPYD Gene May Have the Capacity to Enhance PC Cell Proliferation and Differentiation, Thereby Reducing Patient Survival Time

Five genes, including DPYD, FXYD6, MAP6, FAM110B, and ANK2, were discovered to have survival differences after the 26 genes examined in this study were validated in the GEPIA website. Survival analysis curves revealed that only the DPYD gene negatively impacts the survival of PC patients, ie, DPYD can stimulate the proliferation and differentiation of PC cells, reducing patient survival time.

The protein that DPYD expresses is dihydropyrimidine dehydrogenase. This protein is a catabolic enzyme for pyrimidine. It is a rate-limiting factor and an initiator in the uracil and thymidine catabolic pathway.[42,43] DPYD is currently of interest primarily in cancer patients receiving 5-fu uracil chemotherapy due to the increased risk of chemotherapy toxicity caused by DPYD deficiency resulting from the deficiency of this gene.[44–46] It has also been shown that elevated DPYD expression in hepatocellular carcinoma cells promotes the formation and progression of hepatocellular carcinoma and increases the likelihood of cancer cell infiltration and spread.[47] In a study using mucuna proteins, a chemopreventive drug for PC, reduction of DPYD led to an elevation of STAT3 expression, along with PC cell proliferation, and inhibited the capacity of cancer cells to penetrate and metastasize.[48] This is consistent with our finding that DPYD can stimulate the proliferation and differentiation of PC cells. In addition, there appears to be a relationship between DPYD and the invasive capacity of cancer cells, which requires additional investigation.

HPA Database Validation of DPYD Gene and Immunohistochemical Validation of Clinical Cases Suggest That the DPYD Gene May Become a New Specific Marker for PC and a New Target for PC Treatment

Comparative analysis was performed using the IHC profiles of PC and normal pancreatic tissues from the HPA database. The comparison revealed that DPYD was highly expressed in the cytoplasm of PC cells but nearly nonexistent in normal pancreatic tissue cells; heterogeneity of the DPYD gene in PC cells was also observed in clinical PC cases. Consequently, it is evident that the DPYD gene has the potential to become a potential novel marker for PC and may become a new therapeutic target for PC research.

Conclusion

We used a combination of bioinformatics analysis, experimental verification, and dataset cross-validation to study PC-related DEGs and obtained one novel hub gene (DPYD) that may be associated with the prognosis of PC. Based on these results, the present study provides new ideas for studying the molecular mechanisms of PC.

Previous studies on WGCNA usually only focused on clinical data of genes to discover clinically relevant hub genes. However, in this study we used WGCNA and immune-related data to identify hub genes with immune relevance. The results of this study may provide new ideas for studying the molecular mechanism of PC and provide valuable clues for its treatment.

Ethics Approval and Consent to Participate

This study was conducted in accordance with the declaration of Helsinki.This study was conducted with approval from the Ethics Committee of The First of Affiliated Hospital of Dali University. A written informed consent was obtained from all participants.

Disclosure

The authors report no conflicts of interest in this work.

References

1. KLein AP. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. *Nat Rev Gastroenterol Hepatol*. 2021;18 (7):493–502. doi:10.1038/s41575-021-00457-x
2. Bisht S, Feldmann G. Novel targets in pancreatic cancer therapy - current status and ongoing translational efforts. *Oncol Res Treat*. 2018;41 (10):596–602. doi:10.1159/000493437

3. Tanaka S. Molecular pathogenesis and targeted therapy of pancreatic cancer. *Ann Surg Oncol*. 2016;23(Suppl 2):S197–205. doi:10.1245/s10434-015-4463-x

4. Matsuki R, Arai T, Kogure M, et al. Trends in the treatment of pancreatic cancer in Japan. *Biosci Trends*. 2021;15(3):135–137. doi:10.5582/bst.2021.01103

5. Neesse A, Algül H, Tuveson DA, et al. Stromal biology and therapy in pancreatic cancer: a changing paradigm. *Gut*. 2015;64(9):1476–1484. doi:10.1136/gutjnl-2015-309304

6. Nussinov R, Tsai CJ, Jang H. Anticancer drug resistance: an update and perspective. *Drug Resist Updat*. 2021;59:100796. doi:10.1016/j.drup.2021.100796

7. Randazzo O, Papini F, Mantini G, et al. "Open Sesame?": biomarker status of the human equilibrative nucleoside transporter-1 and molecular mechanisms influencing its expression and activity in the uptake and cytotoxicity of gemcitabine in pancreatic cancer. *Cancers*. 2020;12(11):3206. doi:10.3390/cancers12113206

8. Sciarrillo R, Randazzo O, Peters GJ, et al. Exploring splicing modulation as a novel strategy against pancreatic cancer. Design Of Sf3b1 Subunit Modulators Of The Sf3b Spliceosome Complex; 2021:135.

9. Jiang Y, Chang Y-D, Wang M, et al. Exploring the molecular mechanism of Radix Astragali on colon cancer based on integrated pharmacology and molecular docking technique. *World J Tradit Chin Med*. 2022;8(4):502. doi:10.4103/2311-8571.355594

10. Xiao S-X, Li S-J, Fang W-X, et al. Exploring the mechanism of Tripterygium Wilfordii against cancer using network pharmacology and molecular docking. *World J Tradit Chin Med*. 2022;8(3):417. doi:10.4103/2311-8571.344544

11. Huang J, Li M, Zhoua W-J, et al. Integrated miRNA and mRNA analysis identified potential mechanisms and targets of qianggan extracts in preventing nonalcoholic steatohepatitis. *World J Tradit Chin Med*. 2022;8(1):77. doi:10.4103/wjtcm.wjtcm_48_21

12. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 2008;9:559. doi:10.1186/1471-2105-9-559

13. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007

14. Liu S, Wang Z, Zhu R, et al. Three differential expression analysis methods for RNA sequencing: limma, EdgeR, DESeq2. *J Vis Exp*. 2021;175:1.

15. Zhang Y, Luo J, Liu Z, et al. Identification of hub genes in colorectal cancer based on weighted gene co-expression network analysis and clinical data from the cancer genome atlas. *Biosci Rep*. 2021;41:7. doi:10.1042/BSR20211280

16. Chen S, Yang D, Lei C, et al. Identification of crucial genes in abdominal aortic aneurysm by WGCNA. *PeerJ*. 2019;7:e7873. doi:10.7717/peerj.7873

17. Zeng D, Ye Z, Shen R, et al. IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures. *Front Immunol*. 2021;12:687975. doi:10.3389/fimmu.2021.687975

18. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612. doi:10.1038/ncomms3612

19. Tang Z, Li C, Kang B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45(W1):W98–w102. doi:10.1093/nar/gkx247

20. Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419. doi:10.1126/science.1260419

21. Huang H, Zhu L, Huang C, et al. Identification of hub genes associated with clear cell renal cell carcinoma by integrated bioinformatics analysis. *Front Oncol*. 2021;11:726655. doi:10.3389/fonc.2021.726655

22. Deng H, Hang Q, Shen D, et al. Low expression of CHRDL1 and SPARCL1 predicts poor prognosis of lung adenocarcinoma based on comprehensive analysis and immunohistochemical validation. *Cancer Cell Int*. 2021;21(1):259. doi:10.1186/s12935-021-01933-9

23. Yang Y, Deng X, Chen X, et al. Landscape of active enhancers developed de novo in cirrhosis and conserved in hepatocellular carcinoma. *Am J Cancer Res*. 2020;10(10):3157–3178.

24. Chen Y, Chen H, Mao B, et al. Transcriptional characterization of the tumor immune microenvironment and its prognostic value for locally advanced lung adenocarcinoma in A Chinese population. *Cancer Manag Res*. 2019;11:9165–9173. doi:10.2147/CMAR.S209571

25. Zhou P, Li B, Liu F, et al. The epithelial to mesenchymal transition (EMT) and cancer stem cells: implication for treatment resistance in pancreatic cancer. *Mol Cancer*. 2017;16(1):52. doi:10.1186/s12943-017-0624-9

26. Jiang Y, Zhan H. Communication between EMT and PD-L1 signaling: new insights into tumor immune evasion. *Cancer Lett*. 2020;468:72–81. doi:10.1016/j.canlet.2019.10.013

27. Yan Z, Ohuchida K, Fei S, et al. Inhibition of ERK1/2 in cancer-associated pancreatic stellate cells suppresses cancer-stromal interaction and metastasis. *J Exp Clin Cancer Res*. 2019;38(1):221. doi:10.1186/s13046-019-1226-8

28. Biffi G, Oni TE, Spielman B, et al. IL1-Induced JAK/STAT signaling is antagonized by TGFβ to Shape CAF heterogeneity in pancreatic ductal adenocarcinoma. *Cancer Discov*. 2019;9(2):282–301. doi:10.1158/2159-8290.CD-18-0710

29. Lanfredini S, Thapa A, O'Neill E. RAS in pancreatic cancer. *Biochem Soc Trans*. 2019;47(4):961–972. doi:10.1042/BST20170521

30. Mortazavi M, Moosavi F, Martini M, et al. Prospects of targeting PI3K/AKT/mTOR pathway in pancreatic cancer. *Crit Rev Oncol Hematol*. 2022;176:103749. doi:10.1016/j.critrevonc.2022.103749

31. Zhao H, Wu L, Yan G, et al. Inflammation and tumor progression: signaling pathways and targeted intervention. *Signal Transduct Target Ther*. 2021;6(1):263. doi:10.1038/s41392-021-00658-5

32. Sun X, He X, Zhang Y, et al. Inflammatory cell-derived CXCL3 promotes pancreatic cancer metastasis through a novel myofibroblast-hijacked cancer escape mechanism. *Gut*. 2022;71(1):129–147. doi:10.1136/gutjnl-2020-322744

33. Zhang Y, Chandra V, Riquelme Sanchez E, et al. Interleukin-17-induced neutrophil extracellular traps mediate resistance to checkpoint blockade in pancreatic cancer. *J Exp Med*. 2020;217:12. doi:10.1084/jem.20190354

34. Hu F, Guo F, Zhu Y, et al. IL-17 in pancreatic disease: pathogenesis and pharmacotherapy. *Am J Cancer Res*. 2020;10(11):3551–3564.

35. Kamisawa T, Wood LD, Itoi T, et al. Pancreatic cancer. *Lancet*. 2016;388(10039):73–85. doi:10.1016/S0140-6736(16)00141-0

36. Gao Q, Chen X, Duan H, et al. FXYD6: a novel therapeutic target toward hepatocellular carcinoma. *Protein Cell*. 2014;5(7):532–543. doi:10.1007/s13238-014-0045-0

37. Cuveillier C, Delaroche J, Seggio M, et al. MAP6 is an intraluminal protein that induces neuronal microtubules to coil. *Sci Adv*. 2020;6(14):eaaz4344. doi:10.1126/sciadv.aaz4344

38. Lin H, Wu T, Peng L, et al. Lnc-MAP6-1:3 knockdown inhibits osteosarcoma progression by modulating Bax/Bcl-2 and Wnt/β-catenin pathways. *Int J Med Sci*. 2020;17(15):2248–2256. doi:10.7150/ijms.47405

39. Hauge H, Patzke S, Aasheim HC. Characterization of the FAM110 gene family. *Genomics*. 2007;90(1):14–27. doi:10.1016/j.ygeno.2007.03.002

40. Xie M, Cai L, Li J, et al. FAM110B inhibits non-small cell lung cancer cell proliferation and invasion through inactivating Wnt/β-Catenin Signaling. *Onco Targets Ther*. 2020;13:4373–4384. doi:10.2147/OTT.S247491

41. Jin X, Simmons SK, Guo A, et al. In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science*. 2020;370:6520. doi:10.1126/science.aaz6063

42. Diasio RB, Harris BE. Clinical pharmacology of 5-fluorouracil. *Clin Pharmacokinet*. 1989;16(4):215–237. doi:10.2165/00003088-198916040-00002

43. Longley DB, Harkin DP, Johnston PG. 5-fluorouracil: mechanisms of action and clinical strategies. *Nat Rev Cancer*. 2003;3(5):330–338. doi:10.1038/nrc1074

44. Mattison LK, Fourie J, Desmond RA, et al. Increased prevalence of dihydropyrimidine dehydrogenase deficiency in African-Americans compared with Caucasians. *Clin Cancer Res*. 2006;12(18):5491–5495. doi:10.1158/1078-0432.CCR-06-0747

45. Etienne MC, Lagrange JL, Dassonville O, et al. Population study of dihydropyrimidine dehydrogenase in cancer patients. *J Clin Oncol*. 1994;12 (11):2248–2253. doi:10.1200/JCO.1994.12.11.2248

46. Ogura K, Ohnuma T, Minamide Y, et al. Dihydropyrimidine dehydrogenase activity in 150 healthy Japanese volunteers and identification of novel mutations. *Clin Cancer Res*. 2005;11(14):5104–5111. doi:10.1158/1078-0432.CCR-05-0217

47. Zhu WP, Liu ZY, Zhao YM, et al. Dihydropyrimidine dehydrogenase predicts survival and response to interferon-α in hepatocellular carcinoma. *Cell Death Dis*. 2018;9(2):69. doi:10.1038/s41419-017-0098-0

48. Kato H, Naiki-Ito A, Suzuki S, Et A. DPYD, down-regulated by the potentially chemopreventive agent luteolin, interacts with STAT3 in pancreatic cancer. *Carcinogenesis*. 2021;42(7):940–950. doi:10.1093/carcin/bgab017