Discover Oncology

ANALYSIS Open Access

Integrating bulk and single-cell RNA sequencing reveals intratumor heterogeneity phenotypes and immune infiltration in pancreatic cancer

Li Zhao[1†], Yao Huang[3†], Jiang Li[2], Rui Xie[2*] and Kaiming Li[2*]

[†]Li Zhao and Yao Huang contributed equally to this work.

*Correspondence:
Rui Xie
20173188@zcmu.eud.cn
Kaiming Li
likaiming2016@126.com
[1]Pancreas Center, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China
[2]Department of Pharmacy, The First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine), Hangzhou, China
[3]Guangzhou University of Chinese Medicine, Guangzhou, China

Abstract

Background Although the significance of intratumoral heterogeneity (ITH) has been widely acknowledged in various cancers, its role in pancreatic cancer (PC) remains underexplored and warrants further investigation.

Methods Pancreatic cancer transcriptomic data were acquired from the TCGA and GEO databases. The DEPTH2 algorithm, in combination with differential expression analysis, was used to identify genes associated with intratumoral heterogeneity (ITH). We applied univariate Cox regression analysis and multiple machine learning techniques to establish a reliable prognostic model. Patients were then stratified according to their ITH scores, and differences between subgroups were examined through pathway enrichment analysis, immune cell infiltration profiling, and drug response prediction. Furthermore, we conducted subcellular localization and differential analysis of ITH using single-cell data, followed by cell-cell communication analysis to explore interaction relationships and identify key pathways.

Results Patients exhibiting lower intratumoral heterogeneity (ITH) levels demonstrated poorer clinical outcomes. The constructed 11-gene signature successfully differentiated individuals into high- and low-risk categories with significant survival differences. Immune profiling revealed notable differences in immune cell composition between the two groups, with patients in the high ITH cohort exhibiting enhanced immune activation. Drug sensitivity analysis indicated a differential response to therapies, with high-risk patients more resistant to certain drugs. Single-cell RNA sequencing identified a greater ITH score in epithelial cells, highlighting key interactions, particularly involving Galectin signaling pathways.

Conclusion Our results highlight intratumoral heterogeneity's prognostic and therapeutic relevance in pancreatic cancer, suggesting its potential utility in guiding individualized treatment approaches.

Keywords Intratumoral heterogeneity, Pancreatic cancer, Prognostic model, Immune infiltration, Drug sensitivity

© The Author(s) 2025. **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/.

1 Introduction

With a five-year survival rate below 10%, pancreatic cancer is considered one of the most aggressive malignancies on a global scale [1]. Its high mortality and therapeutic resistance pose significant challenges in oncology research. Epidemiological trends show a consistent upward trajectory in both the incidence and mortality of pancreatic cancer, with forecasts predicting it may rank as the second-leading cause of cancer-related deaths by 2030. The lack of early-specific symptoms and reliable screening methods leads to over 80% of patients being diagnosed at advanced or metastatic stages, preventing the possibility of curative surgery. Although progress has been made in surgical techniques, radiotherapy, chemotherapy, and targeted therapies, the overall prognosis for patients remains poor, emphasizing the urgent need to identify new biomarkers and treatment strategies [2].

Intratumoral heterogeneity (ITH), a hallmark of tumor evolution, has emerged as a frontier in cancer research [3]. ITH manifests as multidimensional diversity among tumor cells, encompassing genomic, epigenomic, transcriptomic, and proteomic variations that drive malignant progression, therapy resistance, and metastatic dissemination [3]. The ITH score, a quantitative metric, systematically evaluates clonal architecture complexity, reflecting tumor evolutionary trajectories and treatment susceptibility. To date, ITH scoring has demonstrated prognostic value in breast [4], colorectal cancer [5], and other solid tumors [6, 7], showing promise for guiding personalized therapies.

Despite PC's pronounced ITH features, systematic studies on ITH scoring in this malignancy remain limited. PC's unique characteristics—including its dense desmoplastic stroma, highly heterogeneous tumor microenvironment, and complex clonal dynamics—make it an ideal model for investigating ITH. Growing evidence indicates that intercellular heterogeneity influences primary tumor growth kinetics, chemoresistance, and distant metastasis. Thus, comprehensively characterizing ITH scores across molecular subtypes and their clinical implications may offer new avenues to overcome therapeutic barriers. Although intratumoral heterogeneity (ITH) plays a critical role in pancreatic cancer biology, related research remains limited, and its clinical translation potential has yet to be fully explored.

Rapid advancements in bioinformatics have opened new opportunities for research on various diseases [8, 9]. Integrative analysis of multi-omics data (e.g., whole-exome sequencing, scRNA-seq, and spatial transcriptomics) combined with machine learning algorithms enables multidimensional dissection of PC heterogeneity. Furthermore, leveraging clinically annotated datasets from TCGA, ICGC, and other large-scale repositories allows the construction of prognostic models linking ITH scores to patient outcomes while elucidating underlying molecular mechanisms [10]. As artificial intelligence and multicenter data are integrated, bioinformatics will become more essential in decoding the biological essence and clinical relevance of intratumoral heterogeneity in pancreatic cancer. This study systematically evaluated ITH levels in pancreatic cancer using established mRNA-based ITH assessment algorithms (DEPTH2) with TCGA and GEO datasets, demonstrating that elevated ITH significantly correlates with poor clinical outcomes. The machine learning-constructed ITH risk model showed robust prognostic predictive performance while providing novel insights into tumor immunosuppressive microenvironments and drug sensitivity evaluation. Further validation through single-cell RNA sequencing confirmed the biological significance of ITH-associated signature

genes, providing theoretical support for applying ITH scoring in precision therapy for pancreatic cancer. Our study offers new research directions for developing individualized treatment strategies for pancreatic cancer.

2 Materials and methods

2.1 Patient dataset

Data for 178 pancreatic cancer patients, including gene expression profiles, clinical information, and somatic mutations, were sourced from the Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/). The Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) provided two independent datasets: one single-cell RNA sequencing (scRNA-seq) dataset and one bulk transcriptome dataset. The scRNA-seq dataset GSE212966 comprises expression profiles from six samples, including three PDAC tissues and three matched adjacent noncancerous tissues. GSE62452, a bulk transcriptome dataset, contains gene expression data from 69 pancreatic tumor tissues and 69 adjacent non-tumor tissues.

2.2 Differential expression analysis

Differential expression analysis was performed using the 'limma' package in R, excluding genes with a false discovery rate (FDR) greater than 0.05.

For the remaining differentially expressed genes (DEGs), genes with a log2 fold change (log2FC) ≥ 1 were considered upregulated, and those with a log2FC ≤ -1 were considered downregulated.

2.3 scRNA-Seq data preprocessing

The pipeline for processing single-cell RNA sequencing (scRNA-seq) data in pancreatic adenocarcinoma (PAAD) is described below. The raw data were imported using the Read10X function, and a Seurat object was constructed with the Seurat R package for further analysis. Quality control thresholds were defined according to the distribution of detected features and counts per cell. Low-quality cells were filtered out, characterized by high mitochondrial and ribosomal gene expression proportions. The data was normalized with the NormalizeData function, followed by selecting the top 2,000 highly variable genes using the FindVariableFeatures function. Principal component analysis (PCA) was performed with the RunPCA function, with the top 20 principal components chosen for further examination. Cell clustering was achieved with the FindNeighbors and FindClusters functions, followed by dimensionality reduction and visualization via Uniform Manifold Approximation and Projection (UMAP) using the RunUMAP function. Cell type annotation was performed using the SingleR R package. Differential expression analysis across cell clusters was done using the Wilcoxon rank-sum test. Pseudotime trajectory analysis was conducted using the Monocle R package, and cell–cell communication analysis and network visualization were executed with the CellChat R package.

2.4 Construction of prognostic models

This study aimed to construct and evaluate various prognostic models based on a diverse set of 101 machine learning algorithms. Expression data from the training and validation cohorts were normalized before analysis. Feature variables were then extracted through

101 combinations of classical and ensemble machine learning algorithms, including Cox regression, random forest (RF), LASSO, gradient boosting machine (GBM), support vector machine (SVM), and Bayesian additive regression trees (BART), among other methods. Prognostic models were trained on the TCGA cohort, and their performance was evaluated using the concordance index (C-index). Model performance was further validated in an external GEO validation cohort. The best-performing prognostic model was selected, and its signature genes were retained for downstream biological functional analysis. The R programming environment was used for all data processing and model construction, with custom scripts handling core functions. LASSO regression analysis was employed to identify the most effective prognostic model. Multivariate Cox regression analysis was then carried out, identifying 11 key genes and their associated coefficients. Risk scores for each patient were calculated using the formula: Risk score = (expression of $RNA_1 \times Coef_1$) + (expression of $RNA_2 \times Coef_2$) +… + (expression of $RNA_n \times Coef_n$). We stratified the training cohort into high- and low-risk subgroups using the median risk score as the cutoff. The model's prognostic performance was rigorously evaluated through Kaplan-Meier survival analysis (log-rank test) and time-dependent ROC analysis (AUC at 1/3/5 years). To further validate the model's predictive capability, its prognostic accuracy, sensitivity, and specificity were evaluated in the validation set.

2.5 Nomogram construction

Univariate and multivariate Cox regression analyses were conducted to evaluate the ITH signature's independent prognostic value for OS in PC. The multivariate model incorporated clinical parameters (age, sex, grade, stage, T stage) alongside the ITH risk score. Using the rms and regplot packages in R, we developed a comprehensive prognostic nomogram incorporating all independent predictors to estimate 1-, 3-, and 5-year overall survival probabilities. The prognostic nomogram was calibrated via 1,000 bootstrap resamples to generate calibration plots. The model's discriminatory power was assessed by calculating the C-index and 95% confidence interval. Time-dependent ROC analysis was conducted to validate its performance further, and AUCs were determined for 1-, 3-, and 5-year overall survival.

2.6 Immune microenvironment analysis

Immune infiltration data were derived from previously estimated immune cell compositions in tumor tissues using established deconvolution algorithms, including CIBERSORT, EPIC, and xCell (immune estimation file: infiltration_estimation_for_tcga.csv). Only samples in the risk score data and immune infiltration profiles were retained for downstream analysis to ensure consistency in sample matching. Spearman correlation coefficients were calculated across various immune cell types to investigate the relationship between intratumoral heterogeneity (ITH) risk scores and immune cell infiltration. The correlation coefficient and corresponding p-value were calculated for each immune cell type. For immune cell types showing significant correlations with the risk score, scatter plots were generated to visualize the relationship, accompanied by linear regression fitting and annotated correlation values. To provide a comprehensive overview, a bubble plot displayed all significantly correlated immune cell types. The y-axis denoted immune cell categories, whereas the x-axis reflected their corresponding correlation values. Color coding was used to differentiate the source algorithms (e.g., CIBERSORT,

EPIC, xCell), facilitating a clear visual representation of the correlation magnitude and direction between each immune cell type and the ITH risk score. ImmuneScore, StromalScore, and ESTIMATEScore were calculated using the ESTIMATE algorithm and compared across high- and low-risk cohorts to characterize tumor microenvironmental differences.

2.7 Analysis of immune functional status

Using ssGSEA (GSVA package), we identified significant [positive/negative] correlations between the risk score and immune functional states, revealing potential immune evasion mechanisms. Immune functional states were assessed by calculating ssGSEA enrichment scores. The resulting ssGSEA scores were subjected to min-max normalization, and non-tumor samples were excluded from subsequent analyses. Differences in immune function scores between the groups were assessed using the Wilcoxon rank-sum test, and boxplots were generated to depict the distribution of scores.

2.8 Differential analysis of immune checkpoint genes

A list of immune checkpoint genes was obtained based on previously published literature [11]. Expression levels of these genes were retrieved from the transcriptomic matrix and integrated with risk group data. The Wilcoxon rank-sum test was applied to identify immune checkpoint genes that showed significant expression differences between the high-risk and low-risk groups. Boxplots were used to visualize these significant alterations in expression.

2.9 Assessment of immunotherapy response

The potential response to immune checkpoint blockade therapy was assessed using TIDE (Tumor Immune Dysfunction and Exclusion) scores, obtained from the TIDE database [12]. We integrated these scores with the risk group classifications to assess the variation in immune evasion potential between the high-risk and low-risk groups. Violin plots were generated to visualize the distribution of TIDE scores across groups, and statistical significance was assessed using appropriate non-parametric tests.

2.10 Functional enrichment analysis

The "clusterProfiler" R package was employed for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses [12], and the findings were visualized using the "circlize" R package. To investigate pathway enrichment differences between the high-risk and low-risk groups, GSEA (Gene Set Enrichment Analysis) was performed with the "c2.cp.kegg.Hs.symbols.gmt" file from MSigDB [13].

2.11 Tumor mutational burden analysis

Somatic mutation profiles for PAAD samples were obtained from TCGA in Mutation Annotation Format (MAF). The TMB(Tumor mutational burden) scores for PAAD patients were computed using the "maftools" R package [14]. We evaluated the association between tumor mutational burden (TMB) and risk scores, followed by Kaplan-Meier survival analysis (log-rank test) to assess TMB's prognostic value in PAAD.

2.12 Chemotherapy sensitivity analysis

We utilized the Genomics of Drug Sensitivity in Cancer (GDSC) database (https://www.cancerrxgene.org/) to estimate the half-maximal inhibitory concentration (IC50) values for frequently used chemotherapeutic agents in PAAD, employing the R packages pRRophetic [15] and oncoPredict [16]. These IC50 values were applied to predict chemotherapeutic sensitivity in patients across various risk groups. The Wilcoxon rank-sum test assessed potential differences in drug sensitivity between the groups classified based on gene expression.

2.13 Statistical analysis

All statistical analyses were executed in R software (version 4.2.2), where p-values and false discovery rate (FDR) q-values below 0.05 were deemed statistically significant.

3 Result

3.1 Overview of the intratumor heterogeneity score in PC

To quantify ITH in pancreatic cancer samples from the TCGA database, we first utilized the DEPTH2 software. Using the surv_cutpoint() function, we established the optimal cutoff for ITH scores and subsequently divided the samples into high-ITH and low-ITH groups with the surv_categorize() function. After performing a distribution analysis, survival curves were generated based on the chosen threshold. As depicted in Fig. 1A, the low-ITH group demonstrated a significantly worse prognosis.

No statistically significant differences in age, gender, or T/N/M stages were observed when comparing the clinical characteristics of the high- and low-ITH groups (Fig. 1B, C, E, F, H). Patients with higher tumor grades and advanced clinical stages exhibited

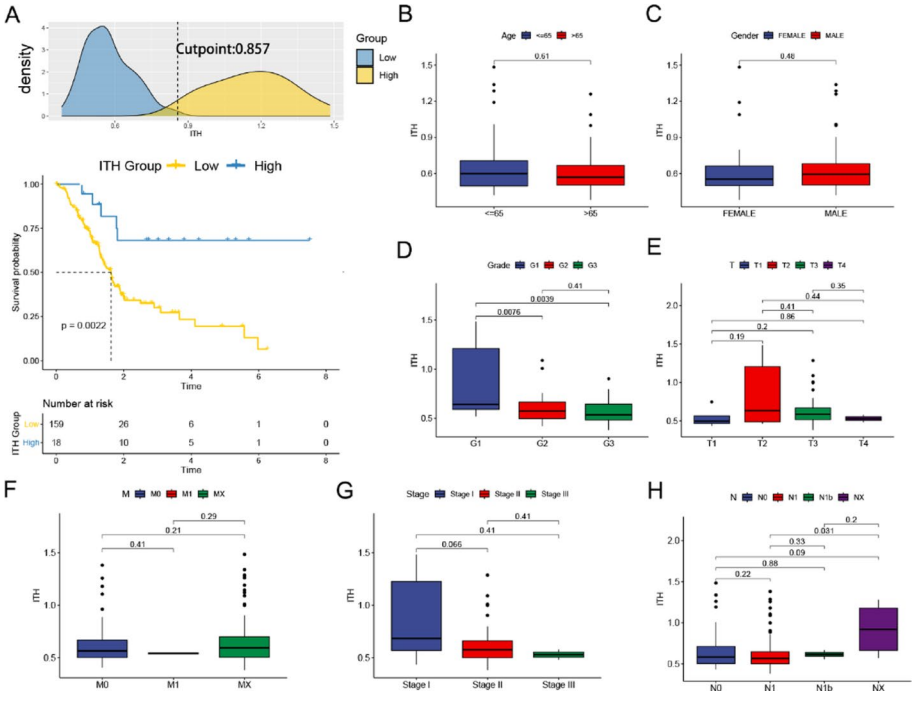


Fig. 1 Correlations between ITH scores and clinicopathological features in PAAD patients. A. Distribution of ITH scores and selection of the optimal cut-off point. Survival analysis for patients with high or low expression of ITH scores. Differences in Age **(B)**, Gender**(C)**, Grade**(D)**, T**(E)**, M **(F)**, Stage**(G)**, and N**(H)** between high-ITH and low-ITH groups

significantly lower ITH scores (Fig. 1D, G), which aligns with the poorer overall survival observed in the low-ITH group. These results suggest the need for further investigation into the molecular mechanisms and differentially expressed genes related to ITH.

We performed differential gene expression analysis comparing high-ITH and low-ITH samples to investigate further the factors contributing to the poor prognosis observed in the low-ITH group. One thousand eight hundred forty-eight differentially expressed genes (DEGs) were identified, of which 345 were upregulated and 1,503 were downregulated in the low-ITH group. A heatmap depicting the top 50 DEGs is shown in Fig. 2A. A total of 302 genes were identified through univariate Cox regression analysis as significantly associated with overall survival. A forest plot (Fig. 2B) illustrates the top 20 genes. The correlation analysis among these prognostic genes revealed mostly positive associations (Fig. 2C). Twelve genes were classified as risk factors, whereas eight were associated with a favorable prognosis.

3.2 Biological functions analyses

We conducted enrichment analyses based on KEGG pathways and GO functional annotation to better understand the biological functions of differentially expressed genes (DEGs) in pancreatic cancer progression. The GO analysis categorized the DEGs into three areas: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC). Biological process (BP) analysis revealed significant enrichment of differentially expressed genes (DEGs) in pathways critical to pancreatic cancer progression, including cell migration, extracellular matrix (ECM) remodeling, angiogenesis, and inflammatory regulation. These findings implicate DEGs mediating tumor invasiveness, metastatic dissemination, and immune microenvironment modulation. In the MF category, the most

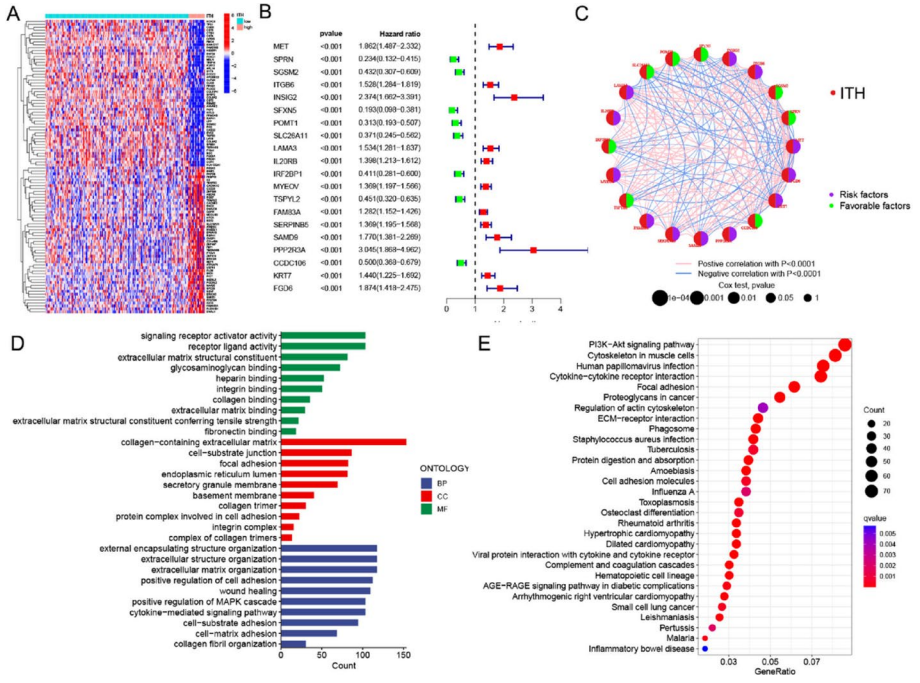


Fig. 2 Differential gene expression and functional analysis; **(A)** The top 50 differentially expressed genes are displayed in a heatmap. **(B)** Forest plot of the Univariate Cox regression analysis; **(C)** Displays the correlation analysis between ITH-related differentially expressed genes; **(D)** GO term enrichment showing biological processes, molecular functions, and cellular components; **(E)** KEGG pathway mapping

significantly enriched functions included structural molecule activity, protein binding, and integrin binding, reflecting their roles in maintaining cellular structural integrity and mediating signal transduction. Regarding CC, DEGs were predominantly localized to the extracellular region, extracellular matrix, and cytoplasm, indicating their potential involvement in the tumor microenvironment and signal transduction processes(Fig. 2D).

KEGG enrichment analysis showed that DEGs were predominantly mapped to pathways critical for pancreatic cancer biology. Highly enriched terms encompassed structural dynamics (muscle cytoskeleton, ECM-receptor binding) and pro-tumorigenic signaling (PI3K-Akt activation, cytokine networks, and proteoglycan-driven malignancy), reinforcing their roles in stromal remodeling and oncogenic progression (Fig. 2E).

The convergent evidence from KEGG and GO analyses provides a systems-level understanding of pancreatic cancer biology, uncovering canonical and novel disease-relevant pathways. These mechanistic insights create actionable opportunities for targeted drug discovery and personalized therapeutic strategies.

3.3 Construction of an integrated machine Learning-Based prognostic signature

Employing a multi-algorithmic machine learning framework, we derived the ITH-S prognostic signature from 302 candidate genes initially selected via univariate Cox analysis. This framework incorporated 10 different machine learning algorithms and 101 algorithmic combinations. Patients with PC from the TCGA dataset were randomly assigned to a training set, and an independent GEO dataset was employed for external validation.

We applied 10-fold cross-validation to develop 101 predictive models and assessed their performance by calculating the concordance index (C-index) across training and validation sets (Fig. 3A). Using 10-fold cross-validation, the optimal λ value in the LASSO regression was selected based on the minimum partial likelihood deviance (Fig. 3B, C). A subsequent stepwise multivariate Cox regression analysis of genes with non-zero coefficients identified 11 genes with significant independent prognostic relevance. These 11 identified genes with the most robust prognostic value were used to develop the final risk model. The individual risk score was calculated as the sum of the products of each gene's expression level and its corresponding Cox regression coefficient (Fig. 3D, E).

Patients were classified into high- and low-risk groups using the median risk score as a cutoff. Figure 3F demonstrates a clear trend of increasing mortality with higher risk scores, underscoring the prognostic significance of the risk model. The predictive capability of the risk model was validated using Kaplan–Meier analysis and time-dependent ROC curves. In the training cohort, the model yielded AUCs of 0.797, 0.734, and 0.786 for 1-, 3-, and 5-year OS predictions, respectively (Fig. 3G). Significant survival differences were observed between high- and low-risk groups in both the training and external validation sets, with high-risk patients experiencing markedly worse outcomes (Fig. 3H, I).

3.4 Construction and evaluation of the prognostic model

Univariate and multivariate Cox regression analyses confirmed the risk score as an independent prognostic factor for PC (Fig. 4A, B). To assess the model's discriminative ability

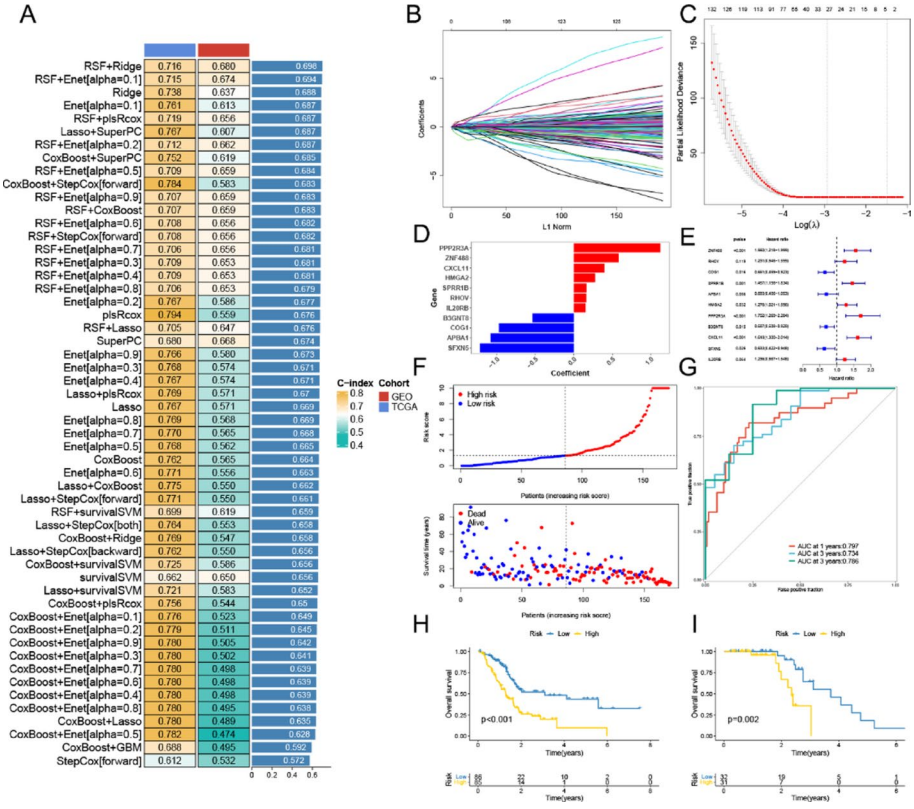


Fig. 3 Construction and evaluation of prognostic model. **(A)** The concordance index(C-index) heatmap of machine learning algorithms; **(B)** Lasso coefficient profiles of ITH-related prognostic genes; **(C)** Identification of the lasso optimal parameter (lambda); **(D)** Bar chart displaying Univariate Cox coefficient profiles; **(E)** Univariate Cox regression results visualized by forest plot; **F.**The riskScore distribution and OS status distribution; **G.** Survival prediction performance of ITH-risk model assessed by time-dependent ROC analysis at three clinical endpoints (1/3/5-year OS); **H.** Survival Analysis of ITH-risk High and Low Grouping Patients with PAAD in TCGA; I. Survival analysis of high and low ITH-risk subgroups in PAAD patients in GEO

in survival prediction, time-dependent concordance index (C-index) curves were plotted. Analysis showed that, throughout the timeline, the risk model maintained a consistently higher concordance index compared to traditional clinical parameters (Fig. 4C).

Five key variables, including age, sex, tumor grade, stage, and the risk model, were selected through machine learning and logistic regression analyses to construct a comprehensive clinical prognostic model for pancreatic adenocarcinoma (PAAD). A nomogram was then generated to visualize this integrated predictive model (Fig. 4F). Time-dependent receiver operating characteristic (ROC) curve analysis was performed to evaluate the model's predictive accuracy at 1-, 3-, and 5-year survival intervals(Fig. 4D, G). The study demonstrated that the risk model's area under the curve (AUC) was consistently higher than other clinical variables, indicating its robust predictive value for patient prognosis (Fig. 4E). The robustness and predictive precision of the nomogram were further confirmed by time-dependent calibration curves, which demonstrated close agreement between predicted and observed survival probabilities (Fig. 4H).

3.5 Immune landscape analysis

Emerging evidence has highlighted the critical role of immune cells within the tumor microenvironment in tumor initiation, progression, and patient outcomes. We used a

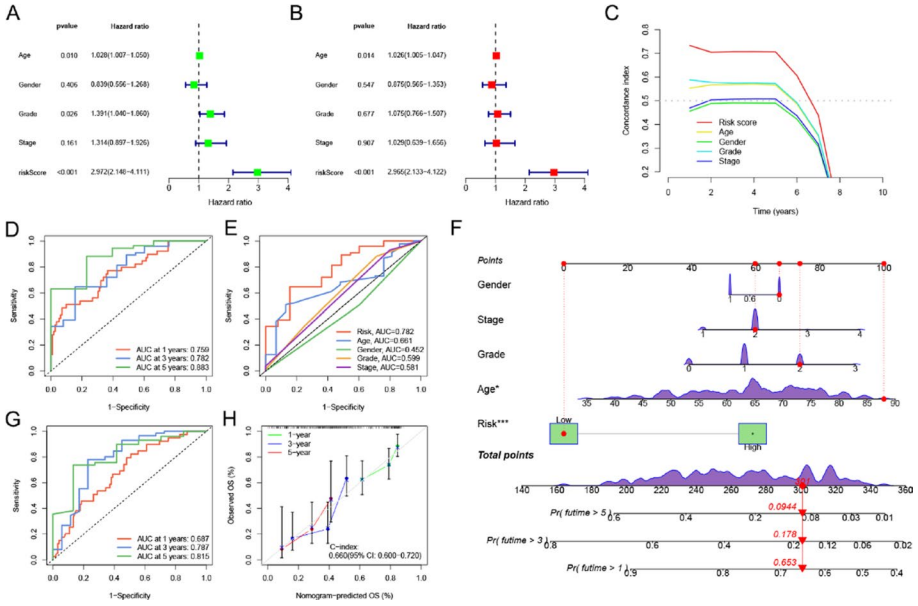


Fig. 4 Construction of a nomogram **(A)** Forest plot of the univariate Cox analysis; **(B)** Forest plot of the multivariate Cox analysis; **(C)** concordance index showed the prognostic predictive ability of ITH-risk model and clinical characteristics; **D**, **E.** ROC curve of risk score at 1, 3, and 5 years; **F** The nomogram combined with the variables; **G.** ROC curve analysis evaluates the prognostic value of risk scoring models and clinical features; **H.** Calibration curve of the nomogram at 1, 3, and 5 years

variety of immune deconvolution algorithms, such as CIBERSORT, EPIC, and TIMER, to examine the immune infiltration patterns between risk groups, providing a detailed comparison of immune cell distribution across different risk categories. Immune landscape analysis revealed substantial differences in cell infiltration patterns between risk groups. The high-risk group exhibited a pro-inflammatory yet immunosuppressive microenvironment, characterized by increased neutrophils and M1 macrophages. Still, reduced NK and memory B cells suggest an immune contexture potentially unfavorable for antitumor immunity (Fig. 5A).

The study identified significant immune functional differences between risk groups, particularly highlighting aberrant activity in APC co-inhibition, parainflammation, and type I interferon response pathways among high-risk individuals(Fig. 5B).

We also analyzed immune checkpoint molecule expression and discovered 12 immune checkpoint genes with significant differential expression. Immunocheckpoint analysis revealed distinct expression patterns between risk groups, with seven inhibitory checkpoints (including TNFRSF4, IDO-2, and CD160) showing significantly higher expression in the low-risk group. In comparison, five stimulatory checkpoints (such as CD276, CD274, and VTCN1) were overexpressed in the high-risk group. These findings imply differential immune escape mechanisms and potential responsiveness to immunotherapy in each group(Fig. 5C).

The TIDE algorithm revealed distinct immune evasion profiles, with high-risk patients showing significantly lower composite scores (reflecting dysfunction and exclusion mechanisms) than low-risk counterparts (Fig. 5D). This pattern suggests the high-risk subgroup may exhibit more favorable tumor-immune microenvironments for checkpoint blockade therapy.

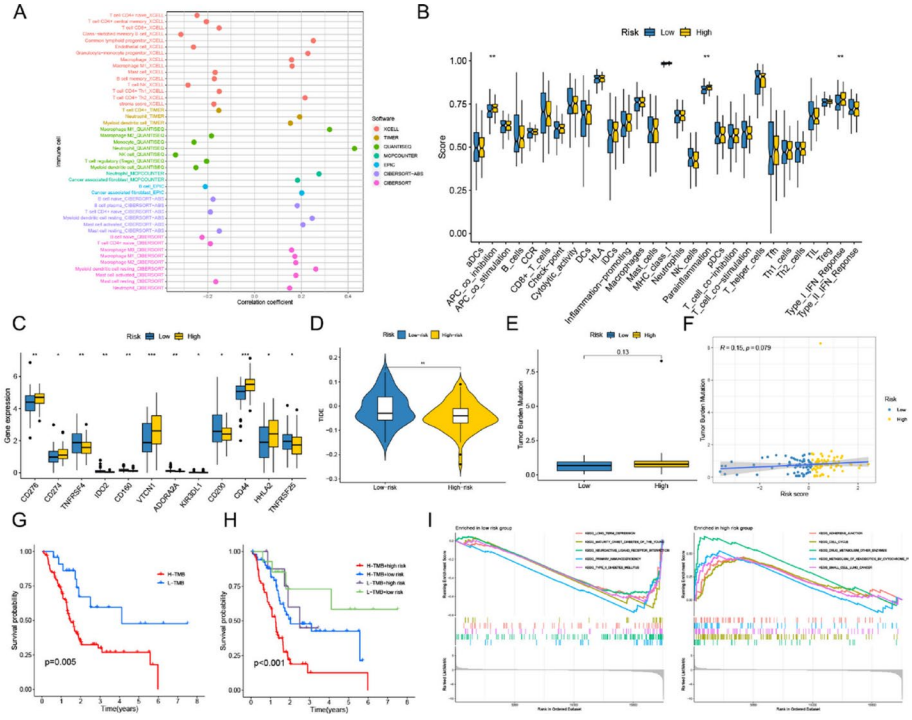


Fig. 5 Immune microenvironments Analysis in different groups **(A)** Various algorithms estimates tumor-infltrating immune cells; **(B)** Diferences in immune-related functions between two risk groups; **(C)** Diferences in immune checkpoints between two risk groups; **(D)** Analysis of the possibility of tumor immune escape by TIDE; **(E)** Comparison of tumor gene mutation burden between high and low risk groups; **(F)** Correlation analysis of tumor gene mutation burden in high and low risk groups; **(G)** Stratified survival analysis based on TMB threshold; **(H)** Differential survival outcomes by TMB status within each ITH stratum; **(I)** GSEA analysis for high and low risk groups

While tumor mutation burden (TMB) did not differ significantly between risk groups ($p = 0.13$, Wilcoxon test; Fig. 5E-F) nor correlate with risk scores ($r = 0.15$, $p = 0.079$), survival analysis revealed a synergistic prognostic impact: patients with concurrent high-risk scores and high TMB showed markedly reduced overall survival compared to those with low-risk/low-TMB profiles (Fig. 5G-H), suggesting combined effects of molecular risk signatures and genomic instability.

GSEA revealed distinct pathway activation patterns between risk groups (FDR < 0.05). The high-risk group showed significant enrichment in proliferation-related pathways (Adherens Junction; Cell Cycle) and xenobiotic metabolism (Cytochrome P450). Conversely, the low-risk group exhibited enrichment in immune regulation (Primary Immunodeficiency) and neuroendocrine signaling (Neuroactive Ligand-Receptor Interaction) pathways (Fig. 5I).

These findings highlight the significant differences between high-risk and low-risk groups in immune cell infiltration, immune functions, and underlying biological behaviors. These results emphasize the potential of the risk score as not only a reliable prognostic marker but also a valuable predictor of immunotherapeutic responsiveness in patients with pancreatic adenocarcinoma (PAAD).

3.6 Drug sensitivity analysis

To investigate the risk model's clinical applicability, we analyzed differential drug sensitivity between high- and low-risk groups using the pRRophetic algorithm. Based on their transcriptomic profiles, this computational approach estimated the half-maximal

inhibitory concentration (IC_{50}) of standard chemotherapeutic and targeted therapies for pancreatic adenocarcinoma (PAAD) patients.

Distinct drug response patterns were observed across risk groups, highlighting potential variations in treatment efficacy. High-risk patients showed enhanced sensitivity (reduced IC_{50}) to tyrosine kinase inhibitors (e.g., Erlotinib, Sapitinib). In contrast, low-risk patients responded more to PARP and BCR-ABL inhibitors (Olaparib and Nilotinib, respectively) (Fig. 6A).

These findings suggest that risk stratification may inform personalized chemotherapy regimens.

3.6.1 Analysis of ITH-Related changes based on Single-Cell transcriptomics

Building upon the insights from bulk RNA sequencing, we sought to explore the distribution of ITH across distinct cell populations at the single-cell resolution. To enable

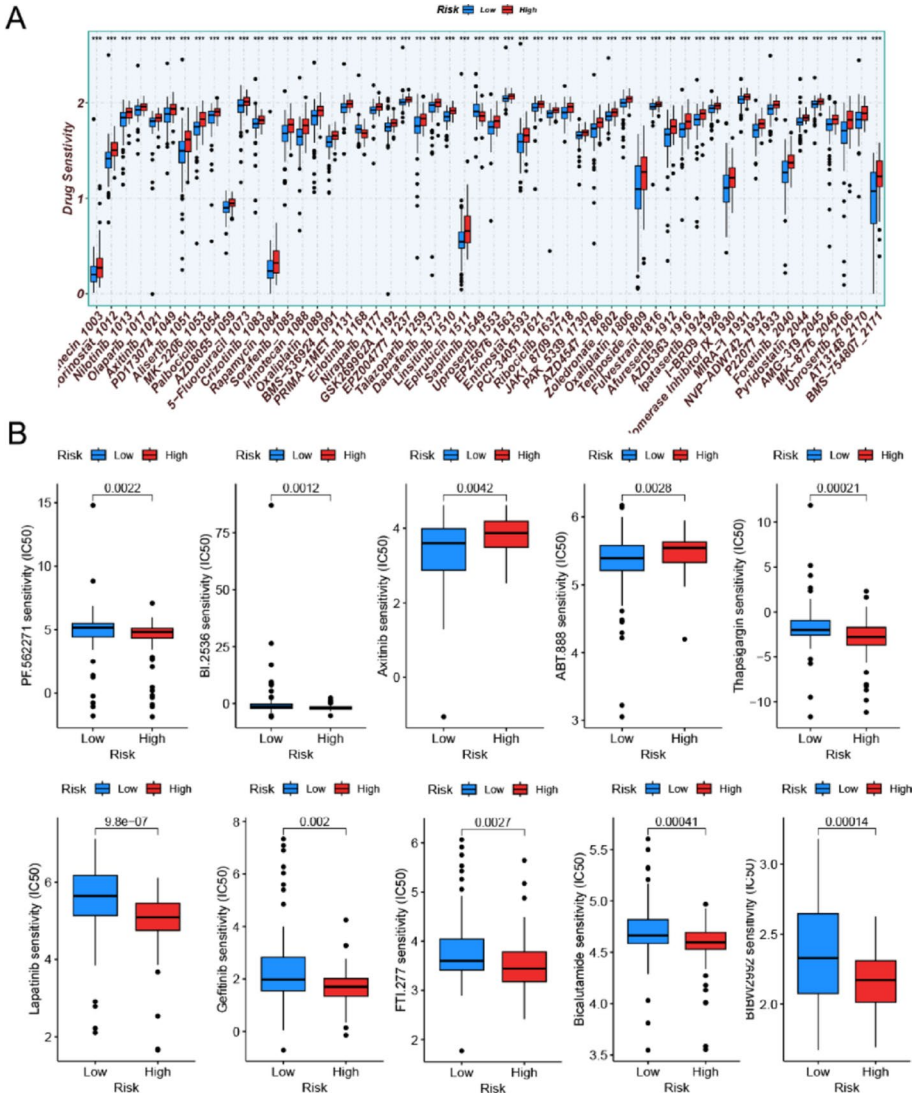


Fig. 6 Chemotherapeutic drugs sensitivity analysis. **(A)** Boxplot visualization of estimated IC50 values for common anti-tumor agents across pancreatic cancer patient subgroups stratified by ITH score; **(B)** A comparative analysis of drug sensitivity was conducted between the two risk groups

downstream analysis, we processed the single-cell RNA-seq dataset GSE212966, which included three pancreatic cancer and three peri-cancerous tissue samples.

After quality control (Figure S1-D), dimensionality reduction via t-SNE and UMAP identified 21 distinct cell clusters manually annotated using established cell-type-specific markers (Figure S2). The spatial distribution of these clusters was visualized using t-SNE (Fig. 7A-B) and UMAP plots (Fig. 7C). To quantify ITH at the single-cell level, we employed the "AddModuleScore" function to score each cell based on a curated set of ITH-related signature genes. The results demonstrated that ITH scores were significantly elevated in tumor tissues (Fig. 7D), particularly in epithelial cells (Fig. 7E-G). Subsequently, we identified 10 distinct epithelial cell subpopulations (Fig. 7H). The first epithelial cluster exhibited the highest ITH score (Fig. 7I). To further explore the dynamic progression of epithelial cells, we constructed developmental trajectories to infer the differentiation states between tumor and adjacent normal tissues. Pseudotime analysis revealed three distinct cellular states across the epithelial subpopulations. Epi2–6 subtypes were predominantly enriched in states 2 and 3, whereas the Epi0 subtype was confined mainly to state 1. Based on pseudotime annotations, state four was considered a potential progenitor/root-like state (Fig. 7J).

Based on transcriptomic features, endothelial cells were categorized into three ITH levels: high, medium, and low. To investigate intercellular communication, we performed CellChat analysis to examine interaction differences among endothelial subgroups and other cell types. As shown in Fig. 8A, the number of receptor–ligand interactions and the cumulative strength of intercellular communication were quantified across cell

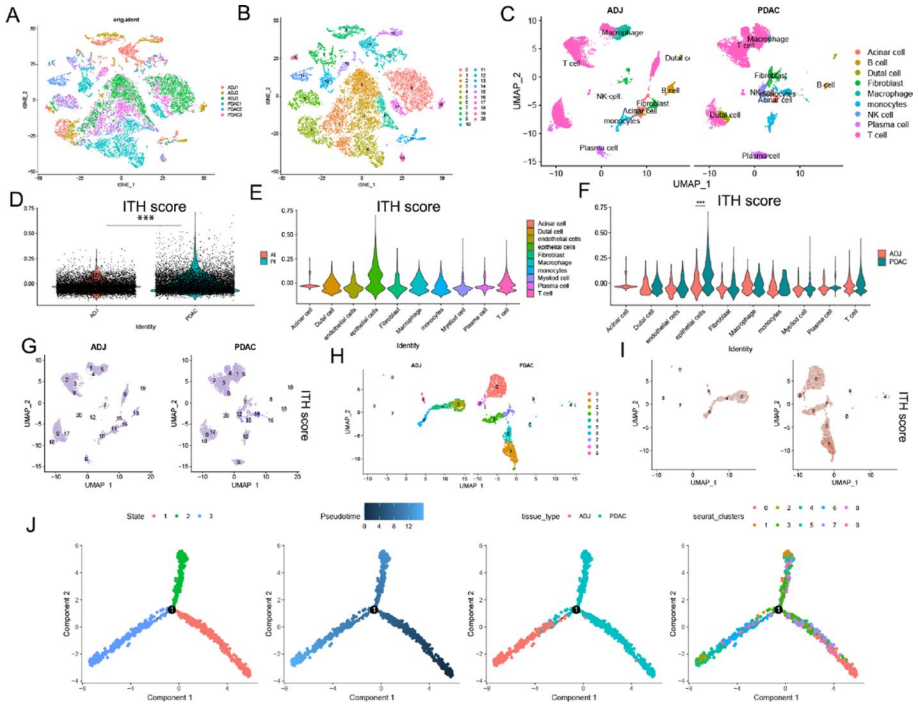


Fig. 7 Single-cell analysis of ITH score in PAAD **(A)** t-SNE plot of normal and tumor cells; **(B)** t-SNE plot of 21 cell clusters; **(C)** UMAP of nine cell types; **(D)** A comparative analysis of ITH score between PAAD and adjacent tissues; **(E)** Violin plots showing ITH score in different clusters of PAAD; **(F)** Violin plots showing ITH score in PAAD and adjacent tissues; **(G)** UMAP projection showing ITH score in different clusters; **(H)** Display clusters of endothelial cells by t-SNE plot; **(I)** Display the ITH scores of different endothelial cell clusters; **(J)** Differentiation trajectory colored for states, pseudotime, tissue types, and clusters

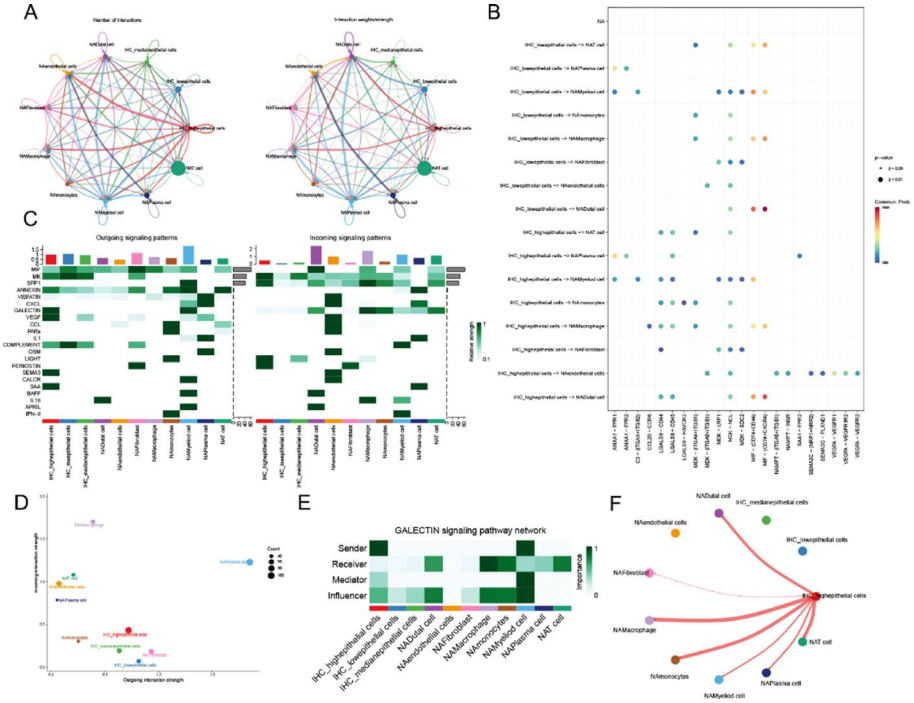


Fig. 8 Cell communication analysis of scRNA-seq data. **(A)** Number of interactions and intensity of action of different cell types in the normal and tumor groups; **(B)** Bubble diagram of receptor-mediated cellular communication; **(C)** The contribution of cell populations to all signaling pathways; **(D)** Visualize the primary signal sender. (source) and signal receiver (target); **E.** Display of GALECTIN SIGNALING PATHWAY among different cell types; **F.** ITH highepithelial cell communication networks with other cell types;

types. Thicker lines represent stronger interactions; line colors match the source cell population, while the opposing nodes denote the target population.

ITH-high endothelial cells exhibited significantly enhanced interaction strength with most other cell types. Ligand–receptor correlation analysis suggested that ITH-associated epithelial cells may influence ductal cells and macrophages through signaling axes such as MIF-(CD74 + CXCR4) and MIF-(CD74 + CD44). (Fig. 8B)

Further analyses highlighted Galectin signaling as the dominant outgoing communication pathway in ITH-high endothelial cells, whereas SPP1-related pathways were identified as major contributors to incoming signals(Fig. 8C). Scatterplots were used to visualize key sender–receiver cell types(Fig. 8D). Compared to other cell populations, ITH-high endothelial cells were identified as significant sources of GALECTIN signaling, which predominantly targeted macrophages. Additionally, GALECTIN pathways mediated interactions between ITH-high endothelial cells and monocytes, T cells, and ductal epithelial cells, suggesting a critical role of this signaling axis in promoting PAAD progression(Figs. 8E-F).

4 Discussion

This study explores ITH in PC, examining its prognostic significance, influence on the tumor immune microenvironment, and implications for treatment response. Our results demonstrate that ITH strongly correlates with survival outcomes in pancreatic cancer, supporting its potential clinical utility as a robust prognostic biomarker for patient stratification.

Interestingly, in contrast to many other cancer types, our analysis revealed that lower transcriptomic ITH was associated with worse prognosis in pancreatic cancer. One possible explanation is that low ITH reflects a monoclonal, immune-evading phenotype that dominates in the hostile stromal and hypovascular environment characteristic of pancreatic tumors. Such tumors may have undergone clonal selection and immune editing, favoring aggressive behavior despite apparent transcriptomic uniformity. Furthermore, late-stage tumors may exhibit transcriptional convergence toward stress-adaptive pathways, reducing ITH scores. These findings align with recent spatial transcriptomic studies that report poor outcomes in phenotypically homogeneous yet biologically aggressive PDAC subregions [17–19]. Lower ITH scores were notably found in patients with higher tumor grade and advanced clinical stage, further corroborating the link between ITH and tumor aggressiveness.

Comparative analysis revealed 1,848 ITH-associated DEGs, with functional annotation implicating oncogenic pathways - particularly PI3K-Akt signaling and ECM remodeling- as key molecular determinants of tumor heterogeneity-driven progression. These findings suggest that the molecular mechanisms underlying ITH in pancreatic cancer involve pathways critical to tumor invasion, metastasis, and immune evasion. We notably uncovered multiple genes comprising detrimental and beneficial factors that may represent viable targets for therapy or serve as biomarkers to assess the progression associated with intratumoral heterogeneity.

We constructed a robust intratumoral heterogeneity-related signature (ITH-S) using an integrated machine learning framework to better stratify patients based on ITH-related risks. The risk model demonstrated strong prognostic performance, with significantly shorter overall survival in high-risk patients versus low-risk counterparts. Our model's accuracy was confirmed through multiple validation cohorts, and the C-index consistently exceeded that of other clinical variables, reinforcing the model's clinical applicability. After controlling for conventional clinical variables, the risk score exhibited strong prognostic significance, underscoring its value as an independent predictor of survival in PAAD. This comprehensive model integrates ITH with other clinical features, providing a refined tool for individualized prognosis prediction. Several genes included in our 11-gene prognostic model are supported by prior studies for their involvement in pancreatic cancer progression, immune modulation, or therapeutic resistance. For instance, LGALS1 (Galectin-1) is a well-characterized immunosuppressive mediator that promotes T cell dysfunction and stromal remodeling in PDAC. CTSB and S100A4 facilitate invasion and metastasis through ECM degradation and EMT induction, respectively, and are linked to chemoresistance. IFI16, an innate immune DNA sensor, modulates interferon pathways and may affect immune evasion. TSPAN1 is overexpressed in PC and promotes proliferation, cancer stemness, and drug resistance. These findings support the risk model's biological plausibility while identifying novel candidate genes for further validation.

The TME has emerged as a critical determinant of tumor behavior, with its composition and functional state strongly correlating with treatment efficacy and survival outcomes in cancer patients [20, 21]. The observed immune infiltration disparities between risk groups imply a link between ITH and immune landscape remodeling. Specifically, low-ITH tumors may create an immunologically 'cold' environment that facilitates immune evasion rather than effective immune activation. Furthermore, significant

differences in immune checkpoint expression between the two groups imply that ITH-related stratification could help predict response to immunotherapies, with the low-risk group potentially benefiting more from immune checkpoint blockade.

Consistent with the immune profiling, TIDE analysis indicated that low-ITH patients might benefit more from immune checkpoint blockade due to their lower dysfunction and exclusion scores, underscoring the potential of ITH-guided immunotherapy stratification.

The ITH-based stratification revealed distinct drug response profiles: low-risk patients demonstrated increased susceptibility to chemotherapeutics targeting the cell cycle and DNA repair mechanisms. At the same time, high-risk individuals responded more favorably to agents targeting immune checkpoint pathways and cellular metabolism. These findings indicate the potential for tailoring chemotherapy regimens based on ITH-related risk stratification. Moreover, the observed therapeutic sensitivity differences suggest that personalized treatments incorporating conventional and immune-targeted therapies could improve treatment outcomes in pancreatic cancer patients.

Several limitations should be noted. The retrospective design based on publicly available datasets (TCGA and GEO) may introduce inherent biases in sample selection and processing. Although our model demonstrated good performance when validated using the GEO dataset, we acknowledge that the sample size of the validation cohort was relatively small, and no external multi-center datasets were included for further validation. This limitation may affect the generalizability and robustness of our findings across different populations and clinical settings. While our drug sensitivity analysis provided useful in silico insights, we acknowledge the limitation of not having matched clinical drug response data, such as progression-free survival after gemcitabine or erlotinib treatment. Future validation using patient-level treatment outcomes will be essential to confirm the predictive utility of the ITH-based stratification model and its potential application in guiding personalized chemotherapy in pancreatic cancer.

While our exploration of immune infiltration and drug response shed light on the immunological landscape of pancreatic cancer, further research is essential to clarify the molecular drivers of immune dysregulation and treatment resistance in tumors with low intratumoral heterogeneity.

5 Conclusion

Our study highlights the clinical significance of intratumoral heterogeneity in pancreatic cancer, both as a prognostic marker and determinant of immune response and therapeutic sensitivity. By integrating ITH and machine learning techniques, we have created a robust prognostic signature capable of predicting patient outcomes and informing personalized treatment strategies. Further research into the molecular and immune mechanisms underlying ITH will be critical for developing more effective therapies for pancreatic cancer.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-03080-3.

Supplementary Material 1

Acknowledgements
We sincerely thank the contributors of the TCGA and GEO databases for providing invaluable data that greatly supported this research.

Author contributions
The conception and design of the study were contributed to by all authors. Li Zhao, Jiang Li, Rui Xie, and Kaiming Li handled the preparation of materials, data collection, and analysis. Li Zhao wrote the initial manuscript draft, with input and revisions from all authors based on earlier versions. Kaiming Li supervised the project and provided additional revisions to the manuscript. Yao Huang contributed significantly to the revision of the manuscript. Specifically, Yao Huang provided critical feedback and carried out substantial modifications in response to reviewers' comments, which greatly improved the clarity, depth, and scientific rigor of the final version. We acknowledge Yao Huang as a co-author in recognition of these essential contributions during the revision process. All authors have read and approved the final manuscript.

Funding
The joint Funds of the Zhejiang Provincial Natural Science Foundation of China under Grant No. LYY21H300002, and the Wu Jieping Medical Foundation Special Fund for Scientific Research (Grant No. 320.6750.2024-18-11) funded this research.

Data availability
Data can be obtained from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The relevant guidelines and regulations were followed in all methods.

Consent for publication
Not application.

Competing interests
The authors declare no competing interests.

Conflict of interest
The author reports no conflicts of interest in this work.

Received: 3 May 2025 / Accepted: 25 June 2025
Published online: 13 July 2025

References

1. Klein AP. Pancreatic cancer epidemiology: Understanding the role of lifestyle and inherited risk factors. Nat Rev Gastroenterol Hepatol. 2021;18(7):493–502.
2. Stoffel EM, Brand RE, Goggins M. Pancreatic cancer: changing epidemiology and new approaches to risk assessment, early detection, and prevention. Gastroenterology. 2023;164(5):752–65.
3. Hinohara K, Polyak K. Intratumoral heterogeneity: more than just mutations. Trends Cell Biol. 2019;29(7):569–79.
4. Hurson AN, Hamilton AM, Olsson LT, Kirk EL, Sherman ME, Calhoun BC, et al. Reproducibility and intratumoral heterogeneity of the PAM50 breast cancer assay. Breast Cancer Res Treat. 2023;199(1):147–54.
5. Hirose Y, Taniguchi K. Intratumoral metabolic heterogeneity of colorectal cancer. Am J Physiol Cell Physiol. 2023;325(4):C1073–84.
6. Quan Y, Zhang H, Wang M, Ping H. Visium Spatial transcriptomics reveals intratumor heterogeneity and profiles of Gleason score progression in prostate cancer. iScience. 2023;26(12):108429.
7. Anwaier A, Xu W, Liu W, Wei S, Tian X, Qu Y, et al. Tumor microenvironment-based signatures distinguish intratumoral heterogeneity, prognosis, and Immunogenomic features of clear cell renal cell carcinoma. J Natl Cancer Cent. 2023;3(3):236–49.
8. Liao J, Qian J, Fang Y, Chen Z, Zhuang X, Zhang N, et al. De Novo analysis of bulk RNA-seq data at spatially resolved single-cell resolution. Nat Commun. 2022;13(1):6498.
9. Zhang Y, Huang X, Yu M, Zhang M, Zhao L, Yan Y, et al. The integrated profiling of single-cell and Spatial transcriptome RNA-seq reveals tumor heterogeneity, therapeutic targets, and prognostic subtypes in CcRCC. Cancer Gene Ther. 2024;31(6):917–32.
10. Yu Z, Zhang J, Zhang Q, Wei S, Shi R, Zhao R, et al. Single-cell sequencing reveals the heterogeneity and intratumoral crosstalk in human endometrial cancer. Cell Prolif. 2022;55(6):e13249.
11. Kalbasi A, Ribas A. Tumour-intrinsic resistance to immune checkpoint Blockade. Nat Rev Immunol. 2020;20(1):25–39.
12. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16(5):284–7.
13. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417–25.
14. Ferrer-Bonsoms JA, Jareno L, Rubio A. Rediscover: an R package to identify mutually exclusive mutations. Bioinformatics. 2022;38(3):844–5.
15. 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(9):e107468.

16. Maeser D, Gruener RF, Huang RS. OncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. Brief Bioinform 2021, 22(6).
17. Pandiani C, Strub T, Nottet N, Cheli Y, Gambi G, Bille K, et al. Single-cell RNA sequencing reveals intratumoral heterogeneity in primary uveal melanomas and identifies HES6 as a driver of the metastatic disease. Cell Death Differ. 2021;28(6):1990–2000.
18. Thomas DD, Lacinski RA, Lindsey BA. Single-cell RNA-seq reveals intratumoral heterogeneity in osteosarcoma patients: A review. J Bone Oncol. 2023;39:100475.
19. Heming M, Haessner S, Wolbert J, Lu IN, Li X, Brokinkel B, et al. Intratumor heterogeneity and T cell exhaustion in primary CNS lymphoma. Genome Med. 2022;14(1):109.
20. Khosravi GR, Mostafavi S, Bastan S, Ebrahimi N, Gharibvand RS, Eskandari N. Immunologic tumor microenvironment modulators for turning cold tumors hot. Cancer Commun (Lond). 2024;44(5):521–53.
21. Jarosz-Biej M, Smolarczyk R, Cichoń T, Kułach N. Tumor microenvironment as A game changer in Cancer radiotherapy. Int J Mol Sci 2019, 20(13).

Publisher's note

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