Machine learning algorithm integrates bulk and single-cell transcriptome sequencing to reveal immune-related personalized therapy prediction features for pancreatic cancer

Longjun Zang[1,*], Baoming Zhang[1,*], Yanling Zhou[2], Fusheng Zhang[3], Xiaodong Tian[3], Zhongming Tian[1], Dongjie Chen[4,5,6], Qingwang Miao[1]

[1]Department of General Surgery, Taiyuan Central Hospital, Taiyuan 030009, Shanxi, P.R. China
[2]University of Shanghai for Science and Technology, Shanghai 200093, P.R. China
[3]Department of General Surgery, Peking University First Hospital, Beijing 100034, P.R. China
[4]Department of General Surgery, Pancreatic Disease Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, P.R. China
[5]Research Institute of Pancreatic Diseases, Shanghai Key Laboratory of Translational Research for Pancreatic Neoplasms, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, P.R. China
[6]State Key Laboratory of Oncogenes and Related Genes, Institute of Translational Medicine, Shanghai Jiao Tong University, Shanghai 200025, P.R. China
*Equal contribution and share first authorship

Correspondence to: Zhongming Tian, Dongjie Chen, Qingwang Miao; **email:** 55124823@qq.com, https://orcid.org/0009-0007-1586-1953; chen_dj@sjtu.edu.cn, https://orcid.org/0000-0001-9992-5019; Drmqw1971@163.com, https://orcid.org/0009-0005-3120-1694

ABSTRACT

Pancreatic cancer (PC) is a digestive malignancy with worse overall survival. Tumor immune environment (TIME) alters the progression and proliferation of various solid tumors. Hence, we aimed to detect the TIME-related classifier to facilitate the personalized treatment of PC. Based on the 1612 immune-related genes (IRGs), we classified patients into Immune_rich and Immune_desert subgroups via consensus clustering. Patients in distinct subtypes exhibited a difference in sensitivity to immune checkpoint blockers (ICB). Next, the immune-related signature (IRS) model was established based on 8 IRGs (SYT12, TNNT1, TRIM46, SMPD3, ANLN, AFF3, CXCL9 and RP1L1) and validated its predictive efficiency in multiple cohorts. RT-qPCR experiments demonstrated the differential expression of 8 IRGs between tumor and normal cell lines. Patients who gained lower IRS score tended to be more sensitive to chemotherapy and immunotherapy, and obtained better overall survival compared to those with higher IRS scores. Moreover, scRNA-seq analysis revealed that fibroblast and ductal cells might affect malignant tumor cells via MIF-(CD74+CD44) and SPP1-CD44 axis. Eventually, we identified eight therapeutic targets and one agent for IRS high patients. Our study screened out the specific regulation pattern of TIME in PC, and shed light on the precise treatment of PC.

INTRODUCTION

As a malignant tumor of the digestive system, pancreatic cancer (PC) poses a serious challenge to human health with an extremely low five-year survival rate. In the past 30 years, the incidence of pancreatic cancer has steadily increased worldwide [1]. In addition, it is the fourth leading cause of cancer death

among men and women of all ages in the United States [2]. Among the traditional treatment modalities including surgical resection and radiotherapy, early surgical resection of pancreatic cancer is considered to be the only possible cure for the malignancy [3]. Noteworthy, only 20% of patients diagnosed with pancreatic cancer can be treated surgically, and even after surgery, most patients will recur and eventually have a seriously poor prognosis. Unfortunately, radiotherapy and chemotherapy for PC also provide limited benefits to patients [4]. Interestingly, advances in immunotherapies, especially immune checkpoint blockade (ICB), have broadened therapy strategies for some historically chemotherapy-refractory malignancies and brought new hope to oncology patients [5]. However, in terms of PC, it has been significantly refractory to ICB therapy. In single-agent ICB and dual-agent ICB studies with anti-PD-1 and anti-cytotoxic t lymphocyte-4 antibodies, overall response rates (ORLs) were 0% and 3%, respectively [6, 7]. These disappointing results (in contrast to the remarkable efficacy of ICBs in other solid tumors) have driven the identification and development of novel immune-related markers in PC that may be key to unlocking immunotherapy as a viable treatment option for pancreatic cancer. Therefore, exploring novel prognostic signatures and drug screening based on the immune level is urgently necessary for delaying the occurrence and development of PC.

Tumor immune microenvironment (TIME) is an indispensable part of tumor progression by providing sufficient nutrients for tumor cell growth and development. The in-depth study of the nature of TIME in the complex evolution of cancer led to a shift from a tumor cell-centered view of cancer development to the concept of a complex tumor ecosystem that supports tumor growth and metastatic spread [8, 9]. The composition of heterogeneous TIME is extremely complex and contains a variety of immunosuppressive cells, including tumor cells, cancer-associated fibroblasts (CAFs), vascular endothelial cells, inhibitory myeloid cells, regulatory T cells (Tregs), and regulatory B cells [10]. These cells and cancer cells can secrete extracellular components, such as extracellular matrix (ECM), matrix metalloproteinase (MMP), growth factors, and transforming growth factor-β (TGFβ), to maintain or disrupt the dynamic equilibrium of the microenvironment and ultimately affect tumor progression [11]. It has been known that these tumor-associated immune cells may possess tumor-antagonizing or tumor-promoting functions. Numerous studies have indicated that the microenvironment plays a vital role in PC progression [12]. Two major features of the pancreatic cancer microenvironment, dense desmoplasia and extensive immunosuppression,

facilitated PC cell proliferation and mediated the immune escape via inhibiting the anti-tumor immunity or induction of the proliferation of immunosuppressive cells. Given the temporal heterogeneity, the application of ICB may not be sufficient to maximize the benefit of immunotherapy in PC, and the use of tumor biomarkers involved in maintaining the immunosuppressive microenvironment should also be considered for better outcomes and safety. Hence, it's necessary for us to explore distinct TIME-related features to guide clinical practice.

In this study, we aimed to explore the immune characteristics of TIME in order to inform the personalized and precise treatment of PC. We identified the immune-related dysregulated genes and constructed the TIME subtype. Additionally, we utilized multiple machine learning algorithms to construct an immune-related signature (IRS) to characterize the relationship between infiltration of immune cells and TIME subtypes and to validate the predictive efficacy of IRS on PC survival outcomes in different cohorts. In fact, we evaluated the sensitivity of chemotherapy and immunotherapy between IRS_high and _low subgroups, and explored the underlying mechanism of how IRS contributes to TIME in PC was also explored based on the results of single-cell sequencing analysis. Eventually, pharmacogenomic datasets are employed to identify potential drug targets and agents and inform immune personalized therapy for pancreatic cancer.

RESULTS

### Immune-related differential expression genes in pancreatic cancer

PC is known as the "immune desert" due to its unique TIME characteristics. The abundant bone marrow-derived cells and Treg cells in PC can mediate tumor immune escape and cause different levels of immunotherapy resistance through different mechanisms. To further explore the characteristics of TIME in PC, based on meta-cohort, Estimation of STromal and Immune cells in MAlignant Tumours using Expression data (ESTIMATE) algorithm was conducted to calculate the immune and stromal scores of each PC patient (Figure 1A), and the result revealed a high level of infiltrating stromal cells in PC. Subsequently, Differential gene expression analysis in the immune high/low group, stromal high/low group and tumor/normal group suggested (Figure 1B) that 1238 up-regulated and 1824 down-regulated genes were screened compared to the immune low group (Supplementary Table 1). Meanwhile, there were 1919 upregulated and 2324 downregulated genes compared to the stromal-low group (Supplementary Table 2). Additionally, we detected

3731 upregulated and 3463 downregulated genes between pancreatic cancer samples and normal pancreatic samples (Supplementary Table 3). After converging all these three types of differential expression genes (DEGs), 1612 immune-related genes (IRGs) were identified for the following study (Supplementary Table 4). Functional enrichment analysis suggested the function of IRG with potential tumor regulatory mechanisms, and the results suggested IRGs mainly enriched in adaptive immune response and regulation of T cell activation of Gene Ontology (GO) terms (Figure 1C), immune cell receptor signaling and antigen binding pathways of Kyoto Encyclopedia of Genes and Genomes (KEGG) terms (Figure 1D). All these results suggest that PC progression may be related to immune response in the tumor microenvironment to varying degrees.

Generation of TIME subtype

Emerging evidence demonstrates that specific expression patterns of TIME could influence the clinical treatment strategies for PC. Hence, we separated PC patients into two clusters (Figure 2A and Supplementary Figure 1A), namely Cluster_1 (n=145, 51.06%) and Cluster_2 (n=139, 48.94%), by an unsupervised consensus clustering algorithm and according to the expression level of 1612 IRGs (optimal cutoff k=2). Interestingly, PCA analysis showed that there are significant differences between these two clusters (Supplementary Figure 1B).

To further identify the correlation between the regulation of immune cells and clusters, Tumor Immune
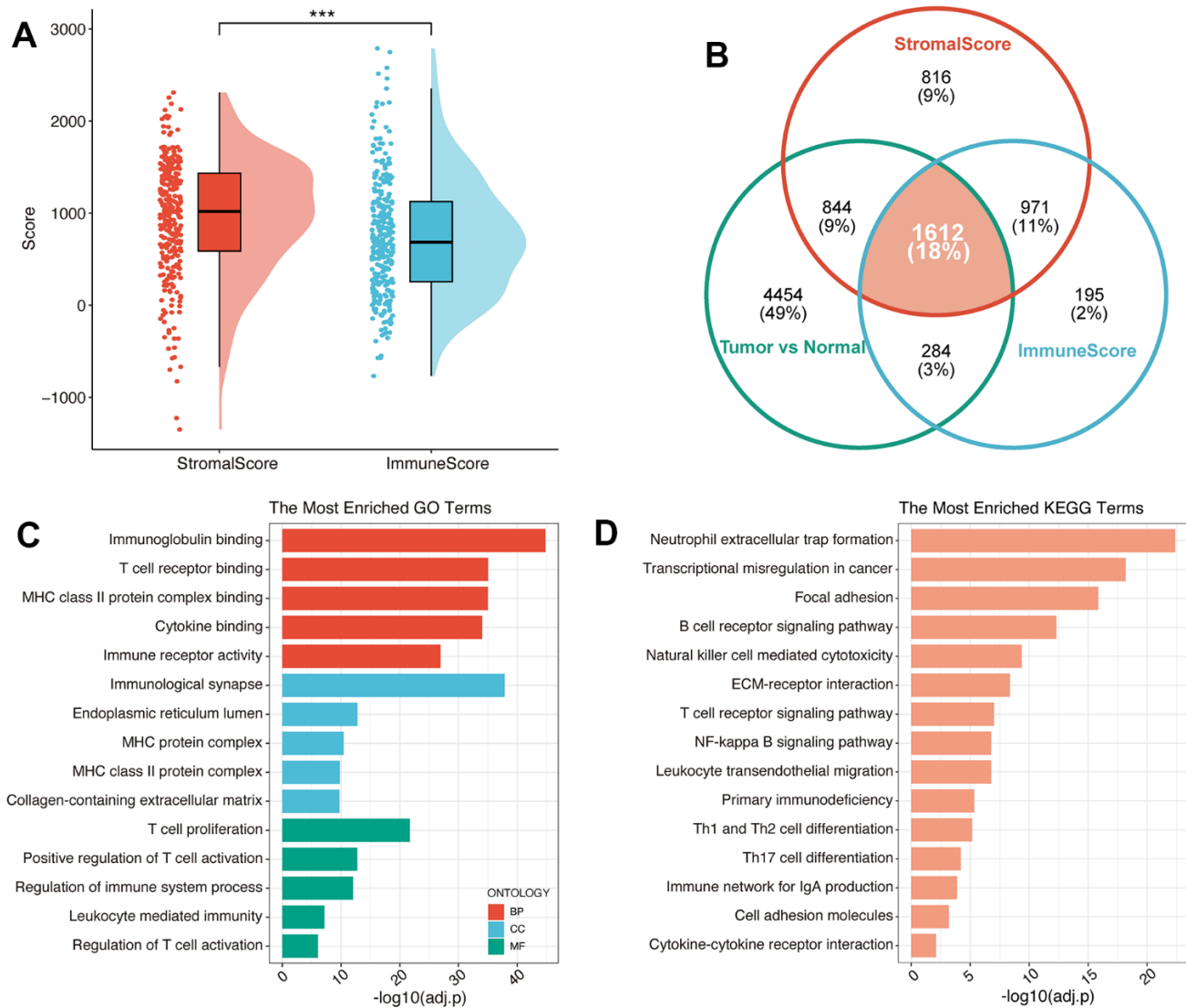
Figure 1. Identification of immune-related genes in PC. (**A**) The results of ESTIMATE on the meta-cohort. (**B**) Venn plot exhibited the converged IRGs. GO (**C**) and KEGG (**D**) analysis of IRGs.

Estimation Resource (TIMER) algorithm was applied to evaluate the abundance of immune cells. As illustrated in Figure 2B, Cluster_2 displayed significantly higher infiltration of immune cells (B cell, CD4+ T cell, CD8+ T cell, neutrophil and myeloid dendritic cell) compared with Cluster_1. Moreover, we performed Tumor Immune Dysfunction and Exclusion (TIDE) algorithm to predict the sensitivity of response to immune checkpoint blockade, including anti-PD1 and anti-CTLA4. Patients in Cluster_2 tend to obtain lower TIDE scores, which means patients in Cluster_2 were sensitive to anti-ICB therapy (Figure 2C). Similarly, we also accessed the diversity in the expression of ICB between Cluster_1 and Cluster_2. Results showed that the expression level of ICB (PDCD1, CD274,

HAVCR2, LAG3, TIGIT and GTLA4) in Cluster_2 was obviously upregulated compared to Cluster_1, suggesting that patients in Cluster_2 were more likely to be targeted (Figure 2D). Therefore, regarding the characteristics between those two clusters mentioned above, we manually defined the Cluster_1 as Immune_desert subtype, and Cluster_2 as Immune_rich subtype. ssGSEA analysis also confirmed that the Immune_rich subtype possessed a significant level of innate and adaptive immune cells, including natural killer cells, immature B cells and T cells (all $p < 0.0001$, Figure 2E). Of note, tumor-suppressing Th1 cells were considerably enriched in the Immune_rich subtype ($p = 2.96e-32$) compared to tumor-promoting Th2 cells ($p = 0.256$).
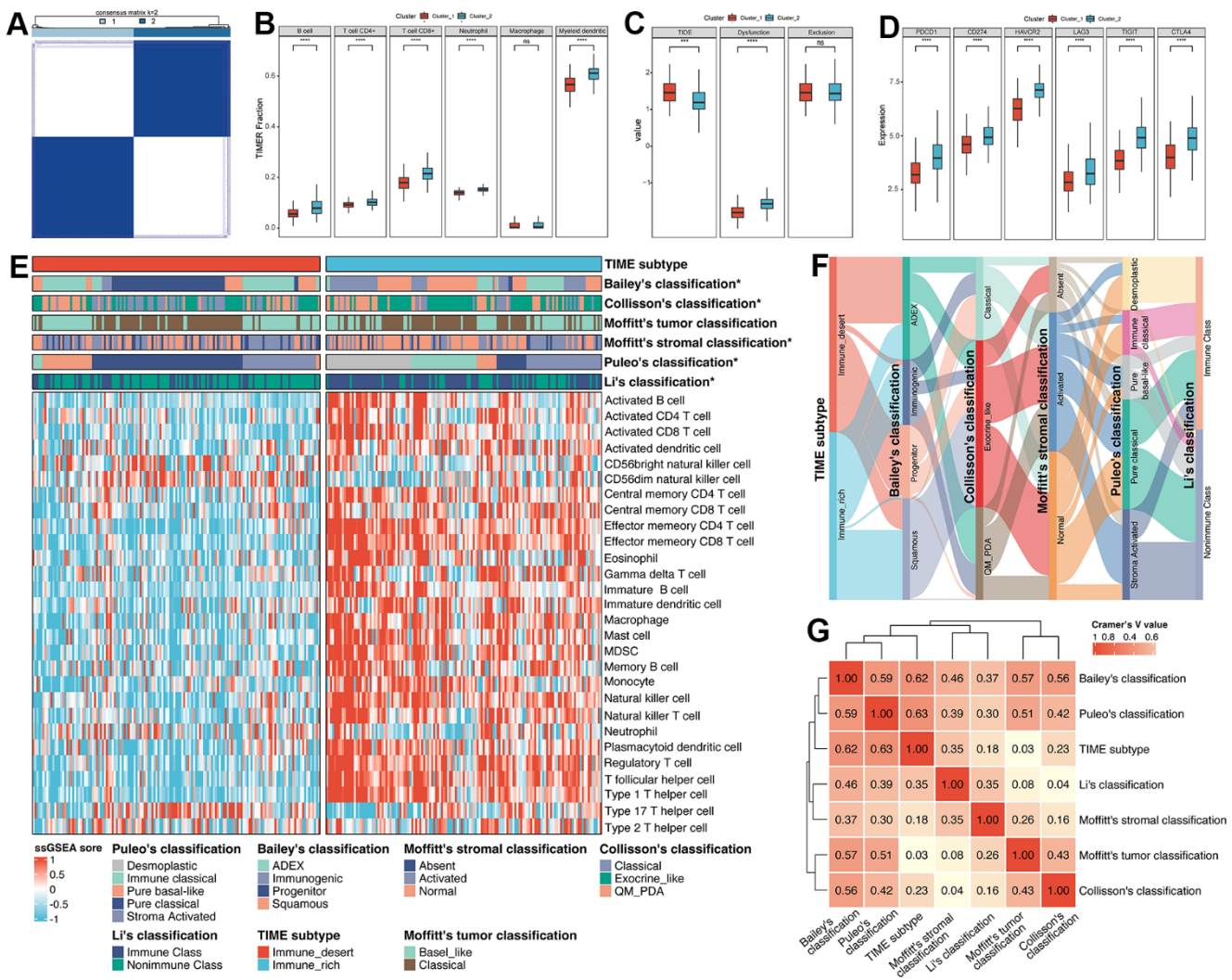
Figure 2. Exploration of the relationships between the regulation of immune cells and clusters. (**A**) Unsupervised consensus clustering based on 1612 IRGs. (**B**) The fractions of immune cells between Cluster_1 and Cluster_2. (**C**) The TIDE score between Cluster_1 and Cluster_2. (**D**) The differences in expression of common ICBs among distinct clusters. (**E**) ssGSEA analysis was utilized to estimate the abundance of immune cells. (**F**) The alluvial plot displayed the relationship between the TIME subtype and other molecular classifications. (**G**) Heatmap of Cramer's V statistic reflected the corrections between seven PC molecular classifications.

Then, we compared the identified TIME subtype with classical molecular classifications in PC. The marker of Bailey's classification, Collisson's classification, Moffitt's tumor classification, Moffitt's stromal classification and Li's classification were utilized to cluster PC patients in the meta-cohort (Supplementary Figure 2A–2E and Supplementary Table 5), and Puleo's classification was predicted followed the pipeline in Materials and Methods (Supplementary Table 5). Results illustrated that there was no significant difference between Moffitt's tumor classification and TIME subtype (p = 0.70, Supplementary Table 6), while Bailey's classification (p < 0.0001), Collisson's classification (p = 0.0006), Moffitt's stromal classification (p = 0.0101), Puleo's classification (p < 0.0001) and Li's classification (p < 0.0001) obtained significant similarity (Supplementary Table 6). For the comparison of Bailey's classification, results showed that the proportion of immunogenic subtype was higher and the percentage of progenitor subtypes was lower in Immune_rich subtype versus Immune_desert subtype (35.25% vs 3.45%, 1.44% vs 40.69%, p < 0.0001, Supplementary Table 6). Interestingly, Collisson's classification, we observed that Immune_rich subtype was composed of a more exocrine-like subtype and a less classical subtype compared to Immune_desert subtype (56.12% vs 41.38%, 14.39% vs 33.79%, p < 0.01, Supplementary Table 6) of Puleo's classification. For Moffitt's stromal classification, results demonstrated that Immune_rich subtype possessed a more normal subtype and less activated subtype than Immune_desert subtype (51.80% vs 35.17%, 36.69% vs 44.14%, p < 0.05, Supplementary Table 6). With respect to Puleo's classification, the frequency of desmoplastic and immune classical was higher within Immune_rich subtype (30.94% vs 0.69%, 23.74% vs 2.76%, p < 0.0001, Supplementary Table 6). On the contrary, we also found a lower frequency of pure basal-like and pure classical subtypes in Immune_rich subtype versus Immune_desert subtype (7.19% vs 18.62%, 10.79% vs 52.41%, p < 0.0001, Supplementary Table 6). In terms of the integration of Li's classification, we observed that Immune_rich subtype had a positive tendency to enrich in immune class and a negative correlation with nonimmune class, compared to Immune_desert subtype (68.35% vs 33.1%, 31.65% vs 66.90%, p < 0.0001, Supplementary Table 6). Moreover, the correlation between TIME subtype and other published molecular subtypes was quantified by Cramer's V (Figure 2F, 2G). Results revealed that TIME subtype had the highest correlation with Puleo's classification (Cramer's V value = 0.63) and the lowest relationship with Moffitt's tumor classification (Cramer's V value = 0.03), probably owing to the deconvolution algorithm applied on tumor cells by Moffitt et al. Additionally, after integrating the TIME

subtype and Puleo's classification, we found that patients with Immune_rich and immune classical subtypes obtained the best survival, while the patients with Immune_desert and pure basal-like subtype had the worst survival (only one patient with Immune_desert and desmoplastic subtype was excluded) (p < 0.0001, Supplementary Figure 3), implying that combination of TIME subtype and Puleo's classification may guide the prognostic prediction of PC.

Recognization of key IRGs and construction of IRS for the prognostic prediction of PC

In order to quantize the distinct characteristics among Immune_rich and Immune_desert subtypes, we applied multiple machine-learning algorithms to construct the prognostic signature based on 1612 IRGs. Before proceeding, a filtering procedure was applied to remove genes with low variability and the mean and variance of each gene were standardized to zero and one, respectively. A total of 284 patients in meta-cohort were divided into training set (n=200) and testing set (n=84) at the ratio of 7:3. Robust prognostic IRGs in PC samples were identified using multi-step processes. First, preliminary screening was performed to include 337 prognosis-related IRGs in meta-cohort via univariate Cox regression analysis. Next, bootstrapping method was used to test the genes which passed initial filtering for robustness. We extracted 70% of samples randomly from the training set and performed univariate Cox regression analysis on these samples to assess the correlation between the gene expression and prognosis. This procedure was repeated 1000 times and the 52 IRGs that were incorporated in 90% of resample runs (achieved P < 0.05 in robustness testing) were kept for next step analysis. Then, the random survival forest (RSF) analysis was independently repeated 1000 times, and 8 IRGs with the largest concordance index (C-index) were considered IRS, namely SYT12, TNNT1, TRIM46, SMPD3, ANLN, AFF3, CXCL9 and RP1L1 (Figure 3A, 3B). A risk prediction score model was then developed by these 8 genes using multivariate Cox regression, and the IRS score for each patient was determined by taking the sum of the regression coefficient for each gene multiplied by its corresponding expression value. The IRS score was then normalized from 0 to 1. According to the optimal cutoff value, PC patients were divided into IRS_high and IRS_low subgroups.

To validate the prognostic efficiency of IRS, the survival analysis was performed on the training set, testing set, meta-cohort, International Cancer Genome Consortium (ICGC), Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) datasets,

respectively. In the 6 internal and external datasets, Kaplan–Meier (KM) curves revealed that the IRS performed well in distinguishing patients with different prognostic statuses (Figure 3C–3H). Also, the univariate Cox analysis showed that SYT12, TNNT1, ANLN, CXCL9 and RP1L1 were risk factors with HR > 1, while TRIM46, SMPD3 and AFF3 were protective factors with HR < 1 (Supplementary Figure 4A),

meanwhile, the survival analysis also demonstrated this result (Supplementary Figure 4B–4I). In addition, the receiver operating characteristic (ROC) curve was utilized to verify the prediction ability of the IRS. As shown in Figure 3I, 3J, the IRS model was confirmed effective in predicting the survival of PC patients in 1 year (training set, AUC = 0.832; testing set, AUC = 0.617), 2 years (training set, AUC = 0.804; testing set,
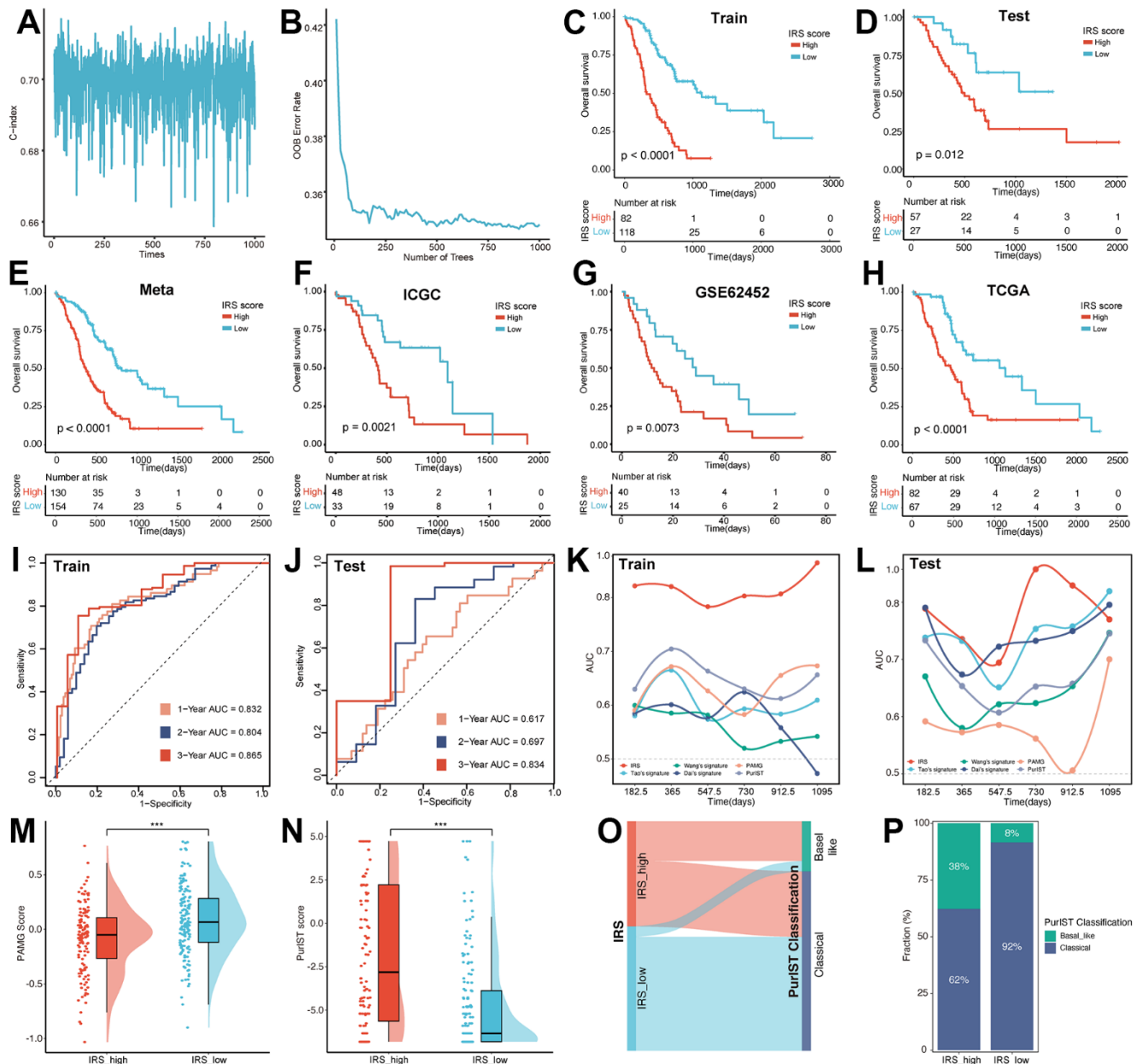
Figure 3. Construction of the IRS. The alteration of C index (**A**) and OOB rate (**B**) during 1000-times iteration. The survival analysis was performed based on the training set (**C**), testing set (**D**), meta-cohort (**E**), ICGC (**F**), GEO (**G**) and TCGA (**H**) datasets, respectively. The predictive efficiency of IRS was validated in training set (**I**), testing set (**J**). AUC value was used for the comparison of IRS with other five signatures in the training (**K**) and testing set (**L**). The distribution of PAMG score (**M**) and PurIST score (**N**) between IRS_high and IRS_low group. (**O**) Sankey plot illustrated the distribution of IRS and PurIST classification. (**P**) Distribution of PurIST classification was compared between IRS_high and IRS_low group.

AUC = 0.697) and 3 years (training set, AUC = 0.865; testing set, AUC = 0.834).

Previous studies have established several prognostic signatures for PC patients, including Wang's signature, Tao's signature, Dai's signature, pancreatic adenocarcinoma molecular gradient (PAMG) signature and PurIST signature. ROC analysis was performed to confirm whether IRS possessed superior survival prediction ability in PC compared to the five signatures mentioned above. The AUC of the IRS were higher than those of the other three prognostic models in the training set (Figure 3K). Notably, in the testing set, the predictive efficiency was far from satisfactory in 3-year survival, possibly due to the limited number of patients (Figure 3L). To further compare IRS with PAMG and PurIST classification, the PAMG score and PurIST were computed. Results illustrated that the IRS_high subgroup possessed a lower PAMG score and higher

PurIST score than the IRS_low subgroup (Figure 3M, 3N). Coincidentally, the Sankey diagram (Figure 3O) and distribution plot (Figure 3P) revealed that the percentage of the Basal-like subtype was significantly lower, and the proportion of Classical subtype was higher within IRS_low subgroup versus IRS_high subgroup (8.44% vs 37.69%, 91.56% vs 62.31%, p < 0.001). The above results fully verified the robustness and predictive effectiveness of our IRS.

Analysis and validation of differential expression for IRS

As mentioned above, a total of eight genes were selected to construct the IRS based on machine-learning algorithm. Then, we distinguished the aberrant expression of these IRGs in PC and normal pancreatic samples. As illustrated in Figure 4A, all of these IRGs were upregulated in PC samples. qRT-PCR was also
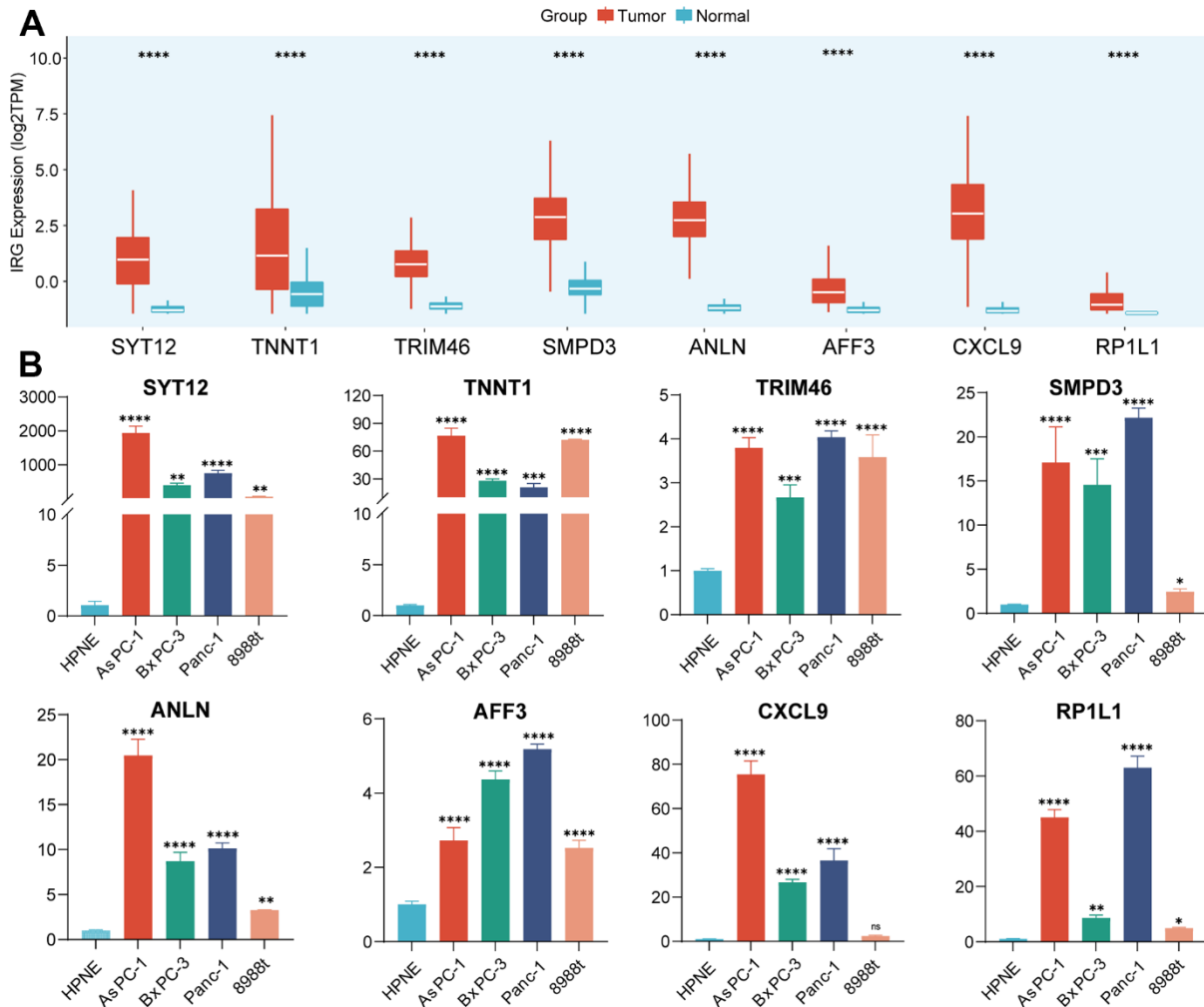

Figure 4. Validation of the expression of IRGs. (**A**) The expression of 8 IRGs genes in PC and normal pancreatic tissues. (**B**) RT-qPCR was conducted to validate the expression of 8 IRGs.

performed to validate the differential expression patterns of IRGs between normal pancreatic cell line (hTERT-HPNE) and 4 PC cell lines (AsPC-1, BxPC-3, PANC-1 and PaTu 8988t) (Figure 4B), results suggested that the expression of these eight genes was higher in all four types of pancreatic cancer cells than in normal pancreatic cells. Owing to the significant upregulation of these hub genes, they may serve as potential targets of PC which suggested further research.

Exploration of IRS-based chemotherapy prediction and potential immunotherapeutic response

As the IRS was established based on prognostic IRGs, we first analyzed the relationship between the IRS score and the infiltration of immune cells. The IRS score was positively correlated with neutrophils, myeloid-derived suppressor cells (MDSCs) and M2 macrophages. On the contrary, CD8+ T cells and CD4+ T cells displayed a negative relationship with the IRS score (Figure 5A, 5B). Moreover, we suggested that the IRS may predict the sensitivity of chemotherapy by comparing the IC50

of multiple chemical compounds between IRS_high and _low groups. As shown in Figure 5C, patients who gained a lower IRS score tended to be more sensitive to chemotherapy. In terms of the prediction value of IRS in the treatment of ICB, we calculated the immunophenoscore (IPS), IPS-CTLA4, IPS-PD1 and IPS-PD1-CTLA4 scores, which are quantitative indexes to access the treatment of ICBs, were higher in the IRS_high group (Figure 5D). Furthermore, we compared the distribution of Tumor microenvironment (TME) and tumor mutation burden (TMB) scores in IRS_high and _low subgroups to evaluate whether the IRS could predict the clinical response to ICB therapy. Results exhibited that the TME and TMB scores were higher in the IRS_high subgroup and both had a positive correlation with the IRS score (Figure 5E, 5F). Since the IRS had a remarkable correlation with the TIME of PC mentioned above, we further determined whether the IRS could predict immunotherapeutic response in PC via SubMap analysis. We evaluated the similarity of the expression module of immune-related gene expression profiles between our cohorts and a cohort of 32 melanoma patients receiving ICB therapy
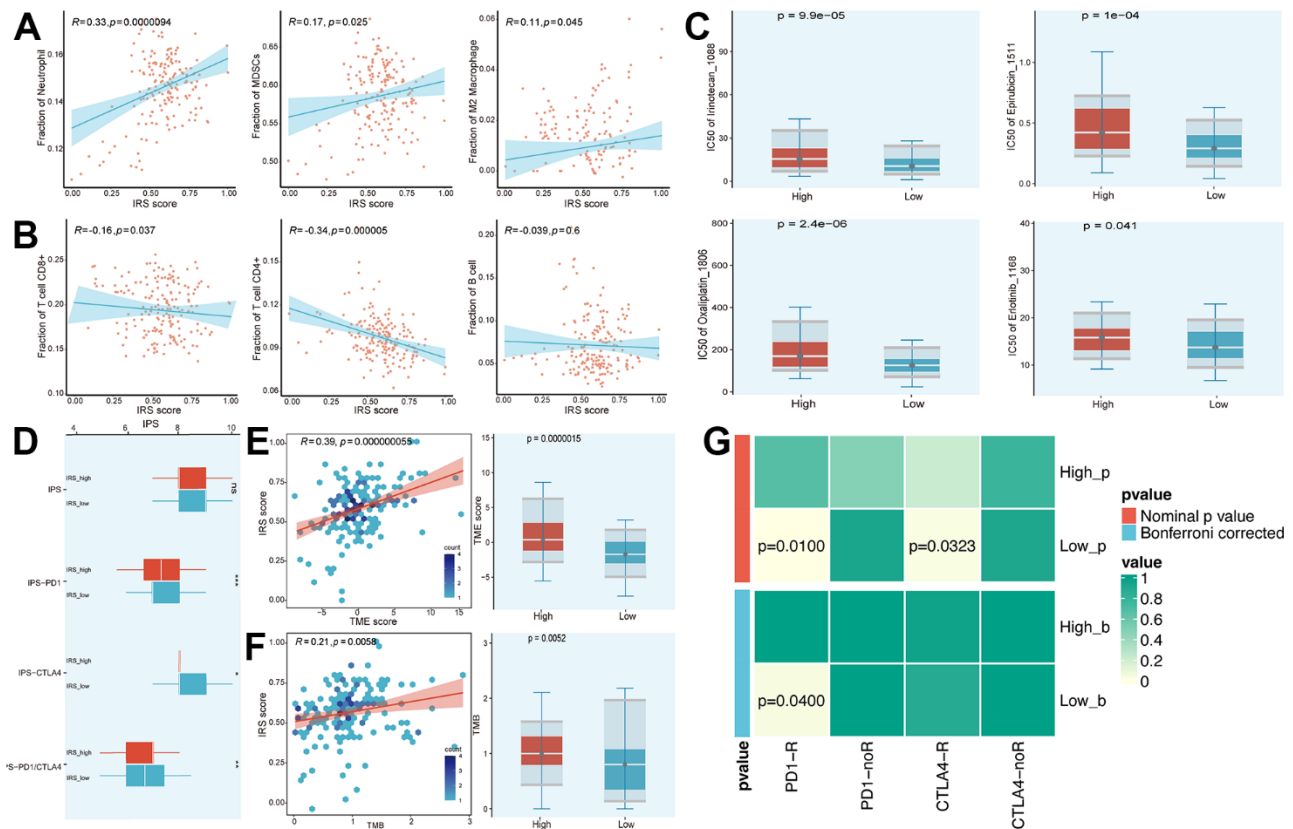
Figure 5. Prediction of chemotherapy and immunotherapy. (**A**, **B**) The relationship between the IRS score and the infiltration of immune cells. (**C**) Estimated IC50 of common chemical compounds between IRS_high and _low groups. (**D**) IPS scores between IRS_high and _low groups. (**E**, **F**) The correlation between IRS score and TME/TMB score. (**G**) Contingency table between immunotherapy responses (anti-PD-1 and anti-CTLA-4) and IRS groups based on SubMap analysis.

[13]. Results illustrated the similarity between patients in the IRS_low group and patients who responded to anti-PD-1 and anti-CTLA4 immunotherapy (Figure 5G). All these results implied that the IRS_low group may have better feedback in immunotherapy compared to IRS_high group, which needs to be further validated in immunotherapy cohorts of PC.

Single-cell sequencing reveals potential mechanism of TIME regulation by IRS

To further recognize the TIME personalized features of pancreatic cancer, scRNA-seq data from 24 PC samples were utilized to reveal the potential mechanism of IRS-promoted PC progression. 22910 cells were screened after quality checks according to the aforementioned research methods. According to the marker genes extracted from the literature, 9 clusters were determined and then annotated to 9 cell types (Figure 6A–6C):

malignant, fibroblast, stellate cell, T cell, endothelial cell, macrophage, ductal, B cell and endocrine cell. To unveil the mechanism of the IRS, we evaluated the distribution of IRS scores in 9 cell types. As illustrated in Figure 6D, higher IRS scores were mainly congregated in malignant cells, which could explain the poor prognosis of high IRS PCs. We also checked the expression of IRS in 9 cell clusters. SMPD3 and TNNT1 were significantly expressed by malignant cells, while AFF3 and ANLN were mainly expressed by B cells (Figure 6E). Therefore, we suggested that malignant cells may contribute to the specific TIME in IRS. Cell-cell communication demonstrated that fibroblast had a significant influence on malignant cells via MIF-(CD74+/CD44) interactions, and ductal cells affected malignant cells through SPP1-CD44 interactions (Figure 6F). Meyer-Siegler KL [14] found that blocking MIF-CD74 interactions may provide new targeted specific therapies for androgen-independent prostate cancer. SPP1-CD44 axis was reported to
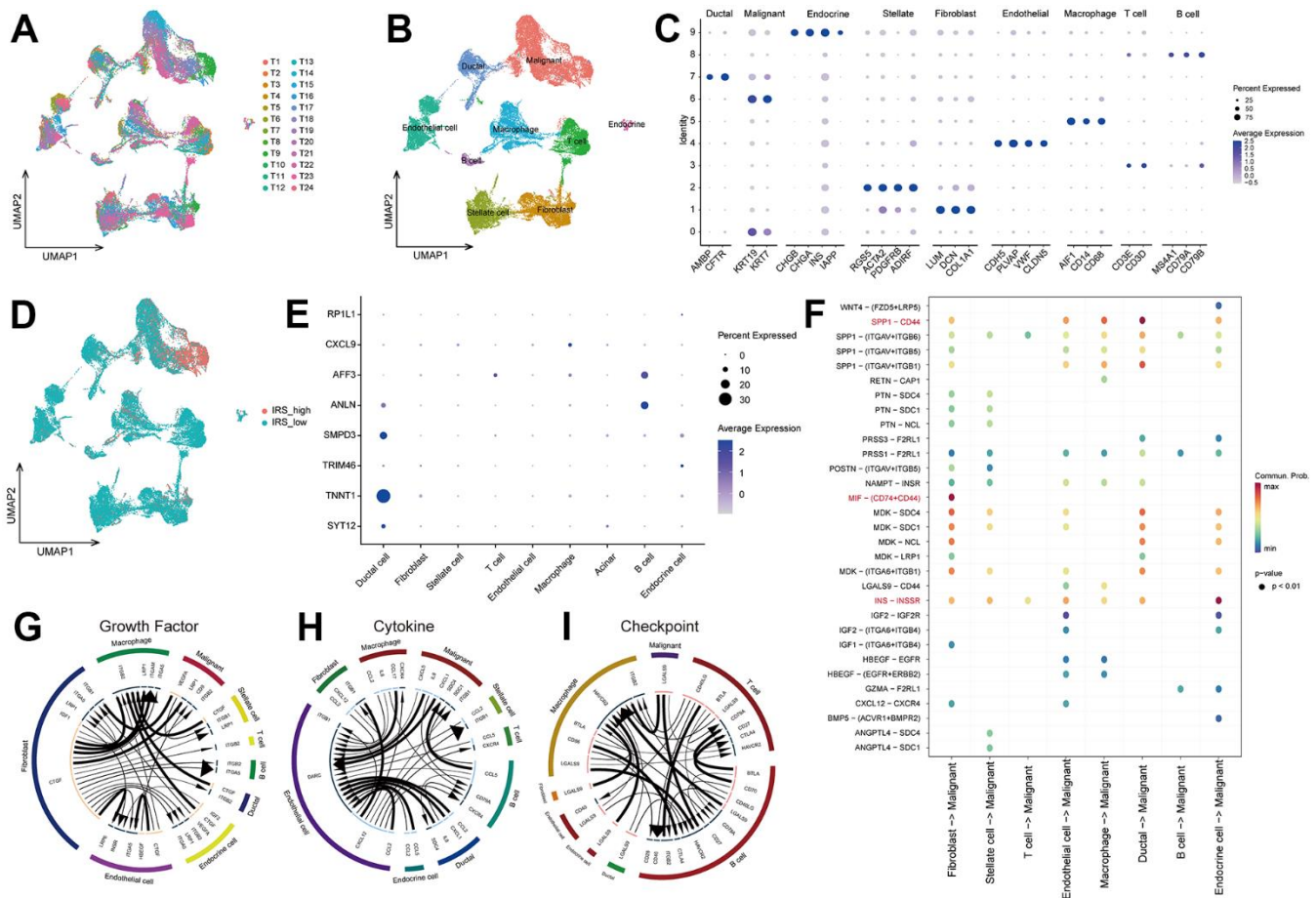
Figure 6. The underlying mechanism of IRS via scRNA-seq analsysis. (**A**) The distribution of cells after quality check. (**B**) UMAP plot showing the cells were clustered into 9 types. (**C**) The expression of marker genes in 9 types of cells. (**D**) The distribution of IRS score among 9 types of cells. (**E**) The expression of 8 IRGs in 9 types of cells. Cell-cell communication analysis were implemented via "CellChat" (**F**) and "iTALK" (**G**–**I**) algorithms.

promote the interplay between CAF and enrichment of stemness population in PC [15]. These results demonstrated that fibroblast cells and ductal cells might promote the progression of cancer via MIF-CD74 and SPP1-CD44 axis, respectively. In addition, CTGF-LRP1 interactions between fibroblast cells and malignant cells could also cause the development of cancer (Figure 6G, 6H). In fact, the expression of HAVCR2 and ITGB2 was higher in macrophage and B cells, while the expression of LGALS9 was upregulated in fibroblast cells and malignant cells (Figure 6I). These results revealed that fibroblast cells might prohibit the activation of immune cells, leading to the "immune dessert" status of PC. Hence, IRS may promote the progression of tumor and suppress the immune system in TIME.

Identification of IRS-related biological processes and drug targets

In order to investigate which biological process plays a critical role in poor prognostic of PC patients who gained high IRS scores, Pathifier and Gene set enrichment analysis (GSEA) analyses were performed to elucidate the potential mechanisms involved in the regulation of PC progression by IRS. Based on gene expression data from both pancreatic cancer and normal pancreatic samples, pathway deregulation score (PDS) was computed via "Pathifier" R package. The correlation between PDS scores and IRS scores helps to evaluate whether a pathway (biological process) may be responsible for the poor prognosis of patients with high IRS scores. "Apoptosis", "TNFA signaling via NFKB", "G2M checkpoint" and "DNA repair" pathways ranked top, which means these three pathways may contribute to the malignant phenotype in patients with high PPS scores (Figure 7A). Next, we performed GSEA analysis to validate the above conclusion. Enrichment score of each gene set was calculated and adjusted P-value less than 0.05 was considered significantly enriched. As expected, genes with positive correlation coefficients were also enriched in those four pathways (Figure 7B). Taken together, the dysregulation of apoptosis and cell cycle-related process might play a vital role in the poor prognosis of high IRS patients.

In high IRS patients, Genes significantly positively correlated with IRS may be potential targets for pancreatic cancer precision therapy. To identify targetable proteins (genes) with potential therapeutic implications in high IRS score PC patients, we conducted Spearman correlation analysis between the protein abundance of targetable genes and PPS. A protein with a correlation coefficient more than 0.3 (with $P < 0.05$) was considered as a poor prognosis-related drug target. Next, we calculated the IRS score

for each PC cell line from the Cancer Cell Line Encyclopedia (CCLE) project, and performed the correlation analysis between the CERES score and PPS score based on these cell lines. A lower CERES score of a gene indicates a higher likelihood that this gene is dependent on a given cancer cell line (CCL). Therefore, we considered a gene with a correlation coefficient less than -0.3 (with $P < 0.05$) as a poor prognosis-dependent drug target. Potential therapeutic drug targets in high IRS score PCs were then considered as targets identified by both analyses above. Finally, 8 potential targets (CCNA2, EPHB4, INCENP, NCF2, PLOD1, PLK1, PANX1 and CCNB1) were screened out (Figure 7C, 7D) and the correlated target drugs were also identified (Figure 7E), which meant that targeting these genes may facilitate the treatment of high IRS PCs.

Identification of potential agents for high IRS score PCs

In the past decade, high-throughput sequencing analysis of large samples has greatly advanced the molecular biology of PC. Hence, we try to detect the potential small molecular compounds for high IRS PCs. The information on compounds in the Cancer Therapeutics Response Portal (CTRP) and Profiling Relative Inhibition Simultaneously in Mixtures (PRISM) database were selected for subsequent analysis after removing the duplicated compound information in the two databases (excluding hematopoietic and lymphoid tissue-derived CCLs) (Figure 8A).

For drug response prediction, many machine learning (ML) methods have been reported, ranging from multivariate linear regression and support vector machine (SVM) to RF and k-nearest neighbours (KNN). Among ML methods, linear regression methods, such as ridge regression and elastic net, tend to exhibit good and robust performance in different settings [16]. Therefore, ridge regression model located in the "oncoPredict" package, which has been applied to multiple studies and proven to be reliable, was applied to estimate drug response of clinical samples in this study [17]. Before selecting the compounds, we further validated the predicted drug sensitivity (AUC) in our cohort. Selumetinib, a PI3K pathway inhibitor, was reported to improve the prognosis in the treatment of KRAS-mutant patients compared to those without KRAS mutations [18]. Thus, we classified PC patients into KRAS altered and KRAS unaltered subgroups. The AUC of PC patients in the KRAS altered group was significantly decreased (Figure 8B, $P = 2.4e-05$), which was consistent with the clinical findings of Simertinib above. Finally, 1 compound from the CTRP database (Canertinib) and 6 compounds from the PRISM database (PP-1, YM-976, CHIR-98014, GW-788388,
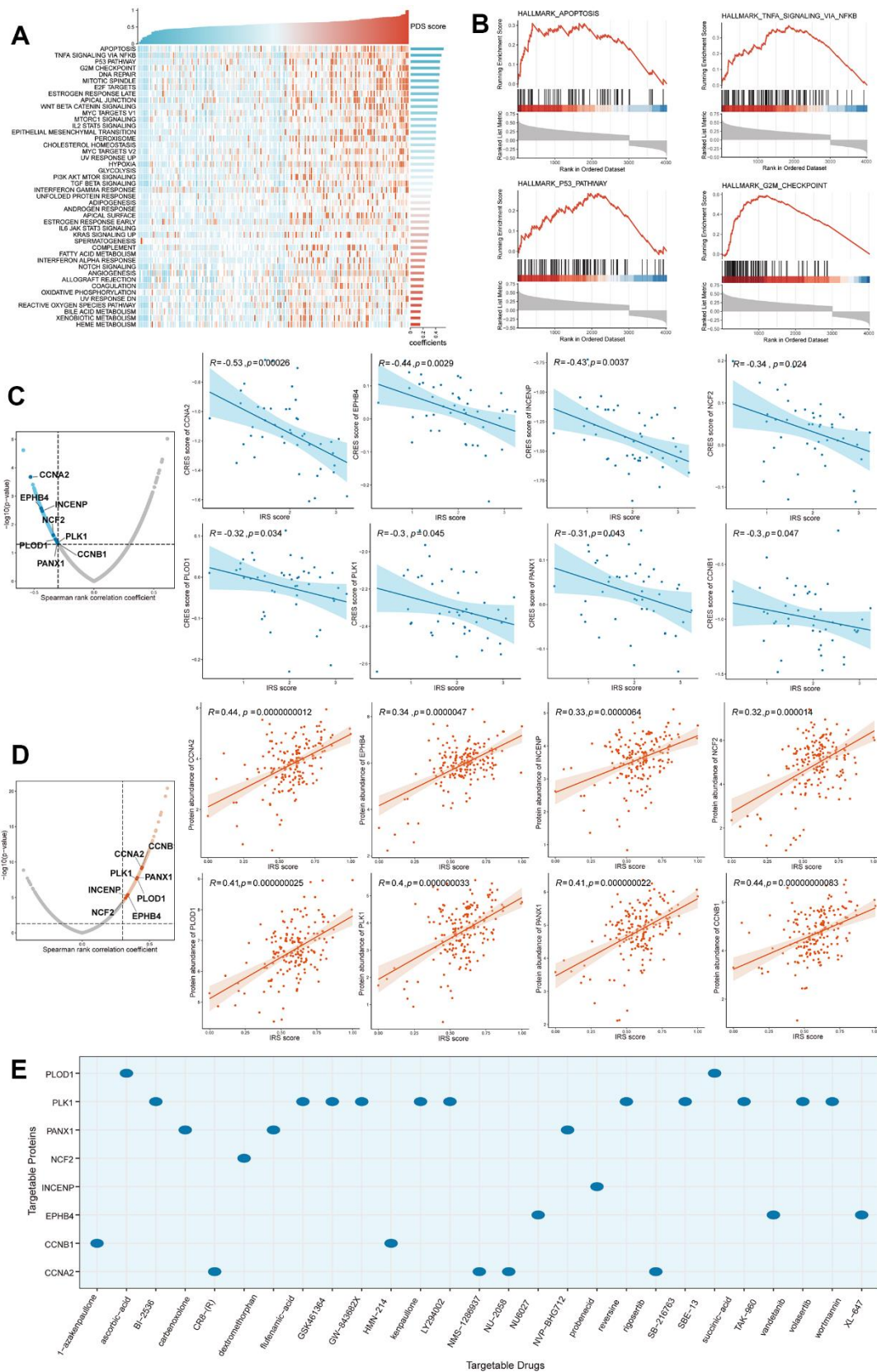
Figure 7. Identification of IRS-related biological processes and drug targets. (**A**) Pathifier analysis showing the IRS-related biological processes. (**B**) Top 4 pathways were validated in GSEA analysis. Potential drug targets were screened out form CTRP (**C**) and PRISM (**D**) datasets. (**E**) The corresponding relationship between drug targets and potential targeted drugs.

Brigatinib and Vincamine) were obtained following the protocol described in Materials and Methods (Figure 8C).

Although these 7 compounds had lower predictive AUC values in the samples with higher PPS scores and their predictive AUC values were significantly negatively correlated with IRS scores, the above analysis alone could not support the conclusion that these compounds had therapeutic effects on PCs. Hence, CMap analysis was utilized to find the most reliable compounds. Among the 7 candidate compounds identified before, Canertinib and PP-1 showed relatively low
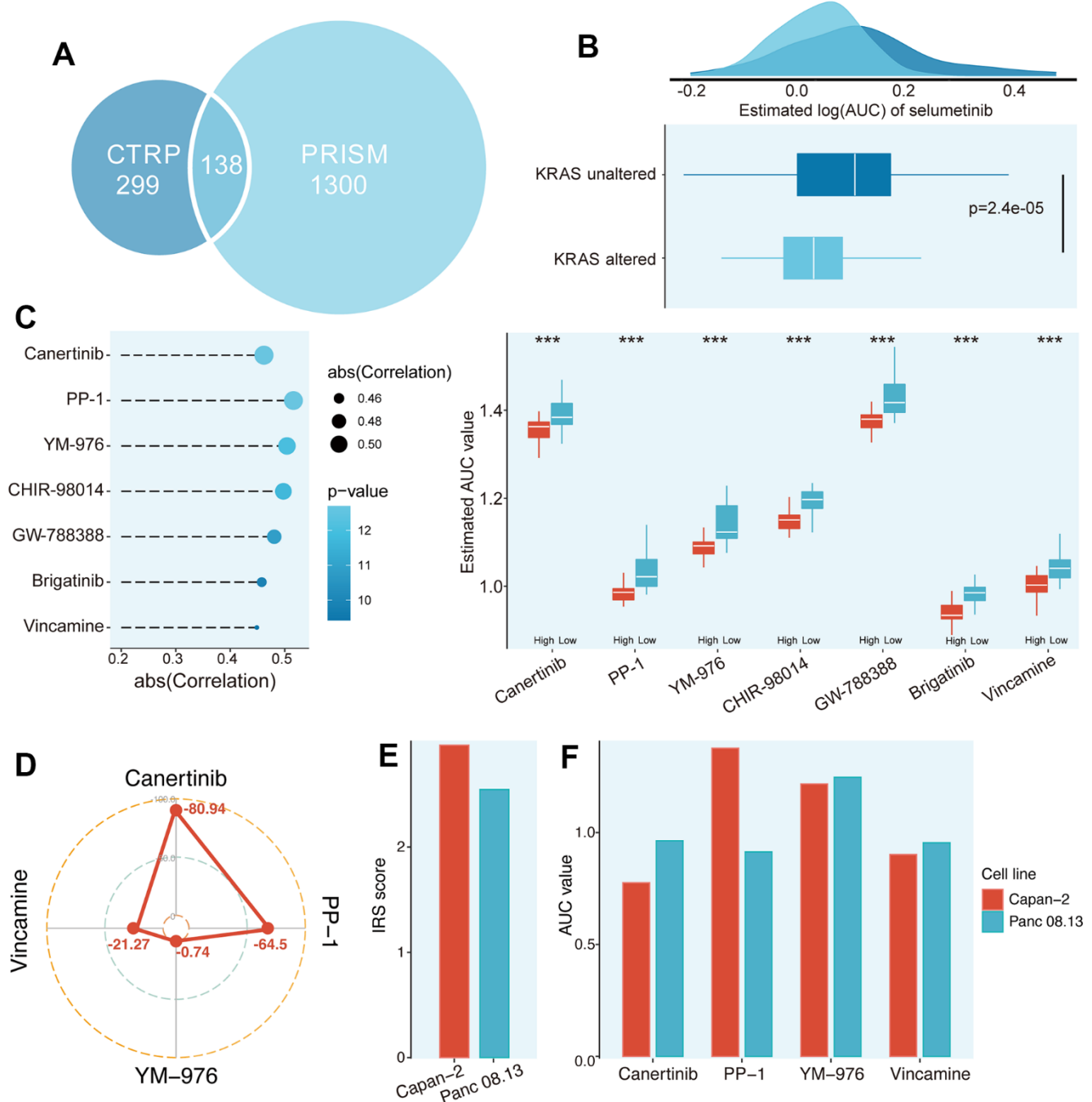
Figure 8. Identification of potential agents for high IRS score PCs. (A) Data availability of drug sensitive data in CTRP and PRISM datasets. (B) Validation of the predicted drug sensitive data based on literature. (C) Potential agents for high IRS patients were identified from CTRP and PRISM datasets. (D) The CMap score of four candidate compounds. (E) The IRS score between Capan-2 and Panc 08.13 cell lines. (F) The AUC value of four candidate compounds in Capan-2 and Panc 08.13 cell lines.

CMap scores (Canertinib, -80.94; PP-1, -64.5), indicating its therapeutic potential (Figure 8D and Supplementary Table 7). To further test the efficiency of these candidates, two PC cell lines (Capan-2 and Panc 08.13) in CTRP and PRISM have been extracted for the following analysis. We first calculated the IRS score of these two cell lines, and the Capan-2 possessed a relatively higher IRS score than Panc 08.13 (Figure 8E and Supplementary Table 8). Secondly, the AUC value of these candidates between Capan-2 and Panc 08.13 were compared. The results indicated that only the AUC value of Canertinib was significantly lower in the Capan-2 compared to Panc 08.13, implying that Canertinib might be the promising potential treatment compound targeted high IRS score PCs (Figure 8F and Supplementary Table 8).

DISCUSSION

As the most common malignant tumor among solid tumors, the complex crosstalk in the microenvironment of pancreatic cancer poses a serious challenge for personalized treatment of patients [19]. With the development of high-throughput sequencing analysis, subtyping cancers on the basis of molecular similarities and clinical characteristics could improve the existing morphological and imaging methods for personalized treatment and risk stratification [20]. Up to now, PC can be divided into multiple molecular subtypes (MS), including Bailey's classification, Collisson's classification, Moffitt's tumor classification, Moffitt's stromal classification, Puleo's classification and Li's classification. Bailey's classification includes Squamous, Pancreatic progenitor, Immunogenic, and Aberrantly Differentiated Endocrine Exocrine (ADEX). Among them, the Squamous subtype enriched for inflammation, metabolic reprogramming, cell pro-liferation and epigenetic downregulation of endodermal genes, which possessed the worst prognosis [21]. Collisson's classification includes Classical subtype related to adhesion and epithelialization, Exocrine-like subtype related to mesenchymal transition, and QM-PDA related to tumor-derived digestive enzymes [22]. Moffitt's tumor classification includes the Classical subtype and Basal-like subtype, and the latter is associated with poor survival of PC [23]. Moffitt's stromal classification contains Absent, Activated and Normal subtypes [23]. Puleo's classification includes Desmoplastic, Immune classical, Pure basal-like, Pure classical and Stroma Activated subtypes [24]. Li's classification defines the TIME of PC as Immune Class and Nonimmune Class [25]. However, the molecular typing of PC is in its infancy. Hence, novel molecular signatures are still necessary to provide opportunities to advance the therapeutic development of PC.

It is reported that the interactions between cancer cells and proximal immune cells can ultimately lead to an environment that promotes the growth and metastasis of PC [26]. Furthermore, a deeper understanding of IRGs involved in the TIME could help illustrate their regulatory mechanisms in TIME and develop novel treatment strategies. Numerous types of research have demonstrated that immune and stromal cells are two major components of TIME [27–29]. Hence, we identified the TIME-related differentially expressed IRGs particularly from PC samples by evaluating tumor-infiltrating immune cells and stromal cells via the ESTIMATE algorithm. Additionally, identifying IRGs that are differentially expressed in tumor and normal tissues could be conducive to selecting dysregulated genes in PC. Therefore, we extracted IRGs with differential expression in the TIME and PC samples that may effectively reflect the characteristics of TIME in PC. We further confirmed their functions in the immune system via pathway enrichment analysis. Unsupervised consensus clustering algorithm was implemented to classify PC patients into two TIME subtypes. TIMER analysis exhibited that the Immune_rich subtype possessed higher infiltration of CD8+ T cells. Currently, according to the specific tumor environment and immune contexture, three main subtypes of tumor; the immune hot, altered and cold tumors were determined [30]. The terms "hot" and "cold" are defined by T cell-infiltrated, inflamed but non-infiltrated, and non-inflamed tumors [31]. Hence, the Immune_rich subtype was correlated to the "hot" tumor and the TIDE algorithm exhibited that the "hot" tumor tended to gain lower TIDE scores, implying their sensitivity to ICB treatment was higher. Moreover, the immune signatures of T cells (TH1, IFNγ, GNLY, PRF1, GZMs) were associated with prolonged survival and more sensitive to anti-PD1 treatment [32–35]. Although the abundance of T cells was higher in Immune_rich subtype, most of them were in a dysfunctional state, leading to a lower TIDE score. Therefore, the risk stratification merely based on the infiltration of T cells is too limited to guide the clinical strategies of immunotherapy, and our novel classifica-tion may pose new directions in the future.

Despite the rapid development of diagnostic methods and therapeutic strategies for PC, the high degree of heterogeneity in PC still makes its prognosis prediction and treatment efficacy face great challenges. In the past decade, many researchers have done a lot of work to develop immune-related prognostic prediction models. However, the construction method of those prognostic markers is relatively single and only applies to the whole PC population, without individualized clinical management analysis for high-risk groups, which is not enough for accurate risk stratification of PC patients. In

fact, with the rapid development of artificial intelligence in the biomedical field, machine learning, as an important branch of artificial intelligence, has been widely used in bulk transcriptomics, single-cell transcriptomics, spatial transcriptomics, radionics and other fields. Chen P [36] and his team showed that through deep learning models (belonging to the branch of machine learning), they developed a model named DeepMACT, which can systematically analyze the size, shape, spatial distribution and other characteristics of tumors, as well as the degree of targeted metastasis by therapeutic monoclonal antibodies. It is an important discovery of the target antibody in the preclinical stage. Boris V J et al. [37] summarized the important role of image-based machine learning algorithms in predicting the clinical outcome of PDAC patients. Among 25 studies based on machine learning algorithms published from 2019 to 2020, 9 models effectively predicted the clinical outcome (AUC: 0.78-0.95, C-index: 0.65-0.76). Therefore, in order to develop a quantified signature to stratify PC patients, we selected machine learning algorithms to construct the IRS model to determine immune-related risk classification in PC patients. Among the 8 genes in the IRS model, SYT12 is reported to play a vital role in oral squamous cell carcinoma (OSCC) progression via CAMK2N1 and could be a new target for OSCC patients [38]. TNNT1 is regulated by miR-873 and confirmed as an oncogene of colorectal cancer (CRC) [39]. TRIM46, which is affiliated in the tripartite motif (TRIM) protein family, acts as an E3 ligase that targets HDAC1 and promotes carcinogenesis and chemoresistance in breast cancer [40]. Similarly, an integrative genomic analysis revealed that SMPD3 is a tumor suppressor gene that could influence the aggressiveness of the hepatocellular carcinoma (HCC) [41]. ANLN promotes the progression of PC via EZH2/miR-218-5p/LASP1 axis, suggesting that ANLN could be served as a potential therapeutic target in PC [42]. CXCL9 was listed as a conserved 4-chemokine signature marks resectable and metastatic PC tumors with an active antitumor phenotype [43].

Meanwhile, we also explored the relationships between IRS and TIME. The IRS score was positively correlated with neutrophils, MDSCs and M2 macrophages, while negatively related to CD8+ T cells and CD4+ T cells. Neutrophils, accounting for 70% of circulating leukocytes, exhibit an N1 (tumor-suppressive) or N2 (tumor-promoting) phenotype in the context of cancer [44]. We suggested that N1 type of neutrophils were abundant in high IRS patients according to the results mentioned above. MDSCs could lead to immuno-suppression, including T cell suppression and innate immune regulation via multiple mechanisms in TIME [45]. Most importantly, MDSCs strengthened cell

stemness and promoted the metastatic process by promoting EMT through IL-6 secretion in tumors [46]. Macrophages can be polarized into inflammatory M1 (classically activated) or immune-suppressive M2 (alternatively activated). Based on the secretion of IL-4, TIME enhanced the immune suppressive M2 which in turn enables tumor growth and progression [47]. CD8+ T cells along with CD4+ T cells are contributed to adaptive immunity and anti-tumor immunity [48]. scRNA-seq was also applied to further explore the underlying mechanism of how IRS leads to the diversity of TIME. The malignant cells tended to possess higher IRS score and may contribute to the specific TIME in IRS. Additionally, Cell-cell communications illustrated that fibroblast and ductal cells could contribute to the development of tumor cells by targeting the SPP1-CD44 and MIF-CD74 axis. In general, our IRS performed well in predicting prognosis and the sensitivity of immunotherapy in PC patients. However, the ultimate goal of clarifying risk stratification is to achieve individualized and personalized treatment, so the screening of drug targets and potential agents has become the main breakthrough.

With the development of next-generation sequencing genomics, researchers can rapidly identify genetic differences between tumor cells and normal cells, genomic mutations, and changes in downstream pathways, which provides convenience for the development of drug targets. Currently, many types of malignant tumors (e.g., breast and ovarian cancer) benefit from "precision medicine" with targeted drugs. However, few targeted drugs have been approved for PC, and it only marginally prolongs patient survival [49]. Hence, based on pharmacogenomic databases, we identified 8 drug targets and 1 potential agents for high IRS patients in PC.

In terms of 8 targets screened from Drug Repurposing Hub and CCLE datasets, it is reported that the high expression of CCNA2 is associated with a worse prognosis in PC and is correlated with advanced tumor stage [50]. Inhibition of EPHB4 combined with radiation can modulate the microenvironment response post-radiation, contributing to increased tumor control in PC [51]. PLOD genes or PLOD family genes also could be served as potential prognostic biomarkers for PC [52]. PLKT1 suppresses PC progression and inhibits NF-κB activity, and targeting PLKT1 can alleviate the sensitivity of immunotherapy in PC [53]. The up-regulation of PANX1 was correlated with poor out-comes and immune infiltration in PC [54]. CCNB1 silencing suppresses cell proliferation and promotes cell senescence by activating the p53 signalling pathway in PC [55]. Although INCENP and NCF2 haven't been reported in PC, it needs further exploration could

concentrate on these two novel targets. Moreover, we identified Canertinib as the most reliable agent targeting high IRS score PCs based on CTRP and PRISM datasets. Canertinib, an EGFR inhibitor, has been demonstrated effective in pNETs according to available genetic atlas data [56]. But unfortunately, its clinical efficiency in PC has been insofar moderate. Current work provides new insights into improving the therapeutic effect of PC, offering new directions for the precision treatment of PC.

Importantly, our study differed from previous studies in the following aspects: (1) Our established TIME subtype tightly correlated with the classical six classifications, which confirmed the reliability of our classification and shed light on a novel strategy for the treatment of PC. (2) Via multiple machine-learning algorithms, the IRS was constructed and achieved better performance in risk stratification than previous prognostic signatures. (3) Recently, numerous studies have merely focused on subtyping PC at an immunogenic level. However, they failed to deliver precision medicine for PC patients based on their classifications. Apart from being informative regarding TIME and prognosis, IRS can also be implemented for precise oncology, and our results have the potential to refine the status quo of population-based therapies and guide personalized treatment in PC. However, this study has several limitations. For instance, our research merely focused on public retrospective datasets, and the predictive efficiency of the IRS in immunotherapy response requires further validation in immunotherapy cohorts of PC. Furthermore, the results of drug targets and agents prediction cannot be verified against each other, which reduces the power of the conclusions.

In conclusion, we classified two TIME subtypes with specific tumor microenvironments and accessed the differences in potential response among these two subtypes. Additionally, we developed a novel immune-related prognostic signature—IRS, and validated it in various cohorts and experiments. Finally, based on multiple drug susceptibility and target databases, we have identified seven potential therapeutic targets and two compounds, which shed new light on the application of precision medicine in PC.

MATERIALS AND METHODS

Data acquisition and preprocessing

The expression profile of pancreatic cancer patients was downloaded from TCGA dataset (https://portal.gdc.cancer.gov/) in the form of Fragments Per Kilobase Million (FPKM) and transformed into log2(TPM+1)

format data. The corresponding clinical data from the TCGA dataset was downloaded from UCSC Xena (https://xenabrowser.net/datapages/). A total of 149 cases in TCGA with corresponding PC tissues and complete clinical data were enrolled in the study [57]. The RNA-Seq data of the CELL cohort (CPTAC3-Discovery project, n = 135) was employed in this study to construct the prognostic signature, which was obtained from Proteomic Data Commons (PDC, https://pdc.cancer.gov/pdc/) and LinkedOmics (http://www.linkedomics.org/data_download/CPTAC-PDAC/). TCGA and CELL cohorts were combined as a meta-cohort (n = 284) to facilitate the model training and the "sva" R package was used to remove the batch effect between two independent datasets (Supplementary Figure 5). Genotype-Tissue Expression (GTEx, https://www.gtexportal.org/) dataset containing the expression data of normal pancreatic was also included (n = 167). Meanwhile, ICGC (https://dcc.icgc.org/, n = 81) and GEO (http://www.ncbi.nlm.nih.gov/geo) datasets (GSE62452, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62452, n = 65) were extracted to validate the efficiency of an established model. The single-cell dataset of PDAC (CRA001160) was extracted from the TISCH database (http://tisch.comp-genomics.org/home/), which contained 24 pancreatic tumor tissues and 11 normal tissues. In terms of studies in personalized treatment, the expression profile of pancreatic CCLs was screened from the Broad Institute CCLE project (https://portals.broadinstitute.org/ccle/, n = 44). Drug sensitivity data of CCLs were achieved from the CTRP v.2.0 (https://portals.broadinstitute.org/ctrp, containing the sensitivity data for 481 compounds over 835 CCLs) and PRISM Repurposing dataset (19Q4, released December 2019, https://depmap.org/portal/prism/, containing the sensitivity data for 1448 compounds over 482 CCLs).

Screening for immune-related genes

ESTIMATE algorithm was applied to calculate the immune scores and stromal scores based on the expression profile of meta-cohort [58]. Then, according to the median value, PC patients were divided into high- and low-immune/stromal score subgroups. Differential expression analysis was performed to screen out the dysregulatory genes among immune score subgroup, stromal score subgroup, and between tumor and normal pancreatic samples via "DESeq2" R package with the criteria of |log2Fold change| >1 and p<0.05 [59]. Importantly, the converged DEGs among three subgroups were defined as IRGs.

Unsupervised clustering analysis

In order to determine the specific patterns of IRGs, the unsupervised consensus clustering algorithm was

implemented via "ConsensusClusterPlus" R package [60]. In addition, principal component analysis (PCA) analysis was also conducted to validate the difference between subtypes.

Enrichment analysis and immune landscape of immune subtypes

GSEA, GO and KEGG annotation were adopted for determining the statistical significance of molecular pathways as well as the consistent heterogeneities between among different groups via "clusterProfiler" R package [61]. A pathway with FDR q < 0.25 and P < 0.05 was defined as statistically significant. single sample Gene Set Enrichment Analysis (ssGSEA) [62] and TIMER [63] algorithm was utilized to evaluate the tumor-infiltrating immune cells among different TIME subtypes. The previously published signatures of immune- and stroma- cells were selected to calculate the abundance of tumor-infiltrating immune cells through ssGSEA [62]. TIDE algorithm was used to predict responsiveness to ICBs between different groups, and lower TIDE scores implied better immuno-therapeutic efficacy [64].

Published PC classifications prediction and comparison

The relationship between our TIME subtype and reported PC molecular classifications was also explored. Six classical PC classifications have been analyzed, including Bailey's classification [21], Collisson's classification [22], Moffitt's tumor classification [23], Moffitt's stromal classification [23], Puleo's classification [24] and Li's classification [25]. Based on published signature genes and algorithms, unsupervised consensus clustering was applied for identifying the subtyping schemas of Bailey's classification, Collisson's classification, Moffitt's tumor classification, Moffitt's stromal classification and Li's classification on meta-cohort using the "ConsensusClusterPlus" package in R. For the prediction of Puleo's classification, we followed the pipeline defined by Puleo [24]. For each sample in meta-cohort, the expression of genes in the centroids was selected, and Spearman rank correlation analysis was conducted between selected genes and 5 centroids. The subtype centroid with the highest correlation is the predicted class of the tested sample. The comparison between the distribution of six predicted classifications and our TIME subtype was measured by Fisher's exact test. R package "ggalluvial" was utilized to plot the Sankey diagram [65]. Cramer's V served as an effect size measurement for the association between TIME subtypes and the other six classifications. It ranges from 0 to 1 where, 0 indicates no association between the two

variables, and 1 indicates a perfect association between them.

Screening, construction and validation of the IRS

PC patients in meta-cohort were categorized into the training set and testing set at the ratio of 7:3. Then, in order to screen the prognostic DEGs, the Bootstrapping univariate Cox analysis was conducted by the "survival" R package. Furthermore, RSF analysis was applied to dimension reduction. The highest C-index of out-of-bag samples was used as the best model and the underlying gene set was observed, and this gene set was defined as IRS. Finally, The IRS scoring model was constructed with the correlation coefficients obtained from multivariate Cox regression, and the formula was as follows:

$$IRS\ score = \sum_{i=1}^{n} Coef_i * x_i$$

Where $Coef_i$ is the multivariate Cox regression coefficient, and x_i is the expression value according to the optimal IRS score, patients were divided into IRS_high and IRS_low group. The area under the curve (AUC) value was used as the criteria to evaluate the effectiveness of the IRS model.

Otherwise, three conventional published signatures (Wang's signature [66], Tao's signature [67] and Dai's signature [68]) and two classical prognostic signatures (PAMG [69] and PurIST [70] signature) in PC were collected to compare the predictive accuracy of the IRS and these signatures. For three conventional signatures, we calculated the risk scores based on the genes and coefficients provided by the articles (Supplementary Table 9). The PAMG score was calculated via the "pdacmolgrad" R package. PurIST score and classification were obtained following the protocol in the original publication [70]. Afterwards, we comprehensively assessed their predictive performance based on AUC values.

IRS-based chemotherapy sensitivity and ICB sensitivity analysis

To predict potential therapeutic effects in different subgroups, "oncoPredict" R package [17] was applied to predict the drug response of PC patients. Moreover, several predicted scores were conducted to evaluate the immunotherapeutic response to IRS, namely, ICB expression, TMB, TME score and IPS. TMB score was computed based on the somatic mutation data from the TCGA dataset, and confirmed as a predictor of immunotherapy. IPS score was downloaded from the

cancer immunome group atlas (TCIA, https://tcia.at/home) after uploading the expression profile of patients [62]. TME score was calculated by "TMEscore" R package [71], revealing the patients' response to ICB. To further validate the role of IRS in the prediction of immunotherapy, we implemented the Subclass mapping (SubMap) to evaluate the expression similarity between IRS_high/IRS_low patients and patients who responded/non-responded to anti-PD-1 and anti-CTLA4 immunotherapy [72].

Single-cell RNA sequencing analysis

The dataset CRA001160 was utilized for scRNA sequencing analysis. UMI count matrices were generated for each sample, and imported into the "Seurat" R package. Low quality cells (<200 genes/cell or >20% mitochondrial genes) were excluded. "Seurat" package was applied for normalization and scaling of the expression matrix, using default settings [73]. Mitochondrial contamination was regressed out by setting "vars.to.regress" parameter. The doublets were cleared out by the "DoubletFinder" R package (version 2.0.3) [74]. To reduce the dimensionality of the expression matrix, PCA analysis was performed based on 2,000 highly variable genes. JackStraw analysis was utilized to identify significant principal components (PC), and PC 1~10 was used for graph-based clustering (res = 0.8) to determine distinct groups of cells. Via previously computed PC 1~10, these groups were projected onto the t-SNE analysis. Subsequently, we used the Seurat FindMarker function to find marker genes of each cell cluster, and defined cell types based on previous datasets and literature [75–78].

Cell–cell interaction analysis

To analyze the cell-cell interactions, R package "CellChat" [79] was employed to predict the major incoming and outgoing intercellular communication networks. In our work, cell-cell interactions were analyzed following the default pipeline. Normalized scRNA-seq counts data were used to create CellChat object with the recommended preprocessing functions. CellChatDB.human was utilized as the database for inferring cell–cell communication with default parameters. "ECM-Receptor" in the database was applied in the analysis. Communications including less than 10 cells were excluded. "iTALK" R package [80] was also used to estimate cell-cell communication. The top 50% of highly expressed genes in each cluster were projected to ligand-receptor pairs in the "iTALK" package. Four categories, including checkpoint protein, cytokine, growth factor, and "other" protein, were employed in our study. The top 30 ligand-receptor pairs for each type were extracted for visualization.

Exploring the biological process of IRS from gene-level to pathway-level

Pathifier analysis was implemented to dig into the differences between IRS_high and _low subgroups via "Pathifier" R package [81]. The Pathifier analysis method is used to identify specific signaling pathways at specific stages of cancer, and can be used in the personalized treatment of cancer. By means of correlation, variance stability and principal component analysis, Pathway Deregulation Score (PDS) was calculated for each PC sample, and then used to estimate the degree to which the activity of a pathway in PC samples deviates from normal samples.

RT-qPCR analysis

All cells in the experiments, including AsPC-1 (RRID: CVCL_0152), BxPC-3 (RRID: CVCL_0186), PANC-1 (RRID: CVCL_0480), PaTu 8988t (RRID: CVCL_1847), and hTERT-HPNE (RRID: CVCL_C466) cells, were purchased from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China) and used for RT-qPCR. All human cell lines have been authenticated using STR profiling within the last three years and that all experiments were performed with mycoplasma-free cells. RNA was reverse transcribed into cDNA using a reverse transcription kit. Gently vortex and then put into the quantitative PCR instrument for amplification. Three technical replicates of each PCR reaction were conducted to ensure the credibility of the experiment. The forward and reverse primers were listed in Supplementary Table 10.

Statistical analysis

Student's t-test was applied in the normal distribution data; Wilcox test was applied for non-normal distribution data between independent groups. Spearman analysis was applied to estimate the correlations between two variables that are not linearly related. The Kaplan–Meier test was utilized to validate the fraction of PC patients living for a certain survival time via the survival package. The log-rank test was conducted to compare the significance of the difference. A two-tailed p-value of less than 0.05 was deemed statistically significant unless specifically stated. See Supplementary File 1 for more information.

Abbreviations

ADEX: Aberrantly Differentiated Endocrine Exocrine; AUC: area under the curve; CAFs: cancer-associated fibroblasts; CCLE: Cell Line Encyclopedia; CTRP: Cancer Therapeutics Response Portal; C-index: concordance index; DEGs: differential expression genes;

ECM: extracellular matrix; ESTIMATE: Estimation of STromal and Immune cells in MAlignant Tumours using Expression data; GEO: Gene Expression Omnibus; GO: Gene Ontology; GSEA: Gene set enrichment analysis; GTEx: Genotype-Tissue Expression; HCC: hepatocellular carcinoma; IPS: immunophenoscore; IRGs: immune-related genes; IRS: immune-related signature; KEGG: Kyoto Encyclopedia of Genes and Genomes; KNN: k-nearest neighbors; MD: Minimal Depth; ML: machine learning; MMP: matrix metalloproteinase; MsigDB: Molecular Signatures Database; ORLs: overall response rates; OSCC: oral squamous cell carcinoma; PAMG: pancreatic adenocarcinoma molecular gradient; PC: Pancreatic cancer; PC: principal components; PCA: Principal component analysis; PDS: Pathway Deregulation Score; RSF: random survival forest; ssGSEA: single sample Gene Set Enrichment Analysis; SubMap: Subclass mapping; SVM: support vector machine; TCGA: The Cancer Genome Atlas; TCIA: the cancer immunome group atlas; TGFβ: transforming growth factor-β; TIDE: Tumor Immune Dysfunction and Exclusion; TIME: Tumor immune microenvironment; TIMER: Tumor Immune Estimation Resource; TMB: tumor mutation burden; Tregs: regulatory T cells; TRIM: the tripartite motif.

AUTHOR CONTRIBUTIONS

DC, MQ and XT conceived and supervised the study. LZ and DC analyzed the data. LZ performed experiments. LZ and DC wrote the draft. ZT, BZ, YZ, FZ, DC, MQ and XT revised and validated the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

REFERENCES

1. Vincent A, Herman J, Schulick R, Hruban RH, Goggins M. Pancreatic cancer. Lancet. 2011; 378:607–20.
https://doi.org/10.1016/S0140-6736(10)62307-0
PMID:21620466

2. Klein AP. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. Nat Rev Gastroenterol Hepatol. 2021; 18:493–502.
https://doi.org/10.1038/s41575-021-00457-x
PMID:34002083

3. Wang Z, Li Y, Ahmad A, Banerjee S, Azmi AS, Kong D, Sarkar FH. Pancreatic cancer: understanding and overcoming chemoresistance. Nat Rev Gastroenterol Hepatol. 2011; 8:27–33.
https://doi.org/10.1038/nrgastro.2010.188
PMID:21102532

4. Wang Y, Yang G, You L, Yang J, Feng M, Qiu J, Zhao F, Liu Y, Cao Z, Zheng L, Zhang T, Zhao Y. Role of the microbiome in occurrence, development and treatment of pancreatic cancer. Mol Cancer. 2019; 18:173.
https://doi.org/10.1186/s12943-019-1103-2
PMID:31785619

5. Asaoka Y, Ijichi H, Koike K. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med. 2015; 373:1979.
https://doi.org/10.1056/NEJMc1510353
PMID:26559583

6. Royal RE, Levy C, Turner K, Mathur A, Hughes M, Kammula US, Sherry RM, Topalian SL, Yang JC, Lowy I, Rosenberg SA. Phase 2 trial of single agent Ipilimumab (anti-CTLA-4) for locally advanced or metastatic pancreatic adenocarcinoma. J Immunother. 2010; 33:828–33.
https://doi.org/10.1097/CJI.0b013e3181eec14c
PMID:20842054

7. Somaiah N, Conley AP, Parra ER, Lin H, Amini B, Solis Soto L, Salazar R, Barreto C, Chen H, Gite S, Haymaker C, Nassif EF, Bernatchez C, et al. Durvalumab plus tremelimumab in advanced or metastatic soft tissue and bone sarcomas: a single-centre phase 2 trial. Lancet Oncol. 2022; 23:1156–66.
https://doi.org/10.1016/S1470-2045(22)00392-8
PMID:35934010

8. Pitt JM, Marabelle A, Eggermont A, Soria JC, Kroemer G, Zitvogel L. Targeting the tumor microenvironment: removing obstruction to anticancer immune responses and immunotherapy. Ann Oncol. 2016; 27:1482–92.
https://doi.org/10.1093/annonc/mdw168
PMID:27069014

9. Tang T, Huang X, Zhang G, Hong Z, Bai X, Liang T. Advantages of targeting the tumor immune microenvironment over blocking immune checkpoint in cancer immunotherapy. Signal Transduct Target Ther. 2021; 6:72.
https://doi.org/10.1038/s41392-020-00449-4
PMID:33608497

10. Mao X, Xu J, Wang W, Liang C, Hua J, Liu J, Zhang B, Meng Q, Yu X, Shi S. Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. Mol Cancer. 2021; 20:131.
https://doi.org/10.1186/s12943-021-01428-1
PMID:34635121

11. Ren B, Cui M, Yang G, Wang H, Feng M, You L, Zhao Y. Tumor microenvironment participates in metastasis of pancreatic cancer. Mol Cancer. 2018; 17:108.
https://doi.org/10.1186/s12943-018-0858-1
PMID:30060755

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

13. Roh W, Chen PL, Reuben A, Spencer CN, Prieto PA, Miller JP, Gopalakrishnan V, Wang F, Cooper ZA, Reddy SM, Gumbs C, Little L, Chang Q, et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. Sci Transl Med. 2017; 9:eaah3560.
https://doi.org/10.1126/scitranslmed.aah3560
PMID:28251903

14. Meyer-Siegler KL, Iczkowski KA, Leng L, Bucala R, Vera PL. Inhibition of macrophage migration inhibitory factor or its receptor (CD74) attenuates growth and invasion of DU-145 prostate cancer cells. J Immunol. 2006; 177:8730–9.
https://doi.org/10.4049/jimmunol.177.12.8730
PMID:17142775

15. Nallasamy P, Nimmakayala RK, Karmakar S, Leon F, Seshacharyulu P, Lakshmanan I, Rachagani S, Mallya K, Zhang C, Ly QP, Myers MS, Josh L, Grabow CE, et al. Pancreatic Tumor Microenvironment Factor Promotes Cancer Stemness via SPP1-CD44 Axis. Gastroenterology. 2021; 161:1998–2013.e7.
https://doi.org/10.1053/j.gastro.2021.08.023
PMID:34418441

16. Cancer Cell Line Encyclopedia Consortium, and Genomics of Drug Sensitivity in Cancer Consortium. Pharmacogenomic agreement between two cancer cell line data sets. Nature. 2015; 528:84–7.
https://doi.org/10.1038/nature15736 PMID:26570998

17. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting *in vivo* or cancer patient drug response and biomarkers from cell line screening data. Brief Bioinform. 2021; 22:bbab260.
https://doi.org/10.1093/bib/bbab260 PMID:34260682

18. Alagesan B, Contino G, Guimaraes AR, Corcoran RB, Deshpande V, Wojtkiewicz GR, Hezel AF, Wong KK, Loda M, Weissleder R, Benes CH, Engelman J, Bardeesy N. Combined MEK and PI3K inhibition in a mouse model of pancreatic cancer. Clin Cancer Res. 2015; 21:396–404.
https://doi.org/10.1158/1078-0432.CCR-14-1591
PMID:25348516

19. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022; 72:7–33.
https://doi.org/10.3322/caac.21708 PMID:35020204

20. Collisson EA, Bailey P, Chang DK, Biankin AV. Molecular subtypes of pancreatic cancer. Nat Rev Gastroenterol Hepatol. 2019; 16:207–20.
https://doi.org/10.1038/s41575-019-0109-y
PMID:30718832

21. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, Miller DK, Christ AN, Bruxner TJ, Quinn MC, Nourse C, Murtaugh LC, Harliwong I, et al., and Australian Pancreatic Cancer Genome Initiative. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature. 2016; 531:47–52.
https://doi.org/10.1038/nature16965 PMID:26909576

22. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, Cooc J, Weinkle J, Kim GE, Jakkula L, Feiler HS, Ko AH, Olshen AB, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med. 2011; 17:500–3.
https://doi.org/10.1038/nm.2344 PMID:21460848

23. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, Rashid NU, Williams LA, Eaton SC, Chung AH, Smyla JK, Anderson JM, Kim HJ, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet. 2015; 47:1168–78.
https://doi.org/10.1038/ng.3398
PMID:26343385

24. Puleo F, Nicolle R, Blum Y, Cros J, Marisa L, Demetter P, Quertinmont E, Svrcek M, Elarouci N, Iovanna J, Franchimont D, Verset L, Galdon MG, et al. Stratification of Pancreatic Ductal Adenocarcinomas Based on Tumor and Microenvironment Features. Gastroenterology. 2018; 155:1999–2013.e3.
https://doi.org/10.1053/j.gastro.2018.08.033
PMID:30165049

25. Li R, He Y, Zhang H, Wang J, Liu X, Liu H, Wu H, Liang Z. Identification and Validation of Immune Molecular Subtypes in Pancreatic Ductal Adenocarcinoma: Implications for Prognosis and Immunotherapy. Front Immunol. 2021; 12:690056.
https://doi.org/10.3389/fimmu.2021.690056
PMID:34335594

26. Hinshaw DC, Shevde LA. The Tumor Microenvironment Innately Modulates Cancer Progression. Cancer Res.

2019; 79:4557–66.
https://doi.org/10.1158/0008-5472.CAN-18-3962
PMID:31350295

27. Hessmann E, Buchholz SM, Demir IE, Singh SK, Gress TM, Ellenrieder V, Neesse A. Microenvironmental Determinants of Pancreatic Cancer. Physiol Rev. 2020; 100:1707–51.
https://doi.org/10.1152/physrev.00042.2019
PMID:32297835

28. Hosein AN, Brekken RA, Maitra A. Pancreatic cancer stroma: an update on therapeutic targeting strategies. Nat Rev Gastroenterol Hepatol. 2020; 17:487–505.
https://doi.org/10.1038/s41575-020-0300-1
PMID:32393771

29. Ullman NA, Burchard PR, Dunne RF, Linehan DC. Immunologic Strategies in Pancreatic Cancer: Making Cold Tumors Hot. J Clin Oncol. 2022; 40:2789–805.
https://doi.org/10.1200/JCO.21.02616 PMID:35839445

30. Galon J, Bruni D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. Nat Rev Drug Discov. 2019; 18:197–218.
https://doi.org/10.1038/s41573-018-0007-y
PMID:30610226

31. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pagès C, Tosolini M, Camus M, Berger A, Wind P, Zinzindohoué F, Bruneval P, Cugnenc PH, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science. 2006; 313:1960–4.
https://doi.org/10.1126/science.1129139
PMID:17008531

32. Goc J, Germain C, Vo-Bourgais TK, Lupo A, Klein C, Knockaert S, de Chaisemartin L, Ouakrim H, Becht E, Alifano M, Validire P, Remark R, Hammond SA, et al. Dendritic cells in tumor-associated tertiary lymphoid structures signal a Th1 cytotoxic immune contexture and license the positive prognostic value of infiltrating CD8+ T cells. Cancer Res. 2014; 74:705–15.
https://doi.org/10.1158/0008-5472.CAN-13-1342
PMID:24366885

33. Mulligan AM, Pinnaduwage D, Tchatchou S, Bull SB, Andrulis IL. Validation of Intratumoral T-bet+ Lymphoid Cells as Predictors of Disease-Free Survival in Breast Cancer. Cancer Immunol Res. 2016; 4:41–8.
https://doi.org/10.1158/2326-6066.CIR-15-0051
PMID:26546451

34. Mulligan AM, Raitman I, Feeley L, Pinnaduwage D, Nguyen LT, O'Malley FP, Ohashi PS, Andrulis IL. Tumoral lymphocytic infiltration and expression of the chemokine CXCL10 in breast cancers from the Ontario Familial Breast Cancer Registry. Clin Cancer Res. 2013;

19:336–46.
https://doi.org/10.1158/1078-0432.CCR-11-3314
PMID:23213058

35. Cristescu R, Mogg R, Ayers M, Albright A, Murphy E, Yearley J, Sher X, Liu XQ, Lu H, Nebozhyn M, Zhang C, Lunceford JK, Joe A, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. Science. 2018; 362:eaar3593.
https://doi.org/10.1126/science.aar3593
PMID:30309915

36. Pan C, Schoppe O, Parra-Damas A, Cai R, Todorov MI, Gondi G, von Neubeck B, Böğürcü-Seidel N, Seidel S, Sleiman K, Veltkamp C, Förstera B, Mai H, et al. Deep Learning Reveals Cancer Metastasis and Therapeutic Antibody Targeting in the Entire Body. Cell. 2019; 179:1661–76.e19.
https://doi.org/10.1016/j.cell.2019.11.013
PMID:31835038

37. Janssen BV, Verhoef S, Wesdorp NJ, Huiskens J, de Boer OJ, Marquering H, Stoker J, Kazemier G, Besselink MG. Imaging-based Machine-learning Models to Predict Clinical Outcomes and Identify Biomarkers in Pancreatic Cancer: A Scoping Review. Ann Surg. 2022; 275:560–7.
https://doi.org/10.1097/SLA.0000000000005349
PMID:34954758

38. Eizuka K, Nakashima D, Oka N, Wagai S, Takahara T, Saito T, Koike K, Kasamatsu A, Shiiba M, Tanzawa H, Uzawa K. SYT12 plays a critical role in oral cancer and may be a novel therapeutic target. J Cancer. 2019; 10:4913–20.
https://doi.org/10.7150/jca.32582 PMID:31598163

39. Chen Y, Wang J, Wang D, Kang T, Du J, Yan Z, Chen M. TNNT1, negatively regulated by miR-873, promotes the progression of colorectal cancer. J Gene Med. 2020; 22:e3152.
https://doi.org/10.1002/jgm.3152 PMID:31830337

40. Zhang Z, Liu X, Li L, Yang Y, Yang J, Wang Y, Wu J, Wu X, Shan L, Pei F, Liu J, Wang S, Li W, et al. SNP rs4971059 predisposes to breast carcinogenesis and chemoresistance via TRIM46-mediated HDAC1 degradation. EMBO J. 2021; 40:e107974.
https://doi.org/10.15252/embj.2021107974
PMID:34459501

41. Revill K, Wang T, Lachenmayer A, Kojima K, Harrington A, Li J, Hoshida Y, Llovet JM, Powers S. Genome-wide methylation analysis and epigenetic unmasking identify tumor suppressor genes in hepatocellular carcinoma. Gastroenterology. 2013; 145:1424–35.e1.
https://doi.org/10.1053/j.gastro.2013.08.055
PMID:24012984

42. Wang A, Dai H, Gong Y, Zhang C, Shu J, Luo Y, Jiang Y,

Liu W, Bie P. ANLN-induced EZH2 upregulation promotes pancreatic cancer progression by mediating miR-218-5p/LASP1 signaling axis. J Exp Clin Cancer Res. 2019; 38:347.
https://doi.org/10.1186/s13046-019-1340-7
PMID:31395079

43. Romero JM, Grünwald B, Jang GH, Bavi PP, Jhaveri A, Masoomian M, Fischer SE, Zhang A, Denroche RE, Lungu IM, De Luca A, Bartlett JM, Xu J, et al. A Four-Chemokine Signature Is Associated with a T-cell-Inflamed Phenotype in Primary and Metastatic Pancreatic Cancer. Clin Cancer Res. 2020; 26:1997–2010.
https://doi.org/10.1158/1078-0432.CCR-19-2803
PMID:31964786

44. Wang J, Jia Y, Wang N, Zhang X, Tan B, Zhang G, Cheng Y. The clinical significance of tumor-infiltrating neutrophils and neutrophil-to-CD8+ lymphocyte ratio in patients with resectable esophageal squamous cell carcinoma. J Transl Med. 2014; 12:7.
https://doi.org/10.1186/1479-5876-12-7
PMID:24397835

45. Kumar V, Patel S, Tcyganov E, Gabrilovich DI. The Nature of Myeloid-Derived Suppressor Cells in the Tumor Microenvironment. Trends Immunol. 2016; 37:208–20.
https://doi.org/10.1016/j.it.2016.01.004
PMID:26858199

46. Condamine T, Dominguez GA, Youn JI, Kossenkov AV, Mony S, Alicea-Torres K, Tcyganov E, Hashimoto A, Nefedova Y, Lin C, Partlova S, Garfall A, Vogl DT, et al. Lectin-type oxidized LDL receptor-1 distinguishes population of human polymorphonuclear myeloid-derived suppressor cells in cancer patients. Sci Immunol. 2016; 1:aaf8943.
https://doi.org/10.1126/sciimmunol.aaf8943
PMID:28417112

47. Lu C, Rong D, Zhang B, Zheng W, Wang X, Chen Z, Tang W. Current perspectives on the immunosuppressive tumor microenvironment in hepatocellular carcinoma: challenges and opportunities. Mol Cancer. 2019; 18:130.
https://doi.org/10.1186/s12943-019-1047-6
PMID:31464625

48. Borst J, Ahrends T, Bąbała N, Melief CJM, Kastenmüller W. CD4+ T cell help in cancer immunology and immunotherapy. Nat Rev Immunol. 2018; 18:635–47.
https://doi.org/10.1038/s41577-018-0044-0
PMID:30057419

49. Huang X, Zhang G, Tang TY, Gao X, Liang TB. Personalized pancreatic cancer therapy: from the perspective of mRNA vaccine. Mil Med Res. 2022; 9:53.
https://doi.org/10.1186/s40779-022-00416-w

PMID:36224645

50. Dong S, Huang F, Zhang H, Chen Q. Overexpression of BUB1B, CCNA2, CDC20, and CDK1 in tumor tissues predicts poor survival in pancreatic ductal adenocarcinoma. Biosci Rep. 2019; 39:BSR20182306.
https://doi.org/10.1042/BSR20182306
PMID:30765611

51. Lennon S, Oweida A, Milner D, Phan AV, Bhatia S, Van Court B, Darragh L, Mueller AC, Raben D, Martínez-Torrecuadrada JL, Pitts TM, Somerset H, Jordan KR, et al. Pancreatic Tumor Microenvironment Modulation by EphB4-ephrinB2 Inhibition and Radiation Combination. Clin Cancer Res. 2019; 25:3352–65.
https://doi.org/10.1158/1078-0432.CCR-18-2811
PMID:30944125

52. Zhang J, Tian Y, Mo S, Fu X. Overexpressing PLOD Family Genes Predict Poor Prognosis in Pancreatic Cancer. Int J Gen Med. 2022; 15:3077–96.
https://doi.org/10.2147/IJGM.S341332
PMID:35330878

53. Zhang Z, Cheng L, Li J, Qiao Q, Karki A, Allison DB, Shaker N, Li K, Utturkar SM, Atallah Lanman NM, Rao X, Rychahou P, He D, et al. Targeting Plk1 Sensitizes Pancreatic Cancer to Immune Checkpoint Therapy. Cancer Res. 2022; 82:3532–48.
https://doi.org/10.1158/0008-5472.CAN-22-0018
PMID:35950917

54. Bao L, Sun K, Zhang X. PANX1 is a potential prognostic biomarker associated with immune infiltration in pancreatic adenocarcinoma: A pan-cancer analysis. Channels (Austin). 2021; 15:680–96.
https://doi.org/10.1080/19336950.2021.2004758
PMID:34796785

55. Zhang H, Zhang X, Li X, Meng WB, Bai ZT, Rui SZ, Wang ZF, Zhou WC, Jin XD. Effect of CCNB1 silencing on cell cycle, senescence, and apoptosis through the p53 signaling pathway in pancreatic cancer. J Cell Physiol. 2018; 234:619–31.
https://doi.org/10.1002/jcp.26816
PMID:30069972

56. Xiao Y, Xu G, Cloyd JM, Du S, Mao Y, Pawlik TM. Predicting Novel Drug Candidates for Pancreatic Neuroendocrine Tumors via Gene Signature Comparison and Connectivity Mapping. J Gastrointest Surg. 2022; 26:1670–8.
https://doi.org/10.1007/s11605-022-05337-6
PMID:35508682

57. Nicolle R, Raffenne J, Paradis V, Couvelard A, de Reynies A, Blum Y, Cros J. Prognostic Biomarkers in Pancreatic Cancer: Avoiding Errata When Using the TCGA Dataset. Cancers (Basel). 2019; 11:126.
https://doi.org/10.3390/cancers11010126
PMID:30669703

58. 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

59. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15:550.
https://doi.org/10.1186/s13059-014-0550-8
PMID:25516281

60. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010; 26:1572–3.
https://doi.org/10.1093/bioinformatics/btq170
PMID:20427518

61. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16:284–7.
https://doi.org/10.1089/omi.2011.0118
PMID:22455463

62. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pancancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 2017; 18:248–62.
https://doi.org/10.1016/j.celrep.2016.12.019
PMID:28052254

63. 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

64. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, Liu J, Freeman GJ, Brown MA, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018; 24:1550–8.
https://doi.org/10.1038/s41591-018-0136-1
PMID:30127393

65. Brunson JC. ggalluvial: Layered Grammar for Alluvial Plots. J Open Source Softw. 2020; 5:2017.
https://doi.org/10.21105/joss.02017 PMID:36919162

66. Wang C, Chen Y, Xinpeng Y, Xu R, Song J, Ruze R, Xu Q, Zhao Y. Construction of immune-related signature and identification of S100A14 determining immune-suppressive microenvironment in pancreatic cancer. BMC Cancer. 2022; 22:879.
https://doi.org/10.1186/s12885-022-09927-0
PMID:35953822

67. Tao S, Tian L, Wang X, Shou Y. A pyroptosis-related gene signature for prognosis and immune microenvironment of pancreatic cancer. Front Genet. 2022; 13:817919.
https://doi.org/10.3389/fgene.2022.817919
PMID:36118860

68. Dai L, Mugaanyi J, Cai X, Lu C, Lu C. Pancreatic adenocarcinoma associated immune-gene signature as a novo risk factor for clinical prognosis prediction in hepatocellular carcinoma. Sci Rep. 2022; 12:11944.
https://doi.org/10.1038/s41598-022-16155-w
PMID:35831362

69. Nicolle R, Blum Y, Duconseil P, Vanbrugghe C, Brandone N, Poizat F, Roques J, Bigonnet M, Gayet O, Rubis M, Elarouci N, Armenoult L, Ayadi M, et al, and BACAP Consortium. Establishment of a pancreatic adenocarcinoma molecular gradient (PAMG) that predicts the clinical outcome of pancreatic cancer. EBioMedicine. 2020; 57:102858.
https://doi.org/10.1016/j.ebiom.2020.102858
PMID:32629389

70. Rashid NU, Peng XL, Jin C, Moffitt RA, Volmar KE, Belt BA, Panni RZ, Nywening TM, Herrera SG, Moore KJ, Hennessey SG, Morrison AB, Kawalerski R, et al. Purity Independent Subtyping of Tumors (PurIST), A Clinically Robust, Single-sample Classifier for Tumor Subtyping in Pancreatic Cancer. Clin Cancer Res. 2020; 26:82–92.
https://doi.org/10.1158/1078-0432.CCR-19-1467
PMID:31754050

71. Zeng D, Wu J, Luo H, Li Y, Xiao J, Peng J, Ye Z, Zhou R, Yu Y, Wang G, Huang N, Wu J, Rong X, et al. Tumor microenvironment evaluation promotes precise checkpoint immunotherapy of advanced gastric cancer. J Immunother Cancer. 2021; 9:e002467.
https://doi.org/10.1136/jitc-2021-002467
PMID:34376552

72. Hoshida Y, Brunet JP, Tamayo P, Golub TR, Mesirov JP. Subclass mapping: identifying common subtypes in independent disease data sets. PLoS One. 2007; 2:e1195.
https://doi.org/10.1371/journal.pone.0001195
PMID:18030330

73. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, et al. Integrated analysis of multimodal single-cell data. Cell. 2021; 184:3573–87.e29.
https://doi.org/10.1016/j.cell.2021.04.048
PMID:34062119

74. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data

Using Artificial Nearest Neighbors. Cell Syst. 2019; 8:329–37.e4.
https://doi.org/10.1016/j.cels.2019.03.003
PMID:30954475

75. Peng J, Sun BF, Chen CY, Zhou JY, Chen YS, Chen H, Liu L, Huang D, Jiang J, Cui GS, Yang Y, Wang W, Guo D, et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. Cell Res. 2019; 29:725–38.
https://doi.org/10.1038/s41422-019-0195-y
PMID:31273297

76. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte AJ, Bhattacharya M. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019; 20:163–72.
https://doi.org/10.1038/s41590-018-0276-y
PMID:30643263

77. Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, Chen J, Jiang W, Yang K, Ou Q, Li X, Wang P, Zhang Y. CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. Nucleic Acids Res. 2023; 51:D870–6.
https://doi.org/10.1093/nar/gkac947 PMID:36300619

78. Schlesinger Y, Yosefov-Levi O, Kolodkin-Gal D, Granit RZ, Peters L, Kalifa R, Xia L, Nasereddin A, Shiff I, Amran O, Nevo Y, Elgavish S, Atlan K, et al. Single-cell transcriptomes of pancreatic preinvasive lesions and cancer reveal acinar metaplastic cells' heterogeneity. Nat Commun. 2020; 11:4516.
https://doi.org/10.1038/s41467-020-18207-z
PMID:32908137

79. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, Nie Q. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021; 12:1088.
https://doi.org/10.1038/s41467-021-21246-9
PMID:33597522

80. Wang Y, Wang R, Zhang S, Song S, Jiang C, Han G, Wang M, Ajani J, Futreal A, Wang L. iTALK: an R Package to Characterize and Illustrate Intercellular Communication. BioRxiv. 2019; 507871.
https://doi.org/10.1101/507871

81. Drier Y, Sheffer M, Domany E. Pathway-based personalized analysis of cancer. Proc Natl Acad Sci USA. 2013; 110:6388–93.
https://doi.org/10.1073/pnas.1219651110
PMID:23547110
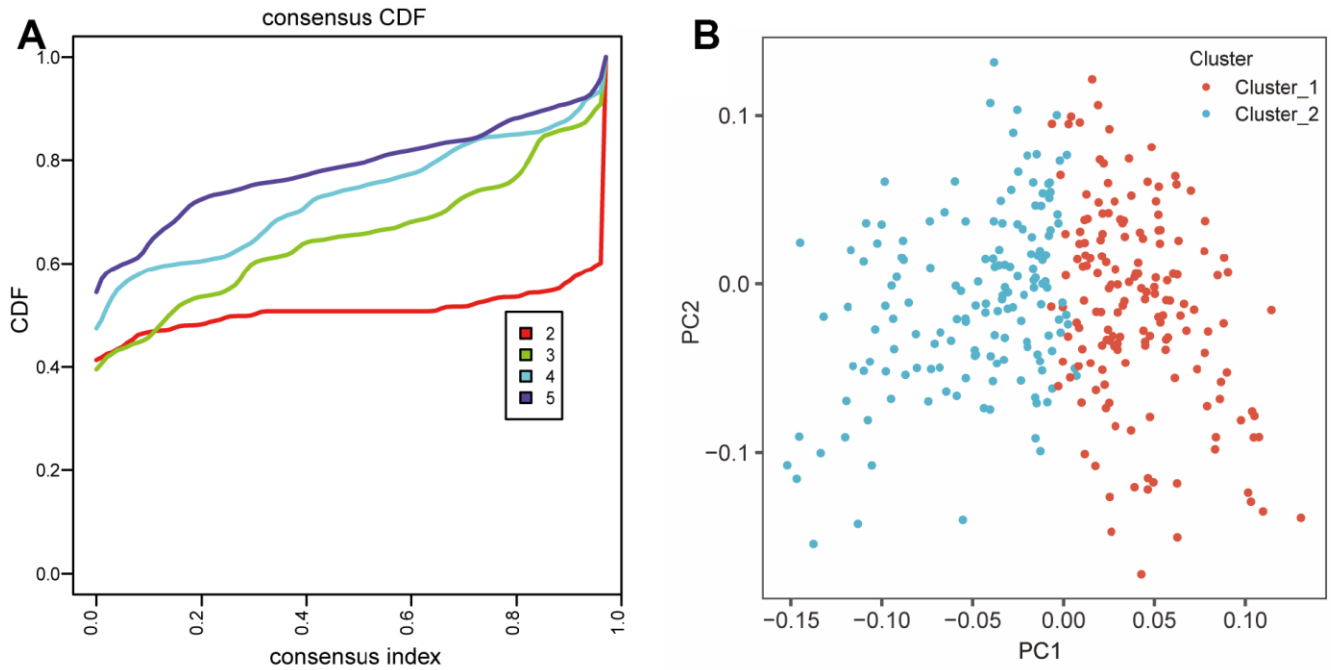
Supplementary Figure 1. Identification of TIME-related subtypes. (**A**) The CDF plot of consensus clustering. (**B**) The PCA plot demonstrated the differences between the two clusters.
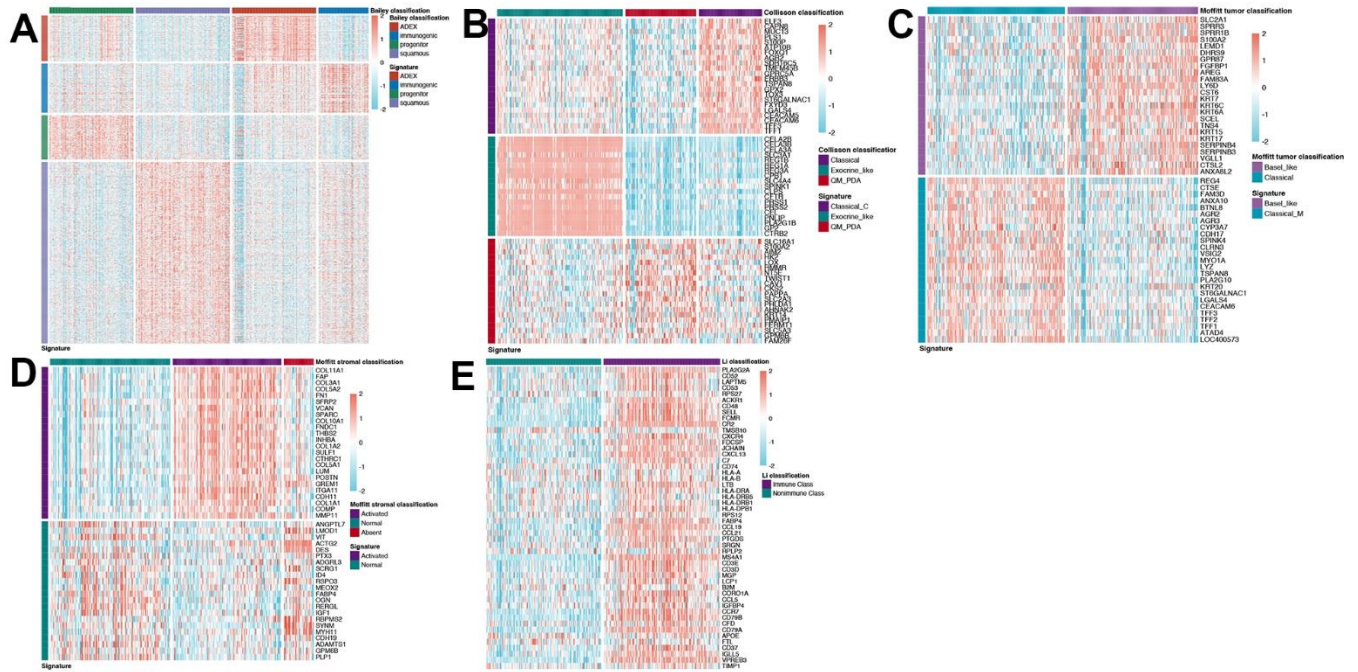
Supplementary Figure 2. Published PC subtypes prediction. Heatmap showed that we defined Bailey's classification (**A**), Collisson's classification (**B**), Moffitt's tumor classification (**C**), Moffitt's stromal classification (**D**) and Li's classification (**E**) based on the published classifier exemplar genes.
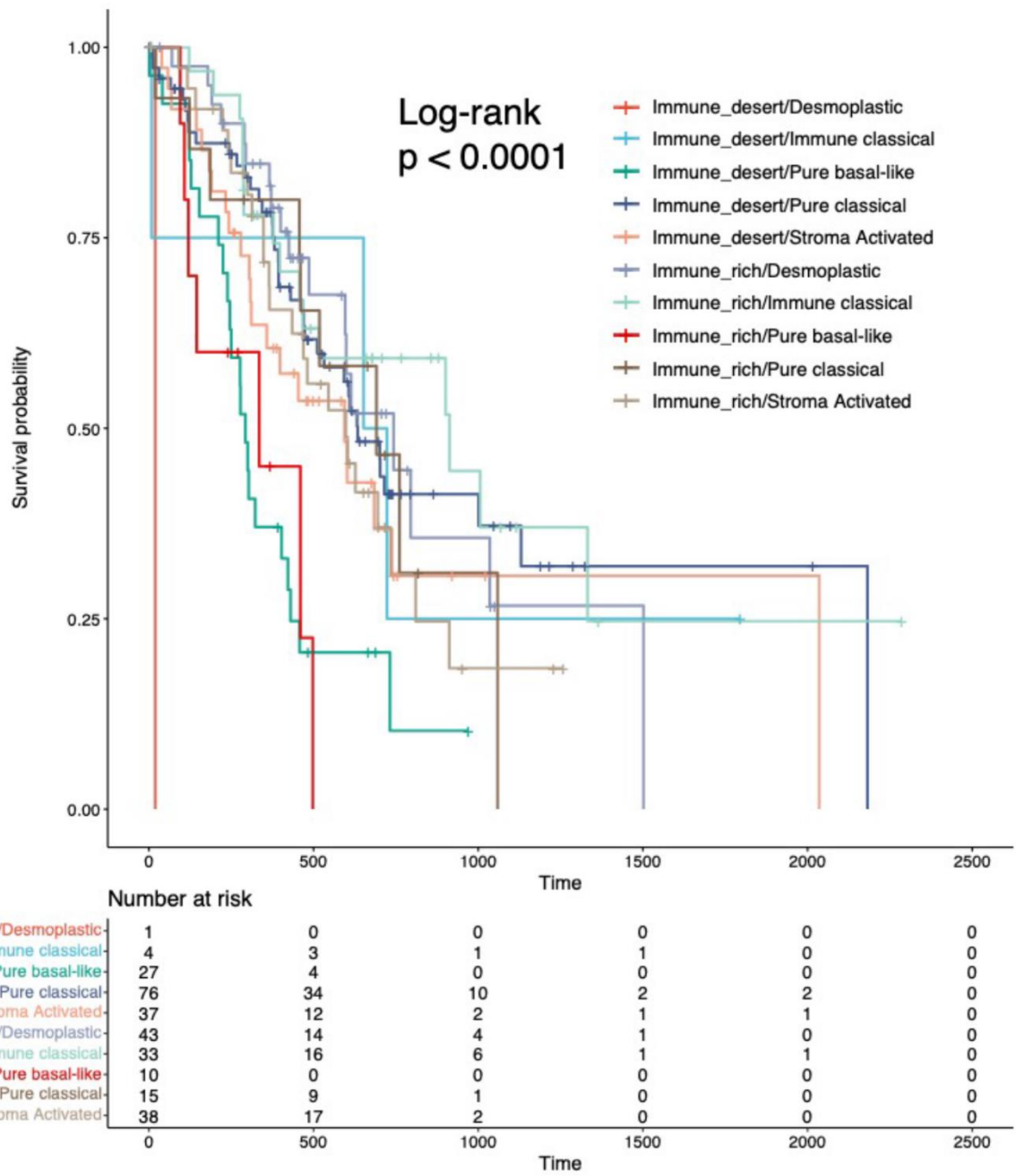
Supplementary Figure 3. Kaplan-Meier curves of overall survival were plotted according to the TIME subtype and Puleo's classification.
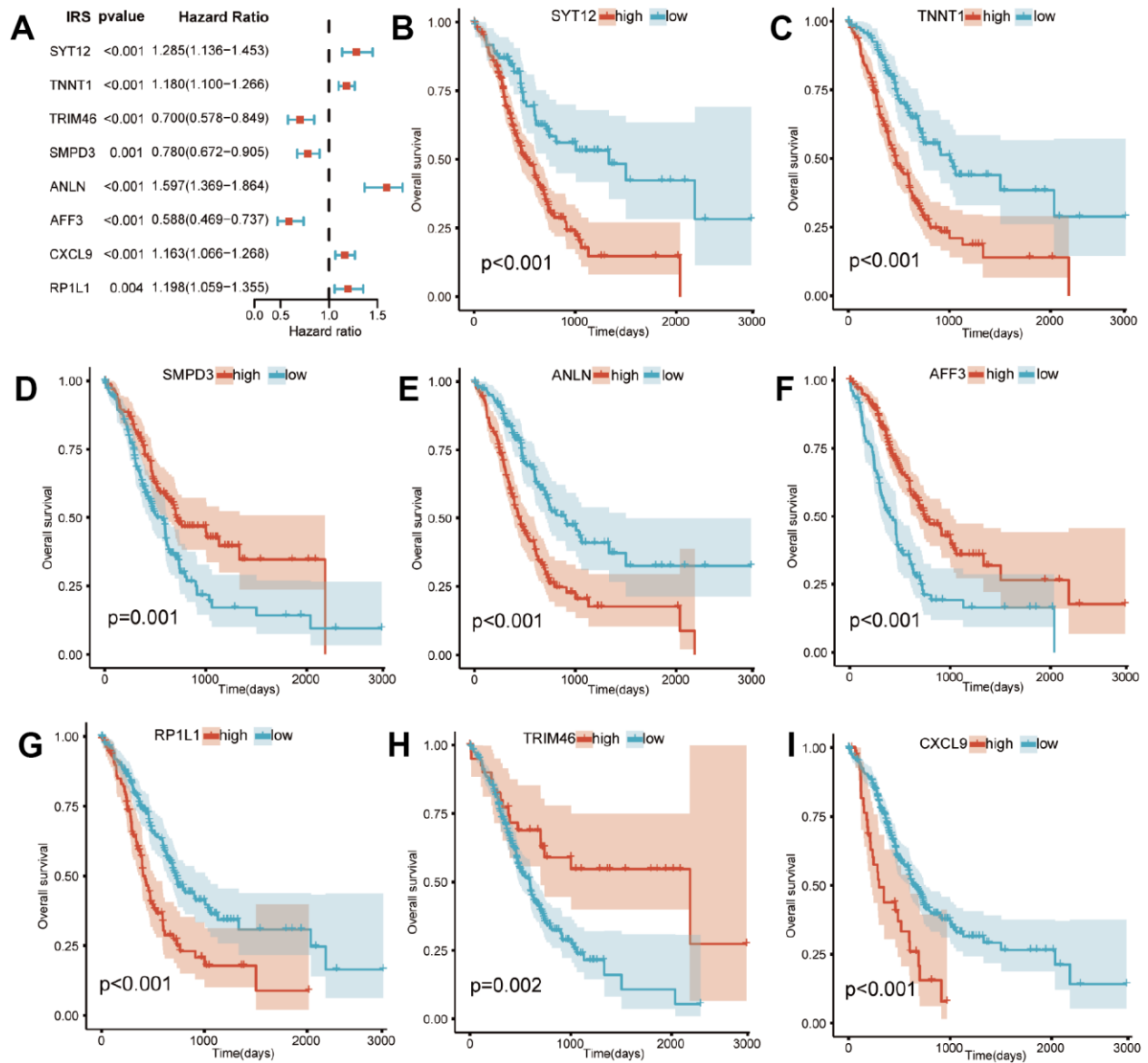
Supplementary Figure 4. Validation of IRS. (**A**) Forest plot of univariate cox analysis based on the training set. (**B**–**I**) The survival analysis is based on the expression of IRS genes.
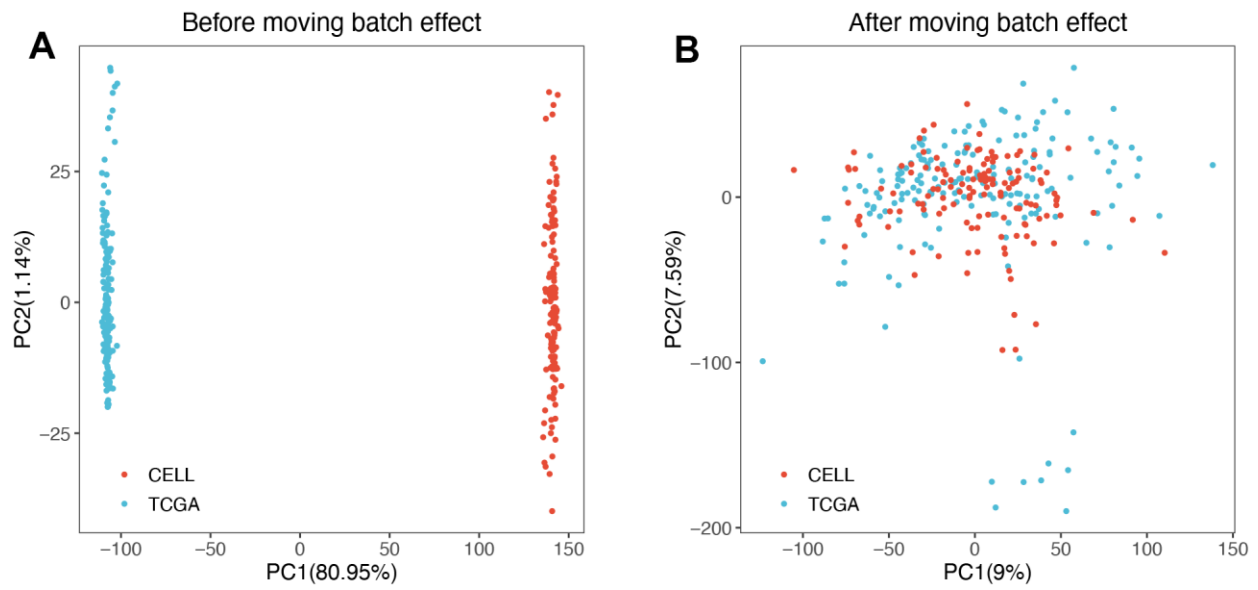
Supplementary Figure 5. Evaluation of the results of batch effect correction. The principal component analysis (PCA) before (**A**) and after (**B**) batch effect correction on Meta-cohort.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1–5.

Supplementary Table 1. Differential genes between immune_high and immune_low samples.

Supplementary Table 2. Differential genes between stromal_high and stromal_low samples.

Supplementary Table 3. Differential genes between pancreatic cancer and normal pancreatic samples.

Supplementary Table 4. The 1612 IRG signature.

Supplementary Table 5. Published PC classifications prediction.

Supplementary Table 6. Comparison of the TIME subtype with other pancreatic molecular classifications.

Classifications		Overall	Immune_desert	Immune_rich	P-value
		284	145	139	
TIME subtype (%)	Immune_desert	145 (51.06)	145 (100.00)	0 (0.00)	<0.0001
	Immune_rich	139 (48.94)	0 (0.00)	139 (100.00)	
Bailey's classification (%)	ADEX	85 (29.93)	55 (37.93)	30 (21.58)	<0.0001
	Immunogenic	54 (19.01)	5 (3.45)	49 (35.25)	
	Progenitor	61 (21.48)	59 (40.69)	2 (1.44)	
	Squamous	84 (29.58)	26 (17.93)	58 (41.73)	
Collisson's classification (%)	Classical	69 (24.30)	49 (33.79)	20 (14.39)	0.0006
	Exocrine_like	138 (48.59)	60 (41.38)	78 (56.12)	
	QM_PDA	77 (27.11)	36 (24.83)	41 (29.50)	
Moffitt's tumor classification (%)	Basel_like	133 (46.83)	70 (48.28)	63 (45.32)	0.7044
	Classical	151 (53.17)	75 (51.72)	76 (54.68)	
Moffitt's stromal classification (%)	Absent	46 (16.20)	30 (20.69)	16 (11.51)	0.0101
	Activated	115 (40.49)	64 (44.14)	51 (36.69)	
	Normal	123 (43.31)	51 (35.17)	72 (51.80)	
Puleo's classification (%)	Desmoplastic	44 (15.49)	1 (0.69)	43 (30.94)	<0.0001
	Immune classical	37 (13.03)	4 (2.76)	33 (23.74)	
	Pure basal-like	37 (13.03)	27 (18.62)	10 (7.19)	
	Pure classical	91 (32.04)	76 (52.41)	15 (10.79)	
	Stroma Activated	75 (26.41)	37 (25.52)	38 (27.34)	
Li's classification (%)	Immune Class	143 (50.35)	48 (33.10)	95 (68.35)	<0.0001
	Nonimmune Class	141 (49.65)	97 (66.90)	44 (31.65)	

Supplementary Table 7. CMap score of candidate potential agents.

Compounds	CMap score
Canertinib	-80.94
PP-1	-64.5
YM-976	-0.74
CHIR-98014	NA
GW-788388	NA
Brigatinib	NA
Vincamine	-21.27

Supplementary Table 8. AUC values of two PC cell lines.

DepMap_ID	Cell line	RRID	IRS score	Group	Canertinib	PP-1	YM-976	Vincamine
ACH-000417	Capan-2	CVCL_0026	2.96757	IRS_high	0.77602	1.3768761	1.216856	0.8804061
ACH-000107	Panc 08.13	CVCL_1638	2.546593	IRS_low	0.96352	0.9137654	1.246992	0.9541501

Supplementary Table 9. Details of 3 published signatures in PC.

Model	PMID	Type	Author	Coef	Gene
Model-1	35953822	mRNA	Wang C	0.0893	ALB
Model-1	35953822	mRNA	Wang C	0.1935	CXCL10
Model-1	35953822	mRNA	Wang C	0.0635	IAPP
Model-1	35953822	mRNA	Wang C	-0.2633	LIFR
Model-1	35953822	mRNA	Wang C	-0.2338	LYZ
Model-1	35953822	mRNA	Wang C	0.2893	MET
Model-1	35953822	mRNA	Wang C	0.2177	S100A14
Model-2	36118860	mRNA	Tao S	0.2302	GSDMC
Model-2	36118860	mRNA	Tao S	-0.4664	ELANE
Model-2	36118860	mRNA	Tao S	0.3341	IL18
Model-2	36118860	mRNA	Tao S	-0.4324	NLRP1
Model-2	36118860	mRNA	Tao S	0.1297	NLRP2
Model-3	35831362	mRNA	Dai L	0.132	S100P
Model-3	35831362	mRNA	Dai L	0.098	S100A2
Model-3	35831362	mRNA	Dai L	0.095	MMP12

Supplementary Table 10. The forward and reverse primers of hub genes.

Name	Primer(5'to3')
Homo-SYT12-F	CAGAATACCATCTGAGCGTCATC
Homo-SYT12-R	TAGTCGTAATTGGGGAACGGA
Homo-TNNT1-F	TGATCCCGCCAAAGATCCC
Homo-TNNT1-R	TCTTCCGCTGCTCGAAATGTA
Homo-TRIM46-F	TTCCGACCCAAGGGCCTTAT
Homo-TRIM46-R	AGAGTTGACATACCAGGCGTT
Homo-SMPD3-F	GCTGCCCTTTGCGTTTCTC
Homo-SMPD3-R	TCCAGCCGTGAATAGATGTAGG
Homo-ANLN-F	TGCCAGGCGAGAGAATCTTC
Homo-ANLN-R	CGCTTAGCATGAGTCATAGACCT
Homo-AFF3-F	GCCCTACAAGACTAACAAGGGG
Homo-AFF3-R	ACTCCAACGAGATGACTCTGAT
Homo-CXCL9-F	CCAGTAGTGAGAAAGGGTCGC
Homo-CXCL9-R	AGGGCTTGGGGCAAATTGTT
Homo-RP1L1-F	TCCCGGAAGAGTCTTAAAACCC
Homo-RP1L1-R	GCGGCCAGGTTCCTAGTATTC

Supplementary File

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