

# Using ESTIMATE algorithm to establish an 8-mRNA signature prognosis prediction system and identify immunocyte infiltration-related genes in Pancreatic adenocarcinoma

Zibo Meng<sup>1,2,\*</sup>, Dianyun Ren<sup>1,2,\*</sup>, Kun Zhang<sup>3,\*</sup>, Jingyuan Zhao<sup>1,2</sup>, Xin Jin<sup>2,4</sup>, Heshui Wu<sup>1,2</sup>

<sup>1</sup>Department of Pancreatic Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

<sup>2</sup>Sino-German Laboratory of Personalized Medicine for Pancreatic Cancer, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

<sup>3</sup>Department of Otorhinolaryngology-Head And Neck Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

<sup>4</sup>Cancer Center, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

\*Equal Contribution

**Correspondence to:** Heshui Wu, Xin Jin; **email:** [heshuiwu@hust.edu.cn](mailto:heshuiwu@hust.edu.cn), [jinxinunion@hust.edu.cn](mailto:jinxinunion@hust.edu.cn)

**Keywords:** pancreatic cancer, tumor microenvironment, immunocytes infiltration, FOXO1

**Received:** August 22, 2019

**Accepted:** March 9, 2020

**Published:** March 17, 2020

**Copyright:** Meng et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**Objective:** The tumour microenvironment is one of the significant factors driving the carcinogenesis of Pancreatic adenocarcinoma (PAAD). However, the underlying mechanism of how the tumour microenvironment impacts the prognosis of PAAD is not completely clear.

**Results:** The transcriptome and clinical data of 182 PAAD program cases were downloaded from the TCGA database. Three hundred thirty-three differentially expressed genes (DEGs) between high and low stromal groups and 314 DEGs between high and low immune score groups were identified using ESTIMATE score. Based on the 203 genes differentially expressed simultaneously in two score-related comparisons, we established an 8-mRNA signature to evaluate the prognosis of PAAD patients. Kaplan-Meier curves showed significantly worse survival for patients with high-risk scores in both the training and validation groups. The risk score was an independent prognostic factor and had a high predictive value for the prognosis of patients with PAAD. By searching the TCGA database, we showed that CA9, CXCL9, and GIMAP7 from the 8-mRNA signature were associated with the infiltration levels of immunocytes by regulating FOXO1 expression in PAAD.

**Conclusions:** Unlike traditional methods of screening for differential genes in cancer and healthy tissues, we constructed a novel 8-mRNA signature to predict the prognosis of PAAD patients by applying ESTIMATE scoring to RNA-seq-based transcriptome data. Most importantly, we identified CA9, CXCL9, and GIMAP7 from the above eight genes as regulators of immunocyte infiltration by adjusting the expression of FOXO1 in PAAD. Thus, CA9, CXCL9, and GIMAP7 might be the ideal targets of immune therapy of PAAD.

**Methods:** ESTIMATE scoring was used to determine the stromal and immune scores of transcriptome datasets downloaded from the TCGA database. An mRNA-based prognostic signature was built for the training cohort via the LASSO Cox regression model. The signature was verified using a validation cohort. Kaplan-Meier curves

and log-rank analysis were used to identify survival differences. Western blot analysis and RT-qPCR analysis were carried out to analyze the expression of specific proteins and mRNAs. IHC was performed to assess the protein levels of Forkhead box-O 1 (FOXO1), Carbonic anhydrase 9 (CA9), C-X-C motif chemokine ligand 9 (CXCL9), and GTPase, IMAP family member 7 (GIMAP7) in the tissue microarray of PAAD.

## INTRODUCTION

Pancreatic adenocarcinoma (PAAD) is one of the most devastating human malignant tumours in the world [1]. Because of its resistance to chemoradiotherapy and the trend of early metastasis, the 5-year survival rate of PAAD is less than 5% [2, 3]. Approximately 39,590 patients die of this disease in the United States yearly [4]. Besides, the morbidity and mortality rates of PAAD are ninth and sixth, respectively, among Malignant tumours in China, which shows a rapid growth pattern that matches the trend in other countries [5]. Thus, a better understanding of the pathogenesis of cancer and the exploration of more therapeutic strategies is imperative for improving the prognosis of PAAD patients.

Malignant solid tumour tissues consist of tumour cells, tumour-associated normal epithelial, vascular cells, immune cells, and stromal cells [6]. Cancer-associated fibroblasts, adipocytes, pericytes, mesenchymal stem cells, endothelial cells, lymphocytes, and extracellular matrix are components of a tumour microenvironment [7]. PAAD is characterized by an intense stromal desmoplastic reaction around cancer cells [8], which plays an essential role in tumorigenesis and drug resistance [9]. Therefore, the tumour microenvironment is one of the significant factors determining the prognosis of PAAD. However, the mechanism of how the tumour microenvironment engineers the carcinogenesis of PAAD is not entirely clear.

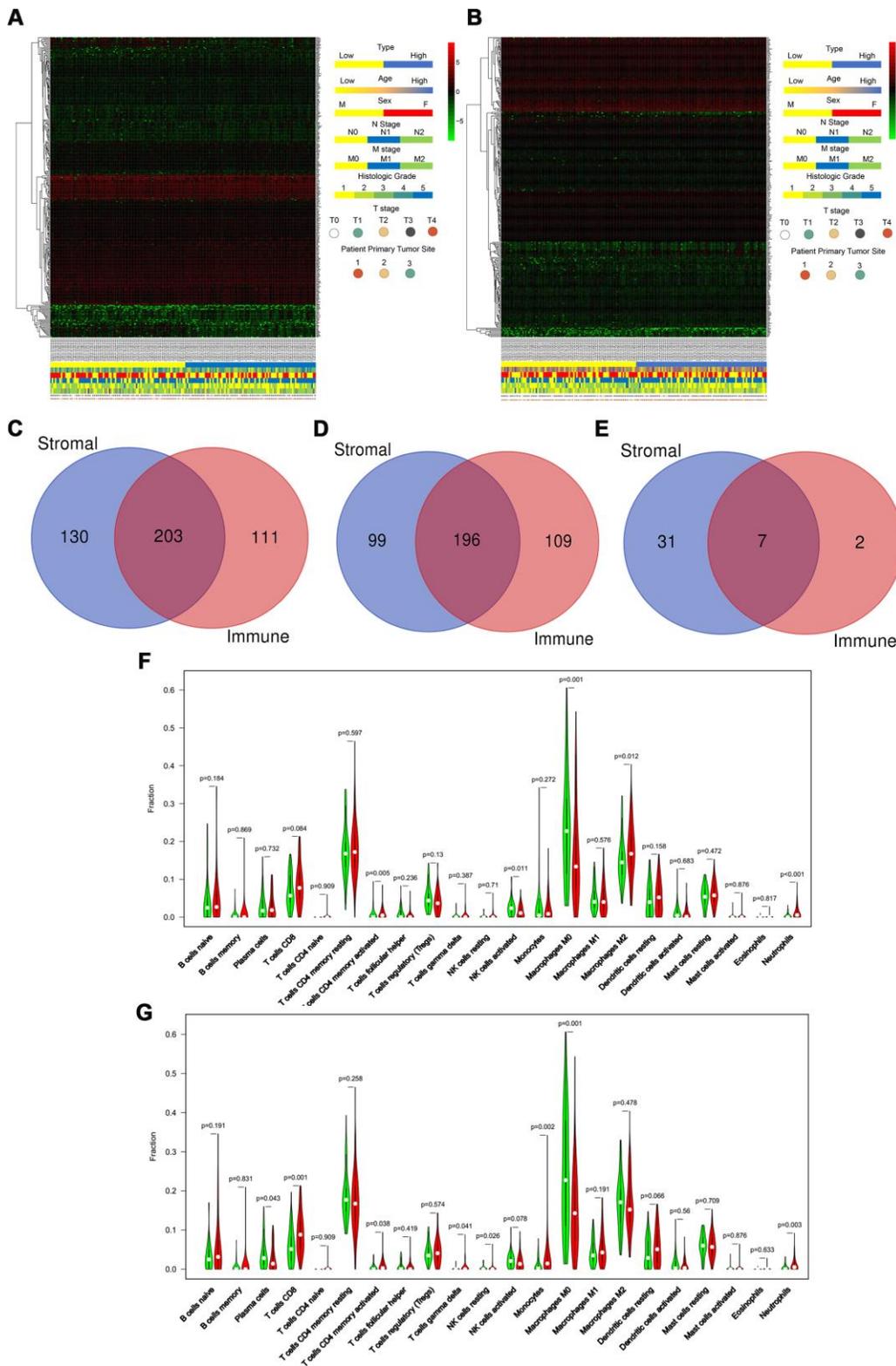
A tumour microenvironment consists of multiple sorts of inflammatory cells and mediators that modulate tumour development, growth, and metastasis [10]. Tumour infiltrated immunocytes encompass macrophages, dendritic cells (DCs), myeloid-derived suppressor cells (MDSCs), and T lymphocytes [10]. Immunocytes in a tumour microenvironment produce a variety of proinflammatory cytokines to maintain chronic inflammation and regulate tumour growth and progression [11]. On the contrary, immunosuppressive Tregs and MDSCs suppress T lymphocyte proliferation [12]. Changes in the proportion and function of different types of tumour-infiltrating immunocytes contribute to PAAD initiation and progression [11, 12]. However, the mechanism regulating the infiltration of immunocytes in the tumour environment of PAAD is poorly understood.

Since immunocytes and stromal cells represent significant components of the tumour microenvironment in PAAD [13, 14], we first calculated the stromal and immune scores of PAAD using 'Estimation of Stromal and Immune cells in Malignant Tumours using Expression data' (ESTIMATE) scoring system to assess the level of infiltrating stromal and immune cells in the tumour microenvironment [15]. Our research using bioinformatics analysis explored differentially expressed genes (DEGs) related to stromal and immune scores in the tumour microenvironment, and we identified eight genes that contributed to the overall survival rate of PAAD. Moreover, we found that three of the eight genes (CA9, CXCL9, and GIMAP7) associated independently with the overall survival rate and have close relationships with immunocyte infiltration in PAAD. Furthermore, we revealed that FOXO1 potentially serves as the downstream target of these three genes in modulating immunocyte infiltration in the tumour microenvironment of PAAD. In summary, our study provided an 8-mRNA signature system to predict the prognosis of PAAD and identified three genes (CA9, CXCL9, and GIMAP7) that were associated with immunocyte infiltration. Most importantly, CA9, CXCL9, and GIMAP7 could serve as the target of immune therapy of PAAD.

## RESULTS

### The ESTIMATE algorithm identifies DEGs associated with stroma and immune scores

We downloaded the transcriptome and clinical data of 182 PAAD program cases from the TCGA database. Based on the ESTIMATE database results and the median values of the stromal and immune scores (excluding the normal pancreas cases), we divided these patients into high and low stromal score groups and high and low immune score groups. We displayed distinct these clinicopathological features, including age, sex, N stage, M stage, histologic grade, T stage, and primary tumor sites, and mRNA expression forms between high and low stromal (Figure 1A) and between high and low immune (Figure 1B) score groups in the form of heatmaps. We found statistically significant results only for sex and M stage between high and low stromal scores ( $P = 0.010$  and  $P = 0.010$ , respectively), whereas only for N stage between high and low immune scores ( $P = 0.049$ ) as shown in Table 1. And the age, M stage, histologic grade, T stage, and primary tumor sites



**Figure 1. ESTIMATE algorithm identifies DEGs associated with stroma and immune scores.** (A and B) Heatmaps displayed distinct mRNA expression forms and clinicopathological features between high and low stromal score groups (A) and between high and low immune score groups (B). (Primary tumor site: 1 for pancreas head, 2 for pancreas body, and 3 for other locations) (C–E) The Venn diagram showed the simultaneously differentially expressed DEGs (C), the simultaneously upregulated differentially expressed DEGs (D), and the simultaneously downregulated differentially expressed DEGs (E) between stromal score and immune score groups. (F, G) CIBERSORT results showed the association between the infiltration levels of immune cells and the stromal-score level (F) and the immune-score level (G).

**Table 1. The clinicopathological characteristics based on low/high score groups in the training set.**

TCGA cohort						
	Groups according to stromal score			Groups according to Immune score		
	Low score (n=89)	High score (n=87)	P	Low score (n=89)	High score (n=87)	P
Age(mean, IQR)	64.5(56.0-74.0)	64.8(58.0-73.0)	0.888	65.0(57.0-73.5)	64.3(56.0-73.0)	0.689
Gender			0.010			0.099
Female	32(36.0)	48(55.2)		35(39.3)	45(51.7)	
Male	57(64.0)	39(44.8)		54(60.7)	42(48.3)	
Race			0.819			0.643
White	79(88.8)	76(87.4)		77(86.5)	78(89.7)	
Other races	10(11.2)	11(12.6)		12(13.5)	9(10.3)	
Primary tumor site			0.135			
Pancreas head	65(73.0)	72(82.8)		63(70.8)	74(85.1)	
Pancreas body	19(21.3)	9(10.3)		21(23.6)	7(8.0)	
Other locations	5(5.0)	6(6.9)		5(5.6)	6(6.9)	
T stage			0.423			0.423
T1	5(5.6)	2(2.3)		5(5.6)	2(2.3)	
T2	13(14.6)	11(12.6)		13(14.6)	11(12.6)	
T3	68(76.4)	72(82.8)		68(76.4)	72(82.8)	
T4	1(1.1)	2(2.3)		1(1.1)	2(2.3)	
TX	2(2.2)	0(0.0)		2(2.2)	0(0.0)	
N stage			0.192			0.049
N0	28(31.5)	21(24.1)		27(30.3)	22(25.3)	
N1	57(64.0)	65(74.7)		57(64.0)	65(74.7)	
NX	4(4.5)	1(1.1)		5(5.6)	0(0.0)	
M stage			0.010			0.085
M0	30(33.7)	49(56.3)		33(37.1)	46(52.9)	
M1	2(2.2)	2(2.3)		3(3.4)	1(1.1)	
MX	57(64.0)	36(41.4)		53(59.6)	40(46.0)	
Grade			0.469			0.329
G1	18(20.2)	12(13.8)		19(21.3)	11(12.6)	
G2	44(49.4)	50(57.5)		45(50.6)	49(56.3)	
G3	24(27.0)	24(27.6)		22(24.7)	26(29.9)	
G4	1(1.1)	1(1.1)		1(1.1)	1(1.1)	
GX	2(2.2)	0(0.0)		2(2.2)	0(0.0)	

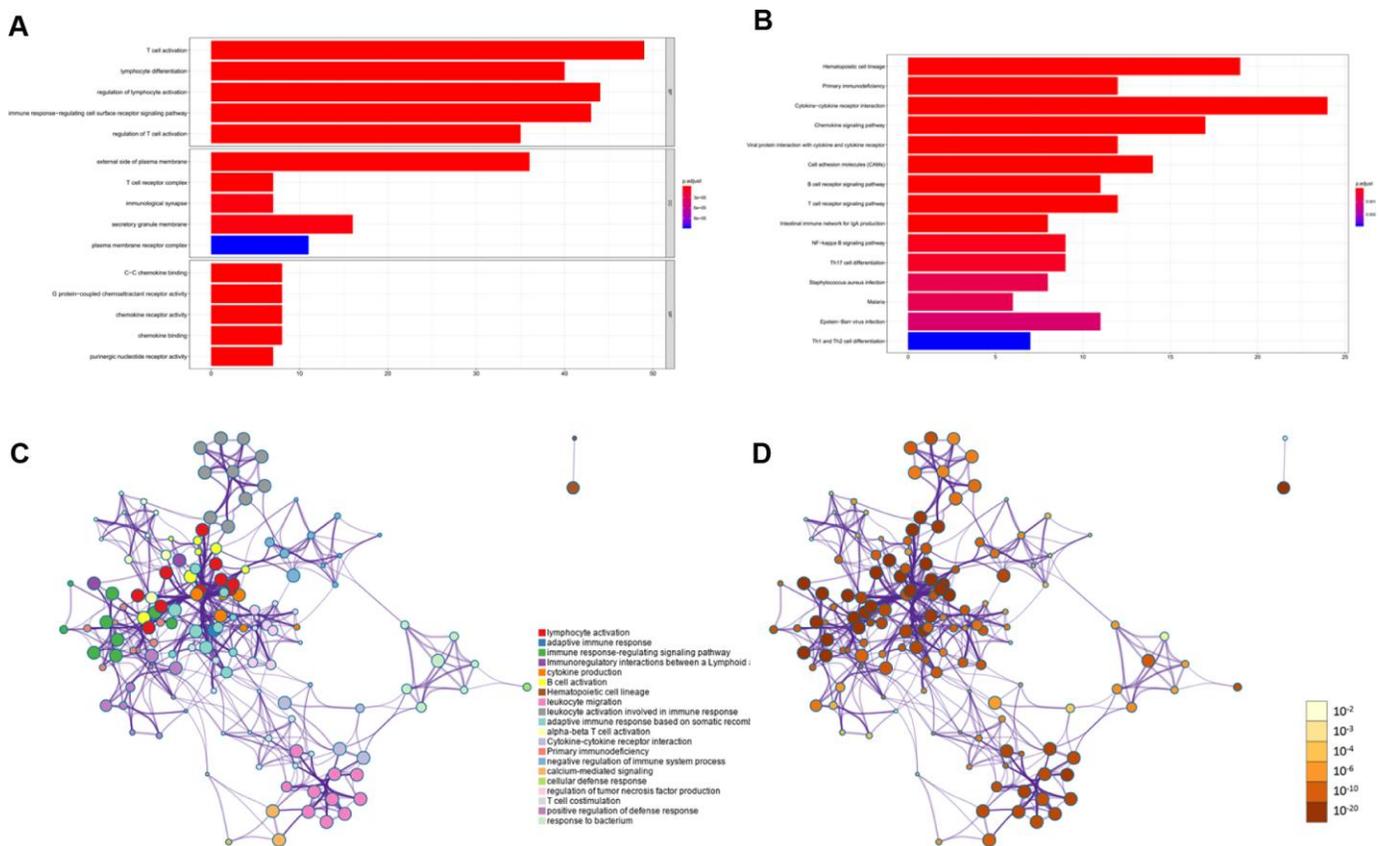
were not significantly different between two groups. Low stromal scores group was characterized by a higher number of male patients and poorly M stages compared with high stromal scores group. Low immune scores

group included more patients with better N stages compared with high immune scores group. Based on our threshold value (log fold change  $\geq 1.5$  and  $P < 0.05$ ), there were 333 DEGs between high and low stromal

groups and 314 DEGs between high and low immune score groups (Supplementary Tables 5 and 6). The Venn diagram (Figure 1C) used to find the co-DEGs in the two groups showed that 203 genes, which accounted for the majority of DEGs, were differentially expressed simultaneously in two score-related comparisons. Moreover, these simultaneously upregulated and downregulated differentially expressed DEGs were 196 and 7, respectively, with the same alternative trend between the two groups (Figure 1D and 1E). As shown in Figure 1F and 1G, CIBERSORT results revealed that CD4<sup>+</sup> activated memory T cells ( $P = 0.005$ ), M2 macrophages ( $P = 0.012$ ), and neutrophils ( $P < 0.001$ ) were upregulated, while activated NK cells ( $P = 0.011$ ) and M0 macrophages ( $P = 0.001$ ) were downregulated in the high stromal-score group. In the high immune-score group, CD8 T cells ( $P = 0.001$ ), gamma delta T cell ( $P = 0.041$ ), resting NK cells ( $P = 0.026$ ), monocytes ( $P = 0.002$ ), and neutrophils ( $P = 0.003$ ) were upregulated, while plasma cells ( $P = 0.043$ ) and M0 macrophages ( $P = 0.001$ ) were downregulated. Hence, we used the ESTIMATE algorithm to identify DEGs associated with stroma and immune scores successfully.

## KEGG and GO functional enrichment analysis for DEGs

To identify differentially regulated genes on immune cell infiltration, we used Metascape tools to conduct the GO and KEGG analyses of the 203 simultaneously expressed DEGs. The Heatmap in Figure 2A exhibits the top 15 enriched GO terms across the DEGs based on Biological Process (BP), Cellular Component (CC), and Molecular Function (MF), and the heatmap in Figure 2B also exhibits the top 15 enriched KEGG pathways across the DEGs. Among them, most terms or pathways were associated with biological processes and lymphocyte activation. We also applied protein-protein interaction networks from the DEGs, which were coloured by different cluster IDs (Figure 2C) and P-values (Figure 2D), and the networks contained more genes that tended to have more significant p-values. Since most networks from the DEGs were related to biological processes and lymphocyte activation, the result confirmed in reverse that the DEGs identified based on stroma and immune scores were accurate. Consequently, combined with the result showed in Figure 1, our study identified that most of the



**Figure 2. KEGG and GO functional enrichment analysis for DEGs.** (A) Heatmap exhibited the top 15 enriched GO terms across the DEGs. (B) Heatmap exhibited the top 15 enriched KEGG pathways across the DEGs. (C–D) Protein-protein interaction networks from the DEGs which coloured by different cluster ID (C) and coloured by p-value (D).

DEGs identified based on the immune and stromal scores were correlated considerably with the immune responses or immune cell infiltration in PAAD.

### **An 8-mRNA signature system was established to predict the overall survival of PAAD patients**

Based on the 203 DEGs obtained from the mRNA difference analysis (Figure 1C), we first conducted Cox univariate regression for initial screening to remove interferences from excessive confounding genes and acquire genes with the most significant impact on prognosis. As shown in Supplementary Table 1, to avoid exclusion of important prognostic genes, 14 of the mRNA with  $P < 0.1$  were then moved into LASSO regression [16, 17]. We presented the LASSO coefficient profiles of the 14 mRNA (Figure 3A) and produced 10-fold cross-validation results that identified optimal values of the penalty parameter  $\lambda$  (Figure 3B).

The forest plot demonstrated the association between each gene and overall survival (Figure 3C). According to these results, we identified an 8-mRNA signature to evaluate the overall survival time of PAAD patients based on the expression of the 8 mRNAs and their regression coefficients as follows: Risk score =  $(-0.01579 \times \text{expression level of ADH1B}) + (0.07838 \times \text{expression level of CA9}) + (-0.76361 \times \text{expression level of CDHR3}) + (0.33812 \times \text{expression level of CXCL9}) + (-0.06606 \times \text{expression level of GIMAP7}) + (-0.70384 \times \text{expression level of ICAM3}) + (-0.11457 \times \text{expression level of LDLRAD1}) + (-0.24179 \times \text{expression level of P2RY8})$ . Patients in the TCGA cohort were divided into a low-risk group ( $N=88$ ) and a high-risk group ( $N=88$ ) utilizing the median risk score as the cut-off value.

The Kaplan-Meier curves showed that high-risk patients had significantly worse survival rates in the training set ( $P < 0.001$ ) (Figure 3D and Supplementary Figure 2A). Moreover, the multivariate analysis revealed that this prognostic signature was an independent prognostic factor in PAAD patients (Table 2).

We conducted a time-dependent ROC analysis at one, two, and three years to assess the prognostic accuracy of the risk score (Figure 3F and Supplementary Figure 2C). We used 69 patients from ICGC PACA-Australia as the validation cohort and calculated the risk score of each case based on our 8-mRNA signature. Per the Kaplan-Meier curves, the high-risk group patients had worse outcomes (Figure 3E and Supplementary Figure 2B); the time-dependent ROC analysis verified that the risk score had a good long-term prognostic accuracy (Figure 3G and Supplementary Figure 2D). We, therefore, successfully established an 8-mRNA signature to assess the prognosis of PAAD patients.

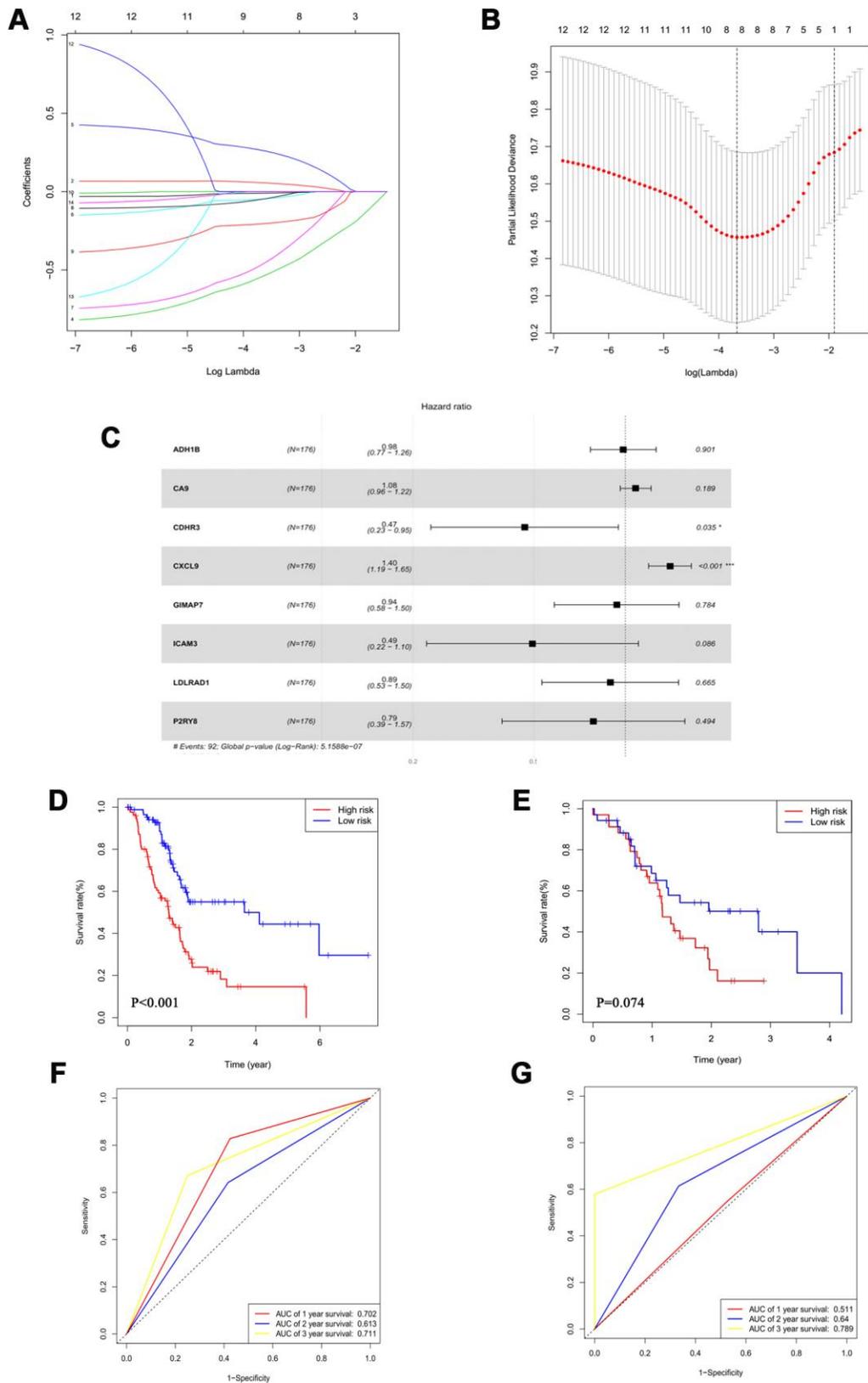
### **CA9, CXCL9, and GIMAP7 have a close relationship with the immune infiltration of the tumour microenvironment in PAAD**

To further explore the regulatory mechanisms for the prognosis of PAAD based on the eight differentially expressed genes in our signature, we searched the GEPIA web tool [18] for an association between the expression of genes at the mRNA level and Overall Survival (OS). The result showed that carbonic anhydrase 9 (CA9), C-X-C motif chemokine ligand 9 (CXCL9), and GTPase, IMAP family member 7 (GIMAP7) were independently associated with the overall survival (OS) rate of PAAD (Figure 4A–4C) but not with the other five factors (Supplementary Figure 1A–1E). So CA9, CXCL9, and GIMAP7 were included for further exploration.

As shown in Figure 1F and 1G, the DEGs were identified based on the immune microenvironment scores, and the GO and KEGG analysis showed that the DEGs correlated significantly with immune reactions or immune cell infiltration in PAAD. So we tried to determine the relationship between the expression levels of these three factors and the immune infiltration level in the tumour microenvironment. The infiltration levels of CD8+ T cells, macrophage cells, neutrophil cells, and dendritic cells were down-regulated by CA9 (Figure 4D). The evaluation of CXCL9 in PAAD showed a positive correlation with infiltration levels of B cells, CD8+ T cells, CD4+ T cells, macrophage cells, neutrophil cells, and dendritic cells (Figure 4E). The analysis of GIMAP7 revealed that the infiltration levels of B cells, CD8+ T cells, CD4+ T cells, macrophage cells, neutrophil cells, and dendritic cells were all up-regulated with the increased expression level of GIMAP7 (Figure 4F). When taken together, these data suggest that CA9, CXCL9, and GIMAP7 contribute to the immune infiltration of the tumour microenvironment in PAAD.

### **CA9, CXCL9, and GIMAP7 regulate the expression level of FOXO1 in PAAD**

Reportedly, the Forkhead box O (FOXOs) family intrinsically influences the anti-tumour immune response and the infiltration levels of immune cells, including CD4+ T cells and CD8+ T cells, B cells, neutrophil cells, macrophage cells, and dendritic cells [19]. Consistent with this report, our data also showed a significant correlation between the expression level of FOXO1 or FOXO3, but not FOXO4, and the infiltration levels of these immune cells (Figure 5A–5C). Furthermore, by searching the GEPIA web tool, we found a negative correlation between the mRNA expression level of CA9 and the FOXOs family in



**Figure 3. An 8-mRNA signature system was established to predict the overall survival of PAAD patients. (A)** LASSO coefficient profiles of 14 mRNA with  $P < 0.01$ . **(B)** 10-fold cross-validations result which identified optimal values of the penalty parameter  $\lambda$ . **(C)** The association between each gene and overall survival. **(D, E)** The Kaplan-Meier curves in the training set **(D)** and the validation set **(E)**. **(F, G)** Time-dependent ROC analysis at 1, 2 and 3 years in the training set **(F)** and the validation set **(G)**.

**Table 2. The univariate and multivariate analysis of prognostic factors in pancreatic cancer patients.**

	TCGA cohort			
	Univariate analysis		Multivariate analysis	
	HR(95%CI)	P	HR(95%CI)	P
Risk score	1.86(1.57-2.20)	<0.001	1.93(1.60-2.34)	<0.001
Age	1.03(1.01-1.05)	0.007		
Tumor location				
Head	1		1	
Body or tail	0.58(0.31-1.06)	0.077		
Other location	0.11(0.02-0.76)	0.025		
Gender				
Female	1			
Male	0.82(0.54-1.24)	0.343		
Race				
White	1			
Others	0.89(0.49-1.64)	0.709		
Stromal score				
Low stromal score	1			
High stromal score	1.07(0.71-1.62)	0.743		
Immune score				
Low immune score	1			
High immune score	0.99(0.65-1.49)	0.945		
Tumor stage				
I	1			
II	2.33(1.07-5.09)	0.033		
III	1.25(0.15-10.26)	0.834		
IV	2.11(0.43-10.28)	0.357		
Grade				
G1	1			
G2	1.98(1.02-3.86)	0.043		
G3	2.62(1.30-5.28)	0.007		
G4	1.65(0.21-12.87)	0.634		

PAAD patients (Figure 5D). On the contrary, the mRNA expression level of CXCL9 correlated positively with FOXOs (Figure 5E), and so did mRNA expression level of GIMAP7 with FOXOs in PAAD (Figure 5F). Consequently, we hypothesised that CA9, CXCL9, and GIMAP7 regulated the infiltration level of immune cells by modifying FOXOs in PAAD.

To determine whether CA9, CXCL9, and GIMAP7 regulated the immune cell infiltration via FOXOs, we first analyzed the relationship between FOXO3 and FOXO4 and these three genes. We knocked down each of CA9, CXCL9, or GIMAP7 using a gene-specific siRNA and detected the changes in the mRNA level of FOXO3 or FOXO4. We found that these three genes (CA9, CXCL9, or GIMAP7) had apparent effects on FOXO3 and FOXO4 in both PANC-1 and SW 1190 cells (Figure 6A–6C). In contrast, the examination of the correlation between

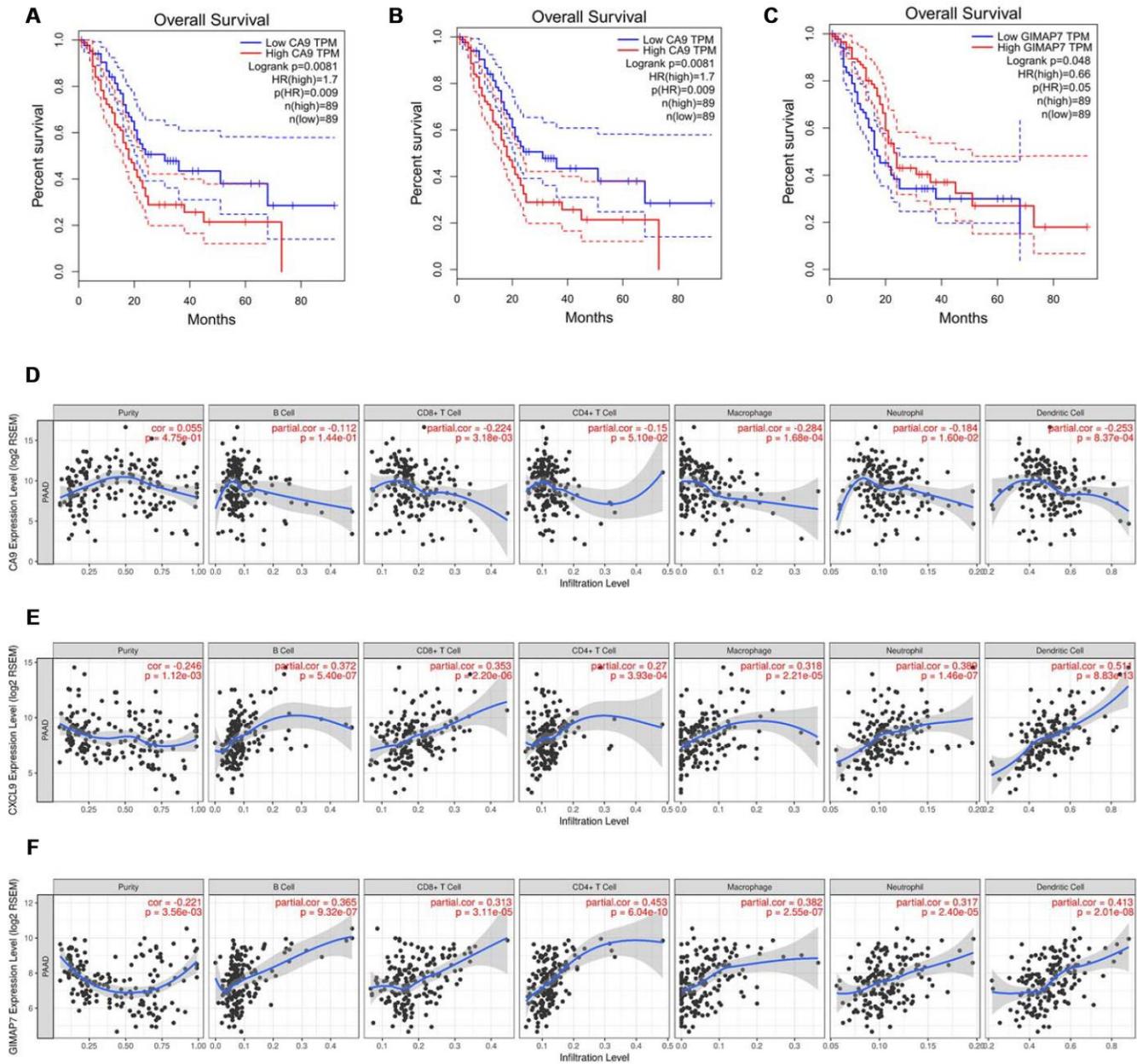
FOXO1 with these three genes indicated that repressing the level of CA9 increased the protein and mRNA levels of FOXO1 in PAAD cells (Figure 6D), whereas, the inhibition of CXCL9 or GIMAP7 by siRNA down-regulated the levels of FOXO1 in PANC-1 and SW 1990 cells (Figure 6E and 6F).

We next examined the downstream inflammatory factors of FOXO1, such as IL-8, IL-10, VEGF, and PD-L1 [20–23] and found that knocking down CA9 significantly decreased the mRNA levels of IL-8, IL-10, VEGF, and PD-L1, but the knockdown of CXCL9 and GIMAP7 substantially increased IL-8, IL-10, VEGF, and PD-L1 expression levels in PAAD cells (Figure 6H and Supplementary Figure 1F). Therefore, our results indicate that CA9 transcriptionally inhibits, while CXCL9 and GIMAP7 transcriptionally promote FOXO1 expression in PAAD.

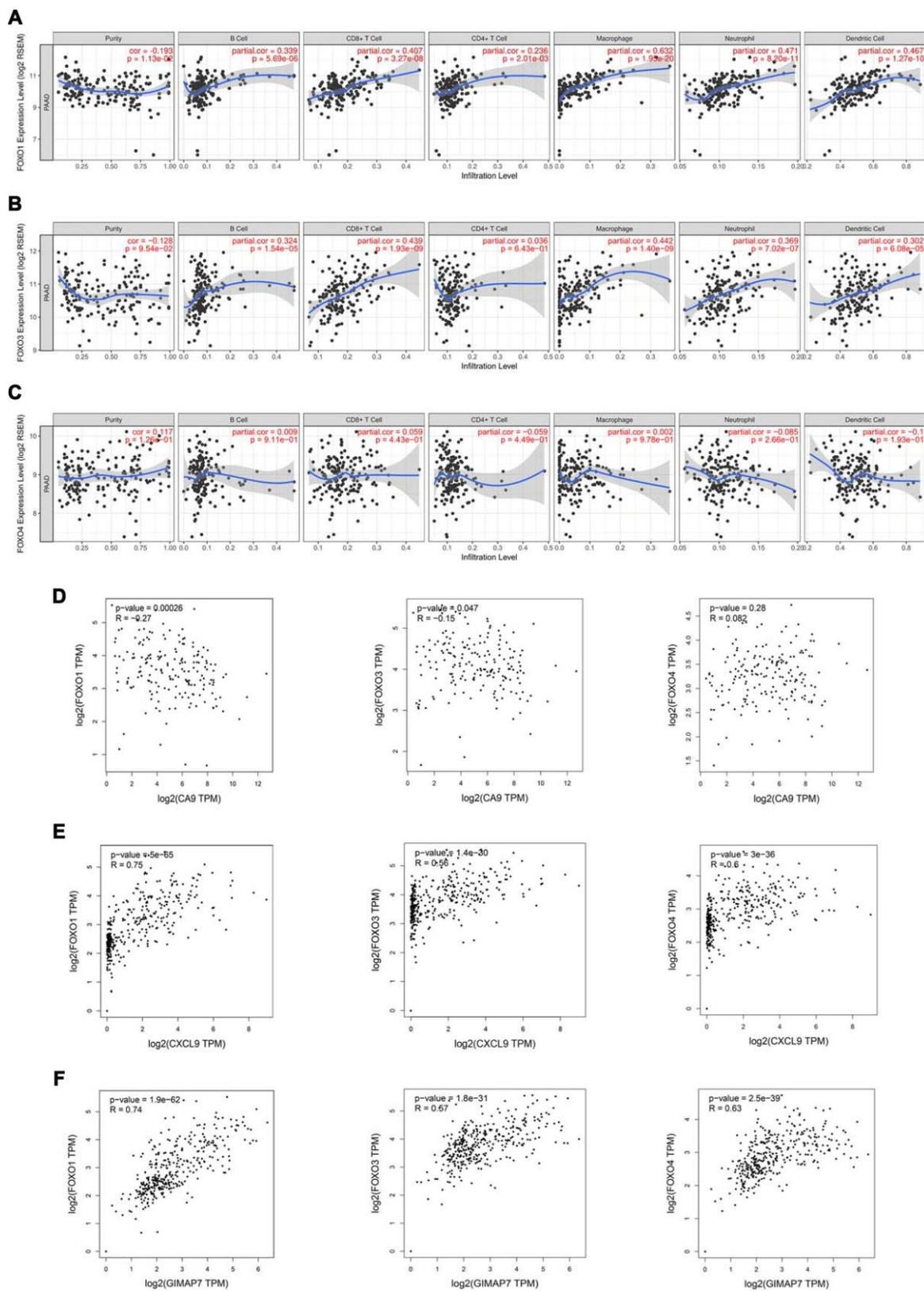
## CA9, CXCL9, and GIMAP7 correlate with FOXO1 in PAAD patient specimens

To determine the correlation between these three proteins and FOXO1 in PAAD specimens, we examined the expression of all four proteins by immunohistochemistry (IHC) on a TMA containing a PAAD samples cohort (n = 31). Representative images of the high and low/no stainings of CA9, CXCL9, GIMAP7, and FOXO1 are shown in Figure 7A, 7D, and

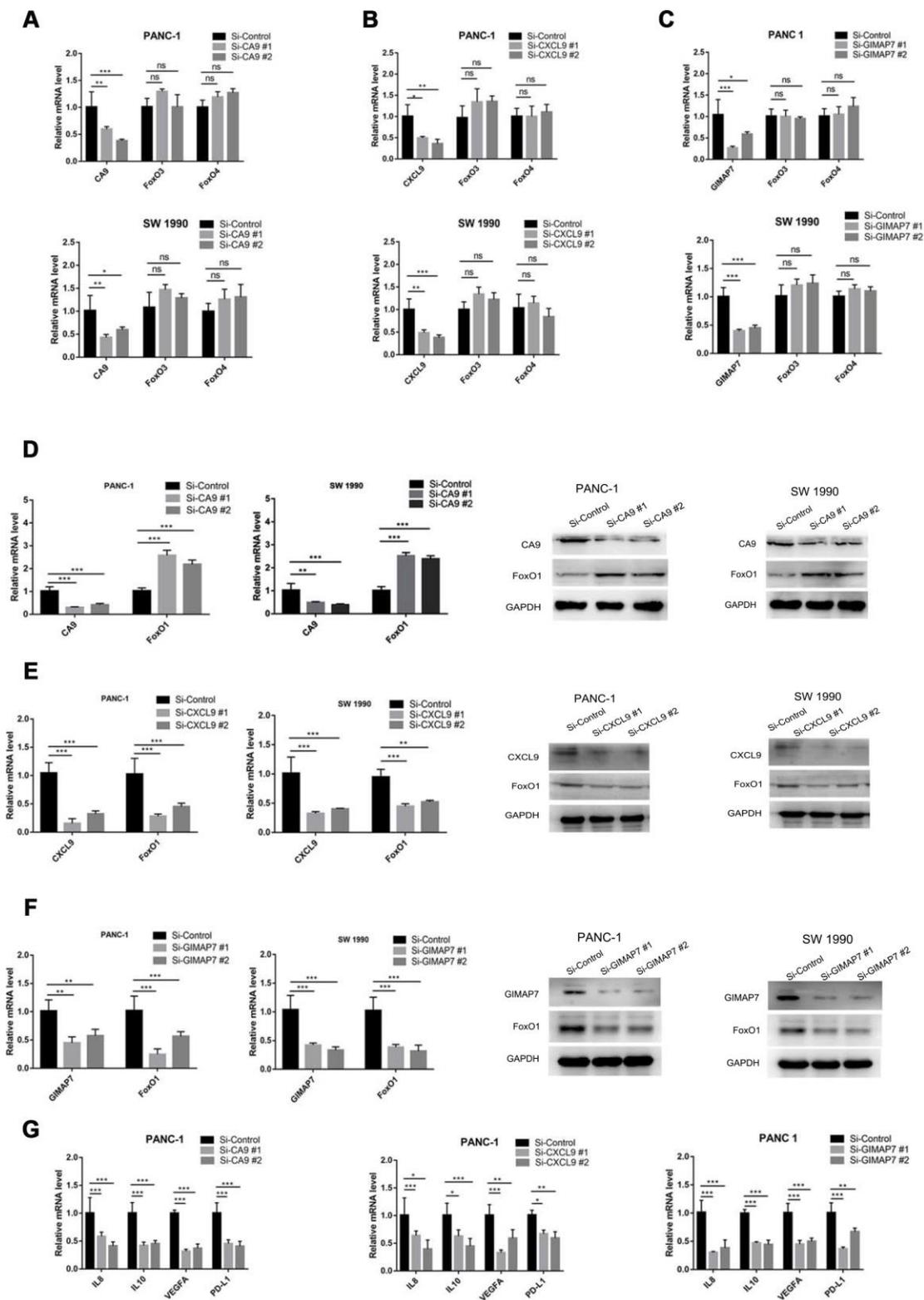
7G. The IHC score was calculated by multiplying the percentage of positively stained cells with the staining intensity. The expression of CA9 correlated negatively with FOXO1 (Spearman's product-moment correlation coefficient  $r = -0.5392$ ,  $P = 0.0017$ ) (Figure 7B, 7C), which was consistent with the protein and mRNA level changes in pancreatic cell lines reported above. Moreover, our data showed that there was a positive relationship between the expression levels of CXCL9 or GIMAP7 and FOXO1 in PAAD patient



**Figure 4. CA9, CXCL9, and GIMAP7 regulate the immune infiltration of tumour microenvironment in PAAD.** (A–C) The overall survival rate of the patients with PAAD were computed with the GEPIA web tool. (D–F) The Timer web tool was used to determine the association between the expression levels of CA9 (D), CXCL9 (E) and GIMAP7 (F) with the infiltration level of immune cells in PAAD samples.



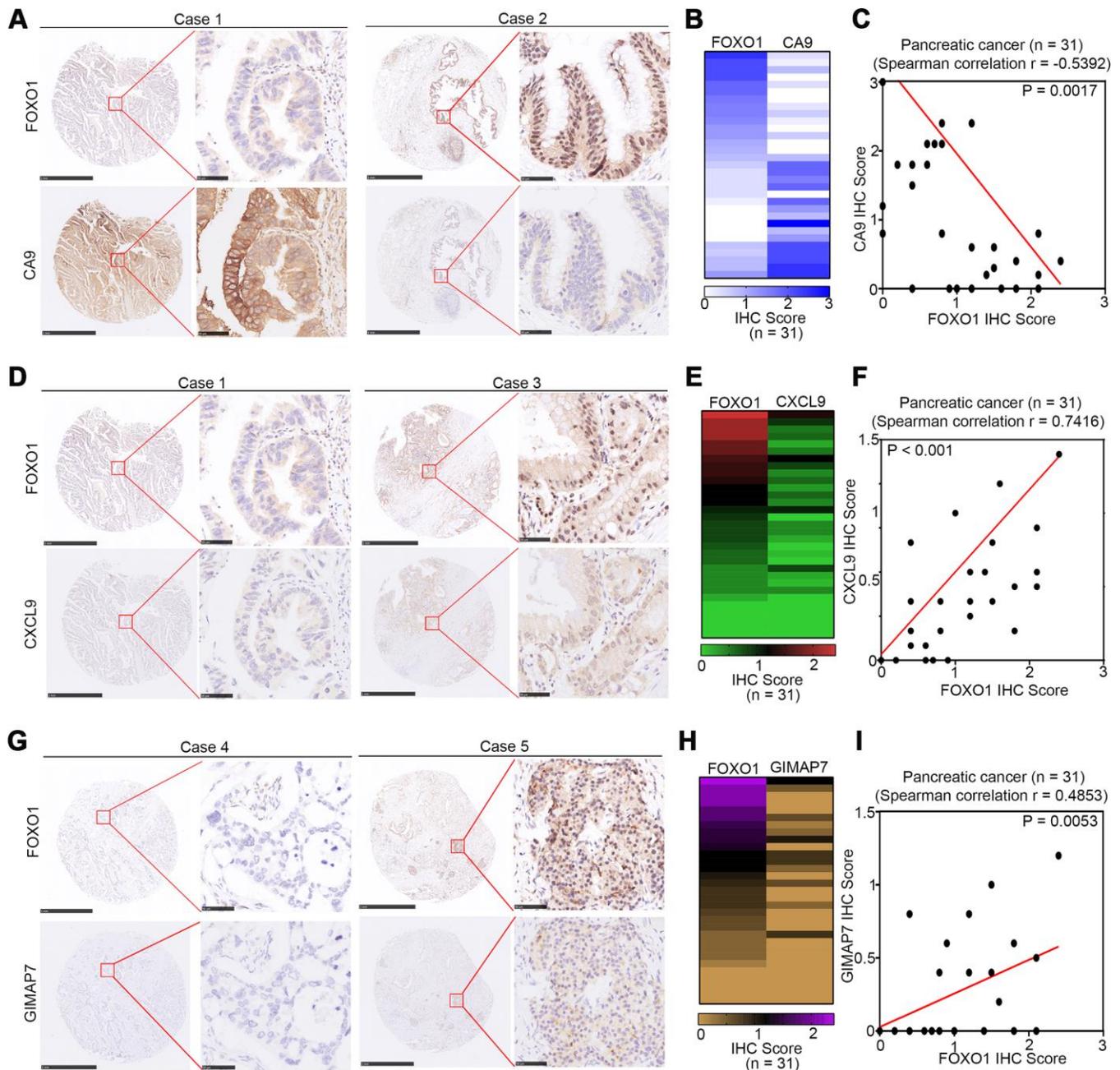
**Figure 5. FOXOs regulate the immune infiltration of tumor microenvironment in PAAD.** (A–C) The TIMER web tool was used to determine the association between the expression levels of FOXO1 (A), FOXO3 (B), and FOXO4 (C) with the infiltration level of immune cells in PAAD samples. (D–F) The GEPIA web tool was used to determine the correlation between the mRNA expression levels of CA9 (D), CXCL9 (E), and GIMAP7 (F) with FoxOs in PAAD samples, respectively.



**Figure 6. CA9, CXCL9, and GIMAP7 regulate the expression level of FOXO1 in PAAD.** PANC-1 and SW 1990 cells were transfected with indicated siRNA. Then, 24 hrs post-transfection, cells were harvested for RT-qPCR. The data shown were the mean values  $\pm$  SD from three replicates. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ . After 48 h post-transfection, cells were harvested for Western Blot analysis. (A–C) PANC-1 and SW 1990 cells transfected with Si-CA9 (A), Si-CXCL9 (B) and Si-GIMAP7 (C) were harvested for RT-qPCR. (D–F) PANC-1 and SW 1990 cells transfected with Si-CA9 (D), Si-CXCL9 (E) and Si-GIMAP7 (F) were harvested for RT-qPCR and western blotting. (G) PANC-1 cells transfected with Si-CA9 (C), Si-CXCL9 (E), or Si-GIMAP7 (G) were harvested for RT-qPCR.

specimens (Spearman's product-moment correlation coefficient  $r = 0.7416$ ,  $P < 0.001$  for CXCL7 and FOXO1, Spearman's product-moment correlation coefficient  $r = 0.4853$ ,  $P = 0.0053$  for GIMAP7 and

FOXO1) (Figure 7D–7G). Consequently, our results suggest that the expression of CA9, CXCL9, and GIMAP7 all correlated with the expression of FOXO1 in PAAD patient specimens.



**Figure 7. CA9, CXCL9, and GIMAP7 are correlated with FOXO1 in PAAD patient specimens respectively.** (A–C) the tissue microarray of PAAD patients (n=31) was stained with FOXO1 and CA9 respectively. The typical image of FOXO1 and CA9 was shown in (A), the IHC scores of FOXO1 and CA9 was shown in (B) and the correlation of these two proteins was shown in (C). (D–F) the tissue microarray of PAAD patients (n=31) was stained with FOXO1 and CXCL9, respectively. The typical image of FOXO1 and CXCL9 was shown in (D), the IHC scores of FOXO1 and CXCL9 was shown in (E) and the correlation of these two proteins was shown in (F). (G–I) the tissue microarray of PAAD patients (n=31) was stained with FOXO1 and GIMAP7, respectively. The typical image of FOXO1 and GIMAP7 was shown in (G), the IHC scores of FOXO1 and GIMAP7 was shown in (H), and the correlation of these two proteins was shown in (I). The scale in A, D, and G represents 1mm or 50um, respectively.

## DISCUSSION

Reportedly, ESTIMATE scores could be used to predict cancer patient survival time and evaluate chemotherapeutic drug resistance [15]. Therefore, applying ESTIMATE to RNA-seq-based transcriptome profiles, as well as clinical data, may help to elucidate the facilitating role of the microenvironment to the infiltration of neoplastic cells and provide new insights into the context in which genomic alterations occur. Stromal and immune cells from the tumour microenvironment play an essential role in tumour initiation and progression and are associated with the prognosis of cancer patients [24]. Since a cancerous pancreas has a high stromal score compared with a healthy pancreas, the ESTIMATE score can be applied for the assessment of the infiltration of immune cells and the presence of stromal cells in tumour samples [25, 26]. In our study, we identified 333 DEGs between high and low stromal groups and 314 DEGs between high and low immune score groups. Among all these genes DEGs, 203 genes were differentially expressed simultaneously in two score-related comparisons.

Several studies have reported signatures that could effectively predict overall patient survival, including a five-miRNA signature [27], a 3-lncRNA signature [28], and a specific four genes signature [29]. However, the differentially expressed molecules in these studies were all conducted based on the expression-related survival risk score, and, due to the heterogeneity among different patients, many genes involved in the development of cancer were ignored. Therefore, we have adopted a new patient grouping method based on the different immune microenvironment scores to explore DEGs.

Based on material from the TCGA database, we obtained a unique 8-mRNA signature from the 203 simultaneously expressed DEGs for the prediction of the overall survival of PAAD. Accordingly, this prognostic signature was an independent prognostic factor in PAAD patients, with excellent long-term prognostic accuracy; the high-risk group patients had worse outcomes. By comparing our finding with other models from literature (AUC of the ROC curve varied from 0.62-0.742) [27–29], our model presented more significant advantages in terms of accuracy in predicting prognosis and was promising to be applied for clinical prognostic evaluation of PC patients. Nevertheless, the existence of limitation couldn't be neglected in our research. The OS and DFS P value of the 8-mRNA model in the validation cohort is not significant as a result of that only 69 patients from ICGC PACA-Australia, which isn't large enough to avoid statistics bias, were conducted in our validation cohort. Thus, further studies should be conducted to verify the predictive efficacy of our signature.

In the study method, several studies had reported that LASSO-penalized regression could increase the accuracy of bioinformatic analysis and allow for simultaneous interpretation of each independent variable to select the most valuable parameters [30–33]. Because there were many confusing features, strong feature selection and shrinkage were still required to prevent overfitting, as well as increase interpretation. To address this issue, the least absolute shrinkage and selection operator (LASSO) Cox regression model, which is suitable for the regression of high-dimensional data [34, 35], was used for the further selection of prognostic mRNAs.

The upregulated CD4+ activated memory T cells, M2 macrophages, and neutrophils in the high stromal-score group may contribute to the poor outcome of PAAD [36]. In an investigation carried out by Tahkola K et al., immune scores were an independent prognostic factor for better disease-free survival and overall survival rates in PAAD patients [37], which may owe much to upregulated CD8 T lymphocytes, gamma delta T lymphocytes, resting NK cells, monocytes, and neutrophils in the high immune-score group [37]. Therefore, understanding the specific mechanism of how the portions of different types of immunocytes are regulated is essential for exploring novel therapeutic strategies to improve the immune response to PAAD. In this study, we revealed that three genes (CA9, CXCL9, and GIMAP7) from the 8-mRNA signature were responsible for the infiltration levels of immunocytes.

It is known that carbonic anhydrase IX (CA9) is a hypoxia-regulated, transmembrane protein associated with neoplastic growth in a broad spectrum of human tumours [38]. CA9 plays a direct role in stimulating an adaptive immune response, and the inhibition of CA9 reduces the capacity of cancer cells to acidify the extracellular environment, which could lead to enhanced immune activity [39, 40]. In stark contrast, CXCL9 is reported to suppress tumours by regulating immune cell migration, differentiation, and activation [41]. Moreover, the GTPases of the immunity-associated protein 7 (GIMAP7) are regulators of lymphocyte survival and homeostasis [42]. Consistent with these facts, our study confirmed that the expression levels of the three genes were related to the infiltration levels of several immune cells.

Forkhead box-O (FOXO) transcription factors have a fundamental role in the development and differentiation of immune cells, including dendritic Cells, T Cells, B Cells, and hematopoietic stem cells [43]. Mammals have 4 FOXO genes, FOXO1, FOXO3, FOXO4, and FOXO6, and these four genes are involved in multiple cellular pathways that regulate proliferation (FOXO1, FOXO3, and FOXO4), oxidative stress resistance (FOXO1 and

FOXO3), metabolism (FOXO1 and FOXO3), cellular differentiation (FOXO3), inflammation (FOXO1, FOXO3, and FOXO4), aging (FOXO1, FOXO3, and FOXO4), and apoptosis (FOXO1, FOXO3, and FOXO4) in mammals [44]. Consistent with the results obtained by searching the TIMER web tool, the immune infiltration level in PAAD correlated with the expression levels of FOXO1 and FOXO3 but not FOXO4. Here, CA9 down-regulated, while CXCL9 and GIMAP7 up-regulated the expression of FOXO1 but not FOXO3 or FOXO4 in PAAD cells. Meanwhile, the IHC staining of PAAD patient specimens also revealed similar findings. Thus, our research demonstrated the critical effect of FoxO1 in regulating the immune response in PAAD.

## CONCLUSIONS

Applying ESTIMATE score to RNA-seq-based transcriptome and based on immune and stromal scores, we identified 333 DEGs between high and low stromal groups and 314 DEGs between high and low immune score groups, of which 203 genes were differentially expressed simultaneously that might regulate the immune response of PAAD patients, including immune cells infiltration level, participating in signalling pathways, and affecting prognosis. Furthermore, we used our results from the LASSO regression to construct a signature to evaluate each patient's risk based on the expression of the 8-mRNA and their regression coefficients, which can predict the prognoses of patients quite accurately, and we hope to apply it in clinical practice. Additionally, we validated that CA9, CXCL9, and GIMAP7 correlated with the OS of PAAD, and these three genes could specifically modulate the expression of FOXO1 to regulate immune infiltration in PAAD. Our investigation, therefore, produced new candidates for improving the immune response to PAAD.

## MATERIALS AND METHODS

### Downloading transcriptome datasets and the stromal and immune scores of PAAD

TCGA PAAD transcriptome FPKM data were downloaded from the GDC Data Portal (<https://portal.gdc.cancer.gov/>). Log<sub>2</sub>-transformed normalized ICGC PACA-Australia data were obtained from UCSC Xena (<https://xenabrowser.net/>). The included criteria for the analysis were: (a) pathological type was pancreatic carcinoma; (b) overall survival (OS) data were available; (c) raw count or normalized gene expression data were available. Per these criteria, we enrolled 176 and 69 pancreatic carcinoma cases for TCGA and UCSC, respectively [29]. Clinical data, such as the age, gender, race, survival time, and status of the selected patients,

were also obtained from the two databases. We got the stromal and immune scores of TCGA PAAD patients from the ESTIMATE database (<https://bioinformatics.mdanderson.org/estimate/index.html>). The TCGA PAAD cohort was set as the training group for prognostic mRNA signature exploration, and the ICGC PACA group was used for validation. The mRNA names of probes were mapped according to their ensemble id, and the median value was used when several inquiries represented the same mRNA.

### Identifying differentially expressed mRNAs based on stromal and immune scores

We separated the TCGA cohort based on the median values of the stromal and immune scores of samples and obtained two groups of DEGs through the linear models for microarray data (LIMMA) R package. Fold change  $\geq 2$  and  $P < 0.05$  were set as the cut-off values for DEGs screening. Heatmaps were drawn using the heatmap R package. We also performed gene ontology (GO) enrichment analysis, KEGG pathway analysis, and protein-protein interaction enrichment analysis for DEGs using the Metascape (<http://metascape.org>) website tool [45].

### Estimating immune cell type proportions

The CIBERSORT R package with an LM22 signature and 1000 permutations was used to calculate the percentages of immune cells in PAAD cases. Per the description of a previous study [46], we inferred the proportions of immune cell types in the transcriptome data of cases and differences in immune cell type composition between different analysis groups (high immune score group vs low immune score group; high stromal score group vs low stromal score group) using the Microenvironment Cell Populations-counter method. To directly visualize the estimation result, we applied the “barplot” and “vioplot” R packages.

### Identifying prognosis-related genes

We selected overall survival (OS) as the endpoint and made verifications using the model for disease-free survival analysis. First, we used the univariate COX regression for the initial mRNA screening. The least absolute shrinkage and selection operator (LASSO) Cox regression model was then employed for the further selection of prognostic mRNAs [47]. We used the “Survival” and “glmnet” R packages for analyses and obtained the final model.

Based on the LASSO Cox regression result, we got a group of mRNAs and built an mRNA-based prognostic signature for the training cohort. Each patient's risk

score was calculated using a combination of the expression levels of mRNAs and LASSO-Cox regression coefficients. We used the median risk score as the cut-off value to divide the training set patients into a high-risk group and low-risk group, and the Kaplan-Meier curves and log-rank analysis were used to identify survival differences. The time-dependent receiver operating characteristic (ROC) curves of the risk scores were generated using the time ROC R package to explore the prognostic accuracy of risk scores. Univariate and multivariable Cox analyses were used to study the independent prognostic values of risk scores compared with other clinical characteristics. Lastly, the ICGC-AU data were used as the validation cohort to verify the effect of the risk-score model.

### Screening DEGs for further exploration

We further screened the genes in the LASSO Cox regression model for their biological regulatory mechanisms in PAAD. We used the webserver GEPIA (<http://gepia.cancer-pku.cn/>) to identify genes that could independently predict overall survival in patients with PAAD [18]. Considering the correlation between these genes and immune components, we also analyzed the relevance of interferon-gamma (as an essential immunoregulatory factor) to these genes. The web server TIMER (<https://cistrome.shinyapps.io/timer/>) was used to investigate the correlation between these genes and tumour-infiltrating immune cells [48].

### Immunohistochemistry (IHC)

To study altered protein expression, an IHC analysis was performed on tissue microarray (TMA) slides purchased from Outdo Biobank (Cat No. XT14-029) (Shanghai, China), using GIMAP7 antibody (Sigma, Cat# HPA020268, 1:50), anti-CXCL9 antibody (Abcam, Cat# ab9720, 1:400), anti-CA9 antibody (Proteintech, Cat# 11071-1-AP, 1:200), and anti-FOXO1 antibody (Proteintech, Cat# 18592-1-AP, 1:1000). Semi-quantitative scoring was performed based upon the staining intensity (negative = 0; weak = 1; moderate = 2; and intense = 3). Two senior pathologists rated the degree of immunostaining of formalin-fixed, paraffin-embedded sections independently in a blinded manner. The scores were determined by the percentage of positive cells multiplied by the staining intensity. The clinical information on Tissue microarray (TMA) slides is shown in Supplementary Table 4.

### Statistical analysis

The R package was used to perform the differential expression genes (DEGs) analysis to obtain the genomic profile between high and low immune/stromal groups.

The LASSO Cox regression model was employed for the further selection of prognostic mRNAs. The Chi-square test was used to calculate differences in clinicopathological variables between groups. The Kaplan-Meier method was used to determine disease-specific survival (DSS) and OS, and the differences between the study groups were compared using the log-rank test. Univariate and multivariate Cox proportional hazard regression models were used to calculate hazard ratios for OS and DSS. All the P values were adjusted using the FDR. A p-value less than 0.05 was considered significant. The described statistical analysis was performed with the IBM SPSS Statistics software version 24 for Windows (IBM Corporation, Armonk, NY, USA).

For RT-PCR and IHC analysis, statistical analysis was carried out using GraphPad Prism 6.0 (GraphPad Software Inc., San Diego, CA). Statistical significance was assessed using the unpaired two-tailed Student t-test between two groups or the one-way ANOVA with Tukey *post hoc* test for multiple comparisons. P values less than 0.05 were considered significant. All the values are expressed as the mean  $\pm$  SD. Asterisks used to indicate significance correspond with \*, P < 0.05; \*\*, P < 0.01; \*\*\*, P < 0.001.

### Abbreviations

ESTIMATE: Estimation of Stromal and Immune cells in Malignant Tumors using Expression data; IHC: immunohistochemistry; DEG: differentially expressed genes; OS: overall survival; TMA: tissue microarray; ADH1B: alcohol dehydrogenase 1B; CA9: carbonic anhydrase 9; CDHR3: cadherin related family member 3; CXCL9: C-X-C motif chemokine ligand 9; GIMAP7: GTPase, IMAP family member 7; ICAM3: intercellular adhesion molecule 3; LDLRAD1: low density lipoprotein receptor class A domain containing 1; P2RY8: P2Y receptor family member 8; FOXOs: Forkhead box-O.

### AUTHOR CONTRIBUTIONS

ZM and DR performed the experiments and wrote the paper, JZ and KZ collected the data. XJ and HW wrote the paper and analyzed the data. All authors read and approved the final manuscript.

### CONFLICTS OF INTEREST

There were no potential conflicts of interest are disclosed.

### FUNDING

This work was supported by grants from the National Natural Science Foundation of China (Grant No. 81702374 (X.J.)).

## REFERENCES

1. Ilic M, Ilic I. Epidemiology of pancreatic cancer. *World J Gastroenterol*. 2016; 22:9694–705.  
<https://doi.org/10.3748/wjg.v22.i44.9694>  
PMID:27956793
2. Goral V. Pancreatic Cancer: pathogenesis and Diagnosis. *Asian Pac J Cancer Prev*. 2015; 16:5619–24.  
<https://doi.org/10.7314/APJCP.2015.16.14.5619>  
PMID:26320426
3. Chu LC, Goggins MG, Fishman EK. Diagnosis and Detection of Pancreatic Cancer. *Cancer J*. 2017; 23:333–42.  
<https://doi.org/10.1097/PPO.000000000000290>  
PMID:29189329
4. Raimondi S, Maisonneuve P, Lowenfels AB. Epidemiology of pancreatic cancer: an overview. *Nat Rev Gastroenterol Hepatol*. 2009; 6:699–708.  
<https://doi.org/10.1038/nrgastro.2009.177>  
PMID:19806144
5. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. *CA Cancer J Clin*. 2016; 66:115–32.  
<https://doi.org/10.3322/caac.21338> PMID:26808342
6. Thomas D, Radhakrishnan P. Tumor-stromal crosstalk in pancreatic cancer and tissue fibrosis. *Mol Cancer*. 2019; 18:14.  
<https://doi.org/10.1186/s12943-018-0927-5>  
PMID:30665410
7. Alderton GK. Microenvironment: an exercise in restraint. *Nat Rev Cancer*. 2014; 14:449.  
<https://doi.org/10.1038/nrc3769> PMID:24898060
8. Ligorio M, Sil S, Malagon-Lopez J, Nieman LT, Misale S, Di Pilato M, Ebricht RY, Karabacak MN, Kulkarni AS, Liu A, Vincent Jordan N, Franses JW, Philipp J, et al. Stromal Microenvironment Shapes the Intratumoral Architecture of Pancreatic Cancer. *Cell*. 2019; 178:160–175.e27.  
<https://doi.org/10.1016/j.cell.2019.05.012>  
PMID:31155233
9. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med*. 2013; 19:1423–37.  
<https://doi.org/10.1038/nm.3394> PMID:24202395
10. Ge P, Wang W, Li L, Zhang G, Gao Z, Tang Z, Dang X, Wu Y. Profiles of immune cell infiltration and immune-related genes in the tumor microenvironment of colorectal cancer. *Biomed Pharmacother*. 2019; 118:109228.  
<https://doi.org/10.1016/j.biopha.2019.109228>  
PMID:31351430
11. Zhang QW, Liu L, Gong CY, Shi HS, Zeng YH, Wang XZ, Zhao YW, Wei YQ. Prognostic significance of tumor-associated macrophages in solid tumor: a meta-analysis of the literature. *PLoS One*. 2012; 7:e50946.  
<https://doi.org/10.1371/journal.pone.0050946>  
PMID:23284651
12. Linehan DC, Goedegebuure PS. CD25+ CD4+ regulatory T-cells in cancer. *Immunol Res*. 2005; 32:155–68.  
<https://doi.org/10.1385/IR.32.1-3.155> PMID:16106066
13. Wörmann SM, Diakopoulos KN, Lesina M, Algül H. The immune network in pancreatic cancer development and progression. *Oncogene*. 2014; 33:2956–67.  
<https://doi.org/10.1038/onc.2013.257> PMID:23851493
14. Xu Z, Pothula SP, Wilson JS, Apte MV. Pancreatic cancer and its stroma: a conspiracy theory. *World J Gastroenterol*. 2014; 20:11216–29.  
<https://doi.org/10.3748/wjg.v20.i32.11216>  
PMID:25170206
15. Liu W, Ye H, Liu YF, Xu CQ, Zhong YX, Tian T, Ma SW, Tao H, Li L, Xue LC, He HQ. Transcriptome-derived stromal and immune scores infer clinical outcomes of patients with cancer. *Oncol Lett*. 2018; 15:4351–57.  
<https://doi.org/10.3892/ol.2018.7855> PMID:29541203
16. Niu Y, Sun W, Chen K, Fu Z, Chen Y, Zhu J, Chen H, Shi Y, Zhang H, Wang L, Shen HM, Xia D, Wu Y. A Novel Scoring System for Pivotal Autophagy-Related Genes Predicts Outcomes after Chemotherapy in Advanced Ovarian Cancer Patients. *Cancer Epidemiol Biomarkers Prev*. 2019; 28:2106–14.  
<https://doi.org/10.1158/1055-9965.EPI-19-0359>  
PMID:31533939
17. Coogan JS, Kim DG, Bredbenner TL, Nicoletta DP. Determination of sex differences of human cadaveric mandibular condyles using statistical shape and trait modeling. *Bone*. 2018; 106:35–41.  
<https://doi.org/10.1016/j.bone.2017.10.003>  
PMID:28987286
18. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017; 45:W98–102.  
<https://doi.org/10.1093/nar/gkx247>  
PMID:28407145
19. Deng Y, Wang F, Hughes T, Yu J. FOXOs in cancer immunity: knowns and unknowns. *Semin Cancer Biol*. 2018; 50:53–64.  
<https://doi.org/10.1016/j.semcancer.2018.01.005>  
PMID:29309928
20. Lappas M. Forkhead box O1 (FOXO1) in pregnant human myometrial cells: a role as a pro-inflammatory mediator in human parturition. *J Reprod Immunol*. 2013; 99:24–32.

<https://doi.org/10.1016/j.jri.2013.04.005>

PMID:[23778262](https://pubmed.ncbi.nlm.nih.gov/23778262/)

21. Hsu P, Santner-Nanan B, Hu M, Skarratt K, Lee CH, Stormon M, Wong M, Fuller SJ, Nanan R. IL-10 Potentiates Differentiation of Human Induced Regulatory T Cells via STAT3 and Foxo1. *J Immunol.* 2015; 195:3665–74.  
<https://doi.org/10.4049/jimmunol.1402898>  
PMID:[26363058](https://pubmed.ncbi.nlm.nih.gov/26363058/)
22. Jeon HH, Yu Q, Lu Y, Spencer E, Lu C, Milovanova T, Yang Y, Zhang C, Stepanchenko O, Vafa RP, Coelho PG, Graves DT. FOXO1 regulates VEGFA expression and promotes angiogenesis in healing wounds. *J Pathol.* 2018; 245:258–64.  
<https://doi.org/10.1002/path.5075> PMID:[29574902](https://pubmed.ncbi.nlm.nih.gov/29574902/)
23. Kawamura A, Kawamura T, Riddell M, Hikita T, Yanagi T, Umemura H, Nakayama M. Regulation of programmed cell death ligand 1 expression by atypical protein kinase C lambda/iota in cutaneous angiosarcoma. *Cancer Sci.* 2019; 110:1780–89.  
<https://doi.org/10.1111/cas.13981> PMID:[30801864](https://pubmed.ncbi.nlm.nih.gov/30801864/)
24. Delitto D, Delitto AE, DiVita BB, Pham K, Han S, Hartlage ER, Newby BN, Gerber MH, Behrns KE, Moldawer LL, Thomas RM, George TJ Jr, Brusko TM, et al. Human Pancreatic Cancer Cells Induce a MyD88-Dependent Stromal Response to Promote a Tumor-Tolerant Immune Microenvironment. *Cancer Res.* 2017; 77:672–83.  
<https://doi.org/10.1158/0008-5472.CAN-16-1765>  
PMID:[27864347](https://pubmed.ncbi.nlm.nih.gov/27864347/)
25. 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](https://pubmed.ncbi.nlm.nih.gov/24113773/)
26. Tahkola K, Mecklin JP, Wirta EV, Ahtiainen M, Helminen O, Böhm J, Kellokumpu I. High immune cell score predicts improved survival in pancreatic cancer. *Virchows Arch.* 2018; 472:653–65.  
<https://doi.org/10.1007/s00428-018-2297-1>  
PMID:[29356891](https://pubmed.ncbi.nlm.nih.gov/29356891/)
27. Shi XH, Li X, Zhang H, He RZ, Zhao Y, Zhou M, Pan ST, Zhao CL, Feng YC, Wang M, Guo XJ, Qin RY. A Five-microRNA Signature for Survival Prognosis in Pancreatic Adenocarcinoma based on TCGA Data. *Sci Rep.* 2018; 8:7638.  
<https://doi.org/10.1038/s41598-018-22493-5>  
PMID:[29769534](https://pubmed.ncbi.nlm.nih.gov/29769534/)
28. Huang GW, Xue YJ, Wu ZY, Xu XE, Wu JY, Cao HH, Zhu Y, He JZ, Li CQ, Li EM, Xu LY. A three-lncRNA signature predicts overall survival and disease-free survival in patients with esophageal squamous cell carcinoma. *BMC Cancer.* 2018; 18:147.  
<https://doi.org/10.1186/s12885-018-4058-6>  
PMID:[29409459](https://pubmed.ncbi.nlm.nih.gov/29409459/)
29. Chen K, He Y, Liu Y, Yang X. Gene signature associated with neuro-endocrine activity predicting prognosis of pancreatic carcinoma. *Mol Genet Genomic Med.* 2019; 7:e00729.  
<https://doi.org/10.1002/mgg3.729> PMID:[31102348](https://pubmed.ncbi.nlm.nih.gov/31102348/)
30. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010; 33:1–22.  
<https://doi.org/10.18637/jss.v033.i01> PMID:[20808728](https://pubmed.ncbi.nlm.nih.gov/20808728/)
31. Cheng P. A prognostic 3-long noncoding RNA signature for patients with gastric cancer. *J Cell Biochem.* 2018; 119:9261–69.  
<https://doi.org/10.1002/jcb.27195> PMID:[30074647](https://pubmed.ncbi.nlm.nih.gov/30074647/)
32. Lao J, Chen Y, Li ZC, Li Q, Zhang J, Liu J, Zhai G. A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. *Sci Rep.* 2017; 7:10353.  
<https://doi.org/10.1038/s41598-017-10649-8>  
PMID:[28871110](https://pubmed.ncbi.nlm.nih.gov/28871110/)
33. Wu Y, Wang PS, Wang BG, Xu L, Fang WX, Che XF, Qu XJ, Liu YP, Li Z. Genomewide identification of a novel six-lncRNA signature to improve prognosis prediction in resectable hepatocellular carcinoma. *Cancer Med.* 2018; 7:6219–33. <https://doi.org/10.1002/cam4.1854>  
PMID:[30378276](https://pubmed.ncbi.nlm.nih.gov/30378276/)
34. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, Liang C, Tian J, Liang C. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. *Radiology.* 2016; 281:947–57.  
<https://doi.org/10.1148/radiol.2016152234>  
PMID:[27347764](https://pubmed.ncbi.nlm.nih.gov/27347764/)
35. Shahraki HR, Salehi A, Zare N. Survival Prognostic Factors of Male Breast Cancer in Southern Iran: a LASSO-Cox Regression Approach. *Asian Pac J Cancer Prev.* 2015; 16:6773–77.  
<https://doi.org/10.7314/APJCP.2015.16.15.6773>  
PMID:[26434910](https://pubmed.ncbi.nlm.nih.gov/26434910/)
36. von Ahrens D, Bhagat TD, Nagrath D, Maitra A, Verma A. The role of stromal cancer-associated fibroblasts in pancreatic cancer. *J Hematol Oncol.* 2017; 10:76.  
<https://doi.org/10.1186/s13045-017-0448-5>  
PMID:[28351381](https://pubmed.ncbi.nlm.nih.gov/28351381/)
37. Tahkola K, Leppänen J, Ahtiainen M, Väyrynen J, Haapasaari KM, Karttunen T, Kellokumpu I, Helminen O, Böhm J. Immune cell score in pancreatic cancer-

- comparison of hotspot and whole-section techniques. *Virchows Arch.* 2019; 474:691–99.  
<https://doi.org/10.1007/s00428-019-02549-1>  
PMID:30843106
38. Logsdon DP, Shah F, Carta F, Supuran CT, Kamocka M, Jacobsen MH, Sandusky GE, Kelley MR, Fishel ML. Blocking HIF signaling via novel inhibitors of CA9 and APE1/Ref-1 dramatically affects pancreatic cancer cell survival. *Sci Rep.* 2018; 8:13759.  
<https://doi.org/10.1038/s41598-018-32034-9>  
PMID:30214007
39. Wang Y, Wang XY, Subjeck JR, Kim HL. Carbonic anhydrase IX has chaperone-like functions and is an immunoadjuvant. *Mol Cancer Ther.* 2008; 7:3867–77.  
<https://doi.org/10.1158/1535-7163.MCT-08-0603>  
PMID:19074859
40. Chafe SC, McDonald PC, Saberi S, Nemirovsky O, Venkateswaran G, Burugu S, Gao D, Delaidelli A, Kyle AH, Baker JH, Gillespie JA, Bashashati A, Minchinton AI, et al. Targeting Hypoxia-Induced Carbonic Anhydrase IX Enhances Immune-Checkpoint Blockade Locally and Systemically. *Cancer Immunol Res.* 2019; 7:1064–78.  
<https://doi.org/10.1158/2326-6066.CIR-18-0657>  
PMID:31088846
41. Tokunaga R, Zhang W, Naseem M, Puccini A, Berger MD, Soni S, McSkane M, Baba H, Lenz HJ. CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation - A target for novel cancer therapy. *Cancer Treat Rev.* 2018; 63:40–47.  
<https://doi.org/10.1016/j.ctrv.2017.11.007>  
PMID:29207310
42. Schwefel D, Arasu BS, Marino SF, Lamprecht B, Köchert K, Rosenbaum E, Eichhorst J, Wiesner B, Behlke J, Rocks O, Mathas S, Daumke O. Structural insights into the mechanism of GTPase activation in the GIMAP family. *Structure.* 2013; 21:550–59.  
<https://doi.org/10.1016/j.str.2013.01.014>  
PMID:23454188
43. Cabrera-Ortega AA, Feinberg D, Liang Y, Rossa C Jr, Graves DT. The Role of Forkhead Box 1 (FOXO1) in the Immune System: Dendritic Cells, T Cells, B Cells, and Hematopoietic Stem Cells. *Crit Rev Immunol.* 2017; 37:1–13.  
<https://doi.org/10.1615/CritRevImmunol.2017019636>  
PMID:29431075
44. Murtaza G, Khan AK, Rashid R, Muneer S, Hasan SM, Chen J. FOXO Transcriptional Factors and Long-Term Living. *Oxid Med Cell Longev.* 2017; 2017:3494289.  
<https://doi.org/10.1155/2017/3494289>  
PMID:28894507
45. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* 2019; 10:1523.  
<https://doi.org/10.1038/s41467-019-09234-6>  
PMID:30944313
46. Zeng D, Li M, Zhou R, Zhang J, Sun H, Shi M, Bin J, Liao Y, Rao J, Liao W. Tumor microenvironment characterization in gastric cancer identifies prognostic and immunotherapeutically relevant gene signatures. *Cancer Immunol Res.* 2019; 7:737–50.  
<https://doi.org/10.1158/2326-6066.CIR-18-0436>  
PMID:30842092
47. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med.* 1997; 16:385–95.  
[https://doi.org/10.1002/\(SICI\)1097-0258\(19970228\)16:4<385::AID-SIM380>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3)  
PMID:9044528
48. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 2017; 77:e108–10.  
<https://doi.org/10.1158/0008-5472.CAN-17-0307>  
PMID:29092952

## SUPPLEMENTARY MATERIALS

### Supplementary Materials and Methods

#### Cell culture and RNA interference

PANC-1 and SW1990 cells were purchased from the Chinese Academy of Science Cell Bank. All cell lines were cultured in RPMI 1640 supplemented with 10% FBS. All cell lines were kept in a 37°C incubator at 5% CO<sub>2</sub>. The gene-specific siRNAs and control siRNA were purchased from RIBOBIO (Guangzhou, China). Pancreatic cancer cell line SW1990, PANC-1 and BxPC-3 were transfected with GIMAP7 siRNA, CXCL9 siRNA, CA9 siRNA or empty vector using Lipofectamine 2000 (Invitrogen Life Technologies) following the manufacturer's instruction. Six hours after transfection, the medium was replaced with fresh 1640 RPMI, containing 10% FBS. After 24–72 h of incubation in a humidified atmosphere (37°C and 5% CO<sub>2</sub>), the experiments were performed. The sequence of siRNA, as indicated in Supplementary Table 2.

#### Western blotting analysis

Cells were washed for three times with phosphate buffer saline (KH<sub>2</sub>PO<sub>4</sub> 2mM, Na<sub>2</sub>HPO<sub>4</sub> 8mM, NaCl 136mM, KCl 2.6mM, pH7.2-7.4) on ice for more than 30 min and the cells were lysed on ice in prechilled lysis buffer (50mM Tris-HCl pH 7.5, 100mM NaCl, 1mM EDTA, 1% NP-40, 0.1% SDS, 20mM sodium fluoride, 2mM sodium orthovanadate, 20mM sodium pyrophosphate), containing a cocktail of protease inhibitors (Roche Diagnostics GmbH). The equal amounts of protein were subjected to SDS-PAGE analysis and transferred to PVDF membranes

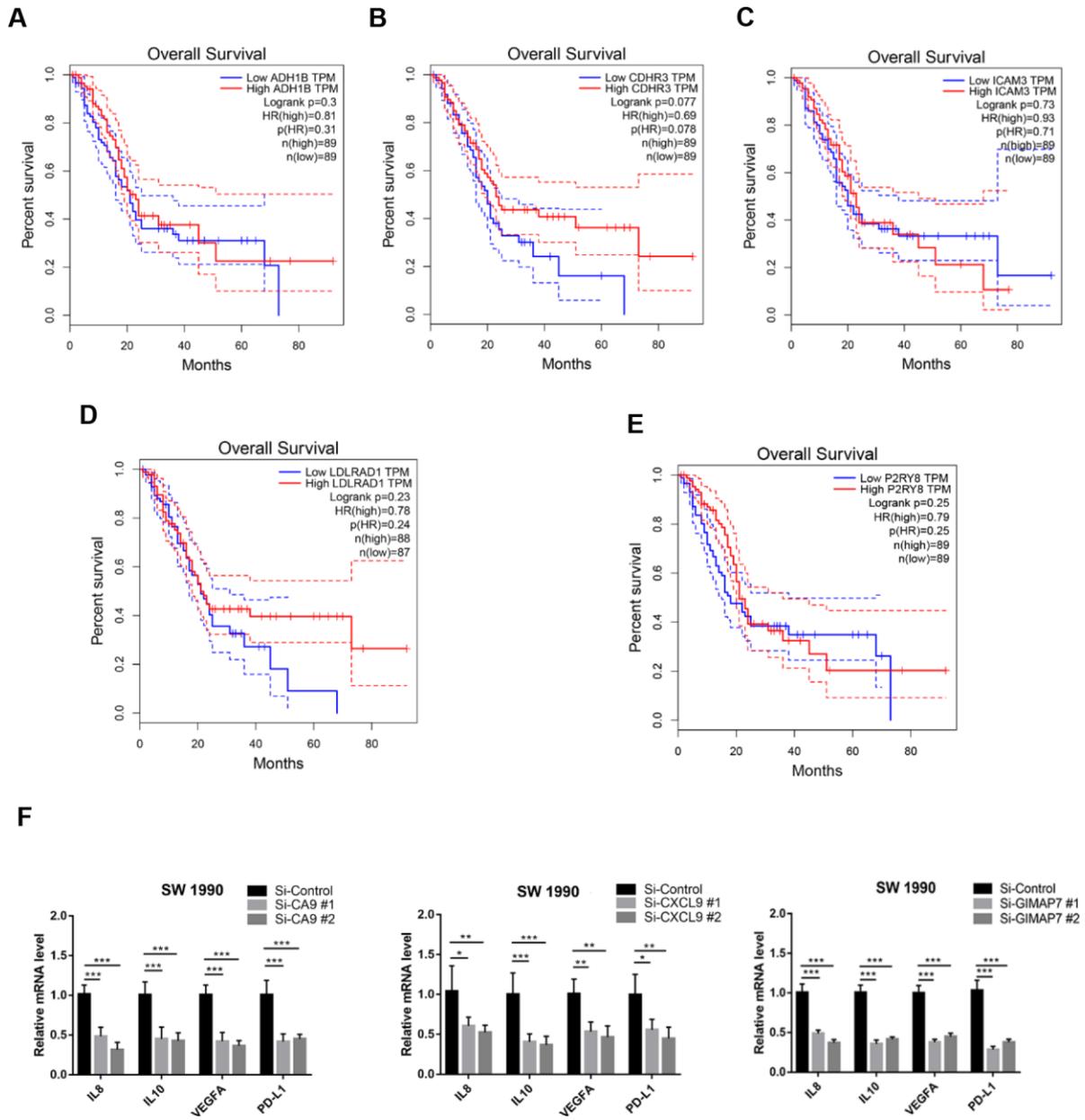
(Pierce Biotechnology, USA). Membranes were fixed in 5% not-fat milk at room temperature for 1 hr and incubated at 4°C overnight with primary antibodies: anti-GIMAP7 (Sigma, Cat# HPA020268), anti-CXCL9 (Abcam, Cat# ab9720), anti-CA9 (Proteintech, Cat# 11071-1-AP), and anti-FOXO1 (Proteintech, Cat# 18592-1-AP). The membranes were stripped and re-probed with an anti-GAPDH antibody (Proteintech, Cat# 10494-1-AP) as a control. After washing for more than three times with TBST, the membranes were incubated for 1 hr with goat anti-rabbit IgG. The protein bands were washed for three times and treated with ECL detection reagents and exposed to X-ray films.

#### RNA extraction and real-time reverse transcription PCR

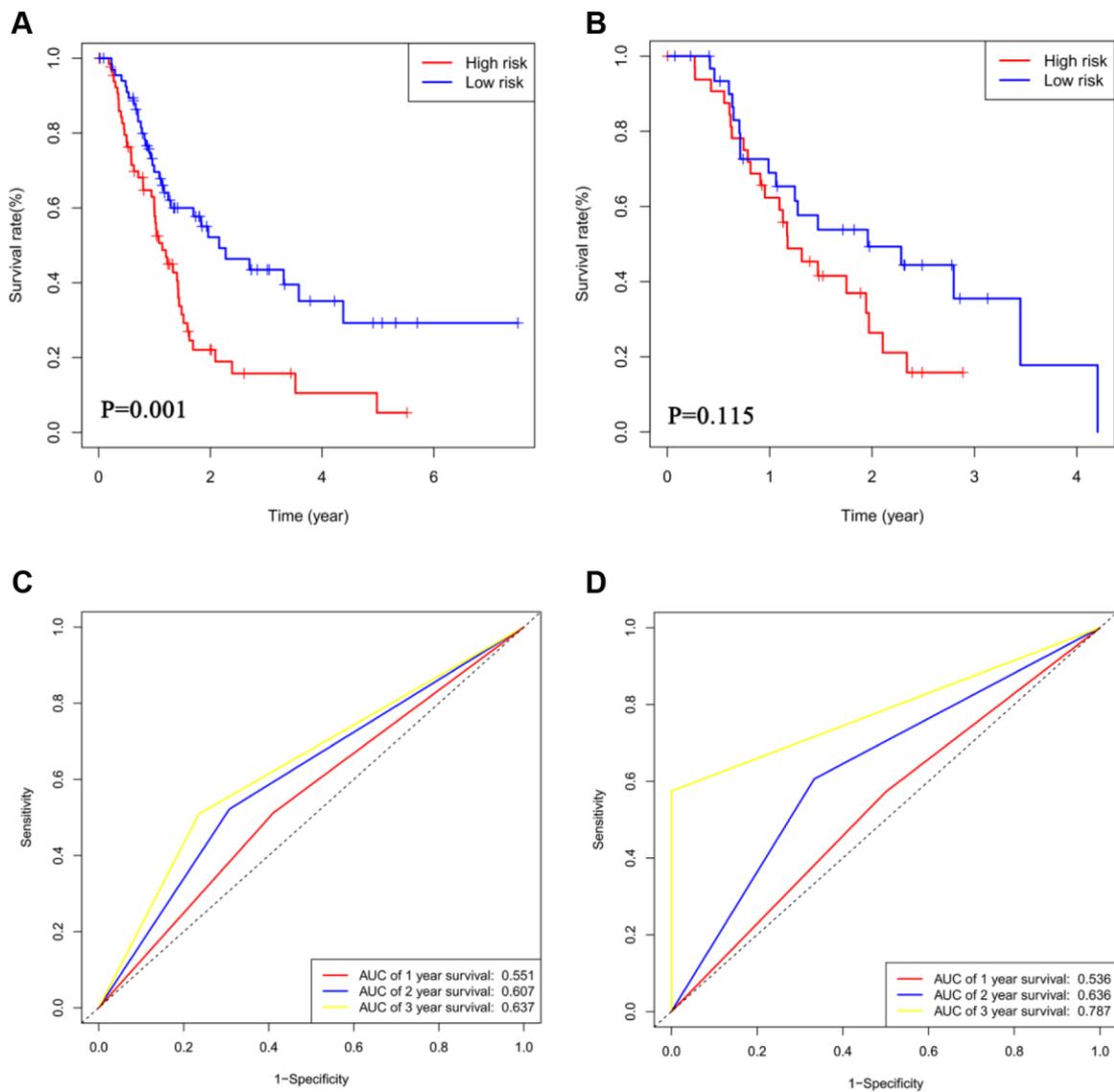
Total cellular RNA was isolated from cells using the RNeasy isolation kit (Qiagen, Cat. # 74106). RNA samples (1 µg) were reverse-transcribed using PrimeScript™ RT reagent Kit (TaKaRa, Cat. # RR037A). Quantitative real-time PCR was performed using One-Step SYBR PrimeScript™ PLUS RT PCR Kit (TaKaRa, Cat. # RR096A).

The sequences of the primers used for qRT-PCR are listed in Supplementary Table 1. Relative quantities of each gene were calculated using the Ct method with GAPDH as a reference. Values represent the average of three technical replicates from at least three independent experiments. RNA extraction and real-time reverse transcription PCR. The sequences of RT-qPCR primers are indicated in Supplementary Table 3.

Supplementary Figures



**Supplementary Figure 1.** (A–E) The overall survival (OS) of the patients with PAAD were computed with the GEPIA web tool. (F) SW 1990 cells transfected with indicated constructs. 48 h post-transfection, cells were harvested for RT-qPCR analysis. The data shown are the mean values  $\pm$  SD from three replicates. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ .



**Supplementary Figure 2.** (A and B) The Kaplan-Meier curves for DFS analysis in the training set (A) and the validation set (B). (C and D) Time-dependent ROC analysis for DFS analysis at 1, 2 and 3 years in the training set (C) and the validation set (D).

## Supplementary Tables

Please browse Full Text version to see the data of Supplementary Table 1

### Supplementary Table 1. Cox univariate regression for initial screening.

### Supplementary Table 2. Sequences of gene-specific siRNAs.

siControl	Provided by RiboBio (Guangzhou, China)
siCA9-1	TTCAGCCGCTACTTCCAAT
siCA9-2	CAGCCGCTACTTCCAATAT
siCXCL9-1	CCAAGGGACTATCCACCTA
siCXCL9-2	GTTCGAAAATCTCAACGTT
siGIMAP7-1	GGCATAGGTTTTTGTGCGAA
siGIMAP7-2	GACACCACCTGCAAGGAAA

### Supplementary Table 3. Sequences of RT-qPCR primers.

Species	Gene	Forward (5'-3')	Reverse (5'-3')
Human	<i>GAPDH</i>	ACCCAGAAGACTGTGGATGG	TTCAGCTCAGGGATGACCTT
Human	<i>CXCL9</i>	CCAGTAGTGAGAAAGGGTTCGC	AGGGCTTGGGGCAAATTGTT
Human	<i>CA9</i>	TTTGCCAGAGTTGACGAGGC	GCTCATAGGCACTGTTTTCTTCC
Human	<i>GIMAP7</i>	TCCAGGGCTCTTTGACACCA	AGAAGTAGGACAATAGCATGGGG
Human	<i>FOXO3</i>	ACGGCTGACTGATATGGCAG	CGTGATGTTATCCAGCAGGTC
Human	<i>FOXO4</i>	CCGGCAAAGCTCTTGGTG	GGTCCACATATCGGCTTCTCA
Human	<i>FOXO1</i>	TCGTCATAATCTGTCCCTACACA	CGGCTTCGGCTCTTAGCAAA
Human	<i>IL-8</i>	ACTGAGAGTGATTGAGAGTGGAC	AACCCTCTGCACCCAGTTTTTC
Human	<i>VEGFA</i>	AGGGCAGAATCATCACGAAGT	AGGGTCTCGATTGGATGGCA
Human	<i>IL-10</i>	GACTTTAAGGGTTACCTGGGTTG	TCACATGCGCCTTGATGTCTG
Human	<i>PD-L1</i>	GCTGCACTAATTGTCTATTGGGA	AATTCGCTTGTAGTCGGCACC

**Supplementary Table 4. Clinical information of Tissue microarray (TMA) slides.**

No.	Surgical organ	Sex(M/F)	Age	Pathological type	Pathological grading	Tumor size(cm)	Tumor location	T	N	M	AJCC 7th Cancer Staging Manual
1	pancreas	M	65	PDAC	II	2*2*2cm	pancreas head	T1	N0	M0	IA
2	pancreas	M	76	PDAC	II-III	1.5*1*1cm	pancreas head	T1	N0	M0	IA
3	pancreas	F	59	PDAC	II	2*2*1cm	pancreas head	T1	N0	M0	IA
4	pancreas	F	68	PDAC	II	3.5*2*2cm	pancreas head	T2	N0	M0	IB
5	pancreas	M	46	PDAC	II	8*6*4cm	pancreas body/tail	T2	N0	M0	IB
6	pancreas	M	74	PDAC	II-III	3.5*3*0cm	pancreas body	T2	N0	M0	IB
7	pancreas	M	50	PDAC	II	6*4*3cm	pancreas body/tail	T3	N0	M0	IIA
8	pancreas	M	53	PDAC	I-II	3.5*3.5*3cm	pancreas head	T3	N0	M0	IIA
9	pancreas	M	66	PDAC	II	3*2.5*2cm	pancreas head	T3	N0	M0	IIA
10	pancreas	F	75	PDAC	I-II	3*3*3cm	pancreas head	T3	N0	M0	IIA
11	pancreas	M	56	PDAC	III	4*3*3cm	pancreas body/tail	T2	N1	M0	IIB
12	pancreas	M	72	PDAC	I-II	5*5*4cm	pancreas head	T2	N1	M0	IIB
13	pancreas	M	64	PDAC	II-III	3*2*2cm	pancreas body	T2	N1	M0	IIB
14	pancreas	M	64	PDAC	II	4*2.5*2.5cm	pancreas head	T2	N1	M0	IIB
15	pancreas	M	72	PDAC	II- III	5*3*3cm	pancreas body/tail	T2	N1	M1	IV
16	pancreas	F	68	PDAC	II-III	3.5*2.5*1.5cm	pancreas head	T3	N1	M1	IV
17	pancreas	M	41	PDAC	II-III	4*3*3cm	pancreas body/tail	T3	N1	M1	IV
18	pancreas	M	59	PDAC	II-III	2.5*1.5*1cm	pancreas head	T3	N1	M1	IV
19	pancreas	M	63	PDAC	I-II	3.5*3*2.5cm	pancreas head	T2	N1	M1	IV
20	Metastases	F	71	PDAC	IV	—	liver Metastases	—	—	M1	IV
21	Metastases	—	—	PDAC	IV	—	liver Metastases	—	—	M1	IV
22	Metastases	M	74	PDAC	IV	—	liver Metastases	—	—	M1	IV
23	Metastases	M	47	PDAC	IV	—	abdominal soft tissue Metastases	—	—	M1	IV
24	pancreas	M	44	PDAC	II	3×1.5×1cm	pancreas head	T2	N1	M0	IIB
25	pancreas	F	52	PDAC	II	6×3×3cm	pancreas head	T2	N1	M0	IIB
26	pancreas	M	60	PDAC	II	4×3×3cm	pancreas head	T2	N1	M0	IIB
27	pancreas	M	62	PDAC	II	0.5×0.5×0.5cm	pancreas head	T1	N1	M0	IIB
28	pancreas	M	61	PDAC	II	7×4×2cm	pancreas head	T2	N1	M0	IIB
29	pancreas	M	52	PDAC	II	4×2×2cm	pancreas	T2	N1	M0	IIB
30	pancreas	M	63	PDAC	II	3.5×2.5×1.5cm	pancreas head	T2	N1	M0	IIB
31	pancreas	M	51	PDAC	II	5×4×3cm	pancreas body/tail	T2	N1	M0	IIB

Please browse Full Text version to see the data of Supplementary Tables 5, 6

**Supplementary Table 5. DEGs between high and low stromal score groups.**

**Supplementary Table 6. DEGs between high and low immune score groups.**