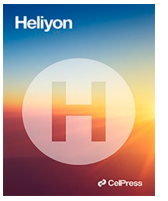
Heliyon 10 (2024) e31302

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

A prognostic binary classifier comprised of five critical mRNAs stratified pancreatic cancer patients following resection

Yueqing Lu, Tong Zhou, Mingshu Lu *

Hepatobiliary and Vascular Surgery, People's Hospital Affiliated to Shandong First Medical University, 271199, Shandong Province, China

ARTICLE INFO

Keywords:
Pancreatic cancer
PDAC
Classifier
Prognostic model
mRNA

ABSTRACT

Background: Pancreatic cancer is characterized by an extremely poor prognosis, even following potentially curative resection. Classical prognostic markers such as histopathological or clinical parameters have limited predictive power. The present study aimed to establish a prognostic model combining mRNA expression data with histopathological and clinical data to better predict survival and stratify pancreatic cancer patients following resection. We pioneered three models in one study and systematically evaluated the clinical benefits of all three models.

Methods: To identify differentially expressed genes in pancreatic cancer, mRNA data from normal (GTEx database) and pancreatic cancer (TCGA database) tissues were used. Survival analysis was carried out to identify prognosis-relevant genes from the identified differentially expressed genes and LASSO regression was used to filter out hub genes. The risk score of several hub genes was calculated according to gene expression and coefficients. Validation was carried out using an independent set of GEO microarray data. Multivariate COX regression was used for identifying independent clinical and pathological risk factors related to patient's survival in the TCGA database and a prognostic model combining mRNA expression data with histopathological and clinical data was established. Another prognostic model using clinicopathological factors from the SEER database was conceived based on multivariate COX regression. NRI (net reclassification improvement) and IDI (integrated discrimination index) were used to compare the predictive capabilities of the different models.

Results: We identified 1589 differentially expressed genes (DEGs) through the comparison of normal and pancreatic cancer tissues, of whom 317 were associated with prognosis($p < 0.05$). LASSO regression identified five hub genes, MYEOV, ANXA2P2, MET, CEP55, and KRT7, that were used for the five-mRNA-classifier prognostic model. The classifier could stratify patients into a short and long survival group: 5-year overall survival in the training set (TCGA, 6 % vs 52 %, $p < 0.001$), test set (TCGA, 18 % vs 55 %,$p < 0.01$) and external validation set (GEO, 0 % vs 25 %, $p < 0.05$). Sensitivity analysis showed that the mRNA model (model 1) was better than the clinicopathological no-mRNA model (model 2) in predicting 5-year survival in the TCGA database (AUC: 0.877 vs 0.718, $z = 3.165$, $p < 0.01$) and better than the multi-factor prognostic model (model 3) from the SEER database (AUC: 0.754, $z = 2.637$, $p < 0.01$). On predictive performance, model 1 improved model 2 (NRI = 0.084, $z = 1.288$, $p = 0.198$; IDI = 0.055, $z = 1.041$,$p = 0.298$) and model 3 (NRI = 0.167,$z = 1.961$,$p = 0.05$; IDI = 0.086, $z = 1.427$, $p = 0.154$).

* Corresponding author. No. 001, Xuehu Street, Changshao North Road, Laiwu District, Jinan City, China.
E-mail addresses: 1066746717@qq.com (Y. Lu), 23025825@qq.com (T. Zhou), lsmingshu@hotmail.com (M. Lu).

https://doi.org/10.1016/j.heliyon.2024.e31302
Received 14 November 2023; Received in revised form 8 May 2024; Accepted 14 May 2024
Available online 17 May 2024

Conclusion: The five-mRNA-classifier is a reliable and feasible instrument to predict the prognosis of pancreatic cancer patients following resection. It might help in patiens counseling and assist clinicians in providing individualized treatment for patients in different risk groups.

1. Introduction

Pancreatic cancer(PC) (in this article refers specifically to "pancreatic ductal adenocarcinoma", PDAC), is a treacherous gastrointestinal malignancy with rapid progress, miserable treatment effect, and dismal prognosis [1,2]. On account of its non-specific symptoms and scarcity of sensitive and specific tumor markers, the patients are difficult to be diagnosed in the early stage. Approximately 55 % of them come to pass distant metastases [3]. Since PDAC exerts violent resistance against chemotherapy and radiotherapy [4], improving the survival rate is in thorny dilemma, with around 9 % of the 5-year survival rate [5]. Curative resection is currently the most effective treatment option, which can notably elevate 20 %–30 % of the 5-year survival rate. Nevertheless, less than 20 % of patients are eligible for this strategy, as most patients have triggered metastasis [6].

The American Joint Committee on Cancer (AJCC) staging system is now extensively employed clinically to assess the prognosis of PC patients [7]. However, the patients with the same stage can have significantly different outcomes in clinical practice. As such, urgent need for effective models to assess PC prognosis and facilitate tailored treatment are required. In recent years, with development of microarray and highthroughput sequencing, studies have focused on mRNAs for the prediction of tumor prognosis. Tian Lan et al. [8] identified a 5-mRNA signature consisted of SLC26A9, SINHCAF, MICB, KRT19, and MT1X for predicting prognosis of Esophageal adenocarcinoma (EAC). In addition, Bo Wu et al. [9] developed an eight-mRNA signature that categorized osteosarcoma patients into high-risk and low-risk groups with significantly different overall survival times. ROC demonstrating the high sensitivity and specificity of the eight-mRNA signature in predicting prognosis for osteosarcoma patients. Hubin Yin et al. [10] built an integrated 13-mRNA signature for prediction of progression of non-muscle invasive bladder cancer (NMIBC) to muscle invasive bladder cancer (MIBC). A nomogram based on 13-mRNA signature might be useful for clinicians to select personalized therapy for patients with bladder cancer.

But there are few mRNA models to assess PC prognosis. Because of a complicated composition, difficultly evaluated clinical utility and no comparisons among models, we put forward a novel binary classifier comprised of five-mRNA risk score, which stratified PC patients into a high-risk group and a low-risk group in terms of survival time after operation. The mRNA data were from The Cancer

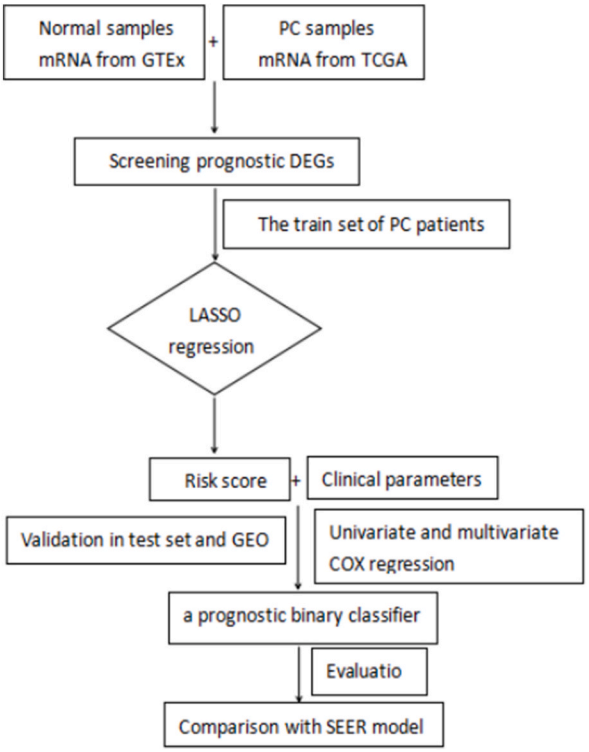


Fig. 1. A flow diagram of the prognostic binary classifier. GTEx = The Genotype-Tissue Expression, TCGA = The Cancer Genome Atlas, PC=Pancreatic Cancer, DEGs = differentially expressed genes, LASSO=The Least Absolute Shrinkage and Selection Operator, GEO=Gene Expression Omnibus.

Genome Atlas (TCGA) and The Genotype-Tissue Expression (GTEx) databases. At the same time, we verified it in an external independent cohort from Gene Expression Omnibus(GEO) and revolutionized its performance from the perspective of patient's profits. Additionally, through comparing it with conventional prognostic pattern based on the AJCC staging system, our five-mRNA model highlights its promise on clinical decisions making and individualized patient management.

2. Methods

2.1. Patients sources and data processing

The elaborate flow diagram of this project is figured out in Fig. 1. Variables annotation for multivariate analysis was in Table 2. We downloaded the fragments per kilobase million(FPKM) gene expression RNAseq data and clinical data of normal healthy individuals from The Genotype-Tissue Expression (GTEx) databases through UCSC Xena (http://xena.ucsc.edu/). FPKM gene expression RNAseq data, clinical and survival data of PC patients were downloaded from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/). The gene name was converted from the Ensembl id to the matrix of the gene symbol through the Ensembl database (http://asia.ensembl.org/index.html). Only mRNAs expression value > 0.5 in all PC samples were enrolled for analysis. Expression was

Table 1
The demographic and clinicopathological factors in TCGA and SEER database.

	Category	TCGA(n = 176)		SEER(n = 9347)			
		Training set(n = 124)	Test set(n = 52)	Training set(n = 6544)	Test set(n = 2803)	P1 value	P2 value
median survival time (years)		1.3	1.3	1.7	1.8	0.917	0.006
dead(%)		54.8	46.2	60.7	60.6	0.293	0.881
age(years)	median	65	65	66	66	0.879	0.203
	>74	24	11	–	–	0.785	
	≤74	100	41	–	–		
	>69	–	–	2449	1058		0.769
	≤69	–	–	4095	1745		
gender	male	71	25	3367	1424	0.264	0.565
	female	53	27	3177	1379		
race	white	–	–	5369	2292		0.751
	others	–	–	1175	511		
site	head	94	34	4705	2002	0.157	0.641
	body and tail	30	18	1839	801		
grading	G1,G2	90	36	4400	1862	0.653	0.446
	G3,G4	34	16	2144	941		
stage	I, II	107	49	5904	2550	0.130	0.256
	III,IV	17	3	640	253		
T	T1,T2	22	9	1414	646	0.945	0.124
	T3,T4	102	43	5130	2157		
N	N0	36	13	2492	1088	0.586	0.503
	N1	88	39	4052	1715		
M	M0	59	20	6131	2660	0.404	0.024
	M1	2	2	413	143		
	Mx	63	30	–	–		
ELN[a]	>21	36	9	–	–	0.104	
	≤21	88	43	–	–		
	>12	–	–	4041	1759		0.360
	≤12	–	–	2503	1044		
PLN*	>2	54	23	–	–	0.934	
	≤2	70	29	–	–		
	>3	–	–	1554	710		0.102
	≤3	–	–	4990	2093		
size(cm)	>4	38	16	–	–	0.987	
	≤4	86	36	–	–		
	>2.3	–	–	4893	2097		0.966
	≤2.3	–	–	1651	706		
Alcohol history	YES	76	36	–	–	0.318	
	NO	48	16	–	–		
Immune score	>-265.06	110	46	–	–	0.962	
	<=-265.06	14	6	–	–		
Resection status	R0	81	38	–	–	0.316	
	R1, R2	43	14	–	–		
Tumor types	PDAC*	111	50	–	–	0.150	
	others	13	2	–	–		

[a] ELN:examined_lymph_nodes.PLN:positive_lymph_nodes. p1,p2 value respectively represented the difference between training set and test set on various variables in TCGA and SEER database.

Table 2
Index annotation of univariate and multivariate regression analysis.

Variable	TCGA		SEER	
	1	0	1	0
age(years)	>74	≤74	>69	≤69
gender	male	female	male	female
race	–	–	white	others
site	head	body and tail	head	body and tail
DG[a]	G3,G4	G1,G2	G3,G4	G1,G2
stage	III,IV	I, II	III,IV	I, II
pT[a]	T3,T4	T1,T2	T3,T4	T1,T2
pN[a]	N1	N0	N1	N0
pM[a]	M1, Mx	M0	M1	M0
ELN[a]	>21	≤21	>12	≤12
PLN[a]	>2	≤2	>3	≤3
size(cm)	>4	≤4	>2.3	≤2.3
alcohol	YES	NO	–	–
immunescore	>-265.06	<=-265.06	–	–
RE[a]	R1, R2	R0	–	–
IDC[a]	YES	NO	YES	NO
riskscore	>0.4	≤0.4	–	–
status	dead	alive	dead	alive

[a] DG:differentiated_grade.pT:pathologic_T.pN:pathologic_N.pM:pathologic_M.ELN:examined_lymph_nodes.PLN:positive_lymph_nodes.RE:residual_edge.IDC:infiltrating_duct_carcinoma.

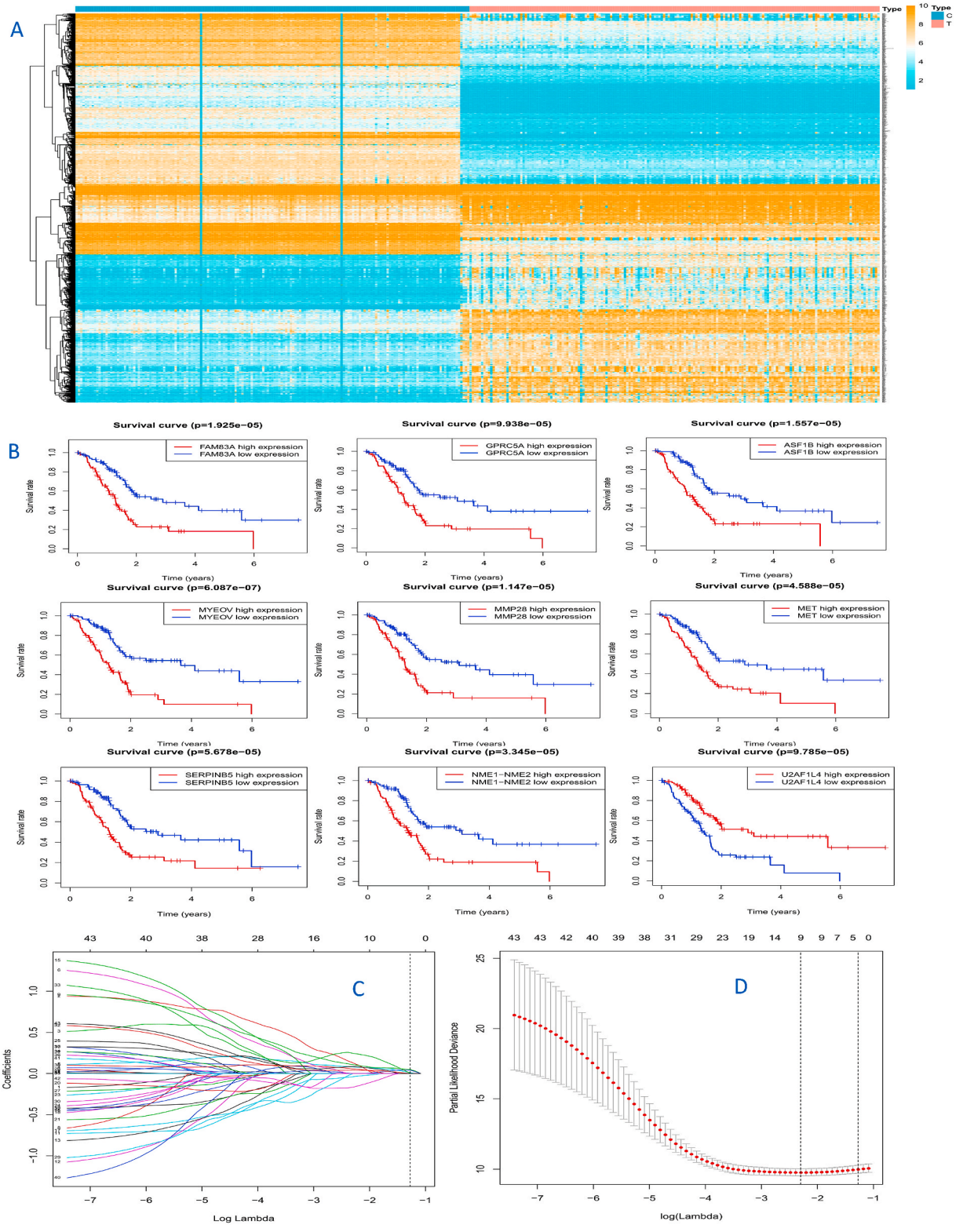
defined as log2 (x+ 1), while x represented the FPKM value. The inclusion criteria were set as follows: 1) Diagnosis of pancreatic cancer; 2) the samples were recorded with complete RNAseq data; 3) the samples were recorded with detailed clinicopathological data including pathological stage, grade, overall survival (OS). We finally obtained 171 normal samples from GTEx and 178 PC samples from TCGA. Table 1 displayed basic information of PC patients from TCGA.

2.2. Construction of the mRNA-related prognostic model

Differentially expressed genes (DEGs) were identified using the R package "limma"in PC versus normal tissue. Selection criteria for significant DEGs were $P < 0.05$ and $|\log FC| > 2$. Then, remove two samples with mRNA average expression below 0.5 in all genes. To further screen DEGs correlating with prognosis, batch survival analysis was performed by R package "survival" in PC samples. Only prognostic DEGs with $P < 0.05$ were considered as statistically significant mRNAs. Those of them ($P < 0.001$) were used for further Least Absolute Shrinkage and Selection Operator (LASSO) regression. It could reduce the model complexity and avoid the overfitting issues. We divided casually the TCGA datasets into train set (70 %,124 samples) and test set (30 %,52 samples) using R package "createDataPartition". LASSO was performed through R software version 3.4.2 and the "glmnet" package, to screen and confirm the selected DEGs in the train set. Risk scores were calculated with the expression level of selected hub prognostic DEGs weighted by their regression coefficients. The formula was set as follow: Risk score = $\alpha 1 \times$gene 1 $+\alpha 2 \times$gene 2 + … $+\alpha n \times$ gene n, where α is the coefficient of each gene, and gene indicates relative expression of gene. On the protein level, we validated the expressions of selected hub genes in Human Protein Atlas (HPA, https://www.proteinatlas.org). It provided immunohistochemistry (IHC) results using a tissue microarray (TMA)-based analysis of the corresponding proteins in PC patients and normal tissues.

2.3. The functional analysis of prognostic DEGs and stratification potential of risk score

We selected the optimum cutoff score for the hub-mRNA risk score using X-tile plots based on the association with survival time. The X-tile software 3.6.1 working depended on the highest χ^2value and minimum p value, defined by Kaplan-Meier survival analysis and log-rank test. the The optimum cutoff score diveded PC patients into a high-risk group and a low-risk group. Gene Set Enrichment Analysis (GSEA) was used to identify hallmark gene sets associated with high-risk populations based on risk score. (https://www.gsea-msigdb.org/gsea/index.jsp). To dissect the functional roles of the prognostic-related DEGs, we, respectively, performed the Gene Ontology(GO) and Kyoto Encyclopedia of Genes and Genomes(KEGG) enrichment analyses. We evaluated the stratification potential of our mRNA prognostic model in PC patients cohort by the optimum cutoff risk score. While, we made a validation on TCGA train set, TCGA test set, and GEO set. The mRNA expression profile of GSE62452 and GSE28735 were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). GSE62452 performed on the platform of Affymetrix Human Gene 1.0 ST Array contained gene expression profile and survival time of 69 pancreatic tumor and adjacent non-tumor tissues. GSE28735 calculated on the platform of Affymetrix Human Gene 1.0 ST Array contained gene expression profile and survival time of 45 pairs of pancreatic tumor and adjacent non-tumor tissues. Samples with both gene expression data and survival time data were included in the study.



(caption on next page)

Fig. 2. Obtain five premium mRNA genes for establishing the predictive prognostic model. **(A)**Heatmap: DEGs between the control group and tumor group. **(B)**Survival curve: extremely relational nine prognostic DEGs ($P < 0.0001$). **(C)**LASSO consequence: from 43 prognostic DEGs to 5 prognostic DEGs. **(D)**10-fold cross-validation: make sure to get 5 vintage prognostic DEGs and minimal penalty factor "lambda". The rightest dashed line in both pictures represented the least model genes as well as one standard error away(1 se) minimal lambda, mini lambda = 0.1004896, lambda.1se = 0.2796182.

2.4. Independent risk factors for PC prognosis and a binary classifier with multiple factors

We screened independent risk factors affecting pancreatic cancer prognosis by univariate and multivariate COX regression analysis. Pancreatic cancer development is dependent on its specific tumor microenvironment. The local immunosuppression and inflammatory response induced by massive immune cell infiltration are an important feature of pancreatic cancer microenvironment. Therefore, immune score was also enrolled into the regression analysis, which was from Estimation of STromal and Immune cells in Malignant Tumor tissues using Expression data(ESTIMATE, https://bioinformatics.mdanderson.org/estimate/). Besides it, other clinical parameters were also used for univariate and multivariate analysis by R package "survival". In a multivariate analysis, those parameters with $P < 0.05$ constituted a binary classifier. The Index of concordance (C-index), was used to speculate the predictive capability of the model in our study. The calibration curve and Hosmer-Lemeshow(HL) test were to estimate the probability that the predicted result was consistent with the actual result. The Receiver Operating Characteristic curve (ROC) and Area Under Curve (AUC) were to evaluate the discrimination (sensitivity and specificity) of a binary classifier. ROC was performed using R package "survivalROC".

2.5. Clinical utility of the binary classifier and subgroup analysis in some prognostic risk variables

Decision Curve Analysis (DCA) could predict clinical outcomes based on the models built with various impact factors and formulates interventions by the probability of adverse events. The Clinical Impact Curve (CIC) could execute the risk stratification of 1000 hypothetical people, computing the ratio of loss to benefit. R package "rmda" and "devtools" helped to achieve them. In order to further evaluate clinical value of the binary classifier, we performed a subgroup analysis for the prognostic risk variables selected from univariate COX regression. To assist clinicians in guiding adjuvant treatment of PC, we also carried out a subgroup analysis of postoperative chemotherapy effect by R package "survminer" "survival ".

2.6. Making a comparison between the model from SEER and the model from TCGA

The Surveillance, Epidemiology, and End Results (SEER) database was used for providing clinical data of population-based cancer registries in the United States between 1975 and 2016 downloaded by SEER*Stat 8.3.5 software. We excluded patients with incomplete clinicopathological parameters and those with missing prognostic follow-up data. R function createDataPartition could aid in measuring off intact cohort into train cohort(70 %) and test cohort(30 %), randomly with double-blind. A bit of missing value was balanced by Mean for measurement data, and by Mode for counting data. The unknown and irrelevant data were removed definitely. As Table 1 showed, total of 176 PC samples from TCGA and 9347 PC samples from SEER were respectively divided into a train set(124 TCGA,6544 SEER) and a test set(52 TCGA,2803 SEER). There were no apparent discrepancy between TCGA and SEER in median survival time and the proportion of dead cases($P > 0.05$). The other baseline clinicopathologic data roughly remained parallel variations among them. Improvement of Net Weight Classification, Reclassification Improvement Index (NRI), and Improvement of Comprehensive Discrimination, Comprehensive Discrimination Improvement Index (IDI) were helpful to judge whether one model had a more robust predictive capability than the other. The former is based on the quantification of the probability gap, and the latter is based on the quantification of the classification change gap. They worked relying on R package "foreign" and R function "NRIcalculate" and "IDIcalculate".

2.7. Statistics

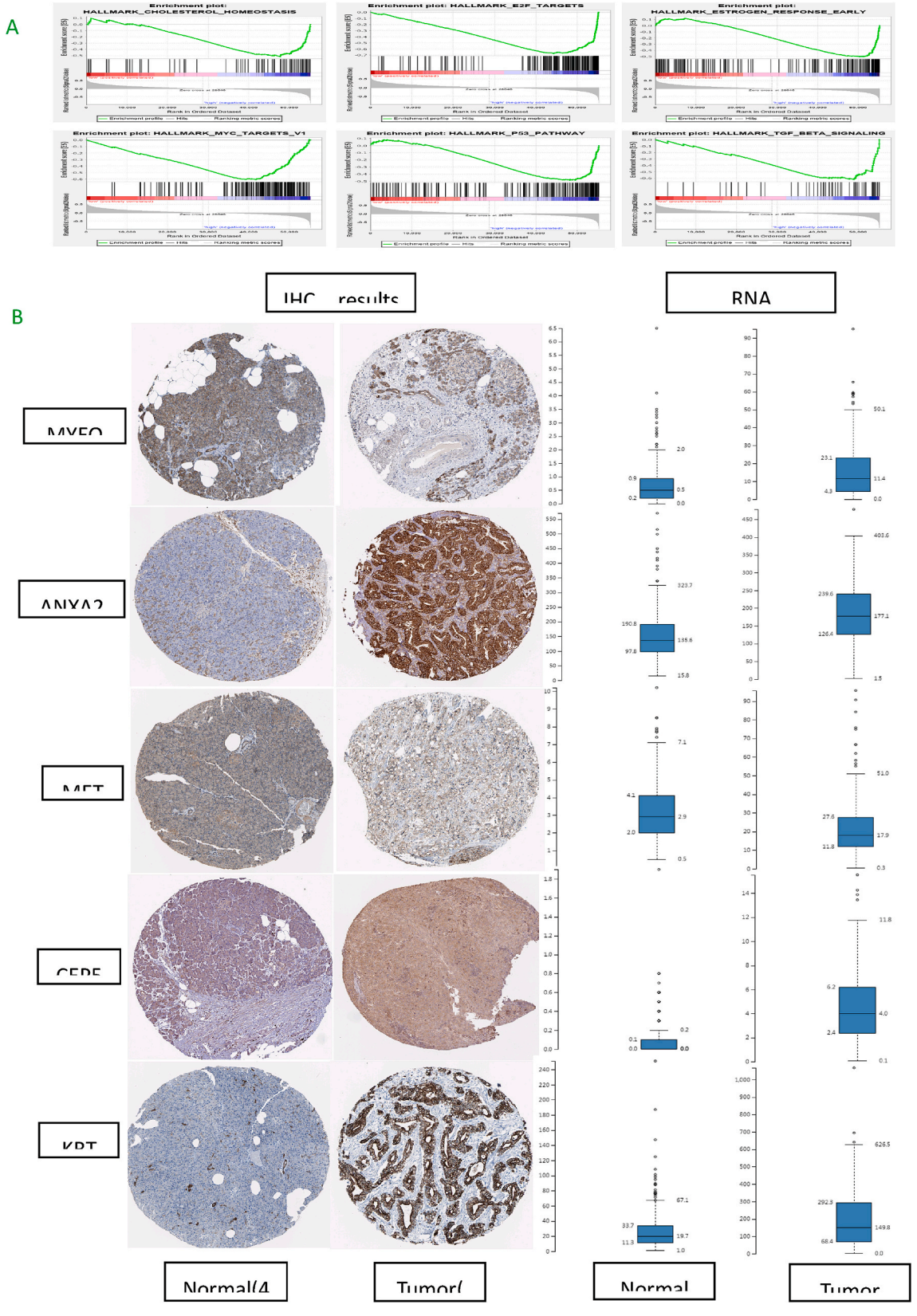
R software 3.4.2 was conducted for statistical analysis and drawing figures. It can perform Kaplan-Meier, LASSO, multivariate COX regression,ROC, calibration curves, DCA, CIC, NRI, IDI, and subgroup forest plots. Hosmer-Lemeshow(HL) test was used for detecting the accuracy of the predictive model. MedCalc software 19.3.1 was used for carrying out DeLong's test to make the comparison of multiple ROC. $P < 0.05$ was regarded as existed a signficant difference statistically.

3. Results

1 Screen differentially expressed genes (DEGs)

Amalgamating 171 normal control samples and 178 tumor samples from GTEx and TCGA, a quantity of 1589 DEGs were filtered out evidently in a condition of logFC>2 or logFC < −2 and pvalue<0.05 (Fig. 2A). Two of the 178 tumor samples were discarded for subsequent analysis, because their average expressions of all genes were low than 0.5.

2 Batch survival analysis



(caption on next page)

Fig. 3. The expression data and function of five core mRNA genes. **(A)** The remaining six enriched hall mark pathway($P < 0.05$).**(B)**Immunohistochemistry(IHC) staining results and coespondent RNA sequence data. Compared with normal samples, PC samples showed more significant expression levels about the five proteins, and the difference was statistically significant.($p < 0.05$).

To obtain DEGs closely related to survival time and status after curative pancreatectomy, the batch survival analysis was carried out in all tumor samples. 317 of the DEGs manifested significant association with patient's prognosis ($P < 0.05$), where it existed especially relevant 43 genes ($P < 0.001$). 9 DEGs, the most interrelated genes to survival rate, was presented well ($P < 0.0001$) (Fig. 2B). The remaining 34 ones were in Fig. S1.

3 LASSO regression for removing expression-related genes each other

Further, we have to disassemble the whole cohort from TCGA,176 tumor samples, into a train cohort for constructing the predictive model and test cohort for verifying internally our established model. R function createDataPartition was used to classify randomly the whole cohort as train set (70 %,124 samples) and test set (30 %,52 samples) by R software 3.6.2. To preclude overfit issue of the model, only 43 DEGs ($P < 0.001$) related to prognosis were executed LASSO regression to discard expression-related genes in the train set (Fig. 2C and D). Ultimately, five mRNA genes, myeloma overexpressed (MYEOV), annexin A2 pseudogene 2(ANXA2P2), met proto-oncogene(MET), centrosomal protein 55(CEP55), keratin 7(KRT7), were selected to cipher its risk score, a binary classifier, in line with respective coefficients. The formula is as follows: Risk score = MYEOV expression level*0.011 + ANXA2P2 expression level*0.048 + MET expression level*0.041 + CEP55 expression level*0.002 + KRT7 expression level*0.001. We selected respectively a classic staining picture from normal samples and tumor samples for each gene. The expressions of five hub genes were stronger in tumor samples than those in normal samples on a protein level (Fig. 3B).

4 Evaluate comprehensively five mRNA genes' functions and their risk score on stratification property

The 176 whole cohort patients were divided into a high-risk and low-risk group, respectively risk score >0.4 and risk score ≤ 0.4. By GSEA analysis, it was significantly discovered that patients in the high-risk group were enriched into 41/50 hallmark gene sets, with 12 obvious ones ($P < 0.05$) (Fig. 4A, Fig. 3A). However, patients in the low-risk group were enriched into 9/50 hallmark gene sets, with only one distinct gene set ($P < 0.05$). Go enrichment analysis found that those prognostic-related DEGs were involved in human immune response, and facilitated the formation of extracellular matrix(Fig. S2 A). KEGG pathway analysis confirmed that they played an important role in pancreatic secretion and protein digestion and absorption (Fig. S2 B). We stratified the whole cohort of patients according to risk score 0.4 and verified the five-mRNA expression among them (Fig. 4B). Then, the five-mRNA-risk score discrimination was evaluated accurately in the TCGA train set, TCGA test set, and GEO validation set. It was displayed to us that our risk score model has relatively robust predictive potentiality for the prognosis of pancreatic cancer patients in the train set and internal testing set (Fig. 4C, left and middle). There are prominent discrepancies in the overall survival rate between the high-risk(risk score>0.4) and the low-risk group(risk score ≤ 0.4) ($P < 0.05$). As exiting the heterogeneity and batch effects of patient samples from various sequencing platforms, the integral mRNA expression level of patients from GEO was slightly higher than ones from TCGA. It suggested that our selected cut-off risk score 0.4 would not be available and feasible for GEO samples. Therefore, another cut-off risk score of 0.8 could distinguish the prognosis of patients from GSE62452 and GSE28735, in the high-risk and low-risk groups (Fig. 4C, right).

5 The independent risk factors from univariate and multivariate COX regression participated in the constitution of the binary classifier.

Given the latest discovery on the intimate correlation between immune stress response and pancreatic cancer genesis, the immune scores of overall TCGA patients derived from ESTIMATE, was incorporated into univariate and multivariate COX regression analysis as well, accompanied by other clinicopathological parameters (Fig. 5A, left and right). Eventually, five-mRNA-risk score, differentiated grade(DG), and positive lymph nodes(PLN) were deemed as relatively independent risk factors for PC progression and damnable prognosis. Surprisingly, the predictive capability of risk score reached 0.777(3-year) and 0.865(5-year) for PC patients to survive. The AUC value of risk score integrated with DG and PLN was palpably increased compared to that of alone clinical factors(DG + PLN)(3-year:0.788vs0.657,5-year:0.877vs0.718) (Fig. 5B, middle and right). Here, we might as well define the combination of risk score, DG, and PLN as Model 1, while one of DG and PLN was Model 2. The C-index of both models were respectively 0.72(95%CI,0.65–0.79) and 0.67(95%CI,0.60–0.75). The calibration curve was reasonably conceived for assessing Model 1 accuracy (Fig. 5C). Hosmer-Lemeshow (HL) goodness of fit confirmed satisfying results for Model 1($X2 = 2.899$,$P = 0.821$),and for Model 2($X2 = 0.079$,$P = 0.981$). Apart from this, we have also verified our constructed classifier ROC in the test cohort (Fig. 6A), and plotted the Model 2 calibration curve (Fig. 6B). DeLong's test suggested the AUC value of risk score + DG + PLN was significantly higher than that of risk score, DG, PLN, DG + PLN ($P < 0.05$).

6 The performance of the binary classifier on clinical utility.

As we intended to set up an uncomplicated forecasting model conducive to taking well-founded and efficient measures on specific patients for clinicians, the classifier practicality had to be estimated conscientiously and rigorously. DCA was plotted to manifest that

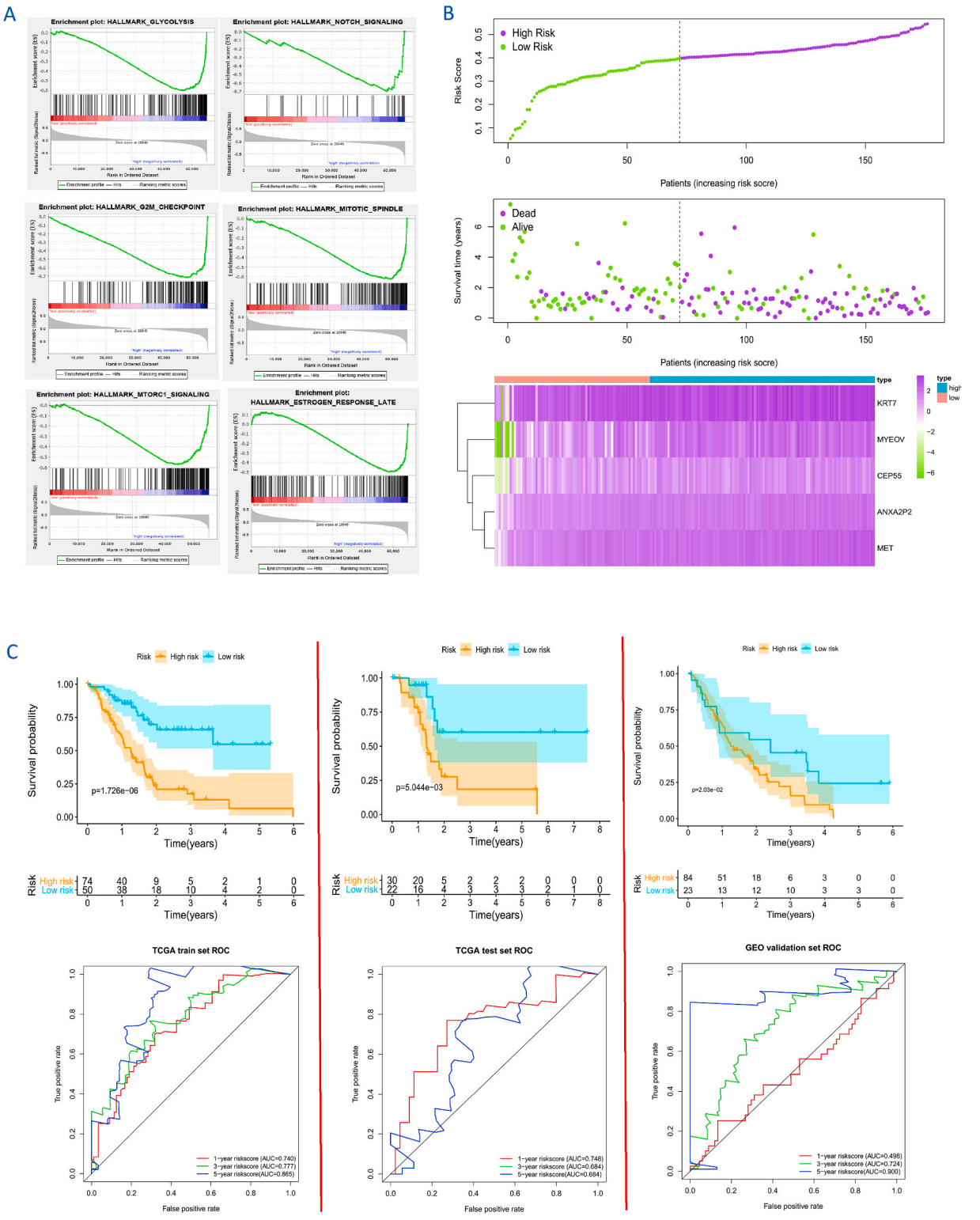
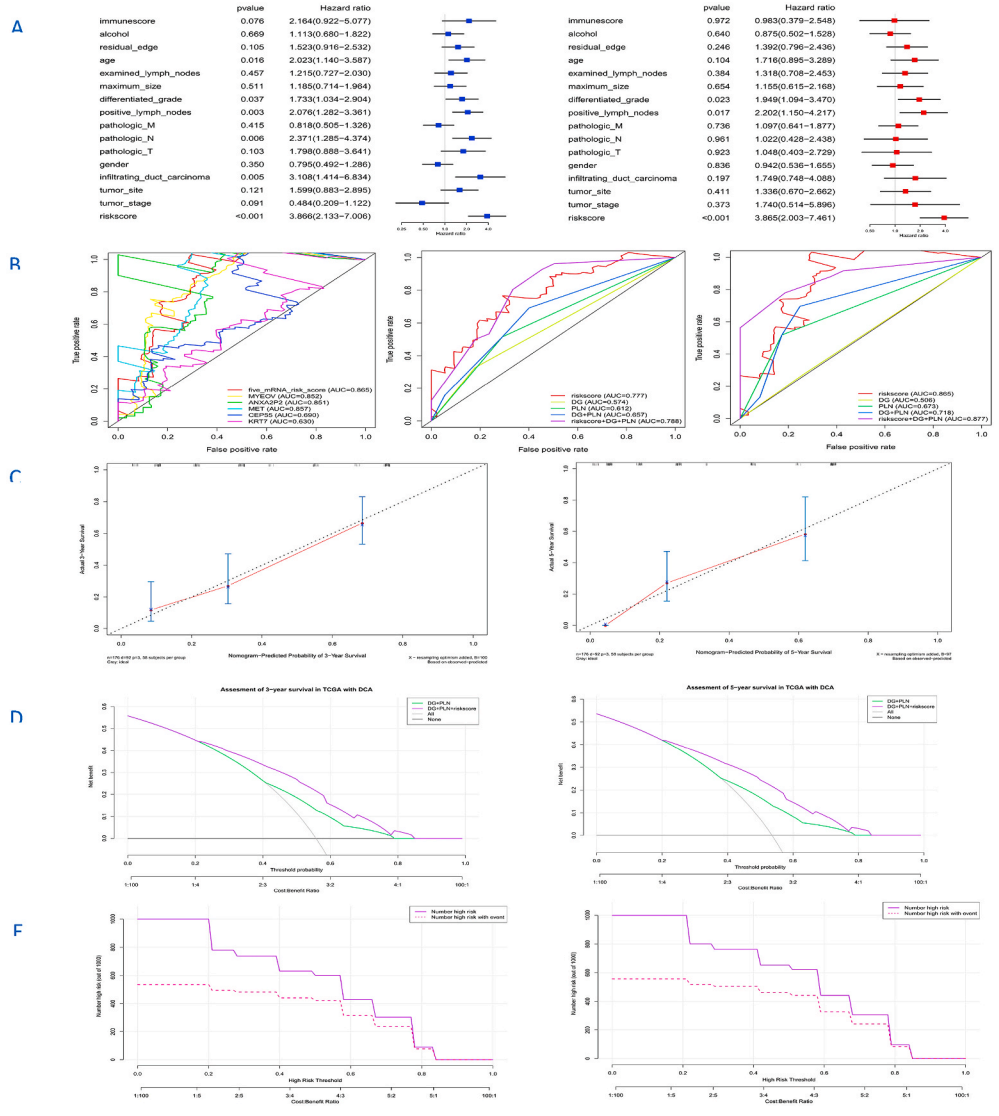


Fig. 4. The function performance and stratification effect of five-mRNA-risk score binary classifier. **(A)**Six most apparent hallmark signaling pathways by GSEA($P < 0.01$). **(B)**The risk score stratification in the whole TCGA cohort. **(C)**The risk score validation respectively in TCGA train set (left), TCGA test set(middle), and GEO external independent validation set(right).



(caption on next page)

Fig. 5. The model performance composed of the five-mRNA classifier combined independent risk factors and relevant subgroup analysis in TCGA. **(A)**univariate (left)and multivariate(right) COX regression in the train set. **(B)**left,5-year ROC of five mRNA and its risk score in the train set. middle,3-year ROC of clinical factors and risk score in the train set. right,5-year ROC of clinical factors and risk score in the trian set. **(C)**3-year(left) and 5-year(right) calibration curve of Model 1 in the whole cohort. **(D)**3-year DCA(left) and 5-year DCA(right) in whole cohort. **(E)**3-year CIC (left) and 5-year CIC(right) of Model 1 in whole cohort. **(F)**Riskscore subgroup analysis on risk elements. **(G)** Chemotherapy subgroup analysis on risk elements.
DG,differentiated_grade.PLN,positive_lymph_nodes.pN,pathologic_N.IDC,infiltrating_duct_carcinoma.*unreliable variables with large Hazard Ratio.

Model 1(the purple line) varied incrementally and approximately in a feasible and secure range from 20 % to 85 % threshold probability(high-risk percentage), while Model 2(the green line) came into force between 40 % and 80 % (Fig. 5D). That was indicated that Model 1 could remind clinicians earlier of adopting appropriate interventions to cut down patient's mortality and to gain larger net benefit. Moreover, CIC was designed further to yield the risk classification of 1000 people, illustrating the proportion of loss vs benefit. Despite it existed undulatory property owing to minor specimen quantity to a certain extent, it was still displayed extensively that both line was not approaching until the high-risk threshold reached 58 % approximately and overlapped at 78 % with a precise predictive ability (Fig. 5E). Its clinical significance was consistent with DCA.

7 Subgroup analysis among clinical and pathological components by risk score and chemotherapy.

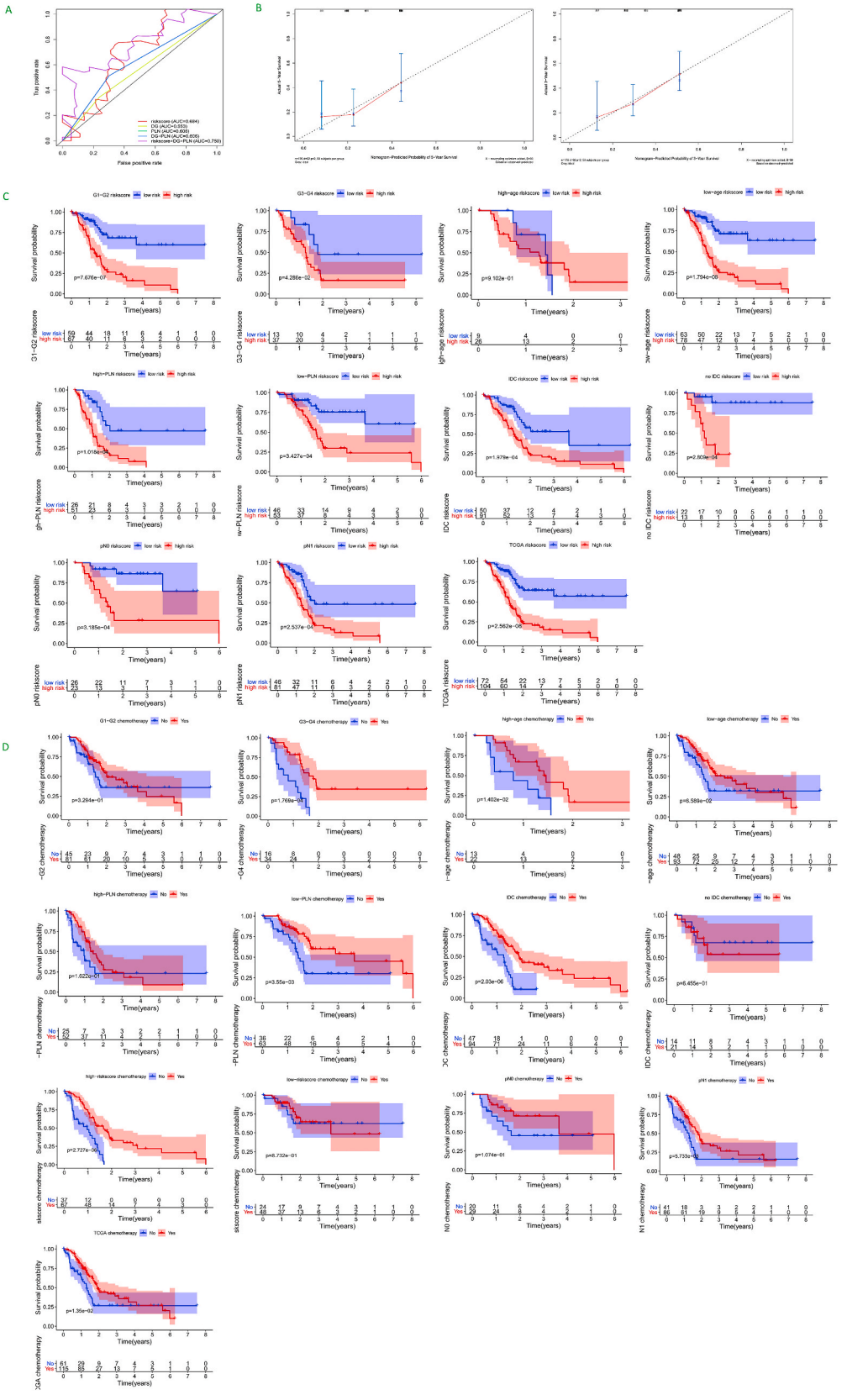
We performed the subgroup analysis on the variables related to prognosis sifted by univariate COX regression. The risk scores of all TCGA patients were calculated on the grounds of the former formula and then, we divided the patients into a high-risk score group and a low-risk score group at a cut-off value of 0.4. So the risk score subgroup analysis suggested that risk score had quite robust discrimination in almost any subgroup, existing extremely conspicuous unlikeness in age($\leq$ 74 years, $P < 0.0001$, HR4.69 (2.60–8.47)), DG(G1-G2, $P < 0.0001$, HR4.06(2.23–7.39)) (Fig. 5F and 6C). To assist clinicians in identifying the applied effects of chemotherapy routinely used in the clinic, we also carried out a subgroup analysis of the clinical availability of chemotherapy. For PC patients with IDC($P < 0.0001$, HR0.33(0.21–0.53)) and high risk score($P < 0.0001$, HR0.31(0.18–0.52)), chemotherapy could achieve remarkable outcome (Fig. 5G and 6D).

8 Establish Model 3 to compare with Model 1 by SEER database.

Multivariate COX regression analysis sieved out eight independent risk divisors involved in Model 3 in the train set, including age, site, DG, stage, pT, pN, ELN, PLN, size (Fig. 7A, down). The 3-year and 5-year AUC of Model 3 were respectively 0.738 and 0.754 in the train set,which were distinctively lower than Model 1(Fig. 7B, upon and down). Its C-index was 0.68 (95%CI,0.67–0.69), weaker than Model 1. HL test of goodness of fit showed $P = 0.258$, $X2 = 10.098$ in the calibration curve (Fig. 7C). Above consequences had been acquired alike in test set (Fig. 8A and B). The selected safety range of Model 3 was around 20%–90 % as DCA, resemble Model 1 (Fig. 7D and 8C). As for CIC, before the high risk threshold of 20 %, namely in an early stage of PC development, the differential value of predicted value reducing real value in Model 3 was visibly less than that of Model 1. Once the patient's carcinoma went into the advanced stage with more than 50 % high-risk threshold, the convergence speed of Model 3 predictive line to accurate line was more sluggish compared with Model 1. The predicted line outskirts to the accurate line at the high-risk threshold of 70 % and coincided at 90 %, lagging behind Model 1, which implied Model 3 forecasting potentiality for high-risk PC patients was not better than Model 1 (Fig. 7E and 8D). At the end of this study, the three model nomograms were painted conveniently to stratify PC patient's risk exponents (Fig. 7F). NRI and IDI were reckoned for the three models as well. Model 1 contrasted to Model 2, "NRI = 0.084, z = 1.288, p = 0.198″, "IDI = 0.055, z = 1.041, p = 0.298″, with slight improvement. However, Model1 compared with Model 3, "NRI = 0.167, z = 1.961, p = 0.05″, "IDI = 0.086, z = 1.427, p = 0.154″, with protuberant advance. It witnessed the superiority of Model 1 again.

4. Discussion

Pancreatic cancer is an aggressive disease with a gloomy prognosis and hidden onset. It is one of the most malignant tumors with a bleak diagnosis rate and survival rate in the early stage [1,2]. More than 85 % of PC are PDAC in the exocrine glands. Owing to the pancreas located behind the peritoneum, therefore, the early symptoms of PC appear inconspicuous, resulted in a difficult early diagnosis [3]. Most PC patients are specifically diagnosed when local advance or distant metastasis takes place [2]. By 2030, PC will turn into the second lethal tumor in Western countries [11]. PDAC clinical guideline(2020 editoion, Spain) pinpointed the risk factors of PDAC, including PC family history, hereditary PC syndrome, complications (diabetes, chronic pancreatitis, obesity], lifestyle (smoking, excessive alcohol intake), occupational factors [12]. Kanno et al. [13] indicated that there were 32 % diabetes, 31 % smoking, 26 % IPMN, 15 % alcohol intake, 13 % chronic pancreatitis, 6.5 % obesity, and 4.5 % PC family history in 200 early PDAC patients. Radical resection (R0 resection) is currently still the most effective means for PC patients to achieve long-term survival. Yet, the peculiar biological behavior of PC leads to blood, lymph, and nerve metastasis and micrometastasis during the early stage. Approximately 80–85 % of patients present with either unresectable or metastatic disease. The 5-year survival rate for patients who can undergo surgical resection is only 10 %–25 % [14]. Neoadjuvant therapy(NAT) can ameliorate the prognosis of PC patients by reducing tumors and micrometastasis, relieving vascular invasion, increasing the R0 resection rate, and alleviating postoperative recurrence and metastasis [15]. However, no consensus has been reached on the formulation of specific clinical strategies, the choice of chemotherapy and radiotherapy, the therapy cycle, and the prediction of efficacy [16].

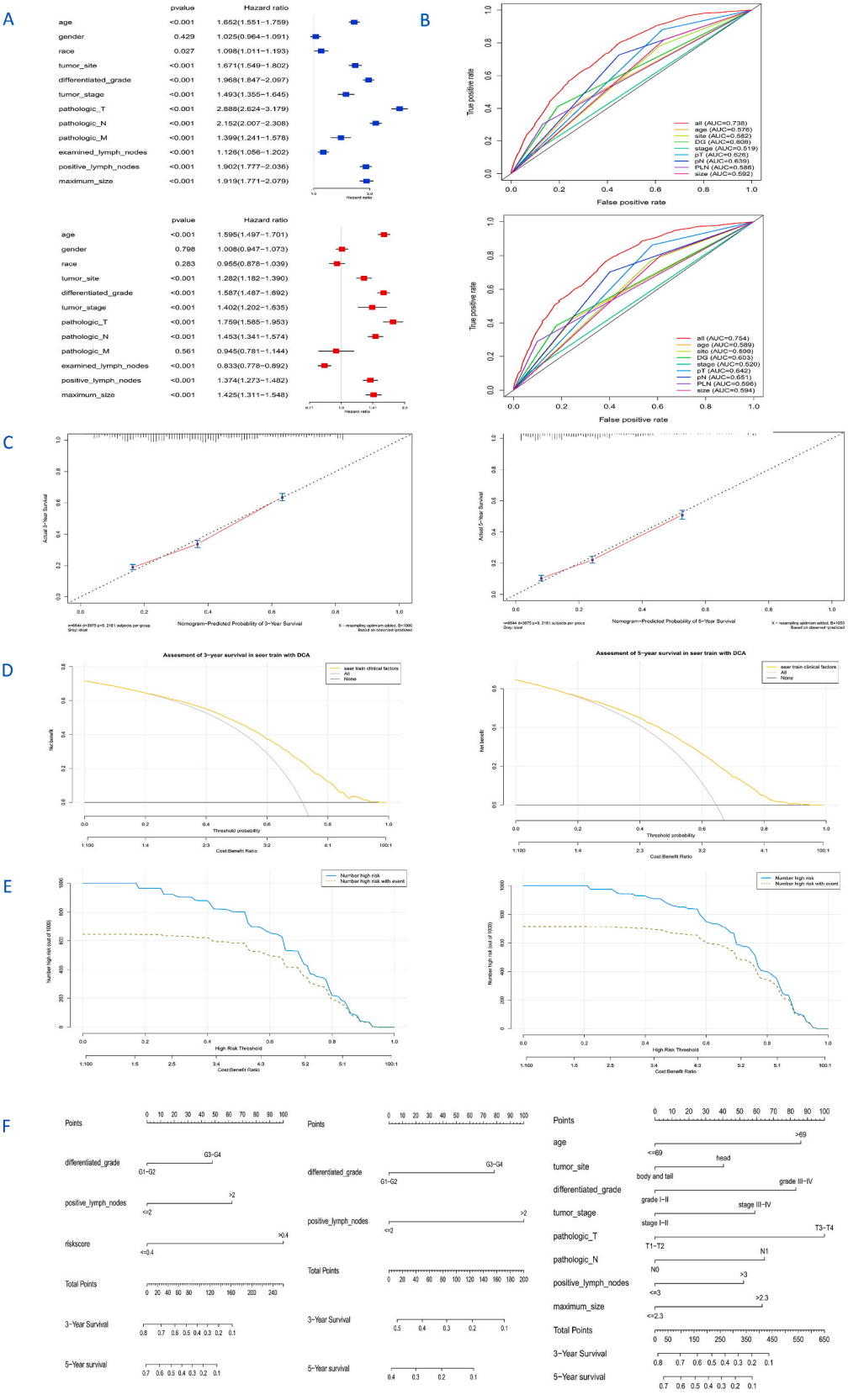


(caption on next page)

Fig. 6. The relevant validation of Model 1 and Model 2 from TCGA cohort. **(A)**3-year and 5-year ROC of Model 1 in test set. **(B)**5-year(left) and 3-year(right) calibration curve of Model 2 in whole cohort. **(C)** risk score survival curves in the subgroups of risk clinical factors. **(D)**survival curves of chemotherapy effect in the subgroups of risk clinical factors.*DG:differentiated_grade. PLN:positive_lymph_nodes.

With penetrated acknowledgement of PC biological characteristics, high-throughput genome sequencing contributes to the diagnosis and treatment of PC more individualized and refinedly [17–19]. In this study, we employed LASSO regression to conceive a binary classifier composed of five hub mRNAs in the TCGA train set for evaluating the postoperative outcomes of PC patients, with punchy verification in the test set and external validation cohort. MYEOV, ANXA2P2, MET, CEP55, KRT7, were filtered out and served jointly as the five-erna-risk score. MYEOV is the first oncogene identified in human orphan genes [20]. MYEOV is overexpressed in hepatocellular carcinoma [21], neuroblastoma [22], non-small cell lung cancer(NSCLC) [23], colorectal cancer [24], multiple myeloma [25] and participates in the regulation of tumor proliferation and invasion. Erbo Liang et al. [26] Discovered that in tumor specimens derived from pancreatic cancer patients, MYEOV was overexpressed and associated with poor prognosis. In addition, MYEOV expression in PDAC was upregulated through promoter hypomethylation. MYEOV depletion impaired metastatic ability and proliferation of PDAC cells both in vitro and in vivo, whereas its overexpression had the opposite effect. Rui Tang et al. [27] demonstrated that MYEOV was higher expressed in tumor tissues compared with adjacent normal pancreas tissues (ANPTs) both in mRNA and protein levels. We also performed bioinformatic analysis and found high MYEOV expression was positively correlated with tumor differentiation ($P = 0.004$), lymph node metastasis ($P = 0.016$) and TNM stage ($P = 0.001$). Shoichiro Tange et al. [28] confirmed that high expression of MYEOV was significantly associated with poor survival and was an independent prognostic factor for disease-specific survival in pancreatic cancer patients. Analysis of multiple cancer samples revealed that the MYEOV promoter region is methylated in noncancer tissues but is demethylated in tumors, causing MYEOV overexpression in tumors. Notably, the knockdown of MYEOV suppressed the expression of MTHFD2 and other folate metabolism-related enzyme genes required for the synthesis of amino acids and nucleic acids and also restored the expression of c-Myc and mTORC1 repressors. Compelling evidence has suggested that pseudogenes play potential roles in the regulation of their cognate wild-type genes, oncogenes, and tumor suppressor genes [29]. Few studies have reported pseudogene ANXA2P2 sense on tumors. Wang Y et al. [30]identified ANXA2P2 used for prognostic ceRNA network in Glioma. Wang QS et al. [31] confirmed ANXA2P2 high expression in hepatocellular carcinoma(HCC) and prognostic value on the risk assessment of recurrence or metastasis of HCC patients. To date, there is no pertinent evidence about the relationship between ANXA2P2 and pancreatic cancer, which may be a significant tendency for PC research not long in the future. MET function study in PC is a hot field at present. Xu Z et al. [32]assessed the effect of HGF/c-MET inhibition plus gemcitabine (G) on the progression of advanced PC. The results proved that even at PC advanced stage, impairing HGF/c-MET pathway combined with chemotherapy, could eliminate metastasis and significantly prevent tumor growth. Firuzi O et al. [33]developed spheroid models to evaluate PSC-PDAC reciprocal interaction, unraveling c-MET inhibition as an important therapeutic option against drug-resistant PDAC. Enliang Li et al. [34] MET was identified as a pancreatic cancer-specific receptor tyrosine kinase(RTK), which is significantly associated with prognosis in both immune "hot" and "cold" pancreatic cancers. MET was observed to be highly upregulated in pancreatic cancer tissues, and positively correlated with PD-L1 levels. Significant benefits of combining MET inhibition with PD-1/PD-L1 blockage were verified in both orthotopic and subcutaneous mouse models of pancreatic cancer. Besides them, in previous studies, CEP55 and KRT7(CK7) served usually as tumor prognostic markers, in agreement with our anticipated results [35–38].Tao Peng et al. [39] also found that cell migration, colony formation, wound-healing, and transwell matrix penetration assays, revealed that upregulation of CEP55 promoted PC cells proliferation, migration, and invasion in vitro, whereas knockdown of CEP55 attenuated it. The nuclear factor κB (NF-κB)/IκBαsignalling pathway, which was activated in CEP55-transduced PC cells and inhibited in CEP55-silenced PC cells, contributed to CEP55-mediated PC cell aggressiveness. This study provided new insights into the oncogenic roles of CEP55 and the mechanism by which the NF-κB pathway is hyperactivated in patients with PC, indicating that CEP55 is a valuable prognostic factor and a potential therapeutic target in PC. Jianlu Song et al. [40] verified that high expression of KRT7 might be responsible for the immunosuppression in the PC microenvironment. KRT7 knockdown was significantly suppressed the abilities of Cell-in-cell (CIC) formation, cell cluster, cell proliferation, migration, and invasion in PC cell lines. Wei Wang et al. [41] study identified that the miR-216a/WT1/KRT7 axis influenced the activity of the PI3K/AKT pathway. To conclude, this study evidenced that miR-216a suppressed WT1 expression and blocked KRT7 transcription, which inactivated the PI3K/AKT signaling and reduced PC progression. At a glance at the five mRNA translation products, there is no relevant literature to confirm whether there is some functional connection between these five hub mRNA. Previous studies have found a correlation between four genes and pancreatic cancer development, infiltration, and metastasis. So whether these five genes are also functionally related to each other is something we should think about. In the next study we hope to delve deeper into this issue.

We integrated five hub-mRNAs risk score as a trick of distinguishing high-risk and low-risk PC patients. High-risk patients obviously focused on 12 hallmark pathways, including glycolysis, NOTCH signaling, G2M checkpoint, mitotic spindle, MTORC1 signaling, estrogen response late, cholesterol homeostasis, E2F targets, estrogen response early, MYC targets, P53 pathway, TGF-βsignaling. Among them, glycolysis [42], NOTCH signaling [43], MTORC1 signaling [44], MYC targets [45], P53 pathway [46], TGF-βsignaling [47] have been identified the roles in PC. The others pathways could provide insightful cognition for specifically targeting PC therapy later. Go enrichment analysis found that those prognostic-related DEGs were involved in human immune response, and facilitated the formation of extracellular matrix. KEGG pathway analysis confirmed that they played an important role in pancreatic secretion and protein digestion and absorption. Previous studies have found that pancreatic cancer cannot occur without alteration in the immune microenvironment. The formation of extracellular matrix promoting immunosuppression is why pancreatic cancer is difficult to be treated well. Our study found that these hub mRNA were closely related to the above behaviors. This was consistent with previous findings.



(caption on next page)

Fig. 7. The evaluated parameters of Model 3 from SEER. **(A)**univariate (upon)and multivariate(down) COX regression in the train set. **(B)**upon,3-year ROC of clinical factors in train set.down,5-year ROC of clinical factors in the trian set. **(C)**3-year(left) and 5-year(right) calibration curve of Model 3 in the train set. **(D)**3-year DCA(left) and 5-year DCA(right) of Model 3 in the train set. **(E)**3-year CIC (left) and 5-year CIC(right) of Model 3 in the train set. **(F)**Nomograms. left, Model 1. middle, Model 2. right, Model 3. DG,differentiated_grade.PLN,positive_lymph_nodes.pT,pathologic_T. pN,pathologic_N.

The five-mRNAs-risk score has robust discrimination, particularly for patients aged $\leq$ 74 years and with G1-G2. Patients with IDC or high risk score received considerable chemotherapy effects. Seldom have prognostic model articles done external verification and subgroup analysis, not to mention patients benefits and clinical practicality [48].ROC is most commonly utilized to evaluate the classifier discrimination, while the calibration curve is for measuring the classifier accuracy. The AUC value of Model 1(our risk score classifier plus independent risk clinical factors by TCGA) was visibly superior to Model 2(only independent risk clinical factors by TCGA) and Model 3(relatively stable one, involving more independent risk clinical factors by SEER), regardless of 3-year (0.788vs0.657vs0.738) or 5-year(0.877vs0.718vs0.754) survival rate. C-index results were consistent with AUC(0.72vs0.67vs0.68). The calibration curve implied that the prediction accuracy of the three models was parallel(Hosmer-Lemeshow test, $P > 0.05$).

Since Vickers AJ [49] invented DCA as a novel tool for accessing predictive models, it frequently assists clinicians in evaluating the appropriateness of treatment strategies for special patients promptly, and tailoring individualized management plans [50]. The prognosis of treatment and mortality of PC are postoperatively often affected by tumor molecular heterogeneity [51,52]. Therefore, there is an urgent demand to better make sense of the molecular pathology of PC to improve the current therapy and lay down novel strategies. A precise molecular classifier can minimize the harm of overtreatment to patients, advise doctors to make decisions, and balance between adverse effects and survival benefits. In our study, two groups of DCA indicated that Model 1 and Model 3 preserved the identical available safe range from 20 % to 90 %, visibly surpassing Model 2(from 40 % to 80 %). In 2016, Kerr KF et al. [53] have further developed the DCA algorithm, which was to draw CIC for more intuitively assessing the performance of prediction models. We demonstrated initially the PC prediction model through CIC, and its appropriate use. In the low-risk stage of PC progression, Model 3 was superior to Model 1 on forecasting accuracy, while it was converse in the high-risk stage as CIC. Pencina MJ et al. [54] exhibited NRI and IDI to redefine various models of improvement on discrimination. Regrettably, the majority of relevant prognostic model articles never involved them till now. In this regard, we seminally attempted NRI and IDI to witness Model 1 robust predictive potentiality advantageous over another two ones.

Although the built five-mRNAs classifier revealed remarkable taxonomy ability and clinical utility, there were still some worth modified shortcomings as below.①As the current study is driven by available retrospective statistics, the optimal cut-off value of risk score is to be determined before it practices clinically.②This binary classifier is established mainly based on TCGA database. Therefore, it is necessary to be verified in multi-center datasets.③Sensitivity analysis can enhance the robustness of the model. Because of the limited sample size of our constructed model, we do not dare to guarantee that our model is sufficiently robust, which is a shortcoming of our model.④Serological detection may be helpful in dynamic surveillance purposes as PC progression markers.⑤Perhaps the experimental studies on five-mRNAs molecular mechanisms further shed light onto the pathogenesis and progress of PC. ⑥If our model was proven to have good discriminatory and generalization ability afterwards, we would like to further simplify and optimize the model before clinical application.But it has to be said that this is currently a challenge for us.

With the continuous updating of pancreatic cancer treatment strategies, new therapeutic modalities, such as neoadjuvant therapy, targeted therapy, immunotherapy, etc. are increasingly used in the clinic. Therefore, there is an urgent need to introduce new predictive elements and improve relevant prediction models with the times to assess the prognostic outcomes of pancreatic cancer. Unlike the empirical judgment of clinicians, the survival prediction model can assess the prognosis of pancreatic cancer patients with quantitative indicators. Its significance is that it can help doctors more accurately assess the patient's disease status and develop personalized treatment plans, thus improving treatment effects and patient survival. Despite our model, the five-mRNA prognostic binary classifier, needs further validation and refinement, but it is a new hope for evaluating PC prognosis.

5. Conclusions

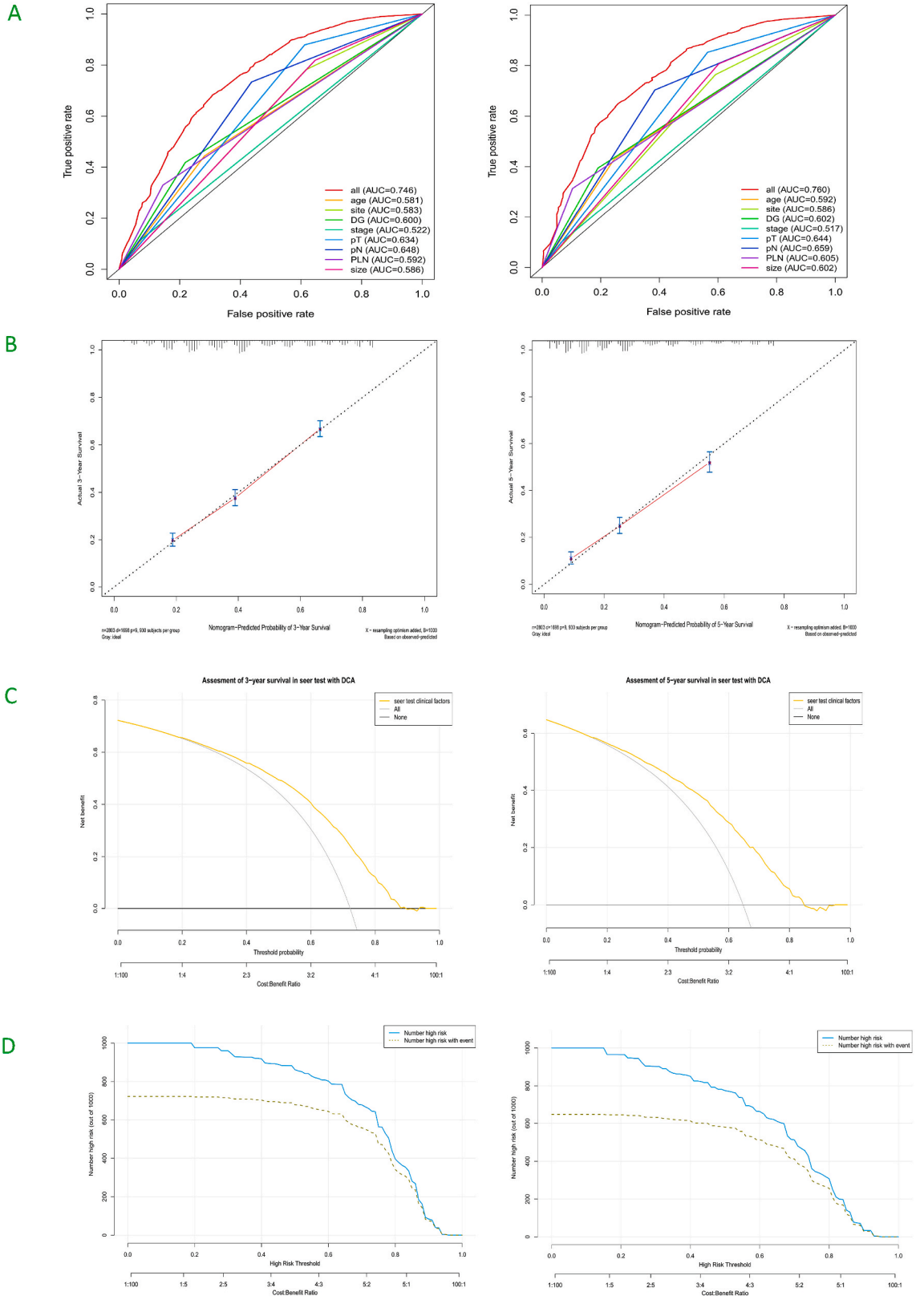
The binary classifier consisting of five core mRNAs was proved as a reliable and feasible instrument to appraise the postoperative prognosis of PC patients. Added with auxiliary clinicopathological factors, it could achieve a better effect. It might enhance patient benefit ratio, facilitate counseling services and connive clinicians to formulate individualized management schemes.

Ethics approval and consent to participate

Since all data are publicly available and they have open-access, ethicalapproval was not necessary in our study.

Methods statement

All methods were carried out by relevant guidelines and regulations.



(caption on next page)

Fig. 8. The evaluated performance of Model 3 from SEER. **(A)**3-year(left) and 5-year(right) ROC of Model 3 in test set. **(B)**3-year(left) and 5-year (right) calibration curve of Model 3 in test set. **(C)**3-year(left) and 5-year(right) DCA of Model 3 in test set. **(D)**3-year(left) and 5-year(right) CIC of Model 3 in test set. DG,differentiated_grade.PLN,positive_lymph_nodes.pT,pathologic_T.pN,pathologic_N.

Consent for publication

All authors have read the manuscript and approved of publication.

Availability of data and materials

These RNAseq expressed data were derived from the following resources available in the public domain: The Cancer Genome Atlas (TCGA) and The Genotype-Tissue Expression (GTEx) databases through UCSC Xena (an online analyzed platform for multi-omic and clinical/phenotype data, http://xena.ucsc.edu/). The Surveillance, Epidemiology, and End Results (SEER) database was used for providing clinical data on PC patients. A validation of the expressions of five hub genes in normal tissues and tumor tissues was done through Human Protein Atlas (HPA, https://www.proteinatlas.org).

Funding

No available funding supported this research.

CRediT authorship contribution statement

Yueqing Lu: Writing – review & editing, Writing – original draft. **Tong Zhou:** Methodology, Data curation. **Mingshu Lu:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Thanks for the RNA sequence data provided with TCGA and GTEx database.Thanks for the SEER database providing clinical data. Thanks to the HPA database providing validated protein data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e31302.

References

[1] C.J. Cabasag, J. Ferlay, M. Laversanne, et al., Pancreatic cancer: an increasing global public health concern, Gut 71 (8) (2022) 1686–1687.
[2] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2019, CA A Cancer J. Clin. 69 (2019) 7–34.
[3] A.J. Grossberg, L.C. Chu, C.R. Deig, et al., Multidisciplinary standards of care and recent progress in pancreatic ductal adenocarcinoma, CA A Cancer J. Clin. 70 (5) (2020) 375–403.
[4] M. Muller, V. Haghnejad, M. Schaefer, et al., The immune landscape of human pancreatic ductal carcinoma: key players, clinical implications, and challenges, Cancers 14 (4) (2022) 995.
[5] J. Huang, V. Lok, C.H. Ngai, et al., Worldwide burden of, risk factors for, and trends in pancreatic cancer, Gastroenterology 160 (3) (2021) 744–754.
[6] T. Kamisawa, L.D. Wood, T. Itoi, K. Takaori, Pancreatic cancer, Lancet 388 (2016) 73–85.
[7] S. Van Roessel, G.G. Kasumova, J. Verheij, et al., International validation of the eighth edition of the American Joint Committee on Cancer (AJCC) TNM staging system in patients with resected pancreatic cancer, JAMA surgery 153 (12) (2018).
[8] T. Lan, W. Liu, Y. Lu, H. Luo, A five-gene signature for predicting overall survival of esophagus adenocarcinoma, Medicine (Baltim.) 100 (14) (2021 Apr 9) e25305.
[9] B. Wu, Z. Wang, N. Lin, X. Yan, Z. Lv, Z. Ying, Z. Ye, A panel of eight mRNA signatures improves prognosis prediction of osteosarcoma patients, Medicine (Baltim.) 100 (14) (2021 Apr 9) e24118.
[10] H. Yin, C. Zhang, X. Gou, W. He, D. Gan, Identification of a 13-mRNA signature for predicting disease progression and prognosis in patients with bladder cancer, Oncol. Rep. 43 (2) (2020 Feb) 379–394.
[11] K.C. Shin, A novel metabolic detour for pancreatic cancer survival, Mol. Cell. 46 (6) (2023) 345.
[12] M.A. Gómez-España, A.F. Montes, R. Garcia-Carbonero, et al., SEOM clinical guidelines for pancreatic and biliary tract cancer (2020), Clin. Transl. Oncol. 23 (2021) 988–1000.
[13] A. Kanno, A. Masamune, K. Hanada, et al., Multicenter studyof early pancreatic cancer inJapan, Pancreatology 18 (1) (2018) 61–67.
[14] J.D. Mizrahi, R. Surana, J.W. Valle, et al., Pancreatic cancer, Lancet 395 (10242) (2020) 2008–2020.
[15] C. Springfeld, C.R. Ferrone, M.H.G. Katz, et al., Neoadjuvant therapy for pancreatic cancer, Nat. Rev. Clin. Oncol. 20 (5) (2023) 318–337.

[16] A.H. Ko, Pancreatic cancer and the possibility of long-term survival: a glimmer of hope? JAMA Oncol. 2 (3) (2016 Mar) 380–381.
[17] S.A. Hosseini, A.S. Jouneghani, M. Ghatrehsamani, et al., CRISPR/Cas9 as precision and high-throughput genetic engineering tools in gastrointestinal cancer research and therapy, Int. J. Biol. Macromol. 223 (2022) 732–754.
[18] K.L. Aung, S.E. Fischer, R.E. Denroche, et al., Genomics-driven precision medicine for advanced pancreatic cancer-early results from the COMPASS trial, Clin. Cancer Res. 24 (6) (2018 Mar 15) 1344–1354.
[19] A. Agathangelidis, E. Vlachonikola, F. Davi, et al., High-Throughput Immunogenetics for Precision Medicine in cancer[C]//Seminars in Cancer Biology, vol. 84, Academic Press, 2022, pp. 80–88.
[20] Y. Shao, C. Chen, H. Shen, B.Z. He, D. Yu, S. Jiang, S. Zhao, Z. Gao, Z. Zhu, X. Chen, et al., GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes, Genome Res. 29 (2019) 682–696.
[21] Z. Deng, J. Wang, B. Xu, et al., Mining TCGA database for tumor microenvironment-related genes of prognostic value in hepatocellular carcinoma, BioMed Res. Int. 2019 (3) (2019) 1–12.
[22] R. Tang, J. Ji, J. Ding, et al., Overexpression of MYEOV predicting poor prognosis in patients with pancreatic ductal adenocarcinoma, Cell Cycle 19 (13) (2020) 1602–1610.
[23] L. Fang, S. Wu, X. Zhu, et al., MYEOV functions as an amplified competing endogenous RNA in promoting metastasis by activating TGF-β pathway in NSCLC, Oncogene 38 (6) (2019) 896.
[24] Z. Zhang, L. Huang, J. Li, et al., Bioinformatics analysis reveals immune prognostic markers for overall survival of colorectal cancer patients: a novel machine learning survival predictive system, BMC Bioinf. 23 (1) (2022) 124.
[25] S. Tange, T. Hirano, M. Idogawa, et al., MYEOV overexpression induced by demethylation of its promoter contributes to pancreatic cancer progression via activation of the folate cycle/c-Myc/mTORC1 pathway, BMC Cancer 23 (1) (2023) 85.
[26] E. Liang, Y. Lu, Y. Shi, Q. Zhou, F. Zhi, MYEOV increases HES1 expression and promotes pancreatic cancer progression by enhancing SOX9 transactivity, Oncogene 39 (41) (2020 Oct) 6437–6450.
[27] R. Tang, J. Ji, J. Ding, J. Huang, B. Gong, X. Zhang, F. Li, Overexpression of MYEOV predicting poor prognosis in patients with pancreatic ductal adenocarcinoma, Cell Cycle 19 (13) (2020 Jul) 1602–1610.
[28] S. Tange, T. Hirano, M. Idogawa, E. Hirata, I. Imoto, T. Tokino, MYEOV overexpression induced by demethylation of its promoter contributes to pancreatic cancer progression via activation of the folate cycle/c-Myc/mTORC1 pathway, BMC Cancer 23 (1) (2023 Jan 25) 85.
[29] S. Kalyana-Sundaram, C. Kumar-Sinha, S. Shankar, et al., Expressed pseudogenes in the transcriptional landscape of human cancers, Cell 149 (7) (2012) 1622–1634.
[30] Y. Wang, X. Liu, G. Guan, et al., Identification of a five-pseudogene signature for predicting survival and its ceRNA network in glioma, Front. Oncol. 9 (2019 Oct 15) 1059.
[31] Q.S. Wang, L.L. Shi, F. Sun, et al., High expression of ANXA2 pseudogene ANXA2P2 promotes an aggressive phenotype in hepatocellular carcinoma, Dis. Markers 2019 (2019 Feb 10) 9267046.
[32] Z. Xu, T.C.Y. Pang, A.C. Liu, et al., Targeting the HGF/c-MET pathway in advanced pancreatic cancer: a key element of treatment that limits primary tumour growth and eliminates metastasis, Br. J. Cancer 122 (10) (2020 Mar 23) 1486–1495.
[33] O. Firuzi, P.P. Che, B.E. Hassouni, et al., Role of c-MET inhibitors in overcoming drug resistance in spheroid models of primary human pancreatic cancer and stellate cells, Cancers 11 (5) (2019) 638.
[34] E. Li, X. Huang, G. Zhang, T. Liang, Combinational blockade of MET and PD-L1 improves pancreatic cancer immunotherapeutic efficacy, J. Exp. Clin. Cancer Res. 40 (1) (2021 Sep 3) 279.
[35] M.S.A. Zaki, M.A. Eldeen, W.K. Abdulsahib, et al., A comprehensive pan-cancer analysis identifies CEP55 as a potential oncogene and novel therapeutic target, Diagnostics 13 (9) (2023) 1613.
[36] J.G. Zhao, Y.J. Li, Y. Wu, et al., Revealing platelet-related subtypes and prognostic signature in pancreatic adenocarcinoma, BMC Med. Genom. 16 (1) (2023) 106.
[37] J. Song, R. Ruze, Y. Chen, et al., Construction of a novel model based on cell-in-cell-related genes and validation of KRT7 as a biomarker for predicting survival and immune microenvironment in pancreatic cancer, BMC Cancer 22 (1) (2022) 1–24.
[38] Y. Lu, D. Li, G. Liu, et al., Identification of critical pathways and potential key genes in poorly differentiated pancreatic adenocarcinoma, OncoTargets Ther. (2021) 711–723.
[39] T. Peng, W. Zhou, F. Guo, H.S. Wu, C.Y. Wang, L. Wang, Z.Y. Yang, Centrosomal protein 55 activates NF-κB signalling and promotes pancreatic cancer cells aggressiveness, Sci. Rep. 7 (1) (2017 Jul 19) 5925.
[40] J. Song, R. Ruze, Y. Chen, R. Xu, X. Yin, C. Wang, Q. Xu, Y. Zhao, Construction of a novel model based on cell-in-cell-related genes and validation of KRT7 as a biomarker for predicting survival and immune microenvironment in pancreatic cancer, BMC Cancer 22 (1) (2022 Aug 16) 894.
[41] W. Wang, J. Wang, C. Yang, J. Wang, MicroRNA-216a targets WT1 expression and regulates KRT7 transcription to mediate the progression of pancreatic cancer-A transcriptome analysis, IUBMB Life 73 (6) (2021 Jun) 866–882.
[42] X. Chang, X. Liu, H. Wang, et al., Glycolysis in the progression of pancreatic cancer, Am. J. Cancer Res. 12 (2) (2022) 861.
[43] J. Xu, W. Xu, X. Yang, et al., LncRNA MIR99AHG mediated by FOXA1 modulates NOTCH2/Notch signaling pathway to accelerate pancreatic cancer through sponging miR-3129-5p and recruiting ELAVL1, Cancer Cell Int. 21 (1) (2021) 1–14.
[44] M.H. Jeong, G. Urquhart, C. Lewis, et al., Inhibition of phosphodiesterase 4D suppresses mTORC1 signaling and pancreatic cancer growth, JCI insight 8 (13) (2023).
[45] M. Ala, Target c-Myc to treat pancreatic cancer, Cancer Biol. Ther. 23 (1) (2022) 34–50.
[46] M.P. Kim, X. Li, J. Deng, et al., Oncogenic KRAS recruits an expansive transcriptional network through mutant p53 to drive pancreatic cancer metastasis, Cancer Discov. 11 (8) (2021) 2094–2111.
[47] X. Ma, Z. Cui, Z. Du, et al., Transforming growth factor-β signaling, a potential mechanism associated with diabetes mellitus and pancreatic cancer? J. Cell. Physiol. 235 (9) (2020) 5882–5892.
[48] M. Strijker, J.W. Chen, T.H. Mungroop, et al., Systematic review of clinical prediction models for survival after surgery for resectable pancreatic cancer, Journal of British Surgery 106 (4) (2019) 342–354.
[49] A.J. Vickers, E.B. Elkin, Decision curve analysis: a novel method for evaluating prediction models, Med. Decis. Making 26 (6) (2006 Nov-Dec) 565–574.
[50] B. Van Calster, L. Wynants, J.F.M. Verbeek, et al., Reporting and interpreting decision curve analysis: a guide for investigators, Eur. Urol. 74 (6) (2018 Dec) 796–804.
[51] A.A. Connor, S. Gallinger, Pancreatic cancer evolution and heterogeneity: integrating omics and clinical data, Nat. Rev. Cancer 22 (3) (2022) 131–142.
[52] L. Lan, T. Evan, H. Li, et al., GREM1 is required to maintain cellular heterogeneity in pancreatic cancer, Nature 607 (7917) (2022) 163–168.
[53] K.F. Kerr, M.D. Brown, K. Zhu, et al., Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use, J. Clin. Oncol. 34 (21) (2016 Jul 20) 2534–2540.
[54] M.J. Pencina, RB Sr D'Agostino, R.B. D'Agostino Jr., et al., Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond, Stat. Med. 27 (2) (2008 Jan 30) 157–172.