

RESEARCH ARTICLE

Open Access



A novel tp53-associated nomogram to predict the overall survival in patients with pancreatic cancer

Xun Liu, Bobo Chen, Jiahui Chen and Shaolong Sun*

Abstract

Background: Gene mutations play critical roles in tumorigenesis and cancer development. Our study aimed to screen survival-related mutations and explore a novel gene signature to predict the overall survival in pancreatic cancer.

Methods: Somatic mutation data from three cohorts were used to identify the common survival-related gene mutation with Kaplan-Meier curves. RNA-sequencing data were used to explore the signature for survival prediction. First, Weighted Gene Co-expression Network Analysis was conducted to identify candidate genes. Then, the ICGC-PACA-CA cohort was applied as the training set and the TCGA-PAAD cohort was used as the external validation set. A TP53-associated signature calculating the risk score of every patient was developed with univariate Cox, least absolute shrinkage and selection operator, and stepwise regression analysis. Kaplan-Meier and receiver operating characteristic curves were plotted to verify the accuracy. The independence of the signature was confirmed by the multivariate Cox regression analysis. Finally, a prognostic nomogram including 359 patients was constructed based on the combined expression data and the risk scores.

Results: TP53 mutation was screened to be the robust and survival-related mutation type, and was associated with immune cell infiltration. Two thousand, four hundred fifty-five genes included in the six modules generated in the WGCNA were screened as candidate survival related TP53-associated genes. A seven-gene signature was constructed: Risk score = $(0.1254 \times \text{ERRFI1}) - (0.1365 \times \text{IL6R}) - (0.4400 \times \text{PPP1R10}) - (0.3397 \times \text{PTOV1-AS2}) + (0.1544 \times \text{SCEL}) - (0.4412 \times \text{SSX2IP}) - (0.2231 \times \text{TXNL4A})$. Area Under Curves of 1-, 3-, and 5-year ROC curves were 0.731, 0.808, and 0.873 in the training set and 0.703, 0.677, and 0.737 in the validation set. A prognostic nomogram including 359 patients was constructed and well-calibrated, with the Area Under Curves of 1-, 3-, and 5-year ROC curves as 0.713, 0.753, and 0.823.

Conclusions: The TP53-associated signature exhibited good prognostic efficacy in predicting the overall survival of PC patients.

Keywords: TP53 mutation, Survival prediction, WGCNA, Nomogram, Pancreatic Cancer

* Correspondence: drsunshaolong@yeah.net

Department of Pancreas and Endocrine Surgery, Shengjing Hospital of China Medical University, No. 36 Sanhao Street, Heping District, Shenyang 110004, Liaoning, China



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Pancreatic cancer (PC) has been the 11th most common cancer in 2012. Both the incidence and mortality rates of PC have been increasing in developed countries. PC was the 3rd leading cause of cancer-related mortality in the United States in 2017, and will grow to be the 2nd leading cause of cancer-related mortality in 2030 [1, 2]. As the malignancy with the highest mortality, the survival rate of PC has not been increased despite years of investigation [3]. The primary treatment approach is still surgery, though only 20% of patients will survive over 5 years after pancreatectomy [4]. The main reasons include delayed diagnosis at an advanced stage, ineffective treatment, and poor prognosis. The diagnosis of pancreatic cancer has been generally difficult because it depended on the clinical symptoms, which were not indicative during onset and gradual progression over years. There may be only midepigastric pain, weight loss, malaise, nausea, and fatigue. Most patients diagnosed with PC have developed metastases, leading to poor prognostic outcomes [5]. The detection of pancreatic cancer at early and resectable stage has been proved to have beneficial effects on long-term survival.

Numerous studies have tried to explore the risk factors for PC. Smoking, Alcohol consumption, Obesity and Dietary factors have been proved to increase the risk of PC [2]. Except for the clinical symptoms, biomarkers in blood or biopsy have been also developed for PC screen and monitor. Serum cancer antigen 19–9 (CA 19–9) is the only approved marker for clinical management of PC patients [6]. With the advances of proteome and genomes, some specific expression profiles have been revealed in PC patients [7]. These expression profiles may assist in interpreting the hereditary incidence, unpredictable efficacy of clinical treatment, as well as the poor outcomes. A recent study has summarized the selected protein biomarkers in tissue, serum, plasma, and pancreatic juice. The combination of CA 19–9 and other emerging biomarkers improved clinical management of PC [7]. The genetic alterations have also been investigated in PC patients. The study proposed several known frequently mutated genes (including KRAS and TP53) and revealed mutations in critical signaling pathways [4]. Further, a recent meta-analysis involving 9040 patients and 12,496 controls reported five new susceptibility loci for PC [8]. The emerging genetic alterations assisted in the better characterization of the complex diseases.

PC is a heterogeneous disease with various subtypes [9]. More understanding may help in improving the disease management [10]. Here, from another point of view, we explored the gene alterations in tumorigenesis and cancer development of PC based on comprehensive bioinformatic analysis. Overall survival-related mutations were screened and a novel gene signature was developed

to predict the overall survival for patients with PC. It may be beneficial to improve the prognosis prediction and post-surgery management.

Materials and methods

Data sources

A total of three cohorts were included in this study. Somatic mutation data ($n = 124$) and RNA-sequencing data ($n = 178$) with corresponding clinical data of the TCGA-PAAD cohort were downloaded from the Cancer Genome Atlas (<https://cancergenome.nih.gov/>). Somatic mutation data of the ICGC-PACA-CA cohort ($n = 263$) and the ICGC-PACA-AU cohort ($n = 373$) with corresponding clinical data were downloaded from the International Cancer Genome Consortium (ICGC) database (<https://www.icgc.org>). RNA-sequencing data of the ICGC-PACA-CA cohort ($n = 182$) with corresponding clinical data were also downloaded from the International Cancer Genome Consortium (ICGC) database (<https://www.icgc.org>).

Data processing and normalization

Somatic mutation data of the TCGA-PAAD cohort were based on VarScan2 [11]. Somatic mutation data of the ICGC-PACA-AU cohort were based on qsnip [12]. Somatic mutation data of the ICGC-PACA-CA cohort were based on MuTect [13]. Synonymous variant data which could not cause the change of amino acid sequence were filtered out. The RNA-sequencing data of the ICGC-PACA-CA and the TCGA-PAAD cohorts were normalized data (FPKM). All expression values were log₂-transformed. The batch effect was eliminated with the SVA R package in R 3.6.1. The somatic mutation data of the three cohorts were used to identify the common survival-related gene mutation. The RNA-sequencing data of the ICGC-PACA-CA and the TCGA-PAAD cohorts were used to explore and validate the multi-gene signature to predict the overall survival in patients with pancreatic cancer.

Identification of survival-related mutations

Somatic mutation data of the three cohorts were extracted and sorted with Perl 5.32.0 (<https://www.perl.org/>). Mutational frequencies were calculated by the counting method. The top 30 genes with the highest mutation frequency were acquired from the three cohorts separately. Waterfall plots of the mutational landscape were generated with the GenVisR R package [14]. The common mutations were selected and drawn by the Venn diagram. In order to identify robust mutations which were associated with the overall survival, Kaplan-Meier (KM) curves comparing the mutated group with the wild group were plotted by the Survival R package in three cohorts separately. $P < 0.05$ were considered to be

indicative of significance. Finally, only TP53 mutation was indicated to be significantly related to the overall survival in all the three cohorts.

Association between TP53 mutation and tumor mutation burden (TMB)

TMB is an important genetic factor in mediating antitumor immunity. In this study, number of non-synonymous single nucleotide polymorphism (SNP) of each sample was calculated with Perl 5.32.0. TMB scores were estimated as number of variants/the length of exons (38 million). Wilcoxon tests were conducted to compare the differences of TMB between the TP53 mutated and wild groups in the three cohorts.

Gene set enrichment analysis (GSEA)

To reveal functional differences between PC patients with and without TP53 mutations, GSEA was done in GSEA 4.0.1 software with the gene set *c2* (cp.kegg.v6.2.-symbols.gmt). A total of 65 patients without TP53 mutations and 77 patients with TP53 mutations in the TCGA-PAAD cohort were included. The normalized RNA-sequencing data (FPKM) were used in the GSEA. The threshold was set at FDR (false discovery rate) < 0.05 and NES (normalized enrichment score) > 1.5. The results were drawn with the ggplot2 R package.

Analysis of total leucocyte infiltration and 22 immune cell types' infiltration patterns

To explore the differences in tumor immune infiltrating cells between TP53-mutated and TP53-wild patients, the estimations of total leucocyte infiltration were performed by the ESTIMATE R package [15], and the fractions of 22 immune cell types (B cells naïve, B cells memory, Plasma cells, T cells CD8, T cells CD4 naïve, T cells CD4 memory resting, T cells CD4 memory activated, T cells follicular helper, T cells regulatory (Tregs), T cells gamma delta, NK cells resting, NK cells activated, Monocytes, Macrophages M0, Macrophages M1, Macrophages M2, Dendritic cells resting, Dendritic cells activated, Mast cells resting, Mast cells activated, Eosinophils and Neutrophils) were calculated with CIBERSORT R script v1.03 [16]. Sixty-five patients without TP53 mutations and 77 patients with TP53 mutations in the TCGA-PAAD cohort were included. Immune scores generated from the ESTIMATE algorithm were used to reflect and compare the total leucocyte infiltrations. The CIBERSORT analysis was conducted by using the default signature matrix at 1000 permutations. Results with $P \geq 0.05$ were excluded. 35 wild and 60 mutated patients were included for the further analysis. Differentially analysis was performed with the Wilcoxon test and plotted by the vioplot R package. $P < 0.05$ were considered to be statistically significant.

Pearson' correlations among the infiltrations of the 22 immune cell types and the immune scores were plotted with the corrplot R package in the wild and mutated groups separately.

Screening of survival related TP53-associated genes

To screen candidate genes which were both associated with TP53 mutations and overall survival, WGCNA was applied with the WGCNA R package [17]. There were 67 wild and 82 mutated patients in the TCGA-PAAD cohort with complete overall survival data included. First, outlier samples were excluded by sample clustering. The sample dendrogram and the clinical-traits heatmap were plotted. Then, the scale-free network was constructed with the appropriate soft-threshold power (β) value. The scale-free topology was plotted to show the constructed scale-free network.

The co-expression modules generated from the scale-free network were further plotted with dynamic tree cutting. Modules were merged if their similarity was greater than 0.75 according to dendrogram height. The Pearson's correlation coefficients between each module and clinical traits were further calculated and plotted. Modules with correlations with TP53 mutation greater than 0.2 ($P < 0.05$) and significantly associated with overall survival ($P < 0.05$) were considered as survival-related TP53-associated modules. The genes included in the modules were screened as candidate survival related TP53-associated genes.

Developing and validation of TP53-associated prognostic signature

In this study, the gene expression data of the ICGC-PACA-AU cohort were mainly based on microarray platform. While, the gene expression data of the ICGC-PACA-CA and the TCGA-PAAD cohorts were all based on RNA-sequencing platforms. The sample size of the ICGC-PACA-CA cohort was larger than the TCGA-PAAD cohort. So, the ICGC-PACA-CA cohort was applied as the training set and the TCGA-PAAD cohort was used as the external validation set. First, the univariate Cox regression analysis was done to further identify the survival related TP53-associated genes in the training set with the cutoff of $P < 0.05$. The LASSO and stepwise regression analyses were applied to construct the best-fit TP53-related prognostic signature, which could estimate the risk score of every patient. Based on the median risk score of the training set, all patients were divided into high- and low-risk groups in both sets. Kaplan-Meier (KM) curves were used to perform the survival analysis which could compare the overall survival of different groups. The predictive performances at different endpoints (1, 2, 3, 4, and 5 years) were assessed with the receiver operating characteristic (ROC) curves

in both cohorts. The alteration trends of risk score, survival time, survival status, and expression levels of the genes included in the signature were further plotted with the pheatmap R package in both cohorts separately.

Independent prognostic prediction analysis

To explore the independence of the TP53-related signature, univariate and multivariate Cox regression analyses were conducted with the Survival R package. One hundred forty-three patients with complete clinical information of the age, gender, stage, and risk score were included in the ICGC-PACA-CA cohort. One hundred and seventy-five patients with complete clinical information of the age, gender, grade, stage, alcohol, family history, and risk score were included in the TCGA-PAAD cohort. The hazard ratio (HR) and *P* values were plotted. *P* < 0.05 were considered statistically significant. 5-year Receiver Operating characteristic Curves (ROC) of the risk score and other clinical features were plotted with the survivalROC R package. Area Under Curves (AUCs) were applied to compare the accuracies of different factors' prediction abilities.

Construction of the prognostic Nomogram

Based on the multivariate Cox regression analysis in the independent prognostic prediction analysis, the risk score was the only robust factor significantly (*P* < 0.05) associated with overall survival in both cohorts. The expression data and the risk scores of the two cohorts were combined. A total of 359 patients were included and the prognostic nomogram was constructed with the rms R package. The survival analysis was carried out by KM plotter. The calibration curve was further plotted using the calibrate function. ROC curves of 1, 3, and 5 years were plotted with the survivalROC R package. Then, we conducted the performance comparison of the signature in our study (referred to as TP53Sig) with eight recently published signatures: 6-mRNA signature from Hou's study (referred to as HouSig) [18], 4-mRNA signature from Meng's study (referred to as MengSig) [19], 8-mRNA signature from Meng's study (referred to as MengSig) [20], 5-mRNA signature from Wu's study (referred to as WuSig) [21], 7-mRNA signature from Wu's study (referred to as WuSig) [22], 10-mRNA signature from Yue's study (referred to as YueSig) [23], 2-mRNA signature from Zhou's study (referred to as ZhouSig) [24] and 6-LncRNA signature from Deng's study (referred to as DengSig) [25]. ROC curves of 3 years were plotted and AUC values were calculated using the survivalROC R package.

Survival analysis in subgroups

To evaluate the predictive level of TP53-related prognostic score in different subgroups (Age > 65, Age ≤ 65,

Male, Female, Stage I, Stage II, Stage III-IV, Grade 1, Grade 2 and Grade 3–4), Kaplan-Meier (KM) curves were plotted with the Survival R package.

Relevance with clinical features

The relationships between the risk score and the other clinical features (Age, Gender, Stage and Grade) were explored by using the ggpubr R package in the combined cohort with Kruskal-Wallis test. *P* < 0.05 indicated statistically significant.

Immune cell infiltration analysis and GSEA

To explore the differences in tumor immune infiltrating cells between the high-risk patients (*n* = 183) and the low-risk patients (*n* = 176), the estimations of total leucocyte infiltration were performed by the ESTIMATE R package and the fractions of 22 immune cell types were estimated with CIBERSORT in R 3.6.1. Immune scores generated from the ESTIMATE algorithm were used to reflect and compare the total leucocyte infiltrations with Wilcoxon test. The CIBERSORT analysis was conducted by using the default signature matrix at 1000 permutations. Results with *P* ≥ 0.05 were excluded. Then, according to the previous Charoentong's study [26], we further explored the differential expression analysis of 96 immunotherapy-related genes between the 183 high-risk and 176 low-risk patients with Wilcoxon Test. Finally, the functional differences between the high-risk and low-risk groups were explored with gene set enrichment analysis in GSEA 4.0.1 software.

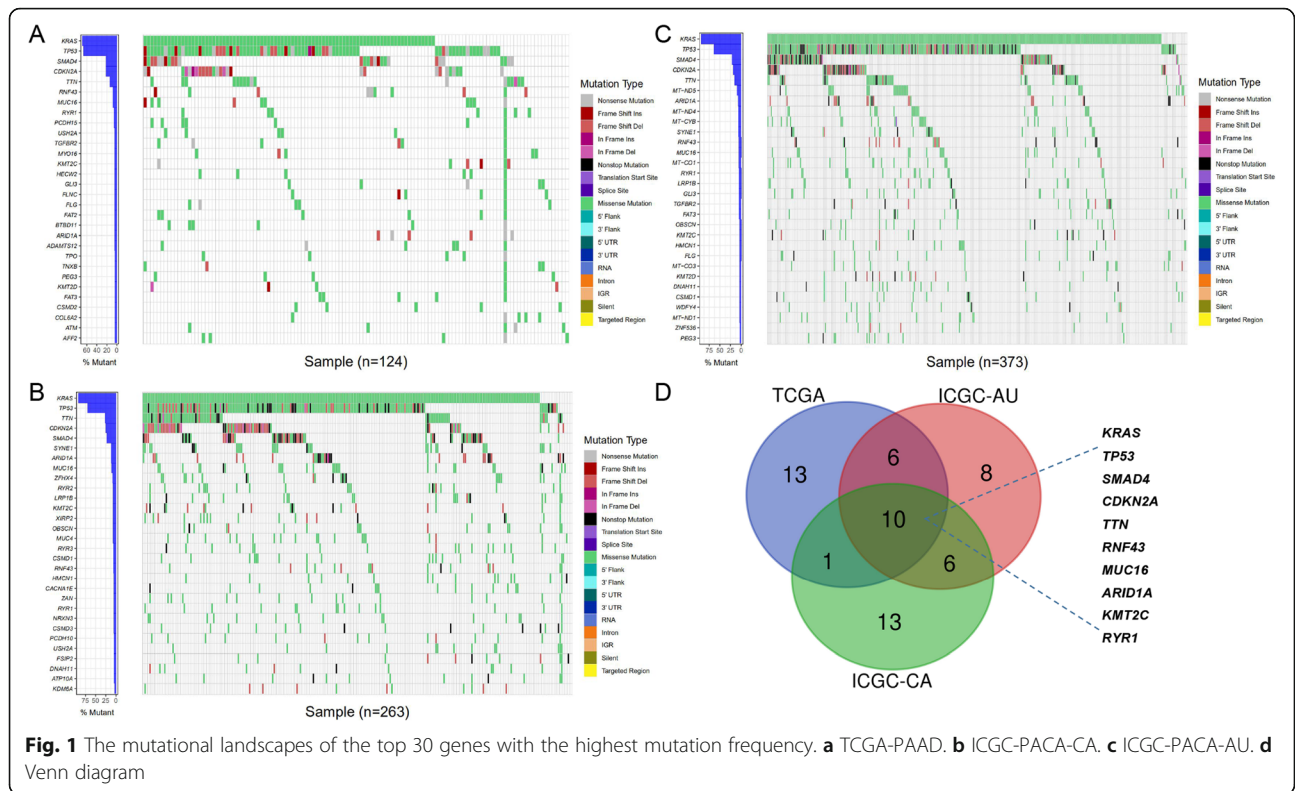
Statistical analysis

The Kaplan–Meier method was used to perform survival analysis, and the log-rank test was used to assess the difference. Differentially analysis of TMB values, immune scores and 22 immune cell types' infiltrations were conducted with the Wilcoxon test. Univariate and multivariate Cox regression analysis were used to assess the independence of the risk score. The performances of the signatures were evaluated by the ROC curves. The relationships between the risk score and the other clinical features were explored with Kruskal-Wallis test. *P* < 0.05 indicated statistically significant.

Results

Screening TP53 mutation to be robust and survival-related mutation type

Mutational landscapes of the top 30 genes with the highest mutation frequency in the three cohorts were plotted separately (Fig. 1a-c). As shown in the results, KRAS and TP53 mutations were the top two mutation types. The mutation frequencies of SMAD4, CDKN2A, and TTN mutations ranked the third to the fifth in the three cohorts. The common ten (KRAS, TP53, SMMAD4,



CDKN2A, TTN, RNF43, MUC16, ARID1A, KMT2C, and RYR1 mutations were drawn by the Venn diagram (Fig. 1d). In order to explore whether these mutations were associated with the overall survival robustly, Kaplan-Meier (KM) curves comparing the mutated group with the wild group were performed in the three cohorts separately. Finally, TP53 mutation was indicated to be the only robust and survival-related mutation type. Kaplan-Meier (KM) curves were shown (Fig. 2a-c). TP53-mutated patients had significantly worse overall survival in the three cohorts ($P < 0.05$).

Association between TP53 mutation and tumor mutation burden (TMB)

TMB values of TP53-mutated group were significantly higher in TCGA-PAAD cohort ($P = 0.001$) and ICGC-PACA-AU cohort ($P < 0.001$). However, no significant difference was found in the ICGC-PACA-CA cohort ($P = 0.93$) (Fig. 2d). Whether TP53 mutation affects TMB needs to be further explored.

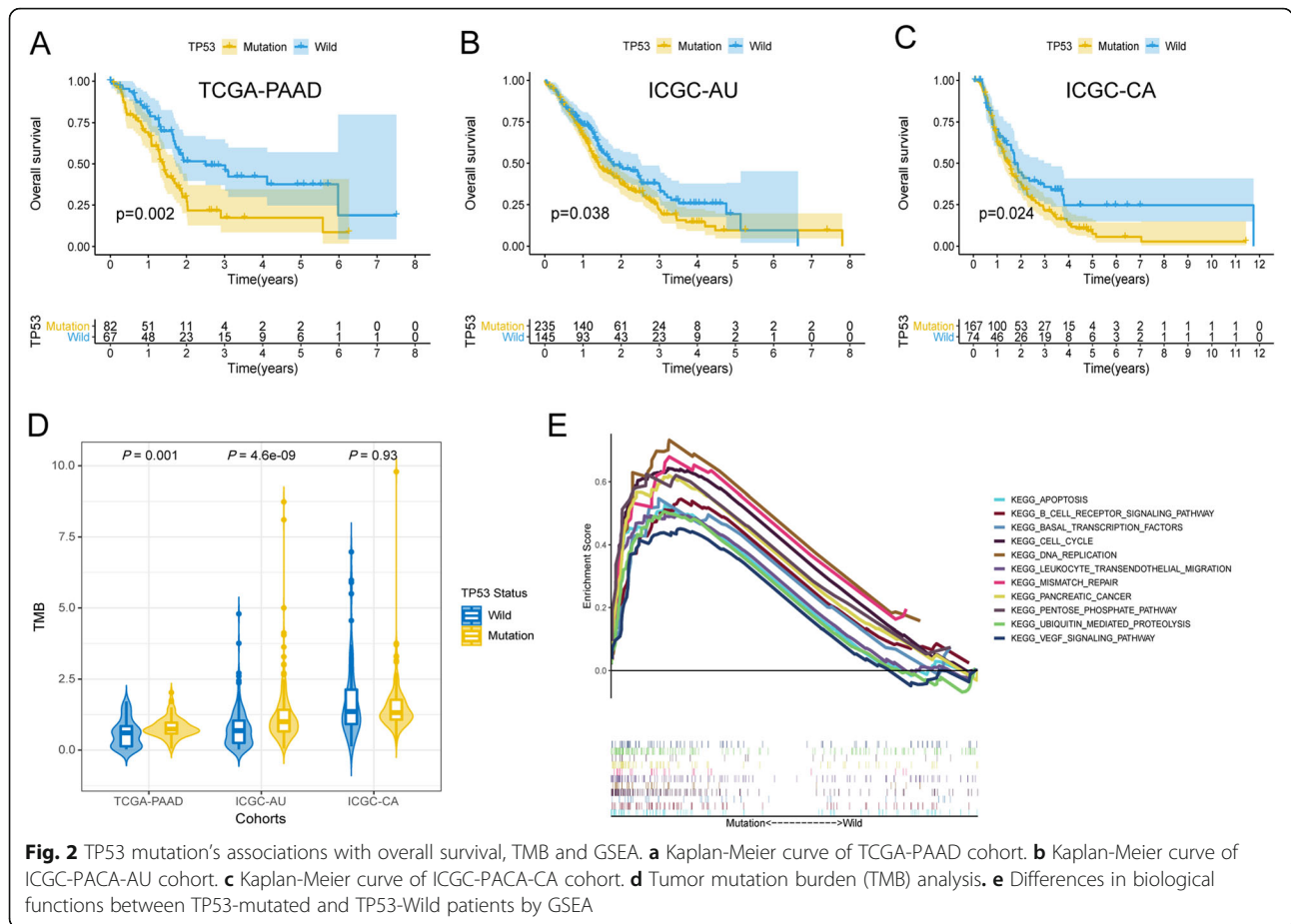
TP53 mutation was associated with immune cell infiltration

GSEA results indicated that the B cell receptor signaling pathway and leukocyte transendothelial migration were significantly enriched in the TP53-mutated patients (Fig. 2e). Immune scores reflecting the total leukocyte infiltrations by ESTIMATE showed no significant

difference between the two groups (Fig. 3a). However, fraction patterns of the 22 immune cell types by CIBERSORT were different. In details, the fractions of plasma cells ($P = 0.023$), T cells CD8 ($P = 0.049$), Monocytes ($P = 0.047$) and Mast cells resting ($P = 0.03$) in TP53-mutated patients were significantly lower than those in TP53-wild patients (Fig. 3b). Moreover, the immune score had the most positive correlation with the fraction of Neutrophils ($R = 0.33$) and the most negative correlation with the fraction of Macrophages M0 ($R = -0.41$) in the wild group (Fig. 3c). And the immune score had the most positive correlation with the fraction of T cells CD4 memory activated ($R = 0.31$) and the most negative correlation with the fraction of B cells naïve ($R = -0.27$) in the mutated group (Fig. 3d).

Screening of candidate TP53-associated genes by WGCNA

To screen candidate TP53-associated genes, WGCNA was performed in the TCGA-PAAD cohort. First, six outlier samples were excluded by sample clustering. The sample dendrogram and the clinical-traits heatmap was plotted (Fig. 4a). Then, the scale-free network was constructed with the soft-threshold power (β) value as 8 (Fig. 4b). The constructed scale-free network was shown by the scale-free topology, with $R^2 = 0.99$ and slope = -1.4 (Fig. 4c). The co-expression modules generated from the scale-free network were further plotted with dynamic tree cutting (Fig. 4d). The Pearson's correlation



coefficients between each module and clinical traits (Additional file 1) were plotted (Fig. 4e). Six Modules (magenta, darkgreen, black, darkorange, steelblue, tan) with correlations with TP53 mutation greater than 0.2 ($P < 0.05$) and significantly related with the overall survival ($P < 0.05$) were considered as survival related TP53-associated modules. A total of 2455 genes included in the six modules were screened as candidate survival related TP53-associated genes.

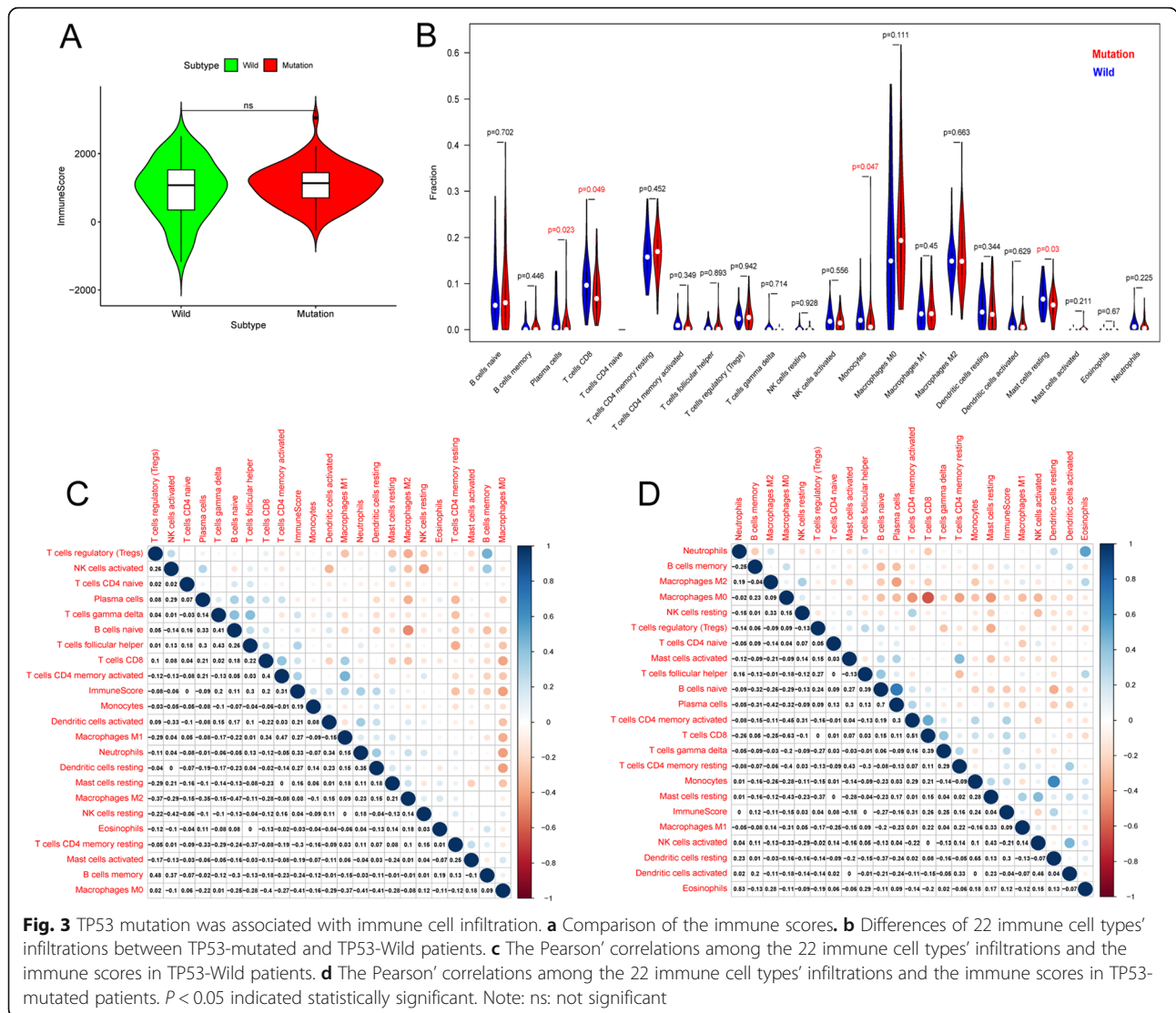
Developing and validation of TP53-related prognostic signature

The ICGC-PACA-CA cohort ($n = 182$) was applied as the training set and the TCGA-PAAD cohort ($n = 177$) was used as the external validation set. Patient characteristics in the training and the testing cohorts were shown (Table 1). First, univariate cox regression analysis was done in the training set and a total of 316 genes were identified with $P < 0.05$. LASSO regression analysis was further applied and 10 genes were screened (Fig. 5a, b). Then, the stepwise regression analysis was used and constructed the best-fit TP53-related prognostic signature calculating the risk score of each patient. Seven genes were included in the signature calculating the risk score

with the formula: risk score = $(0.1254 \times ERFF11) - (0.1365 \times IL6R) - (0.4400 \times PPP1R10) - (0.3397 \times PTOV1-AS2) + (0.1544 \times SCEL) - (0.4412 \times SSSX2IP) - (0.2231 \times TXNL4A)$. The hazard ratios (HRs) were plotted (Fig. 5c). Based on the median risk score 1.005987 of the training set, all patients were divided into high- and low-risk groups in both training and external validation sets. Kaplan-Meier (KM) curves showed the overall survival of the high-risk patients was worse than that of the low-risk patients ($P < 0.001$) (Fig. 5d-e). The ROC curves of 1-, 2-, 3-, 4-, and 5-year were plotted, with the AUC of 0.731, 0.765, 0.808, 0.774, and 0.873 in the training set (Fig. 5f) and the AUC of 0.703, 0.648, 0.677, 0.714, and 0.737 in the external validation set (Fig. 5g). Alteration trends of risk score, survival time, survival status and the included genes' expression levels were further plotted separately (Fig. 6a-f).

Independence of the TP53-related prognostic signature

Univariate and multivariate Cox regression analyses were performed to explore the independence of the TP53-related prognostic signature from other clinical features. The hazard ratio (HR) and P values were plotted. The 5-year Receiver Operating Characteristic Curves (ROC) of



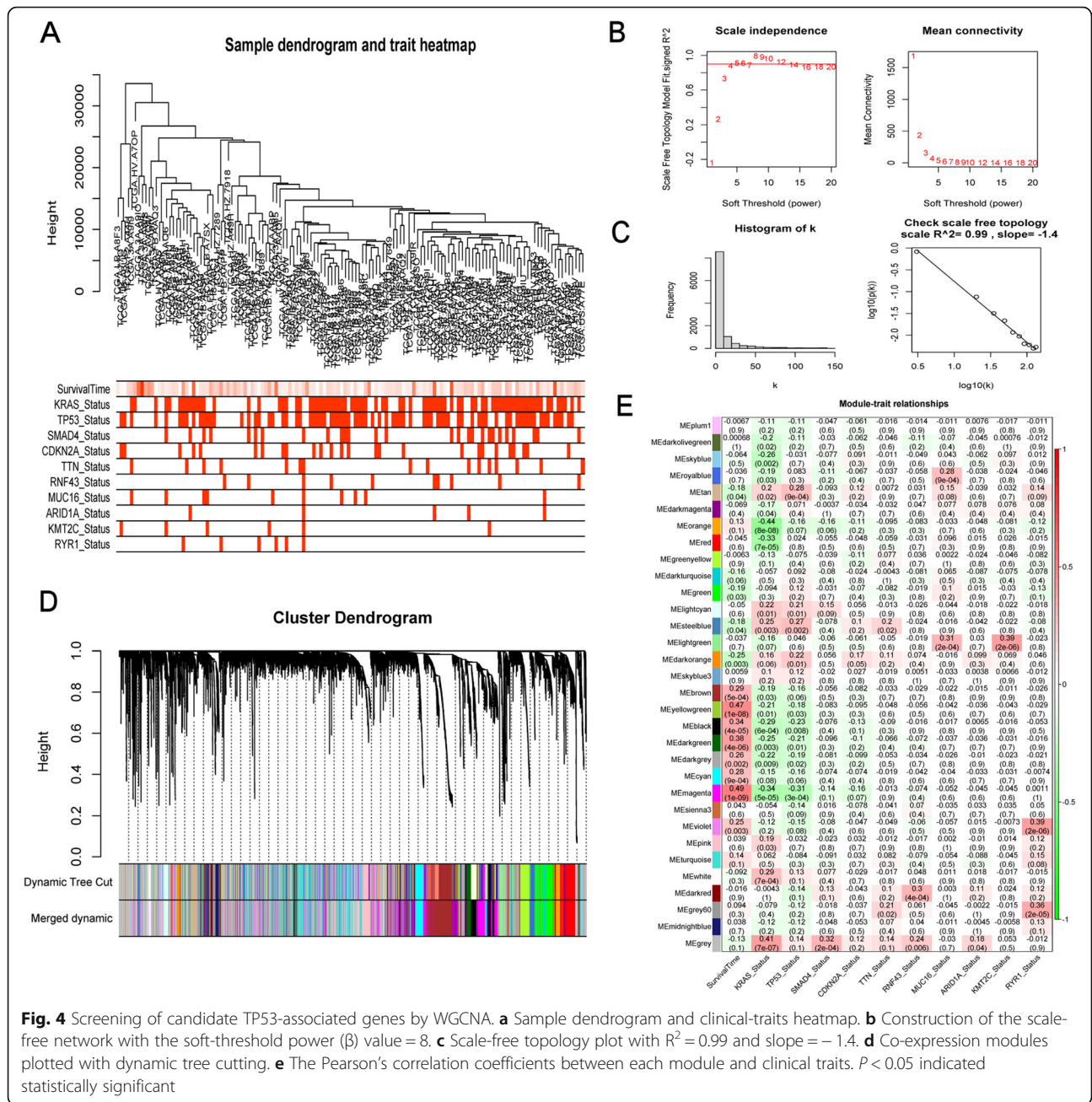
the risk score and other clinical features were plotted. The risk score was proved to be the only variable significantly associated with overall survival in the univariate ($P < 0.001$) and multivariate ($P < 0.001$) Cox regression results in the ICGC-PACA-CA cohort (Fig. 7a, b). The 5-year ROC curve was plotted with the highest value of 0.844 than that of the other clinical features included (Fig. 7c). The risk score ($P < 0.001$), age ($P = 0.012$), and grade ($P = 0.007$) were the three variables significantly associated with overall survival in the univariate analysis of the TCGA-PAAD cohort (Fig. 7d). The risk score ($P < 0.001$) and age ($P = 0.038$) were the two significant variables in the multivariate analysis of the TCGA-PAAD cohort (Fig. 7e). The 5-year ROC curve was also plotted with the highest value of 0.717 than that of the other clinical features included (Fig. 7f). These results indicated that the risk score based on the TP53-related

prognostic signature was independent from other clinical features.

Construction of the prognostic Nomogram

The risk score was the only robust factor significantly ($P < 0.05$) associated with overall survival in both cohorts. The expression data and the risk scores of the two cohorts were combined. Three hundred and fifty-nine patients were included and the prognostic nomogram was conducted (Fig. 8a). The survival analysis was carried out by KM plotter (Fig. 8b). The calibration curve of 3 years was well calibrated (Fig. 8c). The ROC curves of 1-, 3-, and 5-year were plotted, with the AUC of 0.713, 0.753, and 0.823 (Fig. 8d).

ROC curves of 3 years indicated that the signature in our study had the highest AUC value (Fig. 8e).



Kaplan-Meier (KM) curves in subgroups also showed good performance except in Stage III-IV and G3-4 groups (Fig. 9f).

Relevance with clinical features

There were no significant relationships between the risk score and Age or Gender in the combined cohort (Fig. 9a, b). Patients in Stage I had significantly lower risk scores than patients in Stage II ($P < 0.001$) (Fig. 9c). Patients in G1 had significantly lower risk scores than patients in G2 and G3-4 ($P < 0.01$) (Fig. 9d). No significant differences were revealed between Stage I and Stage

III-IV, Stage II and Stage III-IV and G2 and G3-4 groups (Fig. 9c, d). The sample size of Stage III-IV ($n = 15$) was small, which might affect the statistical accuracy. In general, the higher the Stage or Grade is, the higher the risk score will be.

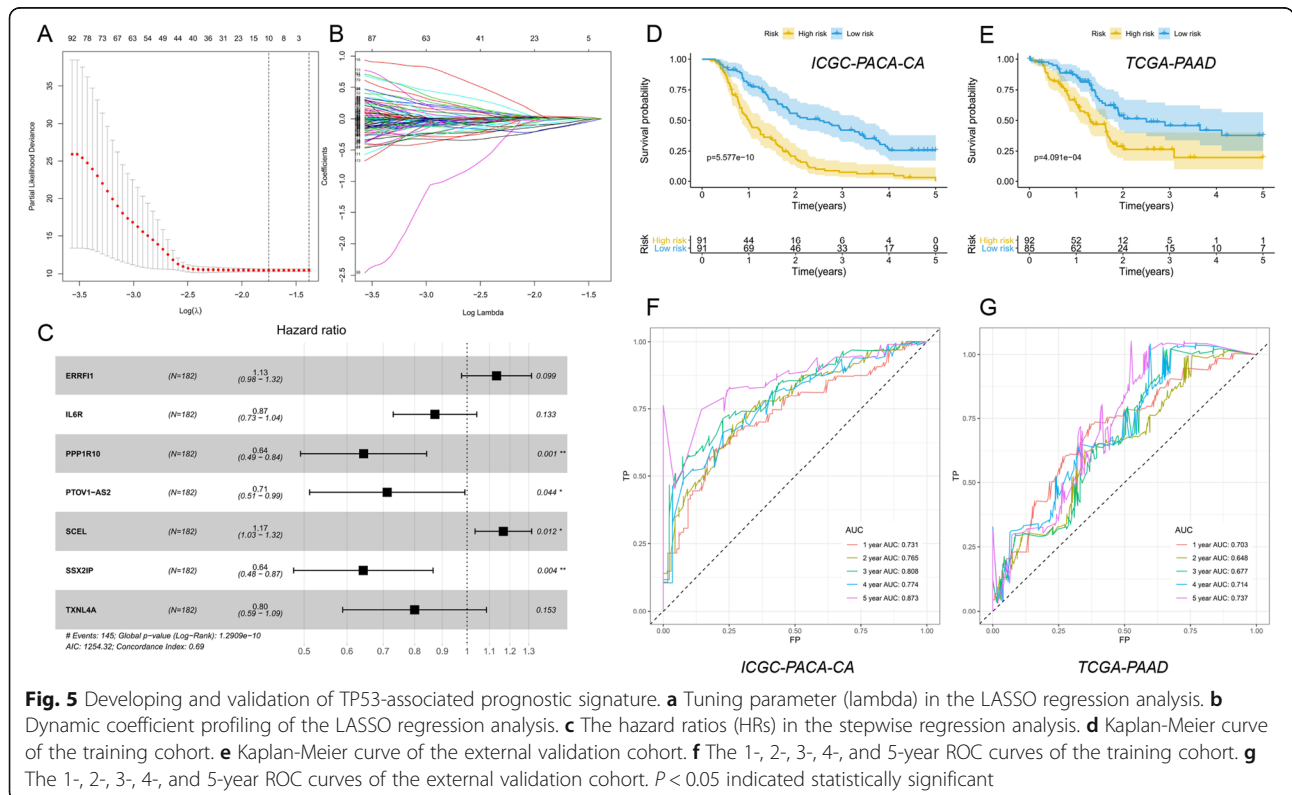
Immune cell infiltration analysis and GSEA

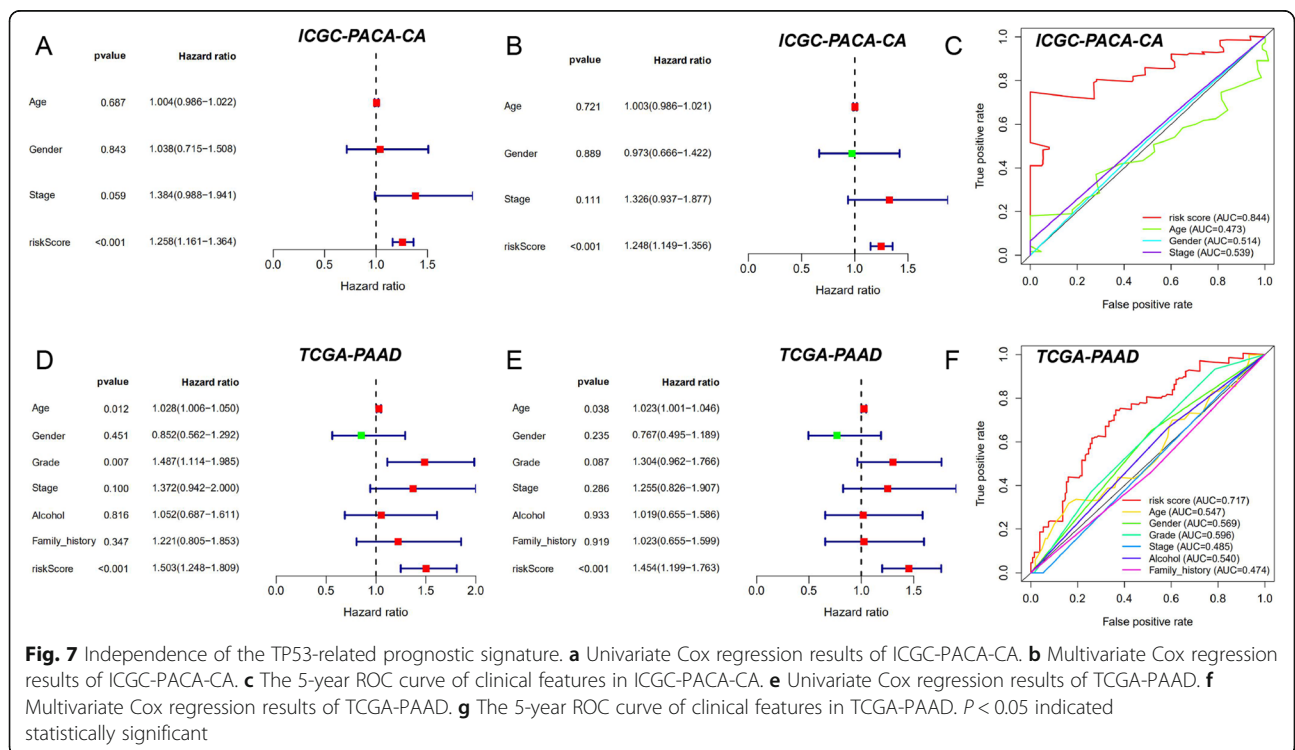
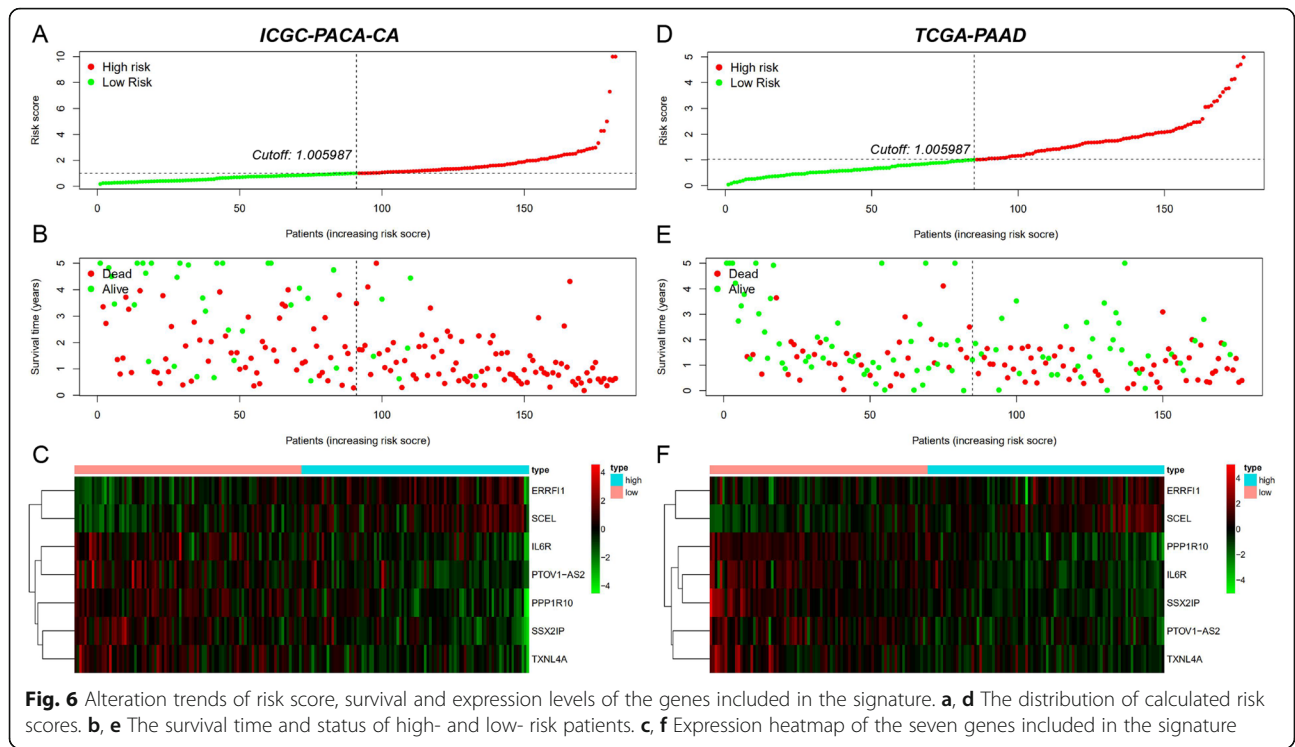
Immune scores by ESTIMATE showed no significant difference between the high-risk and low-risk groups (Fig. 10a). However, fractions of T cells CD8 ($P = 0.011$), T cells CD4 memory resting ($P = 0.012$), and T cells regulatory (Tregs) ($P = 0.007$) in high-risk group were

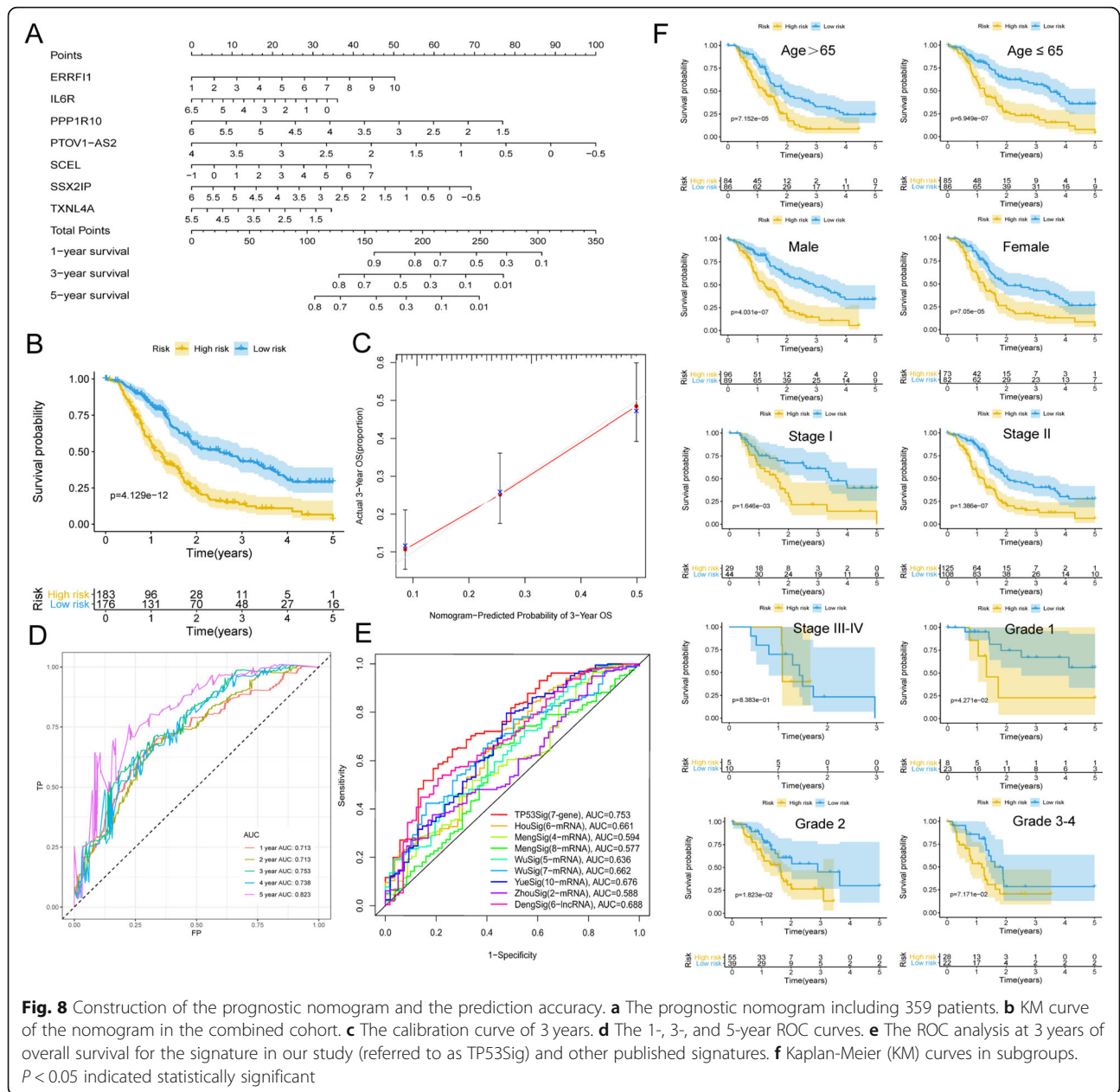
Table 1 Patient characteristics in the training and the testing cohorts

ICGC-PACA-CA cohort (Training)			TCGA-PAAD cohort (Testing)		
Characteristics	Number of cases	% (percentage)	Characteristics	Number of cases	% (percentage)
Age			Age		
<=65	78	42.86	<=65	93	52.54
> 65	86	47.25	> 65	84	47.46
Unknown	18	9.89	Unknown	0	0
Gender			Gender		
Male	99	54.40	Male	97	54.80
Female	82	45.05	Female	80	45.20
Unknown	1	0.55	Unknown	0	0
Stage			Stage		
Stage I	51	28.02	Stage I	22	12.43
Stage II	85	46.70	Stage II	148	83.62
Stage III-IV	8	4.40	Stage III-IV	7	3.95
Unknown	38	20.88	Unknown	0	0
Grade			Grade		
G1	NA	NA	G1	31	17.51
G2	NA	NA	G2	94	53.11
G3-4	NA	NA	G3-4	50	28.25
Unknown	NA	NA	Unknown	2	1.13

Note: NA Not Available







significantly lower (Fig. 10b). The fraction of Macrophages M0 in high-risk group was significantly higher (Fig. 10b). We further explored the differential expression analysis of 96 immunotherapy-related genes (Additional file 2). A total of 31 genes were identified, of which 26 genes' expression levels were lower in the high-risk group. This might indicate that the immune activity in the high-risk group was weaker than that in the low-risk group (Fig. 10c). GSEA results indicated that Epithelial-Mesenchymal Transition and Hypoxia were the most significantly enriched functions in the high-risk patients (Fig. 10d).

Discussion

Gene alterations have been determined in 97% of patients with PDAC, including point mutations, amplifications, deletions, translocations, and inversions [27]. TP53 encoded tumor suppressor TP53, which transcriptionally activated target genes for resisting cellular stresses, then inducing growth arrest or apoptosis [28]. TP53 mutation has been observed in 50–75% of PDAC cases, which would initiate activating mutation of the KRAS gene. Mutant TP53 promoted lymph node metastasis and escaped from KrasG12D-induced growth arrest/senescence in PDAC [29]. In clinical practices, the detection

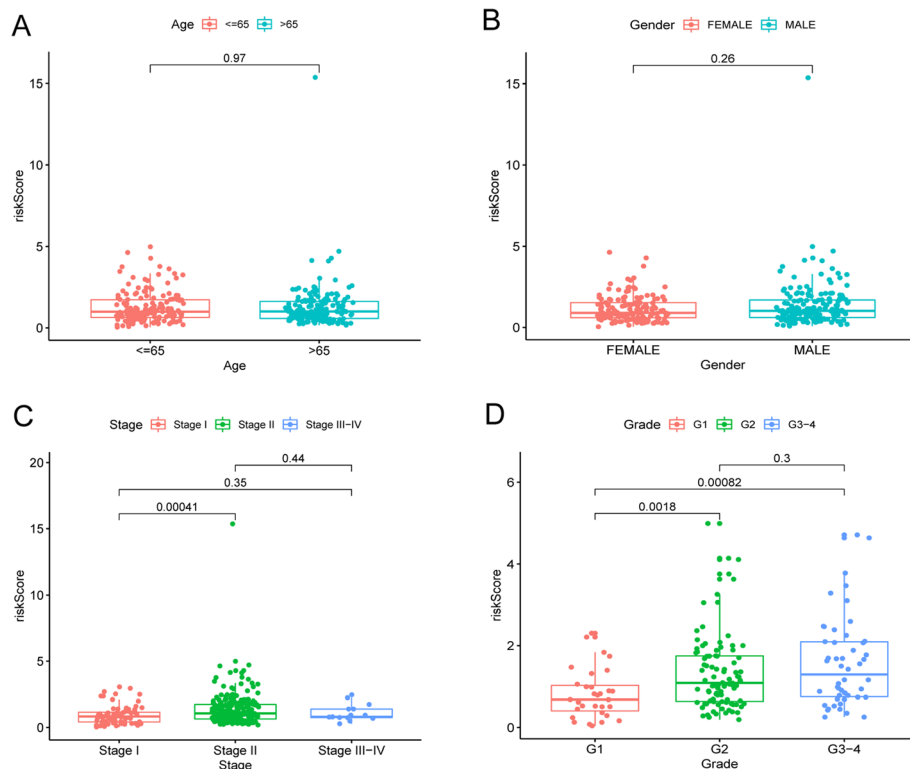


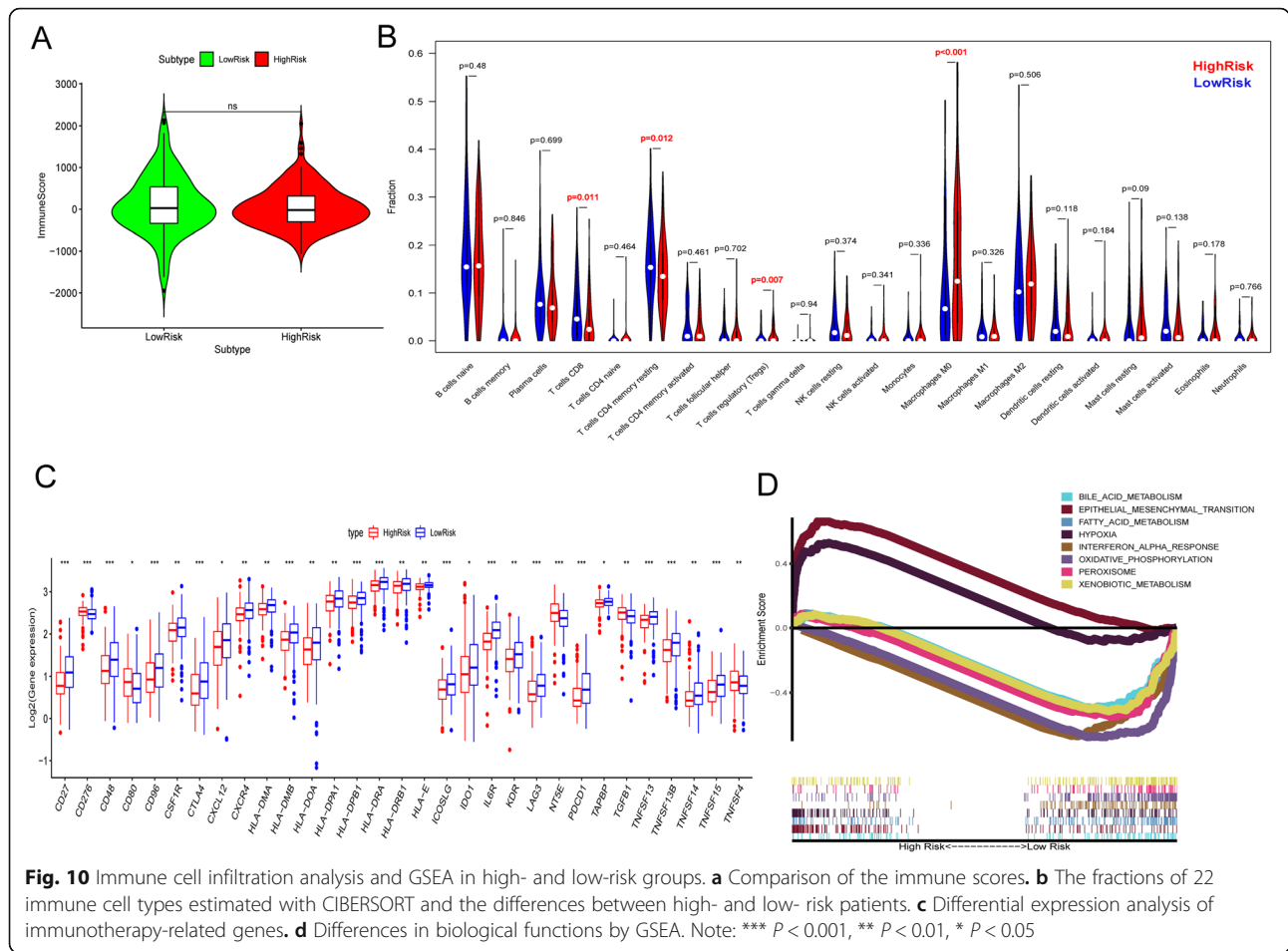
Fig. 9 The relationships between risk score and clinical features. **a** Relationships between risk score and the Age, Gender, Stages in the combined cohort. **b** Relationships between risk score and the Grade in the TCGA-PAAD cohort. $P < 0.05$ indicated statistically significant. Note: the data of Grade were not provided in the ICGC-PACA-CA dataset

of mutant and expression profile of TP53 and associated genes may improve the diagnosis and screen for PC [30, 31]. The sample can be the pancreatic juice, less-invasive serum exosomes or even non-invasive stool [32, 33]. However, until now, although TP53 have been closely associated with PC, the roles of TP53 have not been elucidated. More information on its upstream and downstream genes would be required [27]. Our bioinformatic analysis has also proved the roles of TP53 in PDAC. Firstly, KRAS and TP53 mutations were observed as the top two genes with the highest mutation frequency. Secondly, TP53 mutation was indicated to be the only robust and survival-related mutation type. Thirdly, TP53-mutated patients showed significantly worse overall survival than TP53-wild patients in included cohorts.

Based on above results, more genes associated with TP53 mutations were further identified. Key modules including 2455 genes were preliminarily screened with WGCNA and Pearson's correlation. The key 316 genes were secondly screened with univariate Cox regression analysis. Then, 10 most critical genes were identified with LASSO. Finally, the best-fit TP53-related prognostic signature involved 7 genes were constructed with stepwise regression analysis, as well as corresponding coefficients. In the formula for calculating risk score, the

high levels of *ERRFI1*, *SCEL* increased risk of poor prognosis, while the high levels of *IL6R*, *PPP1R10*, *PTOV1-AS2*, *SSX2IP*, *TXNL4A* suggested decreasing risk. The identified 7 genes were potential candidates of biomarkers for prognosis prediction of PDAC [34].

In previous studies, *ERBB receptor feedback inhibitor 1* (*ERRFI1*) was reported as an important regulatory gene. It regulated AKT/EGFR signaling in an EGFR-dependent manner. In EGFR-low cells, *ERRFI1* activated AKT and promoted proliferation and chemotherapy resistance. In EGFR-high cells, reduced *ERRFI1* led to active EGFR and increased cell proliferation [35]. As a key gene targeted AKT/EGFR signaling, *ERRFI1* may be a binding target for some miRNAs and lncRNAs in various cancers, such as cholangiocarcinoma [36, 37]. Until now, the effects of *ERRFI1* on PDAC have not been revealed, and its effects in other cancers suggested that it may be promising therapeutic target and biomarker. Interleukin-6 receptor (IL-6R) was the receptor of IL-6. The blocking agents that combined with IL-6 and IL-6R may be potential anti-inflammatory drugs, and some of them may be anti-cancer agents [38]. As an important effector in several signaling pathways, IL6R was also proposed as a new therapeutic target for some cancers [39, 40]. For example, IL-6R participated IL-6R /STAT3/miR-204



feedback loop contributed to chemo-resistance of epithelial ovarian cancer cells [41]. IL-6R participated several important pathways in tumor development and chemo-resistance. The suppression of IL-6R may function in these pathways thus affecting the prognosis. The same to ERFF11, the roles of IL-6R in PDAC have not been revealed, which worth further investigation. *SCEL* encoded Sciellin (SCEL), which was a precursor of the cornified envelope firstly identified in mammalian keratinizing tissue [42]. SCEL was a mesenchymal-to-epithelial transition inducer dynamically regulated during the metastasis. Thus it may be a site for regulating the colorectal cancer hepatic metastasis [43]. *SSX2IP* encoded Synovial Sarcoma X breakpoint 2 Interacting Protein (SSX2IP) has been revealed to play various roles in human cancers. SSX2IP was known as leukaemia associated antigen [44, 45]. SSX2IP promoted the tumorigenesis and progression of hepatocellular carcinoma and contributed to the drug resistance, enabling it a new biomarker and specific target in hepatocellular carcinoma [46]. High levels of SSX2IP were associated with aggressive pathological features and poor outcomes in nasopharyngeal carcinoma [47]. Few studies have been

performed on PPP1R10, PTOV1 antisense 2 (PTOV1-AS2) and Thioredoxin-like protein 4A(TXNLA4A). Considering the malignancy of PDAC, relatively less studies could be retrieved, both on the biomarkers or therapeutic targets. The bioinformatic analysis has been a useful tool for exploring promising candidates for further investigation.

Conclusions

With the screened gene candidates, the TP53-associated signature exhibited good prognostic efficacy in predicting the overall survival of PC patients. In the training and validation cohorts, the 1-, 3-, and 5-year ROC curves were plotted according to the risk score, presenting relatively high AUC. The prognostic nomogram including 359 patients was also constructed to calculate the risk score, which would facilitate the further clinical applications. There are also some limitations in our study. Firstly, the gene expression and somatic cell mutations data were obtained from cancerous tissues. However, the results should be further verified in other types of samples, such as serum and stool. As the less-invasive and non-invasive specimens will tend to promote the

early screen and detection of PDAC. Secondly, the association between TP53 mutation and tumor mutation burden should be verified in larger cohort. Thirdly, whether the nomogram could predict immunotherapy or chemotherapy needs further explorations in the future.

Abbreviations

PC: Pancreatic cancer; KM: Kaplan-Meier; TMB: Tumor mutation burden; GSEA: Gene set enrichment analysis; WGCNA: Weighted Gene Co-expression Network Analysis; LASSO: Least absolute shrinkage and selection operator; ROC: Receiver operating characteristic; AUCs: Area Under Curves; TCGA: The Cancer Genome Atlas; ICGC: International Cancer Genome Consortium; FDR: False discovery rate; NES: Normalized enrichment score; HR: Hazard ratio

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-021-08066-2>.

Additional file 1.

Additional file 2.

Authors' contributions

X. Liu and S. Sun designed the study. Collection and assembly of data were performed by B. Chen and J. Chen. Data analysis and interpretation were done by X. Liu. Manuscript writing was finished by X. Liu and S. Sun. All authors approved the manuscript.

Funding

No funding was received.

Availability of data and materials

The data used and analyzed in this study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

None of the authors have competing interests in the manuscript.

Received: 27 December 2020 Accepted: 15 March 2021

Published online: 31 March 2021

References

- 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. *Cancer Res*. 2014; 74(11):2913–21. <https://doi.org/10.1158/0008-5472.CAN-14-0155>.
- Ilic M, Ilic I. Epidemiology of pancreatic cancer. *World J Gastroenterol*. 2016; 22(44):9694–705. <https://doi.org/10.3748/wjg.v22.i44.9694>.
- Ansari D, Tingstedt B, Andersson B, Holmquist F, Stureson C, Williamsson C, et al. Pancreatic cancer: yesterday, today and tomorrow. *Future Oncol* (London, England). 2016;12(16):1929–46.
- Dreyer SB, Chang DK, Bailey P, Biankin AV. Pancreatic Cancer genomes: implications for clinical management and therapeutic development. *Clin Cancer Res*. 2017;23(7):1638–46. <https://doi.org/10.1158/1078-0432.CCR-16-2411>.
- McGuigan A, Kelly P, Turkington RC, Jones C, Coleman HG, McCain RS. Pancreatic cancer: a review of clinical diagnosis, epidemiology, treatment and outcomes. *World J Gastroenterol*. 2018;24(43):4846–61. <https://doi.org/10.3748/wjg.v24.i43.4846>.
- Kim JE, Lee KT, Lee JK, Paik SW, Rhee JC, Choi KW. Clinical usefulness of carbohydrate antigen 19-9 as a screening test for pancreatic cancer in an asymptomatic population. *J Gastroenterol Hepatol*. 2004;19(2):182–6. <https://doi.org/10.1111/j.1440-1746.2004.03219.x>.
- Ansari D, Torén W, Zhou Q, Hu D, Andersson R. Proteomic and genomic profiling of pancreatic cancer. *Cell Biol Toxicol*. 2019;35(4):333–43. <https://doi.org/10.1007/s10565-019-09465-9>.
- Klein AP, Wolpin BM, Risch HA, Stolzenberg-Solomon RZ, Mocci E. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nat Commun*. 2018;9(1):556.
- Torres C, Grippo PJ. Pancreatic cancer subtypes: a roadmap for precision medicine. *Ann Med*. 2018;50(4):277–87. <https://doi.org/10.1080/07853890.2018.1453168>.
- Wu J, Li Z, Zeng K, Wu K, Xu D, Zhou J, et al. Key genes associated with pancreatic cancer and their association with outcomes: a bioinformatics analysis. *Mol Med Rep*. 2019;20(2):1343–52. <https://doi.org/10.3892/mmr.2019.10321>.
- Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22(3):568–76. <https://doi.org/10.1101/gr.129684.111>.
- Kassahn KS, Holmes O, Nones K, Patch AM, Miller DK, Christ AN, et al. Somatic point mutation calling in low cellularity tumors. *PLoS One*. 2013; 8(11):e74380. <https://doi.org/10.1371/journal.pone.0074380>.
- Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213–9. <https://doi.org/10.1038/nbt.2514>.
- Skidmore ZL, Wagner AH, Lesurf R, Campbell KM, Kunisaki J, Griffith OL, et al. GenVisR: genomic visualizations in R. *Bioinformatics*. 2016;32(19):3012–4. <https://doi.org/10.1093/bioinformatics/btw325>.
- Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-García W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4(1):2612. <https://doi.org/10.1038/ncomms3612>.
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–7. <https://doi.org/10.1038/nmeth.3337>.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559. <https://doi.org/10.1186/1471-2105-9-559>.
- Hou J, Wang Z, Li H, Zhang H, Luo L. Gene signature and identification of clinical trait-related m(6) a regulators in pancreatic Cancer. *Front Genet*. 2020;11:522. <https://doi.org/10.3389/fgene.2020.00522>.
- Meng Z, Yuan Q, Zhao J, Wang B, Li S, Offringa R, et al. The m(6) A-related mRNA signature predicts the prognosis of pancreatic Cancer patients. *Mol Ther Oncolytics*. 2020;17:460–70. <https://doi.org/10.1016/j.omto.2020.04.011>.
- Meng Z, Ren D, Zhang K, Zhao J, Jin X, Wu H. Using ESTIMATE algorithm to establish an 8-mRNA signature prognosis prediction system and identify immunocyte infiltration-related genes in pancreatic adenocarcinoma. *Aging* (Albany NY). 2020;12(6):5048–70. <https://doi.org/10.18632/aging.102931>.
- Wu C, Wu Z, Tian B. Five gene signatures were identified in the prediction of overall survival in resectable pancreatic cancer. *BMC Surg*. 2020;20(1):207. <https://doi.org/10.1186/s12893-020-00856-y>.
- Wu M, Li X, Liu R, Yuan H, Liu W, Liu Z. Development and validation of a metastasis-related gene signature for predicting the overall survival in patients with pancreatic ductal adenocarcinoma. *J Cancer*. 2020;11(21): 6299–318. <https://doi.org/10.7150/jca.47629>.
- Yue P, Zhu C, Gao Y, Li Y, Wang Q, Zhang K, et al. Development of an autophagy-related signature in pancreatic adenocarcinoma. *Biomed Pharmacother*. 2020;126:110080. <https://doi.org/10.1016/j.biopha.2020.110080>.
- Zhou S, Yan Y, Chen X, Zeng S, Wei J, Wang X, et al. A two-gene-based prognostic signature for pancreatic cancer. *Aging* (Albany NY). 2020;12(18): 18322–42. <https://doi.org/10.18632/aging.103698>.
- Deng Z, Li X, Shi Y, Lu Y, Yao W, Wang J. A novel autophagy-related lncRNAs signature for prognostic prediction and clinical value in patients with pancreatic Cancer. *Front Cell Dev Biol*. 2020;8:606817. <https://doi.org/10.3389/fcell.2020.606817>.
- Charoentong P, Finotello F, Angelova M, Mayer C, Efreanova M, Rieder D, et al. Pan-cancer Immunogenomic analyses reveal genotype-

- Immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* 2017;18(1):248–62. <https://doi.org/10.1016/j.celrep.2016.12.019>.
27. Cicens J, Kvederaviciute K, Meskinyte I, Meskinyte-Kausiliene E, Skeberdyte A, Cicens J. KRAS, TP53, CDKN2A, SMAD4, BRCA1, and BRCA2 mutations in pancreatic cancer. *Cancers.* 2017;9(5):42.
 28. Levy N, Yonish-Rouach E, Oren M, Kimchi A. Complementation by wild-type p53 of interleukin-6 effects on M1 cells: induction of cell cycle exit and cooperativity with c-myc suppression. *Mol Cell Biol.* 1993;13(12):7942–52. <https://doi.org/10.1128/MCB.13.12.7942>.
 29. Morton JP, Timpson P, Karim SA, Ridgway RA, Athineos D, Doyle B, et al. Mutant p53 drives metastasis and overcomes growth arrest/senescence in pancreatic cancer. *Proc Natl Acad Sci U S A.* 2010;107(1):246–51. <https://doi.org/10.1073/pnas.0908428107>.
 30. Kanda M, Sadakari Y, Borges M, Topazian M, Farrell J, Syngal S, et al. Mutant TP53 in duodenal samples of pancreatic juice from patients with pancreatic cancer or high-grade dysplasia. *Clin Gastroenterol Hepatol.* 2013;11(6):719–730.e715.
 31. Ormanns S, Siveke JT, Heinemann V, Haas M, Sipos B, Schlitter AM, et al. pERK, pAKT and p53 as tissue biomarkers in erlotinib-treated patients with advanced pancreatic cancer: a translational subgroup analysis from AIO-PK0104. *BMC Cancer.* 2014;14(1):624. <https://doi.org/10.1186/1471-2407-14-624>.
 32. Kahlert C, Melo SA, Protopopov A, Tang J, Seth S, Koch M, et al. Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. *J Biol Chem.* 2014;289(7):3869–75. <https://doi.org/10.1074/jbc.C113.532267>.
 33. Lu X, Xu T, Qian J, Wen X, Wu D. Detecting K-ras and p53 gene mutation from stool and pancreatic juice for diagnosis of early pancreatic cancer. *Chin Med J.* 2002;115(11):1632–6.
 34. Mukherjee I, Powell B, Parianos M, Downs D, Ross SB. Available technologies and clinical applications of targeted chemotherapy in pancreatic cancer. *Cancer Genet.* 2016;209(12):582–91. <https://doi.org/10.1016/j.cancergen.2016.08.001>.
 35. Cairns J, Fridley BL, Jenkins GD, Zhuang Y, Yu J, Wang L. Differential roles of ERFF1 in EGFR and AKT pathway regulation affect cancer proliferation. *EMBO Rep.* 2018;19(3):e44767.
 36. Yu Y, Chen Q. Long noncoding RNA ANRIL promotes the malignant progression of cholangiocarcinoma by epigenetically repressing ERFF1 expression. *Cancer Sci.* 2020;111(7):2297–309.
 37. Migliore C, Morando E, Ghiso E, Anastasi S, Leoni VP, Apicella M, et al. miR-205 mediates adaptive resistance to MET inhibition via ERFF1 targeting and raised EGFR signaling. *EMBO Mol Med.* 2018;10(9):e8746.
 38. Heo TH, Wahler J, Suh N. Potential therapeutic implications of IL-6/IL-6R/gp130-targeting agents in breast cancer. *Oncotarget.* 2016;7(13):15460–73. <https://doi.org/10.18632/oncotarget.7102>.
 39. Høgdall D, O'Rourke CJ, Dehlendorff C, Larsen OF, Jensen LH, Johansen AZ, et al. Serum IL6 as a prognostic biomarker and IL6R as a therapeutic target in biliary tract cancers. *Clin Cancer Res.* 2020;26(21):5655–67. <https://doi.org/10.1158/1078-0432.CCR-19-2700>.
 40. Colomiere M, Ward AC, Riley C, Trenerry MK, Cameron-Smith D, Findlay J, et al. Cross talk of signals between EGFR and IL-6R through JAK2/STAT3 mediate epithelial-mesenchymal transition in ovarian carcinomas. *Br J Cancer.* 2009;100(1):134–44. <https://doi.org/10.1038/sj.bjc.6604794>.
 41. Zhu X, Shen H, Yin X, Long L, Chen X, Feng F, et al. IL-6R/STAT3/miR-204 feedback loop contributes to cisplatin resistance of epithelial ovarian cancer cells. *Oncotarget.* 2017;8(24):39154–66. <https://doi.org/10.18632/oncotarget.16610>.
 42. Nagy A, Banyai D, Semjen D, Beothe T, Kovacs G. Scellin is a marker for papillary renal cell tumours. *Virchows Arch.* 2015;467(6):695–700. <https://doi.org/10.1007/s00428-015-1856-y>.
 43. Chou CK, Fan CC, Lin PS, Liao PY, Tung JC, Hsieh CH, et al. Scellin mediates mesenchymal-to-epithelial transition in colorectal cancer hepatic metastasis. *Oncotarget.* 2016;7(18):25742–54. <https://doi.org/10.18632/oncotarget.8264>.
 44. Breslin A, Denniss FA, Guinn BA. SSX2IP: an emerging role in cancer. *Biochem Biophys Res Commun.* 2007;363(3):462–5. <https://doi.org/10.1016/j.bbrc.2007.09.052>.
 45. Guinn B, Greiner J, Schmitt M, Mills KI. Elevated expression of the leukemia-associated antigen SSX2IP predicts survival in acute myeloid leukemia patients who lack detectable cytogenetic rearrangements. *Blood.* 2009;113(5):1203–4. <https://doi.org/10.1182/blood-2008-09-178848>.
 46. Li P, Lin Y, Zhang Y, Zhu Z, Huo K. SSX2IP promotes metastasis and chemotherapeutic resistance of hepatocellular carcinoma. *J Transl Med.* 2013;11(1):52. <https://doi.org/10.1186/1479-5876-11-52>.
 47. Chang SL, Lee SW, Yang SF, Chien CC, Chan TC, Chen TJ, et al. Expression and prognostic utility of SSX2IP in patients with nasopharyngeal carcinoma. *APMIS.* 2020;128(4):287–97. <https://doi.org/10.1111/apm.13023>.

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

