Clinical and Experimental Medicine (2025) 25:253
https://doi.org/10.1007/s10238-025-01801-0

RESEARCH

Identification of a fibroblast-derived gene signature reveals prognostic and therapeutic insights in pancreatic cancer

Wenjia Wang[1] · Heli Li[1] · Jun Gong[2] · Yuhui Liu[2]

Received: 6 June 2025 / Accepted: 30 June 2025
© The Author(s) 2025

Abstract

Fibroblasts play a crucial role in the progression of pancreatic ductal adenocarcinoma (PDAC). This study aimed to develop a prognostic model based on fibroblast-specific gene signatures and identify key genes that may serve as therapeutic targets. Single-cell RNA sequencing data from PDAC and adjacent normal tissues were analyzed using high-dimensional weighted gene co-expression network analysis, and a prognostic signature was constructed using LASSO Cox regression and validated across multiple cohorts. Our findings revealed distinct fibroblast clusters enriched in PDAC and led to the development of a fibroblast-related prognostic signature comprising 14 key genes. The fibroblast-related risk score effectively stratified patients into high-risk and low-risk groups, revealing differences in immune cell infiltration, response to immune checkpoint blockade therapy, and sensitivity to chemotherapeutic drugs. This study presents a novel fibroblast-associated prognostic model with significant clinical utility in predicting PDAC patient outcomes.

Keywords Pancreatic cancer · Single-cell RNA sequencing · hdWGCNA · Immune infiltration · Fibroblast

Introduction

Pancreatic ductal adenocarcinoma (PDAC) is a deadly disease with limited therapeutic options. Despite advances in medical research, the underlying mechanisms of PDAC development and progression remain elusive [1]. The disease often presents with non-specific symptoms, leading to delayed diagnosis [2]. About 80 percent of patients missed the chance of surgical resection at diagnosis, and despite progression in systemic treatment, the 5-year survival rate remains as low as 8% [3, 4]. Given the complex etiology and unfavorable outcomes associated with pancreatic cancer, it is essential to identify novel biomarkers and therapeutic targets that favor early diagnosis and guide personalized treatment strategies to improve the prognosis of PDAC patients.

Tumor microenvironment (TME) is a multicellular dynamic system composed of stromal cells, immune cells, extracellular matrix, and blood vessels [5]. TME influences tumor growth, immune evasion, and therapeutic resistance, playing a crucial role in determining cancer outcomes [6]. Cancer-associated fibroblasts (CAFs), myeloid-derived suppressor cells (MDSCs), macrophages, and lymphocytes in the TME can exert both anti-tumor and pro-tumor effects, further complicating the disease [7]. Specifically, PDAC cells secrete various immunomodulatory factors that shape the immune response, contributing to immune escape and resistance to therapies [8, 9]. Understanding the modulatory networks that regulate the interactions between PDAC cells and their TME is key to developing effective prognostic biomarkers and tailoring therapeutic interventions.

Single-cell RNA sequencing (scRNA-seq) combined with weighted gene co-expression network analysis (WGCNA) has emerged as a powerful approach to unravel the complexity of the TME at a high resolution [10–12]. scRNA-seq allows researchers to profile gene expression at the single-cell level, offering unprecedented resolution into cellular heterogeneity within the TME. By dissecting distinct

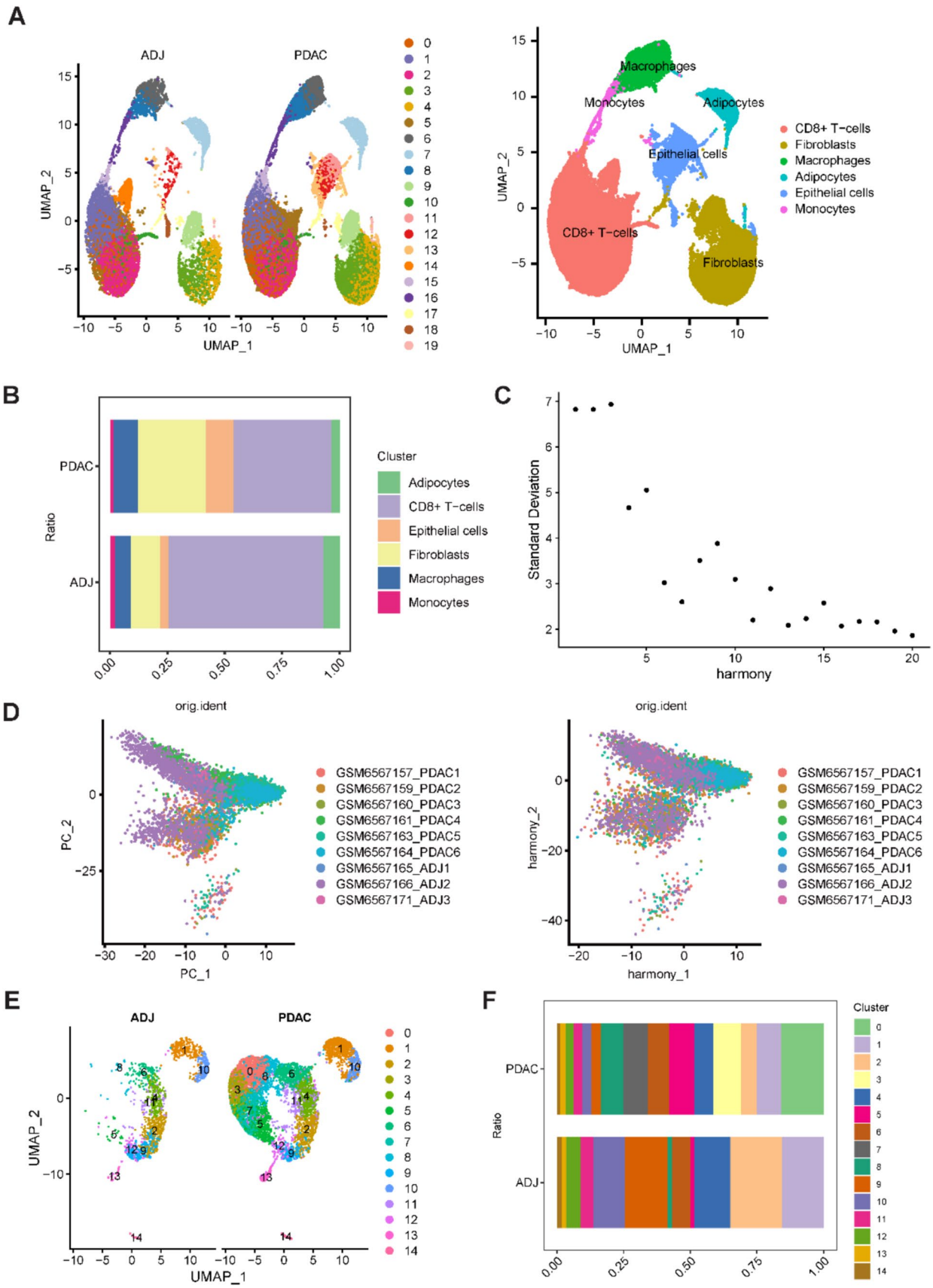
✉ Jun Gong
gjun@hust.edu.cn

✉ Yuhui Liu
2023tj0305@hust.edu.cn

[1] Division of Child Healthcare, Department of Paediatrics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, 1095 Jiefang Ave, Wuhan 430030, Hubei, China

[2] Department of Biliary-Pancreatic Surgery, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, 1095 Jiefang Ave, Wuhan 430030, Hubei, China

Published online: 18 July 2025


A
ADJ
PDAC
UMAP_1
UMAP_2
Macrophages
Monocytes
Adipocytes
Epithelial cells
CD8+ T−cells
Fibroblasts
B
PDAC
ADJ
Ratio
Cluster
Adipocytes
CD8+ T−cells
Epithelial cells
Fibroblasts
Macrophages
Monocytes
C
Standard Deviation
harmony
D
orig.ident
PC_1
PC_2
harmony_1
harmony_2
GSM6567157_PDAC1
GSM6567159_PDAC2
GSM6567160_PDAC3
GSM6567161_PDAC4
GSM6567163_PDAC5
GSM6567164_PDAC6
GSM6567165_ADJ1
GSM6567166_ADJ2
GSM6567171_ADJ3
E
ADJ
PDAC
UMAP_1
UMAP_2
F
PDAC
ADJ
Ratio
Cluster

◂**Fig. 1** Single-cell analysis reveals cell-type composition and fibroblasts heterogeneity in PDAC and adjacent normal pancreas. **A** Cells were categorized into 20 clusters by the UMAP algorithm based on the important components available in PCA. UMAP plot showing the cell-type annotation profiles for the 20 clusters. **B** Proportions of different cell types in PDAC and adjacent normal pancreas. **C** ElbowPlot showing standard deviation of different principle components. **D** Principal component analysis of fibroblasts, with dots representing cells and colors representing samples (left panel). RunHarmony function was used to correct batch effect (right panel). **E** UMAP plot showing 15 clusters of fibroblasts from PDAC and adjacent normal pancreas. **F** Proportions of different clusters of fibroblasts in PDAC and adjacent normal pancreas

cell populations, scRNA-seq uncovers the diverse cellular behaviors that contribute to cancer progression, immune modulation, and therapy resistance [13]. On the other hand, WGCNA is a robust systems biology method that constructs gene co-expression networks from high-dimensional data. It groups genes into modules based on their expression patterns and correlates these modules with clinical features or specific cell populations identified through scRNA-seq [14]. In the context of cancer research, WGCNA can identify gene networks specifically associated with stromal or immune cells within the TME, uncovering cell-type-specific pathways that influence tumor behavior [15, 16]. Several recent studies have successfully employed this combination to uncover immune-associated gene signatures and develop risk models for various cancers, emphasizing the clinical utility of this approach in identifying novel therapeutic targets and prognostic markers [15–17].

In this study, we developed a prognostic model based on fibroblast-specific gene signatures from scRNA-seq data using WGCNA analysis. Further validation was conducted using cohorts from TCGA and GEO databases. Relationship of risk score model with immunotherapy and chemotherapy was investigated, and a nomogram was constructed for predicting PDAC patient outcomes. Our findings provided a novel fibroblast-associated prognostic model with significant clinical utility in clinical decision support.

Materials and methods

Data acquisition

The study's data were sourced from the TCGA database (https://portal.gdc.cancer.gov/), a comprehensive repository of cancer genomic information, including gene expression, miRNA and lncRNA expression, copy number variation, DNA methylation, and SNP data. We acquired processed expression data for pancreatic cancer, comprising 4 normal and 179 tumor samples.

In addition, data from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/info/datasets.html) hosted by NCBI were used. Specifically, we retrieved the GSE212966 dataset for single-cell analysis, consisting of six pancreatic cancer and three adjacent normal tissue samples [18, 19]. We also downloaded Series Matrix Files for GSE57495 and GSE62452, which included expression profile data from 63 and 66 patients, respectively [20, 21], using the GPL15048 and GPL6244 annotation files.

Single-cell analysis

We conducted single-cell RNA sequencing analysis using the Seurat package [22]. Genes with low expression were filtered out (nFeature_RNA > 1000 & nFeature_RNA < 7500 & percent.mt < 15). The data were standardized and normalized, followed by principal component analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP). The optimal number of principal components was determined with an ElbowPlot. Clusters were visualized and annotated using the celldex package to identify key cell types involved in disease progression.

Ligand–receptor interaction analysis

Using the CellChat package, we analyzed intercellular communication networks based on single-cell data. Through network analysis and pattern recognition, CellChat identified the key signaling inputs and outputs of cells and the coordination of these signals [23]. The input data were standardized single-cell expression profiles, with cell subtypes derived from the single-cell analysis. We quantified cell interactions by analyzing their intensity (weights) and frequency, allowing us to observe the activity and influence of different cell types in pancreatic cancer.

High-dimensional weighted gene co-expression network analysis (hdWGCNA)

We employed the hdWGCNA package for network inference, gene module identification, functional enrichment analysis, reproducibility testing, and data visualization [14]. A gene co-expression network was constructed for genes expressed in at least 5% of cells using SetupForWGCNA, with a soft threshold of 14. Co-expression modules were visualized with dendrograms, where each leaf represented a gene, and colors denoted module assignment. The eigengenes (MEs) of the modules were visualized using GetMEs.

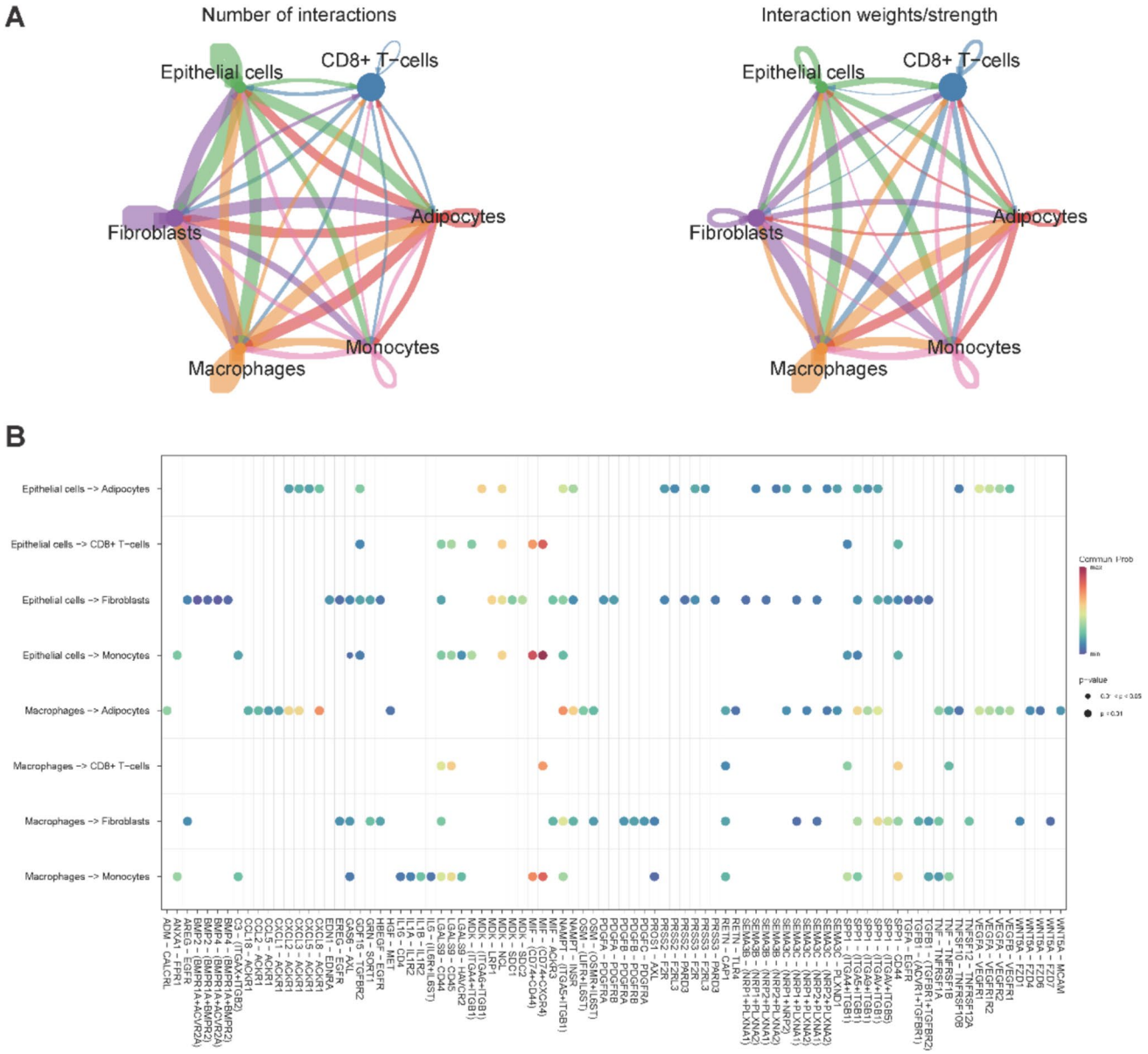


Fig. 2 Cell chat analysis indicates association between fibroblasts and epithelial cells. **A** Network plot showing the communication patterns between fibroblasts and other cell types in PDAC. **B** Dot plot depicting the probability and statistical significance of ligand–receptor pairs involved in communication between different cell types based on CellChat analysis

Prognostic model construction

To construct a prognostic model, we applied Lasso regression to select prognosis-related genes. A risk score formula was generated for each patient, weighted by regression coefficients. Based on the median risk score, patients were classified into low- and high-risk groups. Kaplan–Meier curves and log-rank tests were used to evaluate survival differences between the groups. The model's predictive performance was assessed with Receiver Operating Characteristic (ROC) curve analysis.

Immune cell infiltration analysis

We used the CIBERSORT algorithm, which applies support vector regression to estimate immune cell subtype proportions from RNA-seq data [24]. The relative abundance of 22 immune cell types within the tumor microenvironment was identified. Immune cell composition was compared between high- and low-risk groups, with statistical significance set at $P < 0.05$.

Drug sensitivity analysis

Chemotherapy sensitivity was predicted using the Genomics of Drug Sensitivity in Cancer (GDSC) database (https://www.cancerrxgene.org/) and the pRRophetic R package [25, 26]. IC50 values for various chemotherapeutic agents were estimated via a regression model, and tenfold cross-validation was used to assess model accuracy.

Nomogram model construction

A nomogram was developed based on a multivariate regression model incorporating risk scores and clinical symptoms. Scores were assigned to each predictor based on their regression coefficients, and the total score was used to predict patient outcomes.

Statistical analysis

Survival curves were generated using the Kaplan–Meier method and compared via log-rank tests. Multivariate analysis was performed using Cox proportional hazards models. All statistical analyses were conducted in R (version 4.3.0), with $P < 0.05$ considered statistically significant.

Results

Identification of fibroblasts in PDAC

We performed single-cell analysis on six PDAC tissues and three adjacent normal pancreatic tissues. Cells with nFeature_RNA between 1000 and 7500 and percentage. mt < 15% were retained for further analysis (Fig. S1A, B). The top 10 genes with the highest standardized variance were visualized (Fig. S1C). Batch effects were corrected using principal component analysis (PCA) and the "Harmony" function (Fig. S1D, E). The ElbowPlot determined that five principal components were optimal (Fig. S1F). Seurat's FindClusters function identified 20 cell clusters, which were annotated into six cell types: CD8 + T cells, fibroblasts, macrophages, adipocytes, epithelial cells, and monocytes (Fig. 1A). The proportion of fibroblasts was significantly higher in PDAC tissues compared to adjacent normal tissues (Fig. 1B). Further PCA and Harmony analyses identified 15 fibroblast clusters. Among these, clusters 0, 3, and 7 were newly identified as PDAC-related fibroblasts (Fig. 1C–F). CellChat analysis revealed enhanced communication between fibroblasts and epithelial cells in PDAC (Fig. 2A, B).

hdWGCNA analysis of fibroblasts in PDAC

We utilized high-dimensional weighted gene co-expression network analysis (hdWGCNA) to identify gene modules and explore their functional roles within PDAC-derived fibroblasts. The optimal soft threshold was determined to be 14, with a model fit value exceeding 0.9 (Fig. 3A). Hierarchical clustering of genes across 14 modules was visualized with a WGCNA dendrogram, and the top 10 hub genes from each module were displayed (Fig. 3B, C). We examined the distribution of these hub genes across the 14 fibroblast clusters in PDAC (Fig. 3D). Notably, the magenta, yellow, salmon, and blue modules were significantly enriched in the PDAC-related fibroblast clusters (clusters 0, 3, and 7) (Fig. 3E).

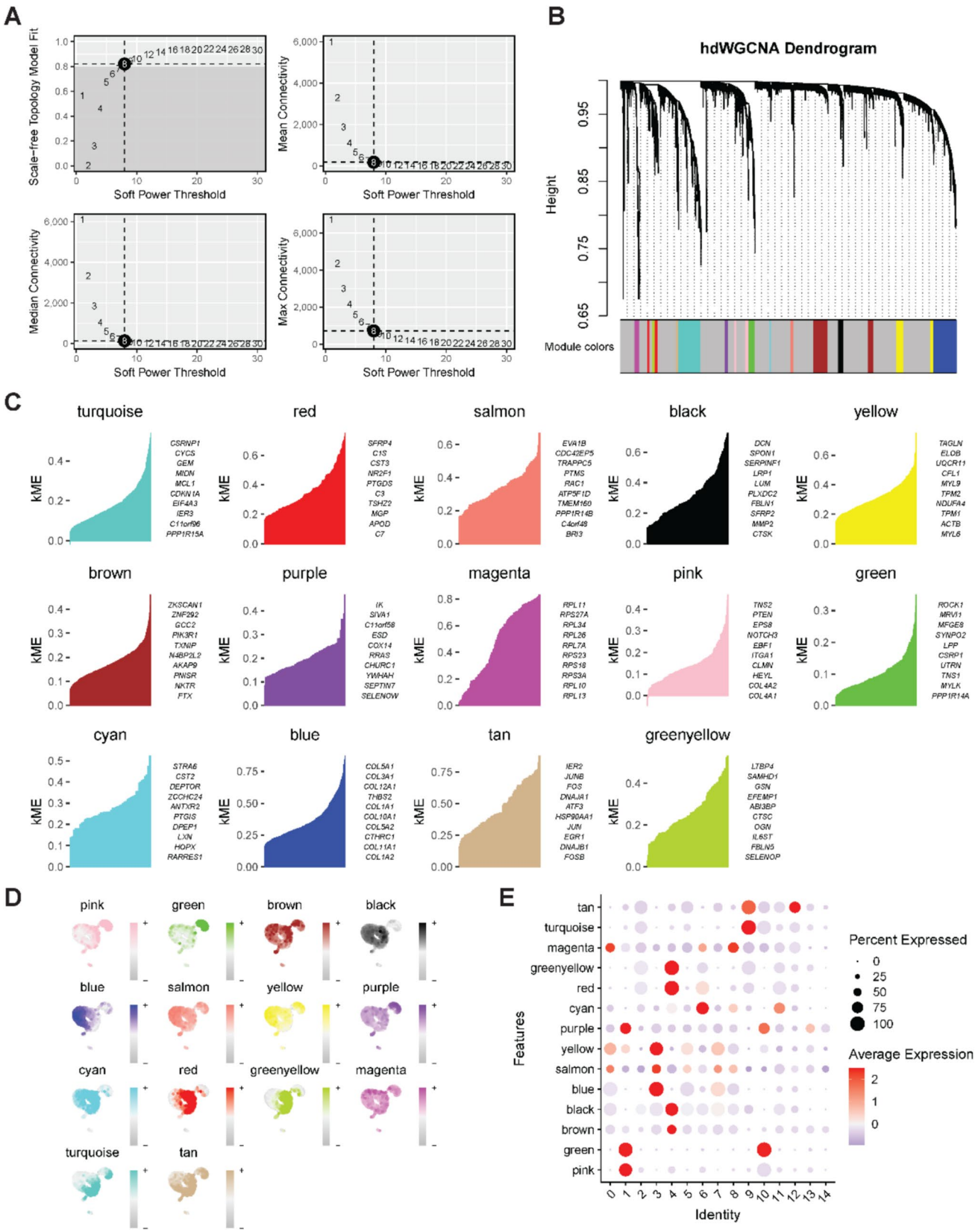
Construction of fibroblast-related prognostic signature

To identify key genes within the identified gene modules, we conducted univariate Cox analysis using clinical data, yielding 42 prognosis-related genes. We then randomly split the dataset into a training cohort and a validation cohort at a 4:1 ratio. A prognostic signature was constructed using LASSO Cox regression analysis, selecting 14 signature genes under optimal regularization parameters (Fig. 4A–C). Risk scores were calculated for each patient, stratifying them into high- and low-risk groups. Kaplan–Meier analysis showed that the low-risk group had significantly better overall survival compared to the high-risk group in both cohorts (Fig. 4D). ROC curve analysis demonstrated strong predictive performance, with AUC values of 0.89, 0.76, and 0.79 at 5, 3, and 2 years, respectively, in the training cohort, and 0.84, 0.84, and 0.74 in the validation cohort (Fig. 4E).

To validate the model, we analyzed two independent datasets (GSE57495 and GSE62452). Consistently, the low-risk group exhibited better survival outcomes compared to the high-risk group in both datasets (Fig. 5A, B). ROC curves confirmed the model's high accuracy in predicting prognosis (Fig. 5C, D). These findings suggest that our fibroblast-related prognostic signature is a robust tool for predicting PDAC prognosis across different cohorts.

Effect of prognostic signature on immunotherapy and chemotherapy

Since tumor-associated fibroblasts and immune cells are key components of the tumor microenvironment, we investigated immune cell infiltration in PDAC using the TCGA-PAAD dataset. Our analysis revealed that


A
Scale-free Topology Model Fit
Soft Power Threshold
Mean Connectivity
Median Connectivity
Max Connectivity
B
hdWGCNA Dendrogram
Height
Module colors
C
turquoise
red
salmon
black
yellow
brown
purple
magenta
pink
green
cyan
blue
tan
greenyellow
kME
D
pink
green
brown
black
blue
salmon
yellow
purple
cyan
red
greenyellow
magenta
turquoise
tan
E
Features
Identity
Percent Expressed
Average Expression

◂**Fig. 3** hdWGCNA reveals module-specific hub genes in PDAC-related fibroblasts. **A** Scale-free topology model fit index and mean connectivity for soft-thresholding powers. **B** hdWGCNA dendrogram visualizing the fourteen modules in a scale-free network. **C** kME (module eigengene-based connectivity) map showing the top 10 hub genes in each module. **D** Distribution of fourteen gene modules in different fibroblast subsets. **E** Dot plot depicting the average expression of module-specific hub genes in the 14 fibroblast clusters

M0 and M1 macrophages, resting NK cells, and resting CD4 memory T cells were more prevalent in the high-risk group, whereas naive B cells and CD8 T cells were more abundant in the low-risk group (Fig. 6A, B). Tumor immune dysfunction and exclusion (TIDE) analysis indicated that high-risk patients might benefit more from immune checkpoint inhibition therapy (Fig. 6C).

We also explored the relationship between our risk score and chemosensitivity using the GDSC database and the pRRophetic R package. High-risk patients were more sensitive to BI.2536, AKT inhibitor VIII, AZD.0530, A.443654, and A.770041, while low-risk patients were more sensitive to ABT.888 (Fig. 6D). These findings suggest that our risk score model could inform personalized treatment strategies, guiding both immunotherapy and chemotherapy approaches for PDAC patients.

Clinical characteristics associated with prognostic signature and validation of prognostic signature via nomogram construction

We then assessed the relationship between PDAC clinical characteristics and the risk score. Higher risk scores were associated with poorer survival and more advanced tumor grades in PDAC (Fig. 7A, B). Univariate and multivariate analyses confirmed that the risk score is an independent prognostic factor for PDAC (Fig. 7C, D). These findings further underscore the strong prognostic capacity of our risk score model.

Next, we developed a nomogram integrating patient age, tumor grade, gender, stage, and risk score. The risk score contributed significantly to the total points (Fig. 8A). Calibration curves for 1-year and 2-year overall survival closely aligned with the nomogram's predictions (Fig. 8B). The ROC curves yielded AUC values of 0.6451, 0.7429, and 0.8187 at 2, 3, and 5 years, respectively, demonstrating high prognostic accuracy (Fig. 8C). Decision curve analysis (DCA) further validated the superior performance of the risk score model (Fig. 8D). Overall, our nomogram, incorporating the risk score, is a reliable predictor of patient prognosis and can support clinical decision-making.

Discussion

In this study, we conducted a comprehensive analysis of pancreatic ductal adenocarcinoma (PDAC) using single-cell RNA sequencing and high-dimensional weighted gene co-expression network analysis (hdWGCNA) to identify distinct cell populations within the tumor microenvironment. Our results revealed a significant increase in fibroblast proportion in PDAC tissues compared to adjacent normal tissues, suggesting their potential involvement in tumor dynamics. Using hdWGCNA, we identified several gene modules enriched in these fibroblast populations, which facilitated the development of a robust prognostic signature comprising 14 key genes. This signature effectively stratified patients into high- and low-risk groups, with clear associations between risk scores and overall survival outcomes. These findings highlight the critical role of fibroblasts-related gene networks in shaping PDAC progression and prognosis.

While our study is based on a single scRNA-seq dataset (GSE212966), we emphasize that this dataset includes six PDAC samples and three matched adjacent normal tissues, making it one of the more comprehensively annotated pancreatic cancer scRNA resources currently available in GEO. We selected it due to its inclusion of paired non-tumor controls and its consistent sequencing protocol, which minimizes batch-related confounders. To address potential heterogeneity, we applied high-stringency filtering and robust batch correction (Harmony), and downstream analyses such as hdWGCNA and CellChat further ensured analytical rigor. Moreover, the fibroblast-based risk model developed from this dataset was validated across three independent bulk transcriptomic cohorts (TCGA, GSE57495, and GSE62452), confirming the generalizability and robustness of our findings.

TME is increasingly recognized as a critical player in cancer progression and patient outcomes across various cancers [9]. Studies have shown that the TME consists of a complex network of stromal cells, immune cells, and extracellular matrix components, all of which influence tumor growth, metastasis, and therapeutic responses [7]. Fibroblasts, particularly cancer-associated fibroblasts (CAFs), are a major component of the TME and have been implicated in shaping tumor behavior [27]. Studies have demonstrated that CAFs play dual roles in PDAC. On the one hand, fibronectin and collagen secreted

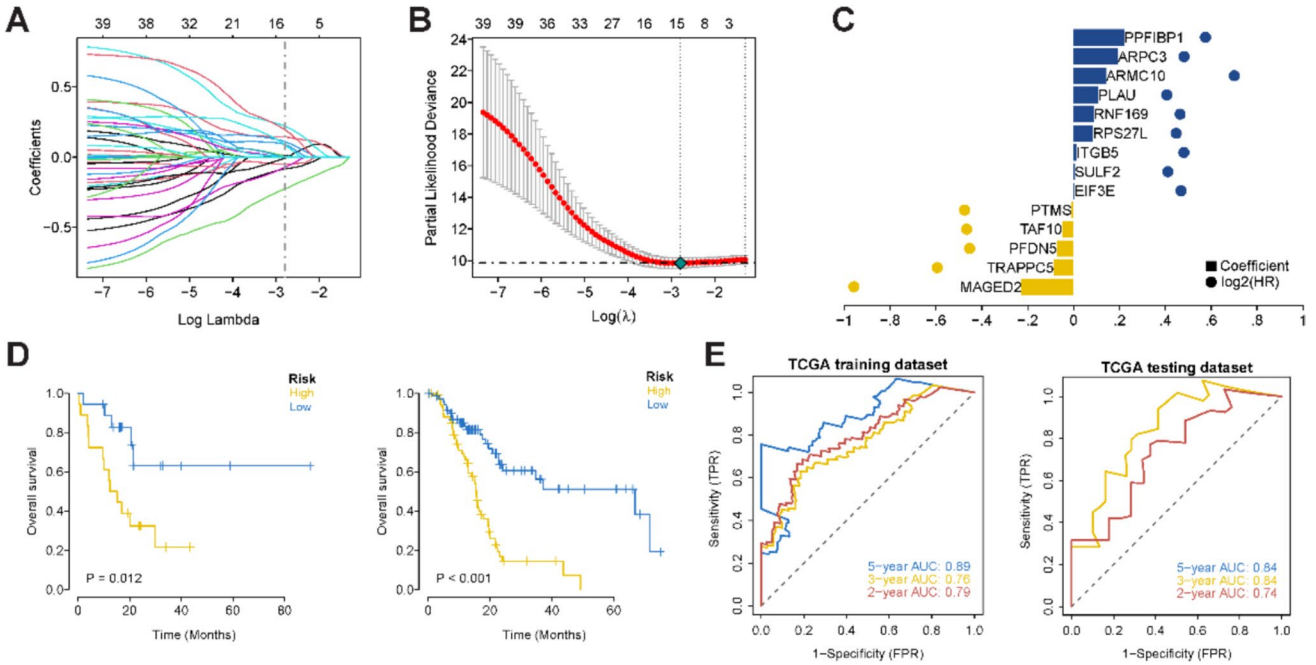


Fig. 4 Construction of the risk score model. **A** LASSO coefficient profiles of the prognostic genes in LASSO-COX regression. **B** Fourteen prognostic genes were selected using LASSO regression analysis with minimum criteria. **C** Detailed parameters of the 14 prognostic genes. **D** Survival plot of risk score model in training cohort (left) and validation cohort (right) using TCGA-PAAD datasets. **E** ROC curve of risk score model in training cohort (left) and validation cohort (right) using TCGA-PAAD datasets

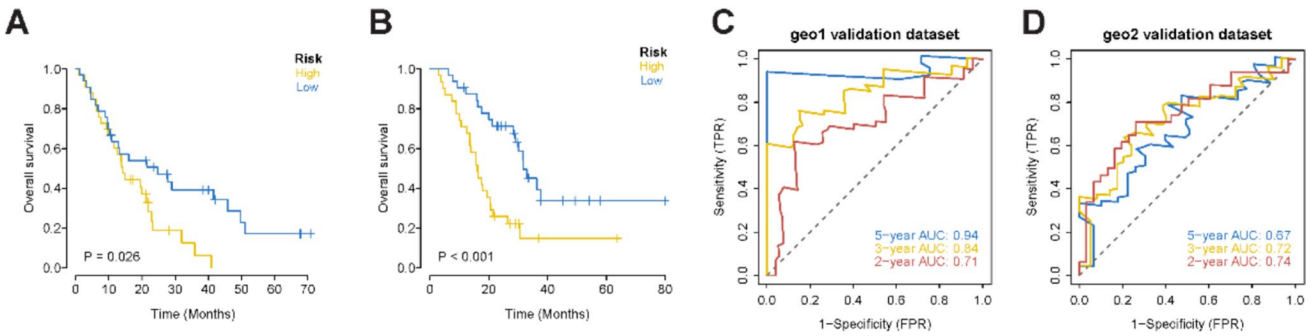


Fig. 5 Validation of the risk score model. **A**, **B** Survival plot of risk score model in GSE57495 (**A**) and GSE62452 (**B**) datasets. **C** and **D** ROC curve of risk score model in GSE57495 (**C**) and GSE62452 (**D**) datasets

by CAFs restricted PDAC tumor growth, proved by the failure of stroma-targeted therapies in clinical trails [28]. On the other hand, CAFs are reported to be related with PDAC development at early stage [29]. Besides, CAFs have a closely interaction with immune cells, suppressing the anti-tumor immune response by modulating the activity of T cells and macrophages, which is associated with more aggressive disease and poorer patient outcomes [30]. Therefore, unraveling the cross-talk between CAFs and tumor cells is crucial to provide novel therapeutic options. In recent years, the use of scRNA-seq has revolutionized the study of the TME by providing high-precision resolution into the cellular heterogeneity of tumors. scRNA-seq allows researchers to profile individual cell populations within the TME, identifying distinct subtypes of fibroblasts, immune cells, and other stromal components [31]. This technology has been pivotal in uncovering the complex cross-talk between cancer cells and their microenvironment [32, 33]. In cancers such as lung adenocarcinoma and hepatocellular carcinoma, scRNA-seq has been used to reveal the roles of NK cells, macrophages, and fibroblasts in driving tumor progression [34, 35]. Similarly, our study utilized scRNA-seq to dissect the heterogeneity of fibroblasts in PDAC,

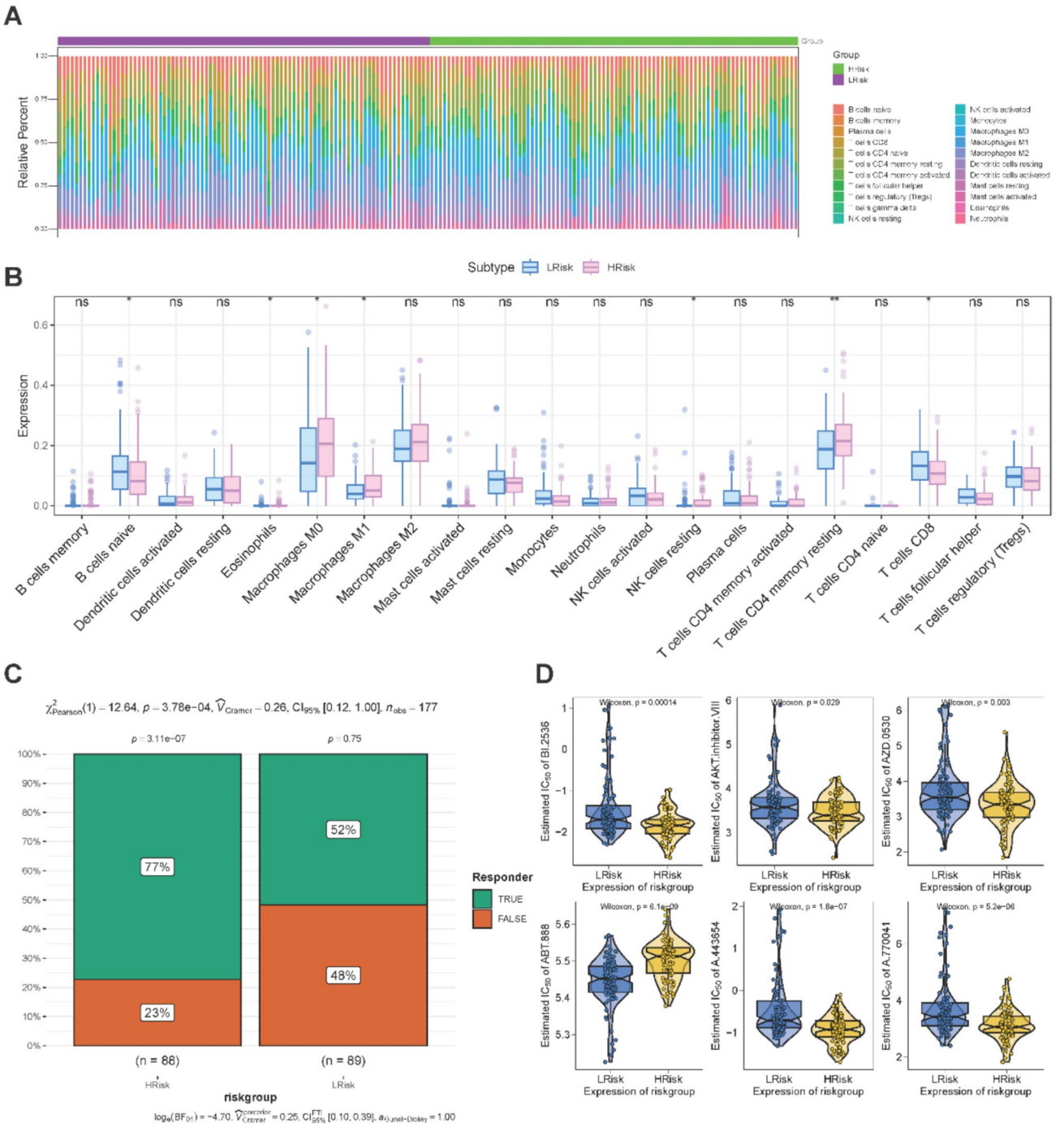


Fig. 6 Predictive value of risk score model on immunotherapy and chemotherapy. **A** Relative percentage of 22 types of immune cells in high- and low-risk group. **B** Comparison of immune cell infiltration between high- and low-risk groups. ns, not significant. $^{*}P < 0.05$ and $^{**}P < 0.01$. **C** Response to immune checkpoint inhibition therapy of high- and low-risk group predicted by tumor immune dysfunction and exclusion (TIDE) database. **D** Response to chemotherapy of high- and low-risk group predicted by genomics of drug sensitivity in cancer (GDSC) database

identifying specific fibroblast subpopulations that are enriched in tumor tissues compared to normal tissues. One previous study combined WGCNA and scRNA analysis to investigate the role of fibroblasts in PDAC [36]. This study identified 17 cell clusters and focused on clusters C11, representing fibroblasts, by analyzing a three-sample scRNA dataset GSE165399. Meanwhile, the authors performed WGCNA on PDAC and normal pancreas

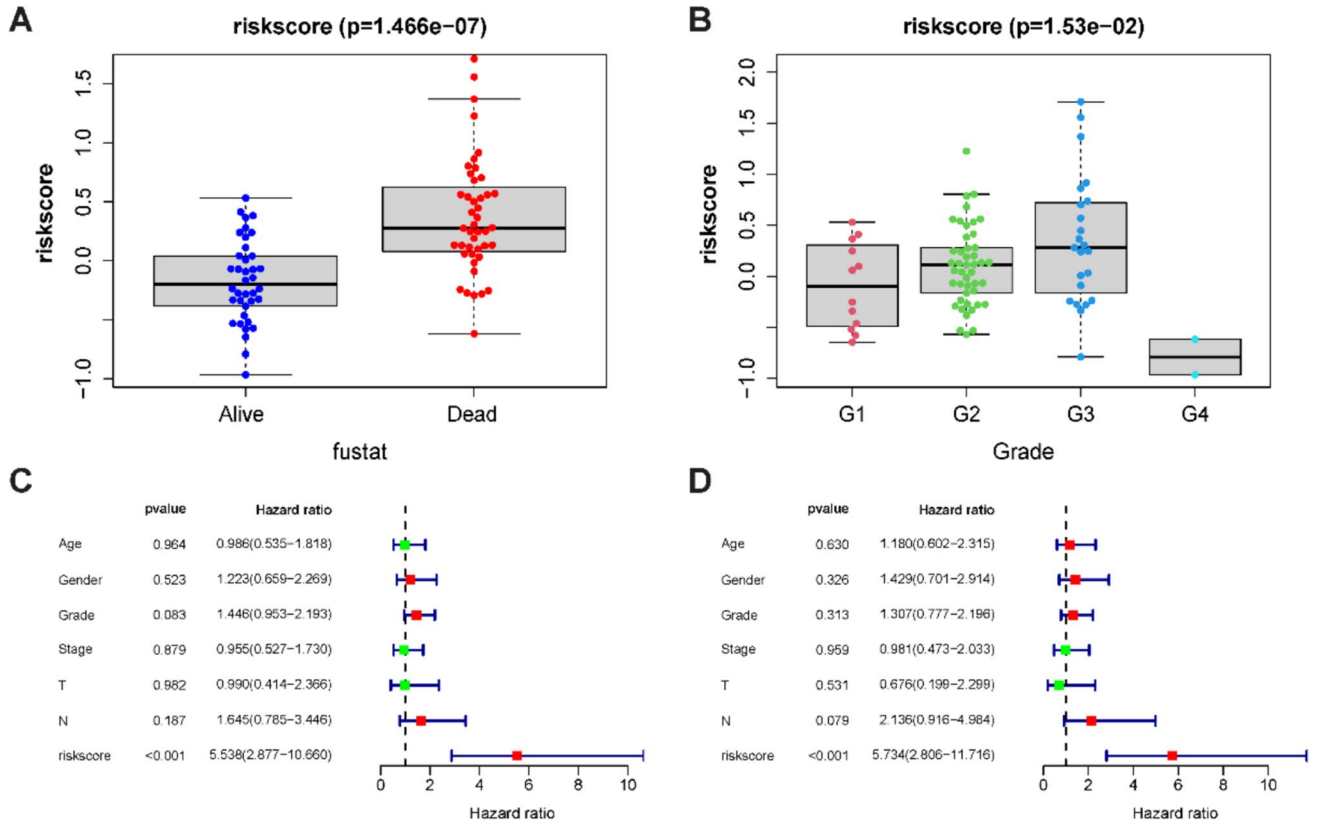


Fig. 7 Clinical application of risk score model in PDAC. **A** Correlation of risk score with survival status. **B** Correlation of risk score with tumor grade. **C** and **D** Forest plot showing result of univariate (**C**) or multivariate (**D**) COX regression analysis of TCGA clinical characteristics and risk score

samples in TCGA and GTEx cohort and focused on a set of brown module genes. By combining fibroblast-related genes, brown module genes and genes upregulated in PDAC compared to normal pancreas, the authors identified 107 genes to construct a prognosis model. In comparison, our study used six PDAC and three paracarcinoma scRNA-seq samples, advantaged in sample size. Besides, we identified clusters C0, C3, and C7 as PDAC-associated fibroblasts and performed hdWGCNA on these subtypes, reaching a high-accuracy analysis of the fibroblast populations. We further analyzed PDAC patients' survival information and constructed our risk model containing 14 key genes. Our in-depth investigation of PDAC-associated fibroblasts provided more detailed evidence and may shed light on the precise targeted therapy on CAFs in PDAC.

Although our prognostic model was constructed using fibroblast-specific genes identified from scRNA-seq data, its validation relied on bulk RNA-seq cohorts. This approach is widely accepted, as bulk transcriptomic data, despite reflecting mixed cell populations, provide essential clinical annotations for outcome modeling. Similar cross-platform strategies have been successfully applied in recent studies to validate cell-type-specific risk models [37, 38]. The reproducible performance of our model across three independent cohorts supports its generalizability. Nevertheless, future studies using spatial transcriptomics or sorted fibroblast data are warranted to further confirm fibroblast specificity.

Apart from evaluating patient prognosis, our fibroblast-related risk score model also showed value in predicting immunotherapy and chemotherapy responses in PDAC patients. Specifically, patients in the high-risk group exhibited lower infiltration of naive B cells and CD8 + T cells, typically linked to anti-tumor immunity. Interestingly, we observed a higher level of respond to immune checkpoint inhibition (ICB) therapies in high-risk group by TIDE analysis. This seemingly paradoxical result may be explained by the principle behind TIDE, which assesses T-cell dysfunction and T-cell exclusion. In the context of low infiltration of CD8 + T cells, TIDE will focus on signatures that related to T-cell exclusion, and a higher respond level suggest that abortion of T-cell exclusion mechanism may enhance the effect of ICB therapies. Therefore, our results indicated that ICB therapies could help overcome the immune evasion mechanisms in high-risk group. This observation underscores the complexity of the tumor

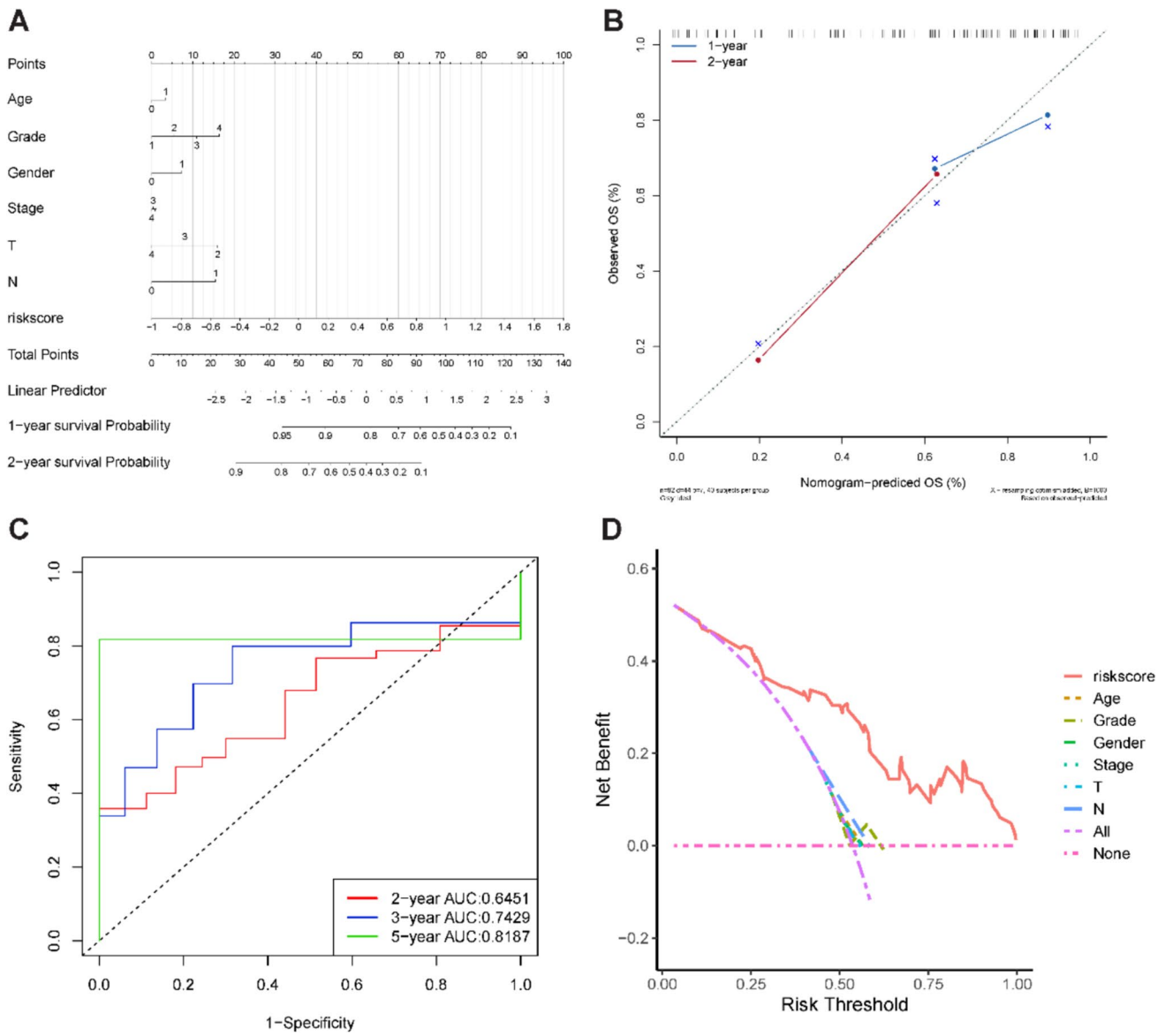


Fig. 8 Construction of the nomogram model. **A** Nomogram for predicting 1-year and 2-year overall survival of PDAC patients. **B** The calibration curve of the nomogram in (**A**). **C** ROC curve of risk score model for predicting 2-year, 3-year, and 5-year overall survival of PDAC patients. **D** Decision curve analysis of clinical usefulness of risk score and PDAC patients' clinical features

microenvironment and highlights the necessity of multi-dimensional analyses like TIDE, which can provide deeper insights into the potential responsiveness to immunotherapy. Additionally, our risk score model showed predictive value for chemotherapy sensitivity. High-risk patients were more sensitive to agents such as BI.2536 and AKT inhibitors, while low-risk patients exhibited greater sensitivity to ABT.888, a PARP inhibitor. These findings suggest that the risk score can be used as a valuable tool for guiding personalized treatment strategies. The integration of this risk model into clinical decision-making could significantly improve the management and outcomes of PDAC patients.

While our study constructed a risk model in PDAC, several limitations should be acknowledged. First, our study relied heavily on in silico analyses based on scRNA-seq and hdWGCNA, without the validation of fibroblast behavior in PDAC. Although we identified key fibroblast-related gene modules and constructed a predictive risk model, further in vivo and in vitro studies are needed to fully elucidate the functional roles of these fibroblast subpopulations in the tumor microenvironment. Secondly, while our prognostic model was validated in independent cohorts, it is important to expand the validation to prospective study or multicenter cohorts to ensure its robustness and applicability across diverse patient populations. Thirdly, while we focus on

14 signature genes within the identified gene modules to construct the risk score model, the detailed function and underlying mechanisms of the selected genes are not studied. Future studies should focus on mechanistic investigations in these key genes.

A limitation of our current study is the lack of functional annotation and spatial localization of specific CAF subtypes such as iCAFs and myCAFs. While our prognostic model was based on fibroblast-specific gene signatures identified from scRNA-seq data, future studies are warranted to validate the spatial distribution and biological roles of these genes within distinct fibroblast subpopulations.

Our study successfully identified key fibroblast-related gene modules in PDAC using scRNA-seq and hdWGCNA. By integrating these findings into a prognostic risk model, we were able to stratify patients into high- and low-risk groups, with the model showing strong predictive power for both survival outcomes and treatment responses. Overall, this work lays the groundwork for developing personalized treatment strategies in PDAC and highlights the importance of targeting specific gene networks within the tumor microenvironment to improve patient outcomes.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10238-025-01801-0.

Acknowledgements We acknowledge TCGA, GTEx, and GEO database for providing their platforms and contributors for uploading their meaningful datasets.

Author contributions YL and JG designed the study and reviewed the manuscript and supervised the study. WW, HL, JG and YL conducted the bioinformatics analysis, performed the experiments and curated the data. WW, JG and YL wrote the original draft. All authors read and approved the final manuscript.

Funding Key Research and Development Program of Hubei Province, 2024BCB053, Hubei Provincial Natural Science Foundation of China, 2024AFB063, National Natural Science Foundation of China, 82404022, Hubei Provincial Postdoctoral Innovation Talent Training Program, 2024HBBHCXB023.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Approval of the research protocol by an Institutional Reviewer Board: N/A. Informed Consent: N/A. Registry and the Registration No. of the study/trial: N/A. Animal Studies: N/A.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/.

References

1. Del Chiaro M, Sugawara T, Karam SD, Messersmith WA. Advances in the management of pancreatic cancer. BMJ. 2023;383: e073995. https://doi.org/10.1136/bmj-2022-073995.
2. Park W, Chawla A, O'Reilly EM. Pancreatic cancer: a review. JAMA. 2021;326:851–62. https://doi.org/10.1001/jama.2021.13027.
3. Halbrook CJ, Lyssiotis CA, di Magliano MP, Maitra A. Pancreatic cancer: advances and challenges. Cell. 2023;186(8):1729–54. https://doi.org/10.1016/j.cell.2023.02.014.
4. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics. CA Cancer J Clin. 2024;74:12–49. https://doi.org/10.3322/caac.21820.
5. Falcomata C, et al. Context-specific determinants of the immunosuppressive tumor microenvironment in pancreatic cancer. Cancer Discov. 2023;13:278–97. https://doi.org/10.1158/2159-8290.CD-22-0876.
6. Huynh D, Winter P, Markl F, Endres S, Kobold S. Beyond direct killing-novel cellular immunotherapeutic strategies to reshape the tumor microenvironment. Semin Immunopathol. 2023;45:215–27. https://doi.org/10.1007/s00281-022-00962-4.
7. Farhangnia P, Khorramdelazad H, Nickho H, Delbandi AA. Current and future immunotherapeutic approaches in pancreatic cancer treatment. J Hematol Oncol. 2024;17:40. https://doi.org/10.1186/s13045-024-01561-6.
8. Yousuf S, et al. Spatially resolved multi-omics single-cell analyses inform mechanisms of immune dysfunction in pancreatic cancer. Gastroenterology. 2023;165(891–908):e814. https://doi.org/10.1053/j.gastro.2023.05.036.
9. Ho WJ, Jaffee EM, Zheng L. The tumour microenvironment in pancreatic cancer—clinical challenges and opportunities. Nat Rev Clin Oncol. 2020;17:527–40. https://doi.org/10.1038/s41571-020-0363-5.
10. Zhong YJ, et al. Integrative analyses of bulk and single-cell transcriptomics reveals the infiltration and crosstalk of cancer-associated fibroblasts as a novel predictor for prognosis and microenvironment remodeling in intrahepatic cholangiocarcinoma. J Transl Med. 2024;22:422. https://doi.org/10.1186/s12967-024-05238-z.
11. Cai R, et al. A high immune-related index with the suppression of cGAS-STING pathway is a key determinant to Herceptin resistance in HER2+ breast cancer. Int J Biol Sci. 2024;20:3497–514. https://doi.org/10.7150/ijbs.94868.
12. Tan Z, et al. Comprehensive analysis of scrna-seq and bulk rna-seq reveals dynamic changes in the tumor immune microenvironment of bladder cancer and establishes a prognostic model. J Transl Med. 2023;21:223. https://doi.org/10.1186/s12967-023-04056-z.
13. Tirosh I, Suva ML. Cancer cell states: lessons from ten years of single-cell RNA-sequencing of human tumors. Cancer Cell. 2024;42:1497–506. https://doi.org/10.1016/j.ccell.2024.08.005.
14. Morabito S, Reese F, Rahimzadeh N, Miyoshi E, Swarup V. hdWGCNA identifies co-expression networks in high-dimensional

transcriptomics data. Cell Rep Methods. 2023;3: 100498. https://doi.org/10.1016/j.crmeth.2023.100498.
15. Chen W, et al. A new 4-gene-based prognostic model accurately predicts breast cancer prognosis and immunotherapy response by integrating WGCNA and bioinformatics analysis. Front Immunol. 2024;15:1331841. https://doi.org/10.3389/fimmu.2024.1331841.
16. Cheng J, et al. An immune-related gene signature for the prognosis of human bladder cancer based on WGCNA. Comput Biol Med. 2022;151: 106186. https://doi.org/10.1016/j.compbiomed.2022.106186.
17. He Z, et al. Comprehensive analyses of a tumor-infiltrating lymphocytes-related gene signature regarding the prognosis and immunologic features for immunotherapy in bladder cancer on the basis of WGCNA. Front Immunol. 2022;13: 973974. https://doi.org/10.3389/fimmu.2022.973974.
18. Chen J, Liu Z, Wu Z, Li W, Tan X. Identification of a chemoresistance-related prognostic gene signature by comprehensive analysis and experimental validation in pancreatic cancer. Front Oncol. 2023;13:1132424. https://doi.org/10.3389/fonc.2023.1132424.
19. Chen K, et al. Immune profiling and prognostic model of pancreatic cancer using quantitative pathology and single-cell RNA sequencing. J Transl Med. 2023;21:210. https://doi.org/10.1186/s12967-023-04051-4.
20. Chen DT, et al. Prognostic Fifteen-Gene Signature for Early Stage Pancreatic Ductal Adenocarcinoma. PLoS ONE. 2015;10: e0133562. https://doi.org/10.1371/journal.pone.0133562.
21. Wang J, et al. Endothelial nitric oxide synthase traffic inducer (NOSTRIN) is a negative regulator of disease aggressiveness in pancreatic cancer. Clin Cancer Res. 2016;22:5992–6001. https://doi.org/10.1158/1078-0432.CCR-16-0511.
22. Hao Y, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42:293–304. https://doi.org/10.1038/s41587-023-01767-y.
23. Jin S, et al. Inference and analysis of cell-cell communication using cell chat. Nat Commun. 2021;12:1088. https://doi.org/10.1038/s41467-021-21246-9.
24. Newman AM, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12:453–7. https://doi.org/10.1038/nmeth.3337.
25. Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. PLoS ONE. 2014;9: e107468. https://doi.org/10.1371/journal.pone.0107468.
26. Ma L, et al. Database commons: a catalog of worldwide biological databases. Genom Proteom Bioinf. 2023;21:1054–8. https://doi.org/10.1016/j.gpb.2022.12.004.
27. Cords L, de Souza N, Bodenmiller B. Classifying cancer-associated fibroblasts-the good, the bad, and the target. Cancer Cell. 2024;42:1480–5. https://doi.org/10.1016/j.ccell.2024.08.011.
28. Ramanathan RK, et al. Phase IB/II randomized study of FOLFIRINOX plus pegylated recombinant human hyaluronidase versus FOLFIRINOX alone in patients with metastatic pancreatic adenocarcinoma: SWOG S1313. J Clin Oncol. 2019;37:1062–9. https://doi.org/10.1200/JCO.18.01295.
29. Zhang F, et al. Cancer associated fibroblasts and metabolic reprogramming: unraveling the intricate crosstalk in tumor evolution. J Hematol Oncol. 2024;17:80. https://doi.org/10.1186/s13045-024-01600-2.
30. Rebelo R, Xavier CPR, Giovannetti E, Vasconcelos MH. Fibroblasts in pancreatic cancer: molecular and clinical perspectives. Trends Mol Med. 2023;29:439–53. https://doi.org/10.1016/j.molmed.2023.03.002.
31. Wang Y, et al. Single-cell analysis of pancreatic ductal adenocarcinoma identifies a novel fibroblast subtype associated with poor prognosis but better immunotherapy response. Cell Discov. 2021;7:36. https://doi.org/10.1038/s41421-021-00271-4.
32. Xu W, et al. Unveiling the role of regulatory T cells in the tumor microenvironment of pancreatic cancer through single-cell transcriptomics and in vitro experiments. Front Immunol. 2023;14:1242909. https://doi.org/10.3389/fimmu.2023.1242909.
33. Liu X, et al. Comprehensive analysis of bulk and single-cell transcriptomic data reveals a novel signature associated with endoplasmic reticulum stress, lipid metabolism, and liver metastasis in pancreatic cancer. J Transl Med. 2024;22:393. https://doi.org/10.1186/s12967-024-05158-y.
34. Lu J, et al. A novel prognostic model based on single-cell RNA sequencing data for hepatocellular carcinoma. Cancer Cell Int. 2022;22:38. https://doi.org/10.1186/s12935-022-02469-2.
35. Si Y, Zhao Z, Meng X, Zhao K. RNA-seq and bulk RNA-seq data analysis of cancer-related fibroblasts (CAF) in LUAD to construct a CAF-based risk signature. Sci Rep. 2024;14:23243. https://doi.org/10.1038/s41598-024-74336-1.
36. Xu Y, Chen X, Liu N, Chu Z, Wang Q. Identification of fibroblast-related genes based on single-cell and machine learning to predict the prognosis and endocrine metabolism of pancreatic cancer. Front Endocrinol (Lausanne). 2023;14:1201755. https://doi.org/10.3389/fendo.2023.1201755.
37. Zhang S, et al. Identification and validation of a prognostic signature based on fibroblast immune-related genes to predict the prognosis and therapeutic response of renal clear cell carcinoma by integrated analysis of single-cell and bulk RNA sequencing data. J Cancer. 2024;15:5942–55. https://doi.org/10.7150/jca.100194.
38. Zhang N, et al. Machine learning integration of single-cell and bulk transcriptomics identifies fibroblast-driven prognostic markers in colorectal cancer. Biomol Biomed. 2025. https://doi.org/10.17305/bb.2025.12038.

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