

Identification and Validation of T Cell Exhaustion Signature for Predicting Prognosis and Immune Response in Pancreatic Cancer by Integrated Analysis of Single-Cell and Bulk RNA Sequencing Data

Yaowu Zhu , [Li Tan](#) , [Danju Luo](#) , [Xiong Wang](#) *

Posted Date: 6 February 2024

doi: 10.20944/preprints202402.0281.v1

Keywords: pancreatic cancer; T cell exhaustion; immunotherapy; risk model; SPOCK2



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Identification and Validation of T Cell Exhaustion Signature for Predicting Prognosis and Immune Response in Pancreatic Cancer by Integrated Analysis of Single-Cell and Bulk RNA Sequencing Data

Yaowu Zhu ^{1,#}, Li Tan ^{2,#}, Danju Luo ^{3,*} and Xiong Wang ^{1,*}

¹ Department of Laboratory Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. Bin Wang, Email: yaowu_zhu@163.com.

² Department of Infection Control, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. Li Tan, Email: tanlidyx@126.com.

³ Department of Pathology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030 China.

* Correspondence: Danju Luo, Email: 395393924@qq.com. Xiong Wang, Email: wangxiong@tjh.tjmu.edu.cn, Tel: +86 027 83663414

Yaowu Zhu and Li Tan equally contributes to this work.

Abstract: Purpose: Pancreatic cancer (PAAD) is one of the most lethal malignancies. Immunotherapy is largely ineffective in PAAD. T-cell exhaustion contributes to the resistance for immunotherapies. We investigated the prognostic potential of T cell exhaustion-related genes (TEXGs). **Methods:** Single-cell RNA (scRNA) sequencing dataset from TISCH, and bulk sequencing datasets from TCGA and GTEx were applied to screen differently expressed TEXGs. Kaplan-Meier survival, LASSO regression, and Univariate/Multivariate Cox regression analysis were performed to construct TEXGs risk model. This model was used to predict the prognosis, tumor immune microenvironment, and immunotherapy response. PAAD cohorts from ICGC and GSE71729 datasets were used to validate this risk model. Pan-cancer expression of SPOCK2 was performed in TISCH database. **Results:** A six-gene (SPOCK2, MT1X, LIPH, RARRES3, EMP1, and MEG3) risk model was constructed. Low-risk PAAD patients had longer survival time in both training dataset (TCGA-PAAD, n=178) and validation datasets (ICGC-PACA-CA, ICGC-PAAD-US, and GSE71729, n=412). Multivariate Cox regression analysis demonstrated that the risk score was an independent prognostic variable for PAAD. High-risk patients correlated with cancer-related KEGG pathway enrichment, higher mutation frequency of KRAS and TP53 genes, and immunosuppressive status. Immunohistochemistry staining confirmed the changes of TEXGs in clinical samples. Moreover, pan-cancer scRNA sequencing datasets from TISCH analysis indicated that SPOCK2 might be a novel marker of exhausted CD8⁺ T cells. **Conclusion:** In conclusion, we established and validated a T cell exhaustion-related prognostic signature for PAAD patients. Moreover, our study suggests that SPOCK2 may be a novel marker of exhausted CD8⁺ T cells.

Keywords pancreatic cancer; T cell exhaustion; immunotherapy; risk model; SPOCK2

Introduction

Pancreatic cancer (PAAD) is the seventh leading cause of cancer-related death, with slightly increased incidence and mortality rates in many countries [1]. PAAD will surpass breast cancer to become the third leading cause of cancer-related death by 2025 [2]. The prognosis of PAAD is poor, and the 5-year survival rate is about 2%~9% [3]. Conventional therapies, including surgical intervention, chemotherapy, radiotherapy, and targeted therapies, are insufficient in most cases [4].

Immunotherapy has had remarkable advances in many malignant tumors, while PAAD is refractory to almost all FDA-approved immunotherapies. PAAD exhibits a “cold” tumor immune microenvironment (TIME) characterized by poor T cell infiltration and low mutational burden. Strategies seeking to combine chemotherapy with immune checkpoint blockade (ICB), and dual checkpoint blockade exhibited minimal activity in PAAD [5]. Personalized immunotherapy for individuals based on PAAD genotypic and immunological heterogeneity may provide novel insights [6].

T cell exhaustion (TEX) is one of the main causes for immunotherapy failure, which is characterized by sustained expression of inhibitory receptors, poor effector function, and distinct transcriptional state different from functional effector or memory T cells [7]. CD8⁺ T cell exhaustion (CD8TEX) is driven by massive immunosuppressive signals including nutrient deficiency, hypoxia, pro-inflammatory cytokines, and chronic TCR signaling in the tumor microenvironment (TME), regarded as the main responder to ICB, and correlated with poor prognosis in numerous cancers [8]. Reinvigoration of CD8TEX indicates great therapeutic potential and may become a breakthrough for improving ICB efficacy [9].

In this study, we extracted TEX-related genes utilizing the single-cell RNA sequencing (ScRNA-seq) dataset from the Tumor Immune Single-Cell Hub (TISCH) database, combined with differentially expressed genes between PAAD tumor and normal tissues from the Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) databases, to construct the TEX prognostic risk model. We revealed that TEX risk model can predict the outcome and TIME status in patients with PAAD. Our analysis show that TEX-related gene signature is an independent and promising prognostic model for PAAD.

Materials and methods

Data source and clinical information

Integrated gene expression and clinical information of PAAD patients and normal subjects in TCGA and GTEx databases were downloaded from Xena (<https://xena.ucsc.edu/>), an online platform for public, multi-omic and clinical/phenotype data. ICGC-PACA-CA and ICGC-PAAD-US PAAD cohorts were retrieved from the International Cancer Genome Consortium (ICGC, <https://dcc.icgc.org/>). GSE71729 dataset was downloaded from the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>). All ScRNA-seq datasets were downloaded from TISCH, a scRNA-seq database including 190 datasets focusing on TME. Adjusted P-value <0.05, and $|\log_2 \text{FC}| > 2$ were used as cutoff values for filtering differentially expressed genes (DEGs) in bulk sequencing datasets. Adjusted P-value <0.05, and $|\log_2 \text{FC}| > 0.3$ were used as cutoff values for filtering DEGs in scRNA-seq datasets. DESeq2 (v4.41.12), edgeR (v3.40.2), and limma (v3.54.2) R packages were used for DEGs analysis. Benjamini & Hochberg (BH) was used to get adjusted p-value in DEG analysis using DESeq2, edgeR, and limma.

Construction of T cell exhaustion signature

DEGs from TCGA-PAAD and GTEx were intersected with DEGs of CD8TEX from TISCH to screen CD8TEX-related DEGs. Univariate Cox regression and Kaplan-Meier survival analysis of overall survival (OS) were performed with P-value cutoff of 0.05 to filter OS correlated TEX-related DEGs using the survival (v3.3.1) R package. LASSO regression followed by stepwise variable selection process to obtain the best candidate for Cox's proportional hazards model using the glmnet (v4.1.7) and My.stepwise (v0.1.0) R packages. The final risk model was constructed using the Multivariate Cox regression analysis using the survival (v3.3.1) R package, and the forest plot was drawn using the forestplot (v3.1.1) R package. The PAAD patients were divided into high- and low-risk subgroups with the cutoff of median risk scores.

Validation of T cell exhaustion-related risk model

In the training TCGA-PAAD dataset, Kaplan-Meier survival analysis was performed to compare the OS between high- and low-risk groups. Risk score curve, patient distribution and expression of TEX-related genes in the risk prognostic model were also compared using tinyarray (v2.2.9) R package. The 1-, 3-, and 5-year survival curves were constructed and the Area Under

Curve (AUC) was determined using the survival (v3.3.1) and survminer (v0.4.9) R packages, and plotted with the timeROC (v0.4) R package [10]. Univariate and multivariate Cox regression analysis were performed to identify independent prognostic factors including age, gender, grade, stage, and risk score. A nomogram was drawn to assess risk based on these characteristics using the rms (v6.6-0) R package. The Decline Curve Analysis (DCA) was used to evaluate the clinical applicability of the nomogram via quantifying the improved benefits with various threshold probabilities using the ggDCA (v1.2) R package.

In the validation cohorts, including ICGC-PACA-CA (n=186), ICGC-PAAD-US (n=112), and GSE71729 (n=114), Kaplan-Meier survival analysis was performed to validate the prognostic prediction potential of the TEX-related gene signature. Risk score curve, patient distribution and expression of TEX-related genes in the risk prognostic model were also compared using the tinyarray (v2.2.9)

Gene set variation analysis (GSVA)

The 186 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were extracted from the Molecular Signatures Database (MSigDB) C2: curated gene sets using the msigdb (v7.5.1) R package. The scores of these KEGG pathways were calculated with the GSVA (v1.46.0) R package [11], and significantly different KEGG pathways between high- and low-risk TCGA-PAAD patients.

Mutation landscape analysis

The mutation and clinical information of TCGA patients were downloaded The Pan-Cancer Atlas (PanCanAtlas, <https://gdc.cancer.gov/about-data/publications/pancanatlas>). The mutation landscape and differences between high- and low-risk patients was summarized and plotted using the maftools (v2.14.0) R package [12]. The Gene Ontology (GO) biological process analysis of differentially mutated genes between high- and low-risk patients was performed on Enrichr [13], an online platform for gene set enrichment analysis (<https://maayanlab.cloud/Enrichr/>), and plotted using ggplot2 (v3.4.2) R package.

Immune cell infiltration estimation

The gene signatures for immune cell type classification and gene signatures for immune function were extracted from previous study [14]. The immune cell abundance and immune function estimation were performed using the 'ssgsea' function with the GSVA R package. Moreover, XCELL, TIMER, QUANTISEQ, and CIBERSORT algorithms were applied to estimate the immune cell infiltration using the immunedeconv (v2.1.3) and CIBERSORT (v0.1.0) R packages [15]. The spearman correlation between the infiltration ratio and risk score was performed using the psych (v2.3.3) R package, and a P-value of 0.05 was used as threshold.

Immune subtype analysis

Tumors could be classified into six immune subtypes [16], named C1 to C6. C1 (Wound Healing) is characterized by high proliferation rate and upregulated expression of angiogenic genes. C2 (IFN- γ Dominant) had a strong CD8 signal and the highest M1/M2 macrophage polarization. C3 (Inflammatory) had low somatic copy number alterations and tumor cell proliferation, while showed elevated Th17 and Th1 genes. C4 (Lymphocyte Depleted) displayed a more prominent macrophage signature and high M2 response. C5 (Immunologically Quiet), was dominated by M2 macrophages and exhibited the highest macrophage responses but the lowest lymphocyte responses. C6 (TGF- β Dominant) had the highest TGF- β signature. The TCGA-PAAD patients were classified using the ImmuneSubtypeClassifier (v0.1.0) R package.

Drug response prediction

The Genomics of Drug Sensitivity in Cancer 2 (GDSC2, <https://www.cancerrxgene.org/>) database consists of 805 types of cells and their response to 198 anti-tumor drugs. The cell line expression and half-maximal inhibitory concentration (IC50) data were downloaded GDSC2. The drug response of the TCGA-PAAD patients to these compounds was predicted using the oncoPredict (v0.2) R package [17]. The correlation between drug response and risk score was calculated using the psych (v2.3.3) R package, and plotted with ggstatsplot (v0.0.6) and ggpubr (v0.6.0) R packages.

Immunohistochemistry staining

The surgical PAAD tumor and adjacent tissues were briefly deparaffinized and rehydrated, followed by incubation with primary antibodies overnight (anti-SPOCK2: 11725-1-AP; anti-MT1X: 7172-1-AP; anti-LIPH: 16602-1-AP) from Proteintech (<https://ptgcn.com/products/>). These sections were rinsed with PBS, incubated with corresponding secondary antibodies at 37° C.

Pan-cancer scRNA-seq analysis of CD8TEX-related genes

The expression of TEX-related genes was explored and plotted in the PAAD-CRA001160 scRNA-seq dataset from TISCH database (<http://tisch.comp-genomics.org/home/>). Then a pan-cancer expression of SPOCK2 cross a variety of cell types was analyzed and plotted in heatmap and violin plots.

Statistical analysis

All statistical analysis was carried out utilizing R software (v4.2.3). Comparison of mean values between different groups was performed using the Student's t-test. The correlation analysis was assessed using Pearson correlation analysis. The log-rank test was used in the Kaplan-Meier survival analysis. P-value of adjusted P-value <0.05 considered statistically significant.

Results

Identification of TEX-related gene signature in PAAD

The general workflow of the current study is depicted in Figure 1.

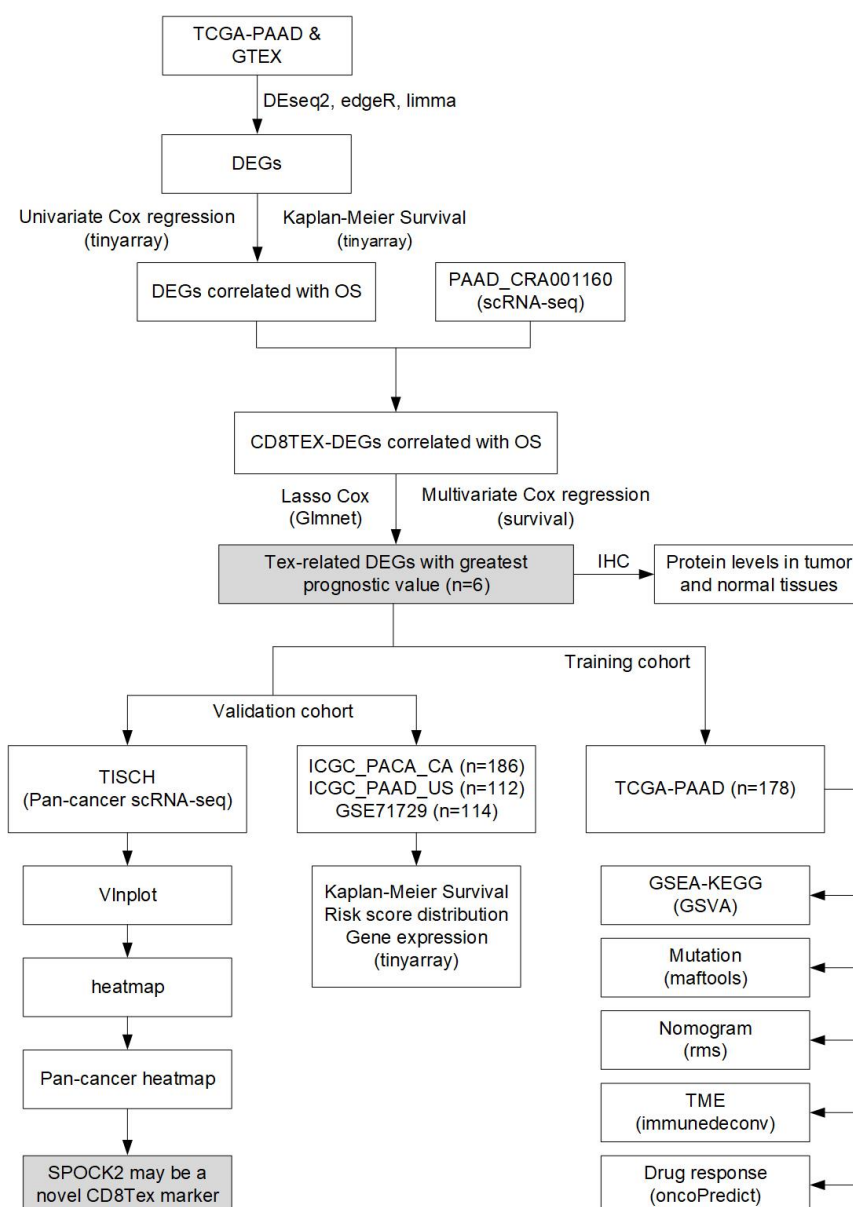


Figure 1. The general workflow of the current study. scRNA-seq, single cell RNA sequencing; DEGs, differently expressed genes; OS, overall survival; LASSO, least absolute shrinkage, and selection operator; IHC, Immunohistochemistry; TME, tumor microenvironment.

Differential expression analysis

In this study, 179 tumor samples and 171 controls were selected from the TCGA and GTEx integrated data. A robust disease-specific model with fewer genes is more suitable for clinical translation, therefore, only those common DEGs which were identified as significant across all three methods (DESeq2, edgeR and limma R packages) were considered. A total of 3156 DEGs were identified across all three different programs, including 1985 upregulated and 1171 downregulated genes, respectively (Table S1).

Prognostic potential analysis of DEGs

We further explored the prognostic potential of these DEGs using univariate Cox regression and Kaplan-Meier survival analysis. The `surv_cox` function from `tinyarray` (v2.2.9) R package was used to calculate the Cox p value and HR of DEGs in univariate Cox regression with p value cutoff of 0.05, and 1150 DEGs with prognostic potential were identified (Table S2). The `surv_KM` function from `tinyarray` (v2.2.9) R package was used to calculate log_rank test p value for DEGs in Kaplan-Meier survival analysis with p value cutoff of 0.05, and 681 DEGs with prognostic potential were identified (Table S3).

Identification of TEX-related gene signature in PAAD

TISCH2 database was a scRNA-seq database including 190 datasets focusing on TME. The Model-based AnalysisEs of Transcriptome and RegulOme (MAESTRO) was utilized by TISCH2 to perform quality control, normalization, clustering and cell type annotation for each collected dataset [18]. A total of 1555 genes associated with CD8TEX were obtained from PAAD-CRA001160 scRNA-seq dataset in TISCH2 database (Figures 2A, B; Table S4). The intersected genes among CD8TEX genes, DEGs with prognostic potential identified by using univariate Cox regression and Kaplan-Meier survival analysis were defined as TEX-related DEGs. A total of 49 TEX-related DEGs were collected (Figure 2C, Table S5).

LASSO regression followed by stepwise variable selection process to obtain the best candidate for Cox's proportional hazards model (Figures 2D, E). The final TEX risk model was constructed using the Multivariate Cox regression analysis including six genes (SPOCK2, MT1X, LIPH, RARRES3, EMP1, and MEG3). Both univariate and multivariate Cox regression analysis confirmed the correlation between these risk genes and OS of PAAD patients (Figures 2F, G).

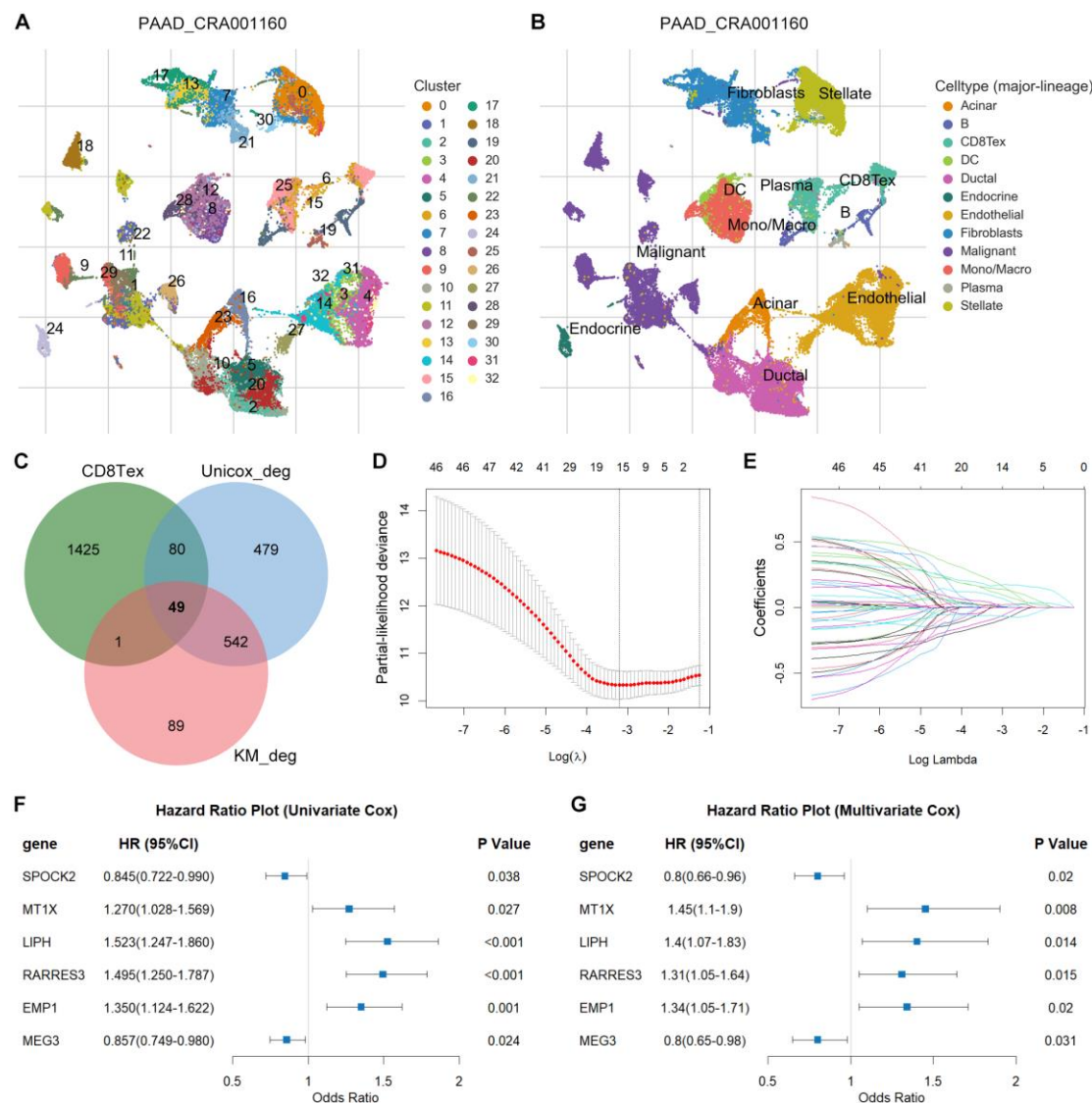


Figure 2. Identification of TEX-related gene signature in PAAD. (A, B) The UMAP plot in PAAD-CRA001160 scRNA-seq dataset from TISCH2 database. The CD8TEX was mainly consisted with cells from cluster 6 and cluster 15. (C) Intersection among CD8TEX-related genes and DEGs with prognostic values in Univariate Cox regression and Kaplan-Meier survival analysis. This intersected venn plot was generated using the jvenn online tools (<https://jvenn.toulouse.inra.fr/app/example.html>). (D, E) The LASSO coefficient profiles generated by 49 prognostic TEX-related DEGs using the survival (v3.3.1) and glmnet (v4.1-7) R packages. The log(λ).min was used to select the optimal gene sets, and a total of 14 TEX-related DEGs were selected for stepwise variable selection process to obtain the best candidate for Cox’s proportional hazards model. Six TEX-related DEGs were finally utilized by stepwise variable selection process with My.stepwise (v0.1.0) R package. (F, G) Univariate and multivariate Cox regression analysis of the six TEX-related DEGs confirmed the significant association with OS in PAAD patients using survival (v3.3.1) and survminer (v0.4.9) R packages. The forest plot was drawn with forestplot (v3.1.1) R package.

Validation of TEX-related gene signature in PAAD

The risk score was calculated with the following formula: $\text{riskscore} = -0.2248802 \times \text{SPOCK2} + 0.3692181 \times \text{MT1X} + 0.3364109 \times \text{LIPH} + 0.2723827 \times \text{RARRES3} + 0.2901199 \times \text{EMP1} - 0.2231695 \times \text{MEG3}$. The TCGA-PAAD patients were used as training dataset, and patients were divided into high- and low-risk groups based on the median risk score. SPOCK2 and MEG3 were downregulated in high-risk group, while MT1X, LIPH, and RARRES3 were upregulated in high-risk group (Figure 3A). The Kaplan-Meier survival analysis showed that the low-risk patients had longer

survival time than the high-risk patients (Figure 3B). Moreover, the ROC curve base on the TEX-related signature revealed the 1-, 3- and 5-year AUC values were 0.77, 0.78, and 0.79, respectively (Figure 3C).

We further explored whether this risk model could act as in independent prognostic variables for PAAD patients. Both univariate and multivariate Cox regression analysis confirmed that risk score and age were significantly correlated with the OS (Figures 3D, E). To make this risk model useful in the clinic, a nomogram was construct to predict 1-, 3-, and 5-year survival in the TCGA-PAAD patients. As depicted in the nomogram, the risk score model had the most weight among all these included factors including age, N stage, T stage, and grade (Figure 3F). The calibration curves showed consensus between the nomogram predicted and actual survival (Figure 3G). The DCA curves demonstrated that the nomogram performed better than other clinical factors in predicting the 1-, 3- and 5-year OS (Figure 3H).

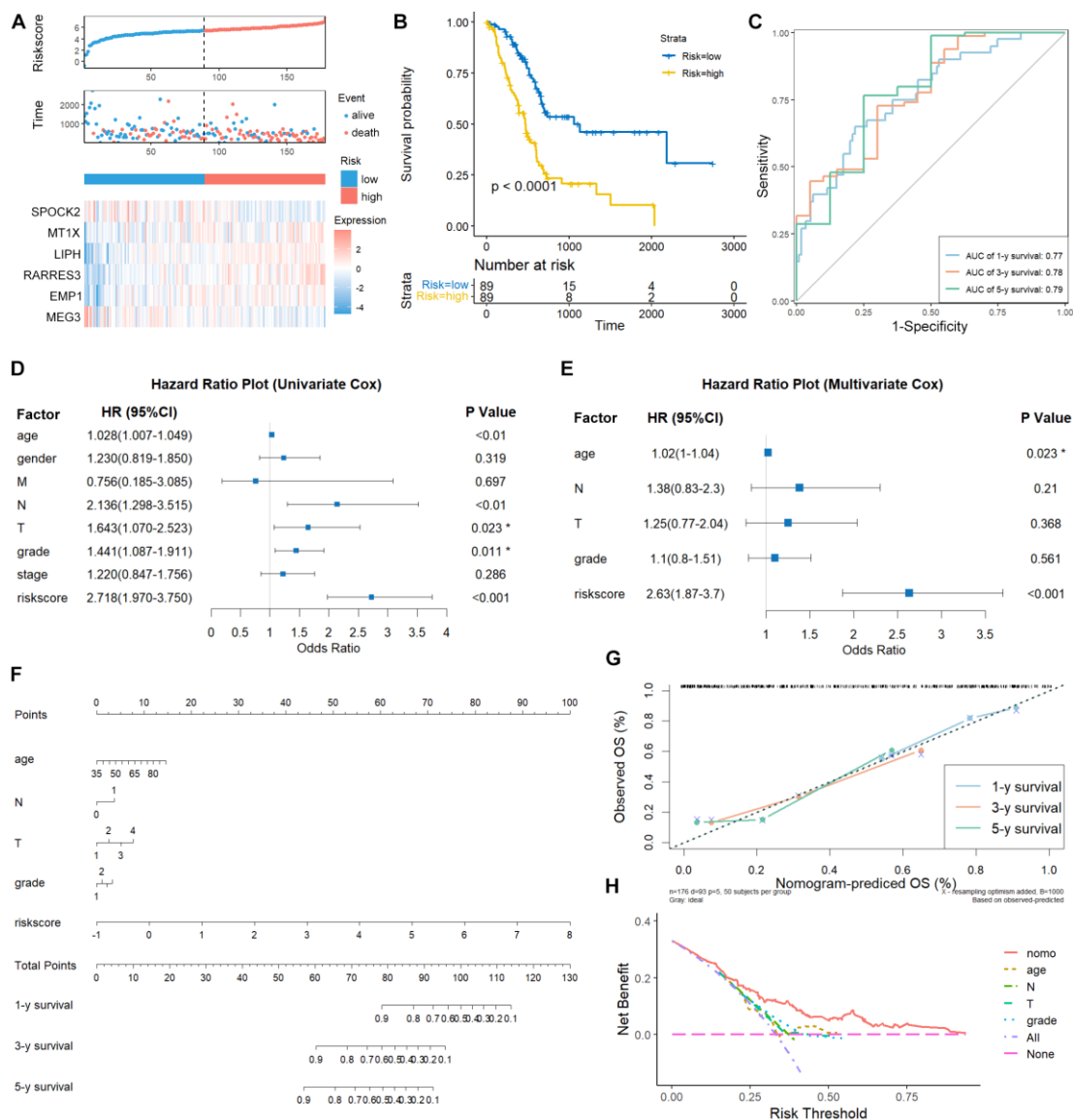


Figure 3. Validation of TEX-related gene signature in PAAD. (A) Risk score and patient distribution, and expression of risk genes in the TCGA-PAAD cohort were shown using the risk_plot function from tinyarray (v2.2.9) R package. (B) Kaplan-Meier curve of the high- and low-risk patients drawn using survival (v3.3.1) and survminer (v0.4.9) R packages. (C) The ROC curve for 1-, 3- and 5-year survival in TCGA-PAAD patients drawn using timeROC (v0.4) R package. (D, E) Univariate and multivariate Cox regression analysis of the risk score and clinical factors demonstrated the risk score was an independent prognostic factor for PAAD patients. (F) The survival (v3.3.1) and survminer (v0.4.9) R packages. (G) The ROC curve for 1-, 3- and 5-year survival in TCGA-PAAD patients drawn using timeROC (v0.4) R package. (H) Univariate and multivariate Cox regression analysis of the risk score and clinical factors demonstrated the risk score was an independent prognostic factor for PAAD patients.

(v0.4.9) R packages were used for analysis, and the forest plot was drawn with forestplot (v3.1.1) R package. (F) The nomogram plot was constructed based on clinical factors and the risk score. (G) Calibration plot of the nomogram in TCGA-PAAD patients. Both nomogram and calibration plots were drawn using rms (v6.6-0) R package. (H) DCA curve of the nomogram, age, N stage, T stage, and grade for the TCGA-PAAD patients. The DCA curve was drawn using ggDCA (v1.2) R package.

Moreover, three external validation datasets (ICGC-PACA-CA, ICGC-PAAD-US, and GSE71729, n=412) were applied to validate the prognostic potential of this TEX-related signature. Consistent with the results in the TCGA-PAAD training dataset, the low-risk patients had longer survival time and the expression pattern of risk genes were similar with the TCGA cohort (Figure 4A-F). Collectively, these results from different cohorts suggest that TEX-related gene signature may be an independent predictive indicator for PAAD patients.

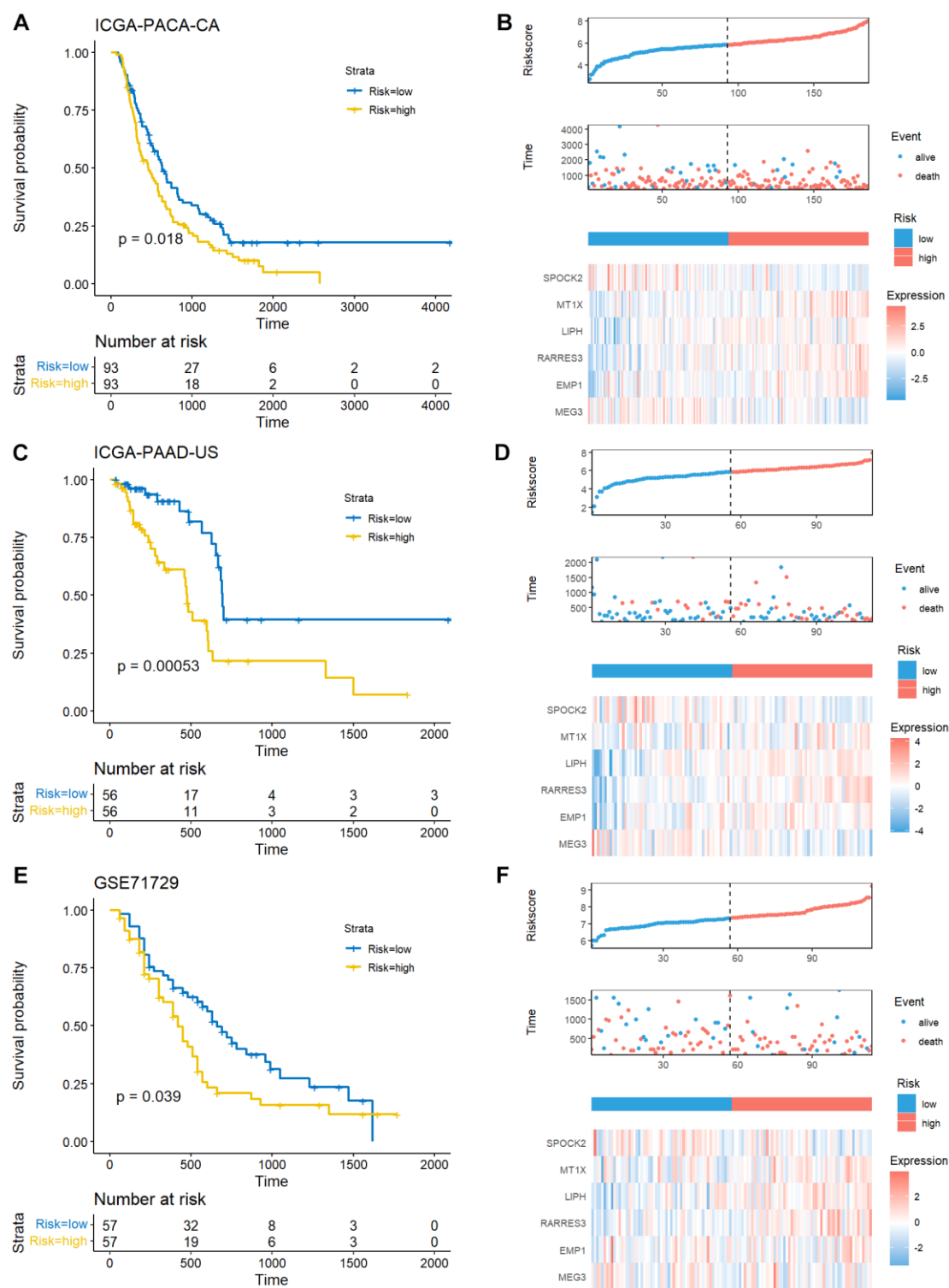


Figure 4. External validation of TEX-related gene signature in PAAD. Kaplan-Meier curve of the high- and low-risk patients, risk score and patient distribution, and expression of risk genes in ICGC-PACA-CA (A, B), ICGC-PAAD-US (C, D), and GSE71729 (E, F) cohorts. The Kaplan-Meier curves were drawn using survival (v3.3.1) and survminer (v0.4.9) R packages. The risk score and patient distribution were drawn using the risk_plot function from tinyarray (v2.2.9) R package.

Mutational landscape difference between high- and low-risk groups

We further explored the molecular characteristics between high- and low-risk patients. GSVA is an unsupervised, non-parametric algorithm for estimating variation of pathway enrichment via transforming gene expression matrix to a gene-set matrix. GSVA was used to calculate the KEGG pathway scores in each sample, and the significant differential KEGG pathways were identified

(Figure 5A, Table S6). The top four KEGG pathways enriched in high-risk group were all involved in cancer development and progression (Figure 5B). Moreover, the mutational landscape difference between high- and low-risk groups were analyzed. *KRAS*, *TP53*, *CDKN2A*, *SMAD4*, and *TTN* were the top 5 genes with mutations in both groups. The mutation proportions of *KRAS*, *TP53*, and *CDKN2A* were remarkably increased in high-risk patients (Figures 5C, D). We systematically evaluated the mutational differences between the two groups, and identified six differentially mutated genes, including *KRAS*, *TP53*, *CDKN2A*, *ADAMTS16*, *FLNA*, and *PCDH9* (Figure 5E). The top 10 GO biological process terms were mainly focused on cell growth and apoptosis (Figure 5F). These data suggest that the TEX-related signature could characterize PAAD patients with different molecular features and mutational landscapes.

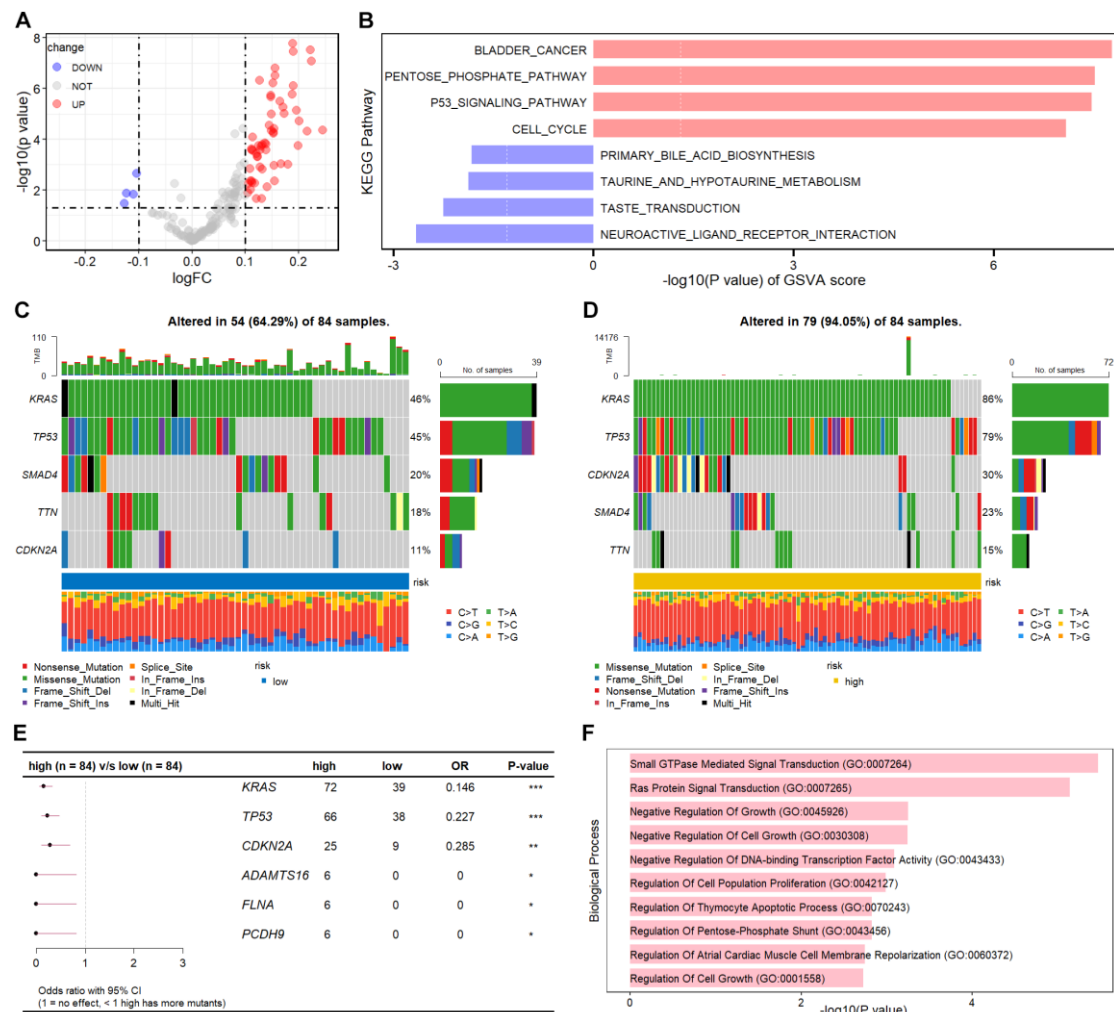


Figure 5. Mutational landscape difference between high- and low-risk groups. (A) Volcano plot of the differential KEGG pathways between high- and low-risk PAAD patients. The C2 category KEGG pathways were collected using msigdb (v7.5.1) R package. The scores of these KEGG pathways were calculated with the GSEA (v1.46.0) R package. The differential pathways were determined using limma (v3.54.2) R package, and the volcano plot was drawn using ggplot2 (v3.4.2) R package. (B) Bar plot of top 4 positively and negatively enriched KEGG pathways in high-risk patients was shown using ggplot2 (v3.4.2) R package. (C) Top 5 mutated gene ordered by mutation rate in low-risk patients. (D) Top 5 mutated gene ordered by mutation rate in high-risk patients. The mutation landscape was summarized using maftools (v2.14.0) R package, and the waterfall plot was drawn using the oncoplot function.

(E) Forest plot of significantly different mutated genes between high- and low-risk patients. The differentially mutated genes were identified through performing fisher test using the mafCompare

function from maftools R package, and the forest plot was drawn using maftools R package. (F) GO biological process enrichment analysis of significantly different mutated genes was performed using Enrichr (<https://maayanlab.cloud/Enrichr/enrichr>), and the top 10 GO terms were shown.

Immune characteristics of high- and low-risk groups

The immune cell abundance and immune function were estimated using previously published gene signatures [14]. The immune cell infiltration and immune function differences between high- and low-risk groups were compared using the Wilcoxon test. The high-risk patients tended to have more macrophages and regulatory T cells (Treg) (Figure 6A). Decreased check point, cytolytic activity, HLA, T cell co-stimulation, and type II IFN response were found in high-risk group (Figure 6B), indicating an immunosuppressive TME in high-risk patients. Additionally, the correlation between immune cell infiltration and risk score was investigated utilizing numerous algorithms. The results showed that the risk score was negatively correlated with NK and CD4⁺ T cells (Figure 6C). Moreover, the PAAD patients were further classified into C1~C6 immune subtypes. Interestingly, the C5 (Immunologically Quiet) was absent from both groups indicating the malignant feature of PAAD. The C4 (Lymphocyte Depleted), which displayed a more prominent macrophage signature and high M2 response, was only found in high-risk group, suggesting the immunosuppressive status (Figure 6D).

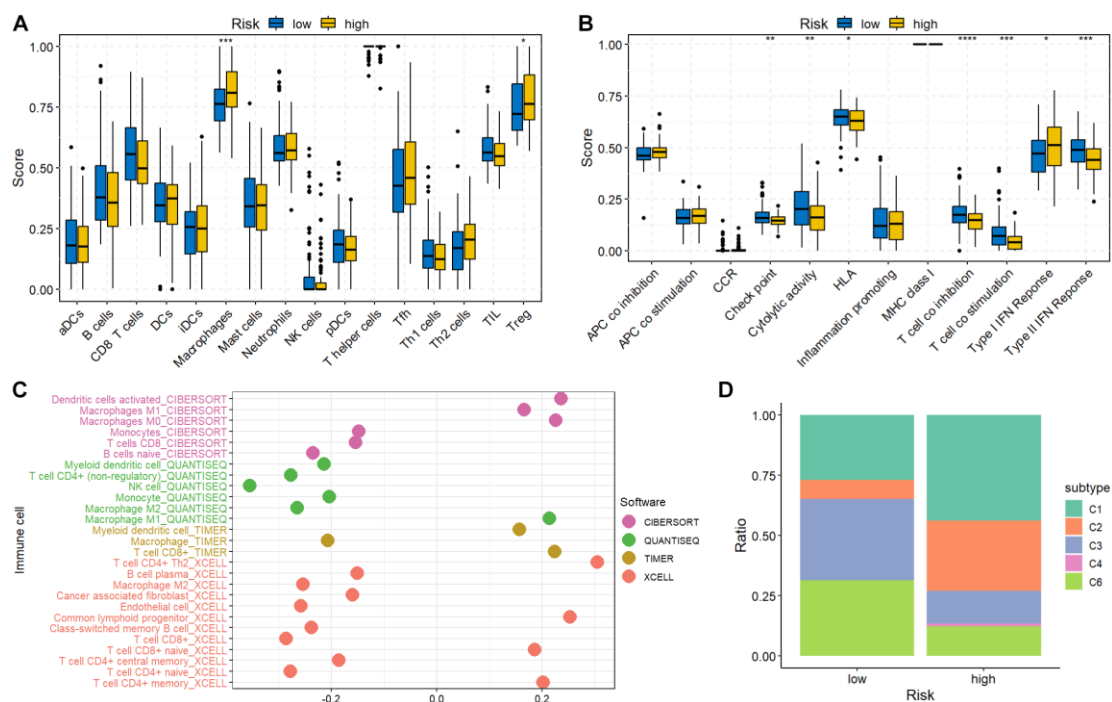


Figure 6. Immune characteristics of high- and low-risk groups. (A) The immune cell infiltration scores between high- and low-risk groups. (B) The immune function scores between high- and low-risk groups. The gene set enrichment was performed using GSVA (v1.46.0) R package, and the box plots were drawn using ggpubr (v0.6.0) R package. (C) Correlation between immune cell abundance and risk score. Immune cell infiltration was calculated using immunedeconv (v2.1.3) and CIBERSORT (v0.1.0) R packages. (D) The distribution of PAAD patients in different immune subtypes was analyzed using ImmuneSubtypeClassifier (v0.1.0) R package, and plotted using ggplot2 R package.

The correlation between drug sensitivity and risk score

The predicted IC₅₀ values of the TCGA-PAAD patients were obtained using the oncoPredict R package based on GDSC2 database. The spearman correlation analysis found that the risk scores were positively correlated with Elefantin, Sorafenib, Entinostat, Vorinostat, Oxaliplatin, while negatively correlated with Acetalax, Trametinib, Dasatinib, Sapitinib, and SCH772984 (Figure 7A).

Moreover, the positively correlated drugs showed significantly decreased IC50 in low-risk group compared with the high-risk group (Figure 7B).

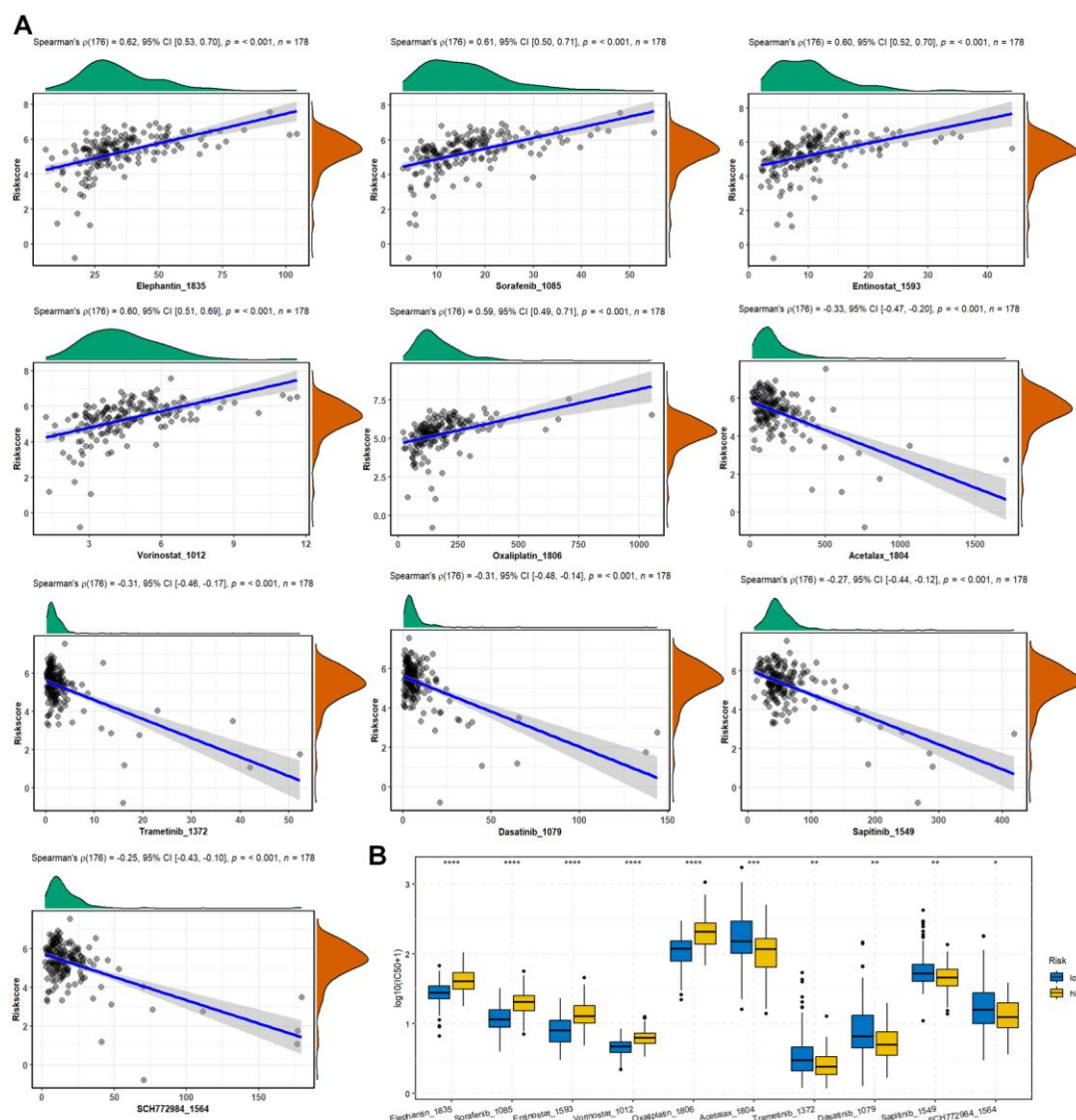


Figure 7. The correlation between drug sensitivity and risk score. (A) Spearman correlation between drug IC50 and risk score was analyzed using the corr.test function from psych (v2.3.3) R package, and the scatter plot was drawn using ggstatsplot (v0.0.6) R package. (B) Differences of drug IC50 between high- and low-risk groups were examined by the Wilcoxon test and the box plot was drawn using ggpubr (v0.6.0) R package.

Expression validation of risk genes in PAAD tissues

We further explored the expression of risk genes in PAAD tissues. The mRNA levels of these genes in TCGA-GTEx cohort showed that SPOCK2, LIPH, RARRES3, and EMP1 were significantly increased in tumor samples compared with normal tissues (Figure 8A). The Kaplan-Meier curve of the high- and low-risk patients based on the expression of risk genes in TCGA cohort found that SPOCK2 and MEG3 were positively correlated with the OS of PAAD patients, while MT1X, LIPH, RARRES3, and EMP1 were negatively correlated with the OS of PAAD patients (Figure 8B). The protein expression of several tumors from Clinical Proteomic Tumor Analysis Consortium (CPTAC) was analyzed by the The University of Alabama at Birmingham CANcer data analysis Portal (UALCAN, <https://ualcan.path.uab.edu/index.html>) [19]. The results from CPTAC database showed that the protein levels of SPOCK2 and LIPH were remarkably increased in PAAD tumor compared with normal tissues, while the protein level of MT1X was significantly decreased in PAAD tumors

(Figure 8C). In addition, our IHC staining demonstrated consistent results with CPTAC database, SPOCK2 and LIPH were increased while MT1X was decreased (Figure 8D).

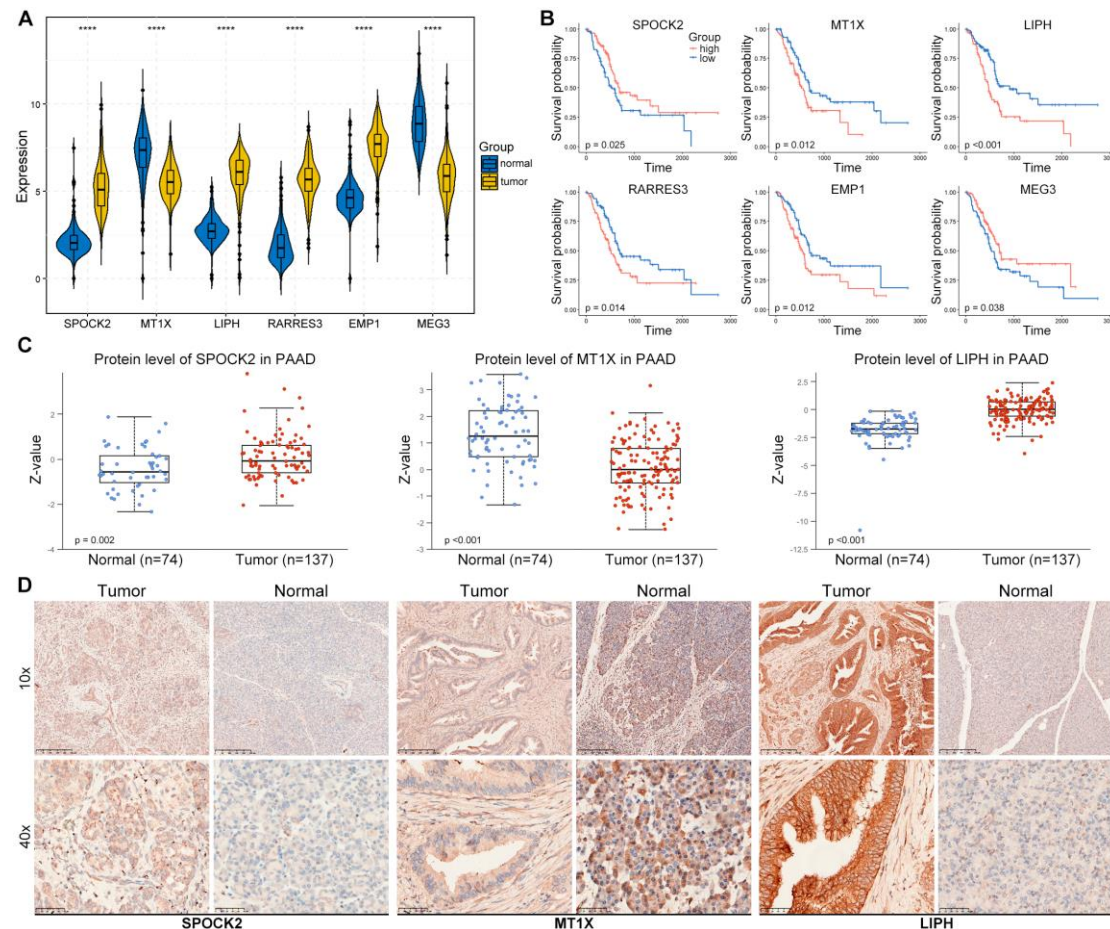


Figure 8. Expression validation of risk genes in PAAD tissues. (A) The mRNA levels of risk genes in TCGA-GTEx cohort. The differences were examined by the Wilcoxon test and the box plot was drawn using ggpubr (v0.6.0) R package. (B) The Kaplan-Meier curve of the high- and low-risk patients based on the expression of risk genes in TCGA cohort drawn using survival (v3.3.1) and survminer (v0.4.9) R packages. (C) The protein levels of SPOCK2, MT1X, and LIPH in PAAD from the CPTAC database (<https://ualcan.path.uab.edu/index.html>). (D) IHC staining of SPOCK2, MT1X, and LIPH in PAAD tumor and adjacent tissues.

SPOCK2 may be a novel CD8TEX cellular marker

ScRNA-seq has promoted the understanding of tumor heterogeneity at the single-cell level [20]. We explored the expression of the risk genes in the TME using PAAD-CRA001160 dataset from TISCH database. Interestingly, the feature plots showed that SPOCK2 was almost exclusively expressed in CD8TEX cells (Figures 2B, 9A). We further explored the expression of SPOCK2 in some tumors, including basal cell carcinoma (BCC), liver hepatocellular carcinoma (LIHC), prostate adenocarcinoma (PRAD), and squamous-cell carcinoma (SCC). Similarly, SPOCK2 was specifically and highly expressed in CD8TEX cells in BCC, LIHC, PRAD, and SCC (Figure 9B). A heatmap of SPOCK2 across numerous tumors was drawn including glioma, kidney chromophobe (KICH), non-small-cell lung carcinoma (NSCLC), and ovarian serous cystadenocarcinoma (OV). The results revealed that SPOCK2 was specifically and highly expressed in CD8TEX cells compared with B cell, dendritic cell, monocyte, macrophage, fibroblast, and malignant cells (Figures 9C, D). Collectively, these scRNA-seq results of numerous tumors strongly suggest that SPOCK2 may serve as a novel CD8TEX cellular marker.

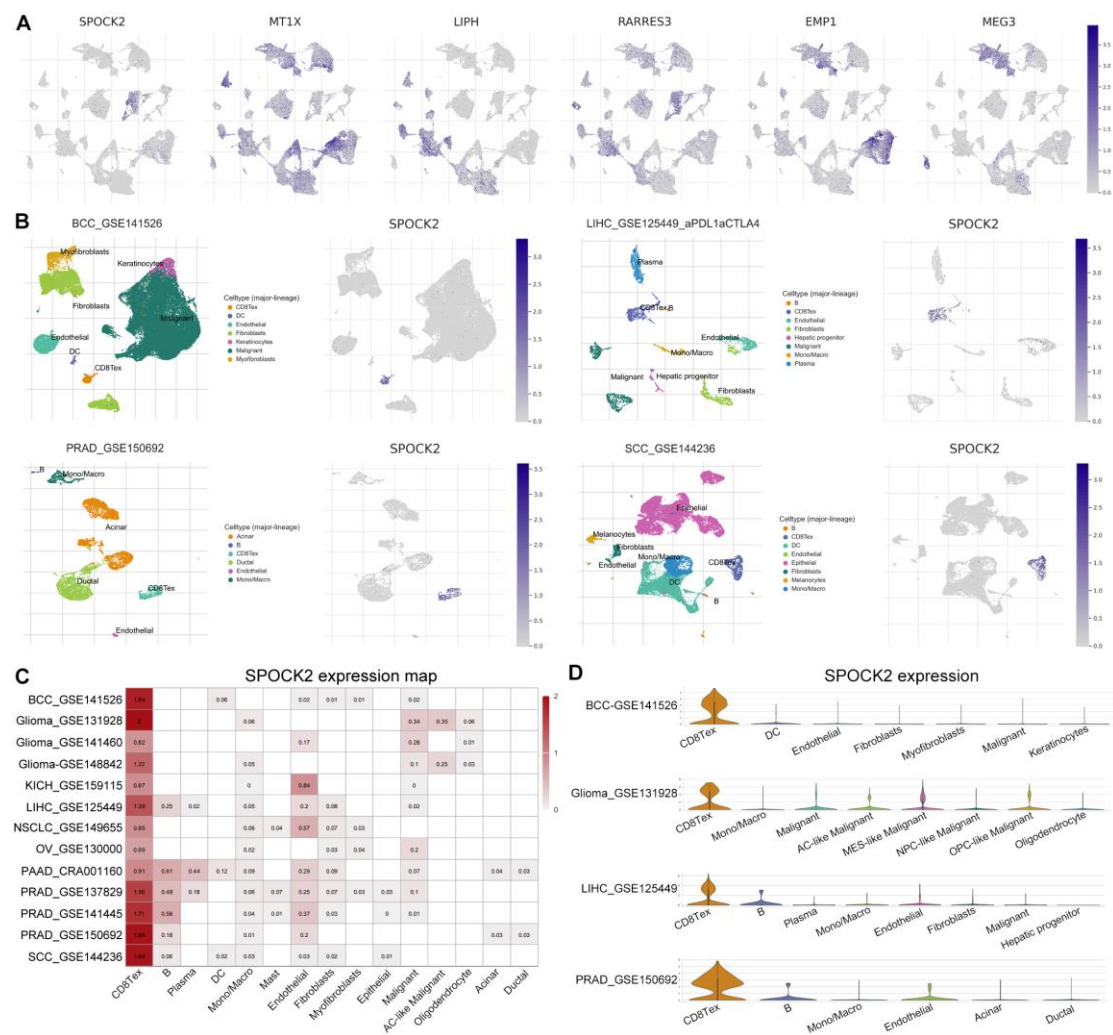


Figure 9. SPOCK2 may be a novel CD8TEX cellular marker. (A) Feature plots showed the expression of risk genes from PAAD_CRA001160 dataset related to Figure 2B. (B) Feature plots of SPOCK2 expression in several tumors. (C) Heatmap SPOCK2 expression across numerous tumors. (D) Violin plots of SPOCK2 expression in BCC, glioma, LIHC, and PRAD. BCC, basal cell carcinoma; LIHC, liver hepatocellular carcinoma; PRAD, prostate adenocarcinoma; SCC, squamous-cell carcinoma; KICH, kidney chromophobe; NSCLC, non-small-cell lung carcinoma; OV, ovarian serous cystadenocarcinoma.

Discussion

The lack of effective treatment makes PAAD one of the most lethal malignancies, although some advances have been achieved in surgical intervention, adjuvant therapy, and chemotherapy [21]. Immunotherapy provides revolutionary strategies for the treatment of cancer and has rejuvenated the tumor immunology field [22], which has been considered as the primary treatment for several tumors [23]. Several types of immunotherapy, including ICIs, T-cell transfer therapy, monoclonal antibodies, treatment vaccines, and immune system modulators, have achieved durable clinical responses [24], however, PDAC is almost unresponsive to immunotherapy due to low mutational loads, desmoplastic dense stroma, low number of tumor neoantigens, and “cold” TIME [25]. With recent advances in scRNA-seq, systematic interrogation of the PAAD TME provided novel insights into the annotation of precise cellular composition in PAAD TME [26]. TEX-related genes could be extracted from CD8TEX cells from scRNA-seq data, and TEX based gene signatures have been widely used to predict the prognostic, molecular and immune features, as well as therapy response in several tumors, including LIHC, lung adenocarcinoma, and NSCLC [27–30]. In this

study, we identified and validated the TEX-related signature for predicting prognosis and immune response in PAAD by integrated analysis of scRNA-seq and bulk RNA sequencing data.

The CD8TEX cell related genes were extracted from PAAD scRNA-seq dataset from the TISCH database, which were merged with DEGs with prognostic potential in TCGA-GTEx cohort. Subsequently, univariate COX regression, LASSO regression, and stepwise multivariate Cox regression analysis were utilized to construct a final TEX-related risk model consisting of six genes. In the training TCGA cohort, we found low-risk patients had longer survival time, and TEX-related risk model was an independent prognostic indicator with more accurate prediction of clinical outcome than other clinical factors such as age, grade, and stage. Moreover, the prognostic potential of this risk model has been further validated in three external cohorts (ICGC-PACA-CA, ICGC-PAAD-US, and GSE71729), indicating the robustness of this model.

To gain more insight into the molecular properties of TEX-related risk model, we systematically evaluated the KEGG pathways using GSVA method. Cancer related KEGG pathways were highly enriched in high-risk PAAD patients, including P53 signaling pathway and cell cycle. As the mutational load was correlated with the prognostics and immune therapy response [31], we further characterized the mutational landscapes in high- and low-risk patients. The top five mutated genes were similar between the two groups, however, the mutations in *KRAS*, *TP53*, and *CDKN2A* were remarkably more prevalent in the high-risk group than the low-risk group. Surprisingly, 86% and 79% high-risk patients harbored *KRAS* and *TP53* mutation, whose mutation rates were more than 1.7 folds of the low-risk patients. PAAD genomic studies revealed activation mutations of *KRAS* and inactivation mutations of *TP53*, *SMAD4*, and/or *CDKN2A* in >50% of cases [32]. Downstream signaling of mutant *KRAS* (m*KRAS*) orchestrates immunosuppressive TIME. m*KRAS* promoted autophagocytosis of cell surface MHC class I to prevent innate and adaptive anti-tumor immunity. m*KRAS* also promotes tumor immunoresistance by stabilizing PD-L1 mRNA [33]. Moreover, m*KRAS* signaling stimulates tumor-intrinsic CXCL1 and GM-CSF expression leading to T-cell exclusion and myeloid-derived suppressor cell (MDSC) infiltration [6]. Most mutations in tumor suppressor *TP53* was loss-of-function mutation [34], especially in the DNA-binding domain, leading to tumor-promoting properties. Cerium oxide nanoparticles targeting mutant p53 protein displayed specific therapeutic efficacy in mutant p53 cancer cells, shedding light to the therapeutic strategy for high-risk patients [35].

The TIME were also investigated to characterize their organization, immune function, and develop new treatments for PAAD. The high-risk patients showed immunosuppressive TME, characterized by increased infiltration of Treg cells and decreased antigen presentation, cytolytic activity, and check point immunological functions. Treg cells, characterized by the expression of FOXP3, are an immunosuppressive subset of CD4⁺ T cells, and play essential roles in maintaining self-tolerance and hampering antitumor immune responses [36]. Moreover, the risk score is negatively correlated with NK and CD4⁺ T cells, essential components of innate and adaptive immune cells. In addition, the high-risk patients had unique immune subtypes. The lymphocyte depleted C4 subtype characterized by high M2 response was only observed in high-risk group.

Some recent studies provide positive results of immune therapy via epigenetic-transcriptional control. Li et al. reported that KDM3A regulated the expression of EGFR to suppress anti-tumor immunity in PAAD. Genetic ablation of KDM3A reduced the infiltration of cancer-associated fibroblasts, and promoted T cell infiltration, and sensitized tumors to combination immunotherapy [37]. We found that the risk scores were positively correlated with several drugs. Moreover, the positively correlated drugs showed significantly decreased IC50 in low-risk group compared with the high-risk group, which may provide combination immunotherapy choice for further investigations. The IC50 of Sorafenib, a kinase inhibitor drug, was positively correlated with the risk score in PAAD patients. A recent meta-analysis of randomized trials of advanced hepatocellular carcinoma (HCC) patients revealed that survival outcome for Sorafenib has improved over time in advanced HCC [38]. The IC50 of histone deacetylase inhibitor Entinostat was positively correlated with the risk score in PAAD patients. Entinostat was reported to convert immune-resistant PAAD

into checkpoint-responsive tumor through reprogramming tumor-infiltrating myeloid-derived suppressor cells [39]. These data may have essential implications for PAAD clinical trial design.

In addition to the mRNA expression of the risk genes, we further validated the protein levels of these risk genes from both proteomic database and our clinical specimens. Consistent with the gene expression changes in PAAD, SPOCK2 and LIPH were increased while MT1X was decreased in PAAD.

Finally, we also investigated the expression pattern of these risk genes within the TME of PAAD, and SPOCK2 was found to be specifically and highly expressed in CD8TEX cells in numerous tumors, including BCC, glioma, KICH, LIHC, NSCLC, OV, PAAD, PRAD, and SCC. Collectively, these scRNA-seq results at single cell level suggest that SPOCK2 may serve as a novel CD8TEX cellular marker.

In conclusion, we identified and validated a TEX-related gene signature with prognostic potential and acts as an independent indicator for predicting the clinical outcome of PAAD patients. Unique molecular and immunological features had been characterized in PAAD based on the TEX-related gene signature, and high-risk patients showed worse outcome and an immunosuppressive TME. This study provides deeper insights into T cell exhaustion in PAAD.

Author contributions: X.W., and D.L. conceived the manuscript. Y.Z., and L.T. collected data. X.W. analyzed data, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding: Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology provided funding for this work (2021141 to XW).

Ethics statement: This study involving human participants was reviewed and approved by The Ethics Committee of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article. All methods were carried out in accordance with relevant guidelines and regulations.

Availability of data and materials: The TCGA and GTEx integrated data were downloaded from Xena (<https://xenabrowser.net/datapages/?cohort=TCGA%20TARGET%20GTEx&removeHub=https%3A%2F%2Fxe-na.treehouse.gi.ucsc.edu%3A443>). ICGC-PACA-CA was downloaded from ICGC (<https://dcc.icgc.org/projects/PACA-CA>). ICGC-PAAD-US was downloaded from ICGC (<https://dcc.icgc.org/projects/PAAD-US>). GSE71729 dataset was downloaded from GEO (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE71729>). Other data and results were displayed in the manuscript and supplementary data.

Acknowledgments: We thank Dr. Jianming Zeng (University of Macau), and all the members of his bioinformatics team, biotrainee, for generously sharing their experience and codes. The Use of the biorstudio high performance computing cluster (<https://biorstudio.cloud>) at Biotrainee and the shanghai HS Biotech Co., Ltd. for conducting the research reported in this paper.

Conflict of interest: The authors declare that no commercial or financial conflict of interest.

Abbreviation

PAAD	Pancreatic cancer
TIME	tumor immune microenvironment
TME	tumor microenvironment
ICB	immune checkpoint blockade
TEX	T cell exhaustion
ScRNA-seq	single-cell RNA sequencing
TISCH	the Tumor Immune Single-Cell Hub
TCGA	the Cancer Genome Atlas
GTEx	the Genotype-Tissue Expression
ICGC	the International Cancer Genome Consortium
GEO	the Gene Expression Omnibus
DEGs	differentially expressed genes
OS	overall survival

BCC	basal cell carcinoma
LIHC	liver hepatocellular carcinoma
PRAD	prostate adenocarcinoma
SCC	squamous-cell carcinoma
KICH	kidney chromophobe
NSCLC	non-small-cell lung carcinoma
OV	ovarian serous cystadenocarcinoma

References

1 Sung: H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209-249, doi:10.3322/caac.21660 (2021).

2 Ferlay, J., Partensky, C. & Bray, F. More deaths from pancreatic cancer than breast cancer in the EU by 2017. *Acta Oncol* **55**, 1158-1160, doi:10.1080/0284186X.2016.1197419 (2016).

3 Ilic, M. & Ilic, I. Epidemiology of pancreatic cancer. *World J Gastroenterol* **22**, 9694-9705, doi:10.3748/wjg.v22.i44.9694 (2016).

4 Singhi, A. D., Koay, E. J., Chari, S. T. & Maitra, A. Early Detection of Pancreatic Cancer: Opportunities and Challenges. *Gastroenterology* **156**, 2024-2040, doi:10.1053/j.gastro.2019.01.259 (2019).

5 O'Reilly, E. M. et al. Durvalumab With or Without Tremelimumab for Patients With Metastatic Pancreatic Ductal Adenocarcinoma: A Phase 2 Randomized Clinical Trial. *JAMA Oncol* **5**, 1431-1438, doi:10.1001/jamaoncol.2019.1588 (2019).

6 Bear, A. S., Vonderheide, R. H. & O'Hara, M. H. Challenges and Opportunities for Pancreatic Cancer Immunotherapy. *Cancer Cell* **38**, 788-802, doi:10.1016/j.ccell.2020.08.004 (2020).

7 Wherry, E. J. T cell exhaustion. *Nat Immunol* **12**, 492-499, doi:10.1038/ni.2035 (2011).

8 Kallies, A., Zehn, D. & Utzschneider, D. T. Precursor exhausted T cells: key to successful immunotherapy? *Nat Rev Immunol* **20**, 128-136, doi:10.1038/s41577-019-0223-7 (2020).

9 Guan, Q. et al. Strategies to reinvigorate exhausted CD8(+) T cells in tumor microenvironment. *Front Immunol* **14**, 1204363, doi:10.3389/fimmu.2023.1204363 (2023).

10 Blanche, P., Dartigues, J. F. & Jacqmin-Gadda, H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* **32**, 5381-5397, doi:10.1002/sim.5958 (2013).

11 Hanzelmann, S., Castelo, R. & Guinney, J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7, doi:10.1186/1471-2105-14-7 (2013).

12 Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* **28**, 1747-1756, doi:10.1101/gr.239244.118 (2018).

13 Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128, doi:10.1186/1471-2105-14-128 (2013).

14 He, Y., Jiang, Z., Chen, C. & Wang, X. Classification of triple-negative breast cancers based on Immunogenomic profiling. *J Exp Clin Cancer Res* **37**, 327, doi:10.1186/s13046-018-1002-1 (2018).

15 Sturm, G., Finotello, F. & List, M. Immunedeconv: An R Package for Unified Access to Computational Methods for Estimating Immune Cell Fractions from Bulk RNA-Sequencing Data. *Methods Mol Biol* **2120**, 223-232, doi:10.1007/978-1-0716-0327-7_16 (2020).

16 Thorsson, V. et al. The Immune Landscape of Cancer. *Immunity* **48**, 812-830 e814, doi:10.1016/j.immuni.2018.03.023 (2018).

17 Maeser, D., Gruener, R. F. & Huang, R. S. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform* **22**, doi:10.1093/bib/bbab260 (2021).

18 Wang, C. et al. Integrative analyses of single-cell transcriptome and regulome using MAESTRO. *Genome Biol* **21**, 198, doi:10.1186/s13059-020-02116-x (2020).

19 Zhang, Y., Chen, F., Chandrashekar, D. S., Varambally, S. & Creighton, C. J. Proteogenomic characterization of 2002 human cancers reveals pan-cancer molecular subtypes and associated pathways. *Nat Commun* **13**, 2669, doi:10.1038/s41467-022-30342-3 (2022).

20 Bai, X., Li, Y., Zeng, X., Zhao, Q. & Zhang, Z. Single-cell sequencing technology in tumor research. *Clin Chim Acta* **518**, 101-109, doi:10.1016/j.cca.2021.03.013 (2021).

21 Rawla, P., Sunkara, T. & Gaduputi, V. Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. *World J Oncol* **10**, 10-27, doi:10.14740/wjon1166 (2019).

22 Zhang, Y. & Zhang, Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cell Mol Immunol* **17**, 807-821, doi:10.1038/s41423-020-0488-6 (2020).

- 23 Cheng, A. L., Hsu, C., Chan, S. L., Choo, S. P. & Kudo, M. Challenges of combination therapy with immune checkpoint inhibitors for hepatocellular carcinoma. *J Hepatol* **72**, 307-319, doi:10.1016/j.jhep.2019.09.025 (2020).
- 24 Yang, X., Yin, R. & Xu, L. Neoadjuvant PD-1 Blockade in Resectable Lung Cancer. *N Engl J Med* **379**, e14, doi:10.1056/NEJMc1808251 (2018).
- 25 Blando, J. et al. 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**, 1692-1697, doi:10.1073/pnas.1811067116 (2019).
- 26 Moncada, R. et al. Author Correction: Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat Biotechnol* **38**, 1476, doi:10.1038/s41587-020-00776-5 (2020).
- 27 Kuang, T., Zhang, L., Chai, D., Chen, C. & Wang, W. Construction of a T-cell exhaustion-related gene signature for predicting prognosis and immune response in hepatocellular carcinoma. *Aging (Albany NY)* **15**, 5751-5774, doi:10.18632/aging.204830 (2023).
- 28 Zhu, Y. P. et al. Cuproptosis-related molecular subtypes direct T cell exhaustion phenotypes and therapeutic strategies for patients with lung adenocarcinoma. *Front Pharmacol* **14**, 1146468, doi:10.3389/fphar.2023.1146468 (2023).
- 29 Chi, H. et al. T-cell exhaustion signatures characterize the immune landscape and predict HCC prognosis via integrating single-cell RNA-seq and bulk RNA-sequencing. *Front Immunol* **14**, 1137025, doi:10.3389/fimmu.2023.1137025 (2023).
- 30 Zhao, Y. et al. Single-Cell Transcriptomics of Immune Cells Reveal Diversity and Exhaustion Signatures in Non-Small-Cell Lung Cancer. *Front Immunol* **13**, 854724, doi:10.3389/fimmu.2022.854724 (2022).
- 31 Samstein, R. M. et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* **51**, 202-206, doi:10.1038/s41588-018-0312-8 (2019).
- 32 Biankin, A. V. et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491**, 399-405, doi:10.1038/nature11547 (2012).
- 33 Coelho, M. A. et al. Oncogenic RAS Signaling Promotes Tumor Immunoresistance by Stabilizing PD-L1 mRNA. *Immunity* **47**, 1083-1099 e1086, doi:10.1016/j.immuni.2017.11.016 (2017).
- 34 Pan, M. et al. TP53 Gain-of-Function and Non-Gain-of-Function Mutations Are Associated With Differential Prognosis in Advanced Pancreatic Ductal Adenocarcinoma. *JCO Precis Oncol* **7**, e2200570, doi:10.1200/PO.22.00570 (2023).
- 35 Zhang, H. et al. Precise pancreatic cancer therapy through targeted degradation of mutant p53 protein by cerium oxide nanoparticles. *J Nanobiotechnology* **21**, 117, doi:10.1186/s12951-023-01867-6 (2023).
- 36 Togashi, Y., Shitara, K. & Nishikawa, H. Regulatory T cells in cancer immunosuppression - implications for anticancer therapy. *Nat Rev Clin Oncol* **16**, 356-371, doi:10.1038/s41571-019-0175-7 (2019).
- 37 Li, J. et al. Epigenetic and Transcriptional Control of the Epidermal Growth Factor Receptor Regulates the Tumor Immune Microenvironment in Pancreatic Cancer. *Cancer Discov* **11**, 736-753, doi:10.1158/2159-8290.CD-20-0519 (2021).
- 38 Tan, D. J. H. et al. Survival Trends in Sorafenib for Advanced Hepatocellular Carcinoma: A Reconstructed Individual Patient Data Meta-Analysis of Randomized Trials. *Liver Cancer* **12**, 445-456, doi:10.1159/000529824 (2023).
- 39 Christmas, B. J. et al. Entinostat Converts Immune-Resistant Breast and Pancreatic Cancers into Checkpoint-Responsive Tumors by Reprogramming Tumor-Infiltrating MDSCs. *Cancer Immunol Res* **6**, 1561-1577, doi:10.1158/2326-6066.CIR-18-0070 (2018).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.