



## OPEN ACCESS

## EDITED BY

Dawei Chen,  
University of Kiel, Germany

## REVIEWED BY

Zhengrui Li,  
Shanghai Jiao Tong University, China  
Ru-chen Zhou,  
Shandong University, China

## \*CORRESPONDENCE

Xiao Hu,  
✉ huxiao202@163.com

†These authors have contributed equally to this work

RECEIVED 18 September 2024

ACCEPTED 11 November 2024

PUBLISHED 21 November 2024

## CITATION

Fu K, Su J, Zhou Y, Chen X and Hu X (2024) The role of epigenetic regulation in pancreatic ductal adenocarcinoma progression and drug response: an integrative genomic and pharmacological prognostic prediction model. *Front. Pharmacol.* 15:1498031. doi: 10.3389/fphar.2024.1498031

## COPYRIGHT

© 2024 Fu, Su, Zhou, Chen and Hu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# The role of epigenetic regulation in pancreatic ductal adenocarcinoma progression and drug response: an integrative genomic and pharmacological prognostic prediction model

Kang Fu<sup>†</sup>, Junzhe Su<sup>†</sup>, Yiming Zhou, Xiaotong Chen and Xiao Hu\*

Department of Hepatobiliary Pancreatic Surgery, The Affiliated Hospital of Qingdao University, Qingdao, China

**Background:** Pancreatic ductal adenocarcinoma (PDAC) is a highly lethal malignancy with poor prognosis. Epigenetic dysregulation plays a crucial role in PDAC progression, but its comprehensive landscape and clinical implications remain unclear.

**Methods:** We integrated single-cell RNA sequencing, bulk RNA sequencing, and clinical data from multiple public databases. Single-cell analysis was performed using Seurat and hdWGCNA packages to reveal cell heterogeneity and epigenetic features. Weighted gene co-expression network analysis (WGCNA) identified key epigenetic modules. A machine learning-based prognostic model was constructed using multiple algorithms, including Lasso and Random Survival Forest. We further analyzed mutations, immune microenvironment, and drug sensitivity associated with the epigenetic risk score.

**Results:** Single-cell analysis revealed distinct epigenetic patterns across different cell types in PDAC. WGCNA identified key modules associated with histone modifications and DNA methylation. Our machine learning model, based on 17 epigenetic genes, showed robust prognostic value (AUC >0.7 for 1-, 3-, and 5-year survival) and outperformed existing models. High-risk patients exhibited distinct mutation patterns, including higher frequencies of KRAS and TP53 mutations. Low-risk patients showed higher immune and stromal scores, with increased infiltration of CD8<sup>+</sup> T cells and M2 macrophages. Drug sensitivity analysis revealed differential responses to various therapeutic agents between high- and low-risk groups, with low-risk patients showing higher sensitivity to EGFR and MEK inhibitors.

**Conclusion:** Our study provides a comprehensive landscape of epigenetic regulation in PDAC at single-cell resolution and establishes a robust epigenetics-based prognostic model. The integration of epigenetic features with mutation profiles, immune microenvironment, and drug sensitivity offers new insights into PDAC heterogeneity and potential therapeutic strategies. These findings pave the way for personalized medicine in PDAC management and highlight the importance of epigenetic regulation in cancer research.

## KEYWORDS

pancreatic ductal adenocarcinoma, epigenetic regulation, single-cell RNA sequencing, machine learning, prognostic model, tumor microenvironment, drug sensitivity

## 1 Introduction

Pancreatic cancer is an extremely dangerous malignant tumor. Despite significant advances in cancer treatment over the past few decades, the prognosis for pancreatic cancer remains poor. Statistics show that the 5-year survival rate for pancreatic cancer patients is only 9% (Rawla et al., 2019). More worryingly, the incidence of pancreatic cancer has been on the rise in recent years. Pancreatic ductal adenocarcinoma (PDAC) is the main type of pancreatic cancer, accounting for over 90% of cases, and is projected to become the second deadliest cancer by 2030 (Nakaoka et al., 2023; Mizrahi et al., 2020). The poor treatment outcomes for PDAC are primarily due to its unique biological characteristics. First, its high metabolic plasticity and adaptability allow it to survive and proliferate rapidly in harsh tumor microenvironments (Bi et al., 2024; Shah et al., 2024). Second, the dense stroma specific to PDAC not only hinders drug penetration but also enables it to evade immune system surveillance (Wilson et al., 2014; Timmer et al., 2021). Third, PDAC exhibits high tumor heterogeneity (Bailey et al., 2016; Wang X. et al., 2023). These characteristics collectively lead to its resistance to traditional treatment methods.

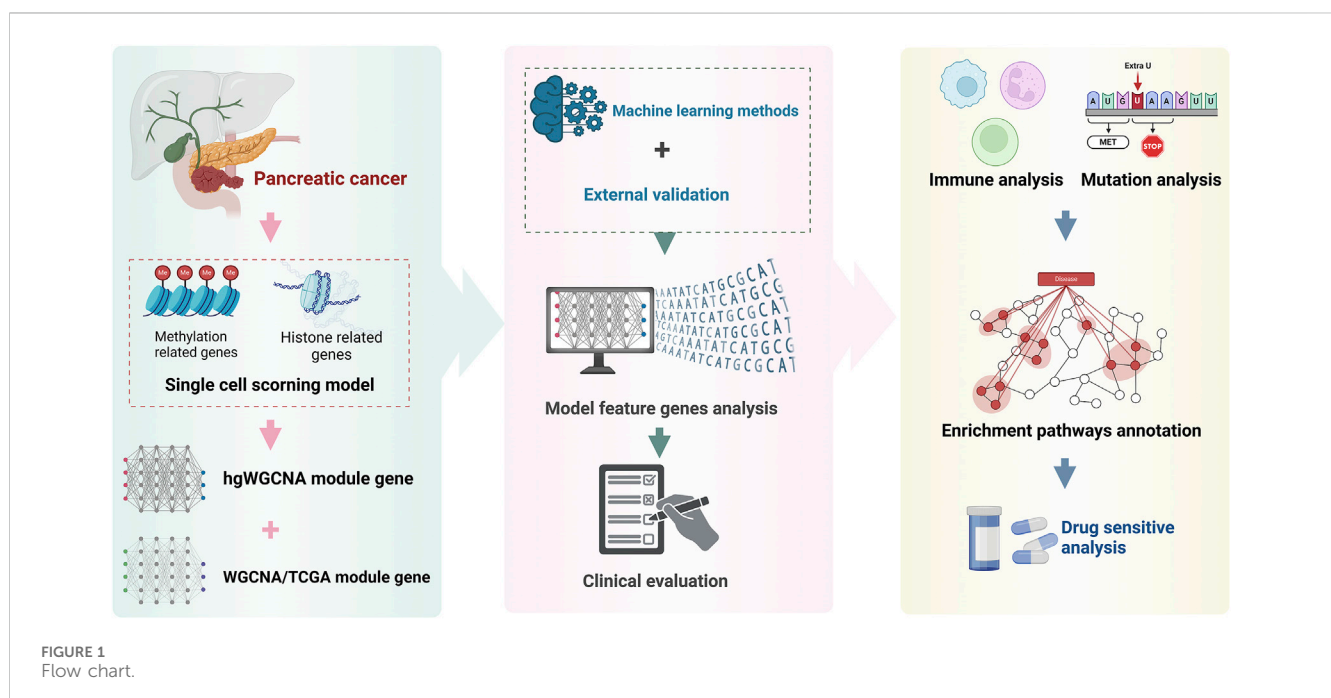
In recent years, two major epigenetic mechanisms - DNA methylation and histone modification - have been recognized as playing crucial roles in the occurrence, progression, and treatment resistance of PDAC. Lomber et al. elucidated that data from many laboratories have demonstrated that oncogenic mutations in PDAC (such as Kras) lead to downstream signaling events that regulate histone and DNA modifications, partly through direct regulation of histones and histone and DNA modifying enzymes, thereby stimulating cell growth (Lomber et al., 2019). Cedar et al. and Liu et al. also described the interdependence and crosstalk between DNA methylation and histone modification patterns (Cedar and Bergman, 2009; Liu et al., 2016). DNA methylation primarily occurs on CpG islands and is usually associated with gene silencing (Nishiyama and Nakanishi, 2021). In PDAC, several

key tumor suppressor genes, such as CDKN2A (Goodwin et al., 2023), RASSF1A (Amato et al., 2016), and BRCA1 (Lai et al., 2021), have been found to be inactivated due to hypermethylation in their promoter regions, leading to dysregulation of important pathways such as cell cycle regulation, DNA repair, and apoptosis, promoting tumor formation and progression. On the other hand, histone modifications regulate gene expression by altering chromatin structure and transcription factor accessibility. In PDAC, abnormalities in histone acetylation and methylation have been widely reported. For example, overexpression of histone deacetylases (HDACs) leads to silencing of multiple tumor suppressor genes (Schneider et al., 2010), while upregulation of the histone methyltransferase EZH2 is associated with increased invasiveness and metastatic potential of PDAC (Versemann et al., 2022).

To date, although there have been numerous studies on the molecular mechanisms of DNA methylation and histone modification in PDAC, no research has constructed a comprehensive epigenetic regulatory landscape and prognostic model. Clinically, there is still no application of epigenetic regulation in disease stratification and treatment. This study aims to construct an epigenetic regulatory landscape of PDAC through integrated analysis of DNA methylation and histone modification data. By applying advanced machine learning algorithms, single-cell analysis, and other techniques, we will identify key epigenetic regulatory modules and explore their associations with gene expression, signaling pathway activation, and clinical phenotypes. Combining patient clinical follow-up data, we will identify PDAC heterogeneity from an epigenomic perspective and develop a prognostic prediction model based on epigenetic features. This model will not only help identify high-risk patients but may also provide guidance for individualized treatment decisions.

## 2 Materials and methods

The flow chart are shown in Figure 1.



## 2.1 Data source

This study integrates multiple public database resources. Specifically, the single-cell sequencing data included was obtained from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>), accession number GSE212966, which contains single-cell sequencing data from 6 PDAC adjacent normal tissue samples and 6 PDAC samples. Additionally, bulk sequencing expression profiles and survival information were sourced from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>), including gene expression data and corresponding survival information for 179 PDAC samples (Tomczak et al., 2015). The external validation set was obtained from the International Cancer Genome Consortium (ICGC) database (<https://dcc.icgc.org/>), comprising 234 Canadian samples and 91 Australian samples (Zhang et al., 2019).

Furthermore, in the selection of specific gene sets, we focused on genes related to epigenetics. Histone modification-related genes were primarily sourced from two origins: first, the cancer-related histone modification study reported by Füllgrabe et al. (2011), and second, genes with a histone modification relevance score greater than 20 in the GeneCards database (<https://www.genecards.org/>). For methylation-related genes, we referenced the research findings of Lee et al. (2020) and supplemented them with genes having a DNA methylation modification relevance score greater than 20 in the GeneCards database.

## 2.2 Single cell analysis and hdWGCNA

The single-cell RNA-seq data in this study was processed using the Seurat package and hdWGCNA package (Stuart et al., 2019; Morabito et al., 2023). First, we read the data in 10X Genomics format from the GSE212966 dataset. Quality control was performed to filter out cells with <100 or >5,000 detected genes, >15% mitochondrial gene content, and <1000 UMI counts. After data filtering and normalization, we performed dimensionality reduction using PCA and UMAP, and integrated different samples using the Harmony algorithm (Cristian et al., 2024). For cell type annotation, we identified specific marker genes for each cell type: cancer cells (EPCAM, KRT19, CEACAM6), fibroblasts (COL1A1, DCN, FAP), endothelial cells (PECAM1, VWF, CDH5), macrophages (CD68, CD163, CSF1R), T cells (CD3D, CD8A, CD4), and B cells (CD79A, CD19, MS4A1). Subsequently, we applied clustering analysis and calculated gene expression differences of the corresponding clusters for cell type annotation. To construct epigenetic scores, we used the ssGSEA method (Chen et al., 2022) to calculate enrichment scores for histone gene sets and methylation gene sets for each cell, and divided cells into high-score and low-score groups based on the median score. On one hand, we performed differential expression analysis using Seurat's FindAllMarkers function with a log<sub>2</sub>FC threshold of 0 and a minimum percentage of 0.35. On the other hand, we applied the hdWGCNA package to construct weighted gene co-expression networks, determine the optimal soft threshold parameters, identify co-expression modules and hub genes. By calculating module characteristic genes and module connectivity, we analyzed the relationships between modules and cell types and phenotypes, thereby comprehensively revealing the expression

patterns of histone modifications and DNA methylation in different cell types and their potential biological functions.

## 2.3 Weighted gene co-expression network analysis (WGCNA)

The Weighted Gene Co-expression Network Analysis (WGCNA) method was used to further explore the association between epigenetics and PDAC (Langfelder and Horvath, 2008). We calculated histone modification and DNA methylation scores for samples using the ssGSEA algorithm with the Gaussian kernel. Subsequently, we extracted gene expression data for PDAC samples from the TCGA database and selected genes related to histone modification and DNA methylation for WGCNA analysis. By determining the optimal soft threshold parameters (power = 5 for both histone modification and DNA methylation analyses), we constructed gene co-expression networks using the unsigned TOM type with a minimum module size of 50 genes. We identified biologically significant gene modules by setting the merge cut height to 0.15. Notably, we found that histone modification-related genes were mainly enriched in the grey module, while DNA methylation-related genes were primarily enriched in the brown module. We further analyzed the relationships between these modules and sample characteristics (such as histone modification scores and DNA methylation scores) using Pearson correlation. Through correlation analysis of Module Membership and Gene Significance, we identified key genes in each module. The correlation between module membership and gene significance for the grey module (histone modification) and brown module (DNA methylation) was visualized using scatter plots. By performing intersection analysis on these genes, we were able to obtain genes that showed high correlation (absolute correlation coefficient >0.4 and p-value <0.05) across multiple states, thereby providing a research foundation for subsequent model construction.

## 2.4 Machine learning based prognosis signature construction

To further evaluate the potential of epigenetic-related genes in PDAC prognosis prediction, this study employed a comprehensive and advanced set of machine learning methods. We integrated multiple algorithms, including Random Survival Forest (RSF) (Ishwaran et al., 2008), Elastic Net (Enet) (Zou and Hastie, 2005; Cho et al., 2009), Stepwise Cox Regression (StepCox) (Liu et al., 2023), CoxBoost (Binder and Binder, 2015), Partial Least Squares Cox Regression (plsRcox) (Bertrand et al., 2022; Bertrand et al., 2014), SuperPC (Bair et al., 2006), Gradient Boosting Machine (GBM) (Ayyadevara and Ayyadevara, 2018), Survival Support Vector Machine (survival-SVM) (Van Belle et al., 2011), Ridge Regression (Arashi et al., 2021), and Lasso Regression (Ranstam and Cook, 2018), to construct a series of prognostic prediction models. To ensure the reliability and generalization ability of the models, we used the TCGA dataset as the training set and selected two independent ICGC datasets as external validation sets. In the data preprocessing stage, we standardized all features to eliminate

the influence of scale differences. Subsequently, we not only evaluated the performance of each algorithm individually but also explored up to 63 algorithm combinations, such as RSF + CoxBoost, Lasso + GBM, etc., aiming to obtain more stable and accurate prediction results. We used the C-index as the primary evaluation metric to comprehensively assess the discriminative ability of each model on both the training and validation sets, and presented the results in heatmap form for quick identification of the best models. For RSF, we used 1,000 trees and optimized the node size. Enet models were built with  $\alpha$  values ranging from 0.1 to 0.9. CoxBoost models were optimized using cross-validation to determine the optimal number of boosting steps. For GBM, we used 10,000 trees with a maximum depth of 3 and a learning rate of 0.001. Lasso and Ridge regression models were fitted using 10-fold cross-validation to select the optimal  $\lambda$  value. Finally, we conducted in-depth analysis of the best-performing models, including feature importance ranking and result visualization, aiming to reveal the crucial role of epigenetic-related genes in PDAC prognosis and provide reliable data support for clinical individualized treatment decisions.

## 2.5 Relevant genes and risk score signature evaluation

For epigenetic-related genes, we conducted a series of network-based analyses. We performed univariate Cox regression analysis using the expression data of these genes, selecting genes significantly associated with patient survival ( $p < 0.05$ ). We calculated their hazard ratios (HR) and p-values. Next, we computed the Spearman correlation coefficients (Pripp, 2018) between these genes and conducted correlation significance tests. Finally, we retained only gene pairs with absolute correlation coefficients greater than 0.4 and p-values less than 0.05 to ensure that the connections in the network have biological significance. Based on these data, we constructed a gene interaction network, where red represents positive correlations, blue represents negative correlations, and the thickness of the lines reflects the strength of the correlation. The igraph package was used for visualization (Csardi and Tamas, 2005).

Furthermore, to evaluate the predictive value of our developed epigenetic-related gene prognostic model in PDAC, we conducted comprehensive Cox regression analyses based on sample clinical characteristics. First, we performed univariate Cox regression analysis on clinical variables, including age, gender, T stage, N stage, M stage, clinical stage, and our risk score. After calculating the hazard ratio (HR), 95% confidence interval (CI), and p-value for each variable, we selected variables with statistical significance in the univariate analysis for multivariate Cox regression analysis. We used the forestplot package to create forest plots to display the HR and 95% CI for each variable.

## 2.6 Enrichment analysis

This study continued to employ various bioinformatics methods to explore the relationship between gene expression patterns and prognostic risk. We used the Gene Set Variation Analysis (GSVA) method (Hänzelmann et al., 2013), utilizing the GSVA package to perform ssGSEA on the PDAC dataset, quantifying the activity of specific pathways in each sample. We chose the Hallmark gene set

(Liberzon et al., 2015) as a reference, ensuring biological relevance of the analysis. Subsequently, we compared the pathway activity differences between high-risk and low-risk groups, using the limma package for differential analysis (Ritchie et al., 2015). To visually present the results, we generated a bar plot showing significantly different pathways and their t-values. Additionally, we calculated the correlation between GSVA scores and risk scores, visualizing these relationships through a heatmap. For pathways with significant statistical differences, we used Cox proportional hazards regression analysis to quantify the association strength between pathway activity and survival risk.

## 2.7 Mutation analysis

To further explore the association between the gene risk model and tumor mutation characteristics, we conducted a comprehensive mutation analysis on the PDAC dataset. First, we calculated the tumor heterogeneity (MATH) score for each sample using the maftools package (Mayakonda et al., 2018), and compared the differences between high and low-risk groups. Subsequently, we divided the samples into high and low-risk groups, generating mutation landscape plots (oncoplots) for each group, showing the 20 most common mutated genes and their frequencies. To understand the mutation patterns more deeply, we also performed somatic mutation interaction analysis, revealing the co-occurrence and mutual exclusivity of gene mutations in high and low-risk groups. This is significant for revealing and understanding the potential link between risk scores and tumor mutation burden and heterogeneity.

## 2.8 Immune analysis

The epigenetic-related gene risk model is closely related to the tumor immune microenvironment. We used the IOBR package (Zeng et al., 2021) to assess ESTIMATE, CIBERSORT, and immune cell subpopulation infiltration in PDAC samples. First, we used the ESTIMATE algorithm to calculate stromal scores, immune scores, and ESTIMATE scores for each sample, and compared the differences between high and low-risk groups. Subsequently, we used the ssGSEA method to perform enrichment analysis on immune-related pathways, and visualized the significantly different pathway activities between high and low-risk groups through heatmaps. To understand the composition of tumor immune cells in more detail, we used the CIBERSORT algorithm to perform deconvolution analysis on 22 immune cell subpopulations (Chen et al., 2018). Through violin plots, we visually demonstrated the differences in immune cell composition between high and low-risk groups. Additionally, we calculated the Spearman correlation between these immune cell subpopulations and risk scores, and used bubble plots to show the correlation strength and statistical significance.

## 2.9 Drug sensitivity analysis

The Genomics of Drug Sensitivity in Cancer (GDSC) database (<https://www.cancerrxgene.org/>) is commonly used to predict tumor sensitivity to drugs (Yang et al., 2013). In this study, we



TABLE 1 Primer sequences for qRT-PCR.

Gene	Direction	Primer sequence (5'-3')
KRTCAP2	Forward	CTCTTCGTGTTCTCGCTCACT
	Reverse	CAGGTGGTGACACAGACTCG
NENF	Forward	AGATCAGCCCATCTACTTGGC
	Reverse	CTTCCCCGTC AAGGCATTG
PSAP	Forward	CCCGGTCCTTGGACTGAAAG
	Reverse	TATGTGCGAGGGAAGGGATTT
MRPL41	Forward	GTTGTCGTCGCCGATCTG
	Reverse	GTAGCTCACGTAGGGCTTGA
S100A16	Forward	ATGTCAGACTGCTACACGGAG
	Reverse	GTTCTTGACCAGGCTGTACTTAG

utilized the GDSC database and the pRRophetic package (Geeleher et al., 2014) to explore the potential application of our epigenetic-related gene risk model in drug sensitivity prediction. We first obtained drug sensitivity data and gene expression data from the GDSC database, then used the pRRopheticPredict function to predict drug sensitivity for each sample in the TCGA-PDAC dataset. Based on some clinical drugs in the GDSC database, we predicted and compared drug sensitivity differences between high and low-risk groups. For each drug, we used the Wilcoxon rank-sum test to compare the differences in predicted IC50 values between high and low-risk groups, and created box plots for visualization (Sebaugh, 2011). These results not only revealed potential connections between our risk model and drug responses but also provided new insights for personalized treatment of PDAC.

## 2.10 Pathological validation

To validate the biological significance of our model and examine the gene expression trends, we conducted further pathological verification on the top five genes with the highest weights in our model. On one hand, we utilized the Human Protein Atlas (HPA) database (<https://www.proteinatlas.org/>) to compare the protein expression levels of these genes in pancreatic cancer tissues and adjacent relatively normal pancreatic tissues (Uhlén et al., 2015). For genes lacking data in the HPA database, we performed immunohistochemical staining verification in our laboratory. All patients provided written informed consent, and this research protocol was approved by the Medical Ethics Committee of the Affiliated Hospital of Qingdao University (approval number: QYFYWZLL28682).

Besides, gene expression was further validated by qRT-PCR. PANC-1 (pancreatic cancer) and hTERT-HPNE (normal pancreatic ductal) cell lines were cultured in DMEM supplemented with 10% FBS and 1% penicillin/streptomycin under standard conditions (37°C, 5% CO<sub>2</sub>). Total RNA was extracted using RNAios Plus reagent following the manufacturer's protocol. RNA was reverse transcribed using ABScript III RT Master Mix, and qRT-PCR was performed using Universal SYBR Green Fast qPCR Mix. The sequences of primers used for qRT-PCR are listed in Table 1. Relative gene expression was calculated using the 2<sup>-ΔΔCT</sup> method.

## 3 Results

### 3.1 Single-cell RNA sequencing reveals cellular heterogeneity and epigenetic characteristics in the PDAC microenvironment

We performed single-cell RNA sequencing analysis on 6 PDAC adjacent normal tissue samples (ADJ) and 6 PDAC samples from the GSE212966 dataset. Fourteen major cell types were identified and annotated (Figure 2A). These cell types include cancer cells, fibroblasts, endothelial cells, smooth muscle cells, and various immune cell subpopulations such as macrophages, T cells, and B cells.

Comparison of cell composition between ADJ and PDAC samples (Figure 2B) revealed significant microenvironment remodeling in PDAC tissues. The proportions of cancer cells, fibroblasts, mast cells, and Treg cells were markedly increased in PDAC samples, while the proportions of endothelial cells and B cells substantially decreased (Figure 2C). To further validate the accuracy of cell type annotation, we plotted a heatmap showing the expression of marker genes for each cell cluster (Figure 2D).

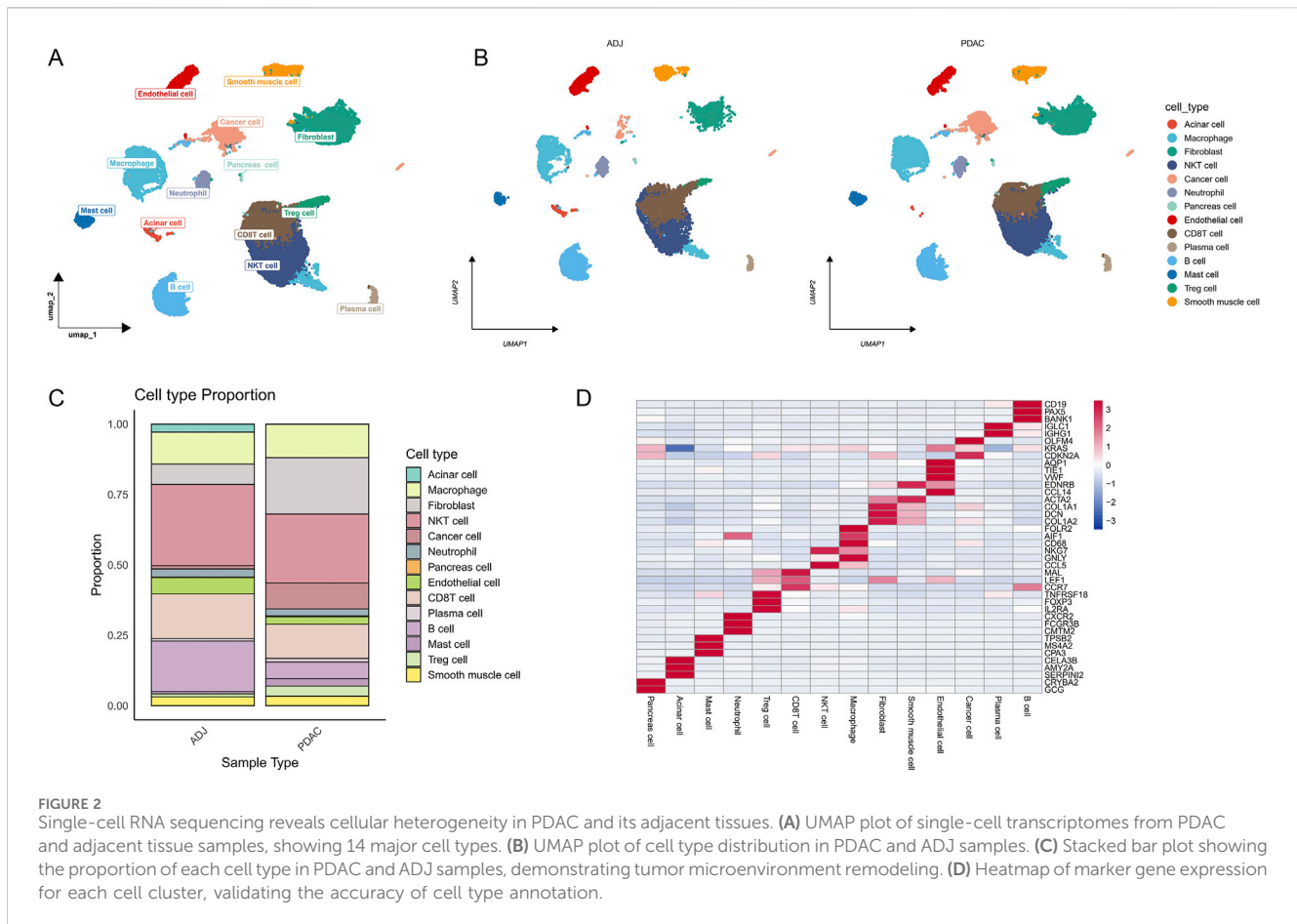
### 3.2 Single-cell level histone modification characteristics in PDAC

To investigate the role of epigenetic regulation in PDAC in depth, we conducted a detailed analysis of the expression patterns of histone modification-related genes. Using the ssGSEA method, we calculated a histone modification score (Histone score) for each single cell. Figure 3A shows the distribution of histone modification scores on the UMAP plot. Further comparison revealed that cells in PDAC samples generally had higher histone modification scores, showing significant differences compared to ADJ samples ( $p = 0.0032$ , Figure 3B). This suggests that PDAC cells may enhance histone modifications to regulate gene expression, thereby promoting tumor progression. Cell type-specific analysis showed no significant differences in histone modification levels among different cell populations in PDAC (Figure 3C).

To further reveal the co-expression patterns of histone modification-related genes, we applied the hdWGCNA method to construct a weighted gene co-expression network. By analyzing soft threshold parameters (Figure 3D), we determined the optimal network construction parameters. The final hierarchical clustering dendrogram (Figure 3E) shows the gene module structure under high histone modification states, with different colors representing different co-expression modules. These modules may reflect functionally coordinated gene sets during PDAC progression.

### 3.3 Single-cell level DNA methylation characteristics in PDAC

Similar to the previous section, we further analyzed the expression patterns of DNA methylation-related genes in PDAC. By calculating a DNA methylation score (Methylation score) for each single cell and displaying it on a UMAP plot (Figure 4A), we found that cells in PDAC samples generally have higher DNA



methylation scores, showing highly significant differences compared to ADJ samples ( $p = 2.4e-11$ , Figure 4B). This finding suggests that PDAC cells may enhance DNA methylation to regulate gene expression. Cell type-specific analysis revealed no significant differences in DNA methylation levels among different cell populations in PDAC (Figure 4C). Subsequently, we applied the hdWGCNA method to construct a weighted gene co-expression network. By analyzing soft threshold parameters (Figure 4D), we determined the optimal network construction parameters. The final hierarchical clustering dendrogram (Figure 4E) shows the gene module structure under high DNA methylation states.

### 3.4 WGCNA reveals key gene modules of epigenetic regulation in PDAC

To further explore the role of epigenetic regulation in PDAC, we performed WGCNA on the TCGA-PDAC dataset. We analyzed histone modification and DNA methylation-related genes separately.

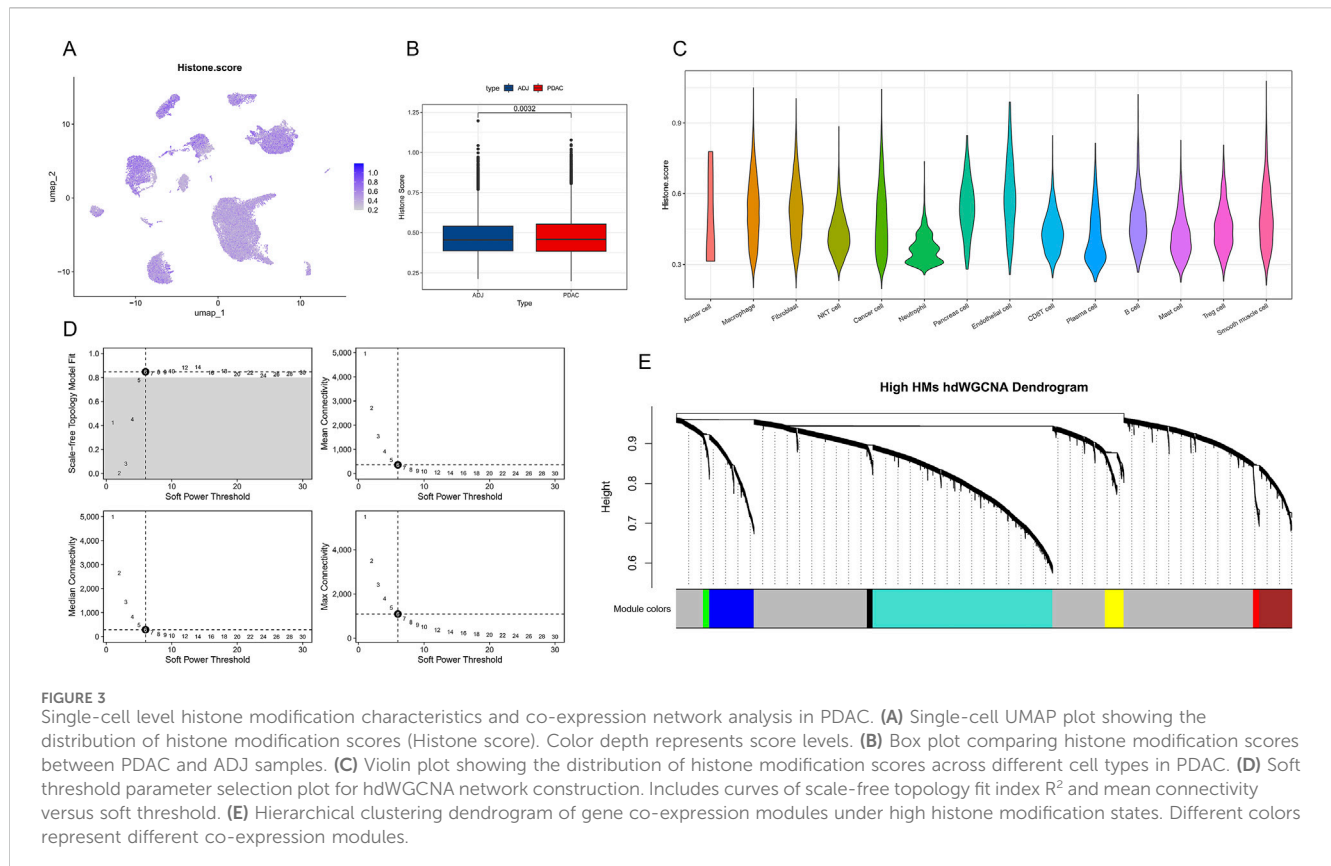
The WGCNA results for histone modification-related genes are shown in Figures 5A–D. Sample clustering and trait heatmap (Figure 5A) display sample similarity and the distribution of histone modification scores. The gene dendrogram and module assignment (Figure 5B) reveal three main co-expression modules.

Module-trait relationship analysis (Figure 5C) indicates that the grey module shows the strongest positive correlation with histone modification scores (correlation coefficient = 0.45,  $p = 1e-09$ ). Further module membership vs. gene significance scatter plot (Figure 5D) shows that genes in the grey module have high module membership and gene significance.

Similarly, WGCNA results for DNA methylation-related genes are shown in Figures 5E–H. Module-trait relationship analysis (Figure 5G) shows that the brown module has the strongest positive correlation with DNA methylation scores (correlation coefficient = 0.55,  $p = 2e-15$ ). The module membership vs. gene significance scatter plot (Figure 5H) further confirms the importance of genes in the brown module.

### 3.5 Integrated analysis reveals key genes and pathways of epigenetic regulation in PDAC

We found that 108 genes were simultaneously identified in differential expression analysis, hdWGCNA, and WGCNA for histone modifications (Figure 6A). In DNA methylation-related analysis, 285 genes were commonly identified (Figure 6B). Functional enrichment analysis revealed the biological processes and molecular functions involved in these epigenetic key genes (Figure 6C). Among them, protein folding and mitochondrial function processes were significantly enriched. KEGG pathway



analysis further identified several signaling pathways closely related to PDAC, including Chemical carcinogenesis - reactive oxygen species, Oxidative phosphorylation, etc. (Figure 6D). Differential expression analysis between normal samples and PDAC are shown in a volcano plot (Figure 6E), where red and blue dots represent upregulated and downregulated genes, respectively. The circular plot (Figure 6F) visually displays the top 50 PDAC DEGs with the most significant expression changes. Furthermore, we compared the overlap between epigenetic-related genes and PDAC DEGs (Figure 6G). The results show that 126 genes are both related to epigenetic regulation and differentially expressed in PDAC. Finally, we constructed a protein-protein interaction network of these key genes (Figure 6H), showing a high positive correlation between *NDUFA13* in protective genes and *PABPC4* in risk genes.

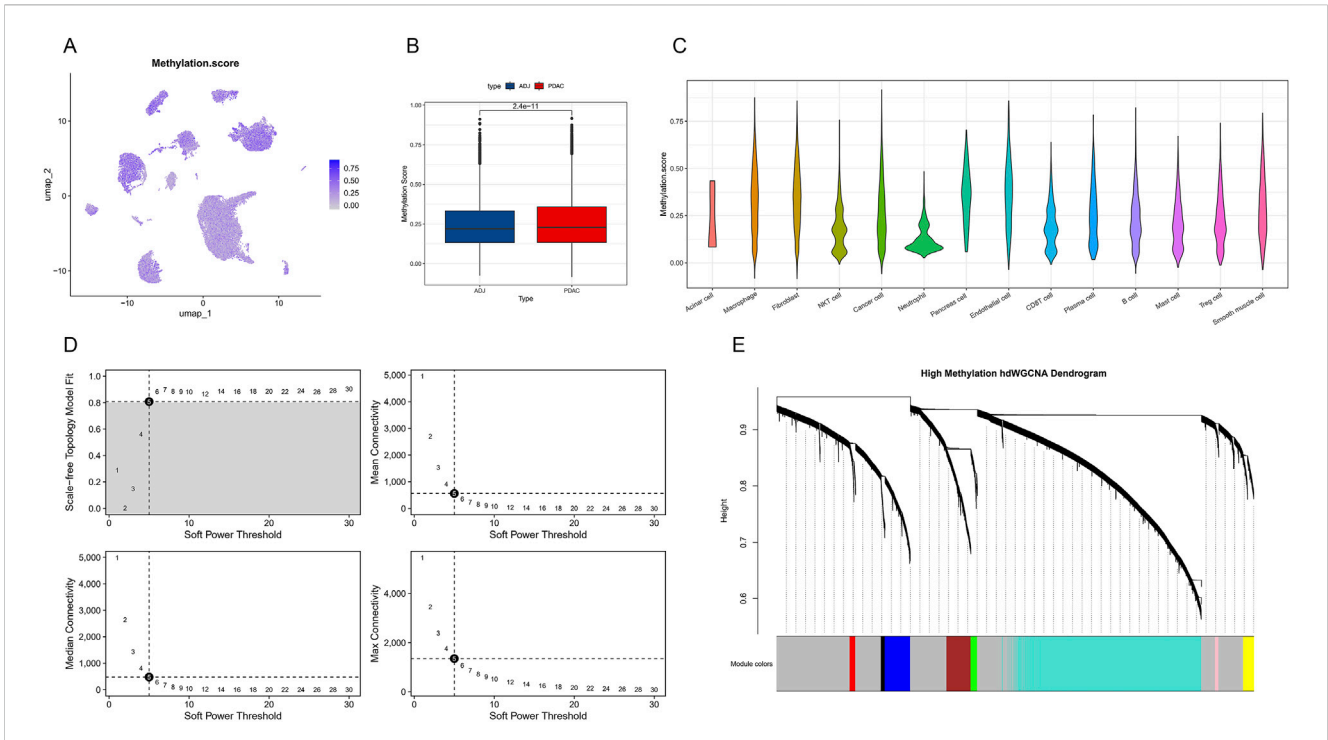
### 3.6 Machine learning models for predicting PDAC patient prognosis

To assess the potential of epigenetic-related genes in predicting PDAC prognosis, we applied various machine learning algorithms to build prognostic models. Figure 7A shows a performance comparison of different algorithm combinations, with the Lasso + RSF combination performing best across multiple evaluation metrics. Figures 7B, C illustrate the feature selection process of Lasso regression. Through L1 regularization (Figure 7B), we gradually increased the penalty coefficient  $\lambda$  to select the optimal feature subset. Figure 7C's partial least squares path graph shows the

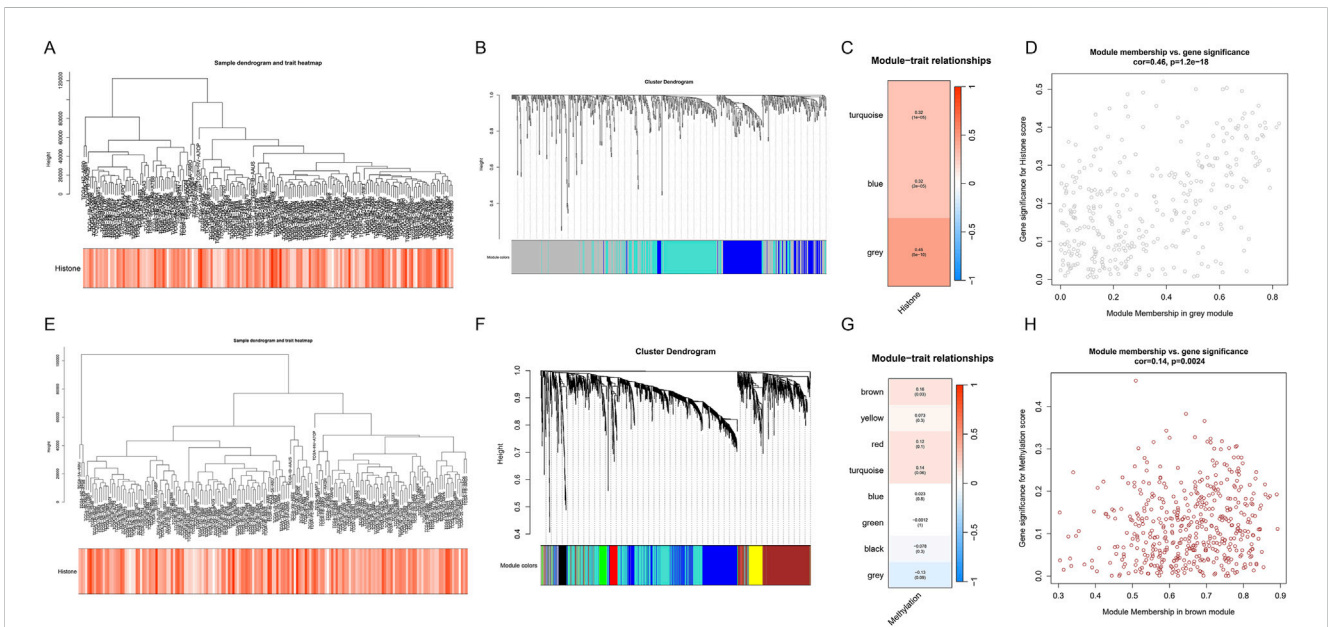
change trend of different feature coefficients as  $\lambda$  increases, helping us identify the most stable and important prognostic-related genes. The final weight calculation formula is as follows: Riskscore = (0.005)\*HSPB1 + (0.0111)\*BST2 + (0.0176)\*BLVRB + (-0.0262)\*TMEM176A + (0.0054)\*IFI27 + (-0.0408)\*PPP2R1A + (0.1781)\*S100A16 + (-0.2303)\*NENF + (0.0952)\*LY6E + (-0.0303)\*TBCB + (0.1254)\*COA4 + (0.0909)\*CEBPB + (-0.1997)\*MRPL41 + (-0.2693)\*KRTCAP2 + (0.1462)\*SNRPG + (0.0278)\*NUPR1 + (-0.2068)\*PSAP.

Based on the selected features, we calculated a risk score for each patient and divided patients into high-risk and low-risk groups. Figure 7D clearly shows the separation of the two patient groups in the risk score distribution plot. Figure 7E's Kaplan-Meier survival curve further confirms the significant difference in survival time between high-risk and low-risk groups ( $p < 0.001$ ). Figure 7F's heatmap shows the expression patterns of key genes in the model for high-risk and low-risk groups. We observed clear differential expression of these genes between the two groups, further supporting the rationality of our risk score model. Finally, Figure 7G's time-dependent ROC curve analysis evaluated the predictive accuracy of our model. The results show that the model demonstrates good discriminative ability in predicting 1-year, 3-year, and 5-year prognosis, with AUC values of 0.746, 0.81, and 0.822, respectively.

These results suggest that machine learning models based on epigenetic-related genes can effectively predict the prognosis of PDAC patients, providing a potential tool for individualized treatment decisions.

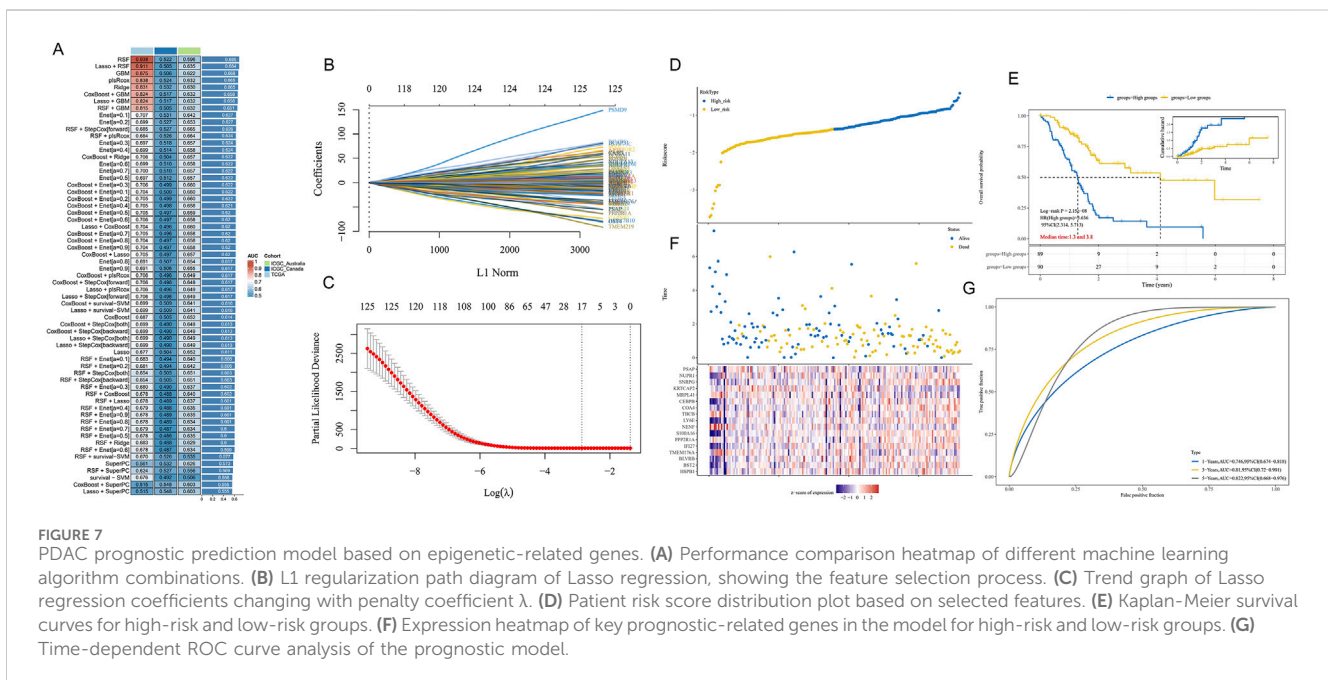
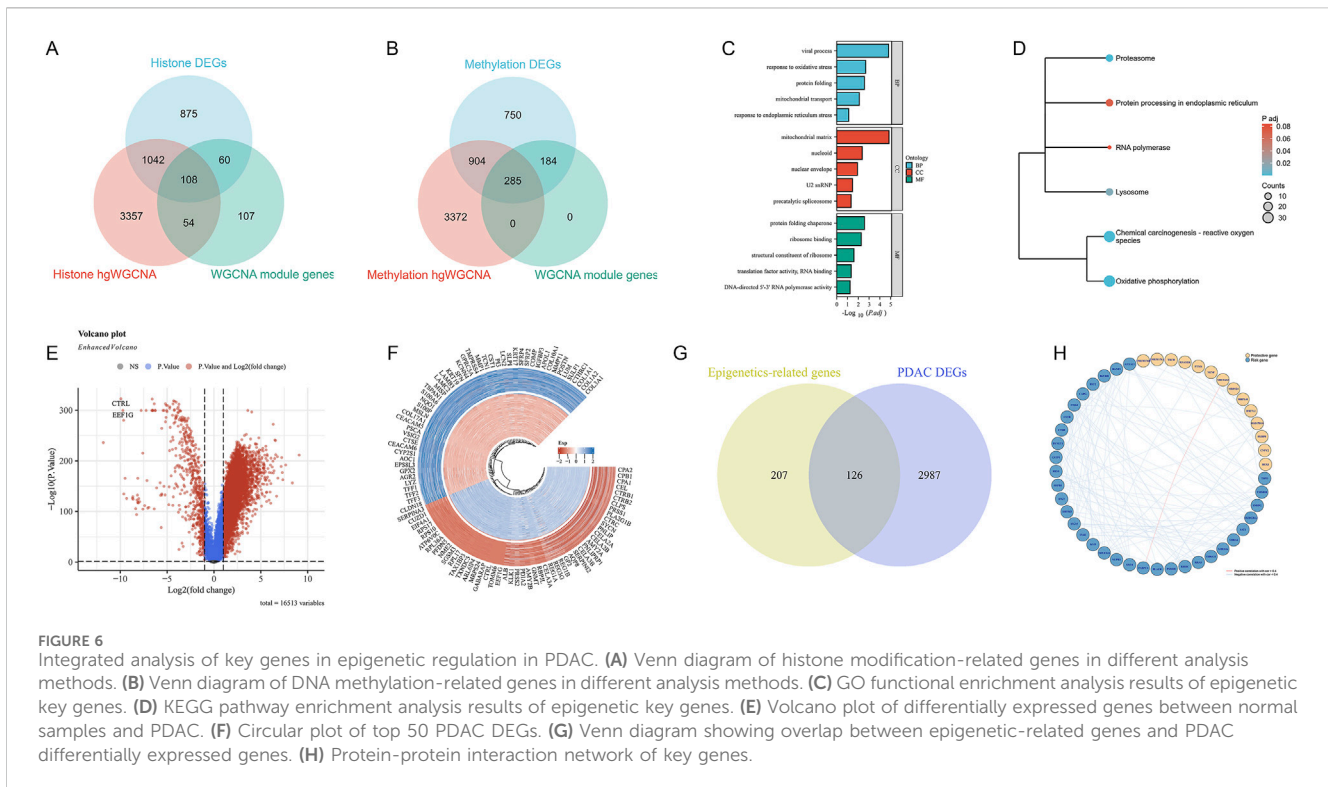


**FIGURE 4** Single-cell level DNA methylation characteristics and co-expression network analysis in PDAC. **(A)** Single-cell UMAP plot showing the distribution of DNA methylation scores (Methylation score). **(B)** Box plot comparing DNA methylation scores between PDAC and ADJ samples. **(C)** Violin plot showing the distribution of DNA methylation scores across different cell types in PDAC. **(D)** Soft threshold parameter selection plot for hdWGCNA network construction. Includes curves of scale-free topology fit index  $R^2$  and mean connectivity versus soft threshold. **(E)** Hierarchical clustering dendrogram of gene co-expression modules under high DNA methylation states.



**FIGURE 5** WGCNA analysis of TCGA-PDAC dataset reveals key gene modules related to epigenetic regulation. **(A–D)** WGCNA analysis of histone modification-related genes. **(A)** Sample clustering dendrogram and histone modification score heatmap. **(B)** Gene dendrogram and module assignment. **(C)** Module-trait relationship heatmap. **(D)** Module membership vs. gene significance scatter plot for the grey module. **(E–H)** WGCNA analysis of DNA methylation-related genes. **(E)** Sample clustering dendrogram and DNA methylation score heatmap. **(F)** Gene dendrogram and module assignment. **(G)** Module-trait relationship heatmap. **(H)** Module membership vs. gene significance scatter plot for the brown module.

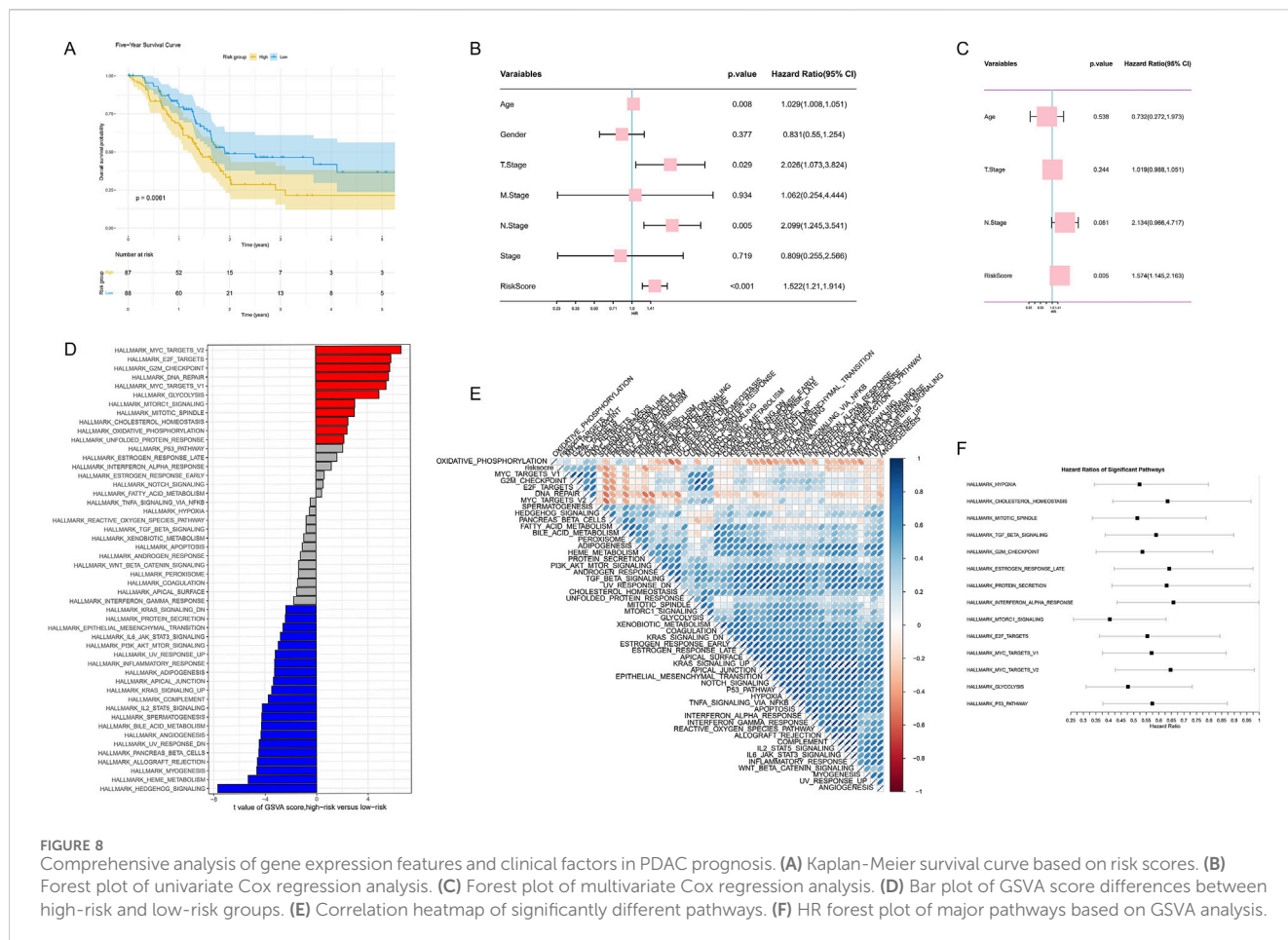




### 3.7 The role of gene expression features and clinical factors in PDAC prognosis

To further validate our prognostic model and explore the role of gene expression features and clinical factors in PDAC prognosis, we conducted a series of analyses. Based on our risk score model, we plotted Kaplan-Meier survival curves, which showed that high-risk

group patients had significantly lower survival rates than the low-risk group ( $p = 0.0061$ , Figure 8A). We performed univariate Cox regression analysis, revealing multiple factors associated with prognosis, including age, T stage, N stage, and our risk score (Figure 8B). Multivariate Cox regression analysis further confirmed that our risk score is an independent prognostic indicator beyond other clinical factors ( $p = 0.005$ , Figure 8C).



To gain deeper insights into molecular differences between high-risk and low-risk groups, we conducted GSVAscore analysis. The results showed significant differences in multiple pathways between high-risk and low-risk groups, with pathways such as MYC V2 targets, E2F targets, G2M checkpoint, and DNA repair enriched in the high-risk group (Figure 8D). We further constructed a correlation heatmap of these significant pathways, revealing potential functional connections between them (Figure 8E). Finally, we plotted a forest plot of Hazard Ratios for major pathways based on GSVAscore analysis, further highlighting the importance of these pathways in PDAC prognosis (Figure 8F).

### 3.8 Mutation characteristic analysis of epigenetic regulation genes in PDAC

To investigate the mutation characteristics of epigenetic regulation genes in PDAC, we conducted a detailed analysis of genomic data from high-risk and low-risk patients. First, we compared the mutation burden between high-risk and low-risk patients. The results showed no significant difference in mutation burden between high-risk and low-risk groups ( $p = 0.8$ , Figure 9A).

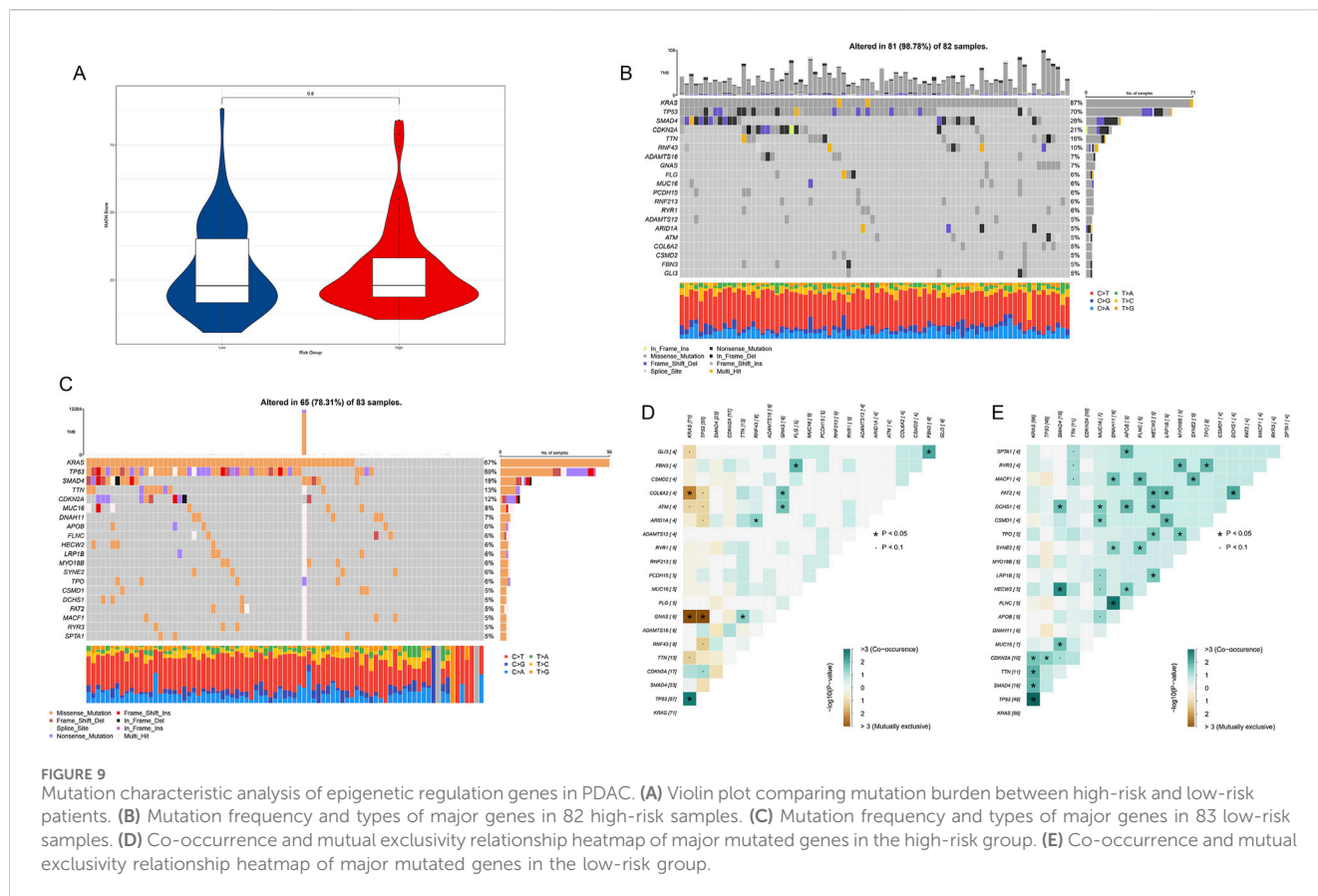
We further analyzed frequently mutated genes in high-risk and low-risk groups. The results showed that in 82 high-risk samples, 81 (98.78%) had at least one mutation in the analyzed genes (Figure 9B). Among them, KRAS, TP53, and SMAD4 were the

three genes with the highest mutation frequencies, occurring in 87%, 70%, and 28% of samples, respectively. In 83 low-risk samples, 65 (78.31%) had at least one mutation in the analyzed genes (Figure 9C). The genes with the highest mutation frequencies were highly consistent with previous analysis results, further confirming the importance of these genes in PDAC development.

We observed some significant gene mutation co-occurrence and mutual exclusivity patterns in both high-risk (Figure 9D) and low-risk groups (Figure 9E). In the high-risk group, KRAS and TP53 frequently co-occurred, while GNAS rarely appeared simultaneously with KRAS and TP53 in the same patient. In low-risk patients, the co-occurrence of KRAS and TP53 remained strong, while more genes showed co-occurrence patterns, and there were no obvious mutual exclusivity patterns.

### 3.9 Association analysis of epigenetic regulation and immune microenvironment in PDAC

To explore the relationship between epigenetic regulation and the immune microenvironment in PDAC, we conducted a series of immune-related analyses on high-risk and low-risk patients. First, we compared the Stromal Score, Immune Score, and ESTIMATE Score between the two groups. Results showed that low-risk patients had significantly higher Stromal Score (Figure 10A,  $p = 3.3e-15$ ),



Immune Score (Figure 10B,  $p = 3.3e-10$ ), and ESTIMATE Score (Figure 10C,  $p = 7.8e-14$ ) than high-risk patients, suggesting that low-risk patients may have richer tumor stromal components and immune cell infiltration.

Correlation analysis (Figure 10D) revealed significant associations between various immune cells and risk scores. Among them, M0 macrophages, memory B cells, follicular helper T cells, and Treg cells showed positive correlations with high risk scores, while CD4<sup>+</sup> active memory T cells, monocytes, and CD8<sup>+</sup> T cells showed positive correlations with low risk scores. The heatmap (Figure 10E) displayed differences in the activity of multiple immune-related pathways between high-risk and low-risk groups. The low-risk group generally showed higher immune pathway activity, consistent with previous immune score results. Finally, we compared the proportions of various immune cell subsets between the two groups (Figure 10F). Results showed significant differences in multiple immune cell subsets between high-risk and low-risk groups. For example, the proportions of CD8<sup>+</sup> T cells, monocytes, and M2 macrophages were significantly higher in the low-risk group, while the proportion of M0 macrophages was significantly increased in the high-risk group.

### 3.10 Drug sensitivity analysis of high-risk and low-risk PDAC patients

To investigate the potential impact of epigenetic regulation patterns on drug responses in PDAC patients, we conducted a series of drug sensitivity analyses on high-risk and low-risk

patients. We selected multiple drugs commonly used in PDAC treatment or clinical trials for evaluation.

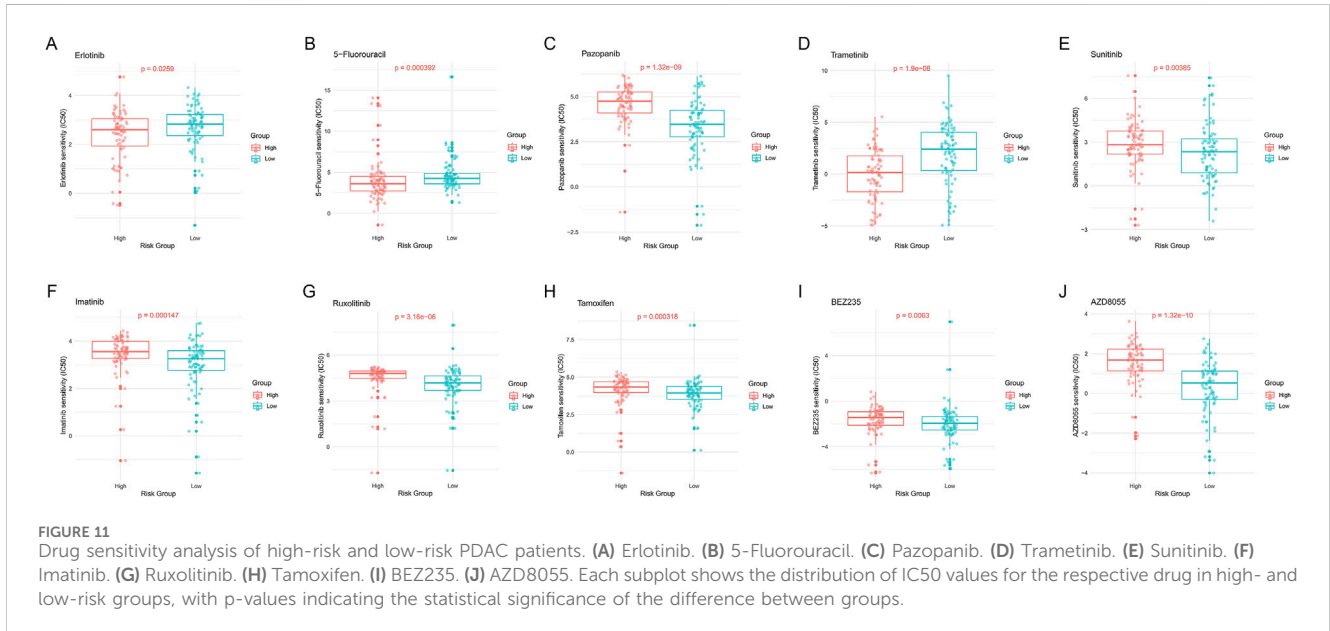
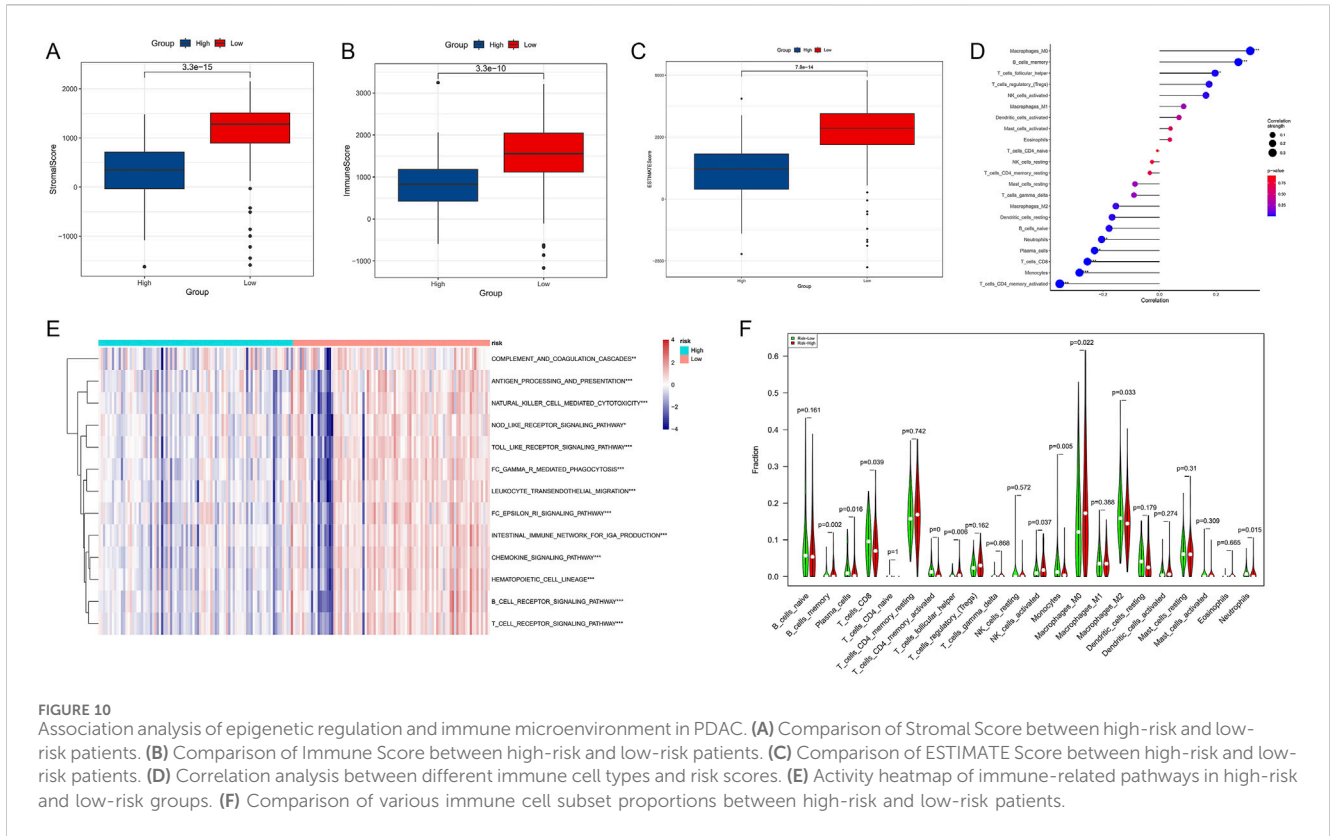
Erlotinib and Trametinib are EGFR and MEK inhibitors, respectively. Results showed that low-risk patients had significantly higher sensitivity to these drugs compared to high-risk patients ( $p = 0.0259$ , Figure 11A;  $p = 1.9e-06$ , Figure 11D). 5-Fluorouracil is a commonly used chemotherapy drug, and we similarly found that low-risk patients had significantly higher sensitivity to this drug ( $p = 0.000392$ , Figure 11B). This suggests that low-risk patients may be more suitable for EGFR and MEK targeted therapies, as well as 5-FU-based chemotherapy regimens.

Pazopanib and Sunitinib are both multi-target tyrosine kinase inhibitors. Analysis results showed that high-risk patients had significantly higher sensitivity to these drugs compared to low-risk patients ( $p = 1.32e-09$ , Figure 11C;  $p = 0.00385$ , Figure 11E). Imatinib, Ruxolitinib, Tamoxifen, BEZ235, and AZD8055 also showed similar differential patterns (Figures 11F–J). This indicates that these drugs may be more suitable for high-risk PDAC patients.

Overall, these drug sensitivity analysis results emphasize the close association between epigenetic regulation patterns and drug responses in PDAC patients, providing new perspectives and potential strategies for precision treatment of PDAC.

### 3.11 Pathological validation

We performed immunohistochemistry validation for the top five weighted genes: KRTCAP2, NENF, PSAP, MRPL41, and S100A16.

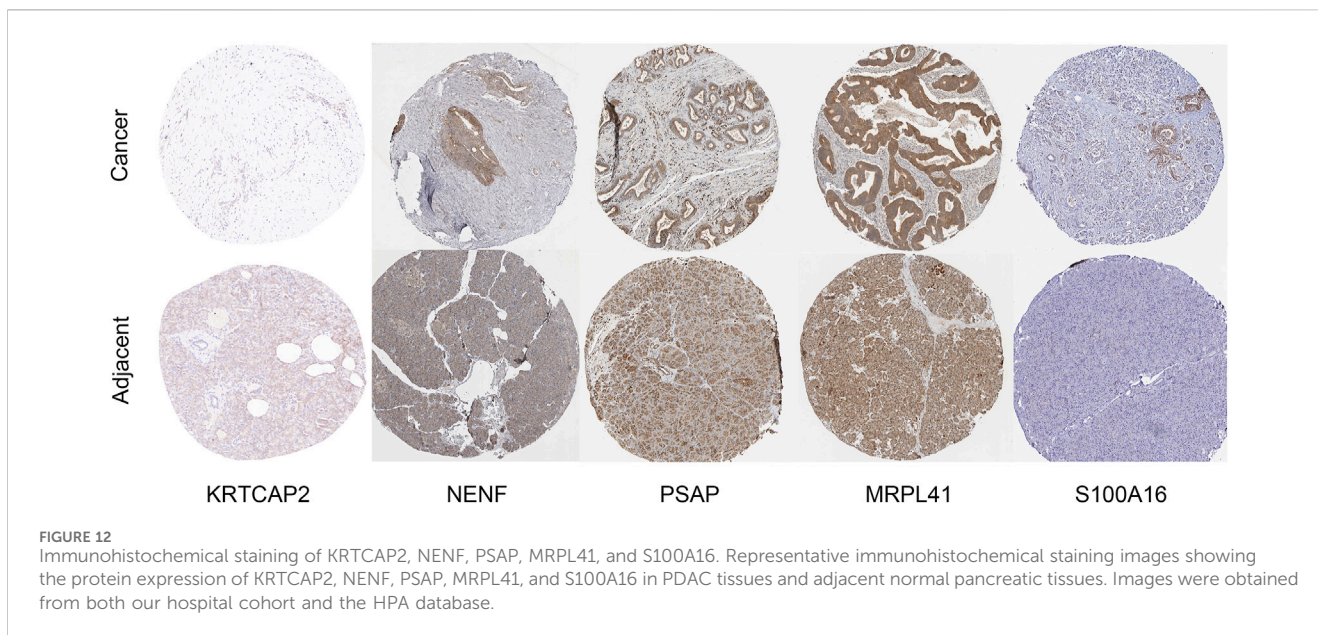


For KRTCAP2, we used patient samples from the hospital. For the remaining 4 genes, we cited results from the HPA database. In Figure 12, compared to adjacent tissue, KRTCAP2 are significantly downregulated in PDAC tissues, NENF, PSAP, MRPL41, and S100A16 are significantly upregulated in PDAC tissues.

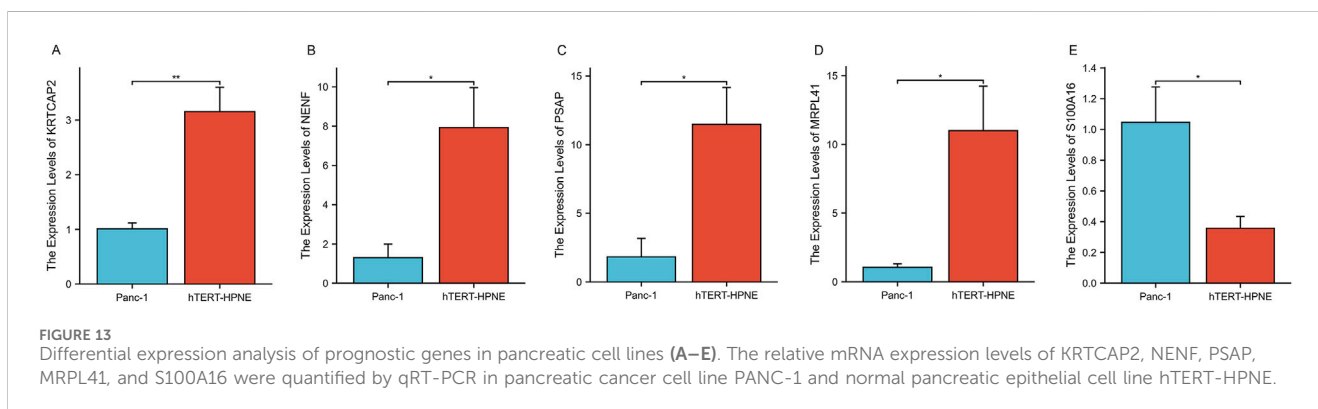
To further validate the expression patterns observed in tissue samples, we examined the mRNA levels of these genes in PDAC cell line PANC-1 and normal pancreatic ductal cell line hTERT-

HPNE using qRT-PCR (Figure 13). Consistent with the tissue results, KRTCAP2 showed significantly lower expression in PANC-1 cells compared to hTERT-HPNE cells (Figure 13A,  $p < 0.01$ ). Conversely, the expression levels of NENF, PSAP, MRPL41, and S100A16 were significantly higher in PANC-1 cells than in hTERT-HPNE cells (Figures 13B–E, all  $p < 0.01$ ). These findings in cell lines were in agreement with our observations in clinical specimens, further confirming the differential expression patterns of these genes in PDAC.





**FIGURE 12** Immunohistochemical staining of KRTCAP2, NENF, PSAP, MRPL41, and S100A16. Representative immunohistochemical staining images showing the protein expression of KRTCAP2, NENF, PSAP, MRPL41, and S100A16 in PDAC tissues and adjacent normal pancreatic tissues. Images were obtained from both our hospital cohort and the HPA database.



**FIGURE 13** Differential expression analysis of prognostic genes in pancreatic cell lines (A–E). The relative mRNA expression levels of KRTCAP2, NENF, PSAP, MRPL41, and S100A16 were quantified by qRT-PCR in pancreatic cancer cell line PANC-1 and normal pancreatic epithelial cell line hTERT-HPNE.

## 4 Discussion

This study comprehensively investigated the role and clinical significance of epigenetic regulation in PDAC by integrating single-cell RNA sequencing, epigenetic analysis, and machine learning methods (Wang Q. et al., 2023; Li et al., 2024). We first performed single-cell RNA sequencing analysis on PDAC and adjacent normal tissues, revealing cellular heterogeneity and compositional changes in the tumor microenvironment. Subsequently, we analyzed histone modifications and DNA methylation characteristics at the single-cell level, discovering that cells in PDAC samples generally exhibited higher levels of epigenetic modifications. Through WGCNA, we identified key gene modules highly correlated with histone modifications and DNA methylation, and conducted functional enrichment analysis. Based on key epigenetic-related genes, we constructed a machine learning model for PDAC prognosis prediction. Additionally, we analyzed the associations between epigenetic features and gene mutation patterns, immune microenvironment, and drug sensitivity.

Overall, the findings of this study not only deepen our understanding of PDAC molecular mechanisms but also provide

new perspectives for PDAC diagnosis, prognosis assessment, and treatment. Epigenetic regulation plays a central role in the occurrence and development of PDAC, influencing multiple stages from tumor initiation to progression (Montalvo-Javé et al., 2023). Our research reveals that PDAC cells generally exhibit high levels of histone modifications and DNA methylation activity, which may confer stronger adaptability and survival advantages to tumor cells. Meanwhile, abnormal epigenetic regulation may also be one of the important factors driving PDAC tumor heterogeneity (Espinet et al., 2022). Different subclones may maintain their unique phenotypic and functional characteristics through specific epigenetic regulatory patterns, thereby promoting the overall adaptability of the tumor (Orlacchio et al., 2024). Furthermore, our analysis revealed significant immune microenvironment differences between risk groups, with the low-risk group showing higher immune and stromal scores, particularly increased infiltration of CD8<sup>+</sup> T cells and M2 macrophages, suggesting stronger anti-tumor immune responses. These findings have crucial clinical implications, providing not only a theoretical basis for differential immunotherapy responses but also new evidence for immunotherapy strategy selection. For instance, high-risk patients

might benefit from additional immune modulatory treatments, particularly immune checkpoint inhibitors, given their higher proportions of immunosuppressive cells. Notably, the association between immune microenvironment differences and epigenetic regulation suggests that epigenetic alterations may influence disease progression by modulating the immune microenvironment, opening new avenues for therapeutic strategies that target epigenetic regulation to enhance anti-tumor immunity (Wang et al., 2024; Zhou et al., 2024; Ding et al., 2024).

This study has validated previous epigenetic research on PDAC in multiple aspects while also presenting some innovative findings. Consistent with previous studies, our research reaffirms the importance of epigenetic regulation in PDAC development, aligning with earlier research results. For instance, we observed generally high levels of histone modifications and DNA methylation in PDAC cells, which is consistent with Lomberk et al.'s findings that KRAS mutations can promote PDAC progression by regulating histone and DNA modifications (Lomberk et al., 2019). The innovative aspects of our study are mainly reflected in the following areas: First, we employed single-cell RNA sequencing technology, enabling us to study PDAC epigenetic characteristics at the single-cell level, revealing similarities and differences in epigenetic regulation across different cell types. This high-resolution analysis method provides a new perspective for understanding PDAC cellular heterogeneity. Second, by integrating various bioinformatics methods, including hdWGCNA and WGCNA, we systematically identified key epigenetic regulatory genes and pathways. This comprehensive analysis approach allows us to understand more fully the role of epigenetic regulatory networks in PDAC. Lastly, we innovatively linked epigenetic features with the immune microenvironment and drug sensitivity, providing new insights for PDAC immunotherapy and personalized medication.

Compared to existing PDAC prognostic models, our epigenetic feature-based prognostic model has several significant advantages. First, most previous PDAC prognostic models are primarily based on gene mutations or transcriptome data (Chen et al., 2021; Zhang et al., 2015), while our model focuses on epigenetic features, providing a new dimension for prognosis assessment. The importance of epigenetic regulation in tumor progression is increasingly recognized, so our model may capture some important information overlooked by traditional models. Second, our model demonstrates good predictive performance across multiple independent datasets, with AUC values exceeding 0.7 for 1-year, 3-year, and 5-year prognosis predictions. The stability and generalizability of our model highlight the potential of epigenetic features in PDAC prognosis prediction. Additionally, our model not only provides prognostic predictions but also combines them with immune microenvironment and drug sensitivity analyses, giving it greater potential in guiding individualized treatment decisions (Lomberk et al., 2016). In contrast, many existing models primarily focus on prognosis prediction, lacking direct applications for treatment guidance.

Although this study provides many valuable findings, some limitations remain. First, our analysis is primarily based on bioinformatics methods, requiring further experimental validation and mechanistic exploration. Second, while our prognostic model

performs well, it still needs validation in larger independent cohorts. Most importantly, the drug sensitivity tests include many drugs not yet clinically applied, necessitating pharmacological research and clinical verification. Future research should focus on in-depth study of the causal relationship between epigenetic regulation and PDAC progression, integrate multi-omics data to construct a more comprehensive PDAC molecular typing system, and investigate the dynamic relationships between epigenetic regulation and PDAC immune microenvironment and drug responses. Based on this, large-scale generalization studies and drug application exploration integrating multi-center data should be conducted.

In conclusion, this study has made significant contributions to epigenetic research in PDAC through multi-level, multi-faceted analysis. We revealed cellular heterogeneity and epigenetic characteristics of the PDAC microenvironment at the single-cell level, systematically identified key epigenetic regulatory genes and pathways, and constructed a high-performance prognostic prediction model based on epigenetic features. Additionally, we uncovered close connections between epigenetic regulation and PDAC mutation characteristics, immune microenvironment, and drug sensitivity. These findings not only deepen our understanding of PDAC molecular mechanisms but also provide new insights for precision medicine, individualized treatment, and immunotherapy optimization. Our research opens new directions for PDAC diagnosis, prognosis assessment, and treatment strategies, potentially advancing clinical practice and ultimately improving patient prognosis and quality of life. However, translating these findings into clinical applications requires further validation and large-scale prospective studies. Overall, this study lays an important foundation for epigenetic research and precision medicine development in PDAC, demonstrating the enormous potential of epigenetics in cancer research.

## Data availability statement

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

## Author contributions

KF: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing. JS: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Visualization, Writing—review and editing. YZ: Data curation, Formal Analysis, Investigation, Methodology, Resources, Visualization, Writing—review and editing. XC: Conceptualization, Data curation, Formal Analysis, Resources, Software, Supervision, Writing—review and editing. XH: Conceptualization, Formal Analysis, Funding acquisition, Methodology, Resources, Supervision, Writing—review and editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the National Natural Science Foundation of China (Project approval number 52075277).

## Acknowledgments

We are so grateful for the selfless help to Zheng Ding from first hospital of China Medical University, his coding experience and thinking system improve the quality of our research greatly.

## References

- Amato, E., Barbi, S., Fassan, M., Luchini, C., Vicentini, C., Brunelli, M., et al. (2016). RASSF1 tumor suppressor gene in pancreatic ductal adenocarcinoma: correlation of expression, chromosomal status and epigenetic changes. *BMC Cancer* 16, 11. doi:10.1186/s12885-016-2048-0
- Arashi, M., Roozbeh, M., Hamzah, N. A., and Gasparini, M. (2021). Ridge regression and its applications in genetic studies. *PLoS One* 16 (4), e0245376. doi:10.1371/journal.pone.0245376
- Ayyadevara, V. K., and Ayyadevara, V. K. (2018). *Gradient boosting machine*. Pro machine learning algorithms: a hands-on approach to implementing algorithms in python and R, 117–134.
- Bailey, P., Chang, D. K., Nones, K., Johns, A. L., Patch, A. M., Gingras, M. C., et al. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* 531 (7592), 47–52. doi:10.1038/nature16965
- Bair, E., Hastie, T., Paul, D., and Tibshirani, R. (2006). Prediction by supervised principal components. *J. Am. Stat. Assoc.* 101 (473), 119–137. doi:10.1198/016214505000000628
- Bertrand, F., Bastien, P., Meyer, N., and Maumy-Bertrand, M. (2014). “plsRcox, Cox-Models in a high dimensional setting in R.”. *Fréd. Bert. and Myr. Mau. Bert.*
- Bertrand, F., Maumy-Bertrand, M., Bertrand, M. F., and Bertrand, M. (2022). Package ‘plsRcox’.
- Bi, X., Wang, J., and Liu, C. (2024). Intratumoral microbiota: metabolic influences and biomarker potential in gastrointestinal cancer. *Biomolecules* 14 (8), 917. doi:10.3390/biom14080917
- Binder, H., and Binder, M. H. (2015). Package ‘CoxBoost’: citeuser.
- Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* 10 (5), 295–304. doi:10.1038/nrg2540
- Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1\_12
- Chen, Y., Feng, Y., Yan, F., Zhao, Y., Zhao, H., and Guo, Y. (2022). A novel immune-related gene signature to identify the tumor microenvironment and prognose disease among patients with oral squamous cell carcinoma patients using ssGSEA: a bioinformatics and biological validation study. *Front. Immunol.* 13, 922195. doi:10.3389/fimmu.2022.922195
- Chen, Z., Han, C., Zhou, X., Wang, X., Liao, X., He, Y., et al. (2021). Prognostic value and potential molecular mechanism of the like-Sm gene family in early-stage pancreatic ductal adenocarcinoma. *Transl. Cancer Res.* 10 (4), 1744–1760. doi:10.21037/tcr-20-3056
- Cho, S., Kim, H., Oh, S., Kim, K., and Park, T. (2009). Elastic-net regularization approaches for genome-wide association studies of rheumatoid arthritis. *BMC Proc.* 3 (7), S25. doi:10.1186/1753-6561-3-s7-s25
- Cristian, P. M., Aarón, V. J., Armando, E. D., Estrella, M. L. Y., Daniel, N. R., David, G. V., et al. (2024). Diffusion on PCA-UMAP manifold: the impact of data structure preservation to denoise high-dimensional single-cell RNA sequencing data. *Biol. (Basel)*. 13 (7), 512. doi:10.3390/biology13070512
- Csardi, G., and Tamas, N. (2005). The igraph software package for complex network research. *InterJournal Complex Syst.* 1695.
- Ding, Z., Chen, J., Li, B., and Ji, X. (2024). Inflammatory factors and risk of lung adenocarcinoma: a Mendelian randomization study mediated by blood metabolites. *Front. Endocrinol. (Lausanne)* 15, 1446863. doi:10.3389/fendo.2024.1446863

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Epinet, E., Klein, L., Puré, E., and Singh, S. K. (2022). Mechanisms of PDAC subtype heterogeneity and therapy response. *Trends Cancer* 8 (12), 1060–1071. doi:10.1016/j.trecan.2022.08.005
- Füllgrabe, J., Kavanagh, E., and Joseph, B. (2011). Histone onco-modifications. *Oncogene* 30 (31), 3391–3403. doi:10.1038/ncr.2011.121
- Geeleher, P., Cox, N., and Huang, R. S. (2014). pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* 9 (9), e107468. doi:10.1371/journal.pone.0107468
- Goodwin, C. M., Waters, A. M., Klomp, J. E., Javadi, S., Bryant, K. L., Stalneck, C. A., et al. (2023). Combination therapies with CDK4/6 inhibitors to treat KRAS-mutant pancreatic cancer. *Cancer Res.* 83 (1), 141–157. doi:10.1158/0008-5472.CAN-22-0391
- Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7
- Ishwaran, H., Kogalur, U. B., Blackstone, E. H., and Lauer, M. S. (2008). Random survival forests. *Ann. Appl. Stat.* 2. doi:10.1214/08-aoas169
- Lai, E., Ziranu, P., Spanu, D., Dubois, M., Pretta, A., Tolu, S., et al. (2021). BRCA-mutant pancreatic ductal adenocarcinoma. *Br. J. Cancer* 125 (10), 1321–1332. doi:10.1038/s41416-021-01469-9
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559
- Lee, C. J., Ahn, H., Jeong, D., Pak, M., Moon, J. H., and Kim, S. (2020). Impact of mutations in DNA methylation modification genes on genome-wide methylation landscapes and downstream gene activations in pan-cancer. *BMC Med. Genomics* 13 (Suppl. 3), 27. doi:10.1186/s12920-020-0659-4
- Li, Z., Wang, Q., Huang, X., Fu, R., Wen, X., and Zhang, L. (2024). Multi-omics analysis reveals that ferroptosis-related gene C1SD2 is a prognostic biomarker of head and neck squamous cell carcinoma. *J. Gene Med.* 26 (1), e3580. doi:10.1002/jgm.3580
- Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004
- Liu, F., Wang, L., Perna, F., and Nimer, S. D. (2016). Beyond transcription factors: how oncogenic signalling reshapes the epigenetic landscape. *Nat. Rev. Cancer* 16 (6), 359–372. doi:10.1038/nrc.2016.41
- Liu, R., Huang, B., Shao, Y., Cai, Y., Liu, X., and Ren, Z. (2023). Identification of memory B-cell-associated miRNA signature to establish a prognostic model in gastric adenocarcinoma. *J. Transl. Med.* 21 (1), 648. doi:10.1186/s12967-023-04366-2
- Lomberk, G., Dusetti, N., Iovanna, J., and Urrutia, R. (2019). Emerging epigenomic landscapes of pancreatic cancer in the era of precision medicine. *Nat. Commun.* 10 (1), 3875. doi:10.1038/s41467-019-11812-7
- Lomberk, G. A., Iovanna, J., and Urrutia, R. (2016). The promise of epigenomic therapeutics in pancreatic cancer. *Epigenomics* 8 (6), 831–842. doi:10.2217/epi-2015-0016
- Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28 (11), 1747–1756. doi:10.1101/gr.239244.118
- Mizrahi, J. D., Surana, R., Valle, J. W., and Shroff, R. T. (2020). Pancreatic cancer. *Pancreat. cancer. Lancet.* 395 (10242), 2008–2020. doi:10.1016/S0140-6736(20)30974-0
- Montalvo-Javé, E. E., Nuño-Lámbarrí, N., López-Sánchez, G. N., Ayala-Moreno, E. A., Gutiérrez-Reyes, G., Beane, J., et al. (2023). Pancreatic cancer: genetic conditions and epigenetic alterations. *J. Gastrointest. Surg.* 27 (5), 1001–1010. doi:10.1007/s11605-022-05553-0

- Morabito, S., Reese, F., Rahimzadeh, N., Miyoshi, E., and Swarup, V. (2023). hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. *Cell Rep. Methods* 3 (6), 100498. doi:10.1016/j.crmeth.2023.100498
- Nakaoka, K., Ohno, E., Kawabe, N., Kuzuya, T., Funasaka, K., Nakagawa, Y., et al. (2023). Current status of the diagnosis of early-stage pancreatic ductal adenocarcinoma. *Diagn. (Basel)* 13 (2), 215. doi:10.3390/diagnostics13020215
- Nishiyama, A., and Nakanishi, M. (2021). Navigating the DNA methylation landscape of cancer. *Trends Genet.* 37 (11), 1012–1027. doi:10.1016/j.tig.2021.05.002
- Orlacchio, A., Muzyka, S., and Gonda, T. A. (2024). Epigenetic therapeutic strategies in pancreatic cancer. *Int. Rev. Cell Mol. Biol.* 383, 1–40. doi:10.1016/bs.ircmb.2023.10.002
- Pripp, A. H. (2018). Pearson's or Spearman's correlation coefficients. *Tidsskr. Nor. Laegeforen* 138 (8). doi:10.4045/tidsskr.18.0042
- Ranstrom, J., and Cook, J. A. (2018). LASSO regression. *Br. J. Surg.* 105 (10), 1348. doi:10.1002/bjs.10895
- Rawla, P., Sunkara, T., and Gaduputi, V. (2019). Epidemiology of pancreatic cancer: global trends, etiology and risk factors. *World J. Oncol.* 10 (1), 10–27. doi:10.14740/wjon1166
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007
- Schneider, G., Krämer, O. H., Fritsche, P., Schüller, S., Schmid, R. M., and Saur, D. (2010). Targeting histone deacetylases in pancreatic ductal adenocarcinoma. *J. Cell Mol. Med.* 14 (6A), 1255–1263. doi:10.1111/j.1582-4934.2009.00974.x
- Sebaugh, J. L. (2011). Guidelines for accurate EC50/IC50 estimation. *Pharm. Stat.* 10 (2), 128–134. doi:10.1002/pst.426
- Shah, A., Ganguly, K., Rauth, S., Sheree, S. S., Khan, I., Ganti, A. K., et al. (2024). Unveiling the resistance to therapies in pancreatic ductal adenocarcinoma. *Drug Resist Updat* 77, 101146. doi:10.1016/j.drug.2024.101146
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., et al. (2019). Comprehensive integration of single-cell data. *Cell* 177 (7), 1888–1902. doi:10.1016/j.cell.2019.05.031
- Timmer, F. E. F., Geboers, B., Nieuwenhuizen, S., Dijkstra, M., Schouten, E. A. C., Puijk, R. S., et al. (2021). Pancreatic cancer and immunotherapy: a clinical overview. *Cancers (Basel)* 13 (16), 4138. doi:10.3390/cancers13164138
- Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol. Pozn.* 19 (1a), A68–A77. doi:10.5114/wo.2014.47136
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347 (6220), 1260419. doi:10.1126/science.1260419
- Van Belle, V., Pelckmans, K., Van Huffel, S., and Suykens, J. A. (2011). Improved performance on high-dimensional survival data by application of Survival-SVM. *Bioinformatics* 27 (1), 87–94. doi:10.1093/bioinformatics/btq617
- Versemann, L., Patil, S., Steuber, B., Zhang, Z., Kopp, W., Krawczyk, H. E., et al. (2022). TP53-Status-Dependent oncogenic EZH2 activity in pancreatic cancer. *Cancers (Basel)* 14 (14), 3451. doi:10.3390/cancers14143451
- Wang, Q., Liu, Y., Li, Z., Tang, Y., Long, W., Xin, H., et al. (2023b). Establishment of a novel lysosomal signature for the diagnosis of gastric cancer with *in-vitro* and *in-situ* validation. *Front. Immunol.* 14, 1182277. doi:10.3389/fimmu.2023.1182277
- Wang, X., Liu, X., Xiao, R., Fang, Y., Zhou, F., Gu, M., et al. (2024). Histone lactylation dynamics: unlocking the triad of metabolism, epigenetics, and immune regulation in metastatic cascade of pancreatic cancer. *Cancer Lett.* 598, 217117. doi:10.1016/j.canlet.2024.217117
- Wang, X., Yang, J., Ren, B., Yang, G., Liu, X., Xiao, R., et al. (2023a). Comprehensive multi-omics profiling identifies novel molecular subtypes of pancreatic ductal adenocarcinoma. *Genes Dis.* 11 (6), 101143. doi:10.1016/j.gendis.2023.101143
- Wilson, J. S., Pirola, R. C., and Apte, M. V. (2014). Stars and stripes in pancreatic cancer: role of stellate cells and stroma in cancer progression. *Front. Physiol.* 5, 52. doi:10.3389/fphys.2014.00052
- Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 41 (Database issue), D955–D961. doi:10.1093/nar/gks1111
- Zeng, D., Ye, Z., Shen, R., Yu, G., Wu, J., Xiong, Y., et al. (2021). IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures. *Front. Immunol.* 12, 687975. doi:10.3389/fimmu.2021.687975
- Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., et al. (2019). The international cancer genome Consortium data portal. *Nat. Biotechnol.* 37 (4), 367–369. doi:10.1038/s41587-019-0055-9
- Zhang, Z. L., Bai, Z. H., Wang, X. B., Bai, L., Miao, F., and Pei, H. H. (2015). miR-186 and 326 predict the prognosis of pancreatic ductal adenocarcinoma and affect the proliferation and migration of cancer cells. *PLoS One* 10 (3), e0118814. doi:10.1371/journal.pone.0118814
- Zhou, Z., Wang, J., Wang, J., Yang, S., Wang, R., Zhang, G., et al. (2024). Deciphering the tumor immune microenvironment from a multidimensional omics perspective: insight into next-generation CAR-T cell immunotherapy and beyond. *Mol. Cancer* 23 (1), 131. doi:10.1186/s12943-024-02047-2
- Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 67 (2), 301–320. doi:10.1111/j.1467-9868.2005.00503.x