Discover Oncology

Analysis

Prognostic model identification of ribosome biogenesis-related genes in pancreatic cancer based on multiple machine learning analyses

Yuan Sun[1] · Yan Li[1] · Anlan Zhang[1] · Tao Hu[1] · Ming Li[1]

Received: 31 December 2024 / Accepted: 16 May 2025
Published online: 24 May 2025
© The Author(s) 2025 OPEN

Abstract

Background Pancreatic cancer is a highly aggressive cancer characterized by low survival rate. Enhanced ribosome biogenesis may be associated with tumor drug resistance and malignant phenotypes, representing a potential therapeutic target in pancreatic cancer. Therefore, exploring the molecular mechanisms of ribosome biogenesis in pancreatic cancer may uncover new biomarkers and potential therapeutic targets, facilitating the development of personalized treatment strategies.

Methods Ribosome biogenesis-related gene signatures were acquired from TCGA and Gene Cards databases. Prognostic gene sets were screened using machine learning algorithms to construct a risk model, which was externally validated via GEO database. Single-cell RNA sequencing analysis (GSE155698 dataset) was performed to assess gene expression patterns and module scores.

Results Sixty ribosome biogenesis-related prognostic genes were identified in pancreatic cancer. Cox regression and machine learning algorithms selected nine pivotal biomarkers (*ECT2; CKB; HMGA2; TPX2; ERBB3; SLC2A1; KRT13; PRSS3; CRABP2*) with high diagnostic and prognostic specificity for PAAD. The machine learning-derived risk score correlated strongly with tumor proliferation pathways and immunosuppression, suggesting dual roles in tumor promotion and immunosuppressive microenvironment remodeling. Single-cell analysis highlighted predominant expression of *CKB, SLC2A1, ERBB3, CRABP2*, and *PRSS3* in pancreatic ductal epithelial cells.

Conclusions Our results shed light on the potential connections between ribosome biogenesis-related molecular characteristics and clinical features, the tumor microenvironment, and clinical drug responses. The research underscores the critical role of ribosome biogenesis in the progression and treatment resistance of pancreatic cancer, offering valuable new perspectives for prognostic evaluation and therapeutic response prediction in pancreatic cancer.

Keywords Pancreatic cancer · Ribosome biogenesis · Prognosis · Tumor microenvironment · Drug sensitivity

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02733-7.

Yuan Sun and Yan Li have contributed equally to this work and share first authorship.

✉ Ming Li, liming@hospital.cqmu.edu.cn | [1]Department of Hepatobiliary Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China.

1 Introduction

Pancreatic cancer (pancreatic adenocarcinoma,PAAD) is the most aggressive malignant tumor in the digestive system and ranks seventh in global cancer-related deaths. Due to the lack of early symptoms and effective diagnostic markers, most patients are diagnosed at an advanced stage, with a 5-year survival rate of only approximately 9% [1]. Pancreatic cancer has a very poor prognosis. Although surgical resection remains the preferred treatment for PAAD, only 15–20% of patients are eligible for surgery at the time of diagnosis [2, 3]. In recent years, new strategies such as FOLFIRINOX combined chemotherapy and neoadjuvant therapy have significantly improved the survival rates of some patients [4]. However, the efficacy of current chemotherapy regimens is limited by drug resistance [5], highlighting the urgent need for new adjuvant treatment strategies and biomarkers to better assess prognosis and guide personalized treatment.

Ribosome biogenesis is the process of assembling ribosomal complexes, a critical and complex process closely related to cell growth and protein synthesis, and is strictly regulated by numerous auxiliary factors [6, 7]. In tumor cells, ribosome biogenesis is often upregulated and regulated by various signaling pathways, such as mTOR and Myc, to meet the high demand for protein synthesis required for rapid tumor growth, proliferation, and migration. This upregulation involves enhanced function of the nucleolus (the primary site of ribosomal RNA synthesis) and the overexpression of genes associated with ribosome assembly [8–10]. In pancreatic cancer research, ribosome biogenesis plays a significant role in the initiation and progression of the disease. For example, it has been reported that oncogenic RAS signaling enhances the RNA-binding activity of the nucleolar protein nucleolin (Ncl), reshaping the RNA-binding protein network of pancreatic cancer cells and leading to the phosphorylation of Ncl downstream of Erk1/2. The phosphorylation of Ncl enhances the binding of precursor rRNA and ribosome biogenesis. This nucleolin-dependent enhancement of ribosome biogenesis is crucial for RAS-induced proliferation of pancreatic cancer cells. Targeted therapies directed at this process can effectively inhibit tumor growth. [11, 12]. Additionally, studies have shown that proteins involved in ribosome assembly, such as RRP9, can activate certain signaling pathways, thereby increasing the resistance of pancreatic cancer cells to the drug gemcitabine [13]. Moreover, ribosome biogenesis inhibitors targeting ribosome biogenesis may prevent the progression of pancreatic cancer by inducing oncogene-induced senescence in combination with ERK hyperactivation [14].

Therefore, this study aims to leverage multiple machine learning approaches to construct a robust prognostic model for assessing the clinical outcomes of pancreatic cancer patients. Additionally, it seeks to elucidate the critical role of ribosome biogenesis in pancreatic cancer initiation and progression, analyze somatic mutations associated with the disease, and identify novel therapeutic targets to advance treatment strategies.

2 Materials and methods

2.1 Data sources

A total of 177 pancreatic cancer tissue samples (after excluding one sample with missing survival data) were selected from The Cancer Genome Atlas (TCGA) database (accessed on September 26, 2024). Clinical and pathological information, including age, sex, tumor grade, tumor stage, tumor type, tumor location, TNM staging, survival days, and survival status, were obtained. Transcriptomic sequencing data (seq-tpm and seq-count), methylation level data and microRNA expression matrix data were also acquired, along with transcriptomic sequencing data (seq-tpm and seq-count) from 180 normal pancreatic tissue samples sourced from the Genotype-Tissue Expression (GTEx) database (accessed on September 26, 2024). All transcriptomic sequencing data were normalized. Ribosome biogenesis-related genes were obtained from the GeneCards website (accessed on September 22, 2024). Selected only protein-coding genes, and set the relevance threshold to above 1.2 from GeneCards.

Validation cohorts, including ICGC-PACA-AU, GSE78229, GSE183795, and GSE62452, were retrieved from the International Cancer Genome Consortium (ICGC) and Gene Expression Omnibus (GEO) databases. These datasets collectively demonstrate the reproducibility of the research findings. Additionally, single-cell RNA sequencing data (scRNA-seq) for 10 PAAD tumor tissues (GSE155698) were used for single-cell analysis. Immunohistochemistry (IHC) data were downloaded from the Human Protein Atlas (HPA, http://www.proteinatlas.org) to verify protein expression in PAAD tissues.

2.2 Methods

2.2.1 Differential gene expression analysis between PAAD and non-tumor tissues

To identify transcriptomic differences between PAAD and non-tumor tissues, 177 PAAD tumor samples were downloaded from the TCGA database (https://www.cancer.gov/ccg/research/genome-sequencing/tcga), and 180 normal pancreatic tissue samples were obtained from the GTEx database (https://gtexportal.org/home/) as controls. Using the DESeq2 package (version 1.42.0) in R, differentially expressed genes (*DEGs*) between tumor and non-tumor samples were identified based on |fold-change|> 1.5 and adjusted p-value < 0.05. The results were visualized using the "**ggplot2** "package (version 3.5.0).

2.2.2 Weighted gene co-expression network analysis (WGCNA) and identification of ribosome biogenesis-related genes in PAAD

WGCNA was applied to further identify gene modules closely related to tumors in the TCGA-PAAD dataset [15]. Using the **WGCNA** package (version 1.73) in R, *DEGs* of PAAD were analyzed. A gene expression matrix was constructed based on 4,725 *DEGs* from 357 samples for subsequent analysis. Since the module-trait correlation is strongest at the corresponding inflection point, the soft thresholding power was set to 5 to construct the gene co-expression network. After that, the gene–gene co-expression similarity and adjacency are calculated and converted into a Topological Overlap.

Matrix (TOM). Hierarchical clustering based on TOM was then used to assign genes to modules. Through this process, gene modules significantly associated with clinical features were identified. In module analysis, key genes were screened by comparing gene significance (GS, indicating the correlation between the gene and clinical features) with module membership (MM, indicating the correlation between the gene and the module). Ultimately, upregulated genes associated with ribosome biogenesis (up CARGs) were extracted from these tumor-related modules. Simultaneously, 1250 ribosome biogenesis-related genes were collected from the GeneCards database (https://www.genecards.org/) with the filtering criteria being the selection of protein-coding genes only and a relevance score threshold set above 1.2. Using the **VennDiagram** package (version 1.7.3) in R, an intersection analysis was performed between upregulated CARGs in PAAD and ribosome biogenesis-related genes, resulting in 60 significantly upregulated ribosome biogenesis-related genes in PAAD, which were prioritized for subsequent analyses.

2.2.3 Construction and validation of prognostic signatures based on multiple machine learning methods for ribosome biogenesis-related genes

To construct a prognostic model based on the identified 60 Ribosome Biogenesis-related genes (CARGs), univariate Cox regression analysis was performed. Subsequently, the R package **Mime1** (version 1.0.0) was employed, leveraging its foundation on the performance of a comprehensive ensemble comprising 100 + machine learning models. By evaluating multiple model combinations, the optimal ensemble was identified and selected, significantly enhancing the overall performance of the model.

These methods included StepCox (stepwise regression), RSF (random survival forest), SuperPC (supervised principal component analysis), Lasso (least absolute shrinkage and selection operator), Survival-SVM (survival support vector machine), GBM (gradient boosting machine), CoxBoost (boosting algorithm for Cox regression), Enet (elastic net regression), Ridge (ridge regression), and plsRcox (partial least squares regression for Cox models).

Models were constructed using different machine learning algorithms, mixing them up, and worked with a dataset that had 177 samples—93 of which were death events. We figured out the concordance index (C-index) for each model. Furthermore, the predictive performance of the models was assessed using the TCGA-PAAD dataset as the training set, while the ICGC-PACA-AU, GSE78229, GSE183795, and GSE62452 datasets were used as independent validation sets. The average C-index across these validation datasets was also computed to ensure robustness.

Based on the median risk score calculated from the best model combination, the TCGA-PAAD training cohort was divided into low-risk and high-risk groups. Kaplan–Meier survival curves were plotted using the **survival** (version 3.6–4) and **survminer** packages (version 0.4.9), and log-rank tests were performed to assess survival differences between the two groups. Receiver operating characteristic (ROC) curves were generated using the **timeROC** package (version 0.4.2) to calculate the area under the curve (AUC), evaluating the model's sensitivity and specificity in predicting 2-year, 3-year, and 5-year survival probabilities. Additionally, we normalized data and corrected batch effects using

the normalizeBetweenArrays function and the ComBat function from the sva package. The validation set, comprising pancreatic cancer tissue samples, was sourced from ICGC-PACA-AU (International Cancer Genome Consortium) and GEO datasets (GSE78229, GSE183795, GSE62452), potentially including samples from diverse countries or institutions. It included gene expression matrices and survival indicators, serving as a counterpart to the TCGA-PAAD training set. Using GEO datasets, we validated the robustness and effectiveness of the CARGs prognostic model. We also analyzed the correlation between risk scores and corresponding clinical indicators using the Wilcoxon rank-sum test and visualized the results with heatmaps generated by the **heatmap** package (version 1.0.12).

We applied univariate and multivariate Cox proportional hazards models to evaluate whether the CARGs prognostic model could independently predict outcomes after adjusting for other clinical and pathological features (such as age, sex, histological grade, pathological stage, and TNM stage). Based on univariate analysis, independent risk factors were integrated to construct a nomogram for predicting patient prognosis. Calibration curves were plotted to assess the consistency between predicted survival rates and actual observations.

2.2.4 Identification of differentially expressed genes between high and low-risk groups and functional enrichment analysis

Using the **DESeq2** package in R, *DEGs* between the high-risk and low-risk groups (DEGs 2) were identified based on |fold-change|> 1 and adjusted p-value < 0.05. Functional enrichment analysis of the *DEGs* was performed using the **clusterProfiler** package (version 4.12.0) to evaluate potential biological functions. Gene Set Enrichment Analysis (GSEA) was conducted using the reference gene sets "c2.cp.reactome.v7.0.entrez.gmt" and "h.all.v7.0.symbols.gmt". Pathways were considered statistically significant if the normalized p-value < 0.05 and the false discovery rate (FDR) q-value < 0.25. The most enriched pathways were ranked based on the normalized enrichment score (NES). The **limma** package (version 3.54.2) was used to quantify the differential pathway activities between the high-risk and low-risk groups.

2.2.5 Tumor microenvironment and immune infiltration analysis

Firstly, we applied the single-sample Gene Set Enrichment Analysis (ssGSEA) algorithm to convert the gene expression data from the TCGA dataset into the abundance of 23 immune cell types. The proportions of infiltrating immune subtypes, including activated B cells, activated $CD4^+$ T cells, M1 macrophages, and plasma cells, were assessed. Based on the differential analysis results, the infiltration levels of these immune cells were compared with the Ribosome Biogenesis risk scores. The **Wilcoxon rank-sum test** and Spearman correlation analysis were employed to further explore the relationship between risk scores and immune checkpoint expression. Additionally, the CIBERSORT algorithm was utilized to perform deconvolution analysis of the tumor immune microenvironment, calculating the infiltration levels of 22 immune cell types in each sample. Spearman correlation analysis was conducted to evaluate the correlation between risk scores and the infiltration levels of these 22 immune cell types. Furthermore, Spearman correlation analysis was performed to examine the relationship between the expression levels of the nine key genes and the infiltration degrees of the 22 immune cell types. Subsequently, the ESTIMATE algorithm was applied to evaluate immune and stromal infiltration scores and their correlation with Ribosome Biogenesis risk scores. The results were visualized using scatter plots and linear regression models. The Tumor Immune Dysfunction and Exclusion (TIDE) algorithm was employed to assess the response to immunotherapy, calculating immunotherapy scores for each sample. Additionally, immune scores under different immune checkpoint inhibitors (such as CD44 and TNFSF9) were compared between the high-risk and low-risk groups using immunotherapy data from the TCIA database. The results were visualized using box plots, and the **Wilcoxon rank-sum test** was used to assess the significance of the differences.

2.2.6 Mutation analysis

We utilized the **maftools** package (version 2.20.0) to analyze and visualize somatic mutation data between the high-risk and low-risk groups in the TCGA dataset, comparing the number and mutation spectra between the two groups. The **maftools**, **remodeling2** (version 1.0.0), and **ggpubr** packages (version 0.6.0) were used to generate waterfall plots illustrating the mutation distribution of key mutated genes in each group. Additionally, we combined risk scores with tumor mutation burden (TMB) data to compare TMB differences between the high-risk and low-risk groups. Box plots displaying TMB differences were generated using the **ggpubr** package, and statistical tests were performed. Based on the median TMB value, samples were further classified into high-TMB and low-TMB groups. The chi-square test was used to evaluate the proportion differences between the high-risk and low-risk groups, and the results were visualized using bar charts.

2.2.7 Drug sensitivity evaluation

We employed the **oncoPredict** package (version 0.2.3) in R to assess the predictive capability of Ribosome Biogenesis risk scores for 198 chemotherapy and targeted therapy drugs. In the TCGA-PAAD dataset (n = 177), we calculated the IC50 (half-maximal inhibitory concentration) for various drugs for each patient and categorized the patients into high-risk (n = 88) and low-risk (n = 89) groups. The **Wilcoxon rank-sum test** was used to analyze the differences in IC50 Z-scores between the high-risk and low-risk groups, exploring the potential value of Ribosome Biogenesis risk scores in predicting drug sensitivity.

2.2.8 Correlation analysis of DNA methylation levels and microrna with key genes and risk groups

Gene methylation levels and miRNA expression data for PAAD patients were obtained from the TCGA database. Differential analysis was conducted using the **limma** package in R, with the criteria of |fold-change|> 1 and adjusted p-value < 0.05. Correlation coefficients between the nine key genes and their corresponding gene methylation levels, as well as differentially expressed miRNAs, were calculated.

2.2.9 Single-cell RNA sequencing analysis and hpa immunohistochemistry validation

scRNA-seq data, including raw sequencing data from 10 PAAD tumor tissues(GSE155698,https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155698), were downloaded from the GEO database. Data processing and analysis were performed using **Seurat** (version 4.3.0.1). Cells with a gene count (nFeature_RNA) greater than 200 but less than 5000, and UMI counts (nCount_RNA) between 200 and 50,000 were selected. Cells with mitochondrial UMI counts (pMT) exceeding 20% and red blood cell UMI counts (pHB) exceeding 5% were removed. Data normalization was performed using the "**NormalizeData"** and **ScaleData** functions, and 2000 highly variable genes were identified using the "**FindVariableFeatures"** function. Dimensionality reduction was conducted using Principal Component Analysis (PCA), and batch effects were corrected using the "**RunHarmony"** function. Cell clusters were visualized using t-SNE and UMAP, and cell clusters were manually annotated based on known marker genes. Additionally, IHC data were downloaded from the HPA (http://www.proteinatlas.org) to evaluate the high and low expression levels of the nine key genes in pancreatic cancer tissues.

2.2.10 Statistical analysis

All statistical analyses were performed using RStudio (version R 4.3.2) and relevant R packages. All p-values were based on two-sided tests, and p-values less than 0.05 were considered statistically significant.

3 Results

3.1 Identification of differentially expressed genes in pancreatic cancer and ribosome biogenesis-related genes identification and enrichment analysis

In the TCGA and GTEx cohorts, a total of 4625 *DEGs* were identified between PAAD samples and normal pancreatic tissues, meeting the criteria of p-value < 0.05 and absolute $\log_2$ fold change > 1.5. Among these, 3701 genes were upregulated, and 924 genes were downregulated (Fig. 1).

3.2 Identification of highly correlated gene modules in TCGA-PAAD and ribosome biogenesis-related genes identification and enrichment analysis

To further uncover key genes associated with tumors, WGCNA was employed to identify gene modules highly correlated with tumors in the TCGA-PAAD dataset. Initially, nine outlier samples were identified and excluded from the analysis through sample clustering (Fig. 2a). Subsequently, a soft-thresholding power of 5 was set (Fig. 2a), and a gene dendrogram was constructed to partition the genes into ten modules: red, purple, pink, black, yellow, magenta, green, turquoise, brown, and blue (Fig. 2b). The heatmap illustrating the correlations between different modules is also presented. The results demonstrated that tumor occurrence was significantly positively correlated with the turquoise module

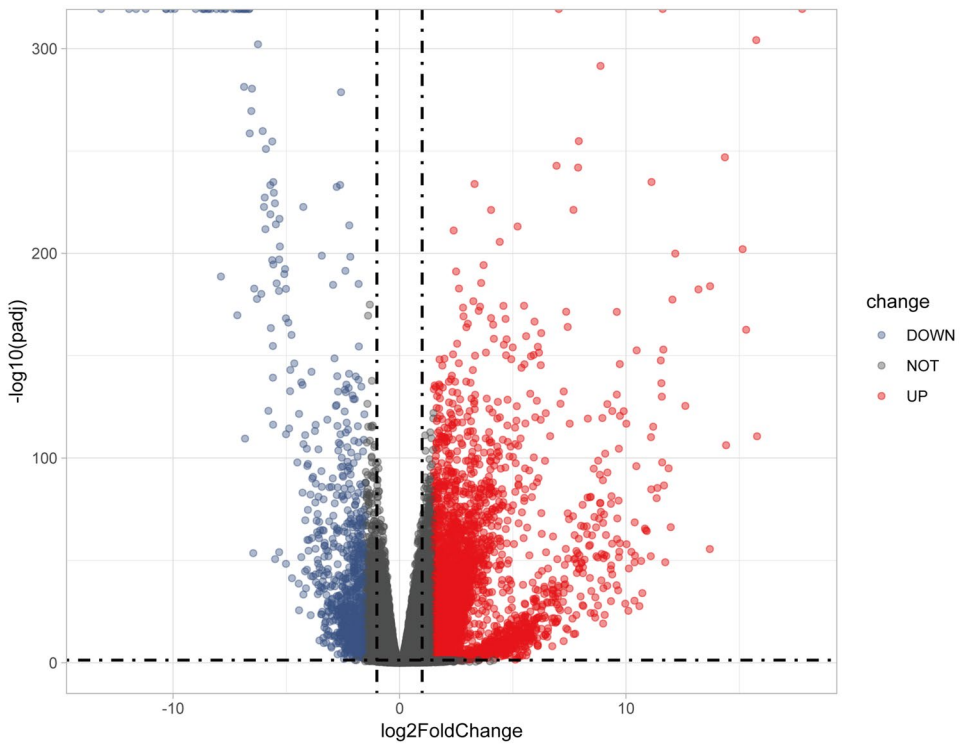


Fig. 1 Identification of DEGs between PAAD tumor and normal tissues

(correlation coefficient = 0.98, p-value = 6e−247) and the blue module (correlation coefficient = 0.88, p-value = 5e−115) (Fig. 2d). Additionally, the gene significance (GS, indicating the correlation between genes and clinical features) and module membership (MM, indicating the correlation between genes and modules) within the turquoise and blue modules were highly correlated, suggesting a strong association of genes in these modules with tumors (Fig. 2c). Genes in the yellow and green modules also exhibited strong correlations with tumors, whereas genes in other modules showed weaker associations (Figures S1). Finally, 2875 upregulated genes (up CARGs) were extracted from the turquoise and blue modules and intersected with 2259 ribosome biogenesis-related genes obtained from the GeneCards database using the **VennDiagram** package in R. This intersection analysis resulted in 60 key ribosome biogenesis-related genes (Fig. 3a). Principal Component Analysis (PCA) further validated that the expression of these 60 CARGs effectively distinguished tumor samples from non-tumor samples (Figures S2a). Additionally, a heatmap illustrated the differential expression of these 60 CARGs between tumor and non-tumor tissues (Figs. 3b).

Gene Ontology (GO) enrichment analysis was performed on these 60 CARGs to explore potential molecular mechanisms associated with PAAD progression. The results indicated that these DEGs were primarily enriched in chromosomal regions, nuclear division, and mitosis-related processes, suggesting that ribosome biogenesis contributes to the synthesis of new proteins that may play roles in cell cycle control and cell division. In terms of cellular components, enrichment analysis revealed that these genes were closely associated with structures such as the spindle, blood microparticles, and the cell cortex, further indicating that the enrichment of ribosome biogenesis in these cellular components may be related to the enhanced proliferation and migratory capabilities of pancreatic cancer. Additionally, the enriched molecular functions were closely related to DNA binding, bending, and other related activities (Fig. 3c). These results collectively suggest that ribosome biogenesis may play a crucial role in PAAD development and progression by regulating the structure and function of genetic material, as well as extracellular signal transduction. Overall, these findings indicate that ribosome biogenesis may influence PAAD development and progression through the regulation of genetic material structure and gene expression.

3.3 Construction and validation of prognostic model for ribosome biogenesis-related genes

Using multiple machine learning methods, we constructed a risk prediction model based on ribosome biogenesis-related genes from the TCGA database and validated it using GEO datasets. Initially, univariate Cox regression analysis of the 60 identified CARGs revealed 28 prognostically significant genes (Fig. 3d). We then constructed models using

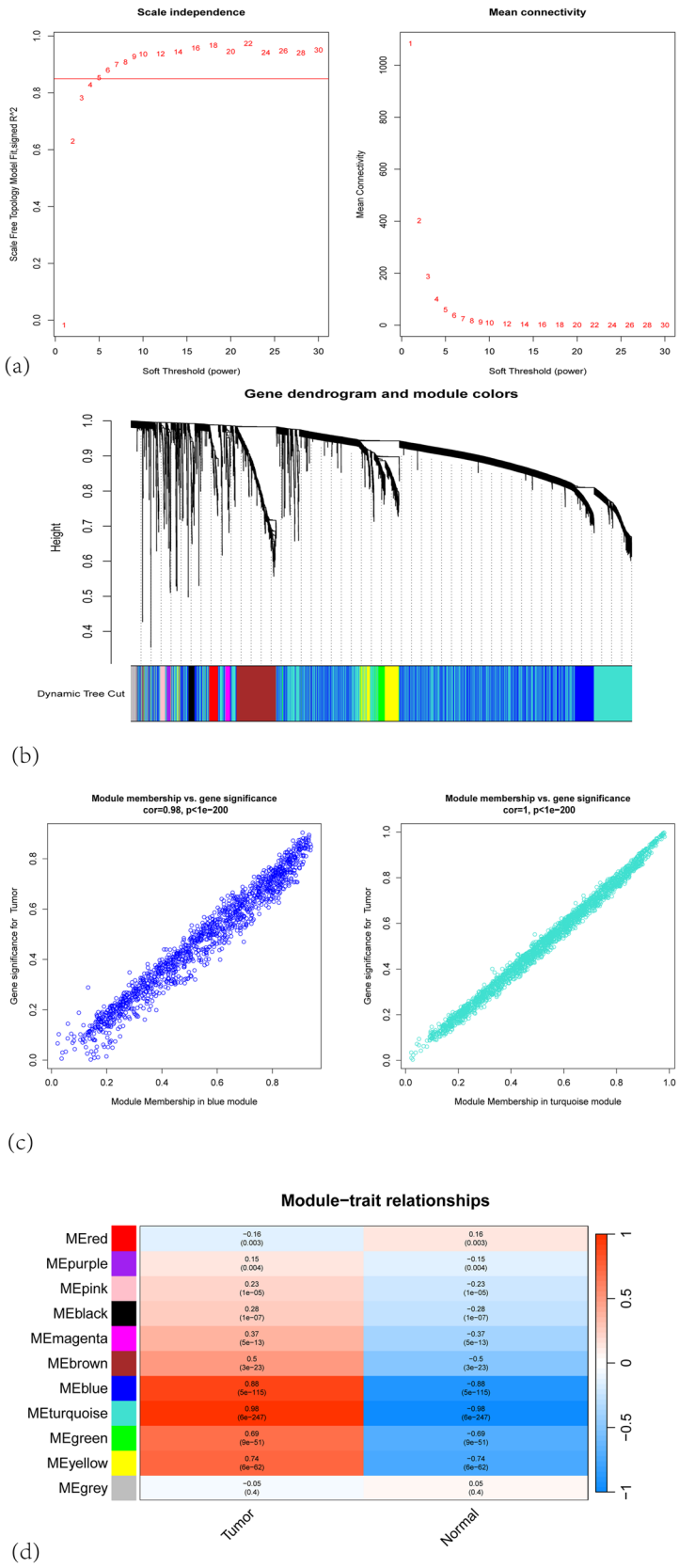


Fig. 2 Identification of the highly correlated gene modules in WGCNA. **a** Soft-thresholding power determination: the left panel's y-axis indicates the scale-free fit index, while the right panel's y-axis indicates mean connectivity. The x-axis represents power values. **b** Dendrogram of differentially expressed genes clustered using a dissimilarity measure (1-TOM), with colors indicating identified gene modules. **c** Gene correlation scatter plots for the turquoise and blue modules: the x-axis represents module membership (MM), indicating the correlation between genes and modules, and the y-axis represents gene significance. **d** Correlations between gene modules and clinical traits: Boolean variables denote clinical phenotypes, representing "Normal" and "Tumor."

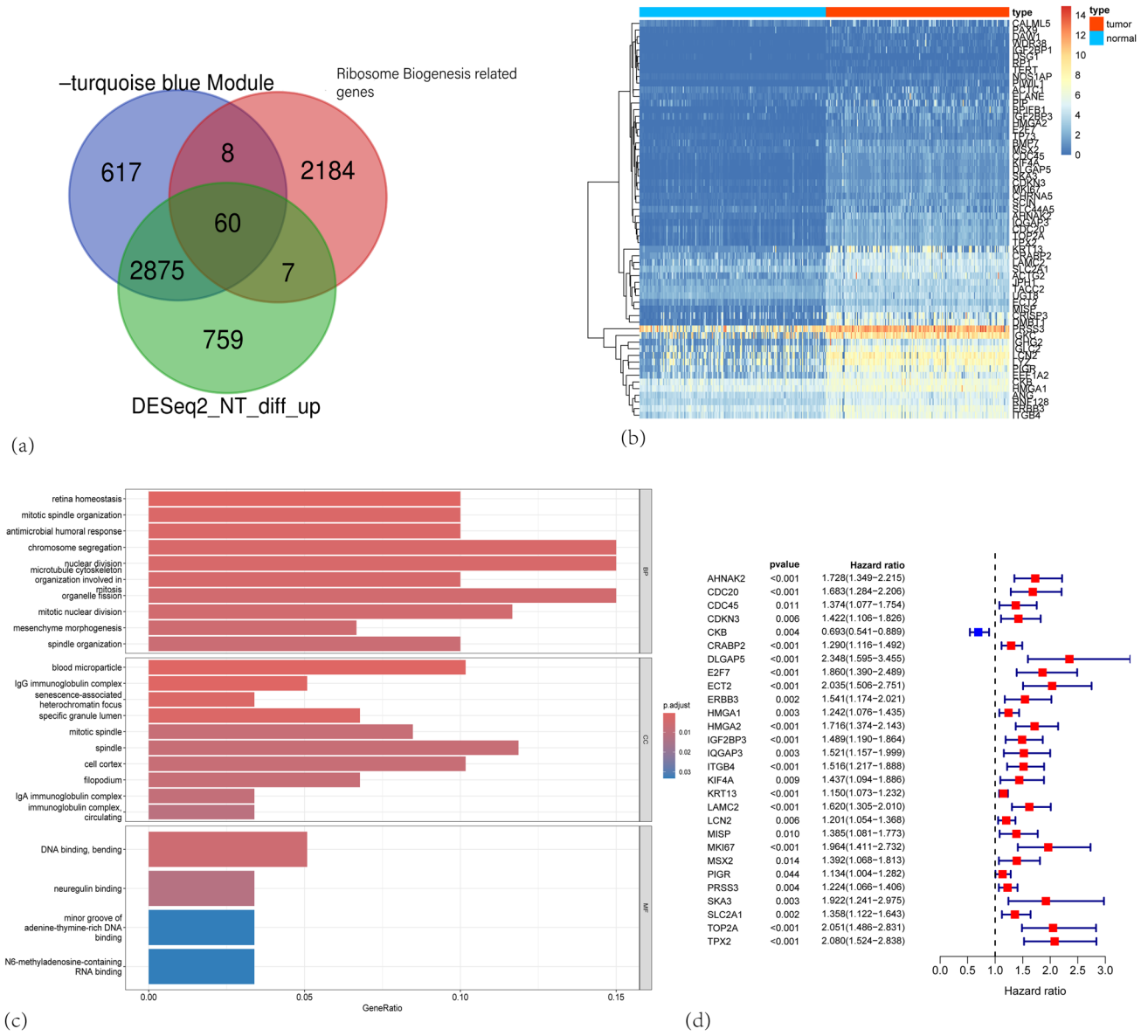


Fig. 3 **a** Identification of the intersection among ribosome biogenesis-related genes, upregulated DEGs, and genes from the WGCNA-identified tumor-associated turquoise and blue modules. **b** Expression levels of Gene 55 in tumor versus normal tissues. **c** GO term analysis of genes identified in the Venn diagram intersection among upregulated DEGs, ribosome biogenesis-related genes, and turquoise and blue modules. **d** Forest plot showing prognosis-related genes in PAAD identified through univariate Cox regression analysis

both individual and combined machine learning algorithms and calculated the concordance index (C-index) for each model. The average C-index across three datasets was computed to evaluate the predictive performance of all models. The optimal model combination was found to be a combination of Lasso (Least Absolute Shrinkage and Selection Operator) and RSF (Random Survival Forest), and the C-index was 0.65 (Fig. 4a). The analytical framework integrated linear screening with nonlinear dynamic modeling through a two-stage approach: initially employing Lasso regression with L1 regularization for secondary dimensionality reduction, where cross-validation identified the optimal λ value (lambda = 0.023) to retain 9 pivotal genes that eliminated multicollinearity while capturing synergistic genomic interactions, subsequently constructing an enhanced random survival forest comprising 1,000 survival decision trees trained via stratified sampling without replacement (SWOR) with 112 samples (63% of total cohort) per bootstrap iteration, constraining terminal nodes to a minimum sample size of 5 to balance overfitting risks (averaging 24.85 nodes per tree), and innovatively implementing a logrank random splitting strategy that randomly selected 3 candidate genomic variables per node evaluation and conducted 10 split attempts at each node to enhance nonlinear relationship detection,

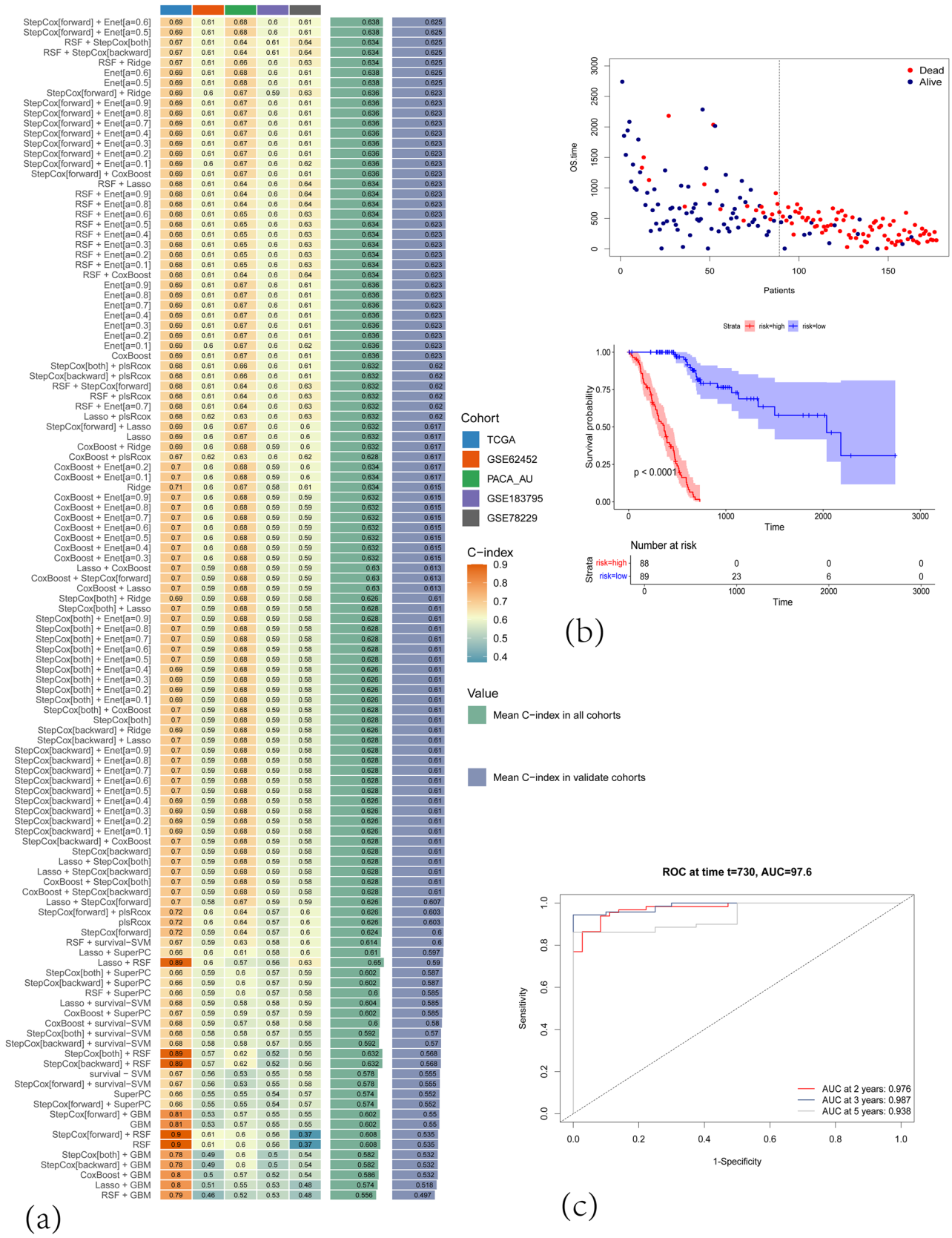


Fig. 4 Construction of a machine learning–based risk-scoring model for PAAD patients using the TCGA database. **a** C-index values for the test and validation prognostic signatures. **b** Patient risk score distribution and Kaplan–Meier survival analysis of TCGA-PAAD patients in high-risk versus low-risk groups. **c** ROC curve analysis of the model's performance for predicting 2-, 3-, and 5-year overall survival (OS)

ultimately achieving a standardized Continuous Ranked Probability Score (CRPS = 0.14) with 45% improvement over conventional Cox models (CRPS > 0.25) and near-perfect alignment between predicted survival curves and actual event distributions (Kolmogorov–Smirnov D = 0.08, p = 0.72), thereby validating the framework's dual capability in genomic signal purification and complex survival pattern reconstruction. These results show that the model has solid predictive capabilities. Ultimately, nine genes were identified as significantly impacting survival outcomes: *CKB*, *ECT2*, *HMGA2*, *TPX2*, *ERBB3*, *SLC2A1*, *KRT13*, *PRSS3*, and *CRABP2* (Figures S3).

Based on this algorithm, risk scores for each sample were calculated, and the median risk score of 40.41 was used as the threshold to classify PAAD patients into low-risk (n = 89) and high-risk (n = 88) groups. Scatter plots demonstrated a significant decline in overall survival (OS) with increasing individual risk scores (Fig. 4b). Kaplan–Meier survival curves and log-rank tests further confirmed that OS was significantly lower in the high-risk group compared to the low-risk group ($p < 0.0001$; Fig. 4b).

To evaluate the performance of the prognostic model, time-dependent ROC curve analysis was conducted to assess the sensitivity and specificity of the model. The area under the curve (AUC) for predicting 2-year, 3-year, and 5-year survival were 0.976, 0.987, and 0.938, respectively, indicating high accuracy in predicting OS (Fig. 4c). The model achieved sensitivities of 0.938, 0.943, and 0.861 for 2-, 3-, and 5-year predictions, respectively, with corresponding specificities of 0.917, 0.995, and 0.986. Subsequently, the feasibility of using ribosome biogenesis features to predict PAAD patient prognosis was validated using three GSE datasets and one ICGC dataset. Kaplan–Meier survival curves showed that patients in the low-risk group had better OS than those in the high-risk group (Fig. 5a, c, e, and g). Robust AUC values for the 2-year, 3-year, and 5-year survival predictions were also observed in both the training and validation cohorts (Fig. 5b, d, f, and h). The model exhibited sensitivities of 0.724 and 0.657 at 2- and 3-year intervals in the ICGC dataset, paired with specificities of 0.600 and 0.833, respectively. Concurrently, the GEO cohort demonstrated robust discriminative power, maintaining consistently high sensitivities (range: 0.62–0.85) and specificities (range: 0.6–0.99) across 2-, 3-, and 5-year prognostic windows (Table 1). Clinical correlation analysis revealed significant associations between risk scores and follow-up tumor status, radiation therapy, and residual tumors (Fig. 6a–d). In summary, the ribosome biogenesis-related prognostic features constructed using multiple machine learning methods demonstrated excellent performance in predicting survival outcomes for PAAD.

3.4 Ribosome biogenesis-related prognostic features as independent prognostic factors in PAAD

To determine independent prognostic factors for PAAD patients, we included ribosome biogenesis-related risk scores and various clinical characteristics (including age, sex, histological grade, T stage, M stage, N stage, history of malignancy, follow-up tumor status, tumor location, DCC histological type, history of radiotherapy, residual tumor, and chronic pancreatitis) in univariate and multivariate Cox regression analyses. The univariate Cox regression results indicated that age (HR = 1.028, 95% CI 1.007–1.049), histological stage (HR = 2.350, 95% CI 1.078–5.126),N stage (HR = 2.0961, 95% CI 1.249–3.519), follow-up tumor status (HR = 2.491, 95% CI 1.077–5.762), radiation therapy (HR = 0.453, 95% CI 0.257–0.798), DCC histological type (HR = 1.770, 95% CI 1.077–2.908), and residual tumor (HR = 1.770, 95% CI 1.097–2.857) were all significant prognostic factors in the TCGA cohort (Fig. 7a). In the multivariate Cox regression analysis, the risk score remained a significant independent prognostic factor closely associated with prognosis (Fig. 7b).

Seven significant clinical variables, along with the risk score, were incorporated into the model. Based on the stepwise Cox regression model, a nomogram was constructed to predict 2-year, 3-year, and 5-year OS. The parameters included in the nomogram were risk score, age, N stage, histological grade, follow-up tumor status, history of radiotherapy, DCC histological type, and residual tumor. The results emphasized the predictive capability of ribosome biogenesis-related risk scores in PAAD prognosis (Fig. 8a). The AUC values for predicting 2-year, 3-year, and 5-year OS were 0.986, 0.990, and 0.949, respectively (Fig. 8b). Subsequently, calibration curves were plotted to assess the performance of the nomogram model. The calibration curves closely aligned with the 45-degree diagonal, indicating a high consistency between the predicted survival probabilities and the actual observed outcomes in the TCGA-PAAD cohort (Fig. 8c).

3.5 Functional analysis of ribosome biogenesis-related gene set and mutation analysis

To investigate the biological behaviors associated with different Ribosome Biogenesis risk score groups, we employed the **GSVA** package in R for functional annotation of TCGA-PAAD samples. The heatmap of GSVA-KEGG results revealed that taurine and hypotaurine metabolism-related pathways were significantly enriched in the high-risk group. Additionally, pathways related to the cell cycle and DNA replication were also markedly enriched in the high-risk group.

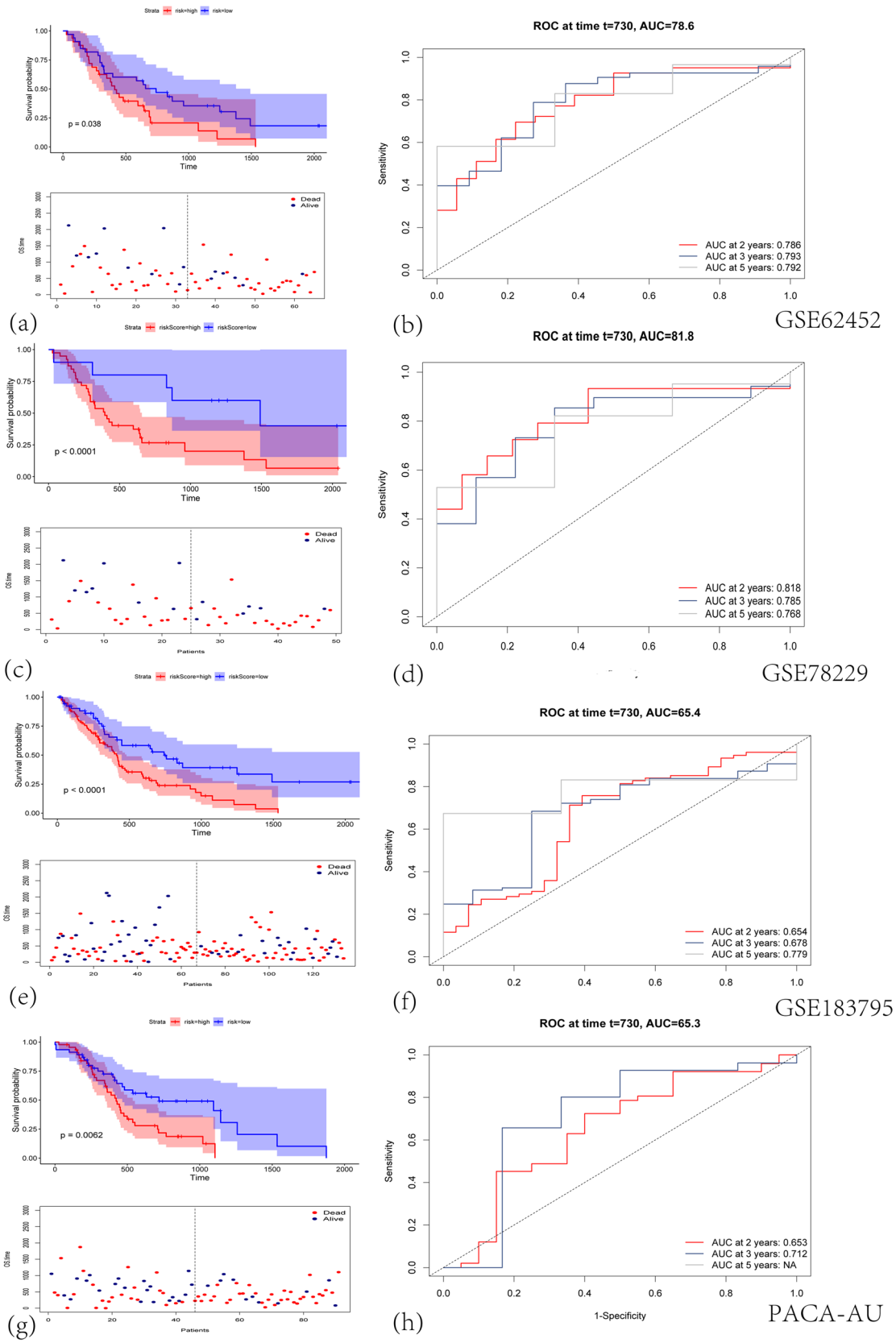


Fig. 5 Survival analysis results of the risk-scoring model for PAAD patients from the GEO and ICGC datasets. **a**, **c**, **e**, **g** Kaplan–Meier survival curves for GEO PAAD patients classified into high-risk and low-risk groups. **b**, **d**, **f**, **h** ROC curve analysis for 2-, 3-, and 5-year overall survival (OS). *OS* overall survival, *ROC* receiver operating characteristic

Table 1 Sensitivities and specificities in both training (TCGA) and validation (GEO and ICGC) cohorts

Cohort	Threshold	Sensitivity	Specificity	Time
TCGA	103	0.93836	0.91667	730
	110	0.94353	0.99535	1095
	123	0.86110	0.98684	1825
ICGC-PACA-AU	Threshold	Sensitivity	Specificity	Time
	58	0.72407	0.60000	730
	47	0.65693	0.83333	1095
GSE183795	Threshold	Sensitivity	Specificity	Time
	84	0.75736	0.60714	730
	79	0.68407	0.75000	1095
	79	0.67323	0.99160	1825
GSE62452	Threshold	Sensitivity	Specificity	Time
	39	0.69528	0.77778	730
	48	0.78803	0.72727	1095
	39	0.68181	0.97412	1825
GSE78229	Threshold	Sensitivity	Specificity	Time
	27	0.65764	0.85714	730
	40	0.85385	0.66667	1095
	27	0.62865	0.95282	1825

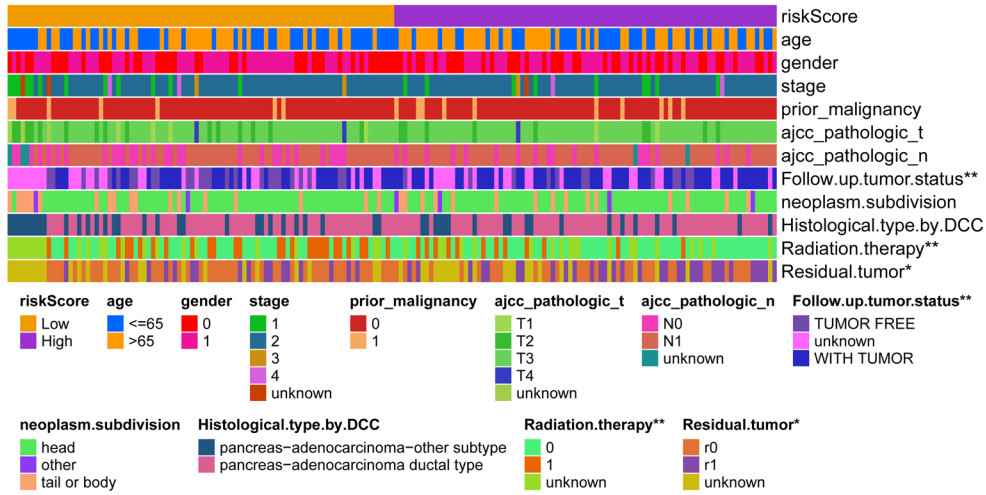
Furthermore, the p53 pathway exhibited higher activity in the high-risk group. In contrast, protein degradation pathways were significantly enriched in the low-risk group (Fig. 9c).

Simultaneously, GSVA-HALLMARK enrichment analysis demonstrated that multiple pathways associated with cancer progression and cell proliferation were significantly enriched in the high-risk group. These pathways include MYC target-related pathways, DNA repair, G2/M checkpoint-related pathways, mTORC1 signaling, hypoxia, and apoptosis-related pathways. Notably, cancer-related pathways such as the p53 pathway, Notch signaling pathway, and TGF-β signaling were more active in the high-risk group compared to the low-risk group (Fig. 9a–b), which was generally consistent with the GSVA-KEGG results. GSEA further indicated significant enrichment of pathways related to the regulation of cell cycle checkpoints, mitosis, and keratinocyte envelope formation (Fig. 9e).

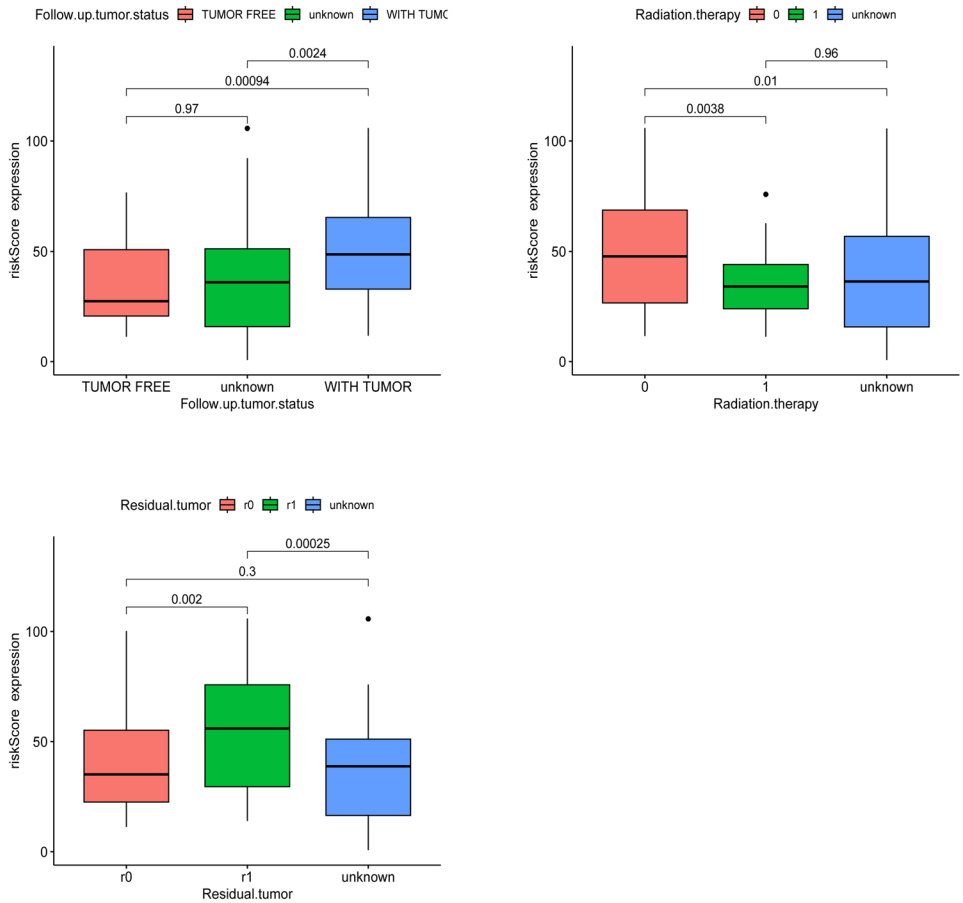
To further explore the differential gene expression between the CARGs and the risk scores in PAAD patients, we identified *DEGs* (DEGs 2) between the high-risk and low-risk groups and performed GO-KEGG enrichment analysis using the "org.Hs.eg.db" and "clusterProfiler" packages. The results revealed that these *DEGs* were enriched in various cellular components, including neuronal cell bodies, synaptic membranes, transmembrane transporter complexes, and ion channel complexes related to neuronal cell membrane structures and channel complexes. In biological processes, these genes were primarily enriched in processes such as the regulation of chemical synaptic transmission, membrane potential regulation, and potassium ion transport. In terms of molecular functions, significant enrichment was observed in potassium and metal ion transmembrane transporter activities, voltage-gated channel activities, and ion channel activities (Fig. 9d). Additionally, KEGG pathway analysis indicated enrichment in neuroactive ligand-receptor interactions and dopaminergic synapse-related pathways (Fig. 9f), suggesting that neuroregulation and metabolic abnormalities may play crucial roles in the initiation and progression of PAAD.

We also analyzed the genomic mutation characteristics associated with the CARGs gene set, including mutation counts and mutation spectra. Utilizing the **maftools** package, we visualized the mutation patterns of 168 patients from the TCGA-PAAD cohort. The results indicated that the overall mutation rate was higher in the high-risk group compared to the low-risk group, suggesting greater genomic instability in the high-risk group. *TP53* and *KRAS* were the most frequently mutated genes in the high-risk group, with mutation rates significantly higher than those in the low-risk group (*TP53*: 70% vs. 51%, *KRAS*: 80% vs. 50%) (Fig. 10a, b). These mutations are closely associated with cancer progression and poor prognosis across various cancers, indicating their potential critical roles in disease progression among high-risk patients. Additionally, *CDKN2A* and *SMAD4* exhibited relatively high mutation frequencies in both groups, with slightly higher rates in the high-risk group, further supporting their importance in tumor suppression and signal transduction. Furthermore, *TTN* also displayed moderately high mutation frequencies in both groups, warranting further investigation.

Fig. 6 **a** A complex heatmap depicting clinical characteristics in high- and low-risk groups. **b** Clinical features associated with the risk score, including tumor status at follow-up tumor status, radiation therapy, and residual tumor



(a)



(b)

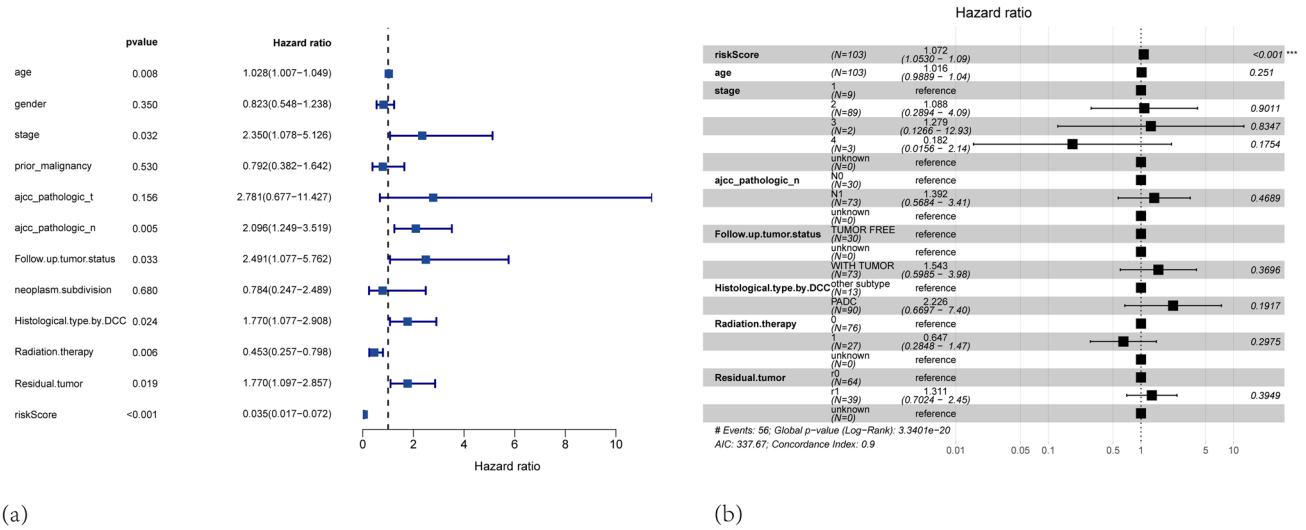


Fig. 7 Univariate (**a**) and multivariate (**b**) regression analysis of the risk score in the TCGA-PAAD cohort; Cox proportional hazards regression

Regarding TMB, the proportion of high-TMB samples was higher in the low-risk group compared to the high-risk group, and this difference was statistically significant based on the chi-square test (Fig. 10c–d, $p = 0.0032$). This suggests that the high-risk group may be associated with poorer responses to immunotherapy.

3.6 Correlation analysis of immune infiltration and tumor microenvironment

To explore the relationships among Ribosome Biogenesis, immune infiltration, and the tumor microenvironment, we first conducted ssGSEA to compare immune cell infiltration levels between the high-risk and low-risk groups. The results indicated that in the high-risk group, the infiltration levels of $CD4^+$ T cells and Type II helper T cells were significantly higher than those in the low-risk group ($P < 0.05$). Conversely, the low-risk group exhibited significantly higher infiltration levels of activated B cells and eosinophils ($P < 0.01$), as well as higher infiltration levels of monocytes and macrophages ($P < 0.05$) (Fig. 11a).

In terms of immune functions, the high-risk group showed elevated levels of antigen-presenting cell co-inhibition, immunosuppressive inflammation, and Type I interferon responses compared to the low-risk group ($P < 0.05$) (Fig. 11b). Using the **ESTIMATE** algorithm, we examined the relationship between Ribosome Biogenesis and the tumor microenvironment in 177 PAAD patients from TCGA. ESTIMATE analysis revealed that the high-risk group had higher tumor purity compared to the low-risk group (Fig. 11f), while the ESTIMATE score and immune score were lower in the high-risk group, with no significant difference in the stromal score between the two groups (Fig. 11e).

In immune checkpoint analysis, CD44, LGALS9, CD160, IDO2, and TNFSF9 were significantly more highly expressed in the high-risk group (Fig. 11c), suggesting that immune checkpoint inhibitors (ICIs) targeting these molecules may offer potential therapeutic benefits. TIDE algorithm, a computational tool for predicting responses to immune checkpoint blockade (ICB), was employed to assess the potential clinical responses to ICB in different risk groups. The results indicated that the relationship between patient immune treatment responses and their risk scores did not reach statistical significance, suggesting that patients in the high- and low-risk groups may not differ significantly in their responses to conventional ICB treatments (Fig. 11f).

Subsequently, using the **CIBERSORT** algorithm, we performed deconvolution analysis of the tumor immune microenvironment and assessed the correlation between infiltration levels and risk scores. The results demonstrated that activated mast cells and activated B cells were negatively correlated with risk scores, while M1 macrophages and activated dendritic cells were positively correlated with risk scores, with all correlations being statistically significant ($P < 0.05$) (Fig. 12a–c). Spearman correlation analysis revealed that among the nine CARGs, the expression levels of *ECT2* were strongly correlated with the infiltration levels of Type II helper T cells, *ERBB3* and *PRSS3* with the infiltration levels of Type 17 helper T cells, *SLC2A1* with $CD56^+$dim natural killer (NK) cells, and *TPX2* with activated $CD4^+$ T cells (Fig. 11d).

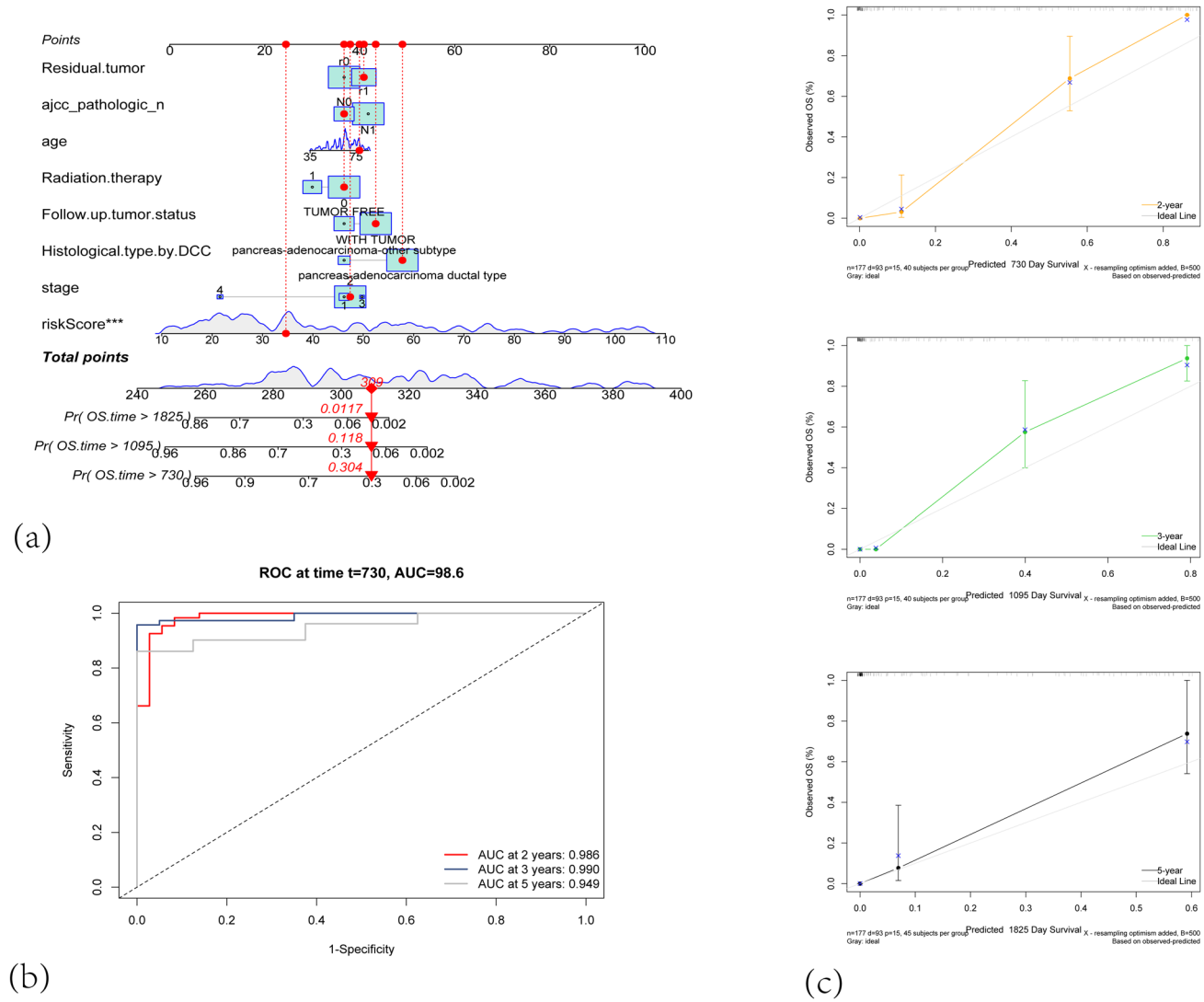


Fig. 8 **a** This nomogram estimates 2-year, 3-year, and 5-year overall survival probabilities for patients with PAAD. The included predictors are risk score, age, stage, AJCC pathological N, follow-up tumor status, WHO histological classification, radiation therapy, and residual tumor status. **b** AUC values of the nomogram for predicting 2-year, 3-year, and 5-year survival. **c** Calibration curves showing the agreement between the nomogram's predicted probabilities and the actual observed outcomes at 2-year, 3-year, and 5-year intervals. The x-axis represents predicted survival probability, and the y-axis represents observed survival probability

3.7 Drug sensitivity prediction and CARGs risk score

We utilized cell line data from the Cancer Genome Project (CGP) database to predict the IC50 of commonly used chemotherapy drugs for different risk groups of PAAD patients. Gemcitabine is a first-line treatment for pancreatic cancer and can be used alone or in combination for PAAD chemotherapy [16]. According to our study, patients in the low-risk group exhibited significantly higher sensitivity to gemcitabine compared to those in the high-risk group ($P < 0.001$), with a correlation coefficient of 0.4 between risk scores and gemcitabine IC50 (Fig. 13a). This suggests that the median inhibitory concentration (IC50) for gemcitabine is significantly lower in the low-risk group, indicating that low-risk patients may be more sensitive to gemcitabine, while high-risk patients may be more prone to developing resistance.

Additionally, the IC50 values for other first-line chemotherapy drugs, such as platinum-based drugs (oxaliplatin, cisplatin) and topoisomerase inhibitors (irinotecan), were also significantly lower in the low-risk group ($P < 0.001$, Fig. 13b–d). Notably, high-risk patients exhibited significantly lower IC50 values for paclitaxel, erlotinib, Laptinib, and

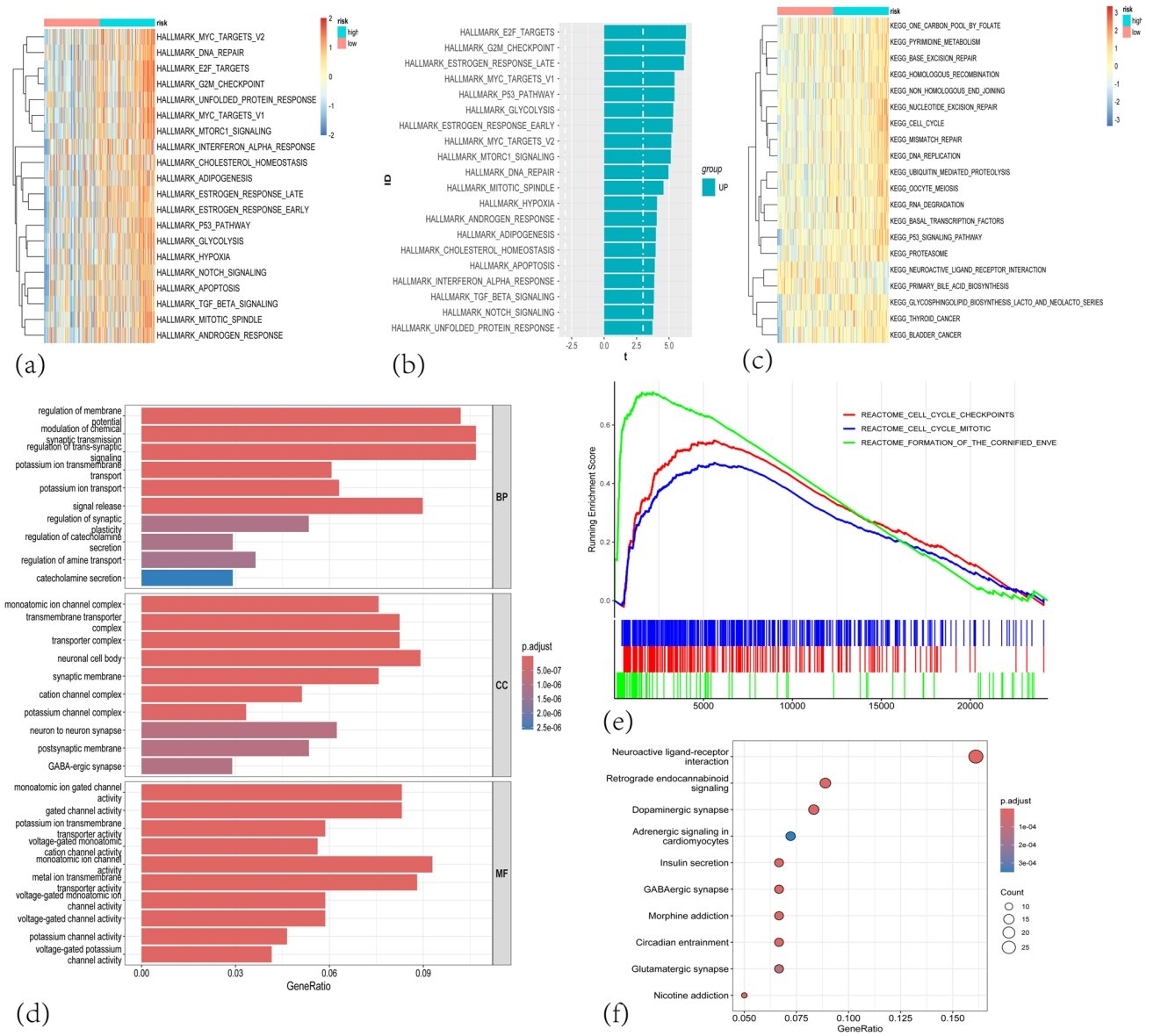


Fig. 9 Enrichment analysis. **a** The top 20 significantly enriched GSVA-HALLMARK pathways in both high- and low-risk groups. **b** The top 20 GSVA-HALLMARK pathways showing increased enrichment in the high-risk group. **c** The top 20 significantly enriched GSVA pathways in the high- and low-risk groups. **d** The most significant and commonly enriched GO terms in the TCGA dataset. **e** The important pathway of GSEA analysis. **f** The most significant and commonly enriched KEGG pathways in the TCGA dataset

Acetalax compared to the low-risk group (Figs. 13e–f and S4a–b), suggesting that these drugs may offer potential therapeutic benefits for patients in the high-risk category.

3.8 Analysis of gene methylation levels and miRNA expression

We conducted differential analysis of gene methylation levels and miRNA expression between the high-risk and low-risk groups. The results indicated that there were no significant differences in the methylation levels of corresponding genes between the high-risk and low-risk groups (Figure S5a). Furthermore, there was no apparent correlation between the methylation levels of the nine key prognostic CARGs and their mRNA expression levels (Figure S5b). This suggests that these genes may not be significantly associated with methylation levels.

Differential analysis of miRNA expression between the high-risk and low-risk groups revealed that miR-203a-3p, hsa-miR-196b-5p, and hsa-miR-205-5p were highly expressed in the high-risk group (Figure S5c). However, the correlation

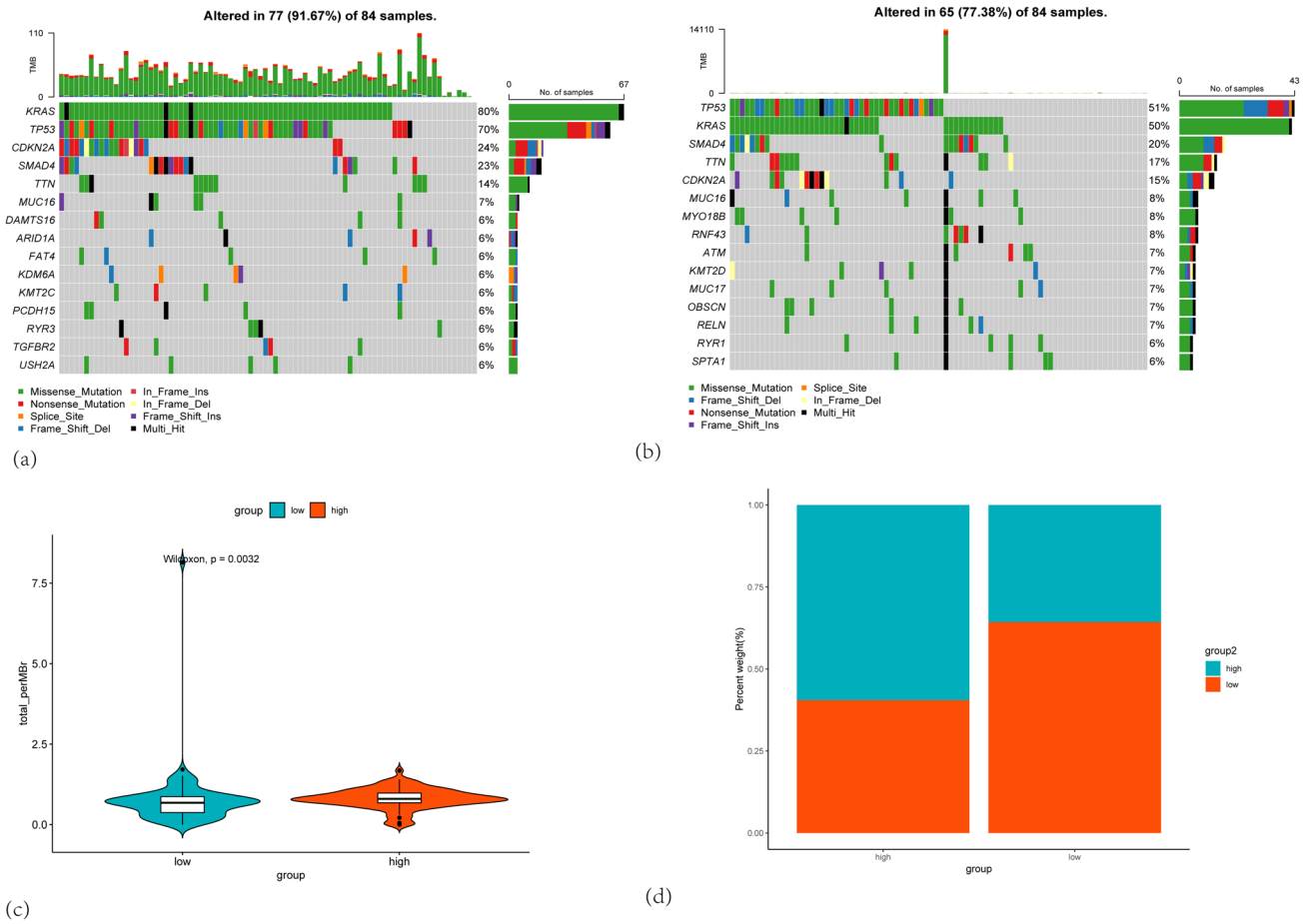


Fig. 10 **a** The top 20 genes with the highest mutation frequencies in the high-risk group. **b** The top 20 genes with the highest mutation frequencies in the low-risk group. **c** A bar chart illustrating the distribution of tumor mutation burden (TMB) across the high- and low-risk groups. **d** TMB distribution differences between the two groups (Wilcoxon test)

between the mRNA expression levels of the nine genes and miRNA expression levels did not reach a correlation coefficient of 0.7 (Figure S5d).

3.9 Validation of ribosome biogenesis-related genes at the single-cell level and protein expression analysis

To further elucidate the specific roles of CARGs in PAAD progression, we performed single-cell RNA sequencing analysis on 10 samples (GSE155698) to study the expression profiles of Biogenesis-related genes in the pancreatic malignant tumor microenvironment. After quality control filtering, we obtained 24,094 cells from 10 primary pancreatic tumors. These cells were subsequently merged, clustered, and annotated (Fig. 14a). Based on cell type-specific marker genes, the cells were classified into T cells, macrophages, endothelial cells, pancreatic ductal cells, NK cells, fibroblasts, B cells, and pancreatic cancer stem cells (Fig. 14b).

Next, we plotted the expression profiles of nine ribosome biogenesis-related genes, including *ECT2*, *CKB*, *HMGA2*, *TPX2*, *ERBB3*, *SLC2A1*, *KRT13*, *PRSS3*, and *CRABP2*. As shown in Fig. 14c, *CKB*, *SLC2A1*, *ERBB3*, and *CRABP2* were predominantly highly expressed in pancreatic ductal cells, while *PRSS3* was highly expressed in both pancreatic acinar and ductal cells. The other genes exhibited only minimal expression (Fig. 14c). To further quantify the overall expression levels of these genes across different cell types, we calculated a module score based on the nine genes. This module score represented the collective expression level of these nine genes within individual cells. The results indicated higher expression levels in ductal epithelial cells and acinar cells (Fig. 14d). These findings suggest that ribosome biogenesis may play critical roles in these cell types.

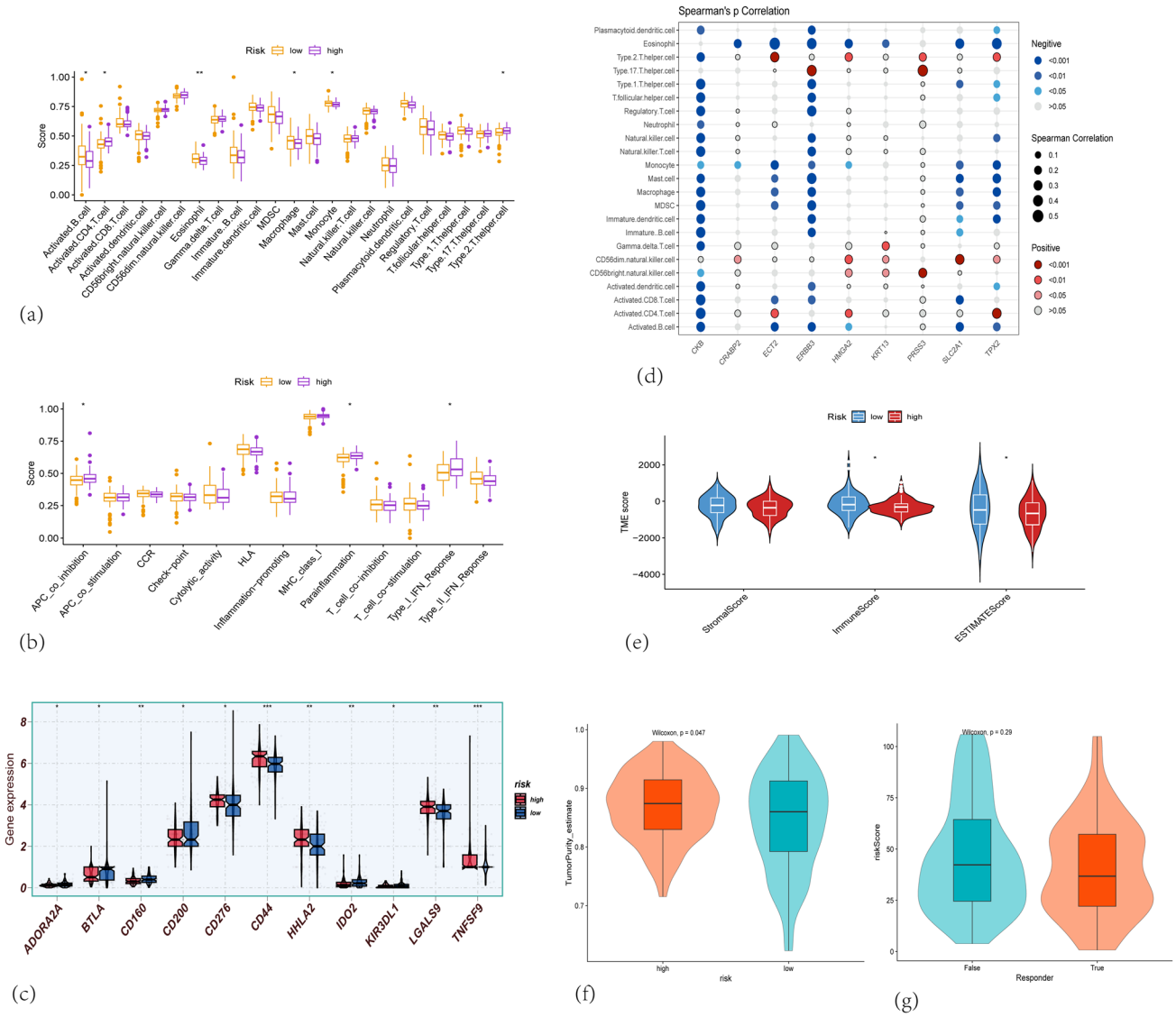


Fig. 11 TCGA tumor microenvironment between high-group and low-risk group. **a** Boxplots of the 23 immune signature ssGSEA scores comparing high-risk and low-risk patients in TCGA. **b** Boxplots of immune function ssGSEA scores comparing high-risk and low-risk patients in TCGA. **c** Differential expression of ADORA2A, BTLA, CD160, CD200, CD276, CD44, HHLA2, IDO2, KIR3DL1, LGALS9, and TNFSF9 between high- and low-risk groups in TCGA. Statistical significance is indicated as follows: *p < 0.05, **p < 0.01, ***p < 0.001, ns = not significant. **d** Spearman's correlation coefficients between the 9 genes and infiltration levels of 23 immune cell types. **e** Differences in stromal score, ESTIMATE score, and immune score between the high-risk (n = 88) and low-risk (n = 89) groups in TCGA PAAD patients. **f** Differences in tumor purity between the high-risk (n = 88) and low-risk (n = 89) groups in TCGA PAAD patients. **g** Differences in TIDE scores between the high-risk (n = 88) and low-risk (n = 89) groups

IHC data were downloaded from the HPA(http://www.proteinatlas.org) to evaluate the expression levels of eight genes (*CKB*, *HMGA2*, *TPX2*, ERBB3, *SLC2A1*, *KRT13*, *PRSS3*, and *CRABP2*) in pancreatic cancer tissues (Figs. 15, 16). IHC analysis demonstrated varying degrees of protein expression, ranging from low to high, in representative PAAD samples.

3.10 Cell communication

The intercellular signaling network diagram (Fig. 17a) reveals complex signal transduction relationships among various cell types, including macrophages, endothelial cells, pancreatic ductal cells, pancreatic acinar cells, and fibroblasts. In this network, node size represents different cell types, while the thickness of the connections indicates the strength of the signaling interactions. Heatmaps illustrating signaling patterns across cell types (Fig. 17b and c) demonstrate that

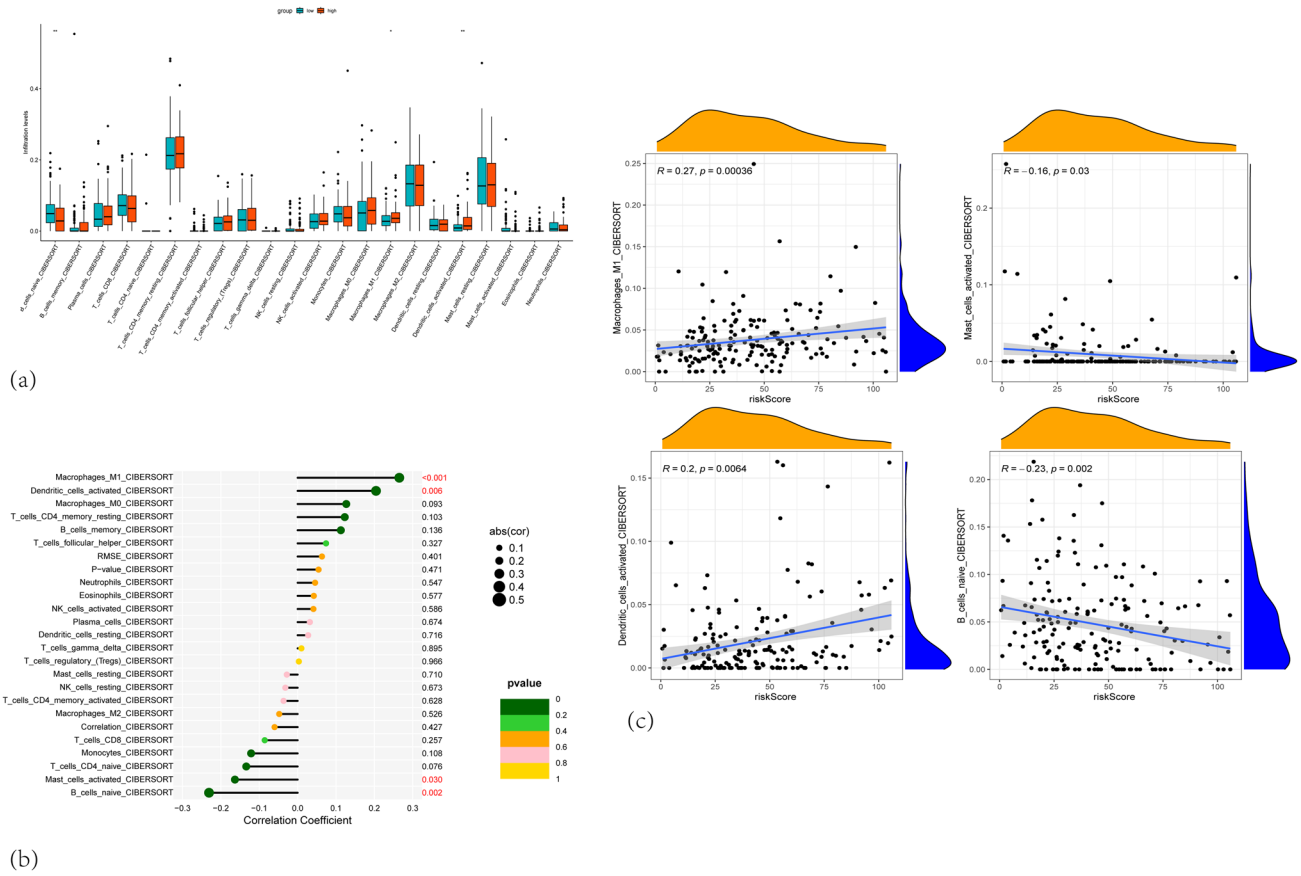


Fig. 12 **a** Boxplots depicting the CIBERSORT scores of 22 immune cells of the high-risk patients compared to low-risk patients in TCGA. **b** The Correlation Coeffcient of 22 immune cells of the high-risk patients with risk score. **c** Correlation of B_cells_naive、Dendritic_cells_activated、Macrophages_M1 and Mast_cells_activated with risk score

high-scoring ductal epithelial cells exhibit strong outgoing signals in pathways such as MIF, MK, and PDGF. Conversely, macrophages show robust incoming signals in pathways including MIF, SPP1, and CSF. These findings highlight the diversity of intercellular communication. Collectively, these results elucidate the complexity of the intercellular communication network and the specificity and intensity of signaling pathways between distinct cell types. The heatmap depicting communication probability and strength between different cell types (Fig. 17d) indicates that high-scoring ductal epithelial cells engage in strong ligand-receptor interactions with B cells and macrophages. This visualization underscores the key communications among various cell types within the tumor microenvironment through diverse signaling pathways. Furthermore, we conducted an analysis of the MIF signaling pathway network (Fig. 17e and f), where nodes represent different cell types and the thickness of the connections reflects the intensity of MIF signal transduction. This analysis emphasizes the complex and critical role of MIF signaling in mediating intercellular communication.

4 Discussion

Using the model, we identified nine genes—*ECT2, CKB, HMGA2, TPX2, ERBB3, SLC2A1, KRT13, PRSS3*, and *CRABP2*—that are strongly linked to the clinical traits and chemotherapy resistance of malignant tumors. Research indicates that these genes demonstrate the following characteristics:

ECT2, a RhoGTPase activator essential for cytokinesis, has been identified as an oncogene with exhibiting distinct oncogenic mechanisms at different cellular locations [17]. In tumors, *CKB* may support rapid proliferation and the protein synthesis demands of ribosome biogenesis by maintaining high levels of energy metabolism [18]. Research indicates that *HMGA2* is involved in maintaining epithelial-mesenchymal transition (EMT) in human pancreatic cancer cells and it is is a potential therapeutic target for pancreatic cancer treatment [19]. *TPX2*, a critical regulator of cell

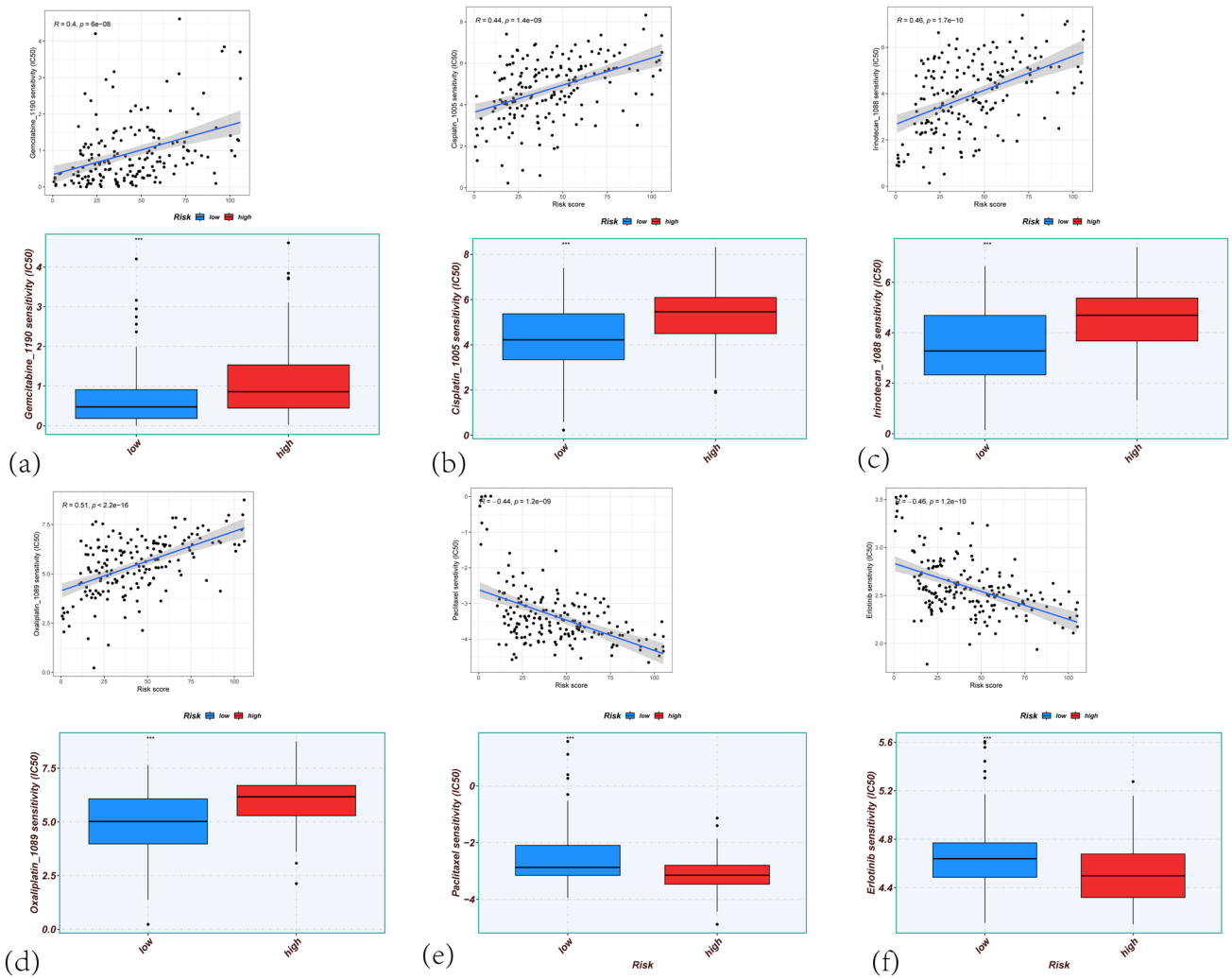


Fig. 13 **a** Correlation analysis of the Risk Score with **Gemcitabine** IC50 and Boxplots of the estimated IC50 of **Gemcitabine** in the high- (n = 88) and low-risk (n = 89) group. **b** Correlation analysis of the Risk Score with **Cisplatin** IC50 and Boxplots of the estimated IC50 of **Cisplatin** in the high- (n = 88) and low-risk (n = 89) group. **c** Correlation analysis of the Risk Score with **Irinotecan** IC50 and Boxplots of the estimated IC50 of **Irinotecan** in the high- (n = 88) and low-risk (n = 89) group. **d** Correlation analysis of the Risk Score with **Oxaliplatin** IC50 and Boxplots of the estimated IC50 of **Oxaliplatin** in the high- (n = 88) and low-risk (n = 89) group. **e** Correlation analysis of the Risk Score with **Paclitaxel** IC50 and Boxplots of the estimated IC50 of **Paclitaxel** in the high- (n = 88) and low-risk (n = 89) group. **f** Correlation analysis of the Risk Score with **Erlotinib** IC50 and Boxplots of the estimated IC50 of **Erlotinib** in the high- (n = 88) and low-risk (n = 89) group

division and a potential oncogenic driver, emerges as a promising therapeutic target for pancreatic cancer [20, 21]. *ERBB3*, which localizes to the nucleolus in various cancers, has been shown to induce apoptosis in pancreatic cancer cells through its downregulation [22, 23]. *SLC2A1* has been demonstrated that promotes metabolic reprogramming and chemoresistance in PAAD [24]. *KRT13* is upregulated in pancreatic cancer stem cell-like cells and is associated with radiotherapy resistance [25]. *PRSS3* promotes tumor cell invasion and metastasis, and its overexpression has been observed in metastatic pancreatic cancer cell lines [26–29]. *CRABP2* promotes ribosome biogenesis, contributing to tumor invasiveness and metastasis [30].

Furthermore, this nine-gene signature risk model can independently predict overall survival outcomes beyond known clinical and pathological risk factors. Through our enrichment analysis that some target genes like *ECT2, ERBB3, HMGA2*, and a few others are getting pretty cozy in key pathways that are related to pancreatic cancer growth. For instance, *ECT2* and *TPX2* are kind of troublemakers when it comes to genomic stability because they mess with how the mitotic spindle is assembled. On the flip side, *ERBB3* and *PRSS3* kick off the *ERBB/MAPK* signaling pathway, which is a big player in making chemotherapy less effective. Then *HMGA2* and *CRABP2* that team up to boost EMT and metastasis through pathways like Wnt/β-catenin and retinoic acid metabolism. Over here, *SLC2A1* and *CKB* are in charge of regulating glycolysis and

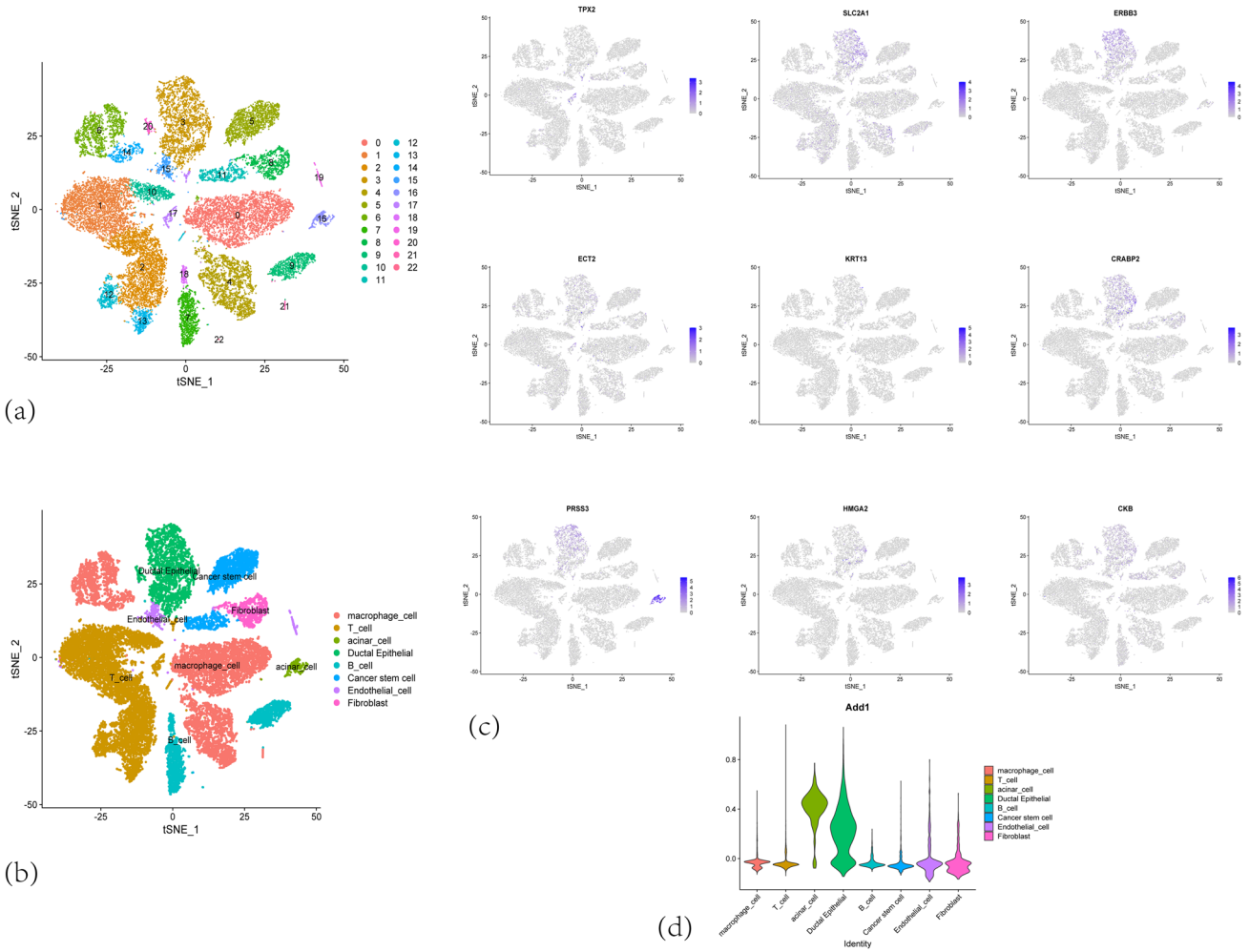


Fig. 14 Expression of nine ribosome biogenesis–related genes in a single-cell RNA sequencing dataset. **a** Cell cluster annotations. **b** Cell type annotations. **c** Expression patterns of *ECT2*, *CKB*, *HMGA2*, *TPX2*, *ERBB3*, *SLC2A1*, *KRT13*, *PRSS3*, and *CRABP2*. **d** Violin plots showing the expression levels of these ribosome biogenesis–related genes across various cell types in pancreatic cancer, as revealed by single-cell RNA sequencing analysis

phosphocreatine metabolism, which helps tumors switch up their metabolism. All this interconnected pathway action really shows how gene interactions push.

Our analysis through the integration of advanced multiple machine learning algorithms to differentiate high and low-risk groups based on risk scores and identify potential drugs significantly affecting disease prognosis. High-risk PAAD patients appear to exhibit resistance to chemotherapy drugs that interfere with DNA synthesis, such as gemcitabine, oxaliplatin, and cisplatin, as well as topoisomerase I inhibitors like irinotecan. Mechanisms underlying resistance to DNA synthesis inhibitors generally include enhanced DNA repair capabilities and altered cellular metabolic pathways. Enhanced DNA repair ability can reverse DNA damage induced by topoisomerase I inhibitors, thereby diminishing the cytotoxic effects of irinotecan [16, 31, 32]. These mechanisms may be potentially linked to our finding that high-risk groups are enriched in DNA repair pathways.

Additionally, we discovered that paclitaxel may offer benefits to high-risk patients. Paclitaxel inhibits cell mitosis by stabilizing microtubule structures, and often used in combination with gemcitabine as a first-line chemotherapy regimen for pancreatic cancer and metastatic pancreatic cancer [33]. This aligns with our enrichment results, where high-risk group patients are more enriched in pathways regulating cell cycle checkpoints and mitosis, potentially serving as potential targets for paclitaxel.

Erlotinib and Acetalax are both EGFR tyrosine kinase inhibitors (TKIs). Erlotinib, in combination with gemcitabine, remains a first-line chemotherapy option for advanced pancreatic cancer. However, compared to the FOLFIRINOX regimen, it has limitations and is suitable for specific patient populations who cannot tolerate multiple chemotherapy

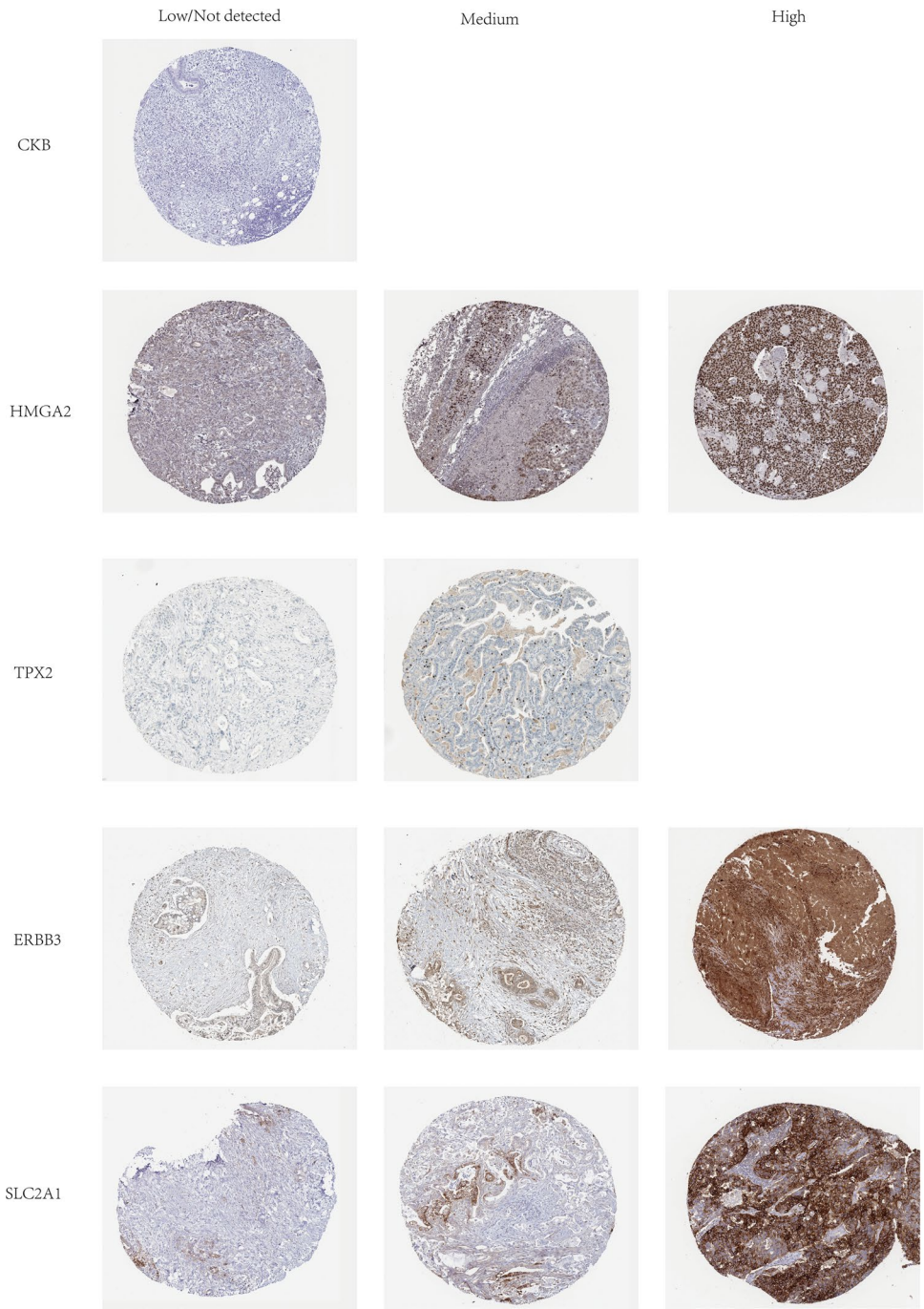


Fig. 15 Protein expression levels of ECT2, CKB, HMGA2, TPX2, ERBB3 and SLC2A1 in PAAD tumor tissue

regimens [34]. In our cohort, high-risk group patients may benefit from erlotinib. Laptinib and Acetalax have primarily been studied for their potential roles in HER2-positive and triple-negative breast cancers, respectively [35, 36]. Their effects on pancreatic cancer require further experimental and theoretical support.

Based on our study cohort's enrichment analysis results, the high-risk group is primarily enriched in MYC target-related pathways, DNA repair, G2/M checkpoint-related pathways, mTORC1 signaling, hypoxia, apoptosis-related pathways, the p53 pathway, Notch signaling pathway, and TGF-β signaling-related Hallmark pathways.

MYC is a crucial transcription factor closely associated with cell growth and proliferation. Multiple biological processes regulated by c-MYC enhance ribosome generation and function [37, 38]. mTORC1 is a key regulator of cellular metabolism, protein synthesis, and cell proliferation [39]. The Notch signaling pathway is believed to regulate tumor development and the tumor microenvironment [40]. Other pathways are closely linked to processes related to cell proliferation and cancer

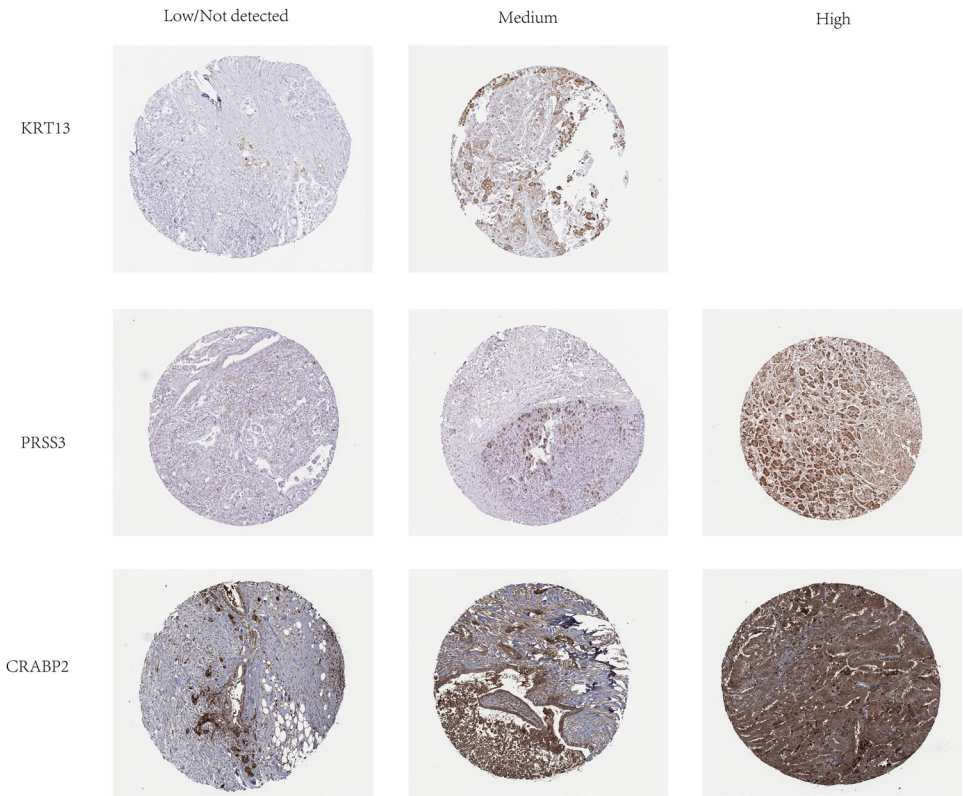


Fig. 16 Protein expression levels of KRT13, PRSS3 and CRABP2 in PAAD tumor tissues

progression. GSEA enrichment analysis also revealed a strong association with cancer cell proliferation pathways related to cell cycle and mitosis. GO-KEGG analysis showed significant enrichment in neuroactive ligand-receptor interactions, transmembrane receptors, ion channels, neurotransmitter-regulated processes, and pathways, indicating a complex interplay between high metabolic activity in tumor cells and neural signaling networks and regulation.

In the cytokine network of pancreatic cancer, immature $CD4^+$ T cells can be activated in various cytokine environments [41]. Studies have shown that in PAAD, cytokines such as TNF-α and IL-1β released by tumor cells can induce cancer-associated fibroblasts (CAFs) to secrete thymic stromal lymphopoietin (TSLP), activating resident dendritic cells (DCs). These DCs release Th2 chemokines (CCL2, CCL17), which recruit Th2 cells-subset differentiated from $CD4^+$ T cells-to the tumor region. Subsequently, Th2 cells secrete cytokines that promote M2 polarization, mediating immune evasion [42].

These findings are consistent with our study, where high-risk patients in our model cohort exhibited higher infiltration levels of $CD4^+$ T cells and Type II helper T cells. In the low-risk group, increased infiltration levels of activated B cells, eosinophils, monocytes, and macrophages are associated with enhanced immune responses and surveillance, as well as pro-inflammatory functions. The M1 subtype of macrophages possesses antitumor activities [43], suggesting that the low-risk group may predominantly harbor M1 macrophages. However, our study also found a positive correlation between M1 macrophages and risk scores, although insufficient to indicate significantly higher infiltration levels in the high-risk group. This slight increase may be related to enhanced metabolism leading to compensatory antitumor activity, warranting further investigation. These results provide a basis for further exploring the potential roles of immune cells in tumor prognosis and therapy.

Tumor immunotherapy, particularly ICB and cell therapy, remains a persistent area of interest in cancer treatment. However, its clinical efficacy is often limited by the immunosuppressive tumor microenvironment. TIDE analysis revealed no statistically significant difference in ICB treatment between high-risk and low-risk groups, due to the complex immune microenvironment of pancreatic cancer and single risk score. Additionally, limitations of sample sizes, data heterogeneity, and lack of experimental validation, may also impact the results. Therefore, future studies should integrate larger-scale clinical data and experimental validations to deeply explore the potential associations between ribosome biogenesis risk scores and ICB treatment responses. Furthermore, the expression of multiple immune checkpoint genes, including *CD44, LGALS9, CD160, IDO2*, and *TNFSF9*, was significantly in high-risk patients, suggesting ICIs targeting these molecules may offer potential therapeutic benefits. However, ICB treatment prediction

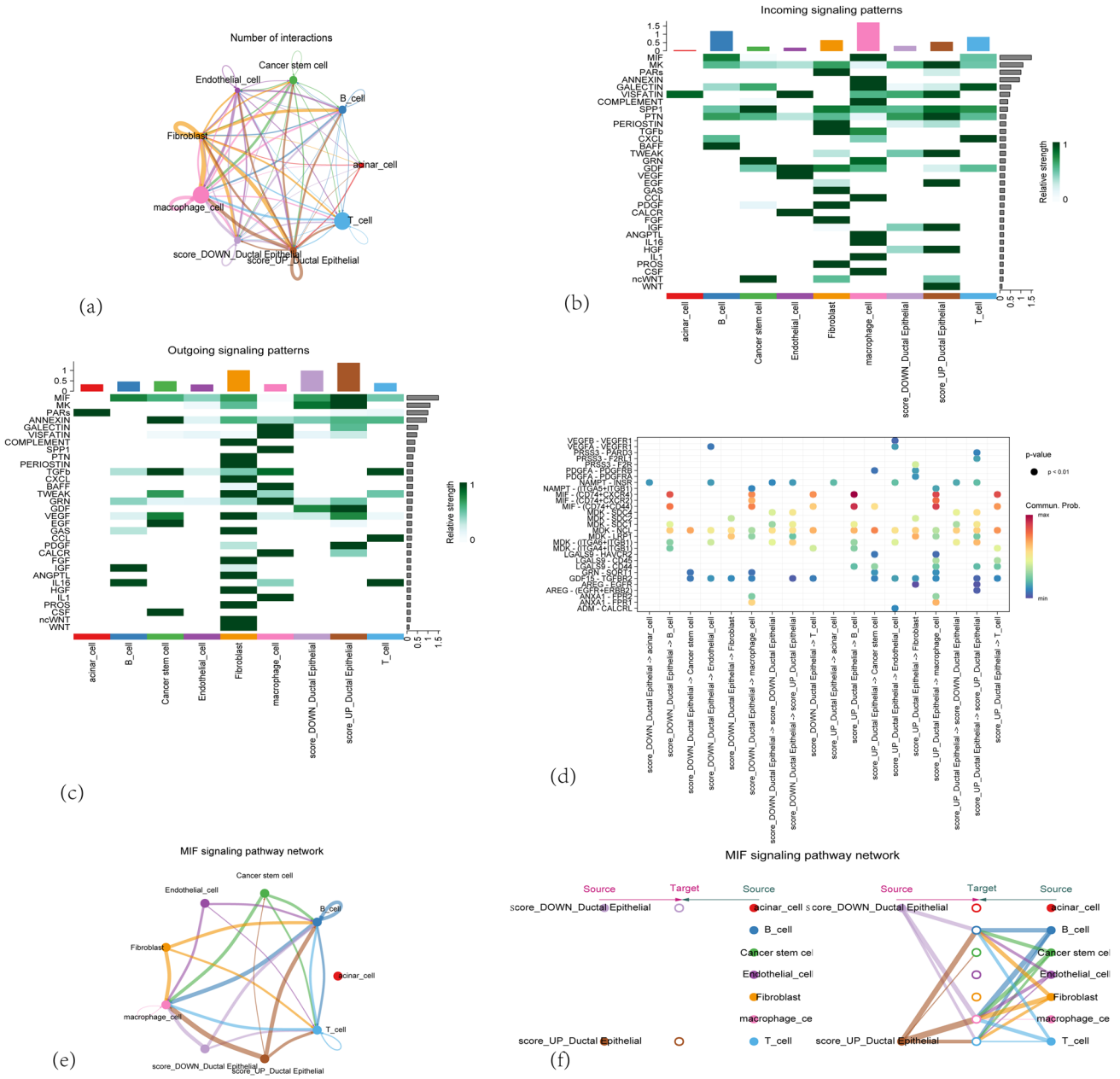


Fig. 17 Single-cell communication analysis **a** network diagram illustrating intercellular signaling pathways **b** Heatmap showing outgoing signaling patterns across different cell types. **c** Heatmap displaying the incoming signaling pathways in various cell types. **d** Bubble plot illustrating the communication probability and strength between different cell types. **e** MIF signaling pathway cell communication network diagram (circular layout). **f** MIF signaling pathway cell communication network diagram (hierarchical layout)

indicated that risk scores did not significantly differ in their association with ICB treatment responses, possibly due to the inherently "cold tumor" characteristics of pancreatic cancer, which render it less responsive to immunotherapy [44]. Immunotherapy for pancreatic malignancies still requires further exploration.

In our analysis of somatic mutations, we found that *TP53* and *KRAS* had significantly higher mutation frequencies in the high-risk group compared to the low-risk group. Both subgroups exhibited high mutation rates of *CDKN2A* and *SMAD4*, with slightly higher rates in the high-risk group. Mutations in *KRAS* and *TP53* have been extensively studied and consistently shown to have profound impacts on the survival of pancreatic cancer patients. Mutations in *CDKN2A* often lead to aberrant cell cycle regulation, while deletions or mutations in *SMAD4* are closely associated with metastatic pancreatic cancer [45]. Moreover, we also found that *TTN* exhibited a moderately high mutation frequency. *TTN*, a structural protein in striated muscle, has the longest exon length in the genome [46]. The predictive significance of

TTN mutations may be primarily driven by statistical mechanisms, but its specific functional mechanisms in certain cancers remain unclear [47].

Future experimental studies are needed to validate how *TTN* mutations affect the structural stability of pancreatic cancer cells and the specific functions of signaling pathways, thereby clarifying their clinical significance. These findings underscore the critical role of key gene mutations in the progression of pancreatic cancer and suggest that therapeutic strategies targeting these gene mutations may help improve patient prognosis and develop personalized treatment plans.

Although some miRNAs have been identified to play roles in pancreatic cancer cell proliferation, invasion, and metastasis, their study remains limited. In our study cohort, we identified that *hsa-miR-203a-3p, hsa-miR-196b-5p,* and *hsa-miR-205-5p* were highly expressed in the high-risk group based on the ribosome biogenesis risk score model, whereas these miRNAs were not highly expressed in the low-risk group.

hsa-miR-203a-3p plays a critical role in various cancers, typically acting as a tumor-suppressive miRNA. In nasopharyngeal carcinoma, it inhibits tumor growth and metastasis by targeting *LASP1* [48]. Its high expression in breast cancer and non-small cell lung cancer is associated with better prognosis [49, 50]. However, our study found that *miR-203a-3p* is highly expressed in the high-risk group of pancreatic cancer, suggesting it may have dual functions. Existing literature indicates that *miR-203a-3p* acts as a tumor suppressor in some cancers, while in others, it may promote tumor progression by enhancing EMT and targeting *PDE4D* to promote colorectal cancer cell proliferation, colony formation, and migration and invasion [51, 52]. This dual functionality may be related to its regulation of different target genes and signaling pathways in various biological environments. Therefore, future studies should further investigate the specific mechanisms of *miR-203a-3p* in pancreatic cancer to elucidate its potential therapeutic value.

miR-196b-5p in pancreatic cancer can regulate the expression of the *ING5* gene, inhibiting apoptosis and promoting cancer cell proliferation [53]. It also enhances the invasiveness and anti-apoptotic capabilities of cancer cells in lung small cell carcinoma and colorectal cancer [54, 55].

hsa-miR-205-5p exhibits dual roles in different tumors, acting both as a tumor suppressor and a tumor promoter. Its specific function depends on the type of cancer and the tissue microenvironment. Currently, there are few studies confirming *miR-205* as a biomarker for pancreatic cancer. Some studies suggest it has ambiguous and pro-tumorigenic roles [56].

Additionally, we analyzed the single-cell sequencing data of the nine genes included in our model and observed differential distribution of cells expressing these molecules in pancreatic cancer. Notably, these cells were primarily distributed in ductal epithelial cells, acinar cells, and a subset of cancer stem cells. Combined with the impact of the risk model on immune infiltration, we speculate that the "cold tumor" characteristics of pancreatic cancer cells may contribute to the formation of an immunosuppressive microenvironment, ultimately resulting in an "immune desert" phenotype. Moreover, our cell communication analysis revealed intricate interactions among cell populations. Ductal epithelial cells associated with hub genes may have complex connections with the MIF signaling pathway, warranting further investigation.

Despite identifying ribosome biogenesis-related risk score groups and prognostic genes in PAAD for the first time, our study has several limitations. First, the number of cohorts and sequencing data used was limited. Second, clinical and pathological feature information of patients was insufficient, necessitating more practical and valuable factors to predict 1- to 5-year survival rates. Lastly, there was a lack of basic experiments to validate the expression of these prognostic genes in pancreatic cancer cell lines, which requires further investigation.

In summary, we identified key prognostic genes associated with ribosome biogenesis and constructed a prognostic risk model for PAAD. The results revealed nine prognostic genes that are highly associated with tumorigenesis and immune cell infiltration. This prognostic risk model demonstrated robust and effective performance in predicting overall survival outcomes for PAAD patients. These findings challenge the traditional perception of ribosomes as mere passive protein factories, redefining them as dynamic hubs that integrate metabolism, immunity, and genomic instability. Overall, our findings hold significant implications for in-depth studies of the molecular pathways and mechanisms underlying PAAD, as well as for the development of targeted therapeutic and prognostic strategies for pancreatic cancer.

5 Limitations

Although this study integrated multiple public databases and employed a series of bioinformatics and machine learning analyses, there are still several limitations. First, all the data used were derived from publicly available databases, potentially leading to issues with data completeness and heterogeneity. Secondly, this study primarily relies on bioinformatics

analyses, and further in vitro and in vivo experimental validation of the molecular biological mechanisms of these prognostic genes in pancreatic cancer cell lines is still needed. Finally, given the significant heterogeneity of pancreatic cancer, further validation in multiple centers with diverse patient populations is required to assess the generalizability and clinical applicability of these findings.

Acknowledgements We would like to thank TCGA, ICGC and GEO databases for the availability of the data

Author contributions Y.S. designed the study concept, performed data analyses, wrote the main manuscript text, and prepared Figs. 1, 2, 3, 4, 5, 6 ,7, 8, 9, 10, 11 and 12. Y.L. wrote the main manuscript text and performed data analyses. A.Z. and T.H. verified the data analysis results and prepared Figs. 13, 14, 15, 16, 17. M.L. provided conceptual guidance and conceived and supervised the study. All authors reviewed the manuscript.

Funding None.

Data availability The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

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. Klein AP. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. Nat Rev Gastroenterol Hepatol. 2021;18(7):493–502. https://doi.org/10.1038/s41575-021-00457-x.
2. Mukund A, Afridi MA, Karolak A, Park MA, Permuth JB, Rasool G. Pancreatic ductal adenocarcinoma (PDAC): a review of recent advancements enabled by artificial intelligence. Cancers (Basel). 2024;16(12):2240. https://doi.org/10.3390/cancers16122240. (**Published 2024 Jun 17**).
3. Principe DR, Underwood PW, Korc M, Trevino JG, Munshi HG, Rana A. The current treatment paradigm for pancreatic ductal adenocarcinoma and barriers to therapeutic efficacy. Front Oncol. 2021;11:688377. https://doi.org/10.3389/fonc.2021.688377. (**Published 2021 Jul 15**).
4. Miller PN, Romero-Hernandez F, Calthorpe L, et al. Long-duration neoadjuvant therapy with FOLFIRINOX yields favorable outcomes for patients who undergo surgery for pancreatic cancer. Ann Surg Oncol. 2024;31(9):6147–56. https://doi.org/10.1245/s10434-024-15579-0.
5 Zeng S, Pöttler M, Lan B, Grützmann R, Pilarsky C, Yang H. Chemoresistance in pancreatic cancer. Int J Mol Sci. 2019;20(18):4504. https://doi.org/10.3390/ijms20184504. (**Published 2019 Sep 11**).
6. Melnikov S, Manakongtreecheep K, Söll D. Revising the structural diversity of ribosomal proteins across the three domains of life. Mol Biol Evol. 2018;35(7):1588–98. https://doi.org/10.1093/molbev/msy021.
7. Kang J, Brajanovski N, Chan KT, Xuan J, Pearson RB, Sanij E. Ribosomal proteins and human diseases: molecular mechanisms and targeted therapy. Signal Transduct Target Ther. 2021;6(1):323 (**Published 2021 Aug 30**).
8. Narla A, Ebert BL. Ribosomopathies: human disorders of ribosome dysfunction. Blood. 2010;115(16):3196–205. https://doi.org/10.1143/blood-2009-10-178129.
9. Steinbusch MMF, Caron MMJ, Surtel DAM, et al. Expression of RMRP RNA is regulated in chondrocyte hypertrophy and determines chondrogenic differentiation. Sci Rep. 2017;7(1):6440. https://doi.org/10.1038/s41598-017-06809-5. (**Published 2017 Jul 25**).
10. Bywater MJ, Poortinga G, Sanij E, et al. Inhibition of RNA polymerase I as a therapeutic strategy to promote cancer-specific activation of p53. Cancer Cell. 2012;22(1):51–65. https://doi.org/10.1016/j.ccr.2012.05.019.
11. Pylayeva-Gupta Y, Grabocka E, Bar-Sagi D. RAS oncogenes: weaving a tumorigenic web. Nat Rev Cancer. 2011;11(11):761–74. https://doi.org/10.1038/nrc3106. (**Published 2011 Oct 13**).
12. Azman MS, Alard EL, Dodel M, et al. An ERK1/2-driven RNA-binding switch in nucleolin drives ribosome biogenesis and pancreatic tumorigenesis downstream of RAS oncogene. EMBO J. 2023;42(11):e110902. https://doi.org/10.15252/embj.2022110902.

13. Zhang Z, Yu H, Yao W, et al. RRP9 promotes gemcitabine resistance in pancreatic cancer via activating AKT signaling pathway. Cell Commun Signal. 2022;20(1):188. https://doi.org/10.1186/s12964-022-00974-5. (**Published 2022 Nov 24**).
14. Rowell MC, Deschênes-Simard X, Lopes-Paciencia S, et al. Targeting ribosome biogenesis reinforces ERK-dependent senescence in pancreatic cancer. Cell Cycle. 2023;22(19):2172–93. https://doi.org/10.1080/15384101.2023.2278945.
15. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. https://doi.org/10.1186/1471-2105-9-559. (**Published 2008 Dec 29**).
16. Stoop TF, Theijse RT, Seelen LWF, et al. Preoperative chemotherapy, radiotherapy and surgical decision-making in patients with borderline resectable and locally advanced pancreatic cancer. Nat Rev Gastroenterol Hepatol. 2024;21(2):101–24. https://doi.org/10.1038/s41575-023-00856-2.
17. Norris K, Hopes T, Aspden JL. Ribosome heterogeneity and specialization in development. Wiley Interdiscip Rev RNA. 2021;12(4):e1644. https://doi.org/10.1002/wrna.1644.
18. Wang Z, Hulsurkar M, Zhuo L, et al. CKB inhibits epithelial-mesenchymal transition and prostate cancer progression by sequestering and inhibiting AKT activation. Neoplasia. 2021;23(11):1147–65. https://doi.org/10.1016/j.neo.2021.09.005.
19. Watanabe S, Ueda Y, Akaboshi S, Hino Y, Sekita Y, Nakao M. HMGA2 maintains oncogenic RAS-induced epithelial-mesenchymal transition in human pancreatic cancer cells. Am J Pathol. 2009;174(3):854–68. https://doi.org/10.2353/ajpath.2009.080523.
20. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat Genet. 2006;38(9):1043–8. https://doi.org/10.1038/ng1861.
21. Warner SL, Stephens BJ, Nwokenkwo S, et al. Validation of TPX2 as a potential therapeutic target in pancreatic cancer cells. Clin Cancer Res. 2009;15(21):6519–28. https://doi.org/10.1158/1078-0432.CCR-09-0077.
22. Andrique L, Fauvin D, El Maassarani M, Colasson H, Vannier B, Séité P. ErbB3(80 kDa), a nuclear variant of the ErbB3 receptor, binds to the Cyclin D1 promoter to activate cell proliferation but is negatively controlled by p14ARF. Cell Signal. 2012;24(5):1074–85. https://doi.org/10.1016/j.cellsig.2012.01.002.
23. Lin T, Ren Q, Zuo W, et al. Valproic acid exhibits anti-tumor activity selectively against EGFR/ErbB2/ErbB3-coexpressing pancreatic cancer via induction of ErbB family members-targeting microRNAs. J Exp Clin Cancer Res. 2019;38(1):150. https://doi.org/10.1186/s13046-019-1160-9. (**Published 2019 Apr 8**).
24. Li Y, Tang S, Shi X, et al. Metabolic classification suggests the GLUT1/ALDOB/G6PD axis as a therapeutic target in chemotherapy-resistant pancreatic cancer. Cell Rep Med. 2023;4(9):101162. https://doi.org/10.1016/j.xcrm.2023.101162.
25. Takenaka W, Yokoyama Y, Ikehata K, et al. KRT13 is upregulated in pancreatic cancer stem-like cells and associated with radioresistance. J Radiat Res. 2023;64(2):284–93. https://doi.org/10.1093/jrr/rrac091.
26. Lin B, Zhou X, Lin S, et al. Epigenetic silencing of PRSS3 provides growth and metastasis advantage for human hepatocellular carcinoma. J Mol Med (Berl). 2017;95(11):1237–49. https://doi.org/10.1007/s00109-017-1578-5.
27. Hockla A, Miller E, Salameh MA, Copland JA, Radisky DC, Radisky ES. PRSS3/mesotrypsin is a therapeutic target for metastatic prostate cancer. Mol Cancer Res. 2012;10(12):1555–66. https://doi.org/10.1158/1541-7786.MCR-12-0314.
28. Jiang G, Cao F, Ren G, et al. PRSS3 promotes tumour growth and metastasis of human pancreatic cancer. Gut. 2010;59(11):1535–44. https://doi.org/10.1136/gut.2009.200105.
29. Nyberg P, Ylipalosaari M, Sorsa T, Salo T. Trypsins and their role in carcinoma growth. Exp Cell Res. 2006;312(8):1219–28. https://doi.org/10.1016/j.yexcr.2005.12.024.
30. Liu J, Tang L, Chu W, Wei L. Cellular retinoic acid binding protein 2 (CRABP2), up-regulated by HPV E6/E7, leads to aberrant activation of the integrin β1/FAK/ERK signaling pathway and aggravates the malignant phenotypes of cervical cancer. Biochem Genet. 2024;62(4):2686–701. https://doi.org/10.1007/s10528-023-10568-6.
31. Maede Y, Shimizu H, Fukushima T, et al. Differential and common DNA repair pathways for topoisomerase I- and II-targeted drugs in a genetic DT40 repair cell screen panel. Mol Cancer Ther. 2014;13(1):214–20. https://doi.org/10.1158/1535-7163.MCT-13-0551.
32. Alagoz M, Gilbert DC, El-Khamisy S, Chalmers AJ. DNA repair and resistance to topoisomerase I inhibitors: mechanisms, biomarkers and therapeutic targets. Curr Med Chem. 2012;19(23):3874–85. https://doi.org/10.2174/092986012802002590.
33. Zhao S, Tang Y, Wang R, Najafi M. Mechanisms of cancer cell death induction by paclitaxel: an updated review. Apoptosis. 2022;27(9–10):647–60. https://doi.org/10.1007/s10495-022-01750-z.
34. Wang Y, Hu GF, Zhang QQ, et al. Efficacy and safety of gemcitabine plus erlotinib for locally advanced or metastatic pancreatic cancer: a systematic review and meta-analysis. Drug Des Devel Ther. 2016;10:1961–72. https://doi.org/10.2147/DDDT.S105442. (**Published 2016 Jun 13**).
35. Suppan C, Balic M. Current standards and future outlooks in metastatic Her2-positive breast cancer. Breast Care (Basel). 2023;18(1):69–75. https://doi.org/10.1159/000528756.
36. Mizunuma M, Redon CE, Saha LK, et al. Acetalax (oxyphenisatin acetate, NSC 59687) and bisacodyl cause oncosis in triple-negative breast cancer cell lines by poisoning the ion exchange membrane protein TRPM4. Cancer Res Commun. 2024;4(8):2101–11. https://doi.org/10.1158/2760-9764.CRC-24-0093.
37. Jha RK, Kouzine F, Levens D. MYC function and regulation in physiological perspective. Front Cell Dev Biol. 2023;11:1268275. https://doi.org/10.3389/fcell.2023.1268275. (**Published 2023 Oct 24**).
38. Acehan D, Petzold C, Gumper I, et al. Plakoglobin is required for effective intermediate filament anchorage to desmosomes. J Invest Dermatol. 2008;128(11):2665–2605. https://doi.org/10.1038/jid.2008.141.
39. Yecies JL, Manning BD. mTOR links oncogenic signaling to tumor cell metabolism. J Mol Med (Berl). 2011;89(3):221–8. https://doi.org/10.1007/s00109-011-0726-6.
40. Meurette O, Mehlen P. Notch signaling in the tumor microenvironment. Cancer Cell. 2018;34(4):536–48. https://doi.org/10.1016/j.ccell.2018.07.009.
41. Qiu Y, Chen T, Hu R, et al. Next frontier in tumor immunotherapy: macrophage-mediated immune evasion. Biomark Res. 2021;9(1):72. https://doi.org/10.1186/s40364-021-00327-3. (**Published 2021 Oct 9**).

42. Halstead SB, Mahalingam S, Marovich MA, Ubol S, Mosser DM. Intrinsic antibody-dependent enhancement of microbial infection in macrophages: disease regulation by immune complexes. Lancet Infect Dis. 2010;10(10):712–22. https://doi.org/10.1016/S1473-3099(10)70166-3.
43. Liu J, Geng X, Hou J, Wu G. New insights into M1/M2 macrophages: key modulators in cancer progression. Cancer Cell Int. 2021;21(1):389. https://doi.org/10.1186/s12935-021-02089-2.
44. Barriga FM, Tsanov KM, Ho YJ, et al. MACHETE identifies interferon-encompassing chromosome 9p213 deletions as mediators of immune evasion and metastasis. Nat Cancer. 2022;3(11):1360–85. https://doi.org/10.1038/s43018-022-00443-5.
45. Hu C, Hart SN, Polley EC, et al. Association between inherited germline mutations in cancer predisposition genes and risk of pancreatic cancer. JAMA. 2018;319(23):2401–9. https://doi.org/10.1001/jama.2018.6228.
46. Ciferri A, Crumbliss AL. The assembling and contraction mechanisms of striated muscles. Front Chem. 2018;6:570. https://doi.org/10.3389/fchem.2018.00570. (**Published 2018 Nov 30**).
47. Jia Q, Wang J, He N, He J, Zhu B. Titin mutation associated with responsiveness to checkpoint blockades in solid tumors. JCI Insight. 2019;4(10):127901. https://doi.org/10.1172/jci.insight.127901. (**Published 2019 May 16**).
48. Jiang N, Jiang X, Chen Z, et al. MiR-203a-3p suppresses cell proliferation and metastasis through inhibiting LASP1 in nasopharyngeal carcinoma. J Exp Clin Cancer Res. 2017;36(1):138. https://doi.org/10.1186/s13046-017-0604-3. (**Published 2017 Oct 5**).
49. Wang L, Wang G, Song J, Yao D, Wang Y, Chen T. A comprehensive analysis of the prognostic characteristics of microRNAs in breast cancer. Front Genet. 2024;15:1293824. https://doi.org/10.3389/fgene.2024.1293824. (**Published 2024 Mar 20**).
50. Zhang Y, Mi K, Li Z, et al. Identification of prognostic miRNAs associated with immune cell tumor infiltration predictive of clinical outcomes in patients with non-small cell lung cancer. Front Oncol. 2021;11:705869. https://doi.org/10.3389/fonc.2021.705869. (**Published 2021 Jul 1**).
51. Tarazona JGR, Abdallah EA, Flores BCT, et al. mir-203a-3p and mmp-2 proteins are highly expressed in circulating tumor cells from patients with pancreatic carcinoma. proteínas mir-203a-3p e mmp-2 são altamente expressas em células tumorais circulantes de pacientes com carcinoma pancreático. Arq Bras Cir Dig. 2022;34(4):1628. https://doi.org/10.1590/0102-672020210002e1628. (**Published 2022 Jan 31**).
52. Chen L, Gao H, Liang J, et al. miR-203a-3p promotes colorectal cancer proliferation and migration by targeting PDE4D. Am J Cancer Res. 2018;8(12):2387–401 (**Published 2018 Dec 1**).
53. Ma D, Chen S, Wang H, et al. Baicalein induces apoptosis of pancreatic cancer cells by regulating the expression of miR-139–3p and miR-196b-5p. Front Oncol. 2021;11:653061. https://doi.org/10.3389/fonc.2021.653061. (**Published 2021 Apr 30**).
54. Xin H, Wang C, Chi Y, Liu Z. MicroRNA-196b-5p promotes malignant progression of colorectal cancer by targeting ING5. Cancer Cell Int. 2020;20:119. https://doi.org/10.1186/s12935-020-01200-3. (**Published 2020 Apr 10**).
55. Huang X, Xiao S, Zhu X, et al. miR-196b-5p-mediated downregulation of FAS promotes NSCLC progression by activating IL6-STAT3 signaling. Cell Death Dis. 2020;11(9):785. https://doi.org/10.1038/s41419-020-02997-7. (**Published 2020 Sep 22**).
56. Chauhan N, Dhasmana A, Jaggi M, Chauhan SC, Yallapu MM. miR-205: a potential biomedicine for cancer therapy. Cells. 2020;9(9):1957. https://doi.org/10.3390/cells9091957. (**Published 2020 Aug 25**).

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