Observational Study

Medicine

OPEN

Prediction of overall survival in pancreatic cancer based on a twenty-four-gene risk model associated with lymph node metastasi

Junwei Fang, MS[a], Meiping Wang, MS[a], Yi Gao, MS[b], Yafeng Qi, MS[a], Weixuan Hong, MS[a], Chunhong Xiao, MS[a,*]

Abstract

Pancreatic adenocarcinoma (PAAD) is a leading cause of tumor-related mortality. Identifying potential prognostic risk genes is crucial for predicting the overall survival of PAAD patients. In this study, we constructed and validated a 24-gene risk score. This risk score stratifies PAAD patients into low-risk and high-risk groups. The model demonstrated excellent prognostic accuracy at different follow-up times (1-year AUC: 0.81, 2-year AUC: 0.85, 3-year AUC: 0.92). PAAD patients from 3 GEO datasets were categorized into low-risk and high-risk groups, with survival analysis revealed significant differences in survival rates between the 2 groups ($P < .01$). Multivariate analysis identified 2 independent risk factors, namely, N stage (HR 2.026, 95% CI 1.139–3.603, $P = .016$) and the 24-gene risk score (HR 0.239, 95% CI 0.148–0.385, $P < .001$). The performance of the nomogram in the TCGA database is commendable (AUC for 1-year, 2-year, and 3-year survival rates = 0.76, 0.77, and 0.86, respectively). In essence, our work establishes a 24-gene risk score and nomogram to facilitate clinicians in predicting the prognosis of individual PAAD patients.

Abbreviations: AUC = area under the curve, DEGs = differentially expressed genes, GO = gene ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes, LASSO = least absolute shrinkage and selection operator, LNM = lymph node metastasis, OS = overall survival, PAAD = pancreatic adenocarcinoma, ROC = receiver operating characteristic, TNF = tumor necrosis factor, WGCNA = weighted gene co-expression network analysis.

Keywords: lymph node metastasis, pancreatic adenocarcinoma, prediction, risk model

1. Introduction

Pancreatic adenocarcinoma (PAAD) is one of the deadliest malignancies, with a 5-year survival rate of approximately 10%.[1] Globally, the burden of PAAD is increasing annually, with projections suggesting that it will become the second leading cause of cancer-related deaths by 2030.[2] This is primarily due to early infiltration of the pancreas into surrounding blood vessels, tissues, and organs, with potential spread along nerves and lymphatics.[3] Current cancer treatments typically involve neoadjuvant therapy, curative surgical resection, and adjuvant chemotherapy or radiation. However, there is no consensus on neoadjuvant therapy for pancreatic cancer; only a minority of patients are eligible for curative surgical resection, and postoperative chemotherapy and radiation have not yielded the expected outcomes.[4] The limited availability of effective treatments remains a significant factor contributing to poor patient prognosis.[5] Presently, the clinical assessment of patient prognosis relies mainly on TNM staging at diagnosis; this method fails to offer accurate personalized prognostic predictions. Increasing evidence suggests that the clinical prognostic factors currently used for PAAD may not consistently predict individual clinical outcomes.[6,7] These findings underscore the need for more sensitive and accurate prognostic markers for PAAD patients.

Transcriptomic analysis has been widely employed to predict the prognosis of cancer patients, leading to the discovery of numerous biomarkers with potential clinical value.[8] Concurrently, machine learning methods have been applied to gene and genomic data to identify molecular features, elucidate complex cellular mechanisms, predict clinical outcomes, and more.[9] In recent years, the analysis of clinical information and expression profiles from public databases has led to the identification of an increasing number of prognostic biomarkers for PAAD.[10,11] One study, through analysis of the TCGA database,

This work was supported by the 900th Hospital of the Joint Logistics Support Force of the Chinese People's Liberation Army (2023GK01) and the Natural Science Foundation of Fujian Province (2022J011095).

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available.

The study protocols were approved by the Ethical Committee Review Board of the 900th Hospital of Joint Logistic Support Force (Fuzhou, Fujian, China).

Supplemental Digital Content is available for this article.

[a] Department of General Surgery, 900th Hospital of Joint Logistics Support Force of People's Liberation Army, Fuzhou, China, [b] Fujian University of Traditional Chinese, Fuzhou, China.

** Correspondence: Chunhong Xiao, Department of General Surgery, 900th Hospital of Joint Logistics Support Force of People's Liberation Army, Fuzhou, Fujian 350000, China (e-mail: xiao84chun@163.com).*

How to cite this article: Fang J, Wang M, Gao Y, Qi Y, Hong W, Xiao C. Prediction of overall survival in pancreatic cancer based on a twenty-four-gene risk model associated with lymph node metastasi. Medicine 2025;104:20(e42448).

Received: 2 October 2024 / Received in final form: 27 December 2024 / Accepted: 27 April 2025

http://dx.doi.org/10.1097/MD.0000000000042448

established a new risk scoring model to predict disease course, immune microenvironment, signaling pathways, tumor mutations, and drug sensitivity in PAAD patients.[12] In another study, the authors developed a reliable and accurate risk scoring model for antiangiogenic strategies in PAAD by analyzing expression profiles from the TCGA database.[13] These studies highlight the great potential of using public database resources to develop predictive risk models. However, the predictive efficacy of the aforementioned methods remains imperfect. Therefore, continued exploration of genetic and polygenic features associated with PAAD prognosis is essential to improve prediction accuracy.

In this study, we aimed to identify and validate robust and reliable prognostic features associated with overall survival (OS). Using least absolute shrinkage and selection operator (LASSO) regression, we successfully established a PAAD predictive model based on 24-genes and validated it with GEO datasets. Our research provides a novel approach for prognostic prediction in clinical PAAD patients and offers additional insights into the molecular mechanisms of this disease.

2. Materials and methods

2.1. Data acquisition and processing

GEPIA2 (http://gepia2.cancer-pku.cn/#index) is an updated version of GEPIA that can be used to analyze RNA expression sequencing data from 9736 tumor samples and 8587 normal samples from the TCGA and GTEx databases. We used this database to screen for differentially expressed genes (DEGs), with normal samples serving as the control group (|logFC| > 1.0, adjusted P value < .01).[14] We collected gene expression profiles from 178 PAAD samples in the TCGA database (https://portal.gdc.cancer.gov/); we also collected gene expression profiles from 3 GEO database cohorts (GSE28735, GSE57495, and GSE62452), totaling 170 PAAD samples, and conducted background adjustment and quantile normalization. Our study excluded samples with missing or insufficient data on variables such as age, stage, recurrence status (none or present), tumor status (tumor or tumor-free), survival status (alive or deceased), and survival time. In correlation analyses with clinicopathological characteristics of PAAD patients, we excluded samples with categories such as "unknown," "TX," and "NX."

2.2. Bioinformatics analysis

Using weighted gene co-expression network analysis (WGCNA), we constructed a gene co-expression network in the TCGA PAAD dataset. Prior to conducting network analysis, we evaluated the PAAD data clustering to identify any clear outliers. We used the "WGCNA" R package to select the optimal soft-thresholding power, aiming to maintain adequate connectivity and bring the gene network closer to a scale-free topology.[15] Additionally, we assessed the correlation between lymph node metastasis (LNM) and different modules to identify the module most significantly associated with LNM. To further explore the biological significance of the DEGs, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. This provided insights into the functional roles and pathways associated with the identified DEGs.

2.3. Construction and validation of the risk model

The key module in WGCNA that showed a significant correlation with LNM underwent LASSO regression analysis to narrow down the range of target genes. LASSO regression analysis was employed to reduce multicollinearity among genes and prevent overfitting of variables in the prognostic risk model. This approach is highly popular in machine learning. The genes selected from the LASSO regression analysis were used to construct risk scoring features. The signature based risk score is calculated via the following formula[16]:

$$\text{Risk score} = \sum n_i = \sum (\text{coef i} * x_i)$$

Coefi is the coefficient and x_i is the z-score transformed relative expression value of each selected gene. We calculated each patient's score and stratified the entire cohort into high-risk and low-risk groups based on the median risk score, thereby establishing a risk prediction model. The difference in OS between the 2 subgroups was compared via the log-rank test.

To validate the risk model, receiver operating characteristic (ROC) curves were plotted via the R package "survivalROC," and the area under the curve (AUC) values for 1-year, 2-year, and 3-year survival rates were calculated.

To further evaluate whether the risk score is an independent risk factor for OS, we performed univariate and multivariate Cox regression analyses on the risk score, combining it with other clinical characteristics to identify independent risk factors. Subsequently, we used the "rms" package to construct a nomogram incorporating these independent prognostic factors.

Calibration curves were plotted, and the concordance index (C-index) was calculated to assess the accuracy of the nomogram. Prognostic risk values for each patient were calculated via the nomogram, and the entire cohort was divided into 2 subgroups based on the total score from the nomogram. The performance of the nomogram was then assessed using tdROC analysis and survival rate estimation.

The area under the tdROC curve (AUC) calculated via the "timeROC" package in R indicates the accuracy of prediction or prognosis. The survival rate estimates for the 2 subgroups were analyzed using Kaplan–Meier analysis with the "survival" package in R.

Three datasets (GSE28735, GSE57495, and GSE62452) were used to confirm the predictive ability of the risk model. Additionally, a $P < .05$ was considered statistically significant.

2.4. Statistical analysis

SPSS version (v. 21.0) and GraphPad Prism (v. 8.0) were used for statistical analysis and generating figures. Paired t tests, unpaired t tests, and one-way analysis of variance (ANOVA) were used to compare gene expression across different groups. Kaplan–Meier survival analysis of data from the TCGA and GEO databases was used to assess the prognostic significance of the risk model, and the log-rank P value was calculated. Cox proportional hazards regression models were utilized for univariate and multivariate survival analyses. Prognostic factors identified in the univariate analysis were included in subsequent multivariate analysis. A $P < .05$ was considered statistically significant.

3. Results

3.1. Identification of DEGs and WGCNA

Overall, we extracted data from 178 PAAD patients with clinical and pathological diagnoses from the TCGA database and analyzed these data,[17,18] as outlined in the flowchart (Fig. 1). Using GEPIA2, we conducted differential expression analysis of RNA sequencing data from pancreatic cancer and normal samples in TCGA and GTEx (|logFC| > 1.0, adjusted P value < .01). A total of 9219 DEGs were identified, including 8740 upregulated and 479 downregulated genes (Fig. 2A and B). To better understand the relationships between LNM and molecular groups, we extracted RNA sequence data and performed WGCNA. Using the WGCNA package, we analyzed the co-expression network of DEGs, constructed these co-expression modules, and divided them into 14 meaningful modules (Fig. 2C). By analyzing the associations between gene modules and LNM, we found that the purple module had the highest correlation with N stage (LNM)

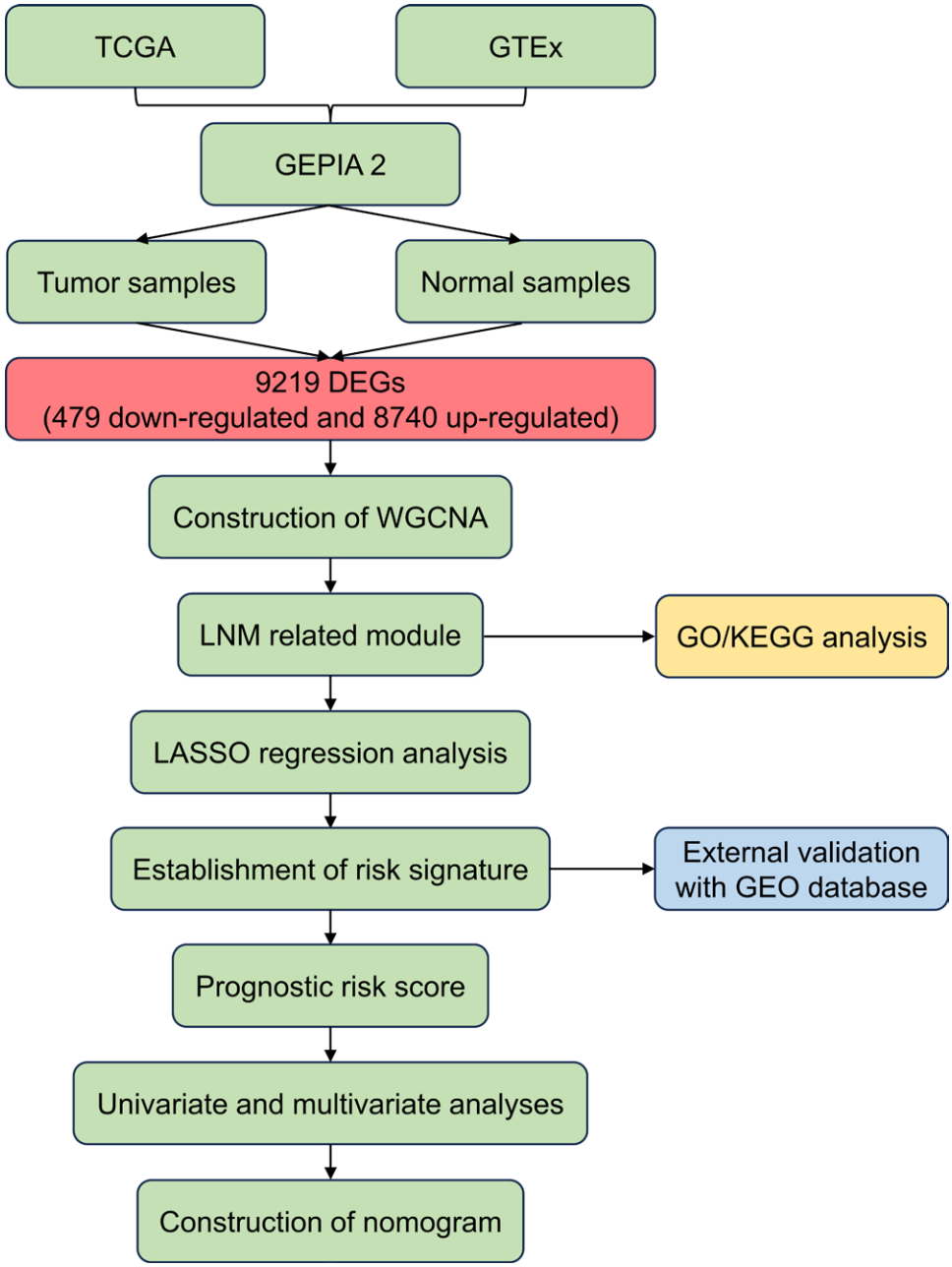


Figure 1. The flow chart of the study design and analysis. DEGs = differentially expressed genes, GEO = Gene Expression Omnibus database, GEPIA2 = Gene Expression Profiling Interactive Analysis, GO = gene ontology, GTEx = Genotype-Tissue Expression, KEGG = Kyoto Encyclopedia of Genes and Genomes, LASSO = least absolute shrinkage and selection operator, LNM = lymph node metastasis, TCGA = The Cancer Genome Atlas, WGCNA = Weighted Gene Correlation Network Analysis.

(COR = 0.16, $P < .05$) (Fig. 2D). There were 155 genes in the purple module, which were further used for subsequent analysis.

3.2. GO/KEGG enrichment analysis of key modules

To elucidate the potential functions of these 155 genes, GO and KEGG analyses were conducted. As shown in Figure 2E, the significant items in the GO analysis varied, with some of them including "activation of immune response," "immune response-regulating signaling pathway," "tumor necrosis factor production," and "regulation of tumor necrosis factor production." Additionally, KEGG analysis revealed that these 155 DEGs were significantly enriched in the "phagosome," "lysosome," "neutrophil extracellular trap formation," and "complement and coagulation cascades" pathways (Fig. 2F).

These results indicate that these DEGs are distributed across pathways related to immune cells and tumor necrosis factors, which have been shown to play a key roles in the onset and progression of PAAD. These functions also suggest mechanisms of disease progression in PAAD patients.

3.3. Construction and validation of the risk model

To establish a risk classifier associated with LNM for predicting the prognosis of PAAD patinets, we utilized the expression profiles of the 155 genes from the purple module to construct a LASSO regression model. Cross-validation was employed to select the penalty parameter lambda, enabling the identification of relatively independent feature genes for subsequent model analysis (Fig. 3A and B). Ultimately, we identified 24 genes with

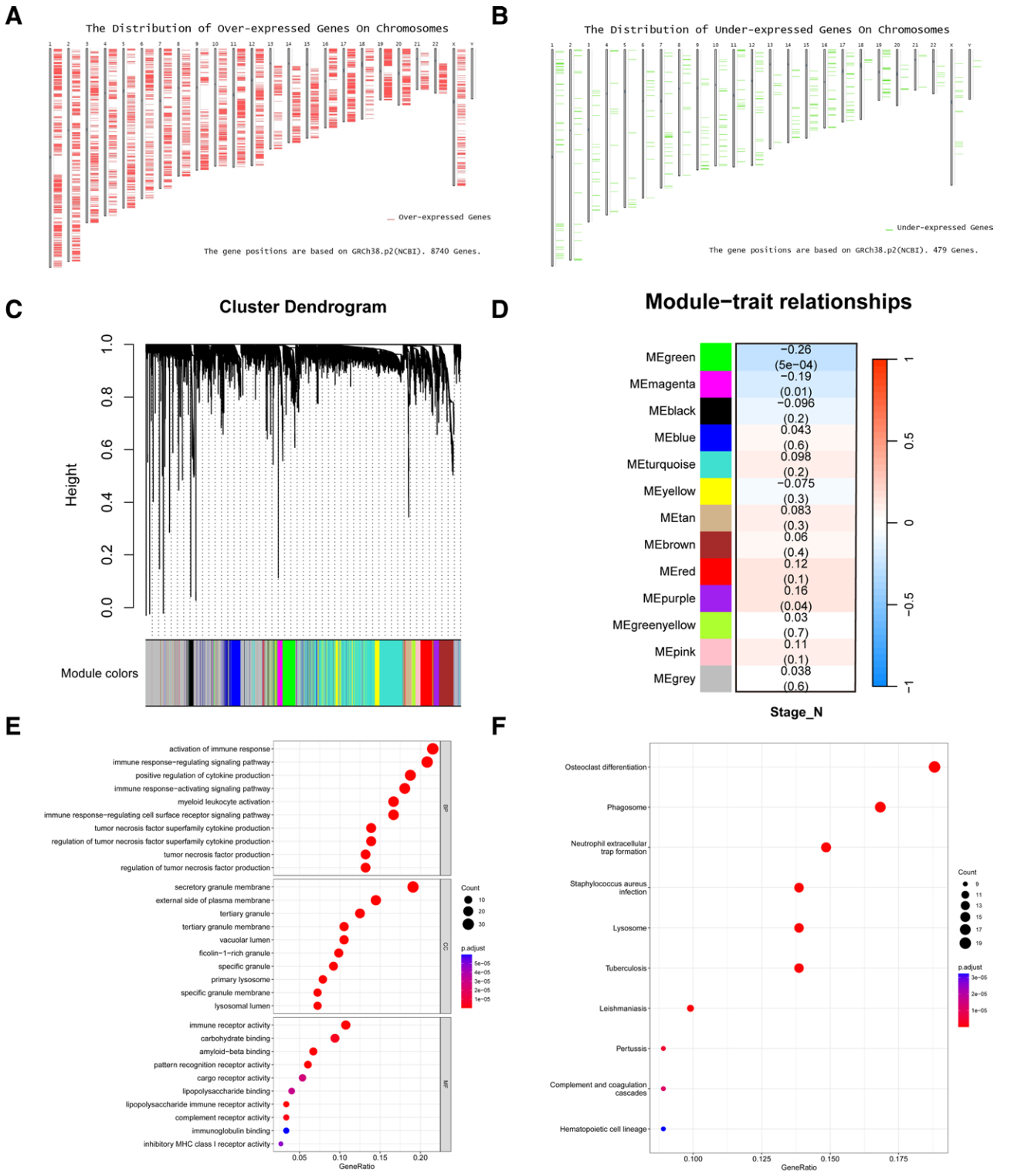
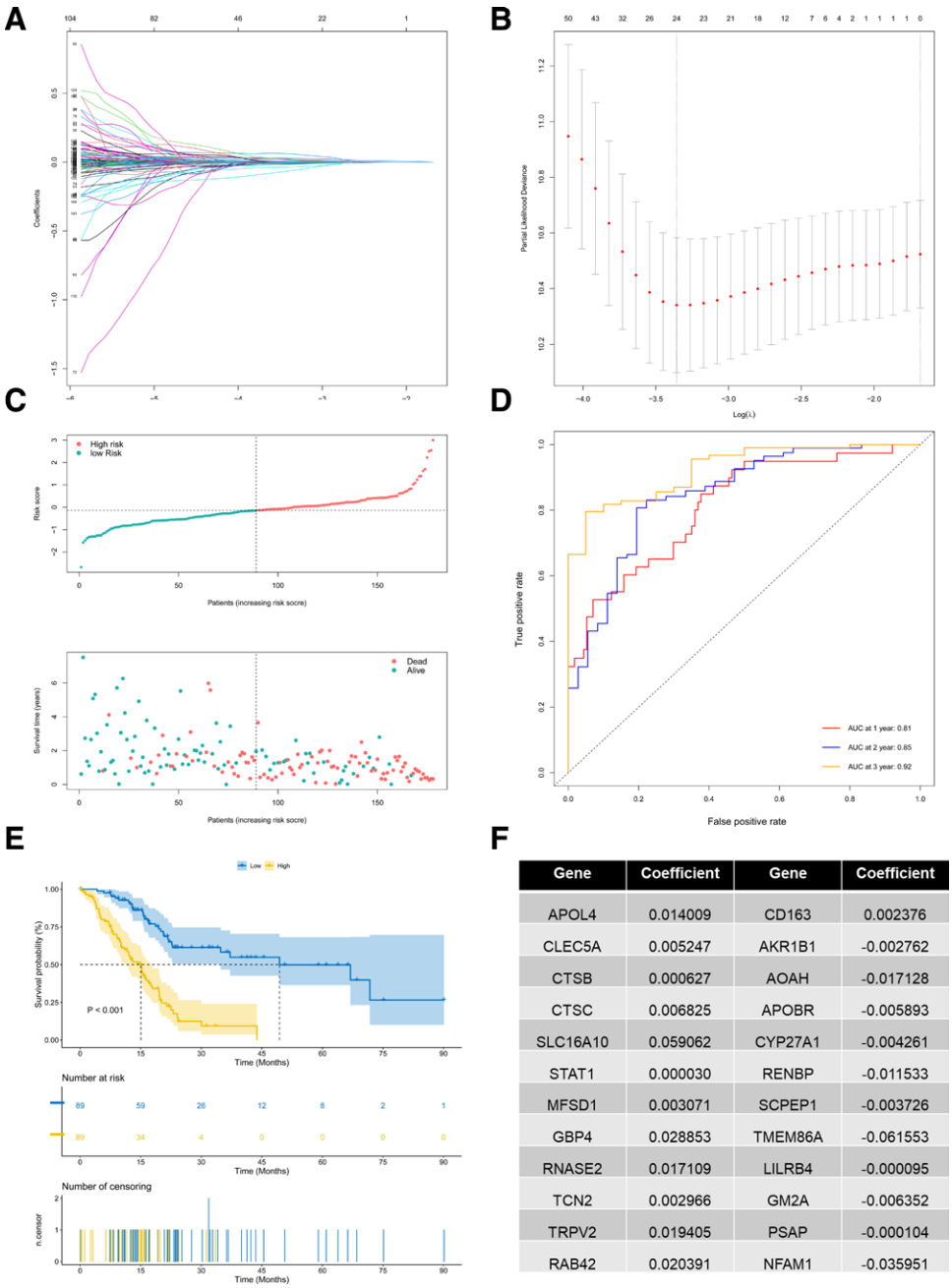


Figure 2. DEGs distribution, WGCNA construction and GO/KEGG analysis. (A) The Distribution of over-expressed Genes on chromosomes. (B) The distribution of under-expressed genes on chromosomes. (C) Clustering dendrogram of genes. (D) Correlation between modules and risk model. (E) GO analysis. (F) KEGG analysis. GO = gene ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes, WGCNA = weighted gene co-expression network analysis.

the largest regression coefficients (Fig. 3F). Based on these 24 independent prognostic feature genes, we computed risk scores for each sample. Consequently, PAAD patients with available follow-up information were stratified into 2 groups: the low-risk group (n = 89) and the high-risk group (n = 89) (Fig. 3C). To validate the risk assessment model, we plotted ROC curves, with the AUC values for 1-year, 2-year, and 3-year survival rates being 0.81, 0.85, and 0.92, respectively (Fig. 3D).Subsequently, Kaplan–Meier curves and log-rank tests demonstrated that patients in the high-risk group had significantly lower overall survival rates compared to those in the low-risk group ($P < .001$) (Fig. 3E). The accuracy of this model and its association with prognosis were also found to be higher than those of previous similar studies.



Gene	Coefficient	Gene	Coefficient
APOL4	0.014009	CD163	0.002376
CLEC5A	0.005247	AKR1B1	-0.002762
CTSB	0.000627	AOAH	-0.017128
CTSC	0.006825	APOBR	-0.005893
SLC16A10	0.059062	CYP27A1	-0.004261
STAT1	0.000030	RENBP	-0.011533
MFSD1	0.003071	SCPEP1	-0.003726
GBP4	0.028853	TMEM86A	-0.061553
RNASE2	0.017109	LILRB4	-0.000095
TCN2	0.002966	GM2A	-0.006352
TRPV2	0.019405	PSAP	-0.000104
RAB42	0.020391	NFAM1	-0.035951

Figure 3. Identification of prognostic genes in PAAD patients. (A and B) LASSO regression model. (C) Risk score distribution and survival status of patients. (D) Time-dependent ROC curve for 1-, 2-, and 3-year survival prediction. (E) Kaplan–Meier survival analysis of the low and high-risk group. (F) Genes from LASSO regression and correlated coefficient value. LASSO = least absolute shrinkage and selection operator, ROC = receiver operating characteristic.

3.4. Development and evaluation of a nomogram for OS prediction

To develop a quantitative method for predicting overall survival probability, we used a nomogram to establish a prognostic model. First, we performed univariable and multivariable analyses to identify independent risk factors for overall survival. In the multivariable analysis, PAAD patients' N stage (HR 2.026, 95% CI 1.139–3.603, $P = .016$) and risk model (HR 0.239, 95% CI 0.148–0.385, $P < .001$) were associated with prognosis (Fig. 4A and B).

Based on the multivariable analysis of overall survival in TCGA PAAD patients ($P < .05$), a nomogram was generated to predict prognosis (Fig. 4C). In the nomogram, the risk model had the greatest impact on prognosis, followed by the N stage. The discrimination and calibration of the nomogram were investigated. The calibration plots showed excellent agreement and acceptable fluctuations between the predicted and observed values of 1-, 2-, and 3-year overall survival in the cohort (Fig. 4D).

Next, the TCGA cohort was divided into 2 subgroups based on the total score from the nomogram. In terms of overall survival, PAAD patients in the high-score group had significantly poorer prognosis than those in the low-score group ($P < .001$; Fig. 4E). To further explore the efficiency of the nomogram, ROC curve analysis was performed, and the AUC for OS was calculated. The results showed that the AUC values for the 1-, 2-, and 3-year survival rates were 0.76, 0.77, and 0.86, respectively (Fig. 4F).

In summary, these results suggest that the nomogram based on the 24-gene risk model and clinical factors has significant predictive utility for the prognosis of PAAD patients.

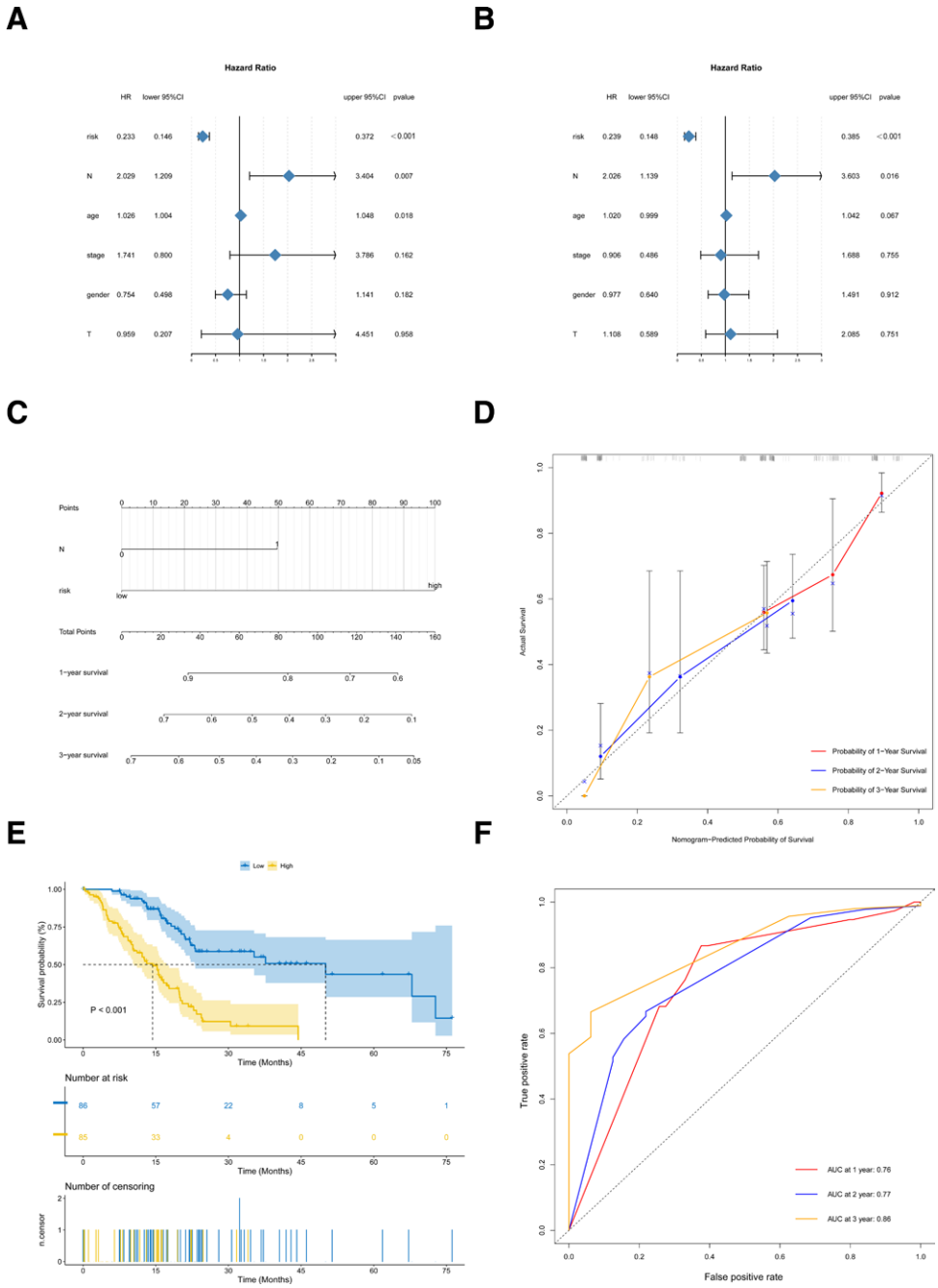


Figure 4. Establishment of a nomogram for survival prediction. (A and B) Univariate and multivariate analyses of the association between clinicopathological factors and overall survival of PAAD patients. (C) Nomogram (D) Calibration plot of the nomogram-predicted probability and actual survival (E) Kaplan–Meier survival analysis of the low and high-risk group. (F) Time dependent ROC curves for the nomogram predicting 1-, 2-, 3-year survival. PAAD = pancreatic adenocarcinoma, ROC = receiver operating characteristic.

3.5. GEO validation of the risk model

To further validate the utility and stability of the 24-gene risk model, we conducted the same analysis on a combined dataset from 3 external cohorts (GSE28735, GSE57495, and GSE62452) from the GEO database.Cox regression analysis demonstrated that the risk model (HR 0.623, 95% CI 0.432–0.899, $P < .05$) was an independent prognostic factor for patient outcomes in the GEO dataset (Figure S1, Supplemental Digital Content, https://links.lww.com/MD/O899). In the external test set (n = 170), the model successfully classified 85 patients into the high-risk group and 85 patients into the low-risk group.

The distributions of risk scores and survival status in the 2 groups of external data (Fig. 5A) showed results consistent with the TCGA PAAD cohort, with lower survival rates observed in patients with high-risk scores compared to those with low-risk scores. Similarly, the validation results on the external validation set demonstrated that patients in the high-risk group (n = 85) had worse overall survival compared to patients in the low-risk group (n = 85).

The survival curve revealed a significantly poorer prognosis in the high-risk group compared to the low-risk group ($P < .01$, Fig. 5B). These validations confirm that the 24-gene risk model exhibits high accuracy in external datasets.

4. Discussion

Evidence suggests that despite advancements in the diagnosis and treatment of PAAD, the prognosis remains poor for patients

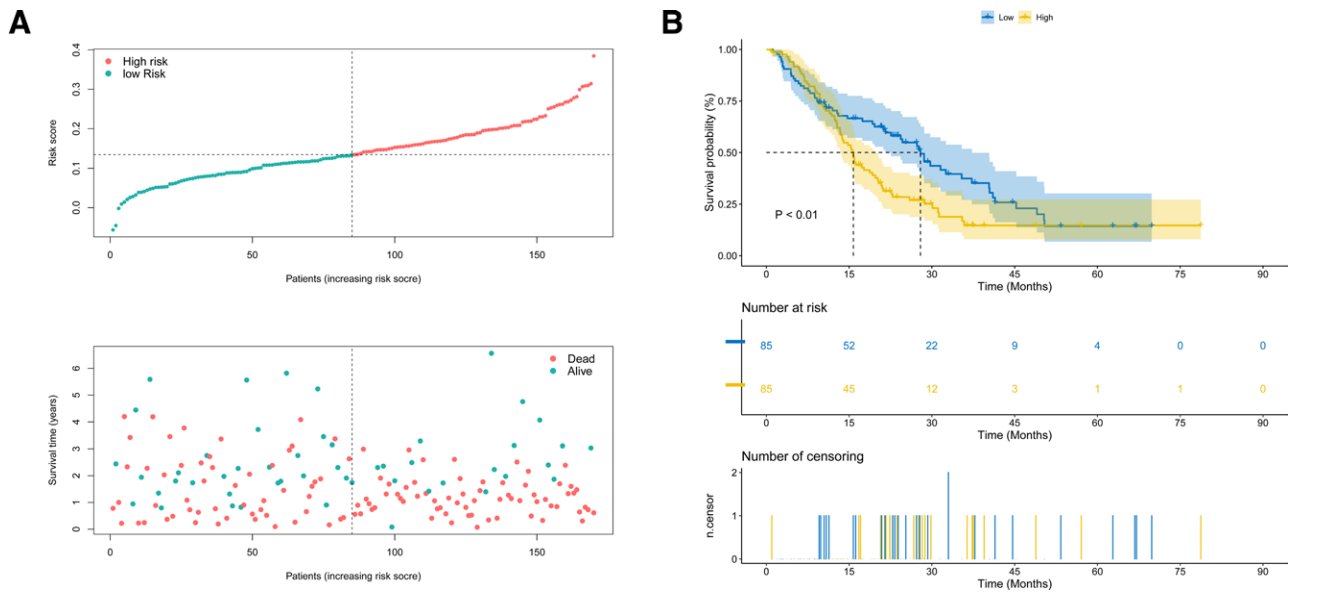


Figure 5. The distribution of risk scores, patients' survival status and ROC curves in the GEO cohort. (A) Distribution of risk scores and urvival status of patients in GEO. (B) Kaplan–Meier survival analysis of GEO. GEO = Gene Expression Omnibus database, ROC = receiver operating characteristic.

with lymph node metastasis.[19] Current clinical prognostic indicators for PAAD primarily rely on patient and cancer-related factors such as TNM stage and grade, but their accuracy and specificity are limited. Numerous studies indicate that dysregulation of gene expression may be associated with the occurrence, progression, and prognosis of tumors, with some genes considered prospective biomarkers for predicting the prognosis of PAAD patients.[20,21] However, these studies still have limitations. Therefore, our study aims to integrate new molecular markers with clinicopathological features to better predict the overall survival of PAAD patients.

In our study, we used WGCNA to explore the relationships between gene modules and LNM. We first filtered DEGs before conducting the WGCNA to maintain the integrity of gene connectivity and relationships within the gene modules. Our model was validated viaexternal data, and demonstrated relatively high accuracy in external validation. Lastly, survival analysis proved that the model to be highly efficient.

In our WGCNA, we identified 13 gene modules. Correlation analysis revealed that the purple module had the strongest association with LNM. Functional enrichment analysis using GO and KEGG indicated that the purple module was enriched in functions related to "immune cells" and "tumor necrosis factor" (TNF).

Pancreatic cancer is characterized by significant infiltration of immune cells, primarily exhibiting immune suppression.[22] The mechanisms of immune suppression in pancreatic cancer mainly involve increased expression of immune inhibitory receptors on antitumor immune cells, the release of immune-suppressive soluble mediators, and inhibition of important metabolic substrates of immune cells.[23] However, the infiltration of anticancer immune cells such as natural killer cells and T cells is relatively low.[24] This contributes to the poor efficacy of immunotherapy in pancreatic cancer; however, studies have shown that when T-cell immunity in pancreatic cancer is adequately induced, T cells can become powerful weapons for tumor treatment.[25]

The TNF family comprises 29 TNF receptors and 19 TNF ligands, which play crucial roles in tumor immunity through coinhibitory or costimulatory pathways.[26] Additionally, TNF has been demonstrated to influence tumor development and metastasis in multiple ways.[27] Therefore, modulating the interactions among TNF family members represents a promising avenue for cancer treatment.[28]

We subsequently performed LASSO regression analysis to identify key genes from the purple module. We identified 24 key genes and calculated a risk score for each patient. The risk score can be used to classify PAAD patients into low-risk and high-risk groups to predict OS. There was a significant difference in OS between the 2 groups, with an AUC of 0.92 for the 3-year survival rate. Additionally, the prognostic value of the 24-gene markers was validated in the GEO dataset, indicating their stability and strong differentiation ability in categorizing PAAD patients into high-risk and low-risk subgroups. Our study demonstrates that this risk score successfully identifies high-risk and low-risk PAAD patients with significant differences in OS and can effectively predict their survival period.

To assess the independence of the 24 gene features in predicting OS, we conducted both univariate and multivariate Cox regression analyses. After adjusting for the effects of age, grade, and tumor pathological stage in the regression analysis, the patient risk model based on the 24 gene features remained significantly correlated with OS. In summary, these results confirm the prognostic ability of the 24-gene model in predicting OS in PAAD patients, independent of other clinical features. Therefore, our predictive features may help identify high-risk PAAD patients and facilitate the development of appropriate clinical follow-up plans.

Other studies have also employed similar methods to construct risk features, such as 5-gene prognostic models, 7-gene prognostic models, and 10-gene prognostic models.[10,29,30] Our study focused on identifying gene modules associated with lymph node metastasis via the WGCNA method. Compared with other studies, our risk model and nomination chart are more comprehensive and have higher predictive accuracy. Importantly, that most of the 24 key genes identified in our study are closely associated with the occurrence of various cancers, including AKR1B1, APOL4, CD163, CLEC5A, CTSC, CYP27A1, STAT1, MFSD1, GBP4, RNASE2, LILRB4, TRPV2, RAB42, and PSAP.

AKR1B1 is highly expressed in various cancers and is associated with tumor development. Studies suggest that AKR1B1 inhibitors have tremendous potential as novel cancer therapeutic agents.[31] APOL4 is associated with tumor grade and immune infiltration, and it may serve as a potential biomarker for treatment and prognosis assessment, particularly in evaluating the efficacy of immunotherapy.[32] CLEC5A is highly expressed in various cancer tissues, and its overexpression indicates poor

prognosis in patients. It may promote tumor development and migration by triggering the AKT/mTOR signaling pathway.[33,34] Alterations in lipid metabolism are considered one of the hallmarks of cancer and critical events in tumorigenesis and progression.[35] Research suggests that CYP27A1 can promote tumor progression by regulating cholesterol metabolism.[36] STAT1 is considered a tumor suppressor in various cell types, and understanding the regulation of STAT1 signaling is crucial for effective cancer therapy. The development of new regulators that modulate STAT1 expression and activity may offer new treatment options for cancer patients.[37] In experimental and spontaneous metastasis mouse models, the loss of MFSD1 increases focal adhesion turnover and reduces the stability of β-1 integrin, leading to increased integrin activation, and subsequently higher levels of tumor metastasis.[38]

Tumor staging is widely used as an important indicator for predicting the survival of PAAD patients. However, at the individual level, the predictive ability of tumor staging is often not accurate enough.[39] By combining the risk model with clinical and pathological features, constructing a nomogram can further improve the predictive ability and specificity of overall survival. Single-factor and multifactor analyses have shown that N stage and the risk model in the TCGA database are independent prognostic factors for OS. To evaluate the accuracy of the predictive signals, time-dependent ROC analysis was conducted, and the AUCs at different time points were calculated. In the TCGA cohort, the AUCs at 1, 2, and 3 years were 0.76, 0.77, and 0.86, respectively. Survival analysis showed significant differences in survival between the high-risk and low-risk groups. These results indicate that our 24-gene risk score and nomogram improve the accuracy of survival prediction for PAAD patients.

To the best of our knowledge, the 24-gene risk model has not been previously reported, and the nomogram combining expression information and clinical-pathological factors will aid clinicians in identifying new prognostic biomarkers for PAAD from both clinical and foundational perspectives. However, it's essential to recognize the limitations of our study. Firstly, it was retrospective in design, with a relatively small sample size in the cohort. Secondly, the expression and function of these 24 genes in patient tissues have not been experimentally validated. Further research is needed to elucidate the interactions among these genes and validate our findings.

5. Conclusion

Most importantly, our study unveiled a 24-gene risk model derived from the TCGA PAAD cohort, which was subsequently validated across multiple datasets. This model demonstrated a high accuracy in predicting the overall survival of PAAD patients. Furthermore, we devised and validated a prognostic nomogram for PAAD, incorporating the 24-gene risk model alongside clinicopathologic features. This nomogram serves as a valuable predictive tool for assessing patient survival and prognosis.

Acknowledgments

The authors thank the staffs from the General Surgery Department at 900 Hospital of the Joint Logistics Support Force for their help.

Author contributions

Conceptualization: Junwei Fang, Yi Gao, Chunhong Xiao.
Data curation: Junwei Fang, Yi Gao, Chunhong Xiao.
Formal analysis: Junwei Fang, Yafeng Qi, Chunhong Xiao.
Funding acquisition: Junwei Fang, Meiping Wang, Chunhong Xiao.
Investigation: Junwei Fang, Meiping Wang, Chunhong Xiao.
Methodology: Junwei Fang, Yafeng Qi, Chunhong Xiao.
Project administration: Junwei Fang, Weixuan Hong, Chunhong Xiao.
Resources: Junwei Fang, Weixuan Hong, Chunhong Xiao.
Software: Junwei Fang, Chunhong Xiao.
Supervision: Junwei Fang, Chunhong Xiao.
Validation: Junwei Fang, Chunhong Xiao.
Visualization: Junwei Fang, Chunhong Xiao.
Writing – original draft: Junwei Fang, Chunhong Xiao.
Writing – review & editing: Junwei Fang, Chunhong Xiao.

References

[1] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74:229–63.
[2] Li J, Yin J, Li W, Wang H, Ni B. Molecular subtypes based on cuproptosis-related genes and tumor microenvironment infiltration characteristics in pancreatic adenocarcinoma. Cancer Cell Int. 2023;23:7.
[3] Xie Z, Chen X, Li J, et al. Salivary HOTAIR and PVT1 as novel biomarkers for early pancreatic cancer. Oncotarget. 2016;7:25408–19.
[4] Fang J, Xiao C, Qi Y, Hong W, Wang M. Influence of pancreaticoduodenectomy for periampullary carcinoma on intestinal microbiome and metabolites. Heliyon. 2024;10:e24393.
[5] Balachandran VP, Beatty GL, Dougan SK. Broadening the impact of immunotherapy to pancreatic cancer: challenges and opportunities. Gastroenterology. 2019;156:2056–72.
[6] Mizrahi JD, Surana R, Valle JW, Shroff RT. Pancreatic cancer. Lancet. 2020;395:2008–20.
[7] Connor AA, Gallinger S. Pancreatic cancer evolution and heterogeneity: integrating omics and clinical data. Nat Rev Cancer. 2022;22:131–42.
[8] Xu D, Wang Y, Chen Y, Zheng J. Identification of the molecular subtype and prognostic characteristics of pancreatic cancer based on CD8 + T cell-related genes. Cancer Immunol Immunother. 2023;72:647–64.
[9] Zhang W-Y, Chen Z-H, An X-X, et al. Analysis and validation of diagnostic biomarkers and immune cell infiltration characteristics in pediatric sepsis by integrating bioinformatics and machine learning. World J Pediatr. 2023;19:1094–103.
[10] Wu Z, Huang X, Cai M, Huang P, Guan Z. Novel necroptosis-related gene signature for predicting the prognosis of pancreatic adenocarcinoma. Aging (Albany NY). 2022;14:869–91.
[11] Wang W, Zhang J, Wang Y, Xu Y, Zhang S. Identifies microtubule-binding protein CSPP1 as a novel cancer biomarker associated with ferroptosis and tumor microenvironment. Comput Struct Biotechnol J. 2022;20:3322–35.
[12] Xue S, Ge W, Wang K, et al. Association of aging-related genes with prognosis and immune infiltration in pancreatic adenocarcinoma. Front Cell Dev Biol. 2022;10:942225.
[13] Cao G, Chang Y, Yang G, Jiang Y, Han K. A novel risk score model based on four angiogenesis long non-coding RNAs for prognosis evaluation of pancreatic adenocarcinoma. Aging (Albany NY). 2022;14:9090–102.
[14] Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res. 2019;47:W556–60.
[15] Xu M, Zhou H, Hu P, et al. Identification and validation of immune and oxidative stress-related diagnostic markers for diabetic nephropathy by WGCNA and machine learning. Front Immunol. 2023;14:1084531.
[16] Zhao Y, Shi W, Tang Q. An eleven-gene risk model associated with lymph node metastasis predicts overall survival in lung adenocarcinoma. Sci Rep. 2023;13:6852.
[17] Zheng L, Zhang J, Ye Y, et al. Construction of a novel cancer-associated fibroblast-related signature to predict clinical outcome and immune response in colon adenocarcinoma. Aging (Albany NY). 2023;15:9521–43.
[18] Gui Z, Ye Y, Li Y, et al. Construction of a novel cancer-associated fibroblast-related signature to predict clinical outcome and immune response in cervical cancer. Transl Oncol. 2024;46:102001.
[19] Kong Y, Li Y, Luo Y, et al. circNFIB1 inhibits lymphangiogenesis and lymphatic metastasis via the miR-486-5p/PIK3R1/VEGF-C axis in pancreatic cancer. Mol Cancer. 2020;19:82.
[20] Zhou C, Wang L, Hu W, et al. CDC25C is a prognostic biomarker and correlated with mitochondrial homeostasis in pancreatic adenocarcinoma. Bioengineered. 2022;13:13089–107.

[21] Fan L, Wang X, Chang Q, Wang Y, Yang W, Liu L. IL2RA is a prognostic indicator and correlated with immune characteristics of pancreatic ductal adenocarcinoma. Medicine (Baltimore). 2022;101:e30966.
[22] Xiang H, Yang R, Tu J, et al. Metabolic reprogramming of immune cells in pancreatic cancer progression. Biomed Pharmacother. 2023;157:113992.
[23] Galluzzi L, Buqué A, Kepp O, Zitvogel L, Kroemer G. Immunological effects of conventional chemotherapy and targeted anticancer agents. Cancer Cell. 2015;28:690–714.
[24] Zheng L, Xue J, Jaffee EM, Habtezion A. Role of immune cells and immune-based therapies in pancreatitis and pancreatic ductal adenocarcinoma. Gastroenterology. 2013;144:1230–40.
[25] Ajina R, Malchiodi ZX, Fitzgerald AA, et al. Antitumor T-cell immunity contributes to pancreatic cancer immune resistance. Cancer Immunol Res. 2021;9:386–400.
[26] Dostert C, Grusdat M, Letellier E, Brenner D. The TNF family of ligands and receptors: communication modules in the immune system and beyond. Physiol Rev. 2019;99:115–60.
[27] Medler J, Kucka K, Wajant H. Tumor Necrosis Factor Receptor 2 (TNFR2): an emerging target in cancer therapy. Cancers (Basel). 2022;14:2603.
[28] Li H, Liu S, Li C, Xiao Z, Hu J, Zhao C. TNF family-based signature predicts prognosis, tumor microenvironment, and molecular subtypes in bladder carcinoma. Front Cell Dev Biol. 2021;9:800967.
[29] Chen X, Yuan Q, Liu J, et al. Comprehensive characterization of extracellular matrix-related genes in PAAD identified a novel prognostic panel related to clinical outcomes and immune microenvironment: a silico analysis with in vivo and vitro validation. Front Immunol. 2022;13:985911.
[30] Chi H, Peng G, Wang R, et al. Cuprotosis programmed-cell-death-related lncRNA signature predicts prognosis and immune landscape in PAAD patients. Cells. 2022;11:3436.
[31] Laffin B, Petrash JM. Expression of the aldo-ketoreductases AKR1B1 and AKR1B10 in human cancers. Front Pharmacol. 2012;3:104.
[32] Zhu H, Hu X, Feng S, et al. APOL4, a novel immune-related prognostic biomarker for glioma. J Clin Med. 2022;11:5765.
[33] Pan Y, Wang C, Wang S, Wu X, Sheng L, Qi Z. CLEC5A regulates the proliferation and migration of colon cancer via the AKT/mTOR signaling pathway. J Gastrointest Oncol. 2023;14:1331–45.
[34] Wang Q, Shi M, Sun S, et al. CLEC5A promotes the proliferation of gastric cancer cells by activating the PI3K/AKT/mTOR pathway. Biochem Biophys Res Commun. 2020;524:656–62.
[35] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–74.
[36] Wang X, Shi J, Huang M, et al. TUBB2B facilitates progression of hepatocellular carcinoma by regulating cholesterol metabolism through targeting HNF4A/CYP27A1. Cell Death Dis. 2023;14:179.
[37] Dong L, Wei X, Yu L, Li Y, Chen L. Inhibition of SIRT7 promotes STAT1 activation and STAT1-dependent signaling in hepatocellular carcinoma. Cell Signal. 2024;114:111005.
[38] Roblek M, Bicher J, van Gogh M, et al. The solute carrier MFSD1 decreases the activation status of β1 integrin and thus tumor metastasis. Front Oncol. 2022;12:777634.
[39] König A-K, Gros H, Hinz U, et al. Refined prognostic staging for resected pancreatic cancer by modified stage grouping and addition of tumour grade. Eur J Surg Oncol. 2022;48:113–20.