Construction of a five-gene prognostic model based on immune-related genes for the prediction of survival in pancreatic cancer

Running Head: A Prognostic Model for Pancreatic Cancer

Bo Liu¹,²,³, Tingting Fu⁴, Ping He¹,², Chengyou Du³, Ke Xu⁵

¹ Department of Hepatobiliary Surgery, Pidu District People’s Hospital of Chengdu, Chengdu, Sichuan, China.
² Department of Hepatobiliary Surgery, The Third Affiliated Hospital of Chengdu Medical College, Chengdu, Sichuan, China.
³ Department of Hepatobiliary Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China.
⁴ Department of Nosocomial Infection Control, The Third Affiliated Hospital of Chengdu Medical College, Chengdu, Sichuan, China.
⁵ Department of Oncology, Clinical Medical College and The First Affiliated Hospital of Chengdu Medical College, Chengdu, Sichuan, China.

* Correspondence: Ke Xu
Department of Oncology, Clinical Medical College and The First Affiliated Hospital of Chengdu Medical College, 278 Baoguang Street, Xindu District, Chengdu 610500, Sichuan, China.
Tel +8618980074098
Email: xuke@cmc.edu.cn
ORCID: 0000-0002-5520-5754

*Email: xuke@cmc.edu.cn is the Institutional email of the corresponding author
Abstract

Purpose: To identify differentially expressed immune-related genes (DEIRGs) and construct a model with survival-related DEIRGs for evaluating the prognosis of patients with pancreatic cancer (PC).

Methods: Six microarray gene expression datasets of PC from the Gene Expression Omnibus (GEO) and ImmPort were used to identify DEIRGs. RNA sequencing and clinical data from The Cancer Genome Atlas Program-Pancreatic Adenocarcinoma (TCGA-PAAD) database were used to establish the prognostic model. Univariate, least absolute shrinkage and selection operator (LASSO) and multivariate Cox regression analyses were applied to determine the final variables of the prognostic model. The median risk score was used as the cut-off value to classify samples into low- and high-risk groups. The prognostic model was further validated using an internal validation set of TCGA and an external validation set of GSE62452.

Results: In total, 142 DEIRGs were identified from six GEO datasets, 47 were survival-related DEIRGs. A prognostic model comprising five genes (i.e., ERAP2, CXCL9, AREG, DKK1, and IL20RB) was established. High-risk patients had poor survival compared with low-risk patients. The 1-, 2-, 3-year area under the receiver operating characteristic curve of the model reached 0.85, 0.87, and 0.93, respectively. Additionally, the prognostic model reflected the infiltration of neutrophils and dendritic cells. The expression of most characteristic immune checkpoints was significantly higher in the high-risk group versus the low-risk group.

Conclusions: The five-gene prognostic model showed reliably predictive accuracy. This model may provide useful information for immunotherapy and facilitate personalized monitoring for patients with PC.

Keywords: pancreatic cancer, immune-related genes, prognostic model, bioinformatics
**Introduction**

Pancreatic cancer (PC) is a highly malignant cancer and the seventh-leading cause of mortality worldwide [1]. It is predicted that PC will become the second-leading cause of cancer-related death by the year 2030 in the USA [2]. Thus far, surgical therapy is the only curative strategy for resectable PC. However, only 10% of the patients are able to undergo standard resection at diagnosis due to the presence of atypical symptoms and the lack of effective imaging examination and diagnostic biomarkers in the early stage of disease [3]. At the time of diagnosis, most patients present with unresectable disease, characterized by nodal metastases, vascular invasion, or distant metastases [4]. Nevertheless, even patients who undergo surgical resection may not achieve satisfactory survival. Therefore, early detection and development of novel therapeutic strategies are urgently warranted to improve the survival of patients with PC.

The immune system plays a pivotal role in tumorigenesis and the progression of human malignancy [5]. A growing body of evidence suggests that the use of immunotherapy could result in favorable outcomes in cancer therapy. Blockade of immune checkpoints has shown substantial survival benefit for patients with several types of cancer, such as hepatocellular carcinoma, renal cell carcinoma, and melanoma [6-8]. Tumor cells could escape recognition and elimination by the immune system, induce immune tolerance, and promote their own growth and metastasis by secretion of immunosuppressive cytokines and regulation of the expression of immunoregulatory molecules [9,10]. Previous studies highlighted that the immune-related genes (IRGs) were associated with the prognosis of several types of cancer [11-13]. However, few studies investigated the role of IRGs in PC. Hence, the identification of genes with prognostic potential and construction of an effective predictive model may be useful for individualized management and assessment of prognosis in patients with PC.

In the present study, we utilized the Robust Rank Aggregation (RRA) method to identify differentially expressed genes (DEGs) between pancreatic tumors and
adjacent normal tissues using six microarray datasets obtained from the Gene Expression Omnibus (GEO). Subsequently, univariate Cox regression was employed to identify survival-related differentially expressed immune-related genes (DEIRGs). Furthermore, least absolute shrinkage and selection operator (LASSO) Cox regression and multivariate Cox regression analyses were utilized to construct a prognostic model comprising survival-related DEIRGs. The median risk score calculated by the model was used to classify patients into high- and low-risk groups. The association between the model and immune cell infiltration was investigated. In addition, the expression of immune checkpoints in the low- and high-risk groups was compared. The aim of this study was to identify the survival-related biomarkers and therapeutic targets, establish a predictive model, and provide a basis for immunotherapy in patients with PC.

Materials and methods

Gene expression datasets

Six gene expression datasets (i.e., GSE15471, GSE60979, GSE62165, GSE71989, GSE91035, GSE102238) of PC were obtained from the GEO. All datasets met the following criteria: 1) included tumor and adjacent tissues of human PC; 2) comprised case and control groups; 3) contained >20 samples. Detailed information regarding these datasets is listed in Supplementary Table 1. A total of 170 PC sample profiles with available survival data were downloaded The Cancer Genome Atlas-Pancreatic Adenocarcinoma (TCGA-PAAD) dataset. A training dataset with 102 samples and an internal validation dataset with 68 samples were randomly generated from the TCGA-PAAD dataset in a ratio of 3:2. In addition, a microarray dataset (GSE62452) containing 64 samples with survival data was obtained from the GEO for external validation. The characteristics of the training and the validation datasets are listed in Table 1.

Identification of DEIRGs
The *Limma* package was utilized to identify DEGs between tumor tissues and adjacent normal tissues of each dataset in the R platform (v3.6.1) [14]. The *RobustRankAggreg* package, which is based on the RRA method, was employed to normalize multiple datasets and conduct gene integration analysis for the identification of the most significant DEGs [15]. Genes with $|\log_2 \text{Fold Change}| > 1$ and adjusted $P$-value $< 0.05$ were selected as significant DEGs.

**Screening for survival-related DEIRGs**

The IRG list (1,811 genes) was obtained from the Immunology Database and Analysis Portal (ImmPort) [16]. DEIRGs were obtained by intersecting the IRG list and DEG list identified from the six GEO datasets. Subsequently, we performed univariate Cox regression analysis of DEIRGs to identify survival-related DEIRGs. Genes with $P$-value $< 0.01$ were selected as survival-related DEIRGs.

**Functional enrichment analysis of DEIRGs**

Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were carried out by the *Clusterprofiler* package to investigate the potential function of DEIRGs [17]. Adjusted $P$ value less than 0.05 was selected as the cut-off criteria for GO terms and KEGG pathways.

**Construction of a prognostic model**

In the training set, LASSO regression through the *glmnet* package was utilized to determine the most powerful prognostic genes among the survival-related DEIRGs [18]. Next, multivariate Cox stepwise regression was applied to determine the best prognostic model. Subsequently, a prognostic model was established using a linear combination of the relative gene expression values ($\text{Exp}_i$) and coefficient ($\beta_i$) generated in the multivariate Cox regression. The risk score calculation formula is as follows: Risk score = $\text{Exp}_1 \times \beta_1 + \text{Exp}_2 \times \beta_2 + \ldots + \text{Exp}_n \times \beta_n$. The median risk score was used as the cut-off value to classify the PC samples into low- and high-risk
groups. Kaplan–Meier curves analysis and log-rank test were performed to identify differences in survival. Time-dependent receiver operating characteristic (ROC) curve analysis was applied to evaluate the predictive ability of the prognostic model via the timeROC package [19]. Furthermore, we performed univariate and multivariate Cox regression analyses to investigate the independent factors between the risk score and clinical parameters, including age, sex, T stage, N stage, AJCC stage, and histologic grade. In addition, we performed gene set enrichment analysis (GSEA) utilizing the "ClusterProfiler" package to investigate the significantly enriched pathways between high- and low-risk groups in TCGA-PAAD dataset. We retrieved KEGG gene sets (c2.cp.kegg.v7.0.symbols.gmt) by using the 'msigdb' package.

**Correlation analysis of risk score and immune cells infiltration**

Tumor Immune Estimation Resource (TIMER, cistrome.shinyapps.io/timer) is a comprehensive analytical web tool, which includes 10,897 samples across 32 cancer types from TCGA to estimate the abundance of six tumor-infiltrating immune cells (TIIC) subsets (B cells, CD4 T cells, CD8 T cells, macrophages, neutrophils, and dendritic cells) [20]. Immune cells infiltration levels of PC patients which obtained from TIMER were applied for exploring the correlation between risk score and immune cells infiltration.

**Comparison of relative expression of immune checkpoints in low- and high-risk group**

Immunotherapy has been achieved promising results in the treatment of many cancers in recent years. The most efficient strategy focused on the blockade of the immune checkpoints [21]. We compared the expression of most characteristic immune checkpoints, including programmed cell death 1 (PDCD1), Programmed Cell Death 1 Ligand 1 (PDCD1-L1), cytotoxic T-Lymphocyte associated protein 4 (CTLA4), CD80, CD86, V-domain immunoglobulin suppressor of T cell activation (VISTA), T-cell immunoglobulin mucin receptor 3 (TIM3), and T cell Ig and ITIM domain (TIGIT),...
between low- and high-risk group aiming to provide information to optimize immunotherapeutic strategies for patients with PC.

**Statistical analysis**

R software v3.6.1 (www.r-project.org) was used for statistical analyses in the present study. The Limma package was used to obtain DEGs. Kaplan–Meier curves analysis and log-rank test were performed using the survival package. Time-dependent ROC curve analysis was applied via the timeROC package. Univariate and multivariate Cox regression were employed to determine independent factors for OS. Spearman correlation analysis was applied to investigate the correlation between risk score and immune cells infiltration. Difference of expression of immune checkpoints in Low- and high-risk Group were compared using Wilcox test. Analysis results with P value <0.05 indicated statistical significance.

**Result**

**Identification of DEGs**

A total of 985 DEGs, including 619 upregulated genes and 366 downregulated genes were identified. The top 20 upregulated and downregulated DEGs are shown in Figure 1A. We obtained a total of 142 DEIRGs by intersecting the DEG list with the IRG list (Figure 1B).

**GO and KEGG enrichment analysis for DEIRGs**

The following GO categories were enriched: biological process (BP), cellular component (CC), and molecular function (MF) (Figures 2A-C). The results showed that the significantly enriched terms were defense response to other organism (BP), extracellular matrix (CC), and receptor ligand activity (MF). Furthermore, according to the KEGG pathway analysis, the cytokine-cytokine receptor interaction pathway was markedly enriched (Figure 2D); the top 10 pathways of DEIRGs are listed in Supplementary Table 2.
Construction and internal validation of prognostic Model

We obtained 47 survival-related DEIRGs via univariate Cox regression analysis (Supplementary Table 3). LASSO Cox regression was applied to narrow down the number of relevant genes in the training dataset. The LASSO coefficient profiles of 47 survival-related DEIRGs are presented in Figure 3A. We obtained 12 genes with minimum partial likelihood deviance according to 10-fold cross-validation results (Figure 3B). Furthermore, the best prognostic model with the smallest Akaike Information Criterion was identified via multivariate Cox stepwise regression analysis.

Finally, a prognostic model involving five genes, namely endoplasmic reticulum aminopeptidases 2 (ERAP2), amphiregulin (AREG), C-X-C motif chemokine ligand 9 (CXCL9), dickkopf-1 (DKK1) and interleukin-20 receptor subunit beta (IL-20RB) was constructed. Figure 3C shows that CXCL9, DKK1, and IL-20RB exhibit the characteristics of independent prognostic factors in the training dataset. The prognostic risk score for each patient was calculated as follows: Risk score = (expression level of ERAP2 × 0.158) + (expression level of CXCL9 × 0.357) + (expression level of AREG × 0.195) + (expression level of DKK1 × 0.172) + (expression level of IL20RB × 0.231).

The patients in the training dataset were divided into low- and high-risk groups applying the median risk score as the cut-off criteria. The Kaplan-Meier curve shows that high-risk patients had a significant worse overall survival (OS) than low-risk group patients (Figure 4A). A time-dependent ROC curve was generated, and the area under the ROC curve (AUC) was calculated to assess the predictive ability of the model. In the training dataset, the 1-, 2-, and 3-year AUCs were 0.85, 0.87, and 0.93 (Figure 4B), respectively. The risk score distribution and the expression of the five genes in the training dataset are shown in Figures 4C and 4D. The prognostic model was further validated using the internal validation dataset. Similarly, patients with higher risk scores were associated with worse OS (Figure 4E). In the internal validation dataset, the 1-, 2-, and 3-year AUCs were 0.79, 0.74, and 0.8, respectively.
(Figure 4F). Figures 4G and 4H demonstrate the distribution of the risk score and a heatmap of the five gene expression data in the validation dataset, respectively. This indicates that this prognostic model is able to predict the OS of patients with PC in TCGA cohort.

**External validation of the prognostic model in GEO dataset**

A GEO dataset (GSE62452) of PC with survival data was used as an external validation dataset to assess the predictive capability of the prognostic model. The risk score of each patient in the dataset was calculated using the formula of the model, and all patients were divided into high- and low-risk groups according to the median risk score. The OS observed in the high-risk group patients was significantly worse than that recorded in the low-risk group. In the external validation dataset, the 1-, 2-, and 3-year AUCs were 0.6, 0.75, and 0.77, respectively. External validation further confirmed the stable and accurate prognostic value of the present model in PC (Figures 4I-L).

**Evaluation of the independence of the prognostic model**

To investigate the independent predictive ability of the prognostic model, we performed univariate and multivariate Cox regression analyses for the relationship between risk score and clinicopathological characteristics. The results showed that age, N stage, histologic grade, and risk score were associated with worse prognosis (Figure 5). Meanwhile, age and risk score were independent prognostic factors for OS. Moreover, higher risk scores were associated with advanced grade of disease in TCGA-PAAD dataset (Supplementary Figure 1F).

**Correlation between the risk score and immune cell infiltration**

We investigated the correlation between the risk score and the abundance of six tumor infiltrating immune cell subsets (i.e., B cells, CD4 T cells, CD8 T cells, macrophages, neutrophils, and dendritic cells). The results shown that the risk score was positively
correlated with the infiltration of neutrophils (R=0.34, P=0.0013) and dendritic cells
(R=0.42, P<0.001) (Figure 6).

**Differences between the expression of immune checkpoints in the low- and high-risk groups**

We compared the relative expression of most characteristic immune checkpoints between the low- and high-risk groups. The expression of PDCD1, PDCD-L1, CTLA4, CD80, CD86, TIM3, VISTA, and TIGIT was significantly higher in the high-risk group compared with the low-risk group, indicating that immunosuppression may contribute to worse OS in high-risk patients (Figure 7).

**GSEA analysis result**

We performed GSEA to further investigate the different functional phenotype between the high- and low-risk groups. The results are listed in Supplementary Table 4. The four most significantly enriched pathways were: pathways in cancer, cytokine-cytokine receptor interaction, regulation of actin cytoskeleton, and chemokine signaling pathway (Figure 8).

**Discussion**

PC remains a disease with dismal prognosis, namely poor surviva and unfavorable therapeutic efficacy. Although numberous studies have investigated the relationship between IRGs and tumor prognosis, only a few focused on PC. In the present study, we identified DEIRGs with prognostic value in PC and established a prognostic model with five IRGs (i.e., ERAP2, CXCL9, AREG, DKK1 and IL20RB). Throughout the course of the study, 142 DEIRGs were identified by integrated analysis of six GEO datasets and intersected with an IRG list. The 47 survival-related DEIRGs were identified by univariate Cox regression. The five-IRG prognostic model was established by LASSO Cox regression and multivariate stepwise Cox regression. Notably, in the validation of the prognostic model, the 1-, 2-, and 3-year
AUCs were 0.79, 0.74, and 0.8 (internal validation) and 0.59, 0.75, and 0.77 (external validation), respectively. The results indicated the reproducibility and generalizability of the prognostic model. Collectively, these results showed that this five-gene prognostic model was an effective predictive tool for OS.

Regarding the prognostic markers for PC, carbohydrate antigen 19-9 (CA19-9) is currently the most widely used marker in clinical practice. CA19-9 has shown a certain prognostic capability in both postoperative and advanced PC [22,23]. However, CA19-9 is not applicable to Lewis antigen-negative individuals, and false positive results may be obtained in cases with biliary infection and other malignant tumors [24,25]. Thus, there is a need to discover accurate prognostic markers for PC. In terms of immune-related molecules, a recent study reported that the expression of MET, 2'-5'-oligoadenylate synthetase 1 (OAS1) and 2'-5'-oligoadenylate synthetase like (OASL) was closely related to the progression of PC. Their expression was up-regulated in PC tissues and was associated with poor prognosis [26]. Similarly, from the perspective of RNA-binding proteins, a close relationship between OAS1 and the prognosis of PC was also found [27]. Guanylate binding protein 2 (GBP2) is a guanylate-binding protein involved in response to viral or microbial infection; it is induced by Type I and Type II interferons. Recently, researchers found that GBP2 was highly expressed in PC and positively correlated with the expression of immune checkpoints (e.g., PDCD1, PDCD1L1, CTLA4, etc.) Patients with PC and high expression of GBP2 were linked to a poor prognosis and an AUC of 0.69 for 3-year survival [28]. In addition, the effect of BRCA1/2 mutation on the prognosis of PC has been a focus of research in this field. Studies found that the BRCA1 rs1799966 polymorphism was correlated with the prognosis of PC. It was demonstrated that patients with advanced PC with BRCA2 mutation have a better response to platinum and a better prognosis [29-31]. Some non-coding RNAs and circulating tumor DNA have also been linked to the prognosis of PC; nevertheless, the usefulness of these markers is compromised by their limited predictive capability [32-35]. Therefore, the
combination of these markers and construction of prognostic models may enhance the predictive capability.

Previously, researchers constructed some predictive models for PC based on prognosis-related genes. Representative studies were performed by Wu et al and Yan et al. The former research group established a nine-gene signature to predict the OS of PC [36]. The latter research group constructed a four-gene prognostic model based on transcription factors and kinases associated with dysregulation genes [37]. However, these two prognostic models did not include immune-related DEG and did not investigate the association between the expression of immune checkpoints and risk score. In addition, the present model exhibited better capability in predicting the 1-, 2-, and 3-year survival of patients with PC.

Among the five genes of the model, ERAP2 is an aminopeptidase which is present in the lumen of the endoplasmic reticulum. It trims and generates peptide ligands for antigen presentation by major histocompatibility class I molecules [38,39]. Previous studies suggested that ERAP2 plays a pivotal role in vessel regeneration by inducing the migration and proliferation of endothelial cells [40,41]. ERAP2 could accelerate anti-tumor immune responses; hence, modulating the activity of ERAP2 may be a novel immunological strategy for cancer immunotherapy [42]. Chemokines are a family of small cytokines inducing directed chemotaxis, which can be found in most types of human cancer [43,44]. A member of the chemokine family, namely CXCL9, recruits leukocytes to sites of inflammation and plays a critical role in tumor progression. Previous studies reported that CXCL9 was correlated with worse OS in renal cell carcinoma, promoted tumor metastasis in melanoma and enhanced the invasive ability of hepatocellular carcinoma [45-47]. Gao et al demonstrated that CXCL9 was overexpressed in PC; this finding was consistent with the results of the present study [48]. They also found that CXCL9 could promote tumor progression in an orthotropic murine PAAD model by regulating the CD8+ T lymphocytes in the
tumor microenvironment (TME). AREG is a ligand of epidermal growth factor receptor (EGFR), which is aberrantly expressed and plays a vital role in numerous types of cancer by mediating the motility, metastasis, and proliferation of cancer cells [49,50]. Stimulation of AREG increased the invasiveness, metastasis, and epithelial-mesenchymal transition of PC cells in vivo [51,52]. DKK1, a member of the DKK family, participates in the WNT/β-catenin pathway [53]. High expression of DKK1 was associated with aggressive features and shorter OS in patients with PC [54]. Previous studies demonstrated that DKK1 was correlated with accumulation of myeloid-derived suppressor cells in PC, contributing to the suppression of the responses of anti-tumor T cells [55]. DKK1 has been utilized as a potential target for immunotherapy in patients with myeloma [56]. IL-20RB, a receptor of the IL20 subfamily, is involved in both amplified inflammatory and anti-inflammatory responses. Dysregulated expression of IL20RB has been observed in various studies, including the present study [57-59]. Overexpression of IL20RB was correlated with poor outcome in patients with papillary renal cell carcinoma, the ability of papillary renal cell carcinoma cells to invade and metastasize could be inhibited by silencing IL20RB in vivo [60].

Previous studies have shown that the TME of PC was infiltrated by immunosuppressive cells, but not effector lymphocytes [61,62]. Moreover, PC was characterized by a low proportion of tumor/stroma ratio in the tumor mass [63]. Notably, the stromal area of the TME was the main site of immune cell infiltration, which contributes to the poor outcome of PC [64]. In the present study, the correlation between the risk score and six subtypes of tumor-infiltrating immune cells was investigated. We observed that our prognostic model was positively associated with the infiltration of dendritic and neutrophil cells. According to the considerable research on immune checkpoints conducted in recent decades, immunotherapy has shown great curative potential for several types of cancer, such as hepatocellular carcinoma, melanoma, and bladder cancer [65-67]. Binding of PDCD1LG1 to its
corresponding ligand PDCD1 negatively regulates the activity of immune cells and 
induces the immune evasion of tumor cells [68]. Previous studies indicated that
PDCD1LG1 was overexpressed in patients with PC, and its down-regulation could
inhibit the proliferation of pancreatic tumor cells [69]. CTLA4 is expressed by
regulatory T cells, which are highly enriched in PC. The binding of CTLA4 to its
ligands CD80 and CD86 leads to tumor cell immunosuppression. PC tumors are
poorly immunogenic; hence, it is important to discover novel immune checkpoints for
immunotherapy and develop more sophisticated treatment strategies. TIM3, belongs
to the immunoglobulin superfamily and plays a dual role in regulating the immune
response. It has been proved to be correlated with worse outcome in several types of
cancer [70,71]. VISTA is a novel immune checkpoint overexpressed on CD68+
macrophages in PAAD. It is a potential immunotherapeutic target based on its high
infiltration of the tumor environment and inhibition of T-cell activation [72]. TIGIT is
expressed in several types of tumor cells and regulatory T cells; it is involved in
immunosuppression and the immune evasion of cancer cells [73,74]. We compared
the expression of immune checkpoints between patients in the low- and high-risk
groups generated from our prognostic risk score model in TCGA-PAAD cohort.
Interestingly, the expression levels of PDCD1, PDCD1-L1, CTLA4, CD80, TIM3,
VISTA, and TIGIT in the high-risk group were notably higher than those measured in
the low-risk group. These results indicated that the immunotherapeutic strategy of
immune checkpoint blockade may be more effective for high-risk patients. GSEA
revealed that the significantly enriched pathways in the high-risk group were
associated with immune-related responses and tumorigenesis.

However, limitations in the present study should be realized. As our study was
driven from statistics analysis of retrospective data, multicenter clinical trials and
prospective research are required to further assess and validate this prognostic model.
Meanwhile, additional experiment should be conducted to evaluate the expression of
survival-related IRGs at protein level. Moreover, the biological function and mechanism of the IRGs in the prognostic model worth to be further elucidated in the future.

**Conclusion**

In this study, survival-related DEIRGs were identified, and a five-gene prognostic model with reliable predictive accuracy was constructed. The risk score calculated from the model is strongly correlated with immune cell infiltration in tumors and the expression of immune checkpoints. This model may provide new insight into the individualization of anti-tumor therapy and facilitate clinical monitoring of PC.

**Data Availability**

The datasets generated and analyzed during the current study are available in TCGA official website repository ([https://portal.gdc.cancer.gov/](https://portal.gdc.cancer.gov/)) and Gene Expression Omnibus (GEO) ([https://www.ncbi.nlm.nih.gov/geo/](https://www.ncbi.nlm.nih.gov/geo/)).

**Competing interests**

The authors declare no potential conflicts of interest.

**Funding**

This study was funded by The Special Scientific Research Foundation of The First Affiliated Hospital of Chengdu Medical College (CYFY2017GLPHX06), The Project of Chengdu Medical Research (2020118) and Scientific Research Program of Sichuan Provincial Health Commission (18J568).

**Author contributions**

Ke Xu designed the study. Bo Liu, Ting-ting Fu and Ke Xu analyzed the data. Bo Liu, Chengyou Du and Ping He prepared the figures and tables. Bo Liu wrote the manuscript.

**Ethics approval and consent to participate**

The ethical approval did not refer to this study and this study did not need the informed consent.

**Consent for publication**

Not applicable

**Code availability**

Not applicable

**Authors' information**

The corresponding author (Ke Xu) would be contacted if necessary.

**Acknowledgements**

We thank all colleagues in the Department of Hepatobiliary Surgery for helping with this study.

**Abbreviations**

PC, pancreatic cancer; PAAD, pancreatic adenocarcinoma; IRGs, immune-related genes; OS,
overall survival; DEGs, differentially expressed genes; DEIRGs, differentially expressed immune-related genes; RRA, Robust Rank Aggregation; TIIC, tumor-infiltrating immune cells; AUC, area under the curve; LASSO, Least Absolute Shrinkage and Selection Operator; PDCD1, programmed cell death 1; PDCD1-L1, Programmed Cell Death 1 Ligand 1; CTLA4, cytotoxic T-Lymphocyte associated protein 4; VISTA, V-domain immunoglobulin suppressor of T cell activation; TIM3, T-cell immunoglobulin mucin receptor 3; TIGIT, T cell Ig and ITIM domain; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; ROC, receiver operating characteristic; ERAP2, Endoplasmic reticulum aminopeptidases 2; AREG, Amphiregulin; CXCL9, C-X-C Motif Chemokine Ligand 9; DKK1, Dickkopf-1; TME, tumor microenvironment.

References

1. Bray F, Ferlay J, Soerjomataram I et al (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68 (6):394-424
2. Rahib L, Smith BD, Aizenberg R et al (2014) Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res 74 (11):2913-2921
3. Strobel O, Neoptolemos J, Jager D et al (2019) Optimizing the outcomes of pancreatic cancer surgery. Nat Rev Clin Oncol 16 (1):11-26
4. Kamisawa T, Wood LD, Itoi T et al (2016) Pancreatic cancer. Lancet 388 (10039):73-85
5. Gentles AJ, Newman AM, Liu CL et al (2015) The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med 21 (8):938-945
6. El-Khoueiry AB, Sangro B, Yau T et al (2017) Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. Lancet 389 (10088):2492-2502
7. Tomita Y, Fukasawa S, Shinohara N et al (2019) Nivolumab versus everolimus in advanced renal cell carcinoma: Japanese subgroup 3-year follow-up analysis from the Phase III CheckMate 025 study. Jpn J Clin Oncol 49 (6):506-514
8. Weber JS, D’Angelo SP, Minor D et al (2015) Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial. Lancet Oncol 16 (4):375-384
9. Martinez-Bosch N, Vinaixa J, Navarro P (2018) Immune Evasion in Pancreatic Cancer: From Mechanisms to Therapy. Cancers (Basel) 10 (1)
10. Gromisch C, Qadan M, Machado MA et al (2020) Pancreatic Adenocarcinoma: Unconventional Approaches for an Unconventional Disease. Cancer Res 80 (16):3179-3192
11. Li B, Cui Y, Diehn M et al (2017) Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-Small Cell Lung Cancer. JAMA Oncol 3 (11):1529-1537
12. Lin P, Guo YN, Shi L et al (2019) Development of a prognostic index based on an
immunogenomic landscape analysis of papillary thyroid cancer. Aging (Albany NY) 11 (2):480-500

13. Wang WJ, Wang H, Hua TY et al (2020) Establishment of a Prognostic Model Using Immune-Related Genes in Patients With Hepatocellular Carcinoma. Front Genet 11:55

14. Ritchie ME, Phipson B, Wu D et al (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43 (7):e47

15. Kolde R, Laur S, Adler P et al (2012) Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics 28 (4):573-580

16. Bhattacharya S, Dunn P, Thomas CG et al (2018) ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. Sci Data 5:180015

17. Yu G, Wang LG, Han Y et al (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 16 (5):284-287

18. Friedman J, Hastie T, Tibshirani R (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 33 (1):1-22

19. Blanche P, Dartigues JF, Jacqmin-Gadda H (2013) Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Stat Med 32 (30):5381-5397

20. Li T, Fan J, Wang B et al (2017) TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer Res 77 (21):e108-e110

21. Salmaninejad A, Valilou SF, Shabgah AG et al (2019) PD-1/PD-1 pathway: Basic biology and role in cancer immunotherapy. J Cell Physiol 234 (10):16824-16837

22. Berger AC, Garcia M, Jr., Hoffman JP et al (2008) Postresection CA 19-9 predicts overall survival in patients with pancreatic cancer treated with adjuvant chemoradiation: a prospective validation by RTOG 9704. J Clin Oncol 26 (36):5918-5922

23. Hess V, Glimelius B, Grøwe P et al (2008) CA 19-9 tumour-marker response to chemotherapy in patients with advanced pancreatic cancer enrolled in a randomised controlled trial. Lancet Oncol 9 (2):132-138

24. Tempero MA, Uchida E, Takasaki H et al (1987) Relationship of carbohydrate antigen 19-9 and Lewis antigens in pancreatic cancer. Cancer Res 47 (20):5501-5503

25. Marrelli D, Caruso S, Pedrazzani C et al (2009) CA19-9 serum levels in obstructive jaundice: clinical value in benign and malignant conditions. Am J Surg 198 (3):333-339

26. Zhang C, Zou Y, Zhu Y et al (2021) Three Immune-Related Prognostic mRNAs as Therapeutic Targets for Pancreatic Cancer. Front Med (Lausanne) 8:649326

27. Wen X, Shao Z, Chen S et al (2020) Construction of an RNA-Binding Protein-Related Prognostic Model for Pancreatic Adenocarcinoma Based on TCGA and GTEx Databases. Front Genet 11:610350

28. Liu B, Huang R, Fu T et al (2021) GBP2 as a potential prognostic biomarker in pancreatic adenocarcinoma. PeerJ 9:e11423

29. Zhu Y, Zhai K, Ke J et al (2017) BRCA1 missense polymorphisms are associated with poor prognosis of pancreatic cancer patients in a Chinese population. Oncotarget 8 (22):36033-36039

30. Golan T, Kanji ZS, Epstein R et al (2014) Overall survival and clinical characteristics of pancreatic cancer in BRCA mutation carriers. Br J Cancer 111 (6):1132-1138

31. Waddell N, Pajic M, Patch AM et al (2015) Whole genomes redefine the mutational landscape of pancreatic cancer. Nature 518 (7540):495-501
52. Wang L, Wu H, Wang L et al (2016) Expression of amphiregulin predicts poor outcome in patients with pancreatic ductal adenocarcinoma. Diagn Pathol 11 (1):60
53. Krupnik VE, Sharp JD, Jiang C et al (1999) Functional and structural diversity of the human Dickkopf gene family. Gene 238 (2):301-313
54. Liu DJ, Xie YX, Liu XX et al (2017) The role of Dickkopf-1 as a potential prognostic marker in pancreatic ductal adenocarcinoma. Cell Cycle 16 (17):1622-1629
55. D’Amico L, Mahajan S, Capietto AH et al (2016) Dickkopf-related protein 1 (Dkk1) regulates the accumulation and function of myeloid derived suppressor cells in cancer. J Exp Med 213 (5):827-840
56. Qian J, Yi Q (2012) DKK1 as a novel target for myeloma immunotherapy. Oncoimmunology 1 (5):756-758
57. Omarini C, Bettelli S, Caprera C et al (2018) Clinical and molecular predictors of long-term response in HER2 positive metastatic breast cancer patients. Cancer Biol Ther 19 (10):879-886
58. Haider S, Wang J, Nagano A et al (2014) A multi-gene signature predicts outcome in patients with pancreatic ductal adenocarcinoma. Genome Med 6 (12):105
59. Baird AM, Gray SG, O’Byrne KJ (2011) IL-20 is epigenetically regulated in NSCLC and down regulates the expression of VEGF. Eur J Cancer 47 (12):1908-1918
60. Cui XF, Cui XG, Leng N (2019) Overexpression of interleukin-20 receptor subunit beta (IL2RB) correlates with cell proliferation, invasion and migration enhancement and poor prognosis in papillary renal cell carcinoma. J Toxicol Pathol 32 (4):245-251
61. Vonderheide RH, Bayne LJ (2013) Inflammatory networks and immune surveillance of pancreatic carcinoma. Curr Opin Immunol 25 (2):200-205
62. Tsujikawa T, Kumar S, Borkar RN et al (2017) Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. Cell Rep 19 (1):203-217
63. Nakamura T, Furukawa Y, Nakagawa H et al (2004) Genome-wide cDNA microarray analysis of gene expression profiles in pancreatic cancers using populations of tumor cells and normal ductal epithelial cells selected for purity by laser microdissection. Oncogene 23 (13):2385-2400
64. Blando J, Sharma A, Higa MG et al (2019) Comparison of immune infiltrates in melanoma and pancreatic cancer highlights VISTA as a potential target in pancreatic cancer. Proc Natl Acad Sci U S A 116 (5):1692-1697
65. Zhu AX, Finn RS, Edeline J et al (2018) Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase 2 trial. Lancet Oncol 19 (7):940-952
66. Romano E, Kusio-Kobialka M, Foukas PG et al (2015) Ipilimumab-dependent cell-mediated cytotoxicity of regulatory T cells ex vivo by nonclassical monocytes in melanoma patients. Proc Natl Acad Sci U S A 112 (19):6140-6145
67. Powles T, Eder JP, Fine GD et al (2014) MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. Nature 515 (7528):558-562
68. Jiang X, Wang J, Deng X et al (2019) Role of the tumor microenvironment in PD-L1/PD-1-mediated tumor immune escape. Mol Cancer 18 (1):10
69. Song X, Liu J, Lu Y et al (2014) Overexpression of B7-H1 correlates with malignant cell proliferation in pancreatic cancer. Oncol Rep 31 (3):1191-1198
70. Nakano M, Ito M, Tanaka R et al (2018) PD-1+ TIM-3+ T cells in malignant ascites predict
Figure Legends

Figure 1. Identification of DEIRGs in GEO datasets. (A) A heatmap of the top 20 up- and down-regulated DEGs in the integrated analysis. Red and blue represent up-regulated and down-regulated genes in each dataset, respectively. The numbers in each rectangle indicate the value of log2FC in each dataset calculated by ‘Limma’ package. The color gradient from green to red represents the log2FC from small to large. (B) Venn diagrams of the DEIRGs between DEGs in the integrated analysis and IRG list obtained from ImmPort. DEGs, differentially expressed genes; DEIRGs, differentially expressed immune-related genes. GEO, Gene Expression Omnibus. IMMPORT, Immunology Database and Analysis Portal.

Figure 2. GO and KEGG analysis of DEIRGs. (A–C) The top 10 terms significantly enriched in the GO categories of biological process (BP), cellular component (CC) and molecular function (MF). (D) KEGG pathway analysis of DEIRGs. The bubble size is proportional to the number of DEIRGs involved. The color gradient from blue to red represents the p-value from small to large, respectively. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes. DEIRGs, differentially expressed immune-related genes.

Figure 3. LASSO profiles and multivariate Cox regression analysis. (A) LASSO coefficient profiles of 47 survival-related DEIRGs. (B) 10-fold cross-validation result.
of the 47 survival-related DEGs. (C) Multivariate Cox regression analysis of the association between the five genes and overall survival in training set from TCGA. DEGs, differentially expressed genes; DEIRGs, differentially expressed immune-related genes. TCGA, The Cancer Genome Atlas.

**Figure 4. Establishment and validation of prognosis model.** Kaplan-Meier survival curves for low- and high-risk group patient in TCGA training set (A), TCGA internal validation set (E) and external validation set (I). Time-dependent ROC curve analysis of the risk score model for predicting 1-, 2-, 3-year OS in TCGA training set (B), TCGA internal validation set (F) and external validation set (J). Distribution of the risk score (C, G, and K). The expression of five survival-related IRGs in TCGA training set (D), TCGA internal validation set (H) and external validation set (L). TCGA, The Cancer Genome Atlas. OS, overall survival. IRGs, immune-related genes.

**Figure 5. Univariate and multivariate Cox regression analysis for prognostic model and clinicopathological characteristics.**

**Figure 6. Association of risk score and immune cells infiltration.** (A) B cells, (B) CD4 T cells, (C) CD8 T cells, (D) Dendritic cells, (E) Macrophages, and (F) Neutrophils.

**Figure 7. Different expression of immune checkpoints between low- and high-risk group.** (A) PDCD1, (B) PDCD-L1, (C) CTLA4, (D) CD80, (E) CD86, (F) TIM3, (G) VISTA, (H) TIGIT.

**Figure 8. The top 4 significantly enriched pathways of this prognostic model.** (A) pathways in cancer, (B) cytokine-cytokine receptor interaction, (C) regulation of actin cytoskeleton, (D) chemokine signaling pathway.
Bioscience Reports. This is an Accepted Manuscript. You are encouraged to use the Version of Record, when published, which will replace this version. The most up-to-date version is available at https://doi.org/10.1042/BSR20204301

A

Coefficients

Log Lambda

B

Partial Likelihood Deviance

Log(λ)

C

Hazard ratio

ERAP2  
(N=102)  
(0.97 - 1.4)  
0.103

CXCL9  
(N=102)  
(1.20 - 1.7)  
<0.001

AREG  
(N=102)  
(0.99 - 1.5)  
0.068

DKK1  
(N=102)  
(1.03 - 1.4)  
0.022 *

IL20RB  
(N=102)  
(1.08 - 1.5)  
0.003 **

Global p-value (Log-Rank): 1.0004e-09
AIC: 414.24; Concordance Index: 0.76
Table 1. The clinicopathological characteristics of patients in training set and internal validation set of TCGA and an external GEO validation set.

| Characteristics          | TCGA training set | TCGA validation set | GSE62452 validation set |
|--------------------------|-------------------|---------------------|-------------------------|
|                          | N=102             | N=68                | N=64                    |
| Age at diagnosis (year)  | 64.23 ± 10.68     | 64.79 ± 11.11       | NA                      |
| Gender (%)               |                   |                     |                         |
| Female                   | 53 (52.0)         | 25 (36.8)           |                         |
| Male                     | 49 (48.0)         | 43 (63.2)           |                         |
| Tumor stage (%)          |                   |                     |                         |
| T1+T2                    | 17 (16.7)         | 10 (14.7)           |                         |
| T3+T4                    | 83 (81.4)         | 58 (85.3)           |                         |
| Not report               | 2 (2.0)           | 0 (0.0)             |                         |
| Node stage (%)           |                   |                     |                         |
| N0                       | 30 (29.4)         | 17 (25.0)           |                         |
| N1                       | 70 (68.6)         | 48 (70.6)           |                         |
| Not report               | 2 (2.0)           | 3 (4.4)             |                         |
| Pathologic stage (%)     |                   |                     |                         |
| I+II                     | 95 (93.1)         | 65 (95.6)           | 47 (73.4)               |
| III+IV                   | 5 (4.9)           | 3 (4.4)             | 16 (25.0)               |
| Not report               | 2 (2.0)           | 0 (0.0)             | 1 (1.6)                 |
| Histologic grade (%)     |                   |                     |                         |
| G1+G2                    | 75 (73.5)         | 44 (64.7)           | 33 (51.6)               |
| G3+G4                    | 25 (24.5)         | 24 (35.3)           | 30 (46.9)               |
| Not report               | 2 (2.0)           | 0 (0.0)             | 1 (1.6)                 |
| Events (%)               |                   |                     |                         |
| Alive                    | 44 (43.1)         | 36 (52.9)           | 16 (25.0)               |
| Dead                     | 58 (56.9)         | 32 (47.1)           | 48 (75.0)               |
| Overall survival time     | 14.12 (9.21-22.02)| 15.86 (11.48-22.84)| 15.45 (9.20-27.63)      |

Abbreviation: NA, Not available.
Supplemental Figure 1. Relationship between the risk score and clinicopathological characteristics: (A) age, (B) gender, (C) tumor stage, (D) node stage, (E) pathologic stage, and (F) histologic grade.
**Supplemental Table 1. Characteristics of the datasets in this study.**

| Dataset        | Normal | Tumor | Platform |
|----------------|--------|-------|----------|
| GSE15471       | 39     | 39    | GPL570   |
| GSE60979       | 12     | 49    | GPL14550 |
| GSE62165       | 13     | 118   | GPL13667 |
| GSE71989       | 8      | 12    | GPL570   |
| GSE91035       | 8      | 25    | GPL22763 |
| GSE102238      | 50     | 50    | GPL19072 |

**Supplemental table 2. The result of KEGG pathway analysis of DEIRGs (ranked by P value).**

| ID     | Description                          | Adjusted P value | Gene count | Gene symbol |
|--------|--------------------------------------|------------------|------------|-------------|
| has04060 | Cytokine-cytokine receptor interaction | 1.05E-13         | 27         | CXCL14, CXCL9, CXCL5, CXCL12, CXCL3, CCL20, IL7R, TNFSF4, CCL18, CXCR4, EPO, GDF10, IL1RN, IL7, INHBA, LIF, TNFRSF11B, TNFSF13B, AMHR2, BMPR2, IL20RB, IL22RA1, IL4R, LIFR, TGFBR1, TNFRSF21, FAS |
| has05330 | Allograft rejection                    | 9.92E-08         | 9          | HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, HLA-F, FAS |
| has05332 | Graft-versus-host disease              | 1.37E-07         | 9          | HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, HLA-F, FAS |
| has04940 | Type I diabetes mellitus              | 1.62E-07         | 9          | HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, HLA-F, FAS |
| has05416 | Viral myocarditis                     | 1.78E-07         | 10         | HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, HLA-F, ICAM1, RAC2 |
| has05320 | Autoimmune thyroid disease            | 7.61E-07         | 9          | HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, HLA-F, FAS, CD74, CTSB, HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1 |
| has04612 | Antigen processing and presentation   | 1.74E-06         | 10         | HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, HLA-F, ICAM1, CXCL5, CXCL12, CXCL3, CCL20, TNFSF13B |
| has05323 | Rheumatoid arthritis                  | 7.52E-06         | 10         | HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, HLA-F, ICAM1, |
Viral protein interaction
hsa04061  with cytokine and cytokine receptor
1.19E-05 10

| Gene symbol | HR  | CI90      | P value |
|-------------|-----|-----------|---------|
| GBP2        | 2.01| 1.43-2.84 | <0.001 |
| PRDX1       | 1.76| 1.2-2.58  | 0.004  |
| TNFRSF21    | 1.64| 1.2-2.24  | 0.002  |
| FAS         | 1.54| 1.15-2.07 | 0.004  |
| S100A10     | 1.53| 1.21-1.94 | <0.001 |
| S100A16     | 1.53| 1.22-1.93 | <0.001 |
| TMSB10      | 1.49| 1.14-1.95 | 0.004  |
| OAS1        | 1.45| 1.19-1.76 | <0.001 |
| PLAU        | 1.42| 1.21-1.68 | <0.001 |
| SEMA3C      | 1.42| 1.19-1.68 | <0.001 |
| PLAUR       | 1.4 | 1.14-1.73 | 0.002  |
| NRP2        | 1.4 | 1.12-1.76 | 0.004  |
| MICB        | 1.39| 1.14-1.69 | 0.001  |
| EDNRA       | 1.38| 1.12-1.69 | 0.002  |
| WNT5A       | 1.36| 1.14-1.62 | 0.001  |
| IL1RN       | 1.36| 1.17-1.58 | <0.001 |
| OASL        | 1.34| 1.16-1.54 | <0.001 |
| LTBP1       | 1.33| 1.09-1.64 | 0.006  |
| IFITM1      | 1.33| 1.1-1.6   | 0.003  |
| BST2        | 1.31| 1.08-1.58 | 0.005  |
| RSAD2       | 1.31| 1.09-1.57 | 0.004  |
| MX2         | 1.3 | 1.07-1.58 | 0.008  |
| PPARG       | 1.3 | 1.11-1.51 | 0.001  |
| AREG        | 1.3 | 1.14-1.48 | <0.001 |
| S100A6      | 1.29| 1.08-1.55 | 0.005  |
| IL20RB      | 1.28| 1.15-1.42 | <0.001 |
| SDC1        | 1.28| 1.08-1.52 | 0.005  |
| NOD2        | 1.27| 1.06-1.51 | 0.008  |
| INHBA       | 1.27| 1.09-1.47 | 0.002  |

Abbreviation: DEIRGs, differentially expressed immune-related genes. KEGG, Kyoto Encyclopedia of Genes and Genomes

Supplemental table 3. The result of Univariate Cox regression analysis for DEIRGs in TCGA PAAD dataset (P < 0.01)

PSMD6, ISG15, IRF7, OAS1, FAS  
CXCL14, CXCL9, CXCL5, CXCL12, CXCL3, CCL20, CCL18, CXCR4, IL20RB, IL22RA1
| Gene   | Fold Change | 95% CI       | p-value |
|--------|-------------|--------------|---------|
| NOX4   | 1.26        | 1.06-1.49    | 0.009   |
| DKK1   | 1.25        | 1.13-1.37    | P < 0.001 |
| ERAP2  | 1.24        | 1.08-1.43    | 0.003   |
| S100A14| 1.23        | 1.09-1.38    | 0.001   |
| IL22RA1| 1.23        | 1.08-1.4    | 0.002   |
| PTGS2  | 1.21        | 1.07-1.35    | 0.002   |
| MMP12  | 1.18        | 1.06-1.32    | 0.002   |
| S100P  | 1.18        | 1.08-1.3    | P < 0.001 |
| GREM1  | 1.18        | 1.06-1.32    | 0.003   |
| CXCL9  | 1.17        | 1.07-1.29    | 0.001   |
| S100A2 | 1.17        | 1.09-1.27    | P < 0.001 |
| PI3    | 1.15        | 1.06-1.26    | 0.001   |
| CTSE   | 1.13        | 1.04-1.24    | 0.005   |
| LCN2   | 1.13        | 1.04-1.24    | 0.006   |
| CXCL5  | 1.12        | 1.04-1.19    | 0.001   |
| CHGA   | 0.91        | 0.85-0.97    | 0.003   |
| PAK3   | 0.85        | 0.76-0.95    | 0.006   |
| NPPA   | 0.77        | 0.65-0.91    | 0.002   |

Abbreviation: DEIRGs, differentially expressed immune-related genes. TCGA, The Cancer Genome Atlas.
Supplemental table 4. The result of GSEA by comparing of high- and low- risk group in TCGA PAAD dataset (Top 10).

| ID                          | SetSize | EnrichmentScore | NES     | P values  | P. adjust | Q values  | Rank  |
|-----------------------------|---------|-----------------|---------|-----------|-----------|-----------|-------|
| PATHWAYS_IN_CANCER          | 325     | 0.47662         | 1.893014| 0.001311  | 0.009578  | 0.006082  | 4515  |
| CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION | 263     | 0.653825        | 2.562224| 0.001348  | 0.009578  | 0.006082  | 2916  |
| REGULATION_OF_ACTIN_CYTOSKELETON | 212     | 0.440692        | 1.70072 | 0.001403  | 0.009578  | 0.006082  | 3188  |
| CHEMOKINE_SIGNALING_PATHWAY | 187     | 0.582715        | 2.220964| 0.001427  | 0.009578  | 0.006082  | 1813  |
| FOCAL_ADHESION              | 199     | 0.54239         | 2.073184| 0.001431  | 0.009578  | 0.006082  | 3119  |
| CELL_ADHESION_MOLECULES_CAMS| 131     | 0.631299        | 2.305935| 0.001508  | 0.009578  | 0.006082  | 2235  |
| JAK_STAT_SIGNALING_PATHWAY  | 155     | 0.495857        | 1.839125| 0.001508  | 0.009578  | 0.006082  | 3277  |
| NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY | 131     | 0.614632        | 2.245055| 0.001508  | 0.009578  | 0.006082  | 3240  |
| T_CELL_RECEPTOR_SIGNALING_PATHWAY | 108     | 0.535505        | 1.903617| 0.001546  | 0.009578  | 0.006082  | 5143  |
| LEUKOCYTE_TRANSEPTHELIAL_MIGRATION | 116     | 0.513062        | 1.839902| 0.001553  | 0.009578  | 0.006082  | 3188  |

Abbreviation: GSEA, Gene set enrichment analysis. TCGA, The Cancer Genome Atlas. PAAD, pancreatic adenocarcinoma.
Supplemental Table 1. Characteristics of the datasets in this study.

| Dataset    | Normal | Tumor | Platform |
|------------|--------|-------|----------|
| GSE15471   | 39     | 39    | GPL570   |
| GSE60979   | 12     | 49    | GPL14550 |
| GSE62165   | 13     | 118   | GPL13667 |
| GSE71989   | 8      | 12    | GPL570   |
| GSE91035   | 8      | 25    | GPL22763 |
| GSE102238  | 50     | 50    | GPL19072 |
Supplemental table 2. The result of KEGG pathway analysis of DEIRGs (ranked by $P$ value).

| ID       | Description                                           | Adjusted $P$ value | Gene count | Gene symbol                                                                 |
|----------|-------------------------------------------------------|--------------------|------------|----------------------------------------------------------------------------|
| hsa04060 | Cytokine-cytokine receptor interaction                | 1.05E-13           | 27         | CXCL14, CXCL9, CXCL5, CXCL12, CXCL3, CCL20, IL7R, TNFSF4, CCL18, CXCR4, EPO, GDF10, IL1RN, IL7, INHBA, LIF, TNFRSF11B, TNFSF13B, AMHR2, BMPR2, IL20RB, IL22RA1, IL4R, LIFR, TGFB1, TNFRSF21, FAS |
| hsa05330 | Allograft rejection                                    | 9.92E-08           | 9          | HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, HLA-F, FAS    |
| hsa05332 | Graft-versus-host disease                             | 1.37E-07           | 9          | HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, HLA-F, FAS    |
| hsa04940 | Type I diabetes mellitus                              | 1.62E-07           | 9          | HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, HLA-F, FAS    |
| hsa05416 | Viral myocarditis                                     | 1.78E-07           | 10         | HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, HLA-F, ICAM1, RAC2 |
| hsa05320 | Autoimmune thyroid disease                            | 7.61E-07           | 9          | HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, HLA-F, FAS    |
| hsa04612 | Antigen processing and presentation                   | 1.74E-06           | 10         | CD74, CTSB, HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, HLA-F |
| hsa05323 | Rheumatoid arthritis                                  | 7.52E-06           | 10         | HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, ICAM1, PSMD6, ISG15, IRF7, OAS1, FAS |
| hsa05169 | Epstein-Barr virus infection                          | 7.52E-06           | 14         | HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DOA, HLA-DQA1, HLA-DQB1, ICAM1, PSMD6, ISG15, IRF7, OAS1, FAS |
| hsa04061 | Viral protein interaction with cytokine and cytokine receptor | 1.19E-05           | 10         | CXCL14, CXCL9, CXCL5, CXCL12, CXCL3, CCL20, CCL18, CXCR4, IL20RB, IL22RA1 |

Abbreviation: DEIRGs, differentially expressed immune-related genes. KEGG, Kyoto Encyclopedia of Genes and Genomes.
Supplemental table 3. The result of Univariate Cox regression analysis for DEIRGs in TCGA PAAD dataset (P < 0.01)

| Gene symbol | HR  | CI90          | P value  |
|-------------|-----|---------------|----------|
| GBP2        | 2.01| 1.43-2.84     | P < 0.001|
| PRDX1       | 1.76| 1.2-2.58      | 0.004    |
| TNFRSF21    | 1.64| 1.2-2.24      | 0.002    |
| FAS         | 1.54| 1.15-2.07     | 0.004    |
| S100A10     | 1.53| 1.21-1.94     | P < 0.001|
| S100A16     | 1.53| 1.22-1.93     | P < 0.001|
| TMSB10      | 1.49| 1.14-1.95     | 0.004    |
| OAS1        | 1.45| 1.19-1.76     | P < 0.001|
| PLAU        | 1.42| 1.21-1.68     | P < 0.001|
| SEMA3C      | 1.42| 1.19-1.68     | P < 0.001|
| PLAUR       | 1.4 | 1.14-1.73     | 0.002    |
| NRP2        | 1.4 | 1.12-1.76     | 0.004    |
| MICB        | 1.39| 1.14-1.69     | 0.001    |
| EDNRA       | 1.38| 1.12-1.69     | 0.002    |
| WNT5A       | 1.36| 1.14-1.62     | 0.001    |
| IL1RN       | 1.36| 1.17-1.58     | P < 0.001|
| OASL        | 1.34| 1.16-1.54     | P < 0.001|
| LTBP1       | 1.33| 1.09-1.64     | 0.006    |
| IFITM1      | 1.33| 1.1-1.6       | 0.003    |
| BST2        | 1.31| 1.08-1.58     | 0.005    |
| RSAD2       | 1.31| 1.09-1.57     | 0.004    |
| MX2         | 1.3 | 1.07-1.58     | 0.008    |
| PPARG       | 1.3 | 1.11-1.51     | 0.001    |
| AREG        | 1.3 | 1.14-1.48     | P < 0.001|
| S100A6      | 1.29| 1.08-1.55     | 0.005    |
| IL20RB      | 1.28| 1.15-1.42     | P < 0.001|
| SDC1        | 1.28| 1.08-1.52     | 0.005    |
| NOD2        | 1.27| 1.06-1.51     | 0.008    |
| INHBA       | 1.27| 1.09-1.47     | 0.002    |
| NOX4        | 1.26| 1.06-1.49     | 0.009    |
| DKK1        | 1.25| 1.13-1.37     | P < 0.001|
| ERAP2       | 1.24| 1.08-1.43     | 0.003    |
| S100A14     | 1.23| 1.09-1.38     | 0.001    |
| IL22RA1     | 1.23| 1.08-1.4      | 0.002    |
| PTGS2       | 1.21| 1.07-1.35     | 0.002    |
| MMP12       | 1.18| 1.06-1.32     | 0.002    |
| S100P       | 1.18| 1.08-1.3      | P < 0.001|
| GREM1       | 1.18| 1.06-1.32     | 0.003    |
| CXCL9       | 1.17| 1.07-1.29     | 0.001    |
| Gene   | Ratio    | 95% CI   | P Value |
|--------|----------|----------|---------|
| S100A2 | 1.17     | 1.09-1.27| < 0.001 |
| PI3    | 1.15     | 1.06-1.26| 0.001   |
| CTSE   | 1.13     | 1.04-1.24| 0.005   |
| LCN2   | 1.13     | 1.04-1.24| 0.006   |
| CXCL5  | 1.12     | 1.04-1.19| 0.001   |
| CHGA   | 0.91     | 0.85-0.97| 0.003   |
| PAK3   | 0.85     | 0.76-0.95| 0.006   |
| NPPA   | 0.77     | 0.65-0.91| 0.002   |

Abbreviation: DEIRGs, differentially expressed immune-related genes. TCGA, The Cancer Genome Atlas.
Supplemental table 4. The result of GSEA by comparing of high- and low- risk group in TCGA PAAD dataset (Top 10).

| ID                              | SetSize | EnrichmentScore | NES   | P values | P. adjust | Q values | Rank |
|---------------------------------|---------|-----------------|-------|----------|-----------|----------|------|
| PATHWAYS_IN_CANCER              | 325     | 0.47662         | 1.89301 | 0.00131  | 0.00957   | 0.00608  | 4515 |
| CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION | 263     | 0.653825        | 2.56222 | 0.00134  | 0.00957   | 0.00608  | 2916 |
| REGULATION_OF_ACTIN_Cytoskeleton | 212     | 0.440692        | 1.70072 | 0.00140  | 0.00957   | 0.00608  | 3188 |
| CHEMOKINE_SIGNALING_PATHWAY     | 187     | 0.582715        | 2.22096 | 0.00142  | 0.00957   | 0.00608  | 1813 |
| FOCAL_ADHESION                  | 199     | 0.54239         | 2.07318 | 0.00143  | 0.00957   | 0.00608  | 3119 |
| CELL_ADHESION_MOLECULES_CAMS    | 131     | 0.631299        | 2.30593 | 0.00150  | 0.00957   | 0.00608  | 2235 |
| JAK_STAT_SIGNALING_PATHWAY      | 155     | 0.495857        | 1.83912 | 0.00150  | 0.00957   | 0.00608  | 3277 |
| Pathway                                      | Gene Set | GSEAX | P-value  | P-value  | P-value  | P-value  | FDR   |
|----------------------------------------------|----------|-------|----------|----------|----------|----------|-------|
| NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY    | 131      | 0.614632 | 2.24505  | 0.00150  | 0.00957  | 0.00608  | 3240  |
| T_CELL_RECEPTOR_SIGNALING_PATHWAY            | 108      | 0.535505 | 1.90361  | 0.00154  | 0.00957  | 0.00608  | 5143  |
| LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION         | 116      | 0.513062 | 1.83990  | 0.00155  | 0.00957  | 0.00608  | 3188  |

Abbreviation: GSEA, Gene set enrichment analysis. TCGA, The Cancer Genome Atlas. PAAD, pancreatic adenocarcinoma.