



Identification of a glycolysis-related gene signature for predicting pancreatic cancer survival

Jiachao Zhang[#], Zhehao Liu[#], Zhensheng Zhang, Rong Tang, Yongchao Zeng, Pingping Chen

Department of Hepatobiliary and Pancreatic Surgery, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, Haikou, China

Contributions: (I) Conception and design: J Zhang; (II) Administrative support: Z Liu; (III) Provision of study materials or patients: Z Zhang; (IV) Collection and assembly of data: R Tang; (V) Data analysis and interpretation: Y Zeng, P Chen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Pingping Chen. Department of Hepatobiliary and Pancreatic Surgery, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, No. 19, Xiuhua Road, Haikou 570311, China. Email: ping52ping@126.com.

Background: Pancreatic cancer (PC) is one of the most common malignant tumors of the digestive tract. Its clinical symptoms are obscure and atypical. It is difficult to diagnose and treat. Tumor cells mainly obtain energy through glycolysis to promote their growth. Inhibiting glycolysis can inhibit proliferation and kill tumor cells.

Methods: Using bioinformatics method, we investigate the relationships between glycolysis-related genes and PC tumor samples' epidemiologic information comprehensively.

Results: Different expression levels of 27 genes were identified. Using bioinformatics methods, we plotted two subgroup curves based on glycolysis-related gene expression level. Potential predictive genes were screened and their prognostic values were analyzed. Survival among high-risk group and low risk group had significant difference. Receiver operating characteristic (ROC) curve analysis indicated that area under curve (AUC) of 10 genes was greater than 0.8. These genes could be used for clinical diagnosis and prediction for PC. Two potential predictors [Kinesin Family Member 20A (KIF20A) and MET Proto-Oncogene, Receptor Tyrosine Kinase (MET)] that met the independent predictive value were selected. In univariate analysis, we screened out 3 regulators MET, protein kinase CAMP-activated catalytic subunit alpha (PRKACA) and KIF20A. According to the 3 regulatory factors, the prognostic signals of PC were constructed, by which the samples with good prognosis and poor prognosis can be clearly distinguished independently of potential confounding factors.

Conclusions: Our results indicate that for PC, glycolysis -related genes could be promising therapeutic targets or prognostic indicators.

Keywords: The Cancer Genome Atlas (TCGA); bioinformatics; pancreatic cancer (PC); glycolysis; prognostic signature

Submitted Nov 25, 2021. Accepted for publication Feb 21, 2022.

doi: 10.21037/jgo-22-17

View this article at: <https://dx.doi.org/10.21037/jgo-22-17>

Introduction

Pancreatic cancer (PC) is one of the most common malignant tumors of digestive system. Its 5-year survival rate is only 10% (1,2). About 90% of PCs are pancreatic ductal adenocarcinoma (PDAC). Its malignancy is very

high, while its curative effect and survival prognosis are very poor (3,4). The etiology of PC is still unclear. Patients with PC are less likely to benefit from the current treatment. The average survival time is about 2 years, even after surgery (5). The main reason is that the early diagnosis of

PC is more difficult. The vast majority of patients have developed metastases or locally advanced state at the time of diagnosis, resulting in the delay of treatment. Meanwhile, radiotherapy and chemotherapy have a poor therapeutic effect on PC. The major risk factors for PC include genetic mutation, smoking, chronic pancreatitis, diabetes, etc. (6,7).

In recent years, tumor cell metabolism has become a research hotspot, as each tumor cell replication cycle means doubling the energy required to maintain cell proliferation. The basis of tumor cell metabolism is the Warburg effect, or aerobic glycolysis, which was first described by Warburg in 1927. He found that the glycolysis ability of tumor cells is significantly higher than that of normal cells even when oxygen is sufficient (8). The occurrence and development of most tumors are related to glycolysis. In addition, glycolysis also affects the prognosis of tumor patients, and the effect of antitumor treatment is also affected by tumor glycolysis (9). Glycolysis plays a very important role in cell metabolism. Glycolysis based tumor targeted therapy is an urgent problem to be solved. For medical workers, it is still difficult to find effective tumor glycolysis therapy targets (10). In acidic and hypoxic tumor microenvironment, the glycolytic metabolism of PC cells is enhanced, thus making cancer cells suitable for tumor microenvironment (11). In PC, hypoxia and hypoxic microenvironment increase the aerobic glycolysis of PC cells. In addition, PC has a dense interstitial fibrosis, resulting in external mechanical stress that impairs tumor vasculature and worsens tissue perfusion, resulting in reduced supply of nutrients and oxygen to tumor cells (12). Therefore, glycolysis is closely related to the occurrence and development of PC, and has become a hot topic in the treatment of PC. However, in the treatment of PC with targeted therapies for glycolysis, it is essential to ensure that glycolytic enzymes or energy metabolism is not impaired by normal glycation or energy metabolism.

The purpose of this study was to evaluate the expression of glycolysis-related genes (GRGs) in PC based on The Cancer Genome Atlas (TCGA) data, and to investigate the relationship between GRGs and PC survival. To this end, gene selection was performed mainly by gene set enrichment analysis (GSEA). Study has focused on differentially expressed genes in tissues to identify biomarkers, but GSEA can improve the statistical analysis of gene expression and biological significance (13). Finally, in our study, we constructed a glycolysis-associated genetic risk profile that effectively predicted patient outcomes. In addition, our gene-based model, acting as an independent predictor, could determine that patients with a high-

risk score had a worse prognosis than those with a low-risk score. In addition, the prognostic performance of the risk model was significantly superior to other clinical features, and showed better performance in predicting clinical outcomes in patients with PC. We present the following article in accordance with the STREGA reporting checklist (available at <https://jgo.amegroups.com/article/view/10.21037/jgo-22-17/rc>).

Methods

Data collection

All subjects in the study were obtained through TCGA (14). There were 189 samples, including 4 adjacent normal tissues and 185 PC tissues. The clinicopathological characteristics of patients with PC were described in detail, as shown in *Table 1*. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

GSEA

To determine whether there were significant differences between the PC tissues and adjacent normal tissues in the identified GRG sets, GSEA (<http://www.broadinstitute.org/gsea/index.jsp>) was performed (15). Statistical significance was defined using the standard of adjusted P value <0.05. The genes of the GRG sets which were significant based on adjusted P value and $\log_2|\text{fold change}| >1$ (16) were collected for subsequent analysis.

Bioinformatics analysis/statistics

R software (version 4.0.2) was used to evaluate the differential expression of GRGs, and 27 genes were identified (*PDHA2*, *HK3*, *P2RX7*, *CXCR4*, *ESRRB*, *CHST2*, *PKP2*, *FKBP4*, *KIF20A*, *QSOX1*, *AURKA*, *GALE*, *ALDH3B1*, *TSTA3*, *MET*, *SPAG4*, *PYGB*, *SDC1*, *ADORA2B*, *AK4*, *CAPN5*, *EFNA3*, *B3GNT3*, *PCK1*, *ALDH3A1*, *TFF3*, and *KIAA1429*) in PC and normal samples. A heatmap of the 27 genes was plotted. Furthermore, we performed Pearson correlation analyses to identify the correlations of these genes.

Protein-protein interaction (PPI)

The 27 genes were imported into the STRING database (17) and the species was defined as "Homo sapiens". Then, the PPI data was obtained. The results were downloaded in TSV

Table 1 Clinicopathological characteristics of pancreatic cancer (PC) patients from The Cancer Genome Atlas (TCGA) database

Characteristics	PC patients (N=185)	
	No.	%
Age, years		
≤65	96	51.89
>65	89	48.11
Gender		
Female	83	44.86
Male	102	55.14
Pathological stage		
I	21	11.35
II	152	82.16
III	4	2.16
IV	5	2.70
Unknown	3	1.62
Grade		
G1	32	17.30
G2	97	52.43
G3	51	27.57
G4	2	1.08
Unknown	3	1.62
Survival time		
Long (>5 years)	8	4.32
Short (<5 years)	177	95.68
Stage T		
T1	7	3.78
T2	24	12.97
T3	147	79.46
T4	4	2.16
TX	1	0.54
Unknown	1	0.54
Stage M		
M0	85	45.95
M1	5	2.70
MX	95	51.35

Table 1 (continued)**Table 1** (continued)

Characteristics	PC patients (N=185)	
	No.	%
Stage N		
N0	50	27.03
N1	130	70.27
NX	4	2.16
Unknown	1	0.54
OS status		
Dead	100	54.05
Alive	85	45.94

OS, overall survival.

format. Combined score and information of node 1 and node 2 were retained in the file. Using Cytoscape 3.6, the network was analyzed and generated.

Principal component analysis (PCA) and survival analyses of subgroups

Using consistent clustering, we detected unknown possible clusters which were composed of items with similar intrinsic characteristics (18). The 27 genes were differentially expressed, and through the ConsensusClusterPlus package in R software, different subgroups of 185 tumor samples were identified and the grouping results were verified using PCA. Survival curves of the 2 subgroups were drawn using the Kaplan-Meier (KM) method.

Prognostic value of GRGs

We selected the candidate genes through univariate Cox regression analysis and set the screening standard as $P < 0.05$. To analyze the high-dimensional data, we performed LASSO regression. In addition, we used the “glmnet” package of R software to select the most useful prognostic factors (19). We screened out 10 genes and calculated the associated risk of these genes. Based on the median expression of GRGs, we divided patients into low-risk and high-risk groups. To investigate the relationship between GRGs and survival rate, we performed KM survival analysis. The P value of the KM survival curve

Table 2 Gene sets enriched in pancreatic cancer

GS follow link to MSigDB	Size	ES	NES	NOM P value	FDR q-value
BIOCARTA_GLYCOLYSIS_PATHWAY	3	0.83	1.24	0.203	0.203
GO_GLYCOLYTIC_PROCESS	106	0.30	0.78	0.742	0.742
HALLMARK_GLYCOLYSIS	200	0.60	1.56	0.025	0.025
KEGG_GLYCOLYSIS_GLYCONEOGENESIS	62	0.45	1.17	0.274	0.274
REACTOME_GLYCOLYSIS	72	0.29	0.71	0.805	0.805

GS, Genomic selection; ES, Enrichment score; NES, Normalized enrichment score; NOM, normal; FDR, false discovery rate.

was calculated according to the log rank test. We plotted receiver operating characteristic (ROC) curves to verify the accuracy of the model. To identify prognostic factors of PC, we performed univariate and multivariate Cox regression analyses. We plotted heatmaps of GRGs and clinical risk factors. Using the “bearm” software package in R, the gene expression between the PC group and normal group in TCGA data set was analyzed.

Construction of the prognostic prediction model

Using multivariate Cox regression analysis, we screened out the regulatory factors of gene expression that were significantly related to the overall survival (OS) of PC patients. Based on LASSO Cox regression, we analyzed the risk factors related to the prognosis of patients.

Nomogram analysis of PC patients

Through the rmsr package in R software, we performed nomogram analysis and analyzed the factors significantly related to OS in PC patients using the nomogram. We used the area under the ROC curve (AUC) to evaluate the predictive ability of the nomograms.

Construction and evaluation of the glycolysis-related risk signature

Using the survival package in R, the patient risk value was output by the multivariate Cox regression method. According to the median value, they were divided into high- and low-risk groups, and the survival curves of the 2 groups were drawn. The effectiveness of the model was judged by the ROC curve, the risk curve was drawn, and independent prognosis was analyzed. We used an online network (<https://www.cbioportal.org/>) to analyze the mutations of the model

genes in the TCGA data set containing PC samples. The mutation frequencies and types of glycolytic genes in tumor tissues were analyzed. The R limma package was used to analyze the differences. The R survival package was used to analyze the survival of patients grouped according to clinical characteristics and to generate the survival curves. At the same time, the survival package was used to analyze the survival of patients in the high and low risk groups according to the prognostic model, and the corresponding curve was drawn. Using Cox regression analysis, the independent prognostic value of other clinical features was assessed.

Statistical analysis

Statistical analysis of gene expression data was conducted by R 4.0.2. In addition, R 4.0.2 was used for image plotting, and considered $P < 0.05$ to indicate statistical significance.

Results

Initial screening of genes using GSEA

We obtained a data set from TCGA containing the clinical information of 185 PC patients and 4 normal controls. We used this data and GSEA to verify which gene sets were significantly different between PC tissues and matched adjacent normal tissues. A significantly enriched gene set ($P < 0.05$) was derived from hallmark_glycolysis (Table 2, Figure 1). A total of 27 genes were selected from the gene set for subsequent analysis.

Expression of GRGs in PC

Based on TCGA dataset, we constructed a heatmap of 27 differentially expressed GRGs (*PDHA2*, *HK3*, *P2RX7*,

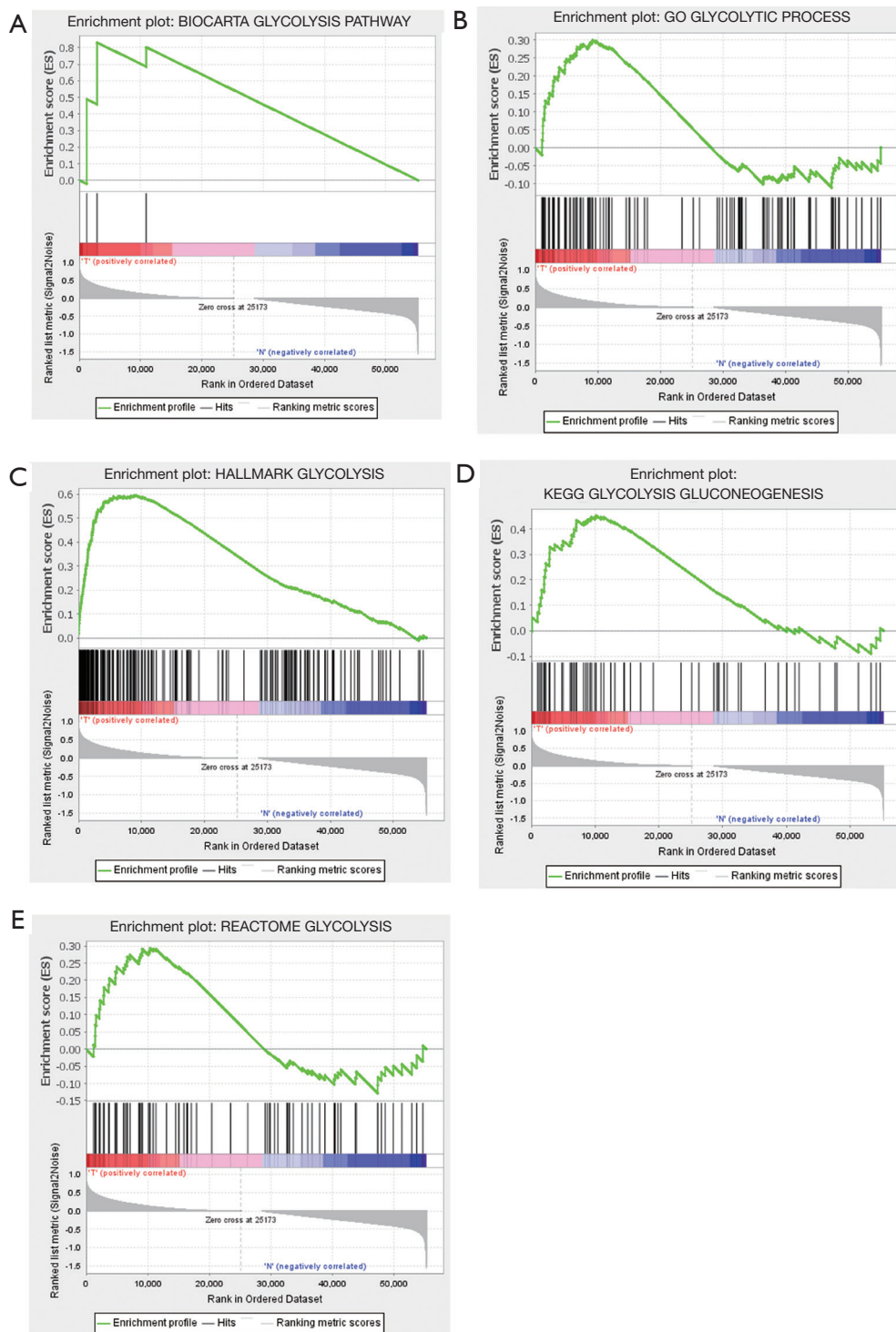


Figure 1 Enrichment plots of 5 gene sets which had significant differences between normal tissues and pancreatic cancer (PC) tissues as determined by gene set enrichment analysis (GSEA).

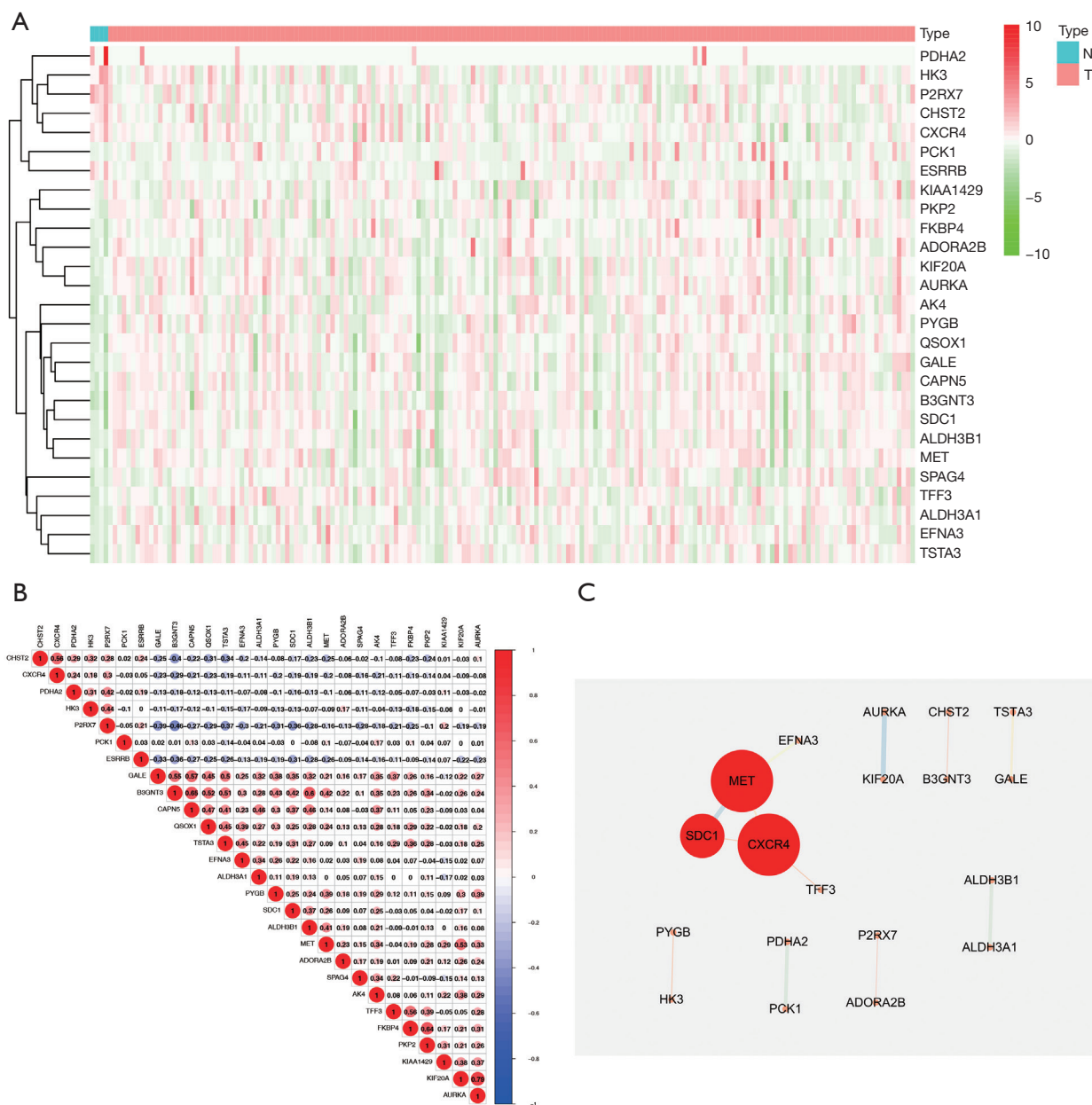


Figure 2 Expression, correlation, and protein-protein interaction (PPI) network of differentially expressed glycolysis-related genes. (A) Heatmaps of differentially expressed glycolysis-related genes expressed in tumors and adjacent normal tissues. (B) Correlation matrix of the interactions between differentially expressed glycolysis-related genes. Correlation coefficients are plotted with negative correlation (blue) and positive correlation (red). (C) Interaction network of proteins related to the differentially expressed glycolysis-related genes. The red nodes represent the big hub nodes. The node size was proportional to the target degree in the network.

CXCR4, *ESRRB*, *CHST2*, *PKP2*, *FKBP4*, *KIF20A*, *QSOX1*, *AURKA*, *GALE*, *ALDH3B1*, *TSTA3*, *MET*, *SPAG4*, *PYGB*, *SDC1*, *ADORA2B*, *AK4*, *CAPN5*, *EFNA3*, *B3GNT3*, *PCK1*, *ALDH3A1*, *TFF3*, and *KIAA1429*), which were significantly

different in tumors compared with normal tissues (*Figure 2A*). As shown in *Figure 2B*, Pearson correlation analysis demonstrated that some of these genes showed positive correlations and some showed negative correlations.

Table 3 Information of different glycolysis-related genes (GRGs)

Gene symbol	Average shortest path length	Betweenness centrality	Closeness centrality	Clustering coefficient	Degree
<i>ADORA2B</i>	1	0	1	0	1
<i>P2RX7</i>	1	0	1	0	1
<i>ALDH3A1</i>	1	0	1	0	1
<i>ALDH3B1</i>	1	0	1	0	1
<i>AURKA</i>	1	0	1	0	1
<i>KIF20A</i>	1	0	1	0	1
<i>B3GNT3</i>	1	0	1	0	1
<i>CHST2</i>	1	0	1	0	1
<i>CXCR4</i>	1.25	0.5	0.8	0.33	3
<i>MET</i>	1.25	0.5	0.8	0.33	3
<i>SDC1</i>	1.5	0	0.67	1	2
<i>TFF3</i>	2	0	0.5	0	1
<i>EFNA3</i>	2	0	0.5	0	1
<i>GALE</i>	1	0	1	0	1
<i>TSTA3</i>	1	0	1	0	1
<i>HK3</i>	1	0	1	0	1
<i>PYGB</i>	1	0	1	0	1
<i>PCK1</i>	1	0	1	0	1
<i>PDHA2</i>	1	0	1	0	1

KIF20A and *AURKA* were identified as having the highest correlation, with a correlation coefficient of 0.79. *B3GNT3* and *CAPN5*, *CHST2* and *CXCR4*, *GALE* and *CAPN5* as well as *BCGNT3* and *ALDH3B1* were also obviously positively correlated, and their correlation coefficients were 0.65, 0.56, 0.57, and 0.6, respectively.

PPI network

There were 19 nodes and 12 edges in the PPI network, as shown in *Figure 2C*, with an average node degree of 0.923 and local clustering coefficient of 0.679. The thickness of the edge indicates the combination fraction. The rougher the edges, the higher the combination score. The size of the node represents the degree value of the protein. The larger the node, the greater the degree value. The hierarchy of each node is shown in *Table 3* and *Figure 2C*.

ROC analysis

The diagnostic significance, specificity, and sensitivity of potential genes were evaluated by ROC analysis, and the diagnostic value was also assessed using the AUC. In our study, genes with an AUC higher than 0.85 were considered to have good predictive power and high sensitivity and specificity. ROC analysis further confirmed the sensitivity and specificity of these potential genes. The AUCs of *AK4*, *ALDH3A1*, *B3GNT3*, *CAPN5*, *EFNA3*, *FKBP4*, *P2RX7*, *PYGB*, *QSOX1*, and *TSTA3* were 0.851, 0.868, 0.937, 0.872, 0.858, 0.869, 0.941, 0.937, 0.855, and 0.876, respectively, as shown in *Figure 3*. These 10 genes were considered to be the most relevant potential genes in PC and may be used for the clinical diagnosis and prognostic prediction of PC. Follow-up clinical trials should be conducted to evaluate their efficacy for further verification.

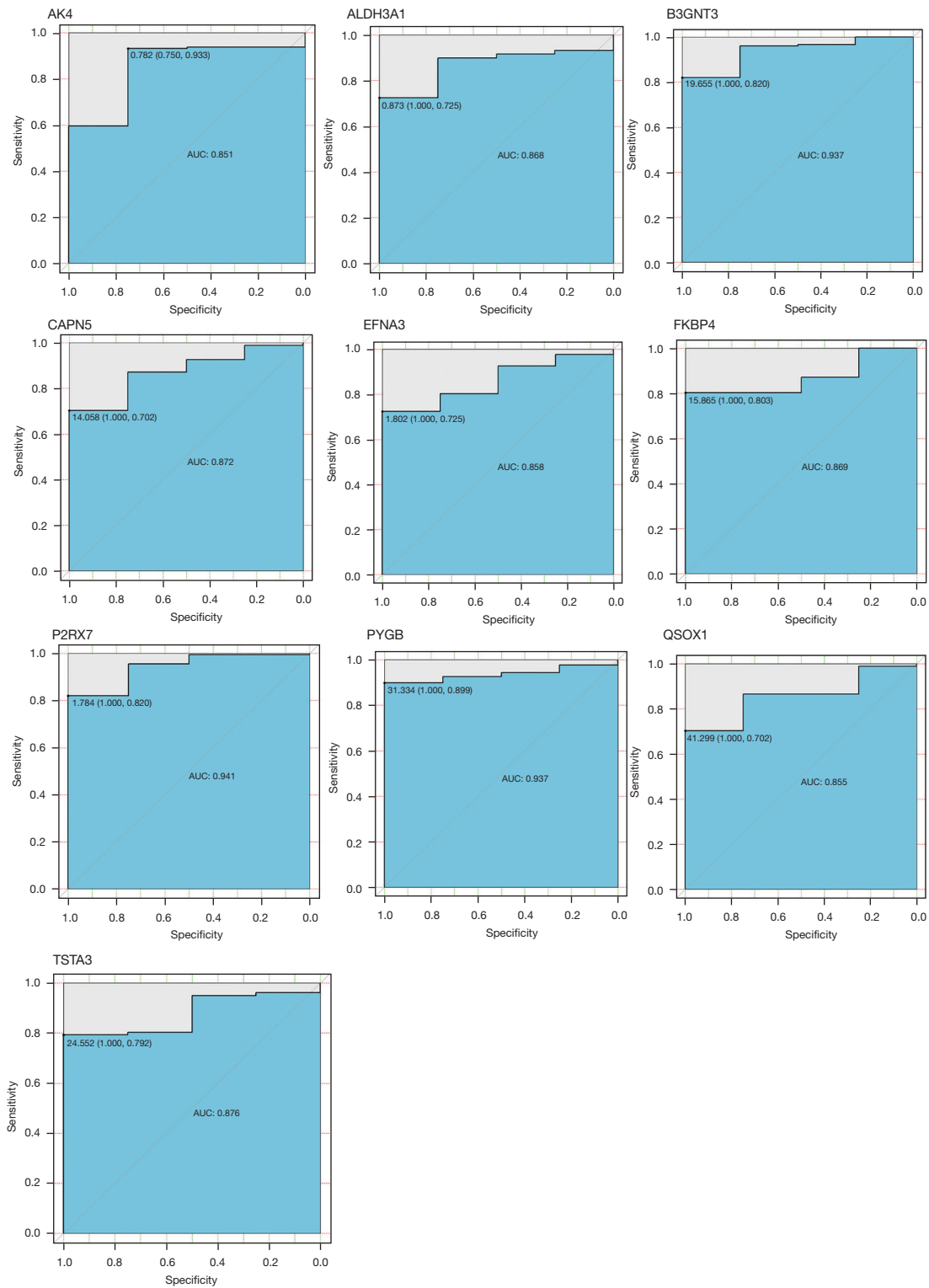


Figure 3 Receiver operating characteristic (ROC) curve of 10 screened marker genes [area under the curve (AUC) >0.85].

Consensus clustering of GRGs identified 2 clusters of PC

All 185 patients with PC were grouped by the consistent clustering method in order to study the relationship between the expression profile of GRGs and the prognosis of these patients. As shown in *Figure 4A-4C*, through expression similarity of GRGs based on TCGA dataset, when the clustering stability increases from $k=2$ to 9, $k=2$ seems to be an acceptable choice. Then, PCA was performed to compare transcriptional profiles between the 2 clusters, as shown in *Figure 4D,4E*. However, we observed that there was no significant difference in OS between cluster 1 and cluster 2 (*Figure 4F*). Further comparison of the clinical characteristics of the 2 subgroups showed significant differences in grade, age, gender, and T, M, and N stage (*Figure 4G*).

Construction of the LASSO model

A total of 27 genes were analyzed by univariate Cox regression, and 9 candidate genes were selected through the screening criteria ($P<0.05$; *Figure 5A*). With LASSO Cox regression model we screened predictive genes as prognostic indicators. We selected λ when the median of the sum of squares of residuals was the smallest (*Figure 5B,5C*). *MET*, *KIF20A*, *B3GNT3*, *AURKA*, *ALDH3B1*, *AK4*, *ESRRB*, *KIAA1429*, and *ALDH3A1* were identified as prognostic factors for PC. At the same time, we calculated the risk scores of 9 genes and performed univariate and multivariate Cox regression analysis. Patients were divided into a low-risk group and high-risk group using the median expression of 9 candidate genes as the critical value of the combination model. The high-risk group had significantly poorer prognosis than the low-risk group ($P=0.0001478$). We generated the survival curve using the KM method (*Figure 5D*). The prognostic efficiency of risk factors was also compared by ROC curves. The results showed an AUC of 0.771 (*Figure 5E*), which suggested that GRGs could be used as biomarkers for the prognosis of PC.

Prognostic value of GRGs

Univariate analysis showed that N stage, age, and risk score of GRGs had an influence on the prognosis of patients ($P<0.05$). Gender, grade, clinical stage, T stage, and N stage were not associated with the prognosis of PC ($P>0.05$; *Figure 6A*). Multiple regression analysis showed that age, N stage, and

risk score of GRGs were independent prognostic factors for PC ($P<0.05$; *Figure 6B*). As shown in *Figure 6C*, *MET* and *KIF20A* were up-regulated in high-risk patients. Grade was correlated with the risk degree, but there were no significant differences in T stage, M stage, N stage, clinical stage, age, and gender ($P>0.05$).

We constructed a nomogram to predict the 1-, 3-, and 5-year OS probabilities, including factors from a multivariate Cox regression analysis such as N stage, age, and risk score, as shown in *Figure 7A*. The AUCs for the prediction of 1-, 3-, and 5-year OS were 0.7, 0.786, and 0.739, respectively. ROC analysis showed that the factors performed well in predicting the OS of PC. These results suggest that risk scores can be used to reinforce clinicopathological characteristics in order to improve the prognosis of PC, as shown in *Figure 7B-7D*.

Construction and evaluation of the glycolysis-related risk signature

Patients with a high-risk score had higher mortality than those with a low-risk score ($P<0.001$, log-rank test; *Figure 8A*). The AUC values of prognostic model for OS were 0.728 (*Figure 8B*). *Figure 8C,8D* show the rank distribution of risk scores and survival status in PC patients. The expression patterns of 3 GRGs in the high/low-risk groups are shown in the heatmap (*Figure 9A*). *TXN*, *MET* and *KIF20A* have varying degrees of mutation frequency in tumor samples, as shown in *Figure 9B-9D*. There were also differences in the expression of *TXN*, *MET*, and *KIF20A* between normal samples and tumor samples ($P<0.05$), as shown in *Figure 10*. We performed univariate and multivariate Cox regression analysis based on the risk scores of 3 genes. Univariate analysis showed that grade, age, and risk score of GRGs had an influence on the prognosis of patients ($P<0.05$). Multiple regression analysis showed that age and risk score of GRGs were independent prognostic factors for PC ($P<0.05$), as shown in *Figure 11*. We constructed a nomogram to predict the 1-, 3-, and 5-year OS probabilities, including age and risk scores, as shown in *Figure 12A*. The AUCs for the prediction of 1-, 3-, and 5-year OS were 0.700, 0.775, and 0.720, respectively. ROC analysis showed that the factors performed well in predicting the OS of PC, as shown in *Figure 12B-12D*. We calculated each patient's risk score based on the genetic profile and screened for 3 genes.

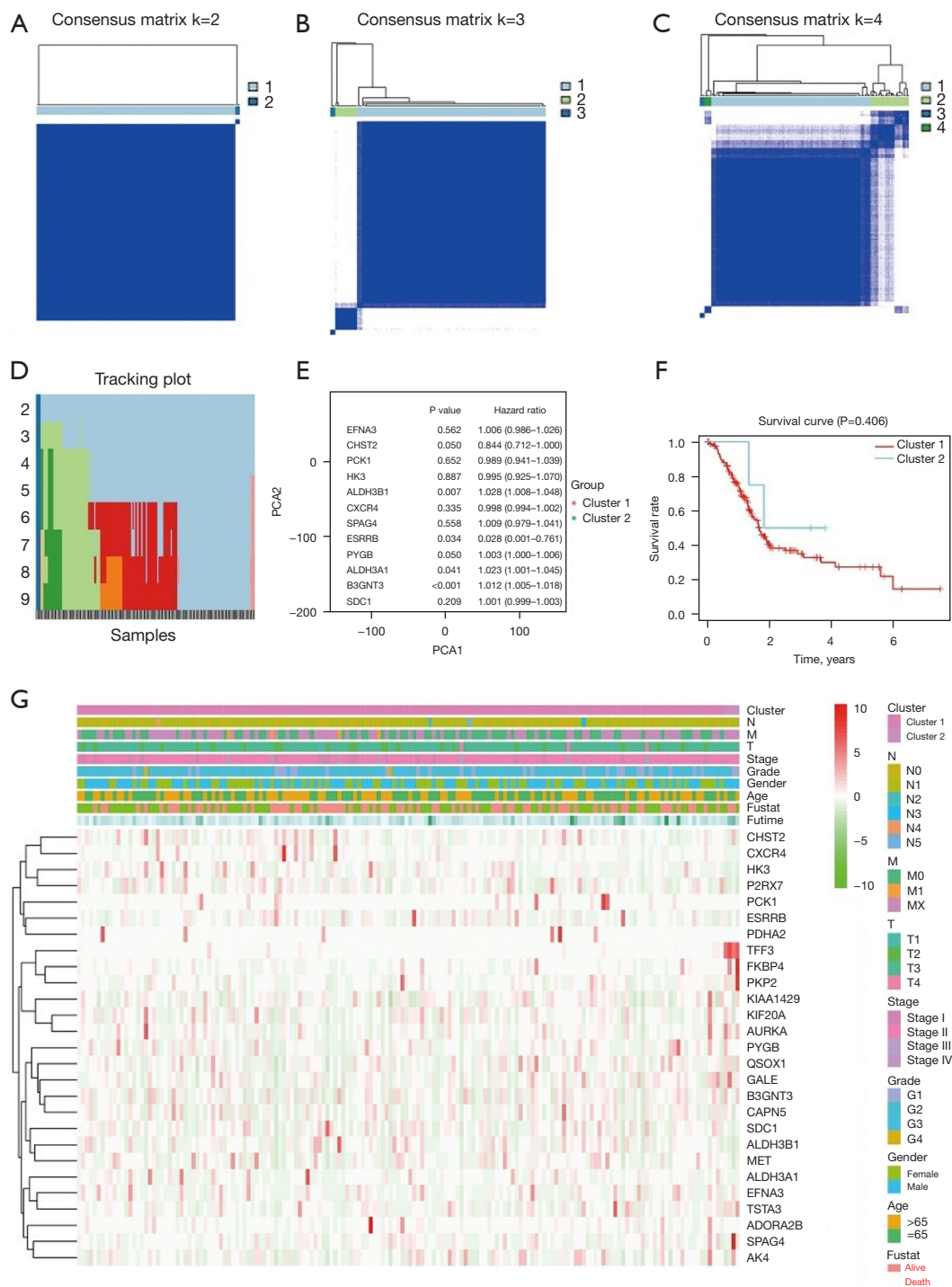


Figure 4 Divergent clinicopathological features and OS of PC in the subgroups. (A-C) Different clinicopathological features and the overall survival (OS) of pancreatic cancer (PC) in the cluster 2-4 subgroups. The patients were divided into 2, 3 or 4 clusters respectively. According to the separation degree between different clusters, the patients were finally divided into 2 clusters. (D) Tracking plot for k=2 to 9. (E) Principal component analysis (PCA) of the 2 subgroups. (F) Kaplan-Meier OS curves for 178 PC patients. (G) Heatmap and clinicopathological characteristics of the 2 clusters defined by the GRG.

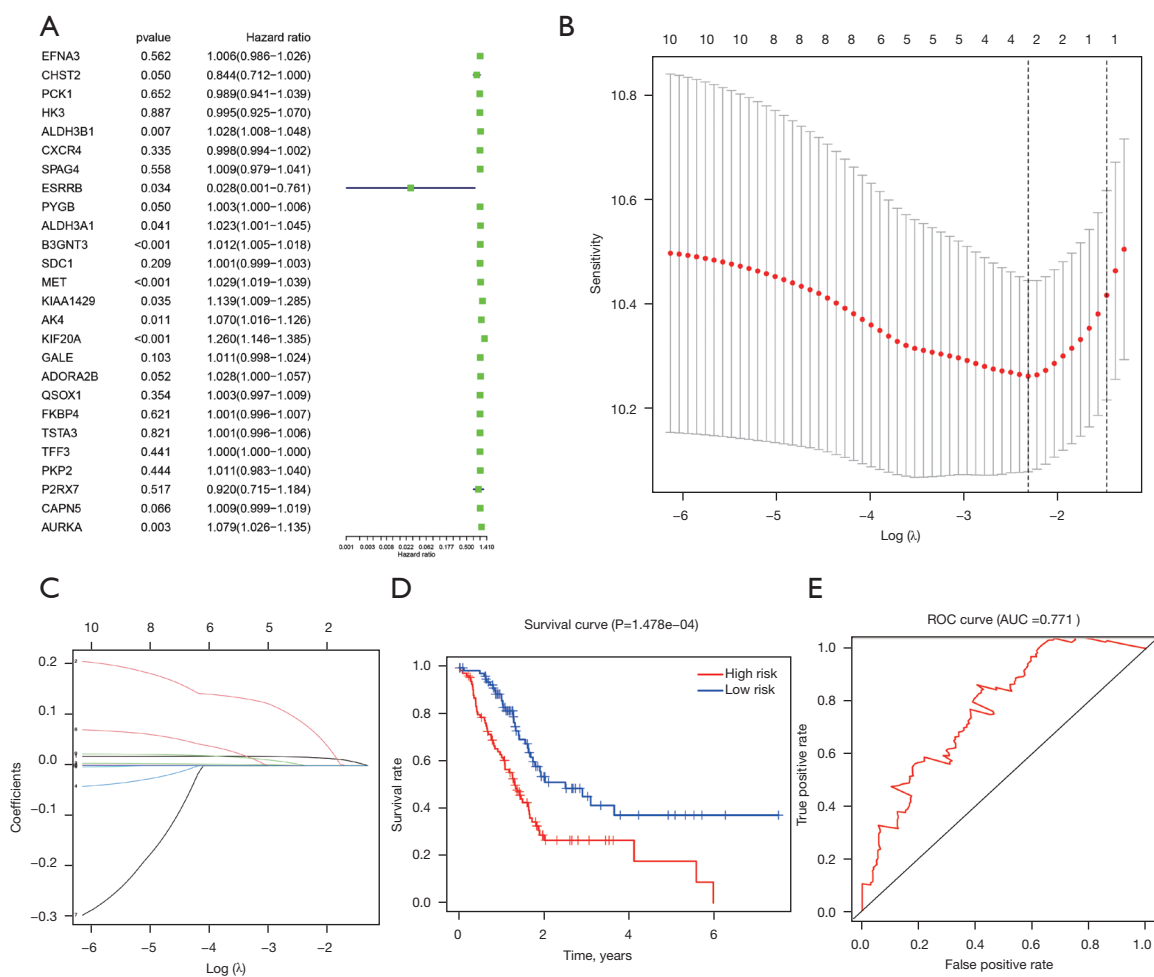


Figure 5 Gene selection and survival analysis for pancreatic cancer (PC) prognostic prediction. (A) Forest plots for hazard ratios (HRs) of survival-associated glycolysis-related genes in PC. (B) Partial likelihood deviance versus $\log(\lambda)$ was drawn using the LASSO Cox regression model. (C) Coefficients of selected features are shown by the λ parameter. (D) Kaplan-Meier survival plots of the 2 groups. (E) Receiver operating characteristic (ROC) curves of the survival model in PC [area under the curve (AUC) = 0.771].

Comparison with other prognostic signatures

Subsequently, we compared the relationship between clinical factors (age, gender, grade, M stage, N stage, T stage, clinical stage) and patient prognosis. The results showed that there were significant differences in survival rates between high-risk and low-risk groups in patients younger than 65 years old and those in M0 stage, N0 stage, N1-3 stage, and T stage, as shown in *Figure 13*.

Discussion

PC develops rapidly and is difficult to diagnose at the early

stage. It is a malignant tumor with poor prognosis (20). Tumor microenvironment is a new concept in tumor research, and it is also an important symbol of tumor research. Tumor microenvironment provides nutrients to cancer cells to ensure their survival, and also provides positive conditions for the invasion and metastasis of cancer cells (21). In recent years, the overall level of medical treatment has been improving, and many new methods have emerged for the treatment of PC. However, the key mechanism of the pathogenesis and metastasis of PC has not been clarified so far. The most effective way to treat PC is surgical resection. With the continuous deepening of research on PC, researchers have found that the tumor

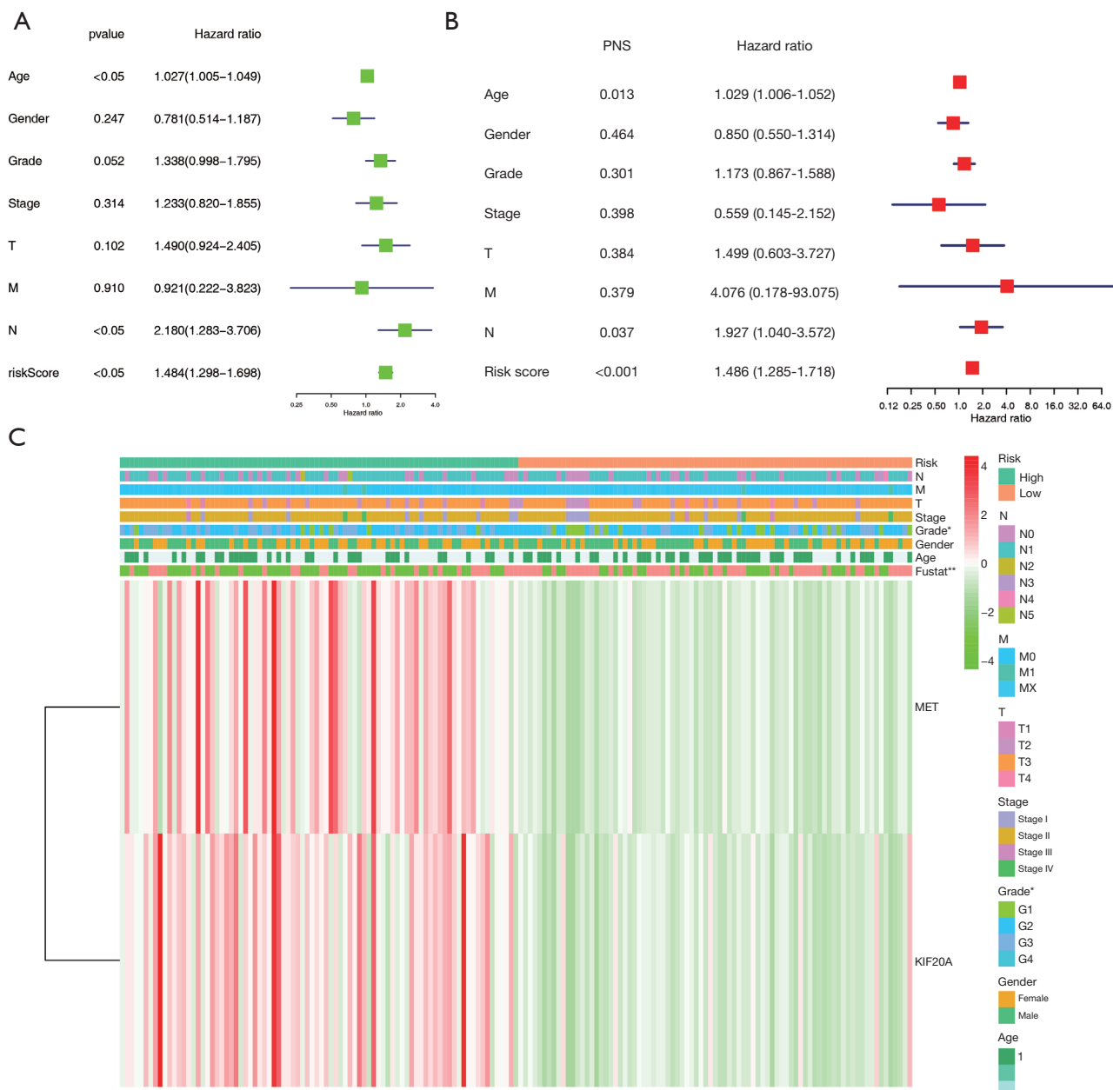


Figure 6 Forest plot and heatmap of glycolysis-related genes and clinical risk factors. (A) Forest plot of univariate Cox regression analysis in pancreatic cancer (PC). (B) Forest plot of multivariate Cox regression analysis in PC. (C) Heatmap of glycolysis-related genes and clinical risk factors. *, P<0.05; **, P<0.01.

microenvironment is the main cause of high mortality and poor prognosis of PC. Therefore, it is very necessary to further study the tumor microenvironment of PC and actively seek effective therapeutic targets. Fibroblasts, cell matrix and other cellular components together constitute the microenvironment of PC, which has a positive effect on

tumor cell proliferation and migration, and can help tumor cells produce drug resistance and escape immune system surveillance (22). The poor vascular system of PC leads to a low tumor microenvironment with low blood supply, so that the glycolytic metabolism of PC cells is strengthened (23). Enhanced glycolysis marks the reprogramming of cancer

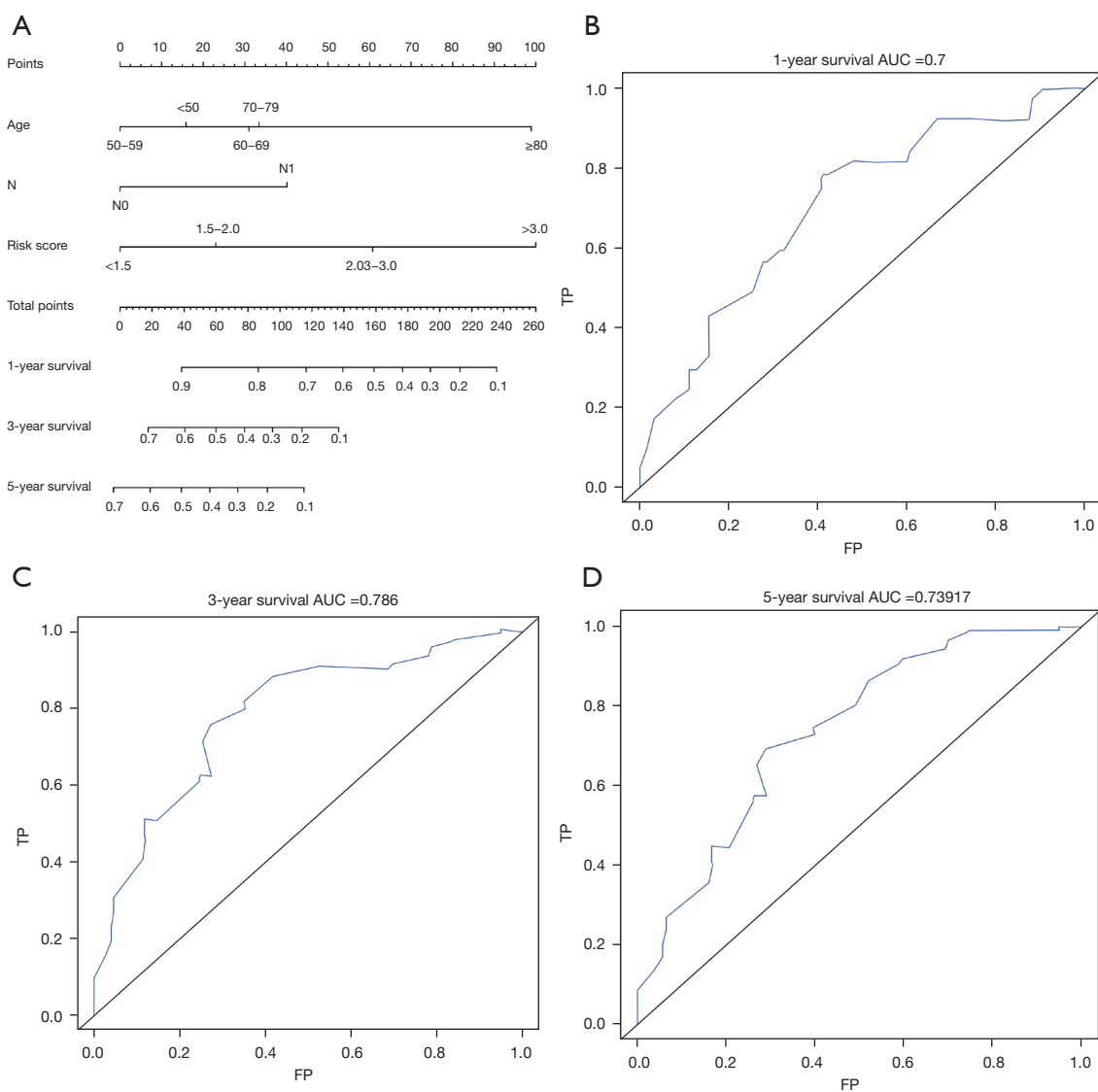


Figure 7 Construction, performance, and validation of the nomogram. (A) Nomogram to predict 1-, 3-, and 5-year survival for pancreatic cancer (PC) patients. (B) Receiver operating characteristic (ROC) curve analysis to assess the 1-year survival of PC patients. (C) Three-year survival of PC patients. (D) Five-year survival of PC patients. TP, true positive; FP, false positive.

metabolism, which further leads to the resistance of cancer cells to chemotherapeutic drugs, thus maintaining a high value-added rate (24).

GSEA can integrate data from different levels and sources, enabling comprehensive evaluation of genome-wide expression profile chip data. In this study, mRNA expression profile data of 185 patients with PC were used to conduct GSEA. One genome with a P value <0.05 showed a significant difference and was selected for

subsequent analysis. Univariate, multivariate, and LASSO Cox regression analyses were performed to determine the prognostic genes of patients with PC. Based on 9 of the most significant biomarkers, we developed and validated a model for predicting clinical outcomes in patients with PC. Survival analysis showed significant differences in prognosis between high-risk and low-risk PC patients. In addition, multivariate Cox analysis showed that the prediction model of PC patients can be used as an independent prognostic

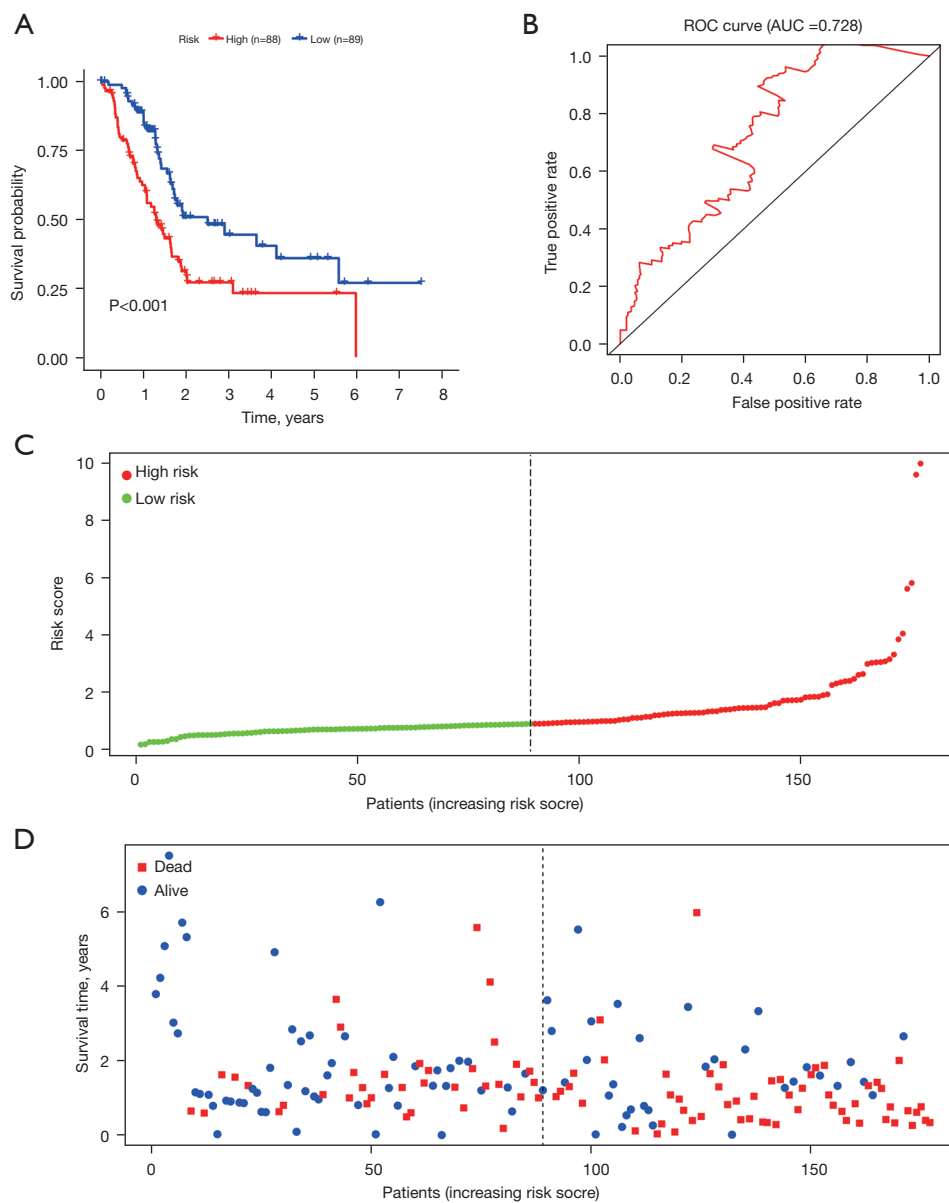


Figure 8 Kaplan-Meier (KM) survival analysis and risk score assessment by the glycolysis related genes (GRG)-related gene signature and time-dependent receiver operating characteristic (ROC) curves in The Cancer Genome Atlas (TCGA) cohort. (A) KM survival analysis of high- and low-risk samples. (B) ROC curve for overall survival. (C) Risk score distribution. (D) Survival status.

tool. We also found that patients with higher risk scores in our predictive model tended to be older, had advanced disease, and had a poorer prognosis. Compared with traditional clinical factors, the prediction model in this study has similar or better clinical application potential. In addition, we constructed a new nomogram combining the predictive model and clinical features. The factors utilize

complementary values of clinical features and predictive models and provide a superior estimate of OS. The results showed that the ROC curve performed well in our study. In addition, genetic markers can further classify clinically defined patient groups (e.g., by age, stage, T/N/M stage) into subgroups with different survival outcomes.

The abnormal expression of certain special genes may be

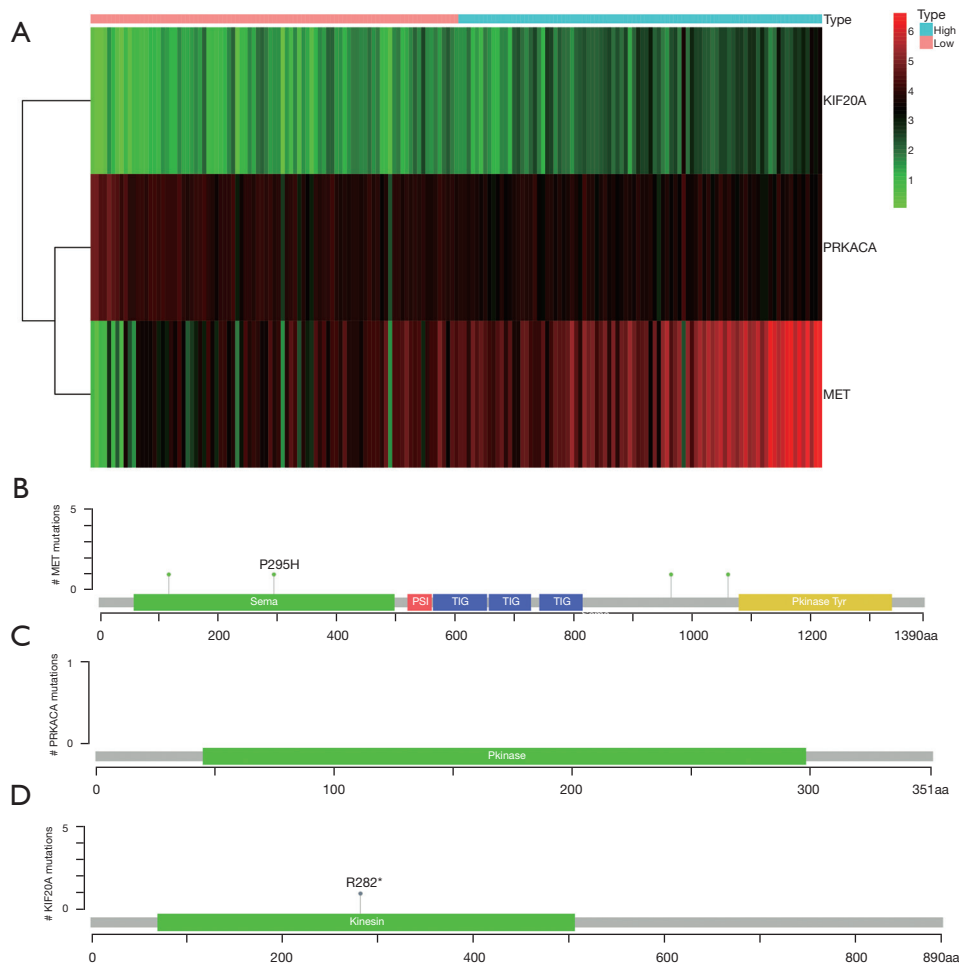


Figure 9 Glycolysis-related genes (GRG) expression patterns for patients in high- and low-risk groups determined by the 3-GRG signature. (A) Heat map of MET, PRKACA and KIF20A in high and low risk groups. (B) Mutation of MET. (C) Mutation of PRKACA. (D) Mutation of KIF20A. *, $P < 0.05$.

an important factor affecting the prognosis of PC patients, and it is also expected to become an effective target for individualized treatment. In recent years, with the rapid development of sequencing technology, high-throughput genomics can effectively mine the key genes involved in tumorigenesis and development, and further analyze their related mechanisms. Prognosis and risk signature analysis indicated that *MET* and *KIF20A* are important genes related to PC. It is an effective target for PC treatment and improving prognosis. *MET* is a proto-oncogene, encoding tyrosine kinase, and is highly expressed in many cancers including human PC (25). *Met* also affects the progress of tumor and the prognosis of patients. At the same time, for the research and development of targeted drugs for solid tumor treatment, c-Met has become the focus of researchers

and medical staff (26). The results of Taniuchi *et al.* Showed that *KIF20A* expression was increased in PC, and the down-regulation of *KIF20A* significantly reduced the proliferation of PC cells, which confirmed for the first time that *KIF20A* had the characteristics of oncogene (27). Our findings suggested that *KIF20A* and *MET* are significantly associated with the prognosis of PC, and may be effective targets for the treatment and prognosis of PC.

Although bioinformatics techniques have been used to identify candidate glycolytic genes associated with PC prognosis in a large number of bioinformatics studies, there are still limitations in this study. It is not enough to use only bioinformatics analysis to obtain and verify the results, but these findings also need to be further studied through clinical trials.

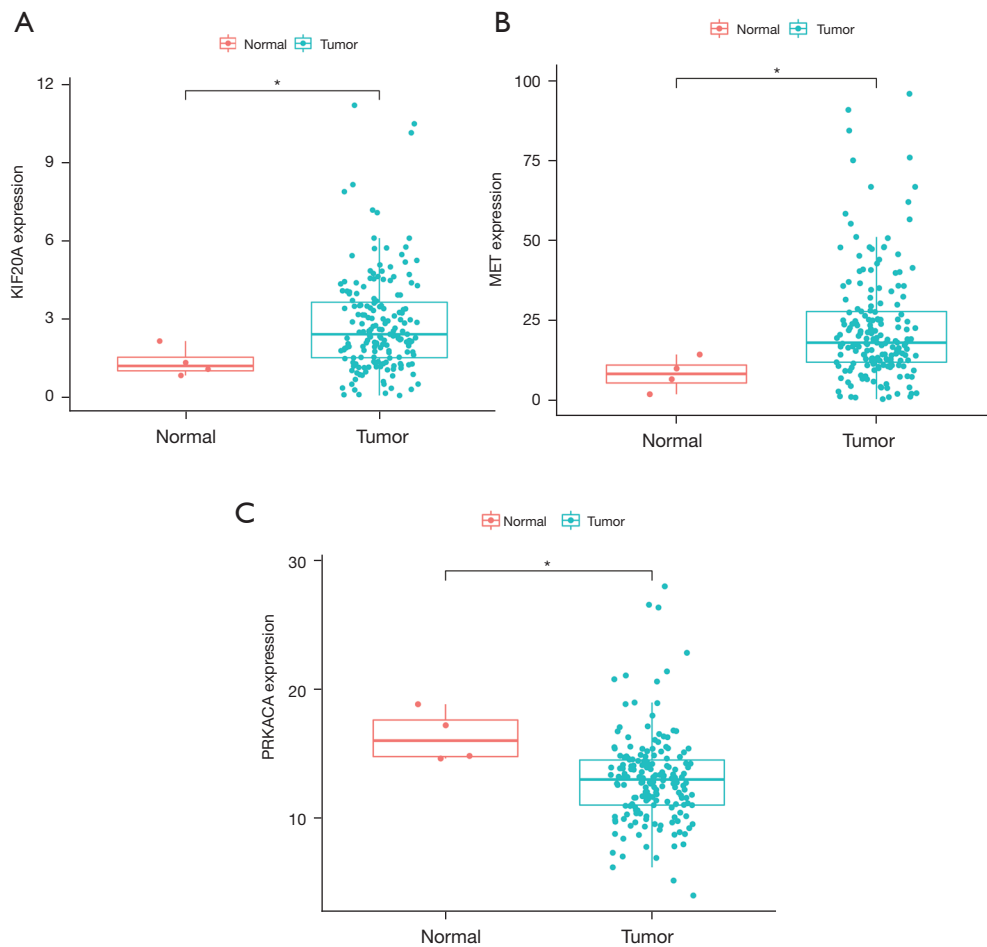


Figure 10 Comparison of the expression of 3 key genes. *, P<0.05.

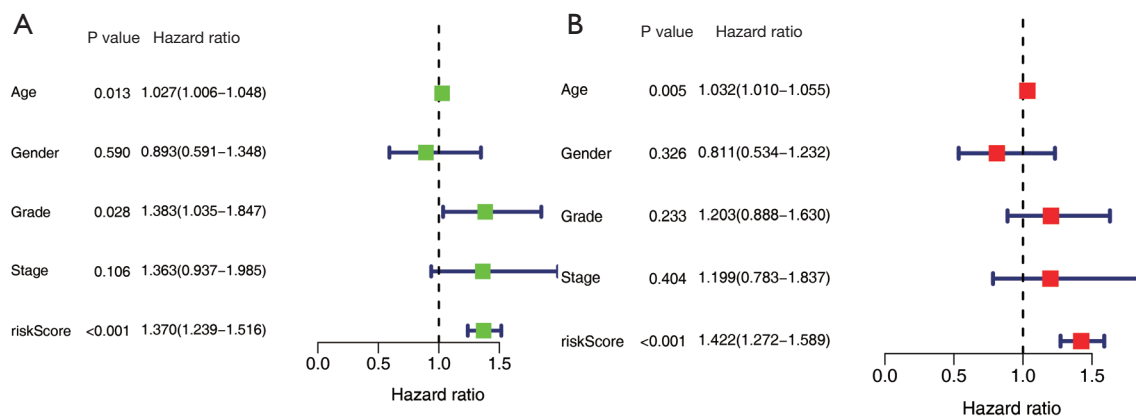


Figure 11 Forest plot and heatmap of glycolysis-related genes and clinical risk factors. (A) Forest plot of univariate Cox regression analysis in PC. (B) Forest plot of multivariate Cox regression analysis in pancreatic cancer (PC).

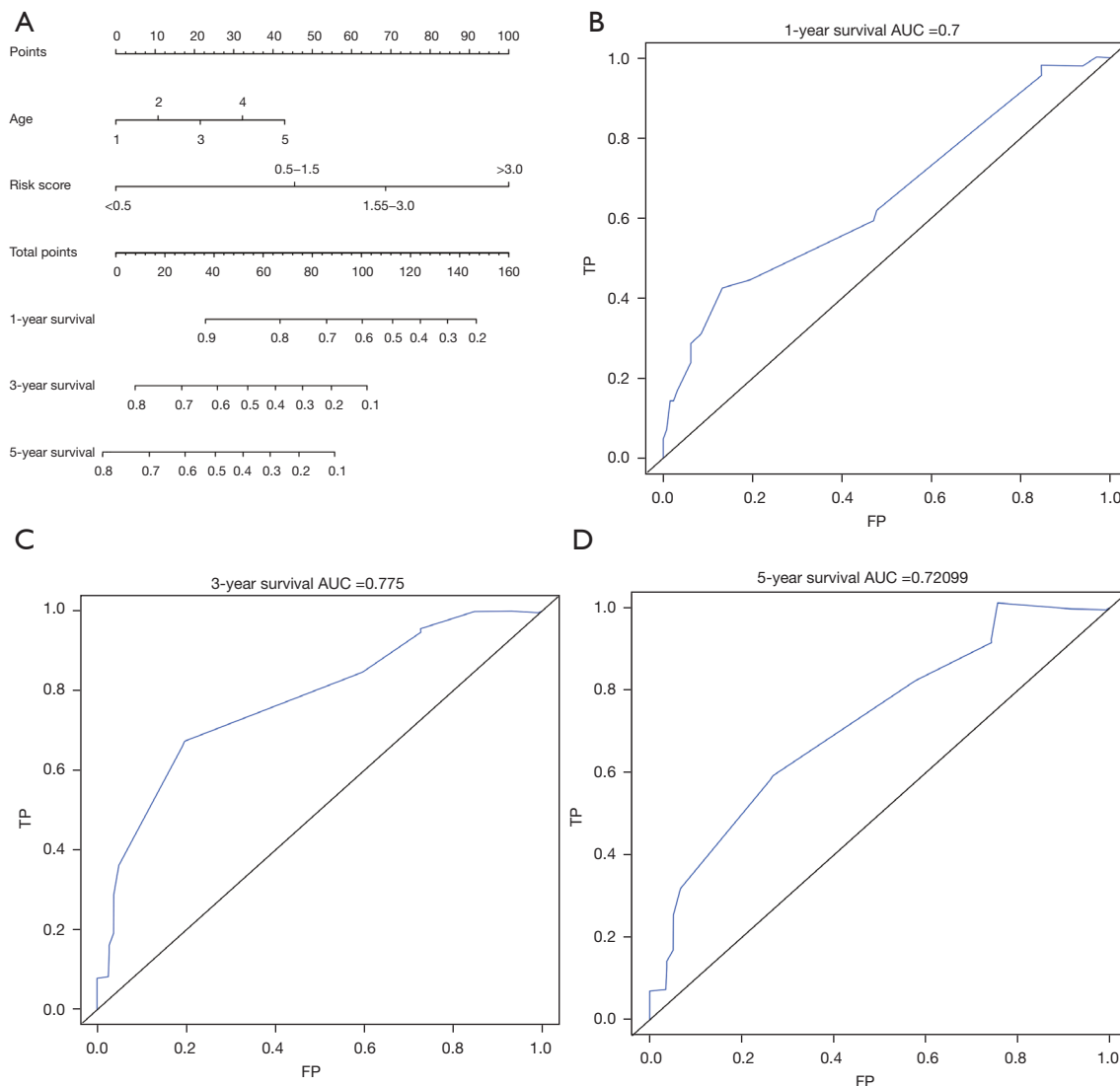


Figure 12 Construction, performance and verification of nomogram. (A) Nomogram to predict the 1-, 3-, and 5-year survival of pancreatic cancer (PC) patients. (B) Receiver operating characteristic (ROC) analysis to assess the 1-year survival of PC patients. (C) Three-year survival of PC patients. (D) Five-year survival of PC patients. TP, true positive; FP, false positive.

Conclusions

In this study, we comprehensively analyzed the relationship between the expression of GRGs and the occurrence and prognosis of PC. Among the GRGs, abnormal expression of *MET* and *KIF20A* was found to be significantly associated

with PC progression, and *MET* and *KIF20A* were identified as key regulators that independently predicted prognosis in patients with PC. This study highlights the important role of glycolysis genes in the development and progression of PC and provides potential biomarkers which will aid in the selection of therapeutic approaches.

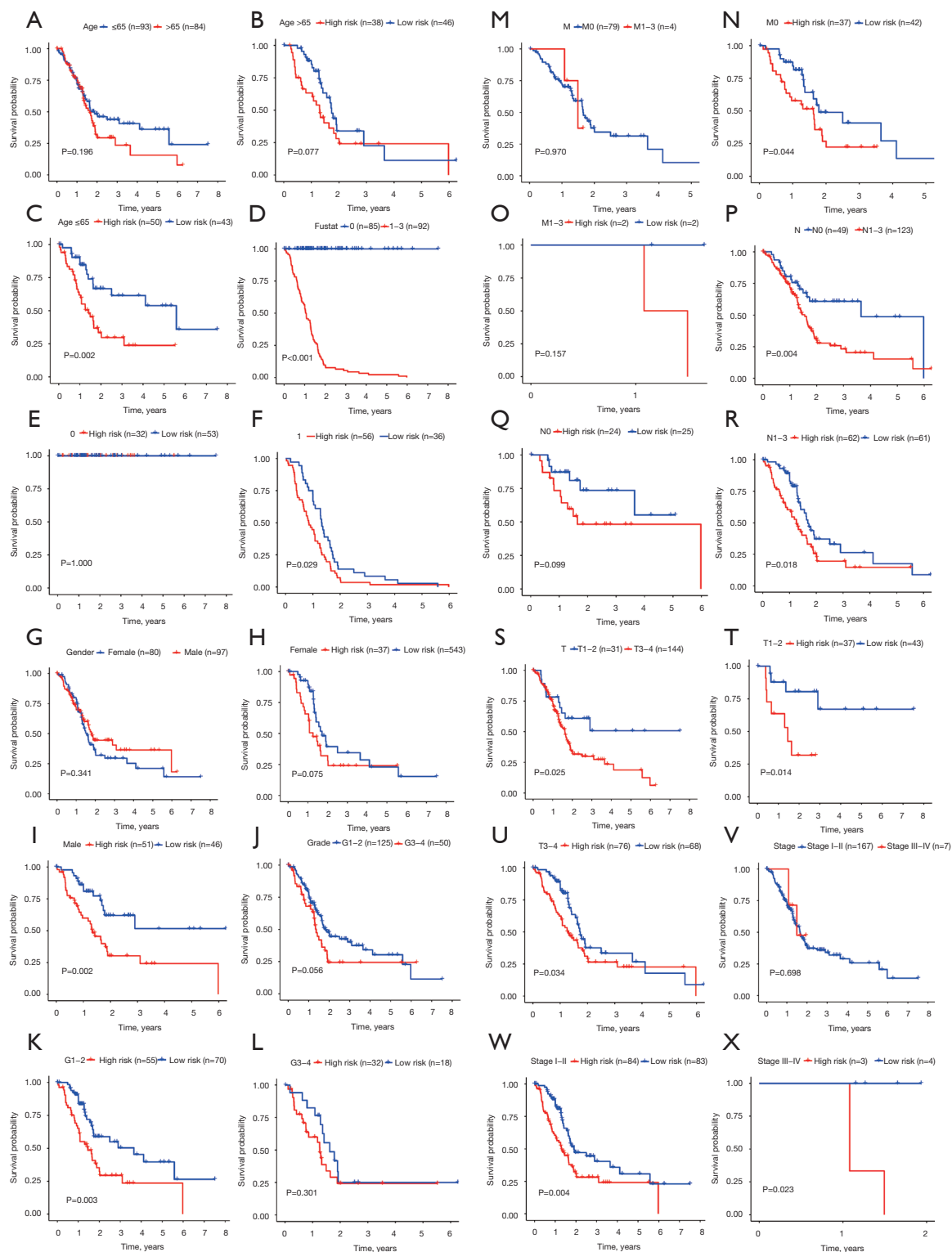


Figure 13 Kaplan-Meier survival analysis for pancreatic cancer (PC) patients with clinical features. (A-C) Age-dependent survival curve. (D) Time-dependent survival curve. (E,F) Risk-dependent survival curve. (G-J) Gender-dependent survival curve. (K,L) Grade-dependent survival curve. (M-O) M stage-dependent survival curve. (P-R) N stage-dependent survival curve. (S-U) T stage-dependent survival curve. (V-X) Clinical stage-dependent survival curve.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the STREGA reporting checklist. Available at <https://jgo.amegroups.com/article/view/10.21037/jgo-22-17/rc>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jgo.amegroups.com/article/view/10.21037/jgo-22-17/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Kleeff J, Korc M, Apte M, et al. Pancreatic cancer. *Nat Rev Dis Primers* 2016;2:16022.
- Siegel RL, Miller KD, Fuchs HE, et al. Cancer Statistics, 2021. *CA Cancer J Clin* 2021;71:7-33.
- Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin* 2016;66:115-32.
- Feig C, Gopinathan A, Neesse A, et al. The pancreas cancer microenvironment. *Clin Cancer Res* 2012;18:4266-76.
- Sohal DPS, Duong M, Ahmad SA, et al. Efficacy of Perioperative Chemotherapy for Resectable Pancreatic Adenocarcinoma: A Phase 2 Randomized Clinical Trial. *JAMA Oncol* 2021;7:421-7.
- Liu J, Luo X, Guo R, et al. Cell Metabolomics Reveals Berberine-Inhibited Pancreatic Cancer Cell Viability and Metastasis by Regulating Citrate Metabolism. *J Proteome Res* 2020;19:3825-36.
- Fischer CG, Wood LD. From somatic mutation to early detection: insights from molecular characterization of pancreatic cancer precursor lesions. *J Pathol* 2018;246:395-404.
- WARBURG O. On the origin of cancer cells. *Science* 1956;123:309-14.
- Jeong H, Kim S, Hong BJ, et al. Tumor-Associated Macrophages Enhance Tumor Hypoxia and Aerobic Glycolysis. *Cancer Res* 2019;79:795-806.
- Wu L, Hollinshead KER, Hao Y, et al. Niche-Selective Inhibition of Pathogenic Th17 Cells by Targeting Metabolic Redundancy. *Cell* 2020;182:641-654.e20.
- Zhao C, Gao F, Li Q, et al. The Distributional Characteristic and Growing Trend of Pancreatic Cancer in China. *Pancreas* 2019;48:309-14.
- Stopa KB, Kusiak AA, Szopa MD, et al. Pancreatic Cancer and Its Microenvironment—Recent Advances and Current Controversies. *Int J Mol Sci* 2020;21:3218.
- Robinson GL, Dinsdale D, Macfarlane M, et al. Switching from aerobic glycolysis to oxidative phosphorylation modulates the sensitivity of mantle cell lymphoma cells to TRAIL. *Oncogene* 2012;31:4996-5006.
- Xiang J, Hu Q, Qin Y, et al. TCF7L2 positively regulates aerobic glycolysis via the EGLN2/HIF-1 α axis and indicates prognosis in pancreatic cancer. *Cell Death Dis* 2018;9:321.
- Wang H, Lengerich BJ, Aragam B, et al. Precision Lasso: accounting for correlations and linear dependencies in high-dimensional genomic data. *Bioinformatics* 2019;35:1181-7.
- Liu J, Sun G, Pan S, et al. The Cancer Genome Atlas (TCGA) based m6A methylation-related genes predict prognosis in hepatocellular carcinoma. *Bioengineered* 2020;11:759-68.
- Thomas MA, Yang L, Carter BJ, et al. Gene set enrichment analysis of microarray data from *Pimephales promelas* (Rafinesque), a non-mammalian model organism. *BMC Genomics* 2011;12:66.
- Liu P, Jiang W, Ren H, et al. Exploring the Molecular Mechanism and Biomarkers of Liver Cancer Based on Gene Expression Microarray. *Pathol Oncol Res* 2015;21:1077-83.
- Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide

- experimental datasets. *Nucleic Acids Res* 2019;47:D607-13.
20. Bhutani MS, Klapman JB, Tuli R, et al. An open-label, single-arm pilot study of EUS-guided brachytherapy with phosphorus-32 microparticles in combination with gemcitabine +/- nab-paclitaxel in unresectable locally advanced pancreatic cancer (OncoPaC-1): Technical details and study protocol. *Endosc Ultrasound* 2020;9:24-30.
 21. Yang X, Yu DD, Yan F, et al. The role of autophagy induced by tumor microenvironment in different cells and stages of cancer. *Cell Biosci* 2015;5:14.
 22. Hosein AN, Brekken RA, Maitra A. Pancreatic cancer stroma: an update on therapeutic targeting strategies. *Nat Rev Gastroenterol Hepatol* 2020;17:487-505.
 23. Akakura N, Kobayashi M, Horiuchi I, et al. Constitutive expression of hypoxia-inducible factor-1alpha renders pancreatic cancer cells resistant to apoptosis induced by hypoxia and nutrient deprivation. *Cancer Res* 2001;61:6548-54.
 24. Wen YA, Xiong X, Scott T, et al. The mitochondrial retrograde signaling regulates Wnt signaling to promote tumorigenesis in colon cancer. *Cell Death Differ* 2019;26:1955-69.
 25. Liu X, Yao W, Newton RC, et al. Targeting the c-MET signaling pathway for cancer therapy. *Expert Opin Investig Drugs* 2008;17:997-1011.
 26. Mo HN, Liu P. Targeting MET in cancer therapy. *Chronic Dis Transl Med* 2017;3:148-53.
 27. Taniuchi K, Nakagawa H, Nakamura T, et al. Down-regulation of RAB6KIFL/KIF20A, a kinesin involved with membrane trafficking of discs large homologue 5, can attenuate growth of pancreatic cancer cell. *Cancer Res* 2005;65:105-12.
- (English Language Editor: C. Betlazar-Maseh)

Cite this article as: Zhang J, Liu Z, Zhang Z, Tang R, Zeng Y, Chen P. Identification of a glycolysis-related gene signature for predicting pancreatic cancer survival. *J Gastrointest Oncol* 2022;13(1):380-399. doi: 10.21037/jgo-22-17