

RESEARCH

Open Access



# Construction of immune-related signature and identification of S100A14 determining immune-suppressive microenvironment in pancreatic cancer

Chengcheng Wang<sup>†</sup>, Yuan Chen<sup>\*†</sup>, Yin Xinpeng<sup>†</sup>, Ruiyuan Xu, Jianlu Song, Rexiati Ruze, Qiang Xu<sup>\*</sup> and Yupei Zhao<sup>\*</sup>

## Abstract

Pancreatic cancer (PC) is a highly lethal and aggressive disease with its incidence and mortality quite discouraging. A robust prognostic signature and novel biomarkers are urgently needed for accurate stratification of the patients and optimization of clinical decision-making. Since the critical role of immune microenvironment in the progression of PC, a prognostic signature based on seven immune-related genes was established, which was validated in The Cancer Genome Atlas (TCGA) training set, TCGA testing set, TCGA entire set and GSE71729 set. Furthermore, S100A14 (S100 Calcium Binding Protein A14) was identified as the gene occupying the most paramount position in risk signature. According to the GSEA, CIBERSORT and ESTIMATE algorithm, S100A14 was mainly associated with lower proportion of CD8 + T cells and higher proportion of M0 macrophages in PC tissue. Meanwhile, analysis of single-cell dataset CRA001160 revealed a significant negative correlation between S100A14 expression in PC cells and CD8 + T cell infiltration, which was further confirmed by tissue microenvironment landscape imaging and machine learning-based analysis in our own PUMCH cohort. Additionally, analysis of a pan-pancreatic cancer cell line illustrated that S100A14 might inhibit CD8 + T cell activation via the upregulation of PD-L1 expression in PC cells, which was also verified by the immunohistochemical results of PUMCH cohort. Finally, tumor mutation burden analysis and immunophenoscore algorithm revealed that patients with high S100A14 expression had a higher probability of responding to immunotherapy. In conclusion, our study established an efficient immune-related prediction model and identified the potential role of S100A14 in regulating the immune microenvironment and serving as a biomarker for immunotherapy efficacy prediction.

**Keywords:** S100A14, Tumor immune microenvironment, CD8 + T cells, Prognostic signature, Immunotherapy, PD-L1, Pancreatic cancer

## Introduction

Pancreatic cancer (PC) is one of the most lethal malignancies, with a five-year survival rate of only 11% in the United States [1]. The global burden of PC has increased dramatically over the past few decades and is expected to become the second most common cause of cancer-related mortality by 2030 after lung cancer [2, 3]. Beyond the scarcity of sensitive screening methods and the

<sup>†</sup>Chengcheng Wang, Yuan Chen and Yin Xinpeng contributed equally to this work.

\*Correspondence: chenyan19961023@hotmail.com; xuqiang@pumch.cn; zhao8028@263.net

Department of General Surgery, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100023, PR China



emergence of chemoresistance, this dismal situation is largely attributable to the lack of effective risk prediction signature and biomarkers [4], which hinders the individualized treatment of PC to some extent. Therefore, it is of great significance to establish validated prediction signature and screen potential novel biomarkers for the accurate assessment of patient's prognosis and optimization of clinical decision-making.

Tumor immune microenvironment (TIME), defined as all immunological components within tumors, mainly comprises innate immune cells, adaptive immune cells, extracellular immune factors and cell surface molecules [5, 6]. Extensive studies have shown that tumor immune microenvironment possessed profound effect on tumor development [7, 8]. For example, IL-10 prevents dendritic cell-mediated CD8+T cell apoptosis, thus potentiating CD8+T cell-mediated antitumor immunity [9]. And transforming growth factor beta (TGF $\beta$ ) functions as an immunosuppressive factor through inhibition of CXCR3 in CD8+T cells, thereby limiting their trafficking into tumors [10]. Therefore, we consider a risk prediction signature based on immune-related genes (IRGs) to better predict the prognosis of PC and assist clinical decision-making. Furthermore, the most paramount gene in the risk signature was further explored, as well as its potential mechanisms and ability to serve as a novel biomarker for the efficiency of immunotherapy.

In the present study, we constructed a prediction signature of PC prognosis consisting of seven IRGs and the corresponding nomogram, which was validated in the training set, testing set, entire set and GSE71729 set respectively. Furthermore, S100A14 (S100 Calcium Binding Protein A14), a highly conserved elongation factor (EF)-hand calcium-binding protein, was identified as the gene occupying the most paramount position in the risk signature. GSEA, CIBERSORT and ESTIMATE algorithm suggested that S100A14 might be closely related to the PC immune microenvironment. Meanwhile, analysis of the single-cell dataset CRA001160 illustrated that CD8+T cells might be the primary target of S100A14 function. Subsequently, tissue microenvironment landscape imaging and machine learning-based analysis were applied to visualize and analyze the immune microenvironment of PC patients in our independent cohort (PUMCH cohort). And the results showed a significant negative correlation between S100A14 expression and CD8+T cell infiltration in tumor. In addition, a pan-pancreatic cancer cell analysis of CCLE database and iRegulon analysis suggested that S100A14-FOXO1-PD-L1 pathway may be a potential signal axes inhibiting the activation of CD8+T cells, which was also verified by immunohistochemical results of PUMCH cohort. Finally, it was predicted that patients with high S100A14

expression had a higher probability of responding to anti-PD-1 and anti-CTLA-4 therapy, suggesting that S100A14 may serve as a potential biomarker for predicting PC immunotherapy efficiency.

## Materials and methods

### Datasets sources and processing

IRGs were extracted and integrated from the ImmPort database (<https://immport.niaid.nih.gov>;  $\leq$  Mar 1, 2021) [11]. The RNA sequencing (RNA-seq) data, mutation profile and clinicopathological information of the patients were downloaded from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov>;  $\leq$  March 1, 2021) and the University of California Santa Cruz (UCSC) Xena website (<https://xenabrowser.net/datapages/>;  $\leq$  March 1, 2021) (Table 1, detailed in Table S1). Patients with incomplete clinical information and follow-up period less than 30 days were excluded. Finally, 166 PC patients were included in the study. The gene expression data were normalized to transcripts per million (TPM) values and transformed to  $\log_2(\text{TPM} + 0.01)$  for further analysis unless otherwise noted.

In addition, GSE15471, GSE28735, GSE62165 and GSE71729 dataset were downloaded from Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) [12–15], which were performed on GPL570, GPL6244, GPL13667 and GPL20769 platform. Expression values were calculated using the robust multi-array average (RMA) algorithm except GSE71729. The normalized expression matrix of microarray data can be directly download from the GEO dataset. Proteomic data from Clinical Proteomic from Tumor Analysis Consortium (CPTAC) were analyzed in the UALCAN database (<http://ualcan.path.uab.edu/index.html>) [16, 17].

Furthermore, the processed single-cell dataset CRA001160 (including over 57,000 cells from 24 primary PDAC samples and 11 normal samples) was downloaded from Tumor Immune Single-cell Hub (TISCH) database (<http://tisch.comp-genomics.org/>) [18, 19] (Table 2, detailed in Table S2). And the RNA-seq data and proteomics data of all pancreatic cell lines was extracted from the Cancer Cell Line Encyclopedia (CCLE) database (<https://portals.broadinstitute.org/ccle>) [20].

### Establishment and validation of the prognostic signature based on IRGs

Limma package was applied to screen differentially expressed genes (DEGs) in GSE15471, GSE28735, GSE62165 and GSE71729 datasets respectively [21].  $|\text{Fold Change}| > 1.5$  and false discovery rate (FDR)  $< 0.05$  were set as the cutoffs for the DEGs. The intersection of the four differential gene sets was then used for least absolute shrinkage and selection operator (LASSO) regression analysis and multivariate Cox regression

**Table 1** Clinical and pathologic information of training set, testing set and entire set

Character	Training set		Testing set		Entire set	
	NUMBER	%	NUMBER	%	NUMBER	%
<b>Age</b>						
<b>Median</b>	65		64.5		65	
<b>Range</b>	35–88		39–81		35–88	
<b>Os (m)</b>						
<b>Median</b>	15.5		15.9		15.6	
<b>Range</b>	1.0–76.2		3.2–91.4		1.0–91.4	
<b>Status</b>						
<b>Alive</b>	57	49.14	19	38.00	76	45.78
<b>Dead</b>	59	50.86	31	62.00	90	54.22
<b>Gender</b>						
<b>Male</b>	65	56.03	25	50.00	90	54.22
<b>Female</b>	51	43.97	25	50.00	76	45.78
<b>AJCC_Stage</b>						
<b>I</b>	12	10.35	6	12.00	18	10.84
<b>II</b>	102	87.93	39	78.00	141	84.94
<b>III</b>	1	0.86	2	4.00	3	1.81
<b>IV</b>	1	0.86	3	6.00	4	2.41
<b>Grade</b>						
<b>G1</b>	17	14.66	9	18.00	26	15.66
<b>G2</b>	63	54.31	28	56.00	91	54.82
<b>G3</b>	34	29.31	13	26.00	47	28.31
<b>G4</b>	2	1.72	0	0.00	2	1.21
<b>T stage</b>						
<b>T1</b>	4	3.45	2	4.00	6	3.61
<b>T2</b>	16	13.79	5	10.00	21	12.65
<b>T3</b>	95	81.90	41	82.00	136	81.93
<b>T4</b>	1	0.86	2	4.00	3	1.81
<b>N stage</b>						
<b>N0</b>	31	26.72	14	28.00	45	27.11
<b>N1</b>	84	72.42	34	68.00	118	71.08
<b>Nx</b>	1	0.86	2	4.00	3	1.81
<b>M stage</b>						
<b>M0</b>	59	50.86	17	34.00	76	45.78
<b>M1</b>	1	0.86	3	6.00	4	2.41
<b>MX</b>	56	48.28	30	60.00	86	51.81

analysis to obtain most effective prognostic model. According to the regression coefficient and the corresponding gene expression in the signature, the risk signature was established as follows: Risk score =  $(\text{expr}_{\text{gene1}} \times \text{Coef}_{\text{gene1}}) + (\text{expr}_{\text{gene2}} \times \text{Coef}_{\text{gene2}}) + \dots + (\text{expr}_{\text{genen}} \times \text{Coef}_{\text{genen}})$ . The Kaplan–Meier (KM) survival analysis, time-dependent receiver operating characteristic (ROC) curves and survival point diagram were utilized to evaluate the predictive efficiency of the risk signature. Meanwhile, univariate and multivariate Cox regression analysis were performed to determine the independence of

the risk signature. Gene mutation status of the seven genes was obtained from the cBioportal database (<http://www.cbioportal.org/>) [22], and protein expression images in normal and tumor tissues was downloaded from the Human Protein Atlas (HPA) database (<https://www.proteinatlas.org/>).

In addition, the nomogram was also established to predict the 1-, 2-, 3-year survival probability based on the risk score and other clinicopathological characteristics. Univariate and multivariate Cox regression analysis were conducted to screen for valid clinicopathological factors

**Table 2** Clinical and pathologic information of CRA001160 set and PUMCH cohort

Character	CRA001160 Set		PUMCH COHORT	
	Number	%	Number	%
<b>Age</b>				
<b>Median</b>	59		65.5	
<b>Range</b>	36–72		38–80	
<b>Gender</b>				
<b>Male</b>	11	45.83	17	44.74
<b>Female</b>	13	54.17	21	55.26
<b>AJCC_Stage</b>				
<b>I</b>	9	37.50	9	23.68
<b>II</b>	12	50.00	21	55.26
<b>III</b>	3	12.50	7	18.42
<b>IV</b>	0	0	1	2.64
<b>Diabetes</b>				
<b>Y</b>	10	41.67	19	50.00
<b>N</b>	14	58.33	19	50.00
<b>T stage</b>				
<b>T1</b>	4	16.67	5	13.16
<b>T2</b>	14	58.33	22	57.90
<b>T3</b>	5	20.83	9	23.68
<b>T4</b>	1	4.17	2	5.26
<b>N Stage</b>				
<b>N0</b>	10	41.67	15	39.47
<b>N1</b>	11	45.83	19	50.00
<b>N2</b>	3	12.5	4	10.53
<b>M Stage</b>				
<b>M0</b>	24	100.00	37	97.36
<b>M1</b>	0	0.00	1	2.64

included in the nomogram, and the corresponding time-dependent ROC curves, C-index calculation and calibration curves were applied to assess the efficiency of the nomogram.

#### GSEA and iRegulon analysis for DEGs between S100A14 high and low expression group

Patients in the TCGA entire set were divided into S100A14 high expression group ( $n=83$ ) and S100A14 low expression group ( $n=83$ ) according to the median expression of S100A14. DEGs between the two groups were obtained by edgeR package in R language ( $|\text{Log}_2\text{FC}|>2$  and  $\text{FDR}<0.001$ ), followed by GSEA analysis to determine the potential role of S100A14 in PC development [23]. The ALL ontology of the DEGs was analyzed by Gene Ontology (GO) [24], while pathway enrichment was analyzed by the Kyoto Encyclopedia of Genes and Genomes (KEGG) [25]. The number of random sample permutations was set at 1000, and NOM

$p$ -value  $<0.05$  and FDR  $q$ -value  $<0.25$  were set as the significance threshold. For transcription factors prediction, DEGs between S100A14 high and low expression group were analyzed by iRegulon tools in Cytoscape (version v3.7.1) [26].

#### Estimation of tumor immune microenvironment

The ESTIMATE algorithm was able to estimate the content of mesenchymal and immune cells in PC tissue based on RNA-seq data [27], resulting in stromal score, immune score and estimate score. And the CIBERSORT algorithm was performed to quantify the relative abundance of 22 immune cells infiltrated in tumor microenvironment by a deconvolution algorithm [28]. These two algorithms were applied in the TCGA entire set and GSE71729 dataset respectively to evaluate the relationship between S100A14 expression and immune infiltration in the two datasets.

#### Tumor mutation burden analysis

The mutation profile was acquired from TCGA data portal (<https://portal.gdc.cancer.gov/>;  $\leq$  March 1, 2021). The landscape of somatic mutation in S100A14 low and high expression group was analyzed and visualized through “maftools” package in R language respectively [29]. Meanwhile, tumor mutation burden (TMB) and mutation frequency of top 10 genes were calculated and compared between S100A14 low and high expression groups.

#### Prediction of the patients’ response to immunotherapy

Immunophenoscore (IPS), a scoring system for the quantification of tumor immunogenicity, has been verified to positively correlated to the responding rate to immunotherapy of PC patients [30]. The IPS data of PC patients was extracted from The Cancer Immunome Atlas (<https://tcia.at/>) for the following analysis, including the scores for anti-PD-1 therapy, anti-CTLA-4 therapy and the combination of the two therapies.

#### Clinical specimens and immunohistochemical analysis

A total of 38 patients with primary PDAC who underwent surgical resection at Peking Union Medical College Hospital (PUMCH) were included in this study (PUMCH cohort, Apr. 2020–Nov. 2020). Informed consent was obtained from all patients, and this study was approved by the ethical committees of Peking Union Medical College Hospital. Manual staining was performed as the protocol previously described [31]. For primary antibody incubation of each patient, sections were incubated with rabbit monoclonal anti-PD-L1 antibody (1:200) (Abcam, ab205921) for 1 h.

### Cell culture

All pancreatic cancer cell lines and pancreatic normal ductal cell line were purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA). Meanwhile, all of the cells were authenticated by short tandem repeat (STR) analysis and regularly tested for mycoplasma contamination. BxPC-3 and HPAF-II cell lines were cultured in RPMI-1640 medium (Corning, #10-040-CV). CFPAC-1 and Capan-1 cell lines were cultured in Iscove's Modified Dulbecco Medium (IMDM; Corning, #15-016-CV). PANC-1, MIA PaCa-2, SW1990, PaTu 8902, AsPC-1, T3M4 and Panc 10.05 cell lines were cultured in high glucose Dulbecco's Modified Eagle Medium (DMEM; Corning, #10-013-CMR). All medium were supplemented with 10% fetal bovine serum (HyClone, #SH30073.03) and 1% Penicillin–Streptomycin (Life Technologies, #15,140-122). Cells were routinely maintained at 37°C with 5% CO<sub>2</sub>.

### RNA isolation and qRT-PCR

The specific operation process has been described in the previous article [32]. All the values were normalized to GAPDH, and the 2- $\Delta$ Ct method was used to quantify the fold change. The primer sequences used for qRT-PCR were as follows:

S100A14: Forward 5'- GAGACGCTGACCCCTTCTG-3',

Reverse 5'- CTTGGCCGCTTCTCCAATCA-3';

GAPDH: Forward 5'-GTCTCCTCTGACTTCAACAGCG-3',

Reverse 5'-ACCACCCTGTTGCTGTAGCCAA-3'.

### Western blot analysis

Western blot analysis was performed as described previously [33]. The blot was cut prior to hybridisation with antibodies during blotting. Primary antibody for S100A14 was purchased from ProteinTech (10,489-1-AP, ProteinTech). And primary antibody for  $\beta$ -actin antibody was purchased from LabLead (A0101, LabLead).

### scRNA-seq data quality control, dimension reduction and cell clustering

The single-cell dataset CRA001160 was already processed by the author [19]. The Seurat package implemented in R language was applied to conduct following analysis. Low quality cells (<200 genes/cell, <3 cells/gene, >5% mitochondrial genes, total expressed genes <200 and total expressed genes >7000) were removed. The gene expression profiles were then normalized and the top 2000 highly variable genes were generated to perform principal component analysis (PCA). Significant principal components were determined using JackStraw analysis and

Elbow plot. PCs 1 to 15 were used for graph-based clustering (res=1.0), which was visualized by the uniform manifold approximation and projection (UMAP) analysis. The cell types were annotated according to the markers provided by the author [19]. Meanwhile, T cells and macrophages were then extracted for further analysis.

### Tissue microenvironment landscape imaging and machine-learning based analysis

4  $\mu$ m sections from the PUMCH cohort were used for tissue microenvironment landscape analysis through multiplex immunohistochemical kit (Panovue Biological Technology, 0,081,100,100). Sections were deparaffinized and tissues were fixed with 10% formalin, followed by antigen retrieval in heated EDTA buffer (pH 9.0, OriGene Technologies, ZLI-9069) for 15 min. Each section was put through three sequential rounds of staining, which includes blocking, primary antibody incubation, secondary horseradish peroxidase-conjugated polymer incubation and covalent binding of a different fluorophore using tyramide signal amplification. Between the two rounds, an additional antigen retrieval in heated EDTA Buffer (pH 9.0) for 15 min was conducted to remove bound antibodies. After all three sequential reactions, sections were counterstained with DAPI and mounted with anti-fade mounting medium. Slides were imaged and analyzed using the Vectra Multispectral Imaging System version 2 (Perkin Elmer) and the supporting software. Filter cubes used for multispectral imaging were DAPI, opal540, opal620, opal690. The corresponding imaging channels and antibody incubation are shown in Table S3.

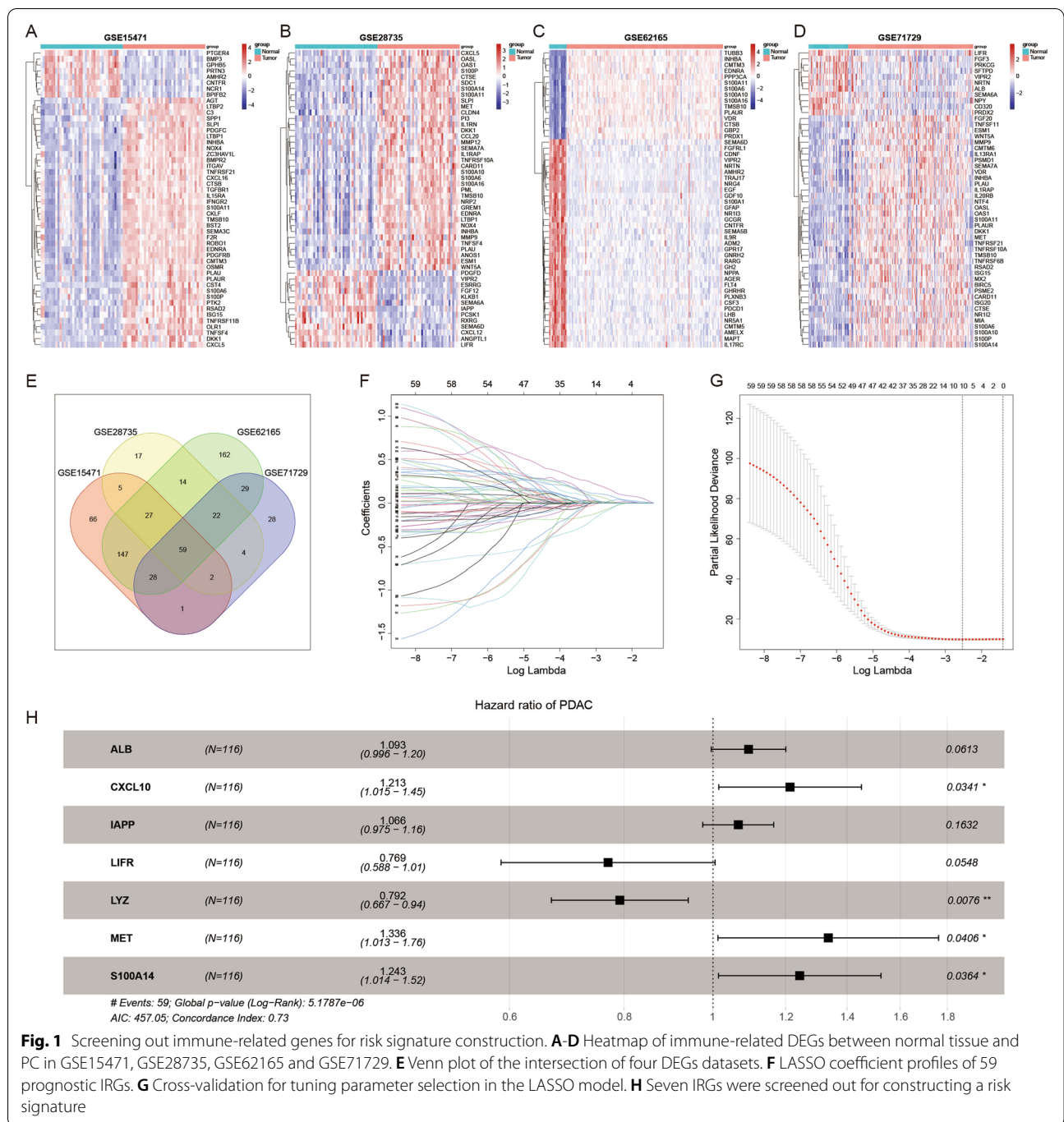
### Statistical analysis

All statistical analysis were performed using R software (version 4.1.0) and GraphPad Prism 8 (version 8.0.1), including DEGs analysis, LASSO regression analysis, multivariate Cox regression analysis, clinicopathological factor analysis, K-M survival analysis, ROC curve analysis and correlation analysis. For qRT-PCR, data are means  $\pm$  Standard Error of Mean (SEM) of three independent experiments. A two-sided *P* value <0.05 was regarded to be statistically significant.

## Results

### Seven immune-related genes were screened out for constructing the risk signature

The process of the whole analysis was illustrated in Figure S1. A total of 1793 IRGs were integrated from the ImmPort database (Table S4). First, DEGs between normal and tumor samples were analyzed by limma package in GSE15471, GSE28735, GSE62165 and GSE71729 datasets. |Fold Change| >1.5 and FDR <0.05 were regarded as statistically significant, and 50 genes with the most



significant differences in each dataset were shown in the heatmap (Fig. 1A-D).

By taking the intersection of these four differential gene sets, 59 common DEGs were obtained (Fig. 1E). Subsequently, LAASO regression analysis was applied to avoid overfitting problems and further screen the candidate genes (Fig. 1F-G, log(lambda.min)=-2.529272). And multivariate Cox analysis was then used to explore

an appropriate gene combination for establishing the risk signature for PC patients. Finally, seven genes (ALB, CXCL10, IAPP, LIFR, LYZ, MET, S100A14) were screened out (Fig. 1H). Among them, LIFR and LYZ were protective factors for PC patients with Hazard Ratio (HR) < 1, and ALB, CXCL10, IAPP, MET and S100A14 were risk factors with HR > 1. Meanwhile, the mutation status and protein expression status of these genes were

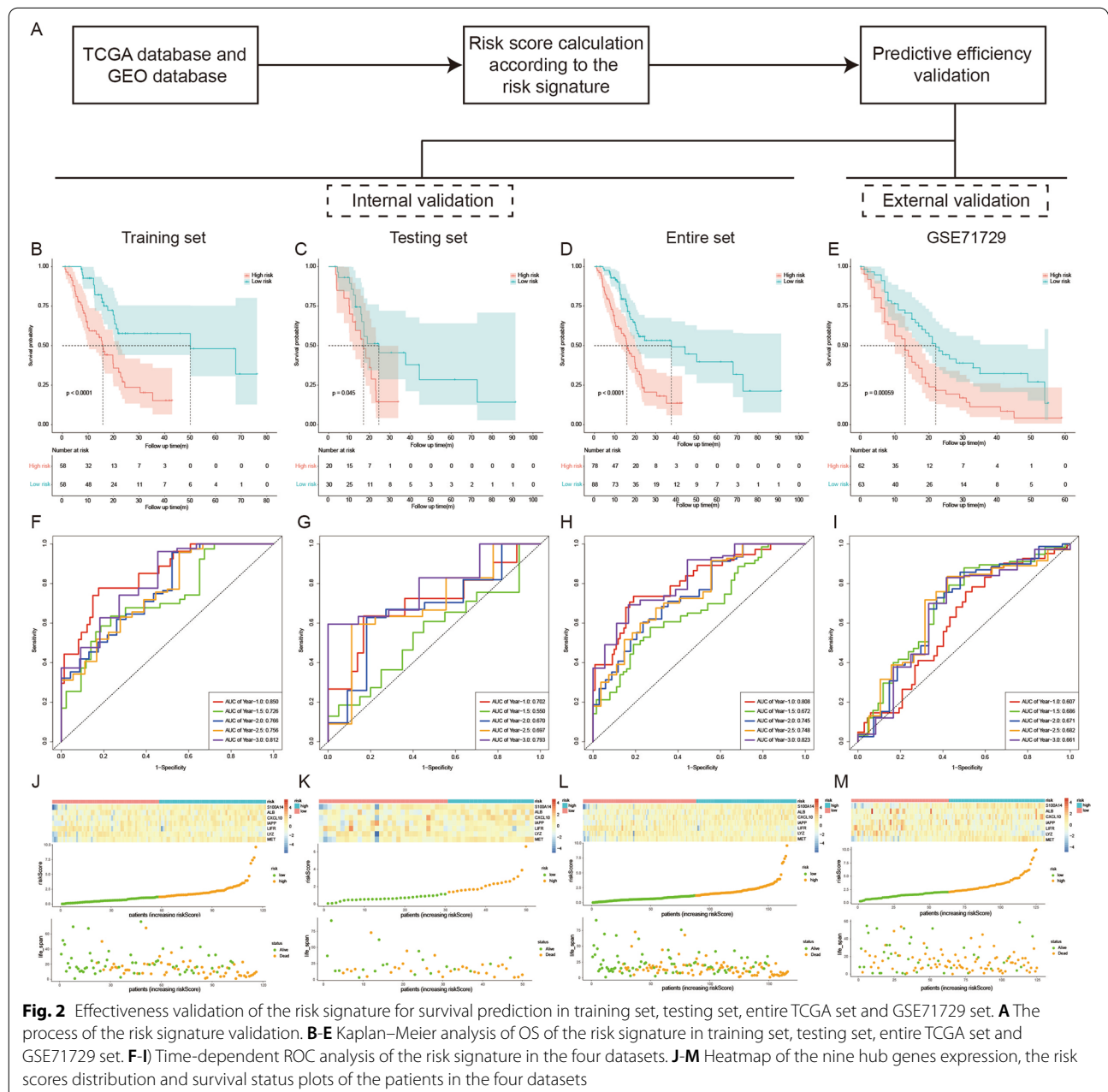
explored through the cBioportal database and HPA database (Figure S2-S3).

**Risk signature construction and evaluation for predicting survival rate of PC**

Based on the gene expression and the regression coefficient derived from the multivariate Cox regression model, a risk-score formula was designed for PC patients' survival prediction. The risk score for each patient was calculated as follows: Risk score =  $(0.0893 \times \text{expression level of ALB}) + (0.1935 \times \text{expression level$

of CXCL10) +  $(0.0635 \times \text{expression level of IAPP}) + (-0.2633 \times \text{expression level of LIFR}) + (-0.2338 \times \text{expression level of LYZ}) + (0.2893 \times \text{expression level of MET}) + (0.2177 \times \text{expression level of S100A14})$ . After that, each patient was assigned a risk score according to the above formula.

In order to verify the validity of the model, we conducted internal validation in training set, testing set and entire set, as well as external validation in GSE71729 set (Fig. 2A). According to the Kaplan–Meier (K–M) curve analysis, patients in the high-risk group had significantly



**Fig. 2** Effectiveness validation of the risk signature for survival prediction in training set, testing set, entire TCGA set and GSE71729 set. **A** The process of the risk signature validation. **B–E** Kaplan–Meier analysis of OS of the risk signature in training set, testing set, entire TCGA set and GSE71729 set. **F–I** Time-dependent ROC analysis of the risk signature in the four datasets. **J–M** Heatmap of the nine hub genes expression, the risk scores distribution and survival status plots of the patients in the four datasets

lower overall survival (OS) than those in the low-risk group in training set, testing set, entire set and GSE71729 set (Fig. 2B-E). At the same time, the area under curves (AUCs) of the risk signature for predicting 1-, 1.5-, 2-, 2.5-, and 3-year survival of PC patients were 0.850, 0.726, 0.766, 0.756, 0.812 in the training set, 0.702, 0.550, 0.670, 0.697, 0.793 in the testing set, 0.808, 0.672, 0.745, 0.748, 0.823 in the entire set, and 0.607, 0.686, 0.671, 0.682, 0.661 in GSE71729 set respectively (Fig. 2F-I). These results demonstrated a high predictive efficacy of the signature in predicting the prognosis of PC. Meanwhile, compared with the low-risk group, the expressions of ALB, CXCL10, IAPP, MET and S100A14 increased in the high-risk group, whereas the expressions of LIFR and LYZ decreased. Consistently, the scatter plot of survival showed a gradual increase in the number of deaths as the risk score rose (Fig. 2J-M).

To further explore whether the predictive power of the signature was independent of other clinicopathological factors, univariate and multivariate Cox regression analysis was conducted in the training set, testing set, and entire set respectively. The results suggested an excellent independence of the risk signature, which was independent of gender, age, grade, AJCC\_stage, T stage and N stage (Figure S4,  $P < 0.05$  in all dataset for risk score). Meanwhile, the predictive efficacy of the signature was also tested in different subgroups stratified by gender, age, tumor grade and T stage. It was found that in almost all subgroups, patients in the high-risk group suffered significantly worse prognosis than those in the low-risk group (Figure S5,  $P < 0.05$  in all subgroup except T1 + T2 group), further confirming the excellent independence of risk signature for PC prognosis prediction.

#### **Establishment and validation of a nomogram based on the seven-gene signature of PC**

In order to optimize the prediction efficiency of the model, clinicopathological factors were also incorporated to construct a nomogram, including gender, age, grade, AJCC\_stage, T stage and N stage. First, univariate Cox regression analysis was applied to preliminarily screen for various clinicopathological factors in training set, and factors with  $P < 0.2$  were included in the multivariate Cox regression analysis (Fig. 3A-B). Concomitantly, we reconfirmed the independence of the risk signature in this process. Finally, risk score, age and N stages were incorporated into the establishment of nomogram for predicting 1-, 2-, and 3-year survival rate of PC. In the nomogram, the patients' survival rates were estimated by the total points obtained by adding up the point of each factor (Fig. 3C). The C-index for the nomogram was 0.727, 0.603 and 0.689 in training set, testing set

and entire set respectively, indicating that the nomogram possessed excellent predictive performance. Subsequently, the predictive power of the nomogram was further evaluated by calibration plot and time-dependent ROC curve. The calibration curves presented satisfied coherence between predicted and actual 1-year, 2-year and 3-year OS in training set, testing set and entire set (Fig. 3D-F). In addition, The AUCs of ROC curves for predicting 1-, 2-, and 3-year survival were 0.791, 0.791, and 0.819 in the training set (Fig. 3G), 0.565, 0.746, and 0.835 in the testing set (Fig. 3H), and 0.728, 0.778, and 0.830 in the entire set, respectively (Fig. 3I).

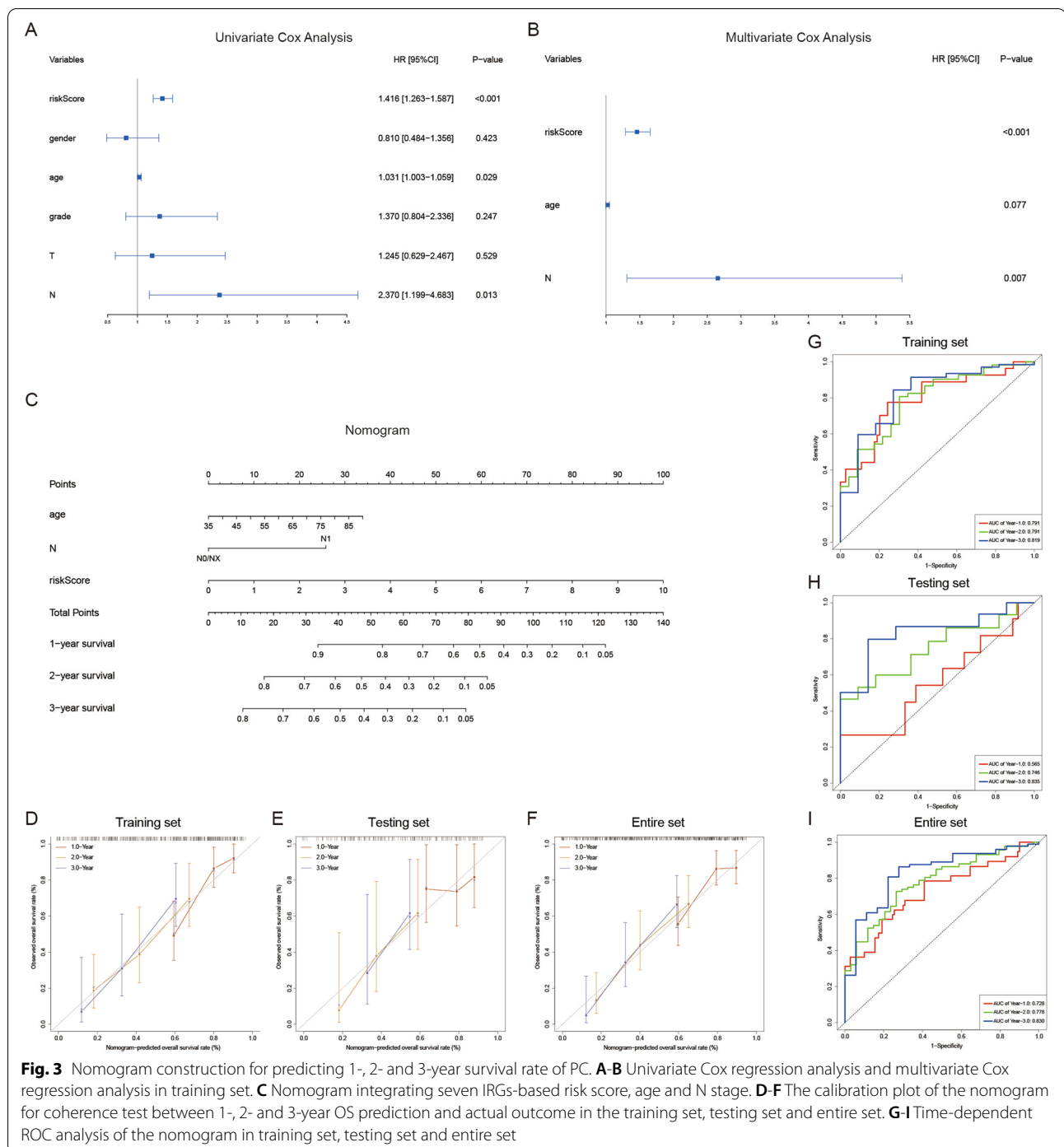
#### **S100A14 was highly expressed and correlates with unfavorable prognosis in PC**

Among the seven hub genes in the risk signature, S100A14 was the risk factor accompanied by the smallest  $P$  value. Furthermore, due to its high HR and coefficient in the risk signature, we tended to consider that S100A14 occupied the most paramount position in the risk signature. A joint analysis of TCGA and GTEx databases confirmed the remarkably high expression of S100A14 in tumor tissues (Fig. 4A). Meanwhile, the analysis from CPTAC database also indicated that S100A14 was significantly overexpressed in tumor tissue at the protein level (Fig. 4B). In addition, patients in the TCGA dataset were divided into high expression group ( $n = 83$ ) and low expression group ( $n = 83$ ) according to the median expression of S100A14. And the prognosis of patients in high expression group was significantly worse than that in low expression group (Fig. 4C). Concomitantly, the relationship between S100A14 expression and clinicopathological information of patients were further analyzed. Notably, the results revealed that the expression of S100A14 increased significantly with the progression of AJCC\_stage, tumor grade, age and T stage (Fig. 4D-I).

#### **S100A14 predicts the infiltration of immune cells into PC microenvironment**

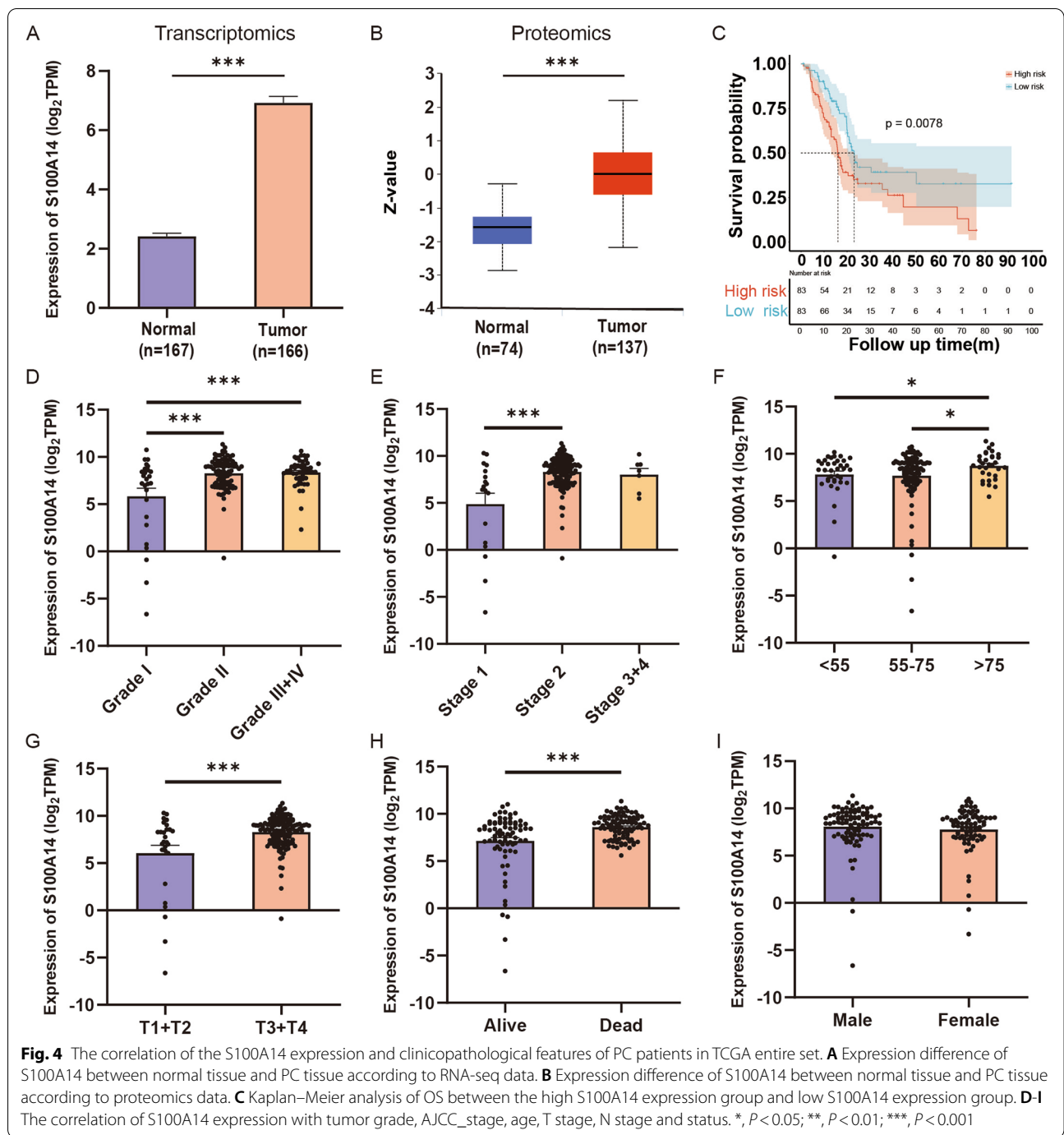
Subsequently, in order to explore the in-depth mechanism of S100A14 leading to poor prognosis of PC, the patients in the TCGA entire set were divided into S100A14 high expression group ( $n = 83$ ) and low expression group ( $n = 83$ ). DEGs analysis was performed on the two groups and GSEA analysis was then conducted (Fig. 5A). Five representative pathways for the Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were presented respectively (Fig. 5B-C). Overall, these pathways suggested that the function of S100A14 in pancreatic cancer may be related to the immunosuppressive tumor microenvironment in vivo.





Therefore, the ESTIMATE algorithm was applied to estimate the proportion of stromal cell and immune cell in PC patients, which revealed that a lower content of stromal cells and immune cells in the S100A14 high expression group (Fig. 5D). And the results of CIBERSORT, an algorithm which was designed to detect the proportions of 22 kinds of immune cells in tissues,

revealed that relatively lower proportion of CD8+ T cells, activated memory CD4+ T cells and monocytes were found in the S100A14 high expression group compared with the low expression group. On the contrary, patients in S100A14 high expression group also possessed higher proportion of M0 macrophages cells and memory B cells (Fig. 5E-F). Moreover, in order

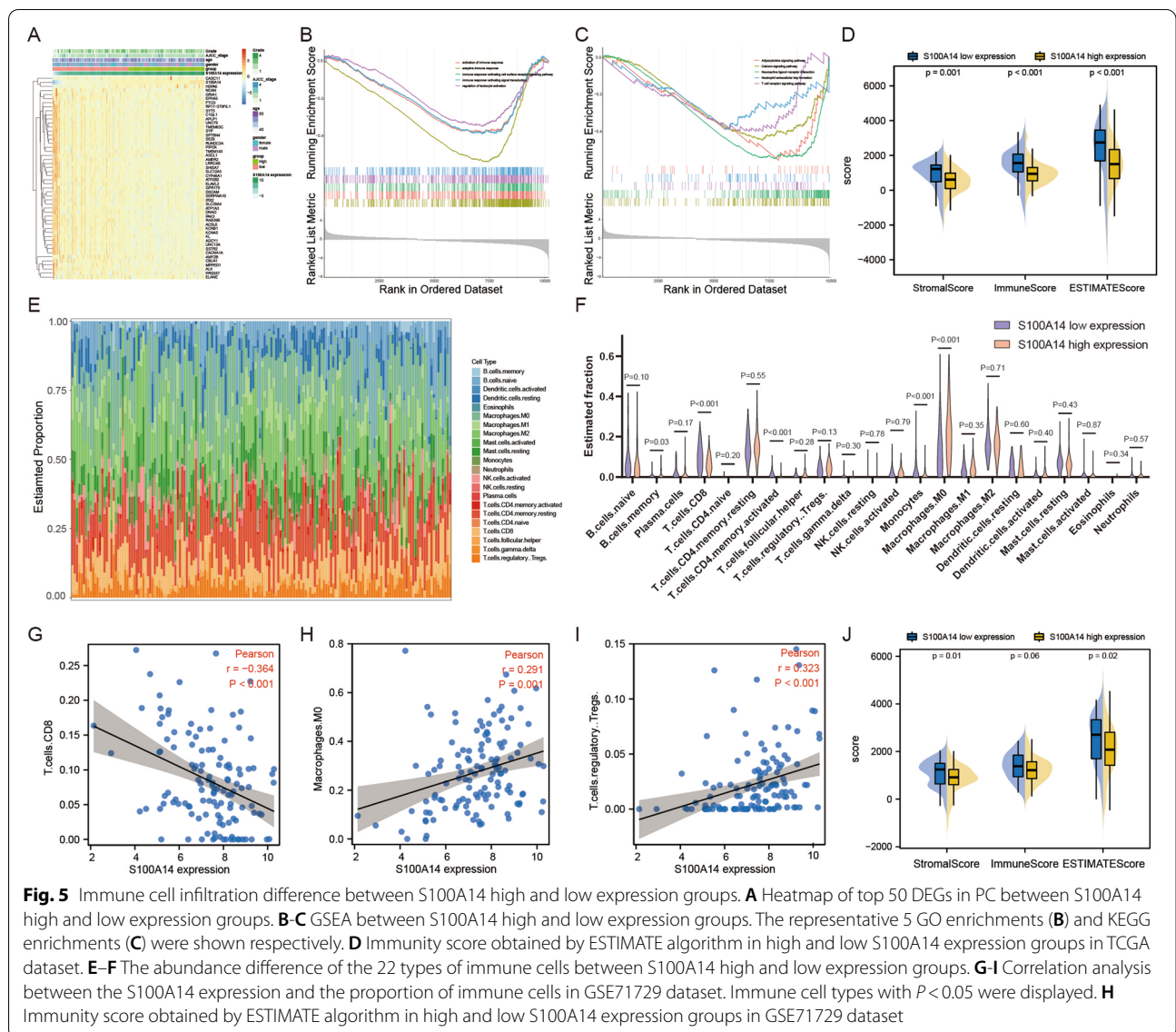


to prove the universality of the results, the immune microenvironment of GSE71729 dataset ( $n = 125$ ) was also analyzed. It was illustrated that the expression of S100A14 possessed a remarkable negative correlation with CD8 + T cells and significant positive correlation with M0 macrophages and Treg cells (Fig. 5G–I). In addition, the score for the tumor microenvironment by

ESTIMATE package was similar in GSE71729 dataset (Fig. 5), which further supported our hypothesis.

#### Verification of the negative correlation between S100A14 expression and CD8 + T cells infiltration in PC

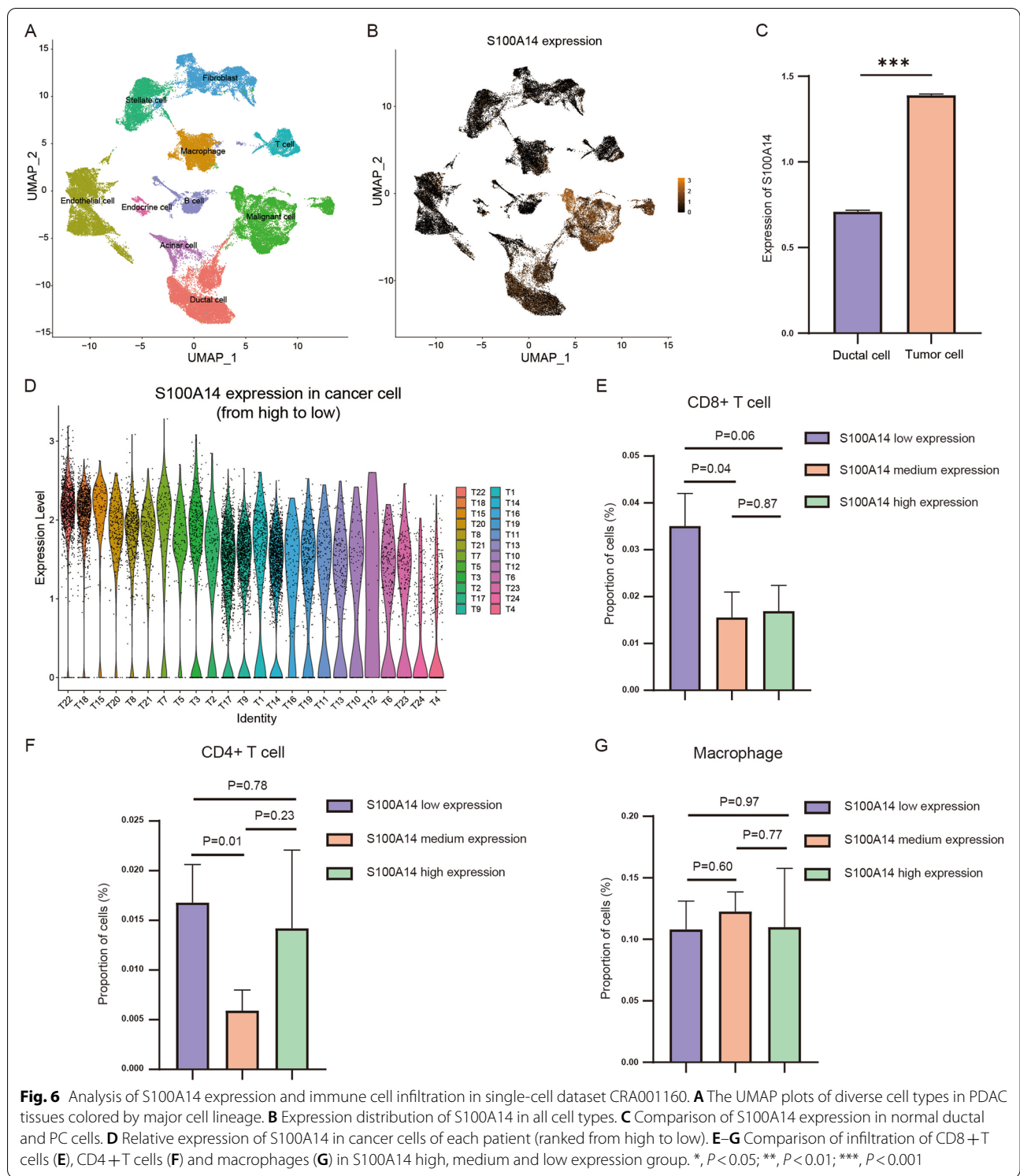
The algorithms above suggested that S100A14 was closely related to the immune microenvironment in PC, especially CD8 + T cells, CD4 + T cells and



macrophages. Here, we further explored the effect of S100A14 on the immune microenvironment at the single-cell level. The single-cell dataset CRA001160 was analyzed in this process, which included 24 tumor samples and 11 normal samples. Firstly, dimensionality reduction clustering and expression distribution analysis of S100A14 showed that S100A14 was mainly expressed in tumor cells (Fig. 6A-B), and the expression of S100A14 in tumor cells was significantly higher than that in normal ductal cells (Fig. 6C). Subsequently, according to the S100A14 expression in tumor cells, patients were divided into S100A14 low expression group, S100A14 medium expression group and S100A14 high expression group (Fig. 6D). The number of CD8 + T cells, CD4 + T cells and macrophages were

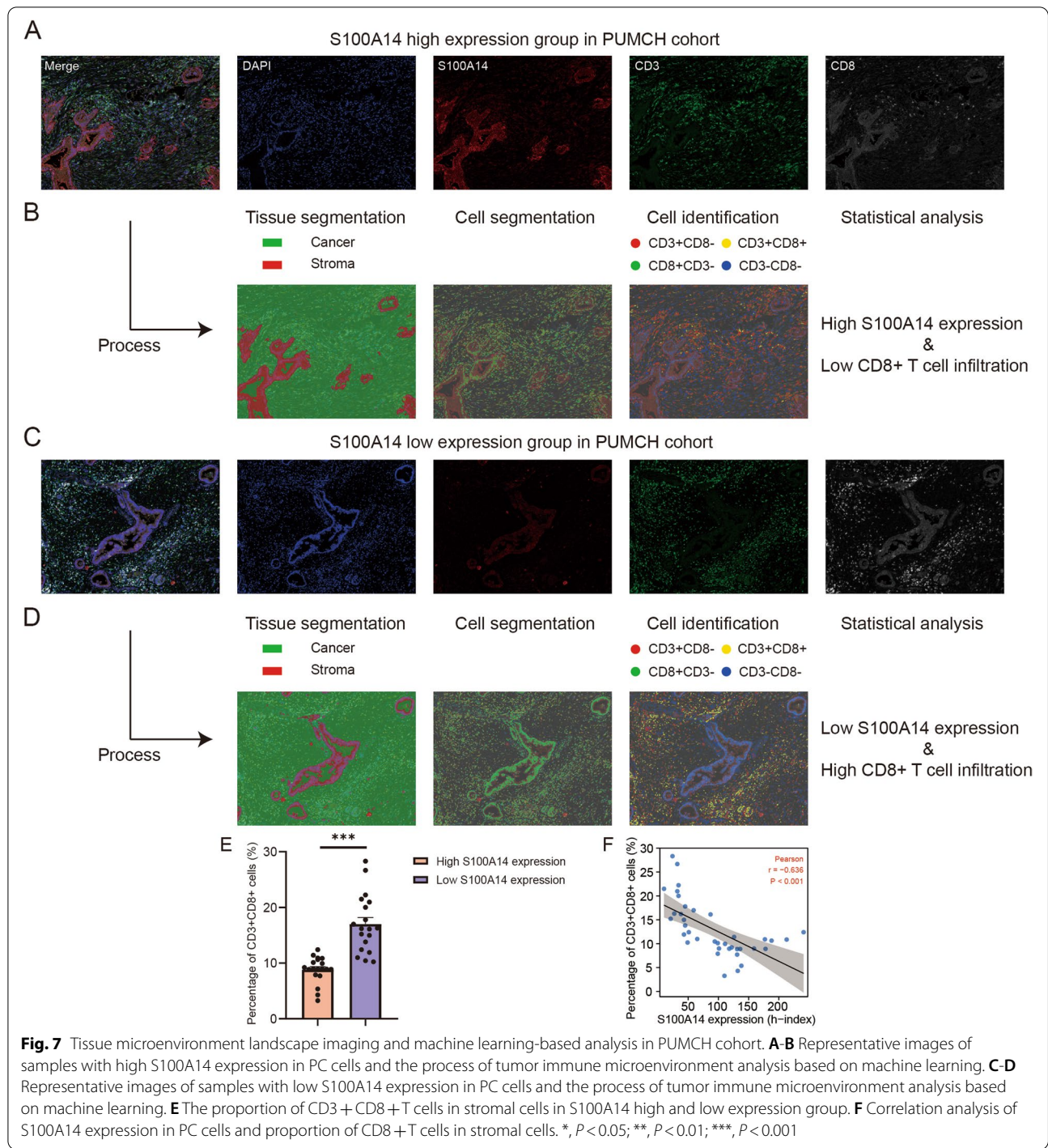
counted in each sample. Consistent with the previous results, with the increase of S100A14 expression, the proportion of CD8 + T cells gradually decreased (Fig. 6E). However, there was no significant trend in the relationship between the expression of S100A14 and the proportion of CD4 + T cells and macrophages (Fig. 6F-G).

To further confirm this conclusion, tissue microenvironment landscape imaging and analysis were conducted on sections of 38 patients in our own PUMCH cohort, in which the association between the expression of S100A14 in tumor cells and the infiltration of CD8 + T cells was explored through machine learning. First, the tumor part and stroma part were separated by machine learning, and all the individual cells in the section were also



separated simultaneously. Tumor cells were automatically scored for S100A14 expression, and cells in the stroma were also identified automatically through the membrane staining (CD3 + CD8 + cells were considered as CD8 + T

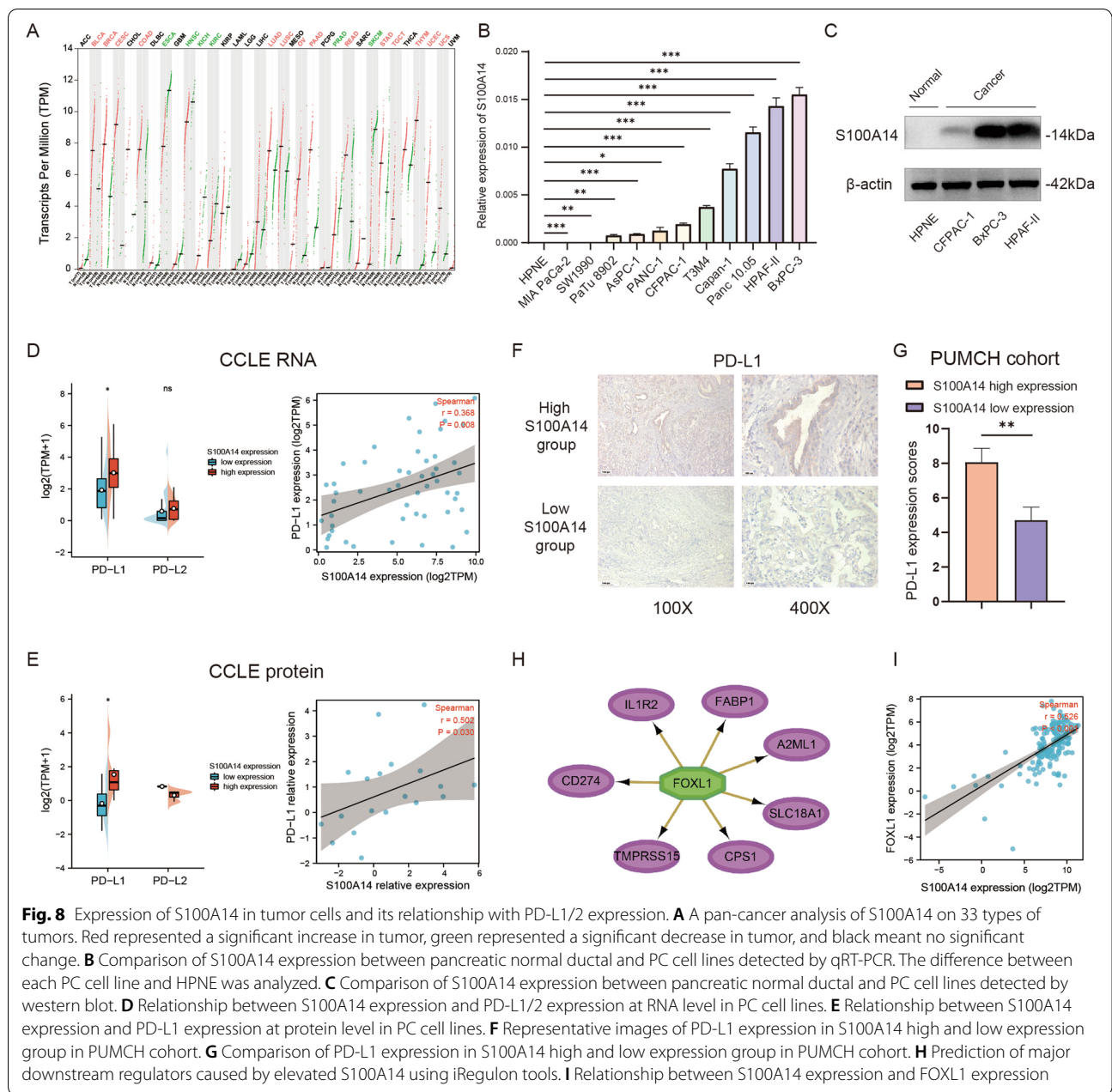
cells). It was discovered that CD8 + T cell infiltration was extremely limited in tissues with high S100A14 expression. Conversely, CD8 + T cell infiltration was abundant in areas with low S100A14 expression (Fig. 7A-D).



Consistently, statistical results also demonstrated that there was a significant negative correlation between the expression of S100A14 in tumor cells and the infiltration of CD8 + T cells in the stroma (Fig. 7E-F).

**The expression of S100A14 was positively correlated with PD-L1 in PC cells**

Since we have proved that S100A14 is associated with reduced infiltration of CD8 + T cells in the immune microenvironment of PC, the mechanism by which S100A14 lead to the deficiency of CD8 + T cells would be further explored. Firstly, a pan-cancer analysis of



S100A14 revealed that PC experienced one of the most remarkably increase of S100A14 expression among all types of cancer (Fig. 8A), which was also confirmed by qRT-PCR and western blot in pancreatic normal ductal cell line and pancreatic cancer cell lines (Fig. 8B-C).

Subsequently, RNA-seq data and proteomics data of all pancreatic cancer cell lines were obtained from the CCLE database to explore whether the deficiency of CD8+ T cells was related to the high expression of immune checkpoints in PC cells. The results showed that the expression of S100A14 was significantly positively correlated

with the expression of PD-L1, but not with the expression of PD-L2 at both RNA and protein levels (Fig. 8D-E). In addition, the results were also validated in our own PUMCH cohort, in which the PD-L1 immunohistochemical score of S100A14 high expression group was significantly higher than that of S100A14 low expression group (Fig. 8F-G).

Meanwhile, we further explored the potential mechanism of elevated PD-L1 expression in patients with high S100A14 expression. The gene co-expression network of patients with high S100A14 expression was analyzed

by iRegulon in Cytoscape, which indicated that the transcription factor FOXL1 (Forkhead Box L1) may play a major regulatory role in the effect caused by high S100A14 expression (Fig. 8H). Moreover, the transcription factor was also predicted to regulate the expression of PD-L1 (Fig. 8H), and it was significantly positively correlated with S100A14 (Fig. 8I), suggesting that the S100A14-FOXL1-PD-L1 pathway may be one of the main signal axes regulating the immune desert microenvironment of PC induced by S100A14.

#### **S100A14 is associated with patients' tumor mutation status and response to immunotherapy**

Mutation profiles of both the S100A14 high expression group and low expression group were analyzed and visualized (Figure S6-S7). For the S100A14 high expression group, the genes with the highest mutation rate were KRAS, TP53, SMAD4, CDKN2A, RNF43, TGFBR2, TTN, GNAS, COL5A1 and FLG. And for the S100A14 low expression group, the top 10 frequently mutation genes were KRAS, TP53, TTN, SMAD4, CDKN2A, MUC16, ARID1A, HECW2, RELN and DCHS1. It's worth noting that although KRAS and TP53 were genes with the highest mutation rate in both groups, the mutation rate was significantly different between the two groups (Fig. 9C, 91%: 63% for KRAS, 76%: 56% for TP53). Therefore, we further calculated the TMB for each patient and found that TMB in S100A14 high expression group was significantly higher than that in S100A14 low expression group (Fig. 9D), although TMB was not significantly associated with patients' prognosis (Fig. 9E).

IPS, a machine learning-based scoring system, was able to predict the efficiency of immunotherapy, including anti-PD1 therapy, anti-CTLA4 therapy and a combination of the two therapies. According to the comprehensive analysis of S100A14 expression and IPS score, patients with high S100A14 expression possessed a relatively higher probability of responding to anti-PD1 therapy, anti-CTLA4 therapy and the combination of anti-PD1 and anti-CTLA-4 therapy (Fig. 9F-K). Therefore, S100A14 is expected to be a predictor of the efficacy of immunotherapy in PC patients, which is consistent with the previous conclusion that S100A14 might contribute to the progression of PC by promoting immune-suppressive tumor microenvironment.

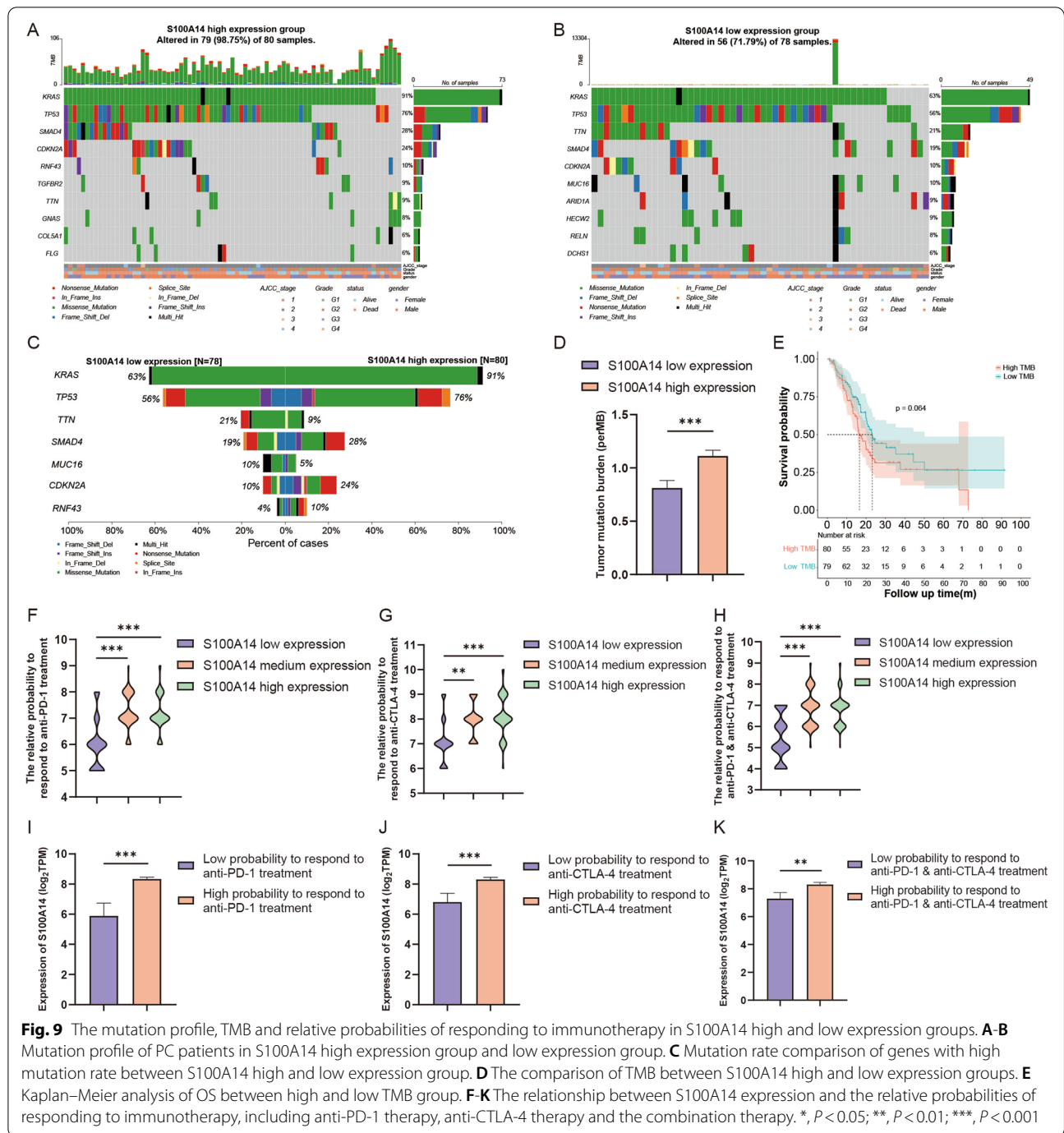
#### **Discussion**

PC is one of the most aggressive malignancies, which is expected to become the second leading cause of cancer-related death by 2030 [3]. The reasons for this situation are various, and the vital one is the lack of reliable predictive models and biomarkers, which hinders individualized treatment of PC. Herein, since the critical role of

tumor immune microenvironment has been gradually revealed in recent years [5, 34], an IRGs-based predictive model was established to accurately assess the prognosis of PC patients, which is also verified in its validity and independence from different dimensions. Among the seven genes in the risk signature, CXCL10, IAPP, LIFR and MET have been reported to be involved in the carcinogenesis and progression of PC [35–38], which also demonstrates the considerable prognostic value of the risk signature to some extent. However, there has been limited coverage of ALB, LYZ and S100A14 in the field of PC, especially in the tumor immune microenvironment. Since S100A14 was the risk factor with the lowest *P* value, and S100A14 possessed relatively large HR and accounted for a high proportion in the risk signature, it was considered that S100A14 might occupy the core position in the risk signature. Therefore, S100A14 was paid special attention in the following analysis.

S100A14, an important member of S100 family proteins, is able to modulate biological process by functioning both as intracellular and extracellular factors [39, 40]. Meanwhile, S100A14 has been reported to be differentially expressed in various human malignancies and implicated in tumorigenesis and tumor progression [41, 42]. Notably, the effects of S100A14 are dual to some extent, depending on different biological processes and tumor types. Li et al. have reported that S100A14 could promote breast cancer metastasis by increasing the expression and secretion of CCL2/CXCL5 via RAGE-NF- $\kappa$ B pathway [43]. Conversely, another study conducted by Meng et al. identified S100A14 as a functional regulator suppressing nasopharyngeal carcinoma metastasis by inhibition of the NF- $\kappa$ B signaling pathway [44]. Here, the results of our study supported its deteriorating role during PC progression. To be specific, in the field of PC, studies have shown that S100A14 was able to promote proliferation, invasion, migration and gemcitabine resistance of PC cells [45]. However, the role of S100A14 in PC immune microenvironment has been limitedly reported, which is also the main content of this manuscript.

GSEA analysis illustrated that the high expression of S100A14 was associated to the weakened immune response in vivo. Therefore, CIBERSORT and ESTIMATE algorithm was further applied to reveal the proportion of various immune cells in PC patients' tissues. Patients with high S100A14 expression had significantly reduced CD8+ T cells, activated memory CD4+ T cells and monocytes, as well as increased M0 macrophages cells and memory B cells. In addition, the results of CIBERSORT algorithm in GSE71729 dataset also showed a similar trend. Based on the results of the two datasets, S100A14 may be associated with reduced CD8+ T cells



**Fig. 9** The mutation profile, TMB and relative probabilities of responding to immunotherapy in S100A14 high and low expression groups. **A-B** Mutation profile of PC patients in S100A14 high expression group and low expression group. **C** Mutation rate comparison of genes with high mutation rate between S100A14 high and low expression group. **D** The comparison of TMB between S100A14 high and low expression groups. **E** Kaplan–Meier analysis of OS between high and low TMB group. **F-K** The relationship between S100A14 expression and the relative probabilities of responding to immunotherapy, including anti-PD-1 therapy, anti-CTLA-4 therapy and the combination therapy. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$

and increased M0 macrophages in the immune micro-environment of PC. Subsequently, analysis of single-cell dataset CRA001160 revealed that S100A14 was mainly expressed in tumor cells. In addition, CD4+T cells, CD8+T cells and macrophages from each patient were extracted in from the single-cell dataset CRA001160. And the percentage of each immune cells to all stromal cells in each patient was then calculated. As a result,

CD8+T cells were significantly reduced in patients with high S100A14 expression, while there were no significant changes in CD4+T cells and macrophages. According to the above results, the effect of high expression of S100A14 in tumor cells on the immune microenvironment of PC was finally focused on CD8+T cells, the immune cell possessing the most prominent tumor killing ability [46, 47].



Tissue microenvironment landscape imaging system is a technology favoring the visualization and analysis of the tumor microenvironment through machine learning, which is able to automatically divide the tumor part and the stroma part, and identify the stroma cell type through the immunofluorescence of the cells. Here, we conducted tissue microenvironment landscape imaging in our PUMCH cohort ( $n=38$ ) to further explore the relationship between S100A14 expression and CD8+T cell infiltration. The machine was designed to automatically segment the tumor cells and score the expression of S100A14 on the tumor cells. Subsequently, for stromal cells, CD3+CD8+ cells were recognized as CD8+T cells and their proportion in stromal cells was counted. Consistent with bioinformatics analysis, patients with high S100A14 expression had a lower proportion of CD8+T cells in their cancer tissues, further confirming the above results from another dimension.

Afterwards, we aimed to explore how S100A14 leads to the deficiency of CD8+T cells. Since the results above showed that S100A14 was mainly expressed in tumor cells, and previous studies had proved that S100 family proteins might up-regulate the expression of PD-L1 [32, 48], we hypothesized that S100A14 might be related to the dysregulation of immune checkpoints including PD-L1 and PD-L2, which had been widely acknowledged to inhibit T cell activation through activation of PD-1 receptors [49–51]. Therefore, the RNA-seq and proteomics data of all pancreatic cancer cell lines in the CCLE database were integrated and analyzed, which proved the expression of S100A14 was significantly positively correlated with the expression of PD-L1 in PC cell lines. Therefore, it was suggested that S100A14 might inhibit proliferation of CD8+T cells and promote apoptosis of CD8+T cells by up-regulating PD-L1 expression in PC cells.

Cancer immunotherapies, which manipulate the immune system to recognize and attack cancer cells, have become a powerful clinical strategy for treating cancer [52, 53]. However, immunotherapy has shown very limited efficacy in PC [54, 55]. Promisingly, a small subset of patients who exhibited high effector T-cell infiltration in tumor had longer overall survival [56, 57], suggesting that immunotherapy is still a potential treatment strategy for PC patients. And what we need to do is to identify patients with a high probability of responding to immunotherapy or explore optimal drug combinations to improve the effectiveness of immunotherapy.

Since we had explored the role of S100A14 in the immune microenvironment of PC, we wondered whether S100A14 might be a promising biomarker for predicting

the immunotherapy response of PC patients. It has been widely reported that patients with high TMB had a relatively higher probability of responding to immunotherapy [58, 59]. Therefore, we explored the relationship between S100A14 expression and TMB, which suggested that patients with high S100A14 expression possessed a higher TMB, indicating a higher response rate to immunotherapy in this group. Consistently, the application of IPS algorithm also reached similar conclusion, indicating that S100A14 could be used as a potential biomarker for predicting the efficacy of immunotherapy in PC patients.

In spite of the positive results, the limitation of current study should be addressed as well. First of all, IPS is an algorithm based on machine learning to simulate the immunotherapy response of PC patients. Although its effectiveness has been widely verified in multiple datasets, it is still different from the real clinical situation to some extent. Secondly, due to the extremely poor prognosis of PC, there are few patients in the cohort with more than three years of survival, which may affect the long-term prediction efficiency of the risk signature.

## Conclusion

In summary, in this study, an IRGs-based prediction signature was constructed and validated in the training set, testing set, entire set and GSE71729 set respectively. Meanwhile, a nomogram was also established to further improve the prediction efficiency of the model. Subsequently, S10014 was identified as the gene occupying the most paramount position in this signature, which was proved to be increased significantly with the progression of AJCC\_stage, tumor grade, age and T stage. GSEA, ESTIMATE and CIBERSORT algorithms demonstrated that the deteriorating effect of S100A14 on PC was mainly related to the dysregulation of immune microenvironment. Exploration of the single-cell dataset CRA001160 further demonstrated that high expression of S100A14 was associated with reduced infiltration of CD8+T cells in tumor tissue. Subsequently, tissue microenvironment landscape imaging and machine learning-based immune microenvironment analysis were conducted in our PUMCH cohort, which confirmed the negative correlation between S100A14 and CD8+T cell infiltration from another dimension. In addition, a pan-pancreatic cancer cell lines analysis, iRegulon analysis and immunohistochemical results of PUMCH cohorts suggested that S100A14-FOX L1-PD-L1 pathway may be a potential signal axes inhibiting CD8+T cell activation. Finally, TMB analysis and IPS algorithm implied that S100A14 may serve as a potential biomarker for predicting the efficacy of immunotherapy in PC.

## Abbreviations

PC: Pancreatic Cancer; S100A14: S100 Calcium Binding Protein A14; TIME: Tumor Immune Microenvironment; IRGs: Immune-related Genes; TMB: Tumor Mutation Burden; IPS: Immunophenoscore; GSEA: Gene Set Enrichment Analysis.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-022-09927-0>.

**Additional file 1: Figure S1.** Flowchart of the whole study. **Figure S2.** The mutation status of the seven hub IRGs in TCGA dataset (A) and ICGC dataset (B). **Figure S3.** Representative images of hub IRGs expression status except CXCL10 and LIFR in normal pancreas and PC tissues. (The expression information of CXCL10 and LIFR was unavailable in HPA database). **Figure S4.** Independence of the risk signature and the other clinical variables, including, age, tumor grade, AJCC\_stage and gender. (A, C, E) Univariate Cox regression analyses in the training set, testing set and entire set. (B, D, F) Multivariate Cox regression analyses in the training set testing set and entire set. **Figure S5.** Stratification analyses of all patients using the risk signature. (A-C) The Kaplan-Meier analysis of the younger stratum (age  $\leq 65$ , n=88), older stratum (age  $>65$ , n=78) and all patients with PC (n=166). (D-F) The Kaplan-Meier analysis of the male stratum (n=90), female stratum (n=76) and all patients with PC (n=166). (G-I) The Kaplan-Meier analysis of the Grade I/II stratum (n=117), Grade III/IV stratum (n=49) and all patients with PC (n=166). (J-L) The Kaplan-Meier analysis of the T1+T2 stratum (n=27), T3+T4 stratum (n=139) and all patients with PC (n=166). **Figure S6.** The mutation profiles of patients in S100A14 high expression group. **Figure S7.** The mutation profiles of patients in S100A14 low expression group. **Figure S8.** Correlation analysis of S100A14 and CD3E in the Bailey P. et al. cohort.

### Additional file 2.

**Additional file 3: Table S1.** Detailed clinical and pathologic information of TCGA entire set. **Table S2.** Detailed clinical and pathologic information of CRA001160 dataset. **Table S3.** The imaging channels for each antibody and antibody incubation information in tissue microenvironment landscape imaging. **Table S4.** Immune-related genes obtained from the ImmPort database.

## Acknowledgements

Not applicable

## Authors' contributions

Study concept and design: CW, YC, XY, RX, JS, RR, QX and YZ. Experimental design and implementation: CW, YC and RX. Drafting of the manuscript: CW, YC and XY. Critical revision of the manuscript for important intellectual content: CW, YC, XY, RX, JS, RR, QX and YZ. Obtained funding: CW, QX and YZ. All authors read and approved the final manuscript.

## Funding

This study was supported by the CAMS Innovation Fund For Medical Sciences (2021, 2021-1-I2M-002, to YZ), National Nature Science Foundation of China (2021, 82102810, to CW) and fellowship of China Postdoctoral Science Foundation (2021, 2021M700501, to CW).

## Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

## Declarations

### Ethics approval and consent to participate

This study conformed to the experimental guidelines of the World Medical Association and the Ethics Committee of Peking Union Medical College Hospital. Written informed consent were obtained from all the patients enrolled

in this study. All methods were performed in accordance with the relevant guidelines and regulations.

### Consent for publication

Not applicable.

### Competing interests

No potential competing interests were disclosed by the authors.

Received: 8 April 2022 Accepted: 20 July 2022

Published online: 11 August 2022

## References

- 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>.
- 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>.
- Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Can Res.* 2014;74:2913–21. <https://doi.org/10.1158/0008-5472.Can-14-0155>.
- Mizrahi JD, Surana R, Valle JW, Shroff RT. Pancreatic cancer. *Lancet (London, England).* 2020;395:2008–20. [https://doi.org/10.1016/s0140-6736\(20\)30974-0](https://doi.org/10.1016/s0140-6736(20)30974-0).
- Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol.* 2013;14:1014–22. <https://doi.org/10.1038/ni.2703>.
- Binnewies M, Roberts EW, Kersten K, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med.* 2018;24:541–50. <https://doi.org/10.1038/s41591-018-0014-x>.
- Fu T, Dai LJ, Wu SY, Xiao Y, Ma D, Jiang YZ, Shao ZM. Spatial architecture of the immune microenvironment orchestrates tumor immunity and therapeutic response. *J Hematol Oncol.* 2021;14:98. <https://doi.org/10.1186/s13045-021-01103-4>.
- Riera-Domingo C, Audigé A, Granja S, Cheng WC, Ho PC, Baltazar F, Stockmann C, Mazzone M. Immunity, Hypoxia, and Metabolism—the Ménage à Trois of Cancer: Implications for Immunotherapy. *Physiol Rev.* 2020;100:1–102. <https://doi.org/10.1152/physrev.00018.2019>.
- Qiao J, Liu Z, Dong C, et al. Targeting Tumors with IL-10 Prevents Dendritic Cell-Mediated CD8(+) T Cell Apoptosis. *Cancer Cell.* 2019;35:901–15.e4. <https://doi.org/10.1016/j.ccell.2019.05.005>.
- Gunderson AJ, Yamazaki T, McCarty K, et al. TGF $\beta$  suppresses CD8(+) T cell expression of CXCR3 and tumor trafficking. *Nat Commun.* 2020;11:1749. <https://doi.org/10.1038/s41467-020-15404-8>.
- Bhattacharya S, Andorf S, Gomes L, et al. ImmPort: disseminating data to the public for the future of immunology. *Immunol Res.* 2014;58:234–9. <https://doi.org/10.1007/s12026-014-8516-1>.
- Badea L, Herlea V, Dima SO, Dumitrascu T, Popescu I. Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. *Hepatogastroenterology.* 2008;55:2016–27.
- Zhang G, He P, Tan H, et al. Integration of metabolomics and transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer. *Clin Cancer Res.* 2013;19:4983–93. <https://doi.org/10.1158/1078-0432.Ccr-13-0209>.
- Janky R, Binda MM, Allemeersch J, Van den Broeck A, Govaere O, Swinnen JV, Roskams T, Aerts S, Topal B. Prognostic relevance of molecular subtypes and master regulators in pancreatic ductal adenocarcinoma. *BMC Cancer.* 2016;16:632. <https://doi.org/10.1186/s12885-016-2540-6>.
- Moffitt RA, Marayati R, Flate EL, 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>.
- Chen F, Chandrashekar DS, Varambally S, Creighton CJ. Pan-cancer molecular subtypes revealed by mass-spectrometry-based proteomic characterization of more than 500 human cancers. *Nat Commun.* 2019;10:5679. <https://doi.org/10.1038/s41467-019-13528-0>.

17. Cao L, Huang C, Cui Zhou D, et al. Proteogenomic characterization of pancreatic ductal adenocarcinoma. *Cell*. 2021;184:5031–52.e26. <https://doi.org/10.1016/j.cell.2021.08.023>.
18. Sun D, Wang J, Han Y, et al. TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic Acids Res*. 2021;49:D1420–30. <https://doi.org/10.1093/nar/gkaa1020>.
19. Peng J, Sun BF, Chen CY, et al. Single-cell RNA-seq highlights intratumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res*. 2019;29:725–38. <https://doi.org/10.1038/s41422-019-0195-y>.
20. Barretina J, Caponigro G, Stransky N, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483:603–7. <https://doi.org/10.1038/nature11003>.
21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res*. 2015;43:e47. <https://doi.org/10.1093/nar/gkv007>.
22. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2:401–4. <https://doi.org/10.1158/2159-8290.Cd-12-0095>.
23. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102:15545–50. <https://doi.org/10.1073/pnas.0506580102>.
24. Harris MA, Clark J, Ireland A, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004;32:D258–61. <https://doi.org/10.1093/nar/gkh036>.
25. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30. <https://doi.org/10.1093/nar/28.1.27>.
26. Janky R, Verfaillie A, Imrichová H, et al. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput Biol*. 2014;10:e1003731. <https://doi.org/10.1371/journal.pcbi.1003731>.
27. Yoshihara K, Shahmoradgolgi M, Martínez E, 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>.
28. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12:453–7. <https://doi.org/10.1038/nmeth.3337>.
29. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018;28:1747–56. <https://doi.org/10.1101/gr.239244.118>.
30. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer 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>.
31. Chen Y, Xu R, Ruze R, Yang J, Wang H, Song J, You L, Wang C, Zhao Y. Construction of a prognostic model with histone modification-related genes and identification of potential drugs in pancreatic cancer. *Cancer Cell Int*. 2021;21:291. <https://doi.org/10.1186/s12935-021-01928-6>.
32. Chen Y, Wang C, Song J, Xu R, Ruze R, Zhao Y. S100A2 Is a Prognostic Biomarker Involved in Immune Infiltration and Predict Immunotherapy Response in Pancreatic Cancer. *Front Immunol*. 2021;12: 758004. <https://doi.org/10.3389/fimmu.2021.758004>.
33. Ren B, Yang J, Wang C, et al. High-resolution Hi-C maps highlight multiscale 3D epigenome reprogramming during pancreatic cancer metastasis. *J Hematol Oncol*. 2021;14:120. <https://doi.org/10.1186/s13045-021-01131-0>.
34. van de Wall S, Santegoets KCM, van Houtum EJH, Büll C, Adema GJ. Sialoglycans and Siglecs Can Shape the Tumor Immune Microenvironment. *Trends Immunol*. 2020;41:274–85. <https://doi.org/10.1016/j.it.2020.02.001>.
35. Hirth M, Gandla J, Höper C, et al. CXCL10 and CCL21 Promote Migration of Pancreatic Cancer Cells Toward Sensory Neurons and Neural Remodeling in Tumors in Mice, Associated With Pain in Patients. *Gastroenterology*. 2020;159:665–81.e13. <https://doi.org/10.1053/j.gastro.2020.04.037>.
36. Permert J, Larsson J, Westermark GT, Herrington MK, Christmansson L, Pour PM, Westermark P, Adrian TE. Islet amyloid polypeptide in patients with pancreatic cancer and diabetes. *N Engl J Med*. 1994;330:313–8. <https://doi.org/10.1056/nejm199402033300503>.
37. Shi Y, Gao W, Lytle NK, et al. Targeting LIF-mediated paracrine interaction for pancreatic cancer therapy and monitoring. *Nature*. 2019;569:131–5. <https://doi.org/10.1038/s41586-019-1130-6>.
38. Guo X, Zhou Q, Su D, et al. Circular RNA circBFAR promotes the progression of pancreatic ductal adenocarcinoma via the miR-34b-5p/MET/Akt axis. *Mol Cancer*. 2020;19:83. <https://doi.org/10.1186/s12943-020-01196-4>.
39. Bresnick AR, Weber DJ, Zimmer DB. S100 proteins in cancer. *Nat Rev Cancer*. 2015;15:96–109. <https://doi.org/10.1038/nrc3893>.
40. Donato R. S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles. *Int J Biochem Cell Biol*. 2001;33:637–68. [https://doi.org/10.1016/s1357-2725\(01\)00046-2](https://doi.org/10.1016/s1357-2725(01)00046-2).
41. Chen H, Yu D, Luo A, et al. Functional role of S100A14 genetic variants and their association with esophageal squamous cell carcinoma. *Can Res*. 2009;69:3451–7. <https://doi.org/10.1158/0008-5472.Can-08-4231>.
42. Zhu M, Wang H, Cui J, Li W, An G, Pan Y, Zhang Q, Xing R, Lu Y. Calcium-binding protein S100A14 induces differentiation and suppresses metastasis in gastric cancer. *Cell Death Dis*. 2017;8:e2938. <https://doi.org/10.1038/cddis.2017.297>.
43. Li X, Wang M, Gong T, et al. A S100A14-CCL2/CXCL5 signaling axis drives breast cancer metastasis. *Theranostics*. 2020;10:5687–703. <https://doi.org/10.7150/thno.42087>.
44. Meng DF, Sun R, Liu GY, et al. S100A14 suppresses metastasis of nasopharyngeal carcinoma by inhibition of NF-κB signaling through degradation of IRAK1. *Oncogene*. 2020;39:5307–22. <https://doi.org/10.1038/s41388-020-1363-8>.
45. Zhu H, Gao W, Li X, Yu L, Luo D, Liu Y, Yu X (2021) S100A14 promotes progression and gemcitabine resistance in pancreatic cancer. *Int J Pancreatol (IAP)* [et al.]. 21: 589–98. <https://doi.org/10.1016/j.pan.2021.01.011>
46. van der Leun AM, Thommen DS, Schumacher TN. CD8(+) T cell states in human cancer: insights from single-cell analysis. *Nat Rev Cancer*. 2020;20:218–32. <https://doi.org/10.1038/s41568-019-0235-4>.
47. Krishna S, Lowery FJ, Copeland AR et al. (2020) Stem-like CD8 T cells mediate response of adoptive cell immunotherapy against human cancer. *Science* (New York, N.Y.). 370: 1328–34. <https://doi.org/10.1126/science.abb9847>
48. Cheng P, Eksioğlu EA, Chen X, et al. S100A9-induced overexpression of PD-1/PD-L1 contributes to ineffective hematopoiesis in myelodysplastic syndromes. *Leukemia*. 2019;33:2034–46. <https://doi.org/10.1038/s41375-019-0397-9>.
49. Daassi D, Mahoney KM, Freeman GJ. The importance of exosomal PDL1 in tumour immune evasion. *Nat Rev Immunol*. 2020;20:209–15. <https://doi.org/10.1038/s41577-019-0264-y>.
50. Chen S, Crabill GA, Pritchard TS, McMiller TL, Wei P, Pardoll DM, Pan F, Topalian SL. Mechanisms regulating PD-L1 expression on tumor and immune cells. *J Immunother Cancer*. 2019;7:305. <https://doi.org/10.1186/s40425-019-0770-2>.
51. Yearley JH, Gibson C, Yu N, et al. PD-L2 Expression in Human Tumors: Relevance to Anti-PD-1 Therapy in Cancer. *Clin Cancer Res*. 2017;23:3158–67. <https://doi.org/10.1158/1078-0432.Ccr-16-1761>.
52. Kennedy LB, Salama AKS (2020) A review of cancer immunotherapy toxicity. *CA: a cancer journal for clinicians*. 70: 86–104. <https://doi.org/10.3322/caac.21596>
53. Riley RS, June CH, Langer R, Mitchell MJ. Delivery technologies for cancer immunotherapy. *Nat Rev Drug Discovery*. 2019;18:175–96. <https://doi.org/10.1038/s41573-018-0006-z>.
54. Schizas D, Charalampakis N, Kole C, Economopoulou P, Koustas E, Gkotsis E, Ziogas D, Psyrris A, Karamouzis MV. Immunotherapy for pancreatic cancer: A 2020 update. *Cancer Treat Rev*. 2020;86: 102016. <https://doi.org/10.1016/j.ctrv.2020.102016>.
55. Bear AS, Vonderheide RH, O'Hara MH. Challenges and Opportunities for Pancreatic Cancer Immunotherapy. *Cancer Cell*. 2020;38:788–802. <https://doi.org/10.1016/j.ccell.2020.08.004>.
56. Morrison AH, Byrne KT, Vonderheide RH. Immunotherapy and Prevention of Pancreatic Cancer. *Trends in cancer*. 2018;4:418–28. <https://doi.org/10.1016/j.trecan.2018.04.001>.

57. Balachandran VP, Łuksza M, Zhao JN, et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature*. 2017;551:512–6. <https://doi.org/10.1038/nature24462>.
58. Jardim DL, Goodman A, de Melo GD, Kurzrock R. The Challenges of Tumor Mutational Burden as an Immunotherapy Biomarker. *Cancer Cell*. 2021;39:154–73. <https://doi.org/10.1016/j.ccell.2020.10.001>.
59. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, Peters S. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Annals Oncol*. 2019;30:44–56. <https://doi.org/10.1093/annonc/mdy495>.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

