Research Paper

A novel gene signature for prognosis prediction and chemotherapy response in patients with pancreatic cancer

Hongcao Lin[1,2,*], Chonghui Hu[3,*], Shangyou Zheng[3,*], Xiang Zhang[1,2], Rufu Chen[3,&], Quanbo Zhou[1,2]

[1]Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, Guangdong Province, China
[2]Department of Pancreatobiliary Surgery, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, Guangdong Province, China
[3]Department of General Surgery, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, Guangdong Province, China
*Equal contribution

Correspondence to: Quanbo Zhou, Rufu Chen; email: zhouqbo@mail.sysu.edu.cn, chenrf63@163.com

ABSTRACT

Pancreatic cancer is a lethal disease. Chemoresistance is one of the characteristics of pancreatic cancer and leads to a poor prognosis. This study built an effective predictive model for personalized treatment and explored the molecular mechanism of chemoresistance. A four-gene signature, including serine peptidase inhibitor Kazal type 1 (*SPINK1*), anoctamin 1 (*ANO1*), desmoglein 3 (*DSG3*) and GTPase, IMAP family member 1 (*GIMAP1*) was identified and associated with prognosis and chemoresistance in the training group. An internal testing dataset and the external dataset, GSE57495, were used for validation and showed a good performance of the gene signature. The high-risk group was enriched with multiple oncological pathways related to immunosuppression and was correlated with epidermal growth factor receptor (EGFR) expression, a target molecule of Erlotinib. In conclusion, this study identified a four-gene signature and established two nomograms for predicting prognosis and chemotherapy responses in patients with pancreatic cancer. The clinical value of the nomogram was evaluated by decision curve analysis (DCA). It showed that these may be helpful for clinical treatment decision-making and the discovery of the potential molecular mechanism and therapy targets for pancreatic cancer.

INTRODUCTION

Pancreatic cancer is one of the highly malignant tumors associated with a 5-year survival rate as low as 9%, and ranks fourth as a common cause of cancer-associated death in the USA [1]. However, it is estimated to become the second common cause of cancer-associated death before 2030 [2]. Surgical resection is considered the only potentially curative treatment, but less than 10% of patients are resectable using the standard resection, and the 5-year survival rate of these early

stage patients is only 24.6% [3]. Systemic chemotherapy, radiotherapy, and targeted molecular therapy are also treatment choices for postoperative patients or patients with unresectable tumors [4]. Recurrence, early metastasis, and resistance to chemoradiotherapy are important characteristics of pancreatic cancer [5]. Therefore, patients with pancreatic cancer should have individualized systemic treatment to prolong survival time and improve quality of life; thus, it is essential to identify an effective predictive prognosis model and biomarkers for guiding

individualized systemic treatment. Clinical characteristics including the American Joint Committee on Cancer (AJCC) stage and pathological type are commonly used as indicators for prognosis and treatment evaluation [6]. With the development of gene chips and high-throughput sequencing technology, prediction tools based on prognosis-related genes have been developed with better accuracy, and may be beneficial to elucidate the molecular mechanism of pancreatic cancer occurrence and progression.

One of the important factors for poor prognosis in pancreatic cancer is chemoresistance, including extrinsic or intrinsic resistance [7, 8]. Intrinsic resistance is driven by multiple mechanisms that are not clearly understood. For instance, Farrell JJ et al. has found that human equilibrative nucleoside transporter 1 (hENT1), a transport protein that transports gemcitabine and other nucleoside analogs into cellular compartments, has variable expression in patients with pancreatic cancer with different responses to chemotherapy [8, 9]. Ju HQ et al. has found that gemcitabine resistance in pancreatic cancer plays a role through redox modulation [10]. Extrinsic resistance results from changes in the stromal microenvironment, including dense fibrotic tumor stroma and immunosuppression [11–13]. At present, gemcitabine or S-1, a fluoro-pyrimidine derivative is commonly used as a first-line chemotherapy drug, while (m)FOLFIRINOX and gemcitabine plus nanoparticle albumin-bound paclitaxel are other choices for patients who can tolerate regimens [4]. With the continuous development of molecular biology technology, the molecular mechanisms of chemoresistance have been studied, and clinical trials of combination therapies, targeted molecular drugs, and immunotherapy are ongoing [14]. However, these studies have not contributed much to improving the survival of patients with pancreatic cancer. Erlotinib, a target inhibitor of the epidermal growth factor receptor (EGFR) tyrosine kinase, is currently recommended as a combination treatment for patients with skin rash, and is beneficial to the overall survival of patients with pancreatic cancer [4, 15]; but, immunotherapy has no efficacy in patients with pancreatic cancer due to the immunosuppressive tumor microenvironment [13].

Many gene signatures have been reported to be related to prognosis [16, 17]. However, they are seldom associated with the chemotherapy response. In this study, a four-gene prognostic signature related to the chemotherapy response was constructed based on the different responses after chemotherapy in The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) project datasets training group, and validated in both an internal testing group and an external Gene Expression Omnibus (GEO) dataset.

Moreover, the potential molecular mechanisms were explored through gene set enrichment analysis (GSEA) and tumor immunity relevance analysis. Two clinical nomogram models were also constructed to predict prognosis and the chemotherapy response and might serve as a reference for chemotherapy in patients with pancreatic cancer.

RESULTS

Identification of resistance-related differentially expressed genes (RRDEGs)

Figure 1 depicts a flowchart of the study. The clinical characteristics of patients with pancreatic cancer in TCGA dataset are shown in Supplementary Table 1. TCGA and GTEx datasets with 165 pancreatic cancer samples and 332 normal samples were used to conduct differential expression analysis between tumor samples and normal samples after removing the batch effect (Figure 2A, 2B); 980 differential expression genes (DEGs) (526 upregulated and 454 downregulated genes) were identified ($|log2FC| > 1$ and $p < 0.05$). As chemoresistance plays a significant role in the progression of pancreatic cancer, the expression of 71 patients with different responses after chemotherapy in the TCGA database was analyzed using a Kruskal-Wallis rank sum test. A total of 128 RRDEGs were identified ($p < 0.05$, Figure 2C, 2D and Supplementary Table 2).

GO and KEGG enrichment analysis of RRDEGs

To explore RRDEGs function, RRDEGs were subjected to GO and KEGG enrichment analysis (Supplementary Table 3). RRDEGs were enriched in biological processes related to neutrophils, such as neutrophil degranulation and activation, which suggested that innate immunity plays an important role in tumor progression and chemoresistance (Figure 3A). GO enrichment analysis in cellular component and molecular function indicated that the RRDEGs were related to cell adhesion, which was consistent with the highly aggressive nature of pancreatic cancer and extrinsic resistance mechanisms (Figure 3B, 3C). Furthermore, KEGG pathway analysis also revealed that RRDEGs were related to cell adhesion (Figure 3D).

A PPI network of RRDEGs was created to identify protein interactions; the top candidate hub included 25 genes that played a significant role in this network (Figure 3E). Module analysis identified a meaningful clustering module with score 4.5 in the PPI network, which was marked with white font on the network (Figure 3E). Functional enrichment analysis of biological processes for these hub genes was associated with the

immune response (Figure 3F), including the innate immune response, cytokine production, and lymphocyte activation. Thus, PPI network analysis showed that RRDEGs promoted pancreatic cancer progression and chemoresistance through immune regulation.

Construction of a four-gene signature associated with chemoresistance

Next, univariate Cox regression analysis was performed to identify potential prognostic RRDEGs in the training group; 51 RRDEGs were identified that were significantly associated with overall survival (Supplementary Table 4). Lasso-penalized Cox analysis was then used to eliminate model over-fitting and showed that the expression of 10 RRDEGs was correlated with overall survival in patients with

pancreatic cancer (Supplementary Figure 1A, 1B and Supplementary Table 5). Subsequently, a prognostic signature composed of four genes, including anoctamin 1 (*ANO1*), desmoglein 3 (*DSG3*), serine peptidase inhibitor Kazal type 1 (*SPINK1*) and GTPase, IMAP family member 1 (*GIMAP1*), was identified using stepwise multivariate Cox regression analysis (Figure 4). The downregulated gene *GIMAP1*, with a hazard ratio (*HR*) < 1, was regarded as a tumor suppressor, while the upregulated genes *ANO1*, *DSG3*, and *SPINK1*, with a *HR* > 1, were considered oncogenes. The risk score was calculated using the following formula:

$$\text{Risk score} = (0.47599 \times \text{Expression value of } ANO1)$$
$$+ (0.24553 \times \text{Expression value of } DSG3)$$
$$+ (0.23389 \times \text{Expression value of } SPINK1)$$
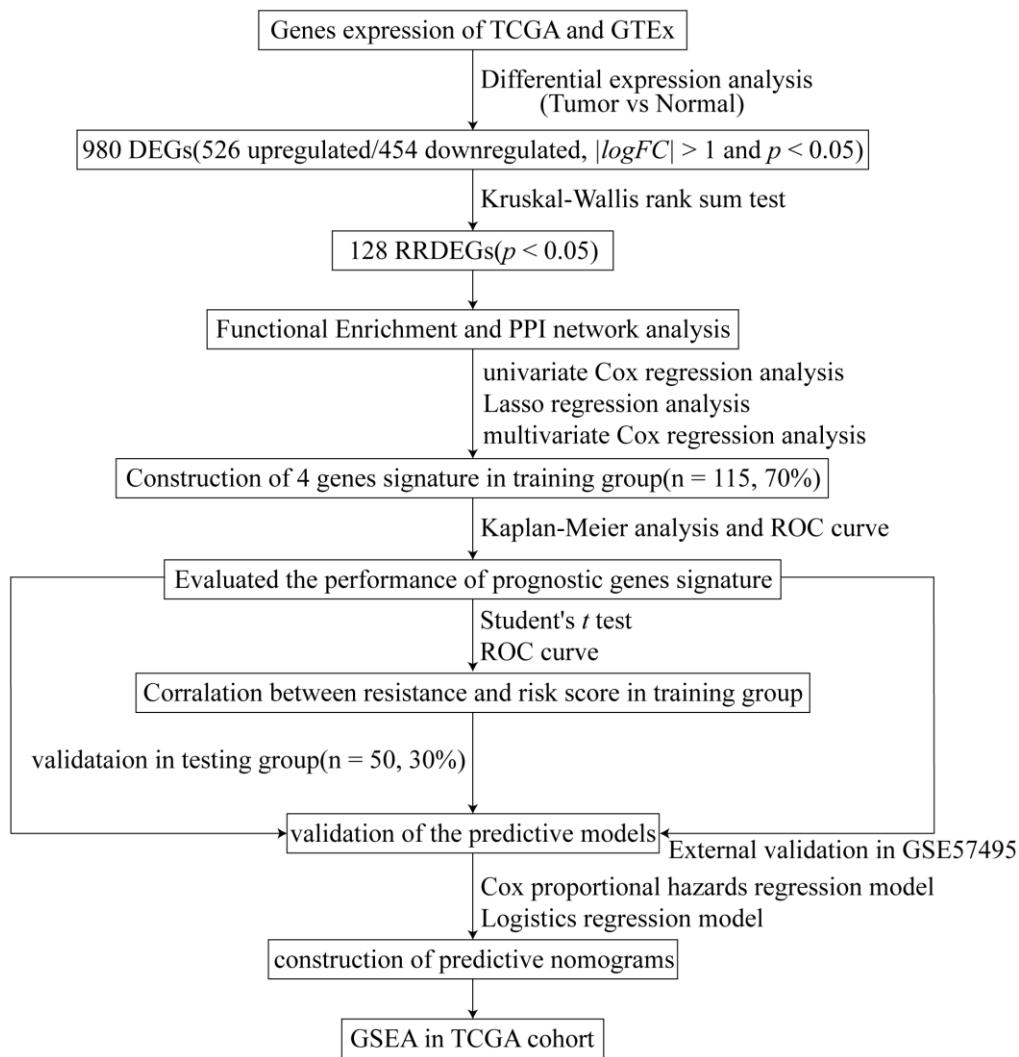$$+ [(-0.59799) \times \text{Expression value of } GIMAP1].$$

Figure 1. Study flowchart.

The patients in the training group were divided into two (cutoff value = 7.03) or three (cutoff values = 6.59 and 7.03, respectively) groups based on the optimal cutoff values generated by the X-tile software. Time-dependent ROC and C-index analyses were used to evaluate the prognostic value of the four-gene signature compared to the AJCC stage. The area under the curves (AUCs) for 1-, 2-, and 3-year overall survival predicted by the risk scores were 0.750 (95% CI: 0.631–0.869), 0.821 (95% CI: 0.725–0.916), and 0.770 (95% CI: 0.629–0.911), respectively (Figure 5A–5C). As a control, the AUCs for

1-, 2-, and 3-year overall survival predicted by the AJCC stage were 0.523 (95% CI: 0.419–0.628), 0.680 (95% CI: 0.569–0.791), and 0.725 (95% CI: 0.589–0.861), respectively. The C-index of the gene signature was 0.724 (95% CI: 0.650–0.798), while the C-index of the AJCC stage was 0.573(95% CI: 0.504–0.643). Kaplan-Meier survival curve analysis showed that patients with lower risk scores had a significantly more favorable prognosis (Figure 5D–5K). The calibration curve for the gene signature demonstrated a satisfactory fit in the training group (Figure 5L).
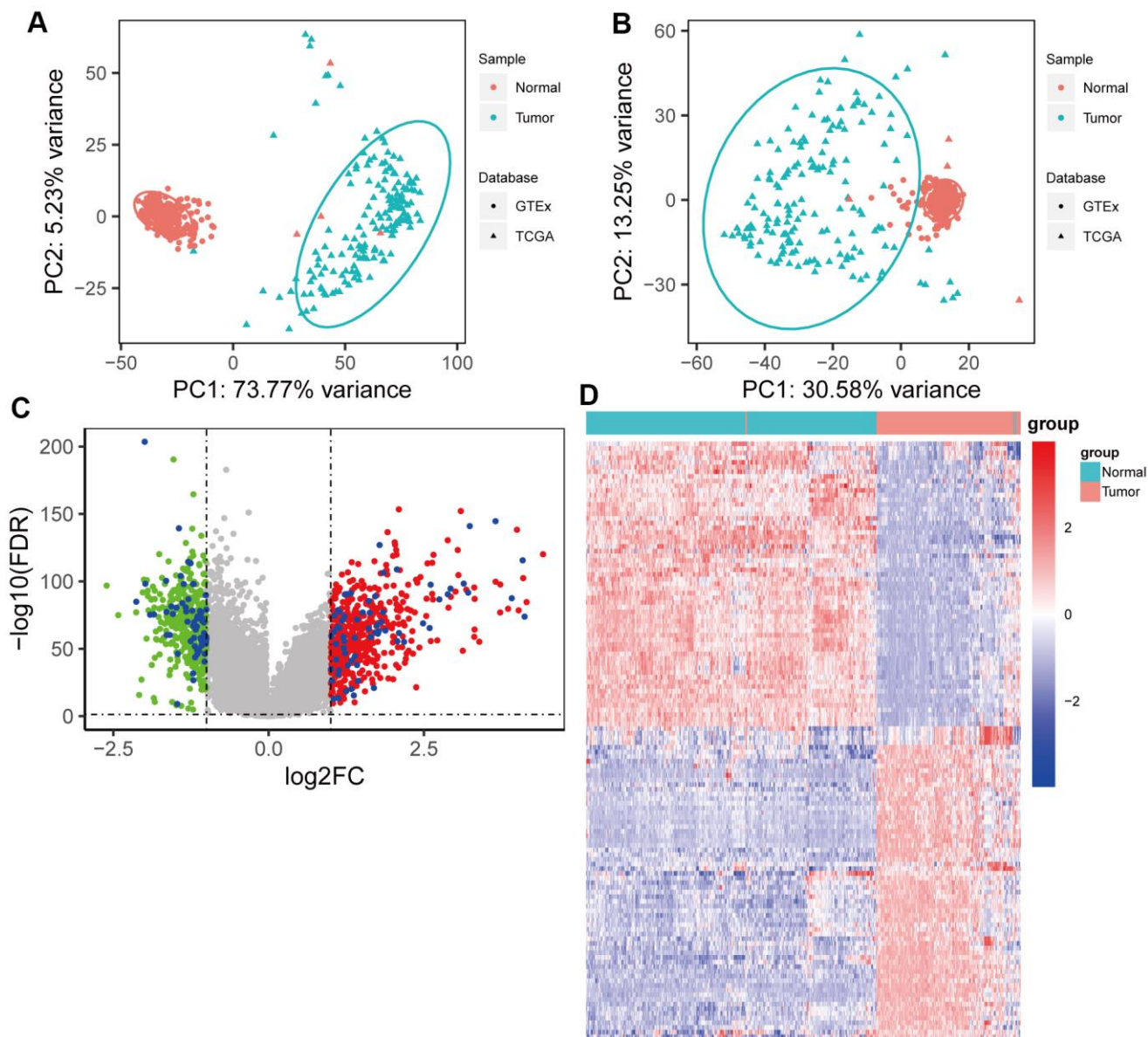

Figure 2. Identification of resistance-related differentially expressed genes (RRDEGs) in pancreatic cancer. (**A**) Principle component analysis (PCA) plot of the data merged from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) datasets before removing the batch effect. (**B**) PCA plot of the data merged from TCGA and GTEx datasets after removing the batch effect. (**C**) Volcano plot of DEGs between tumor tissues and normal tissues. The red and green points are DEGs, and RRDEGs are plotted with blue points. The lines were drawn where the absolute value of log2FC is equal to 1 and FDR is equal to 0.05. (**D**) Heatmap showing the expression of RRDEGs.

Chemoresistance is an important factor for poor progression of pancreatic cancer. In this study, the prognostic risk score based on patients with different responses after chemotherapy was closely related to chemoresistance in the training group ($p < 0.05$, Figure 5M). Moreover, the AUC for chemoresistance predicted by the risk score was 0.711(95% CI: 0.563–0.858, Figure 5N). In general, these data showed that the
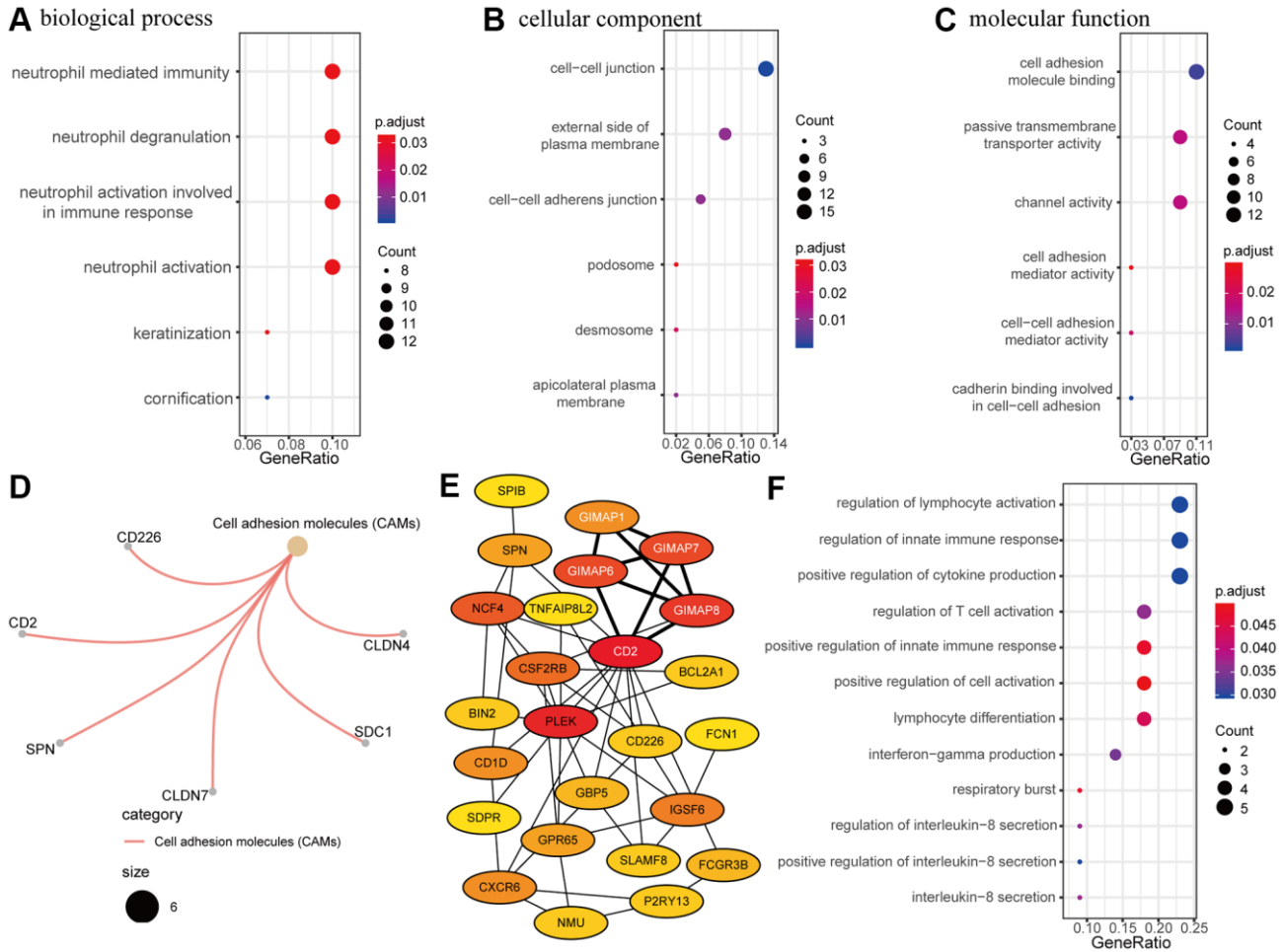

Figure 3. Functional enrichment and protein-protein interaction (PPI) network analysis. (**A**–**C**) Gene ontology (GO) enrichment analysis of resistance-related differentially expressed genes (RRDEGs). (**A**) Biological process. (**B**) Cellular component. (**C**) Molecular function. (**D**) Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of RRDEGs. (**E**) Visualization of the top 25 hub genes in the PPI network. A clustering module with a score of 4.5 is marked with white font. (**F**) GO enrichment analysis of hub genes in the biological process.
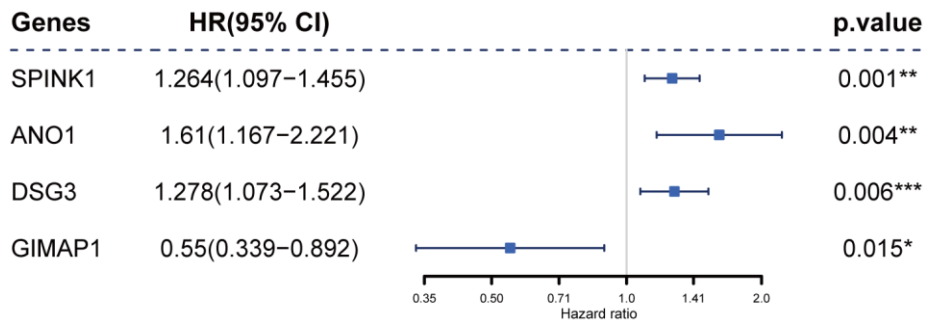

Figure 4. Construction of the prognostic four-gene signature using a Cox regression model.
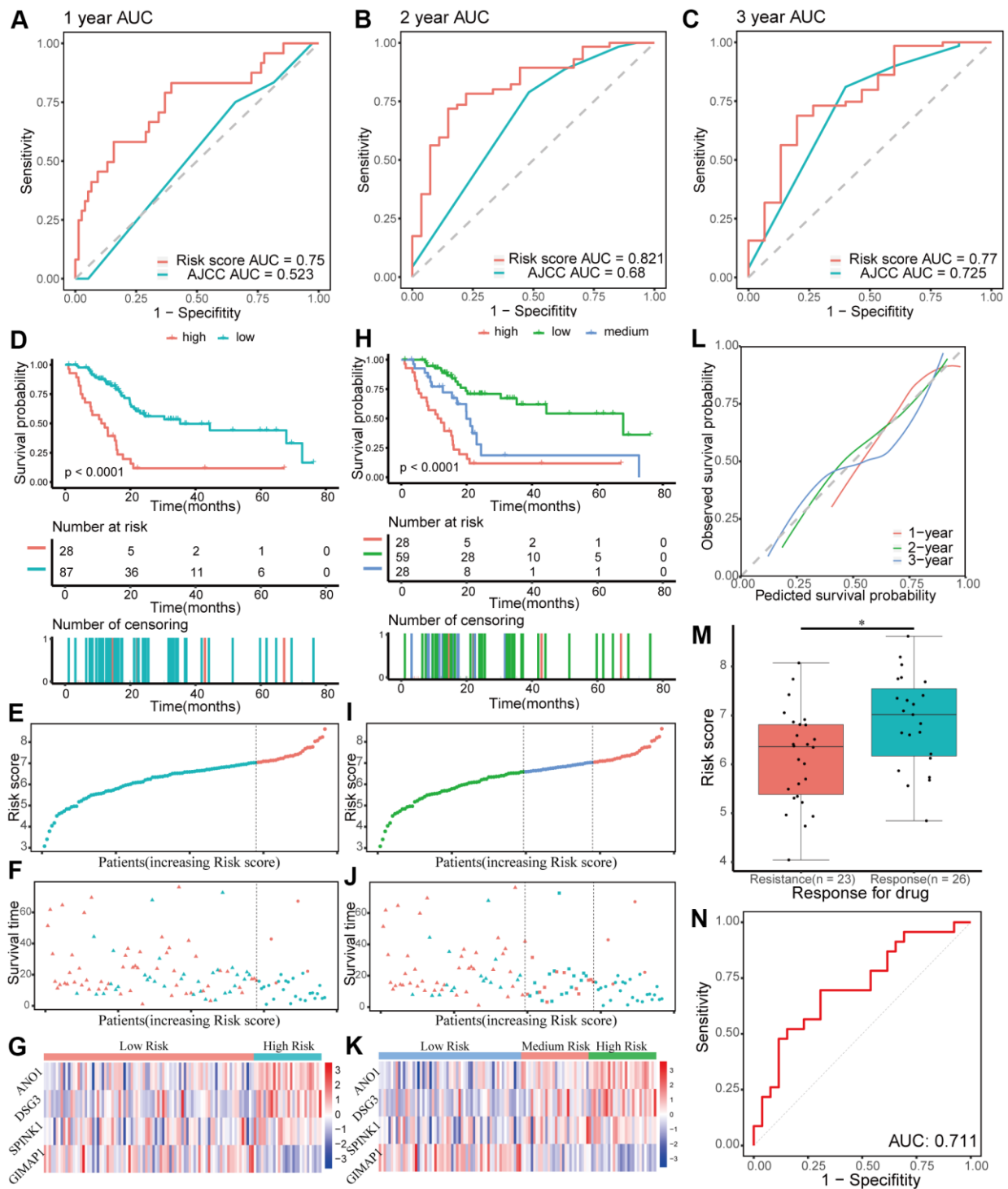
Figure 5. Evaluating the performance of the prognostic gene signature in the training group. (**A**–**C**) Time-dependent receiver operating characteristic (ROC) curves for 1-, 2-, and 3-year overall survival predictions of the gene signature. American Joint Committee on Cancer (AJCC) stage is the control. (**D**–**G**) Kaplan-Meier survival curves of the four-gene signature and distribution of patient survival and risk score in different groups when patients are divided into two groups. (**D**) Kaplan-Meier survival curves. (**E**) Distribution of risk score. (**F**) Distribution of survival time. Circle shape stands for high-risk group while triangle shape for low-risk group. Red stands for survival and green stands for dead. (**G**) Heatmap of the expression of the four genes. (**H**–**K**) Kaplan-Meier survival curves of the four-gene signature and distribution of patient survival and risk score in different groups when patients are divided into three groups. (**H**) Kaplan-Meier survival curves. (**I**) Distribution of risk score. (**J**) Distribution of survival time. (**K**) Heatmap of the expression of the four genes. (**L**) Calibration plot for validation of the gene signature. (**M**) Distribution of risk score in different responses to drug in the training group. (**N**) The ROC curve for the response to drug prediction of risk score in the training group is shown.

four-gene signature performed well in predicting prognosis and chemoresistance of pancreatic cancer.

Internal and external validation of the four-gene signature

The performance of the gene signature was validated in the internal testing group and external GEO dataset, GSE57495. Risk scores were calculated according to the same formula for each patient. In the testing group, patients were divided into two or three groups using the same cutoff values used in the training group. Due to sequencing using different platforms, patients in the GEO dataset were divided into two (cutoff value = 6.71) or three (cutoff values = 6.24 and 6.90, respectively) groups according to the optimal cutoff values. Time-dependent ROC and C-index analyses were utilized to elevate the prognostic predictive value of the four-gene signature compared with the AJCC stage. In the testing group, the AUCs for 1-, 2-, and 3-year overall survival predicted by the risk score were 0.696, 0.713, and 0.693, respectively (Figure 6A–6C). The AUCs for 1-, 2-, and 3-year overall survival predicted by the AJCC stage were 0.455, 0.597, and 0.604, respectively. The C-index of the gene signature was 0.639 (95% CI: 0.540–0.737), while the C-index of the AJCC stage was 0.522 (95% CI: 0.431–0.614). In the external GEO dataset, the AUCs for 1-, 2-, and 3-year overall survival predicted by the risk score were 0.615, 0.653, and 0.674, respectively (Supplementary Figure 2A–2C), while the AUCs for the AJCC stage were 0.575, 0.682, and 0.584, respectively. The C-index of the gene signature for the GEO dataset was 0.603 (95% CI: 0.528–0.677), while the C-index of the AJCC stage was 0.594 (95% CI: 0.509–0.680). Kaplan-Meier survival curve analysis showed significant differences for overall survival between the different risk score groups both in the testing group and the external GEO dataset ($p < 0.05$, Figure 6D–6K and Supplementary Figure 2D–2K). Furthermore, the calibration curve for the gene signature revealed that the predicted overall survival was approximately consistent with the actual overall survival (Figure 6L and Supplementary Figure 2L). However, when the predicted 3-year overall survival in the testing group was higher than 50%, the signature might underestimate the overall survival in patients with pancreatic cancer. In the testing group, the correlation between response to chemotherapy and risk score was also explored and showed that patients with chemoresistance had a higher risk score ($p < 0.01$, Figure 6M). Moreover, the AUC for chemoresistance predicted by the risk score was 0.858 (95% CI: 0.695–1, Figure 6N). The internal and external validation indicated that the four-gene signature performed well in predicting overall survival and chemoresistance in patients with pancreatic cancer.

Building predictive nomograms

Patients with complete clinical information from TCGA dataset were included in the analysis. The clinical characteristics included gender, age, AJCC TNM stage, tumor dimension, tumor site, pathologic grade, chemotherapy, neoadjuvants, radiation, molecular therapy, residual tumor after surgery, tobacco smoking history, alcohol use history, history of chronic pancreatitis, and diabetes. Prognostic factors of overall survival for pancreatic cancer were identified using univariate and stepwise multivariate Cox regression analyses while chemoresistance predictive factors for pancreatic cancer were identified using univariate and stepwise multivariate logistics regression analyses.

Univariate Cox regression analysis revealed that risk score ($p < 0.001$), AJCC stage ($p < 0.05$), T stage ($p < 0.05$), N stage ($p < 0.05$), tumor dimension ($p < 0.05$), tumor site ($p < 0.01$), pathologic grade ($p < 0.05$), chemotherapy ($p < 0.05$), radiation ($p < 0.05$), and molecular therapy ($p < 0.01$) were corrected with overall survival of patients with pancreatic cancer. Stepwise multivariate Cox regression analysis indicated that the risk score ($p < 0.001$), N stage ($p < 0.01$, N1 vs N0), and chemotherapy ($p < 0.001$) were significantly correlated with overall survival of patients with pancreatic cancer. A prognostic nomogram predicting 1-, 2-, and 3-year overall survival based on the stepwise multivariate Cox regression model was developed (Figure 7A). The AUC of the 1-, 2-, and 3-year overall survival predictions for the nomogram was 0.666, 0.721, and 0.752, respectively (Figure 7B). The C-index of the nomogram was 0.791 (95% CI: 0.732–0.849). The calibration plot showed that the nomogram was effective at predicting 1-, 2-, and 3- year overall survival in patients with pancreatic cancer (Figure 7C). In the DCA, the results showed that the nomogram indicated a better net benefit than was achieved with the AJCC stage for predicting OS (Figure 7D).

Univariate logistics regression analysis revealed that risk score ($p < 0.001$), residual tumor after surgery ($p < 0.05$, R1 vs R0), and N stage ($p < 0.05$) were corrected with response after chemotherapy. Additionally, stepwise logistics regression analysis showed that the risk score ($p < 0.01$) and residual tumor after surgery ($p < 0.05$, R1 vs R0) were independent risk factors of disease progression after chemotherapy. A predictive nomogram predicting the probability of response after chemotherapy was built based on the logistics regression model (Figure 7E). The AUC of the disease progression probability predictions for the nomogram was 0.847 (Figure 7F) while the C-index of the nomogram was 0.847. The calibration plot showed good agreement between predictions and observations (Figure 7G).
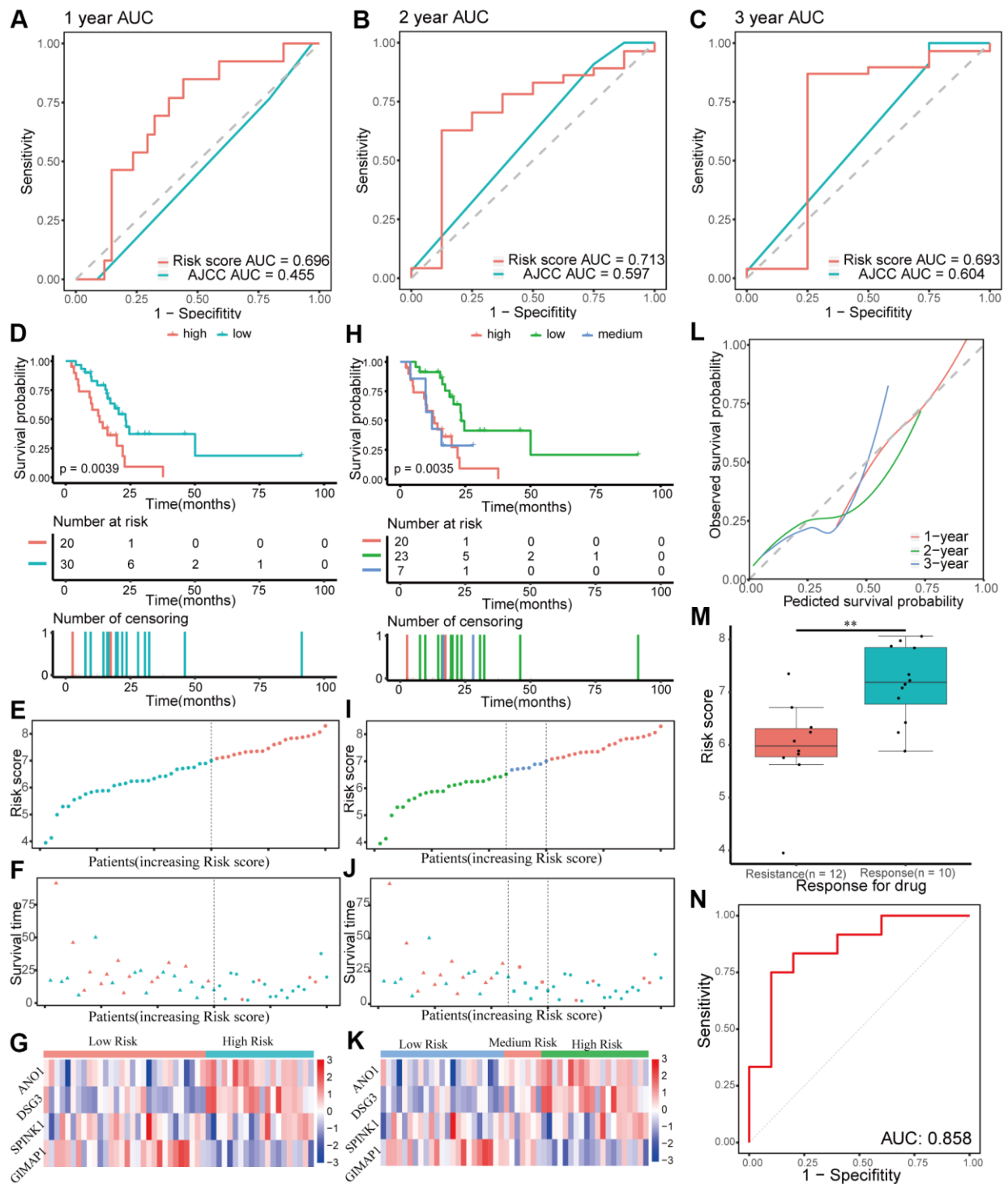
Figure 6. Validation of the prognostic gene signature in the testing group. (**A**–**C**) Time-dependent receiver operating characteristic (ROC) curves for 1-, 2-, and 3-year overall survival predictions of the gene signature in the testing group. American Joint Committee on Cancer (AJCC) stage is the control. (**D**–**G**) Kaplan-Meier survival curves of the four-gene signature and distribution of patient survival and risk score in different groups when patients in the testing group are divided into two groups. (**D**) Kaplan-Meier survival curves. (**E**) Distribution of risk score. (**F**) Distribution of survival time. Circle shape stands for high-risk group while triangle shape for low-risk group. Red stands for survival and green stands for dead. (**G**) Heatmap of the expression of the four genes. (**H**–**K**) Kaplan-Meier survival curves of the four-gene signature and distribution of patient survival and risk score in different groups when patients in the testing group are divided into three groups. (**H**) Kaplan-Meier survival curves. (**I**) Distribution of risk score. (**J**) Distribution of survival time. (**K**) Heatmap of the expression of the four genes. (**L**) Calibration plot for validation of the gene signature in the testing group. (**M**) Distribution of risk score in different responses to drug in the testing group. (**N**) The ROC curve for the drug reaction prediction of risk score in the testing group is shown.

GSEA and tumor immunity relevance of the gene signature

To elucidate the molecular mechanisms of the four-gene signature, 165 patients in TCGA dataset were divided into two groups according to the median of the risk score and GSEA was used to compare the high and low risk groups. The results revealed the malignant characteristics of cancer and immunity relevance and included cholangiocarcinoma, breast cancer, pancreas beta cell, *MYC*, keratinization, and mutation of *P53* and *KRAS* (Figure 8A and Supplementary Figure 3A–3C). Results of the GSEA are shown in Supplementary Table 6.

As shown above, RRDEGs might be related to tumor immunity. To further investigate the tumor immunity relevance of the gene signature, immune scores of TCGA-PAAD samples were downloaded from ESTIMATE, a database evaluating infiltrating immune cells in tumor tissues (https://bioinformatics.mdanderson.org/estimate/), and infiltration information of different immunocytes in TCGA-PAAD samples was downloaded from TIMER (http://timer.cistrome.org/). The immune score was significantly lower in the high-risk group, which revealed that the four-gene signature might play an important role in tumor immune escape (Figure 8B). Immune cell proportions were weakly to
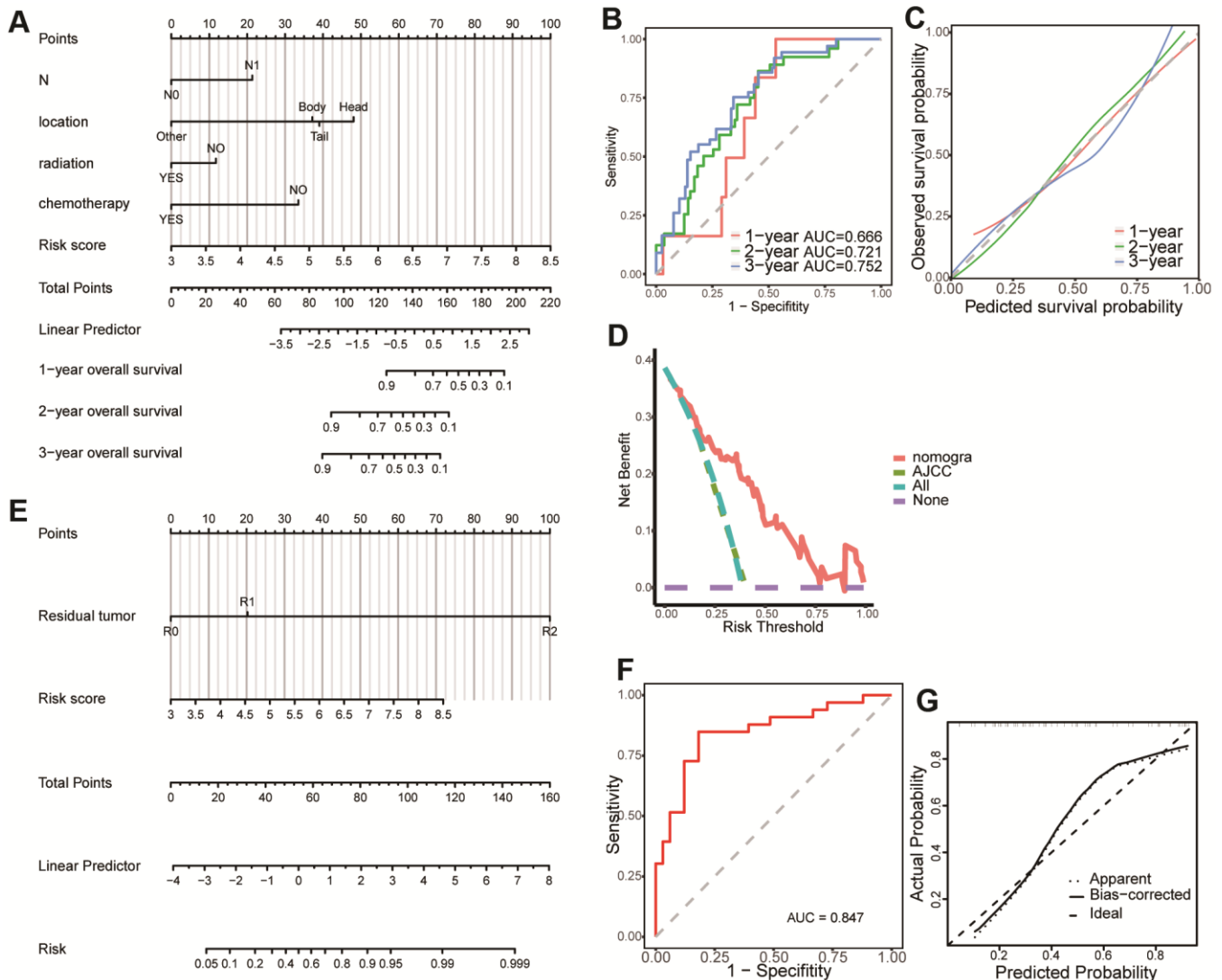

Figure 7. Creation of a predictive nomogram in The Cancer Genome Atlas (TCGA) dataset. (**A**) A prognostic nomogram predicting 1-, 2-, and 3-year overall survival of patients with pancreatic cancer. (**B**) Time-dependent receiver operating characteristic (ROC) curves for 1-, 2-, and 3-year overall survival predictions of the nomogram. (**C**) Calibration plot for the validation of the prognostic nomogram. (**D**) The nomogram predicting OS were compared against AJCC stage by DCA. (**E**) A predictive nomogram predicting drug reaction in pancreatic cancer. (**F**) ROC curve of the nomogram for the prediction of chemotherapy resistance. (**G**) Calibration plot for the validation of the chemotherapy resistance predictive nomogram.

moderately correlated in patients with pancreatic cancer (Figure 8C). Infiltration levels of different immunocytes, including naïve B cells, macrophage M0s, monocytes and CD8+ T cells, were significantly different between different risk groups (Figure 8D, 8E). To evaluate the correlation between risk score and immunotherapy, correlation analysis between the expression of the immune checkpoint programmed cell death protein 1 (PD-1) and risk score was performed; the results showed that risk score was negatively correlated with the expression level of PD-1 (correlation coefficient = -0.351, Figure 8F). However, there was not significance difference between signature and the expression level of PD-L1 (Supplementary Figure 4A).
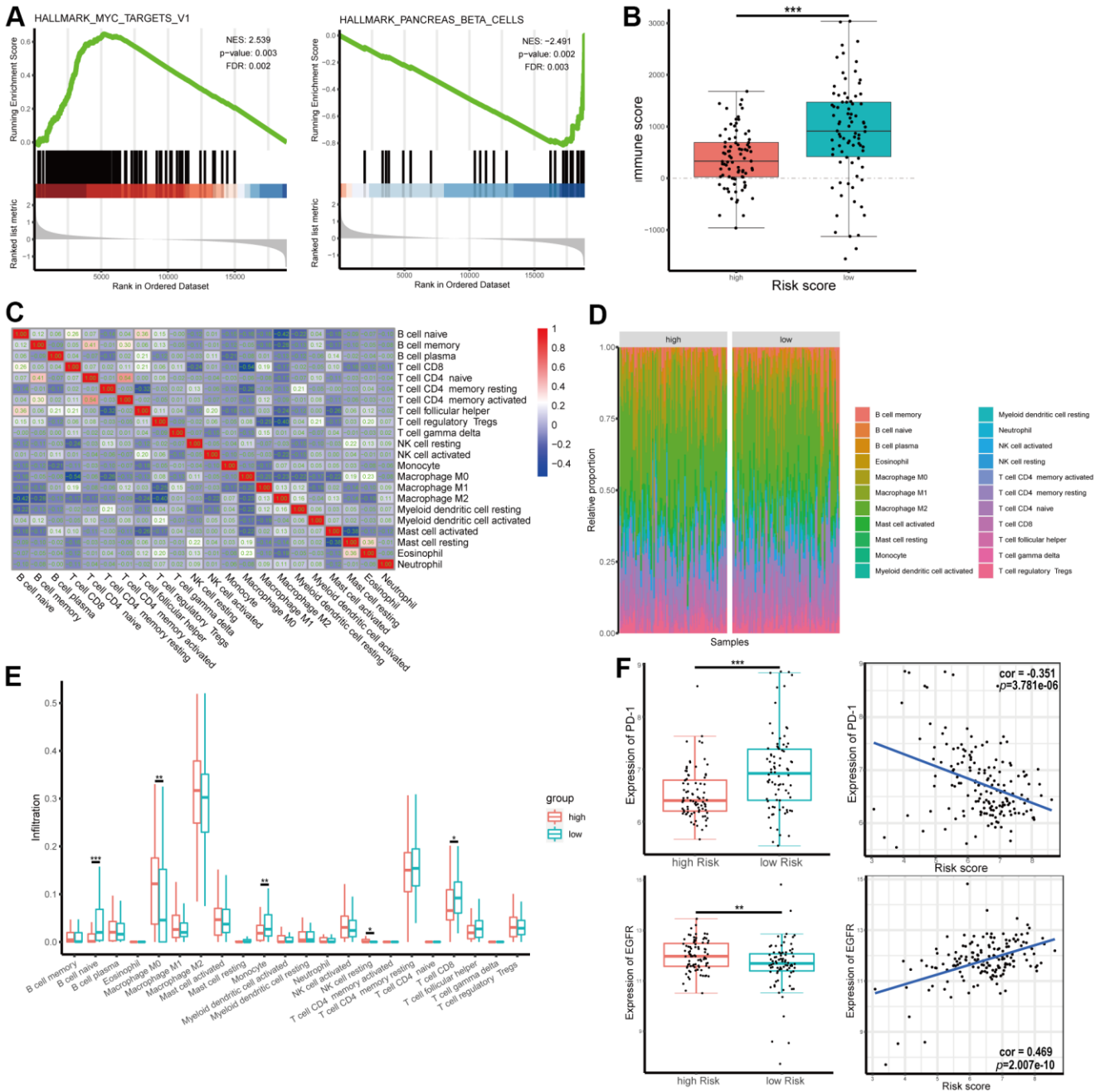

Figure 8. Gene set enrichment analysis (GSEA) and tumor immunity relevance of the gene signature. (**A**) GSEA for hallmark gene sets. Upregulated and downregulated enriched pathways with top normalized enrichment scores are shown. (**B**) Distribution of immune score in high and low risk groups in The Cancer Genome Atlas (TCGA) dataset. (**C**) Correlation of the infiltration proportion between different immunocytes. (**D**, **E**) Infiltration levels of different immunocytes in high and low risk groups. (**F**) Correlation between expression of programmed cell death protein 1 (PD-1) or epidermal growth factor receptor (EGFR) and risk score.

The expression of EGFR, a target molecule of Erlotinib, was positively correlated with risk score (correlation coefficient = 0.469, Figure 8F), which suggested that these risk scores might be used as a reference for Erlotinib-targeted drug therapy. To further explore the role and the clinical relevance of genes, we analyzed the expression level of mRNA from dataset. The mRNA expression of *SPINK1*, *DSG3*, *ANO1* were significantly increased in PAAD tumor tissue while *GIMAP1* were significantly decreased compared with normal tissues (Supplementary Figure 4B–4E).

DISCUSSION

Pancreatic cancer is a lethal disease where the 5-year survival rate is only 9% [1]. However, current treatments, including surgery, chemotherapy, radiation therapy, targeted molecular therapy, and immunotherapy have only led to modest improvements in overall survival. Accurate prognostic models provide a reference for personalized treatments strategies, which are expected to improve patient prognosis. The AJCC stage and other clinical features are commonly used as prognostic markers and are one of the reference indicators for clinical treatment strategies. Additionally, molecular prognostic markers provide good supplementation to the AJCC stage for improving accuracy of prognosis predictions [6]. Moreover, chemoresistance plays a significant role in the progress of pancreatic cancer, leads to poor prognosis, and may have effects through a complex molecular regulatory network [5, 7, 8, 12, 18, 19]. By analyzing the related molecules, these molecules can be used as prognostic molecular markers and help in the understanding of the mechanism of chemoresistance, which may provide new therapeutic targets. Nomograms that combine several prognostic risk factors, including molecular markers and clinical characteristics, are widely used to accurately calculate and visualize the probabilities of clinical events, and contribute to clinical decisions in personalized treatment strategies [20–22].

In this study, a four-gene prognostic signature that included *SPINK1*, *ANO1*, *DSG3*, and *GIMAP1* in patients with pancreatic cancer were identified. The gene signature was related to the chemotherapy response based on different responses in patients after chemotherapy in training group, and was validated in both an internal testing group and an external GEO dataset. The gene signature was closely correlated with prognosis, chemoresistance, and target molecular therapy of patients with pancreatic cancer. To explore the potential mechanisms of these genes in pancreatic cancer progression and chemoresistance, the patients were divided into high and low risk groups based on risk score and GSEA was performed. When using

oncogenic gene sets, a correlation was revealed with *KRAS* and *P53*, common genes with mutation in pancreatic cancer; it has been reported that *KRAS* and *P53* are associated with chemoresistance [23, 24]. Utilizing hallmark gene sets, *MYC*-related pathways and other pathways are enriched, which is consistent with previous studies where *MYC* has been reported to play a role in cell growth, proliferation, and tumorigenesis in pancreatic cancer [25, 26]. Subsequently, it was found that patients with a high-risk score had a lower immune score, which suggested that the high-risk group had a tumor immunosuppressive microenvironment. Furthermore, the relationship between risk score and tumor immunotherapy was explored and the results show that risk score is negatively correlated with the expression level of PD-1, which may be due to the immunosuppressive microenvironment and is consistent with the modest curative effect of anti-PD-1 drugs in pancreatic cancer [27]. Although there is no benefit for immunotherapy in pancreatic cancer [13], new regimens of immunotherapy may be one of the choices in the future.

SPINK1, a trypsin inhibitor, is secreted into pancreatic juice by pancreatic acinar cells. It is encoded by *SPINK1*, located in chromosomal region 5q32. Mutation of this gene is closely associated with idiopathic chronic pancreatitis [28], a risk factor of pancreatic cancer. However, the relationship between *SPINK1* and the development of pancreatic cancer is controversial [29, 30]. Chen F et al. has found that overexpression of *SPINK1* promotes pancreatic cancer aggressiveness, particularly chemoresistance, through the epithelial-endothelial transition mediated by EGFR downstream signaling [31]. However, a meta-analysis shows that *SPINK1* has no correlation with the progression of pancreatic cancer [30]. Overexpression of *SPINK1* has also been discovered in multiple tumors, including colon, lung, breast, and prostate, and is associated with tumor progression and poor prognosis [32, 33]. *ANO1*, a calcium-activated chloride channel, is a biomarker for poor prognosis in pancreatic cancer. Overexpression of *ANO1* promotes pancreatic cancer cell migration via the ligand-dependent EGFR signaling pathway [34, 35]. *ANO1* is also upregulated in multiple tumor tissues, including head and neck squamous cell carcinoma (HNSCC), and prostate and breast cancer [36, 37]. It is reported that *ANO1* interacts with EGFR and affects EGFR-targeted therapy in HNSCC and breast cancer [38, 39]. However, additional experiments are required to elucidate whether this occurs in pancreatic cancer. *DSG3*, encoded by *DSG3*, located in chromosomal region 18q12.1, is a member of the desmoglein family and the cadherin cell adhesion molecule superfamily of proteins that establish links between adjacent cells [40]. *DSG3* has been identified as an autoantigen in the skin

disease pemphigus vulgaris [41] and is also upregulated in several cancers, including squamous cell carcinoma, pancreatic ductal adenocarcinoma, and head and neck cancer [40, 42, 43]. In pancreatic cancer, it has been identified as a negative prognosis biomarker [40]. GIMAP1, a guanosine triphosphatase (GTPase) that is immunity-associated, is involved in Th cell differentiation and development of mature B and T lymphocytes [44]. GIMAP1 has been commonly reported to be related to autoimmune diseases, such as Bechet's disease [45] and type I diabetes [46]. Moreover, GIMAP1 is reported to be downregulated in lymphomas and regulates the B and T lymphocyte cell cycle [47]. However, there are no reports on its effect in pancreatic cancer.

There are limitations to this study. First, the model was conducted and validated in TCGA and GEO datasets and some clinical information was limited, especially the information on chemotherapy. Therefore, it is necessary to validate the model in a larger number of samples with complete clinical information. Second, the potential molecular mechanisms of progression and chemoresistance in pancreatic cancer that were identified in this study need to be verified by further experimental studies.

This study identified a four-gene signature that predicted prognosis and chemoresistance in patients with pancreatic cancer. In addition, two nomograms were established in pancreatic cancer. The gene signature was closely correlated with prognosis, chemoresistance, and target molecular therapy of patients with pancreatic cancer and might be beneficial to the discovery of potential mechanisms of progression and chemoresistance. Two predictive nomograms were created to be used as references for clinical treatment decisions.

MATERIALS AND METHODS

Gene expression data and clinical data

The gene expression profiles (HTSeq-Counts) of patients with pancreatic cancer (TCGA-PAAD) and their associated drug information were downloaded from TCGA using the R package "TCGAbiolinks," while the other clinical information was obtained from the cBioPortal database (https://www.cbioportal.org/). Samples meeting the following criteria were excluded: (1) a metastatic tumor; (2) without pathological information; (3) with a pathological diagnosis of colloid (mucinous non-cystic) carcinoma or undifferentiated carcinoma; and (4) follow-up time less than 30 days. Meanwhile, eight cases without corresponding gene expression data were eliminated. Therefore, 165 cases,

165 tumor samples and 4 normal samples, were included in this study. Gene expression profiles (gene read counts) from normal pancreatic tissues (n = 328) of healthy individuals were downloaded from the GTEx project (https://www.gtexportal.org/). Then, TCGA and GTEx gene expression data were combined for further analysis. The gene expression microarray dataset, GSE57495, associated with follow-up information, was downloaded from GEO (https://www.ncbi.nlm.nih.gov/geo/) for the validation of the prognostic model.

Differentially expressed gene (DEG) identification and analysis

Gene expression data were normalized and a variance stabilizing transformation using the R package "DESeq2" was applied. To reduce the influence of a batch effect, a batch variate giving rise to a different database was designed as a covariant, and the batch effect was further removed using the removeBatchEffect () function in the R package "limma." DEGs between tumor samples and normal samples were identified using "limma." $|log_2FC| > 1$ and $p < 0.05$ were set as the cutoffs for DEGs. Then, RRDEGs were identified using a Kruskal-Wallis rank sum test between patients with different responses after chemotherapy, including "Clinical Progressive Disease", "Partial Response", "Stable Disease" and "Complete Response". $p < 0.05$ was considered statistically different. Furthermore, gene ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed for RRDEGs using the R package "clusterProfiler." A protein–protein interaction (PPI) network was conducted using the search tool for the retrieval of interacting genes/proteins (STRING) database (https://www.string-db.org/) and was visualized using Cytoscape v. 3.7.2 (http://www.cytoscape.org/).

Construction of a risk prognostic signature

A total of 165 tumor samples were randomly divided into two groups, a training group ($n = 115$, 70%) and a testing group ($n = 50$, 30%). There were 49 samples with chemotherapy information in the training group, while 22 samples had chemotherapy information in the testing group. A risk prognostic model was constructed based on the data from the training group and validated in the testing group and the GSE57495 dataset. First, a univariate Cox proportional hazards regression model was performed to identify potential survival-related RRDEGs for subsequent analysis. Lasso-penalized Cox regression analysis was performed to reduce over-fitting based on the "glmnet" package. Next, a stepwise multivariate Cox proportional hazards regression model was used to construct a risk prognostic signature, and

the risk score was calculated through the summation of the gene expression multiplied by the regression coefficients from the multivariate Cox proportional hazards regression model. Patients with different responses after chemotherapy were divided into two groups. Patients with a response from "Clinical Progressive Disease" were assigned to one group and the others were assigned to another group. A student's t-test was performed to evaluate the correlation between risk score and chemoresistance. The receiver operating characteristic (ROC) curve was performed to evaluate the predictive power of the gene signature for chemoresistance. Patients were separated into two or three groups based on an optimal cutoff value risk score using X-tile software (V3.6.1). Kaplan-Meier analysis, ROC curves, Harrell's concordance indexes (C-index), and calibration plots were conducted to evaluate the performance of the prognostic gene signature. The AJCC stage was used as a control. The GSE57495 dataset, with follow-up information, was set as an external validation dataset and the same formula was used to calculate the risk score.

Construction of nomograms and DCA

To identify independent risk factors of prognosis, univariate and multivariate Cox regression analyses were performed, and clinical characteristics included age, gender, AJCC stage, T, N, M, dimension, location, grade, histopathology, neoadjuvant, radiation, molecular therapy, residual tumor, smoking, drinking, pancreatitis, diabetes, chemotherapy, and risk score. $p < 0.05$ was considered statistically different. All independent risk factors were included in the construction of prognostic nomograms using a stepwise Cox regression model in TCGA PAAD dataset. A univariate and multivariate binary logistics regression model was performed to identify chemoresistance risk factors, and all were included in the construction of a predictive nomogram using a stepwise method. The ROC curve, C-index, calibration plots, and DCA were conducted to evaluate the performance of the nomograms.

GSEA

GSEA was used to explore the potential molecular mechanism of a prognostic gene signature. All patients in TCGA dataset were divided into two groups according to the median of the risk score. The R package "limma" was used to acquire the DEGs between high and low risk groups, and GSEA was performed using the R package "clusterProfiler" [48], based on the Molecular Signatures Database v. 7.0. Hallmark gene sets included C2 (curated gene sets), C5 (GO gene sets) and C6 (oncogenic signatures). $p < 0.05$ was regarded as statistically different.

Data of tumor immunity relevance in pancreatic cancer

The data on tumor purity and the presence of infiltrating stromal/immune cells in TCGA PAAD sample tumor tissue were download from the Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) [49] (https://bioinformatics.mdanderson.org/estimate/). The abundance of different immunocyte immune infiltrates' in TCGA PAAD samples was downloaded from TIMER (http://timer.cistrome.org/) [50] and the CIBERSORT method was used to acquire the results [51].

Statistical analysis

Statistical analysis was performed in the programming language R (v3.6.2, https://www.r-project.org/). Pearson's chi-squared test was used to analyze categorical variables. A Wilcoxon signed rank test or Student's t-test was used to compare two groups of continuous variables. A Kruskal-Wallis rank sum test was used to analyze multiple groups of continuous variables. Lasso regression analysis, and univariate and multivariate Cox regression analyses were performed for survival analysis. Univariate and multivariate binary logistics regression analysis were performed for regression analysis of binary category variables. $p < 0.05$ was regarded as statistically different.

Abbreviations

RRDEGs: Resistance-related differentially expressed genes; ROC curve: receiver operating characteristic curve; C-index: Harrell's concordance index; HR: Hazard ratios; CI: Confidence intervals; AUC: Area under the receiver operating characteristic curve; FDR: False discovery rate; NES: normalized Enrichment score. GSEA: Gene set enrichment analysis; PAAD: Pancreatic adenocarcinoma; TCGA: The Cancer Genome Atlas; GTEx: The Genotype-Tissue Expression project; GEO: The Gene Expression Omnibus; DEGs: Differential Expression Genes; PPI: protein–protein interaction; ESTIMATE: Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data.

Editorial note

&This corresponding author has a verified history of publications using a personal email address for correspondence.

AUTHOR CONTRIBUTIONS

RC and QZ: conception, design and supervision. HL, CH and SZ: development of methodology, analysis and interpretation of data. HL and XZ: the draft of the article. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020; 70:7–30.
https://doi.org/10.3322/caac.21590
PMID:31912902

2. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res. 2014; 74:2913–21.
https://doi.org/10.1158/0008-5472.CAN-14-0155
PMID:24840647

3. Strobel O, Neoptolemos J, Jäger D, Büchler MW. Optimizing the outcomes of pancreatic cancer surgery. Nat Rev Clin Oncol. 2019; 16:11–26.
https://doi.org/10.1038/s41571-018-0112-1
PMID:30341417

4. Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic cancer. Lancet. 2016; 388:73–85.
https://doi.org/10.1016/S0140-6736(16)00141-0
PMID:26830752

5. Borazanci E, Dang CV, Robey RW, Bates SE, Chabot JA, Von Hoff DD. Pancreatic cancer: "A riddle wrapped in a mystery inside an enigma". Clin Cancer Res. 2017; 23:1629–37.
https://doi.org/10.1158/1078-0432.CCR-16-2070
PMID:28373361

6. Kamarajah SK, Burns WR, Frankel TL, Cho CS, Nathan H. Validation of the American Joint Commission on Cancer (AJCC) 8th edition staging system for patients with pancreatic adenocarcinoma: A Surveillance, Epidemiology and End Results (SEER) analysis. Ann Surg Oncol. 2017; 24:2023–30.
https://doi.org/10.1245/s10434-017-5810-x
PMID:28213792

7. Binenbaum Y, Na'ara S, Gil Z. Gemcitabine resistance in pancreatic ductal adenocarcinoma. Drug Resist Updat. 2015; 23:55–68.
https://doi.org/10.1016/j.drup.2015.10.002
PMID:26690340

8. Koay EJ, Truty MJ, Cristini V, Thomas RM, Chen R, Chatterjee D, Kang Y, Bhosale PR, Tamm EP, Crane CH, Javle M, Katz MH, Gottumukkala VN, et al. Transport properties of pancreatic cancer describe gemcitabine delivery and response. J Clin Invest. 2014; 124:1525–36.
https://doi.org/10.1172/JCI73455
PMID:24614108

9. Farrell JJ, Elsaleh H, Garcia M, Lai R, Ammar A, Regine WF, Abrams R, Benson AB, Macdonald J, Cass CE, Dicker AP, Mackey JR. Human equilibrative nucleoside transporter 1 levels predict response to gemcitabine in patients with pancreatic cancer. Gastroenterology. 2009; 136:187–95.
https://doi.org/10.1053/j.gastro.2008.09.067
PMID:18992248

10. Ju HQ, Gocho T, Aguilar M, Wu M, Zhuang ZN, Fu J, Yanaga K, Huang P, Chiao PJ. Mechanisms of overcoming intrinsic resistance to gemcitabine in pancreatic ductal adenocarcinoma through the redox modulation. Mol Cancer Ther. 2015; 14:788–98.
https://doi.org/10.1158/1535-7163.MCT-14-0420
PMID:25527634

11. Neesse A, Michl P, Frese KK, Feig C, Cook N, Jacobetz MA, Lolkema MP, Buchholz M, Olive KP, Gress TM, Tuveson DA. Stromal biology and therapy in pancreatic cancer. Gut. 2011; 60:861–68.
https://doi.org/10.1136/gut.2010.226092
PMID:20966025

12. Neesse A, Algül H, Tuveson DA, Gress TM. Stromal biology and therapy in pancreatic cancer: a changing paradigm. Gut. 2015; 64:1476–84.
https://doi.org/10.1136/gutjnl-2015-309304
PMID:25994217

13. Morrison AH, Byrne KT, Vonderheide RH. Immunotherapy and prevention of pancreatic cancer. Trends Cancer. 2018; 4:418–28.
https://doi.org/10.1016/j.trecan.2018.04.001

PMID:29860986

14. Diab M, Azmi A, Mohammad R, Philip PA. Pharmacotherapeutic strategies for treating pancreatic cancer: advances and challenges. Expert Opin Pharmacother. 2019; 20:535–46. https://doi.org/10.1080/14656566.2018.1561869 PMID:30592647

15. Moore MJ, Goldstein D, Hamm J, Figer A, Hecht JR, Gallinger S, Au HJ, Murawa P, Walde D, Wolff RA, Campos D, Lim R, Ding K, et al, and National Cancer Institute of Canada Clinical Trials Group. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol. 2007; 25:1960–66. https://doi.org/10.1200/JCO.2006.07.9525 PMID:17452677

16. Wu M, Li X, Zhang T, Liu Z, Zhao Y. Identification of a nine-gene signature and establishment of a prognostic nomogram predicting overall survival of pancreatic cancer. Front Oncol. 2019; 9:996. https://doi.org/10.3389/fonc.2019.00996 PMID:31612115

17. Sun M, Liu X, Xia L, Chen Y, Kuang L, Gu X, Li T. A nine-lncRNA signature predicts distant relapse-free survival of HER2-negative breast cancer patients receiving taxane and anthracycline-based neoadjuvant chemotherapy. Biochem Pharmacol. 2020; 114285. https://doi.org/10.1016/j.bcp.2020.114285 PMID:33069665

18. Wu X, Yan F, Wang L, Sun G, Liu J, Qu M, Wang Y, Li T. MicroRNA: another pharmacological avenue for colorectal cancer? Front Cell Dev Biol. 2020; 8:812. https://doi.org/10.3389/fcell.2020.00812 PMID:32984321

19. Shi F, Li T, Liu Z, Qu K, Shi C, Li Y, Qin Q, Cheng L, Jin X, Yu T, Di W, Que J, Xia H, She J. FOXO1: another avenue for treating digestive malignancy? Semin Cancer Biol. 2018; 50:124–31. https://doi.org/10.1016/j.semcancer.2017.09.009 PMID:28965871

20. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. Lancet Oncol. 2015; 16:e173–80. https://doi.org/10.1016/S1470-2045(14)71116-7 PMID:25846097

21. Li T, Jiang S, Lu C, Yang W, Yang Z, Hu W, Xin Z, Yang Y. Melatonin: another avenue for treating osteoporosis? J Pineal Res. 2019; 66:e12548. https://doi.org/10.1111/jpi.12548 PMID:30597617

22. Li T, Jiang S, Han M, Yang Z, Lv J, Deng C, Reiter RJ, Yang Y. Exogenous melatonin as a treatment for secondary sleep disorders: A systematic review and meta-analysis. Front Neuroendocrinol. 2019; 52:22–28. https://doi.org/10.1016/j.yfrne.2018.06.004 PMID:29908879

23. Mann KM, Ying H, Juan J, Jenkins NA, Copeland NG. KRAS-related proteins in pancreatic cancer. Pharmacol Ther. 2016; 168:29–42. https://doi.org/10.1016/j.pharmthera.2016.09.003 PMID:27595930

24. Quan MY, Guo Q, Liu J, Yang R, Bai J, Wang W, Cai Y, Han R, Lv YQ, Ding L, Billadeau DD, Lou Z, Bellusci S, et al. An FGFR/AKT/SOX2 signaling axis controls pancreatic cancer stemness. Front Cell Dev Biol. 2020; 8:287. https://doi.org/10.3389/fcell.2020.00287 PMID:32457900

25. Witkiewicz AK, McMillan EA, Balaji U, Baek G, Lin WC, Mansour J, Mollaee M, Wagner KU, Koduru P, Yopp A, Choti MA, Yeo CJ, McCue P, et al. Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. Nat Commun. 2015; 6:6744. https://doi.org/10.1038/ncomms7744 PMID:25855536

26. Li T, Yang Z, Jiang S, Di W, Ma Z, Hu W, Chen F, Reiter RJ, Yang Y. Melatonin: does it have utility in the treatment of haematological neoplasms? Br J Pharmacol. 2018; 175:3251–62. https://doi.org/10.1111/bph.13966 PMID:28880375

27. Feng M, Xiong G, Cao Z, Yang G, Zheng S, Song X, You L, Zheng L, Zhang T, Zhao Y. PD-1/PD-L1 and immunotherapy for pancreatic cancer. Cancer Lett. 2017; 407:57–65. https://doi.org/10.1016/j.canlet.2017.08.006 PMID:28826722

28. Singh VK, Yadav D, Garg PK. Diagnosis and management of chronic pancreatitis: A review. JAMA. 2019; 322:2422–34. https://doi.org/10.1001/jama.2019.19411 PMID:31860051

29. Suzuki M, Shimizu T. Is SPINK1 gene mutation associated with development of pancreatic cancer? New insight from a large retrospective study. EBioMedicine. 2019; 50:5–6. https://doi.org/10.1016/j.ebiom.2019.10.065 PMID:31759918

30. Cazacu IM, Farkas N, Garami A, Balaskó M, Mosdósi B, Alizadeh H, Gyöngyi Z, Rakonczay Z Jr, Vigh É, Habon T, Czopf L, Lazarescu MA, Erőss B, et al. Pancreatitis-associated genes and pancreatic cancer risk: A systematic review and meta-analysis. Pancreas. 2018; 47:1078–86.

https://doi.org/10.1097/MPA.0000000000001145
PMID:30134356

31. Chen F, Long Q, Fu D, Zhu D, Ji Y, Han L, Zhang B, Xu Q, Liu B, Li Y, Wu S, Yang C, Qian M, et al. Targeting SPINK1 in the damaged tumour microenvironment alleviates therapeutic resistance. Nat Commun. 2018; 9:4315.
https://doi.org/10.1038/s41467-018-06860-4
PMID:30333494

32. Tiwari R, Pandey SK, Goel S, Bhatia V, Shukla S, Jing X, Dhanasekaran SM, Ateeq B. SPINK1 promotes colorectal cancer progression by downregulating Metallothioneins expression. Oncogenesis. 2015; 4:e162.
https://doi.org/10.1038/oncsis.2015.23
PMID:26258891

33. Attard G, Parker C, Eeles RA, Schröder F, Tomlins SA, Tannock I, Drake CG, de Bono JS. Prostate cancer. Lancet. 2016; 387:70–82.
https://doi.org/10.1016/S0140-6736(14)61947-4
PMID:26074382

34. Sauter DR, Novak I, Pedersen SF, Larsen EH, Hoffmann EK. ANO1 (TMEM16A) in pancreatic ductal adenocarcinoma (PDAC). Pflugers Arch. 2015; 467:1495–508.
https://doi.org/10.1007/s00424-014-1598-8
PMID:25163766

35. Crottès D, Lin YT, Peters CJ, Gilchrist JM, Wiita AP, Jan YN, Jan LY. TMEM16A controls EGF-induced calcium signaling implicated in pancreatic cancer prognosis. Proc Natl Acad Sci USA. 2019; 116:13026–35.
https://doi.org/10.1073/pnas.1900703116
PMID:31182586

36. Liu W, Lu M, Liu B, Huang Y, Wang K. Inhibition of Ca(2+)-activated Cl(-) channel ANO1/TMEM16A expression suppresses tumor growth and invasiveness in human prostate carcinoma. Cancer Lett. 2012; 326:41–51.
https://doi.org/10.1016/j.canlet.2012.07.015
PMID:22820160

37. Ayoub C, Wasylyk C, Li Y, Thomas E, Marisa L, Robé A, Roux M, Abecassis J, de Reyniès A, Wasylyk B. ANO1 amplification and expression in HNSCC with a high propensity for future distant metastasis and its functions in HNSCC cell lines. Br J Cancer. 2010; 103:715–26.
https://doi.org/10.1038/sj.bjc.6605823
PMID:20664600

38. Kulkarni S, Bill A, Godse NR, Khan NI, Kass JI, Steehler K, Kemp C, Davis K, Bertrand CA, Vyas AR, Holt DE, Grandis JR, Gaither LA, Duvvuri U. TMEM16A/ANO1 suppression improves response to antibody-mediated

targeted therapy of EGFR and HER2/ERBB2. Genes Chromosomes Cancer. 2017; 56:460–71.
https://doi.org/10.1002/gcc.22450
PMID:28177558

39. Bill A, Gutierrez A, Kulkarni S, Kemp C, Bonenfant D, Voshol H, Duvvuri U, Gaither LA. ANO1/TMEM16A interacts with EGFR and correlates with sensitivity to EGFR-targeting therapy in head and neck cancer. Oncotarget. 2015; 6:9173–88.
https://doi.org/10.18632/oncotarget.3277
PMID:25823819

40. Ormanns S, Altendorf-Hofmann A, Jackstadt R, Horst D, Assmann G, Zhao Y, Bruns C, Kirchner T, Knösel T. Desmogleins as prognostic biomarkers in resected pancreatic ductal adenocarcinoma. Br J Cancer. 2015; 113:1460–66.
https://doi.org/10.1038/bjc.2015.362
PMID:26469831

41. Schmidt E, Kasperkiewicz M, Joly P. Pemphigus. Lancet. 2019; 394:882–94.
https://doi.org/10.1016/S0140-6736(19)31778-7
PMID:31498102

42. Savci-Heijink CD, Kosari F, Aubry MC, Caron BL, Sun Z, Yang P, Vasmatzis G. The role of desmoglein-3 in the diagnosis of squamous cell carcinoma of the lung. Am J Pathol. 2009; 174:1629–37.
https://doi.org/10.2353/ajpath.2009.080778
PMID:19342368

43. Chen YJ, Chang JT, Lee L, Wang HM, Liao CT, Chiu CC, Chen PJ, Cheng AJ. DSG3 is overexpressed in head neck cancer and is a potential molecular target for inhibition of oncogenesis. Oncogene. 2007; 26:467–76.
https://doi.org/10.1038/sj.onc.1209802
PMID:16878157

44. Saunders A, Webb LM, Janas ML, Hutchings A, Pascall J, Carter C, Pugh N, Morgan G, Turner M, Butcher GW. Putative GTPase GIMAP1 is critical for the development of mature B and T lymphocytes. Blood. 2010; 115:3249–57.
https://doi.org/10.1182/blood-2009-08-237586
PMID:20194894

45. Lee YJ, Horie Y, Wallace GR, Choi YS, Park JA, Choi JY, Song R, Kang YM, Kang SW, Baek HJ, Kitaichi N, Meguro A, Mizuki N, et al. Genome-wide association study identifies GIMAP as a novel susceptibility locus for Behcet's disease. Ann Rheum Dis. 2013; 72:1510–16.
https://doi.org/10.1136/annrheumdis-2011-200288
PMID:23041938

46. Heinonen MT, Laine AP, Söderhäll C, Gruzieva O, Rautio S, Melén E, Pershagen G, Lähdesmäki HJ, Knip M, Ilonen J, Henttinen TA, Kere J, Lahesmaa R, and Finnish Pediatric Diabetes Registry. GIMAP GTPase

family genes: potential modifiers in autoimmune diabetes, asthma, and allergy. J Immunol. 2015; 194:5885–94.
https://doi.org/10.4049/jimmunol.1500016
PMID:25964488

47. Cambot M, Aresta S, Kahn-Perlès B, de Gunzburg J, Roméo PH. Human immune associated nucleotide 1: a member of a new guanosine triphosphatase family expressed in resting T and B cells. Blood. 2002; 99:3293–301.
https://doi.org/10.1182/blood.V99.9.3293
PMID:11964296

48. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16:284–87.
https://doi.org/10.1089/omi.2011.0118
PMID:22455463

49. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013; 4:2612.
https://doi.org/10.1038/ncomms3612
PMID:24113773

50. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S, Signoretti S, Liu JS, Liu XS. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol. 2016; 17:174.
https://doi.org/10.1186/s13059-016-1028-7
PMID:27549193

51. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015; 12:453–57.
https://doi.org/10.1038/nmeth.3337
PMID:25822800

Supplementary Figures
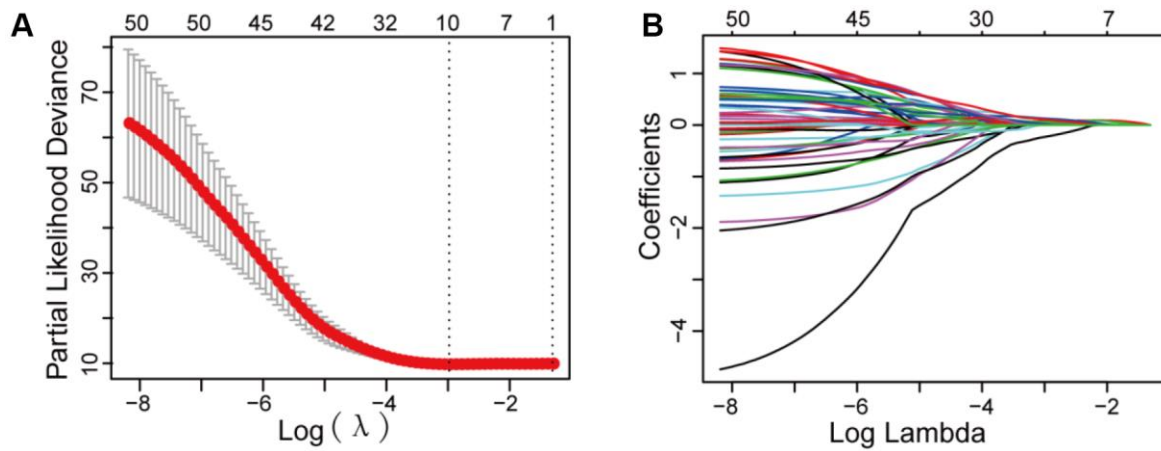
Supplementary Figure 1. Lasso analysis of the prognostic RRDEGs in pancreatic cancer. (**A**) Selection of tuning parameter (λ) in LASSO model. (**B**) LASSO coefficient profiles of all genes.
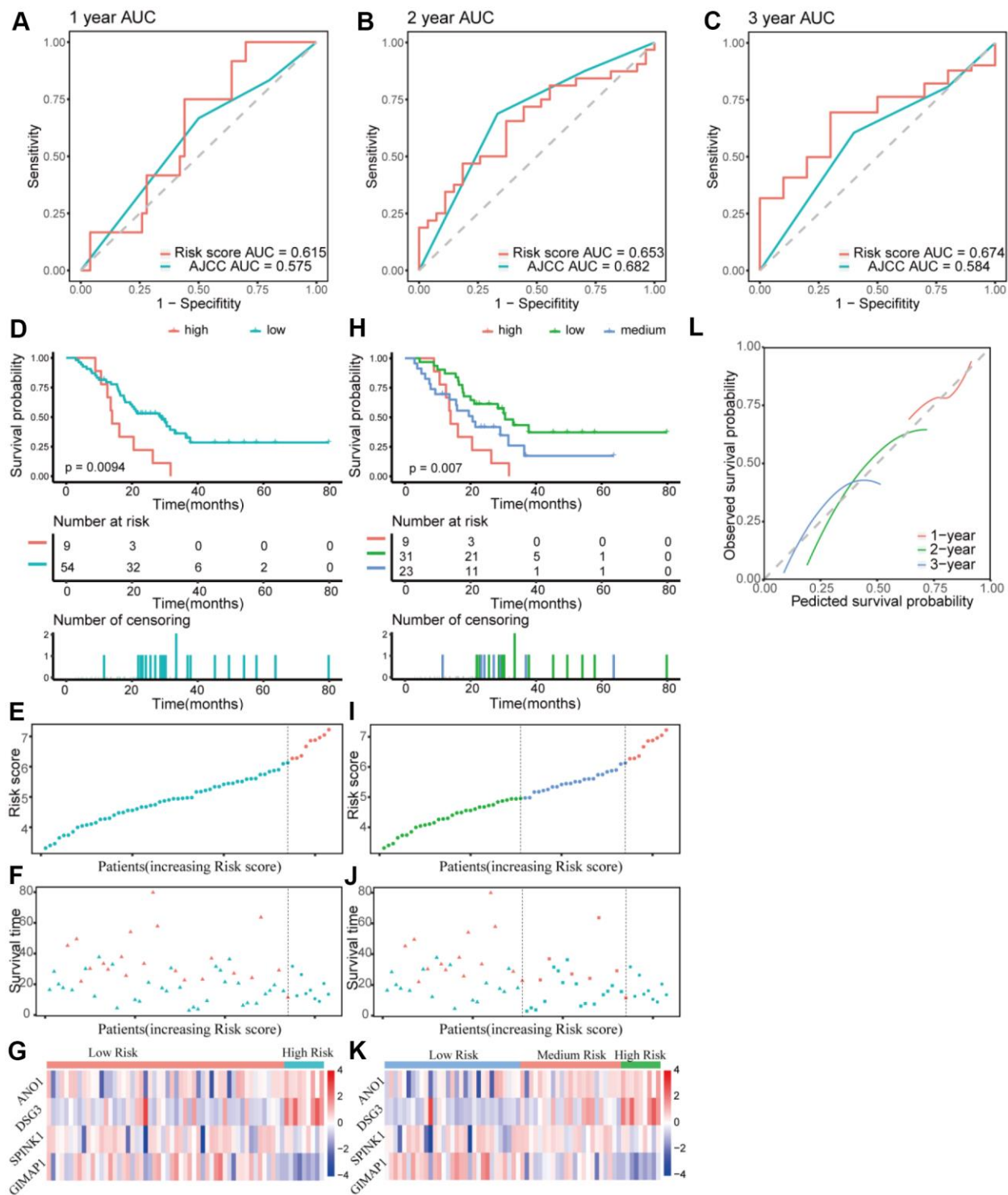
Supplementary Figure 2. External validation of the prognostic gene signature. (**A–C**) Time dependent ROC curves for 1-, 2- and 3- years overall survival predictions of genes signature in GSE57495, and AJCC stage is the control. (**D–G**) Kaplan-Meier survival curves of the four-gene signature and distribution of patient survival and risk score in different groups when patients in GSE57495 were divided into two groups. (**D**) Kaplan-Meier survival curves. (**E**) Distribution of the risk score. (**F**) Distribution of the survival time. Circle shape stands for high-risk group while triangle shape for low-risk group. Red stands for survival and green stands for dead. (**G**) Heatmap of the expression of the four genes. (**H–K**) Kaplan-Meier survival curves of the four-gene signature and distribution of patient survival and risk score in different groups when patients in GSE57495 were divided into three groups. (**H**) Kaplan-Meier survival curves. (**I**) Distribution of the risk score. (**J**) Distribution of the survival time. (**K**) Heatmap of the expression of the four genes. (**L**) Calibration plot for validation of the gene signature in GSE57495.
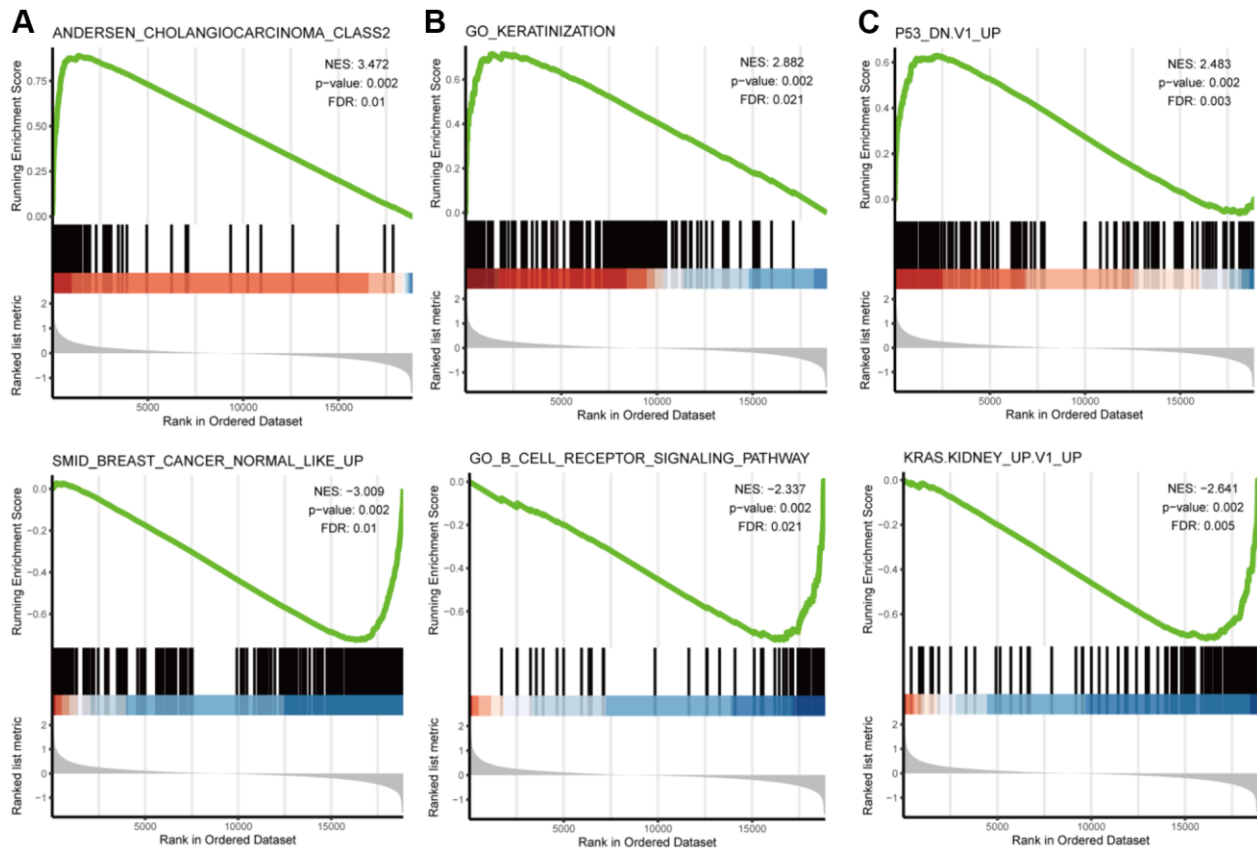
Supplementary Figure 3. GSEA for other MSigDB gene sets. Upregulated and downregulated enriched pathways with top NES were showed. (**A**) C2 (curated gene sets). (**B**) C5 (GO gene sets). (**C**) C6 (oncogenic gene sets).
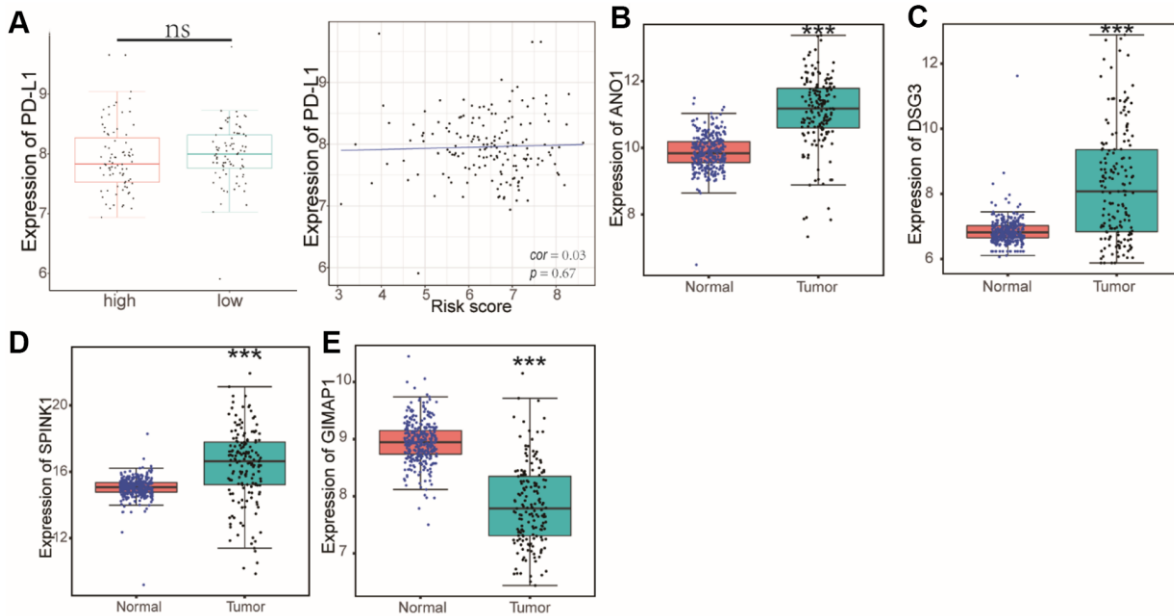
Supplementary Figure 4. Expression of PD-L1 and four genes in signature. (**A**) Correlation between expression of PD-L1 and risk score. (**B–E**) The mRNA expression of four genes in signature.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1–6.

Supplementary Table 1. Clinical characteristics of pancreatic cancer patients in the TCGA dataset.

Supplementary Table 2. Identification of the RRDEGs in pancreatic cancer.

Supplementary Table 3. Functional enrichment analysis of RRDEGs.

Supplementary Table 4. Univariate Cox regression analysis of RRDEGs in training group.

Supplementary Table 5. Mean cross-validation error (CVM) and standard deviation of the CVM (CVSD) for each lambda in the Lasso regression.

Supplementary Table 6. Gene set enrichment analysis between high and low risk groups in 165 samples from TCGA PAAD.