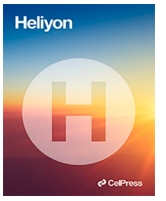
Heliyon 10 (2024) e34284

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

A five-gene prognosis model based on lysine β-hydroxybutyrylation site genes to predict the survival and therapy response in pancreatic adenocarcinoma

Fangfang Hu [a], Zhibin Bai [b], Kai Yan [a], Zheng Zhang [c], Jiahua Zhou [a,*]

[a] Department of Hepatobiliary and Pancreatic Surgery, Zhongda Hospital, Medical School, Southeast University, Nanjing, Jiangsu, 210009, China
[b] Center of interventional Radiology and Vascular Surgery, Department of Radiology, Zhongda Hospital, Medical School, Southeast University, Nanjing, Jiangsu, 210009, China
[c] Medical School, Southeast University, Nanjing, Jiangsu, 210009, China

ARTICLE INFO

Keywords:
Pancreatic adenocarcinoma
Butyrylation
Lysine β-hydroxybutyrylation
Prognosis
Metabolism

ABSTRACT

Background: Pancreatic adenocarcinoma (PAAD) is one of the most malignancy diseases. Lysine β-hydroxybutyrylation (Kbhb) has been reported to involve various metabolism and cancer progression.
Methods: Data from online databases (TCGA and GEO) were retrieved for the selection of differential expressed Kbhb site genes (DTRGs). Univariate cox and LASSO analysis were performed to identify the prognostic DTRGs. Based on the optimal DTRGs, a prognostic risk score model was established. Kaplan-Meier and Receiver operator characteristic analysis were conducted to evaluate the predicting ability of the prognosis model. Generated with clinical data, independent analysis and nomogram model were performed. Finally, the differences of survival, immune cell levels, immunotherapy response, drug sensitivity between high- and low-risk groups were explored.
Results: A total of 63 DTRGs were identified in PAAD, and these genes were related to cell division and apoptosis biological functions. Through univariate cox regression and LASSO analysis, 30 DTRGs were selected to be related to prognosis and five (*KRT18*, *ANLN*, *ECT2*, *RBM5*, and *RBM6*) were identified as the optimal DTRGs in PAAD. Based on the five optimal DTRGs, a prognostic risk score model was constructed, with promising predictive ability in PAAD survival (AUC >0.70). High-risk group showed lower survival rate ($P < 0.05$). Moreover, based on the risk score, a nomogram model was also established, which possessed perfect stability. Finally, lower risk score was related to higher immune cell levels, indicating an immune activation in low-risk status, which maybe the reason for the better survival in low-risk group. Furthermore, the immunotherapy response and drug sensitivity were all higher than that in low-risk groups ($P < 0.05$).
Conclusion: A five-gene prognosis risk model which exhibit promising predictive ability in survival is constructed for patients with PAAD.

* Corresponding author. Department of Hepatobiliary and Pancreatic Surgery, Zhongda Hospital, Medical School, Southeast University, No. 87 Dingjiaqiao, Gulou District, Nanjing, Jiangsu, 210009, China.
E-mail addresses: janovewaldner@163.com (F. Hu), baizhibin1004@sina.com (Z. Bai), yankai5212@163.com (K. Yan), chervymedic@163.com (Z. Zhang), zhoujh@seu.edu.cn (J. Zhou).

https://doi.org/10.1016/j.heliyon.2024.e34284
Received 15 March 2024; Received in revised form 30 June 2024; Accepted 7 July 2024
Available online 8 July 2024

1. Introduction

Pancreatic adenocarcinoma (PAAD) is a highly malignancy disease [1]. The common risk factors of PAAD are gallstones, smoking, alcohol consume, and chronic pancreatitis [2]. Most PAAD arise from microscopic non-invasive epithelial proliferations within the pancreatic ducts, considered as pancreatic intraepithelial neoplasias [3]. PAAD is commonly detected at advanced stage, about 80–85 % of the patients missed optimal surgery period due to absent of typical symptom [4]. Surgery resection is the only potentially curative method, while for patients received surgery resection, the five-year survival is only 20 % [3]. Early-stage PAAD manifests as rapid progression and early metastasis, thereby inducing poor prognosis [4]. An effective prediction tool is necessary for the prognosis and management of PAAD.

Tumor metabolic reprogramming is an important hallmark of cancer, and alteration of metabolism can induce immune escape [5]. Various genes-related to metabolism have been reported as oncogenes or suppressor genes in cancer [6–8]. Lysine β-hydroxybutyrylation (Kbhb) is a novel histone posttranslational modification and is first reported by Xie et al., in 2016 [9]. Kbhb has been reported to involve various metabolism and cancer progression [9,10]. In lung adenocarcinoma, BDH1-mediated LRRC31 regulation dependent on histone Kbhb to promote tumor progression [11]. p53 is a common tumor suppressor, which is also reported to be modified by Kbhb. p53 Kbhb decreases p53 acetylation and downstream genes expression like p21 and PUMA, and thus inhibit apoptosis [12]. These evidences indicate the association of Kbhb with cancer progression, whereas, role of Kbhb in PAAD has not be investigated.

Prognostic models based on the gene chips and high-throughput sequencing emerged recently [13,14]. Nevertheless, the prognostic signature based on the Kbhb have not been reported. In our analyses, data from online databases were involved in a bioinformatic analysis, and prognostic model related to Kbhb genes was mainly established. Moreover, based on the expression of the Kbhb genes, all the samples were divided into high- and low-risk groups to investigate differences of immune response, drug sensitivity, and mutation features between the two risk groups. Five optimal genes (*KRT18*, *ANLN*, *ECT2*, *RBM5*, and *RBM6*) were identified in PAAD. And based on the five optimal genes, a prognostic risk score model was constructed, with a promising predictive ability in PAAD survival. High-risk group showed lower survival rate, lower immune cell levels, higher immunotherapy response and higher drug sensitivity.

2. Methods

2.1. Data source

Data of gene expression, survival information, and clinical feature of PAAD were downloaded from TCGA-PAAD dataset in UCSC Xena database (http://xena.ucsc.edu/). A total of 177 cancer samples and four normal samples in TCGA-PAAD were involved in the training analysis. Sixty-six PAAD samples with survival information from GSE62452 dataset in Gene Expression Omnibus (GEO https://www.ncbi.nlm.nih.gov/) database was also retrieved for the validation analysis. Moreover, the list of 1493 Kbhb site-related genes (TRGs) were downloaded from a previous study [15].

2.2. Differential expression analysis

Based on the gene expression data from TCGA-PAAD dataset, differential expressed genes (DEGs) between cancer and normal samples were identified using limma package (version 3.54.0) [16] in R. The selection threshold was set as $P < 0.05$ & $|\log_2 \text{fold change (FC)}| > 0.5$. Following, the DEGs and TRGs were intersected using VennDiagram package (version 1.7.3) [17] in R, and overlapping genes were identified as differential Kbhb-related genes (DTRGs).

2.3. Functional analysis and protein-protein interaction (PPI) networks analysis

To investigate the biological function and pathways involved in TRGs, the function of the DTRGs were investigated in an online website DAVID (version 6.8.0, https://david.ncifcrf.gov/tools.jsp), containing Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. To investigate the connection of the DTRGs, protein-protein interaction (PPI) was analyzed in STRING database (version 11.0, https://string-db.org/) and PPI network was constructed using Cytoscape software.

2.4. Prognostic analysis

Combined with the expression of DTRGs and survival information of cancer patients in TCGA-PAAD dataset, the prognosis-related DTRGs were further restrained by using univariate cox regression analysis in survival package (version 3.2–7) [18] in R. Based on the prognostic DTRGs, least absolute shrinkage and selection operator (LASSO) regression analysis was used to select the optimal DTRGs and construct the survival risk score model in TCGA-PAAD dataset. The risk core model was established as the following formula:

$$\text{Risk score} = \sum_{i=1}^{n} Coef_i \times Exp_i$$

$Coef_i$ represents the risk coefficient; Exp_i represent the expression value.

The risk score of each sample was calculated according to the above formula. Based on the median of the risk score, all the samples in TCGA-PAAD dataset were divided into two groups: high-and low-risk groups. Kaplan-Meier (KM) analysis was used to investigated

the survival difference between the two risk groups. Receiver operator characteristic (ROC) curve was used to evaluate the robustness of the risk score model using timeROC package (version 0.4) in R. Moreover, the model was validated in GSE62452 validation dataset.

2.5. Independent prognostic analysis

Univariate and multivariate COX regression analysis were used to perform independent prognostic analysis in TCGA dataset, and then RMS package (version 6.2–0) [19] was used to construct a nomogram model for predicting the survival of patients in clinic. The calibration curve, decision curve analysis (DCA) curve, and ROC curve were used to evaluate the predictive ability of the nomogram.

2.6. Gene Set Variation Analysis (GSVA) analysis

GSVA connected the gene expression and pathway change. By downloading gene sets from the MsigDB, ssGSEA algorithm was used to score each KEGG pathway enrichment. limma package (version 3.54.0) was performed to compare the differential KEGG pathways between high- and low-risk groups.

2.7. Immune infiltration, immune checkpoint, immunotherapy response, and drug sensitivity analysis

Estimate package [20] in R was used to investigate the differences of immune, stromal, and ESTIMATE score between different risk groups. ssGSEA was used to evaluate the immune infiltration levels between two risk groups. As for immune checkpoint analysis, firstly, the expression level of immune checkpoint was obtained from TCGA database, and Wilcox rank-sum test in stats package (version 4.2.2) in R was used to analyze the differential of immune checkpoint between two risk groups. Moreover, Tumor Immune Dysfunction and Exclusion (TIDE) analysis tool (http://tide.dfci.harvard.edu/) was used to evaluate the immunotherapy response between two risk groups and Wilcox rank-sum test was used to compare the differences of immunotherapy response between different groups. To evaluate the drug sensitivity, 138 drugs in GDSC database (https://www.cancerrxgene.org/) were downloaded. pRRophetic package (version 0.5) [21] was used to predict the IC50 value of the drugs for each patient based on the gene expression levels.

2.8. Mutation frequency analysis

Based on cBioPortal database (https://www.cbioportal.org/), the domain and the mutation sites of each optimal DTRG was further analyzed, and the mutation frequency in each of the DTRG in patients was evaluated. Moreover, KM curve was analyzed to investigate the survival differences between mutation group and non-mutation group using the survival option in the cBioPortal database.

2.9. Statistical analysis

All the analysis in this study were performed in R software (version 4.2.2, R Foundation for Statistical Computing, Vienna, Austria). $P < 0.05$ was set as statistical significance. Wilcox rank-sum test was used to evaluate the differences between two risk groups. Spearman method was used to evaluate the correlation between two groups (risk score and ESTIMATE score; genes and risk score; genes and immune cells; risk score and drugs). Pearson analysis was used for investigation of the correlation among the five genes and 63 genes in PPI network. The Benjamini & Hochberg procedure was used to correct for multiple testing.

The study design is exhibited in Fig. 1.

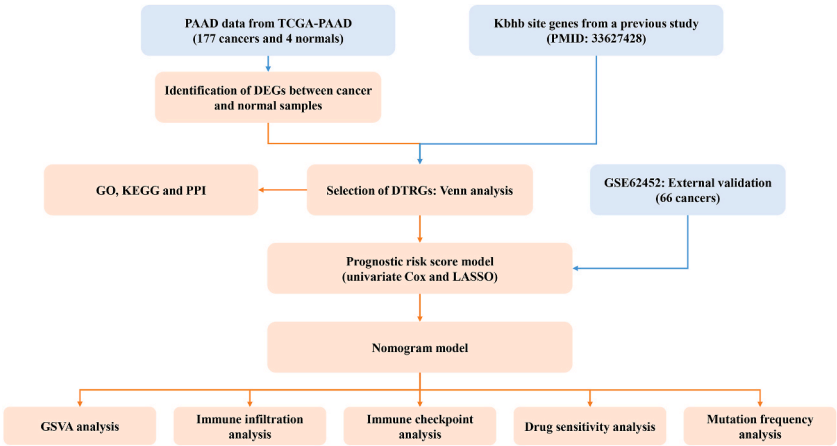


Fig. 1. The study design. PAAD, pancreatic adenocarcinoma; TCGA, the Cancer Genome Atlas Program; DEGs, differential expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; DTRGs, differential Kbhb-related genes; LASSO, least absolute shrinkage and selection operator; GSVA, Gene Set Variation Analysis.

3. Results

3.1. Identification of DTRGs

A total of 1594 DEGs were obtained between tumor and normal samples (Fig. 2A). Through intersecting these DEGs with the 1493 TRGs from previous study [15], a total of 63 overlapping DTRGs were involved into the functional analysis (Fig. 2B). The functional analysis suggested that these DTRGs were mainly enriched into 23 GO biological process (BP), 22 GO cellular components (CC), nine GO molecular functions (MF), and one KEGG pathways. The TOP 5 functions were visualized in Fig. 2C, manifesting as that the 63 DTRGs were mainly involved in cell division and apoptosis-related biological function. PPI network indicated that the 63 DTRGs were significantly correlated, exerting as 110 protein interaction relationship pairs and 43 nodes were concluded in PPI network (Fig. 2D). Among which, H3F3C, HIST2H2BE, PLK1 were closely related to most of the proteins.

3.2. Identification of optimal DTRGs and construction of the prognostic model

In TCGA-PAAD training dataset, multivariate cox regression analysis identified 30 DTRGs that were related to prognosis of patients with PAAD (Fig. 3A, $P < 0.05$). Following, LASSO analysis generated the genes combination into five: *KRT18*, *ANLN*, *ECT2*, *RBM5*, and

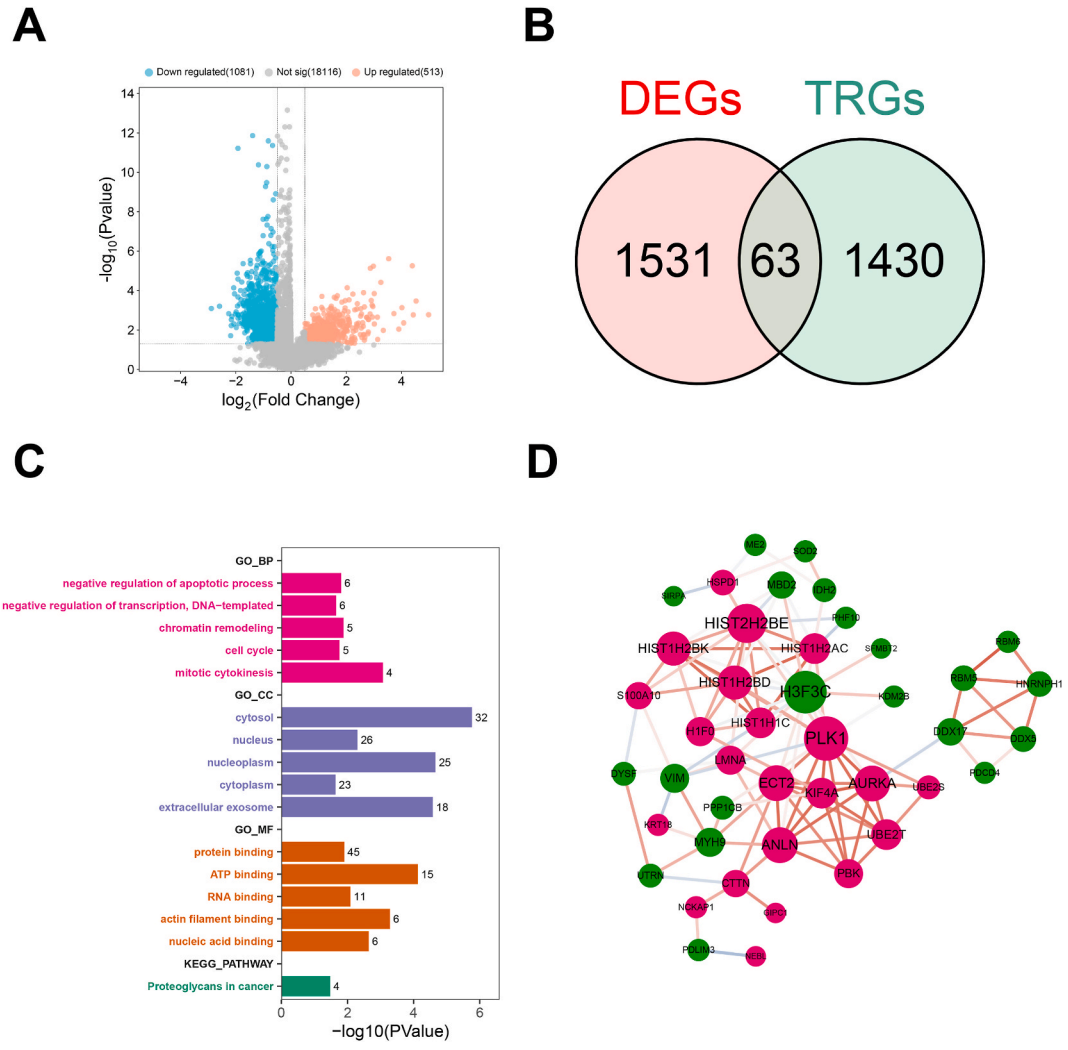


Fig. 2. Identification of the DTRGs. (A) Volcano plot for identification of DEGs between cancer and normal groups. (B) Venn analysis. (C) GO and KEGG results in TOP5. The length of the bar indicates significance; the numbers behind the bats indicate the number of enriched genes. (D) Protein-protein interaction (PPI) networks. Lines indicate the interaction between genes; node size indicates the degree of connectivity; red circle indicates up-regulated genes; green circle indicates down-regulated genes; red lines indicate positive correlation; blue lines indicate negative correlation; the darker the color, the stronger the correlation. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

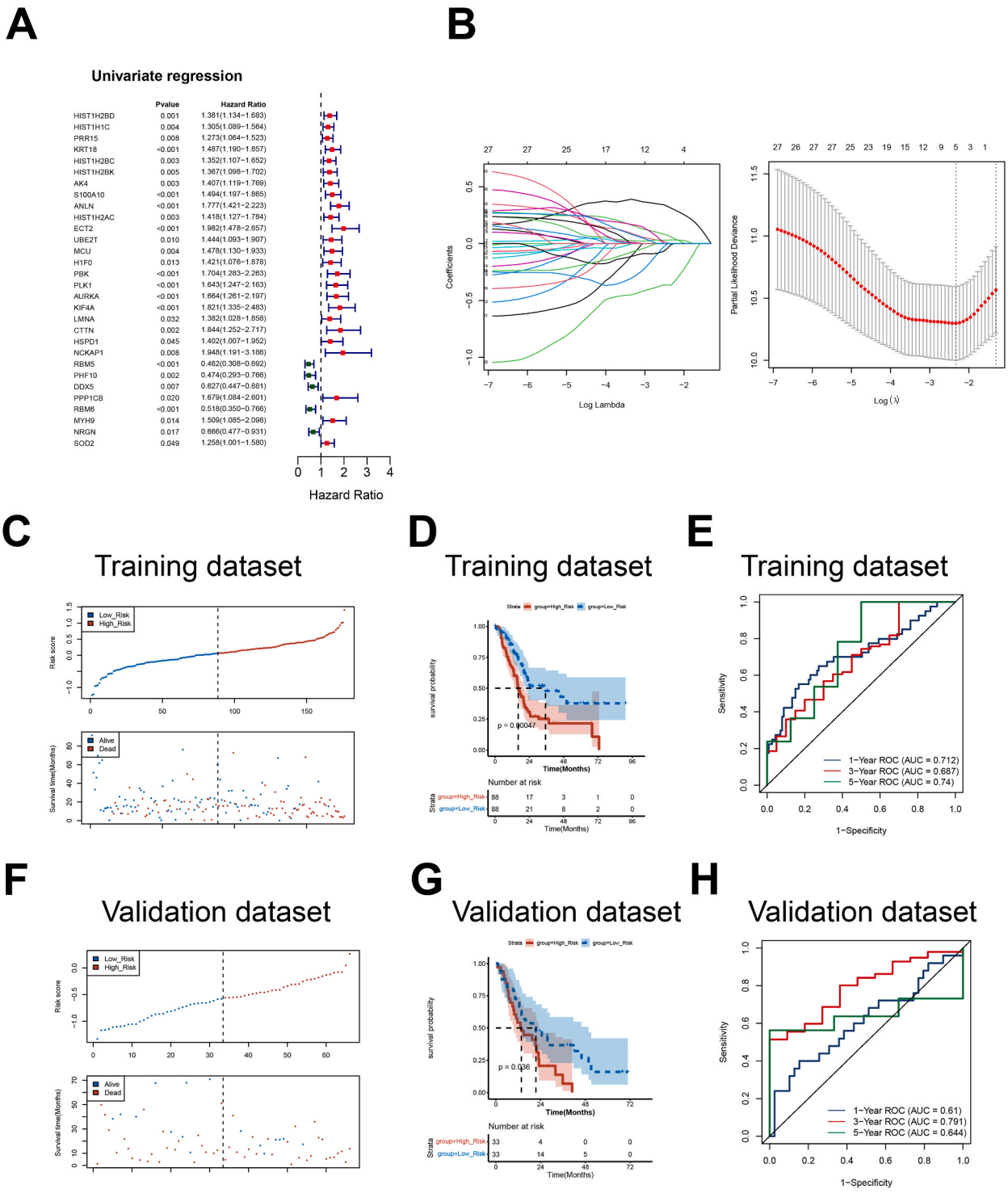


Fig. 3. Construction and evaluation of the prognostic risk score model. (A) Univariate cox regression analysis. (B) LASSO analysis. Left panel represents the coefficients. Right panel represents the selection of λ. The two dashed lines indicate two special values of λ: lambda.min on the left and lambda.1se on the right. (C) Distribution of the risk score in TCGA-PAAD training datasets. (D) Kaplan-Meier (KM) curves in TCGA-PAAD training datasets. (E) ROC curve in TCGA-PAAD training dataset. (F) Distribution of the risk score in GSE62452 validation dataset. (G) Kaplan-Meier (KM) curves in GSE62452 validation dataset. (H) ROC curve in GSE62452 validation dataset.

RBM6, which were considered as the biomakers (Fig. 3B). among which, *ANLN*, *KRT18* and *ECT2* were risk factors; *RBM5*, and *RBM6* were protective factors in PAAD. According to the five optimal DTRGs, a prognostic risk score model was constructed:

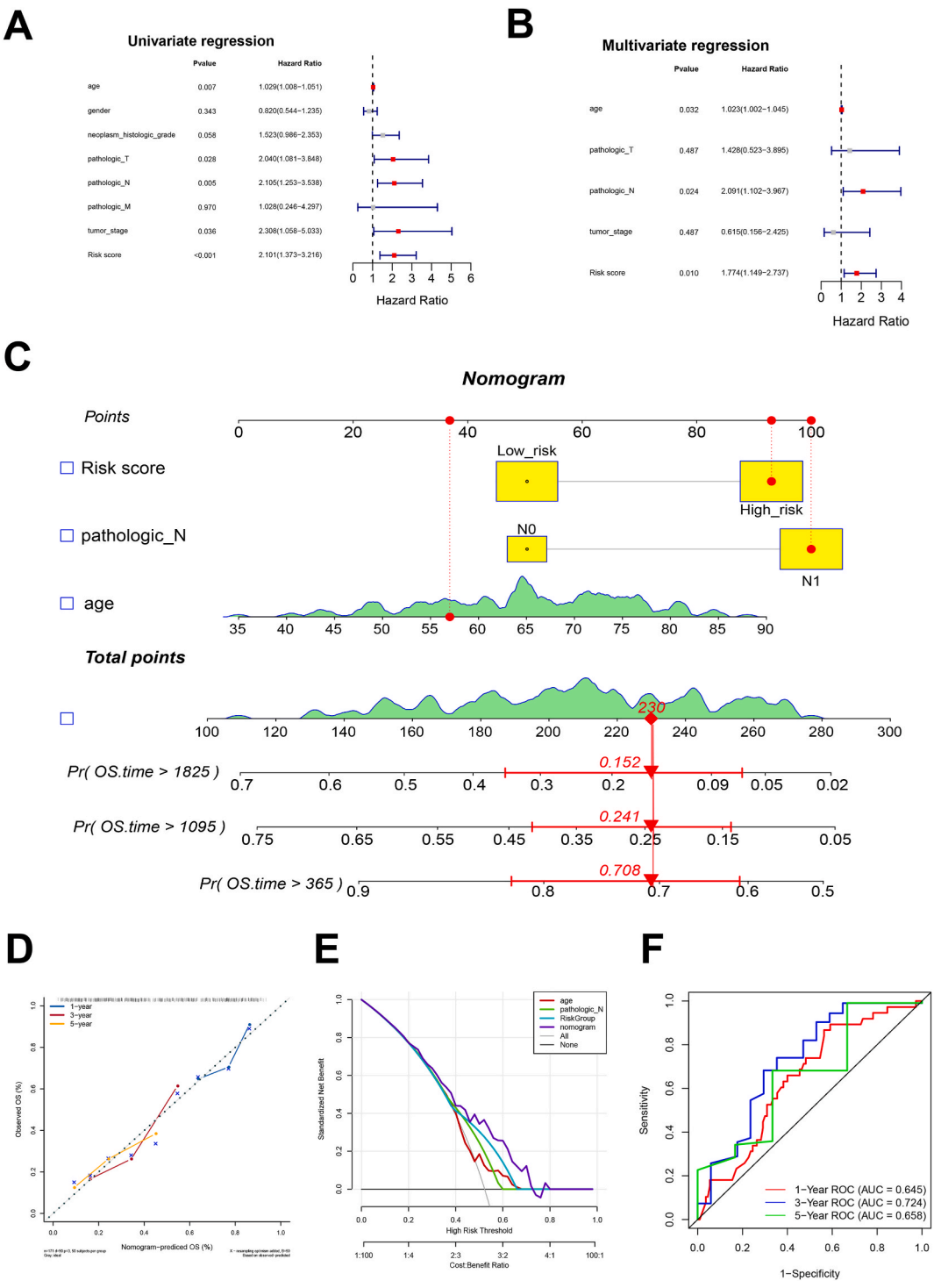


Fig. 4. Construction and evaluation of Nomogram model. (A) Univariate Cox regression analysis. (B) Multivariate Cox regression analysis. (C) Nomogram model for predicting the survival of patients. Each factor corresponds to a score, and the total score of each factor corresponds to the total score. According to the total score, the 1-year, 3-year, and 5-year survival rates are predicted. (D) Calibration curve. (E) decision curve analysis (DCA) curve. (F) ROC curve.

Risk score = 0.058414713 × *KRT18* + 0.311584926 × *ANLN* + 0.03302212 × *ECT2* - 0.363869556 × *RBM5* -0.008985028 × *RBM6*.

Based on the gene expression, the samples were divided into high- and low-expression groups according to median. KM analysis indicated that higher expression of *ANLN*, *KRT18*, and *ECT2* predicted worse overall survival; and higher expression of *RBM5* and *RBM6* predicted better overall survival (Figure S1 A-E), which is consistent with the above univariate cox regression analysis. Correlation analysis showed that the links among these five genes were closely, especially *RBM5* and *RBM6* (Figure S1F).

3.3. Evaluation of the prognostic model

For evaluating the predicting ability of the risk score model, all the cancer samples in TCGA-PAAD training dataset were divided into two groups: high- and low risk groups, with 88 samples in each group (Fig. 3C and D). KM analysis suggested that the high-risk groups exhibited worse survival rate compared with low-risk groups (Fig. 3D, P = 0.00047). ROC cure showed that the risk score model possessed promising predicting ability in survival of patients with PAAD (Fig. 3E, AUC >0.70 in 1- and 5-year survival; and AUC >0.65 in 3-year survival). The prognostic ability of the risk score model were also validated in GSE62452 dataset, and similar results were obtained (Fig. 3F–H).

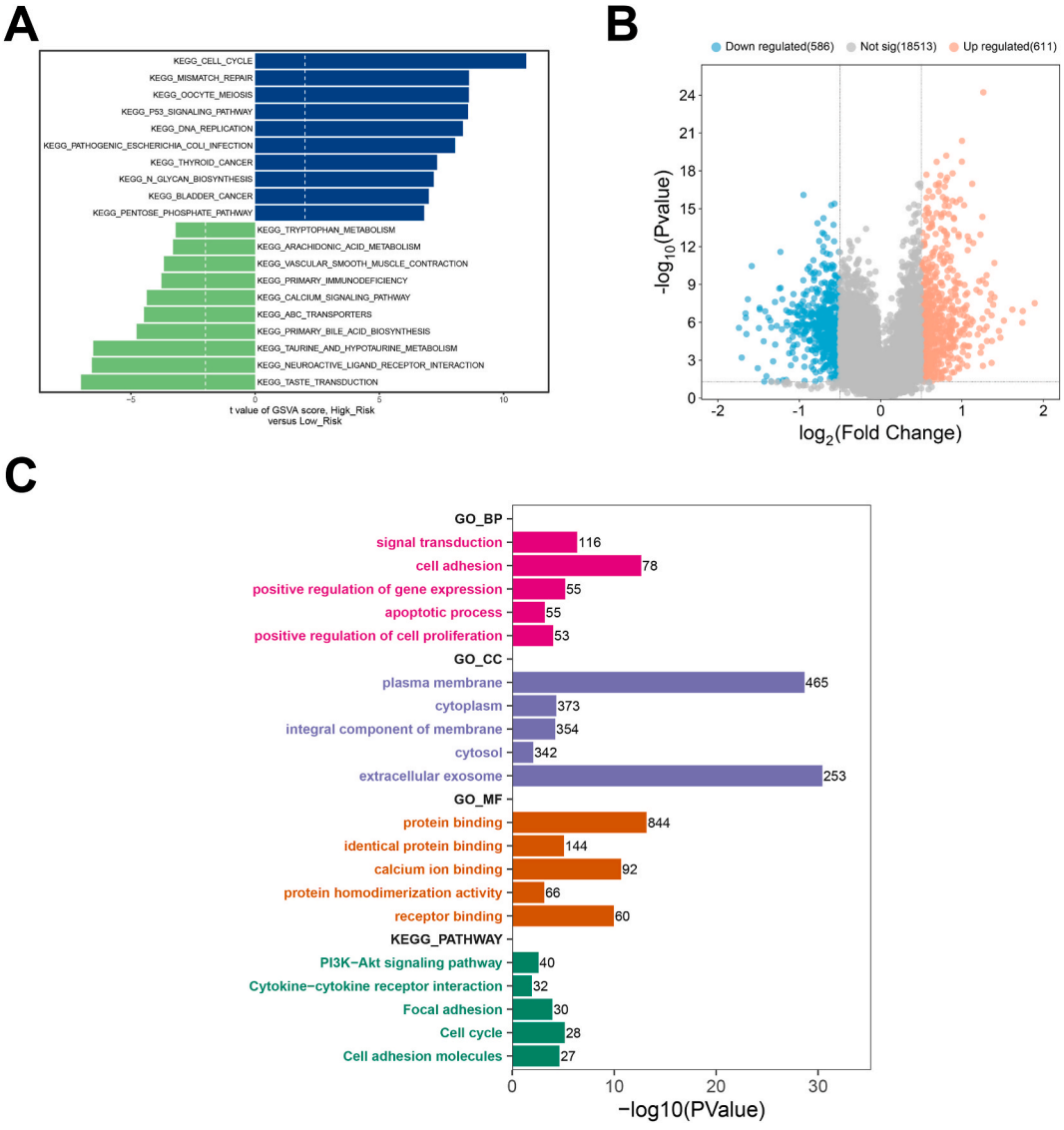


Fig. 5. Function analysis between two risk groups. (A) GSVA results. (B) Volcano plot for identification of DEGs between high- and low-risk groups. (C) GO and KEGG analysis for the DEGs between high- and low-risk groups.

3.4. Construction and evaluation of nomogram model

To investigate the independence of risk score, the clinical elements of cancer patients in TCGA-PAAD training dataset and the risk score were involved into the univariate and multivariate cox regression analysis. In the univariate cox regression analysis, age, pathologic T, pathologic N, tumor stage, and risk score were significantly correlated with the risk of PAAD (Fig. 4A, $P < 0.05$). The five elements were following involved into multivariate cox regression analysis, and the results indicated that age, pathologic N, and risk score exhibited independent prognostic ability in PAAD (Fig. 4B, $P < 0.05$). Based on the three independent factors from multivariate cox regression analysis, a nomogram model that was used to predicting the survival of patients with PAAD in 1, 3, and 5 year was constructed (Fig. 4C).

The calibration curve and DCA curve indicated that the slope between observed overall survival and nomogram-predicted overall survival were close to 1 in 1, 3, and 5 years (Fig. 4D), indicating effectiveness of the nomogram model. The DCA curve also exhibited that the predicting ability of nomogram model was higher than age, pathologic N, and risk score curve (Fig. 4E). In a threshold of 0–0.6, nomogram model possessed more benefit effectiveness. Moreover, the ROC curve also demonstrated the promising predictive ability of nomogram in survival (Fig. 4F).

3.5. Differences of functions between two risk groups

To investigate the differences of pathways between two risk groups, GSVA analysis was performed. The results indicated that high-risk group was related to cell cycle, mismatch repair, P53 signaling pathway etc.; low-risk group was enriched in tryptophan metabolism, taste transduction, neuroactive ligand receptor interaction, and so on (Fig. 5A). Furthermore, to further investigate the difference in function between two risk groups, limma package was used to explore the differential expressed genes between the two risk groups, manifesting that 1197 DEGs were selected (Fig. 5B). GO and KEGG analysis demonstrated that these DEGs between two risk groups were correlated with tumor progression process like cell adhesion, apoptotic process, cell cycle, PI3K-Akt signaling pathways (Fig. 5C).

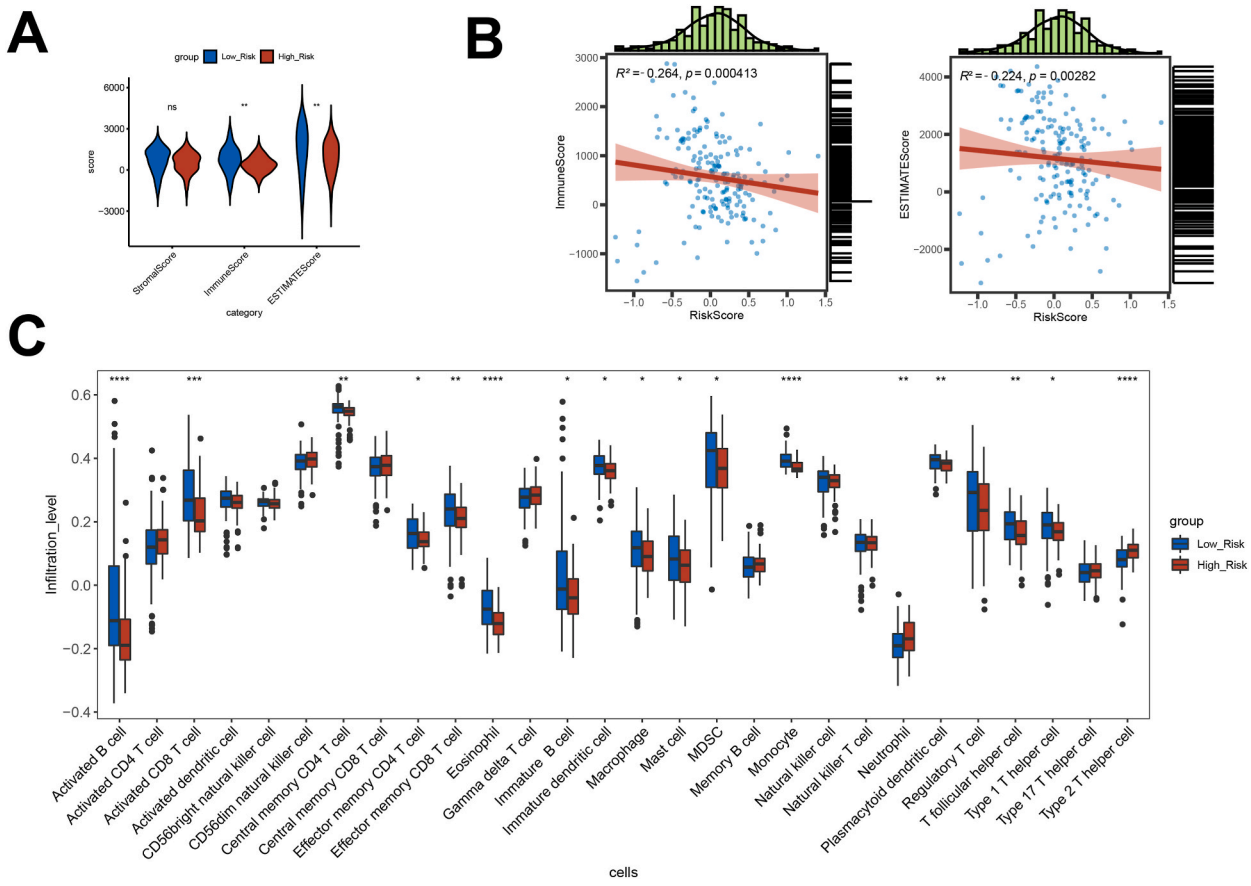


Fig. 6. Immune differences between two risk groups. (A) Stromal, immune, and ESTIMATE score between two risk groups. (B) Correlation of risk scores with immune and ESTIMATE scores. (C) Immune cell levels between two risk groups. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

3.6. Differences of immune feature between two risk groups

Tumor microenvironment is composed of tumor cells, various immune cells, fibroblasts, cytokines and catalytic factors, which is a complex system with dynamic balance. In our results, immune and ESTIMATE score were significantly different between high- and low-risk groups (Fig. 6A). The risk score suggested a significantly negative correlations with immune and ESTIMATE scores (Fig. 6B, $P < 0.001$), indicating that immune status was more active in low-risk score conditions. Moreover, immune cell infiltration analysis indicated 17 significantly different immune cells between high- and low-risk groups (Fig. 6C). For example, immune cell infiltration levels in low-risk groups were relatively higher than that in high-risk groups, exerting as higher activated B cell, activated CD8 T cell, effector memory CD4 T cell, effector memory CD8 T cell, Eosinophil, Immature B cell, immature dendritic cell, macrophage, MDSC, and Monocyte (Fig. 6C). Incidentally, expression levels of the five optimal DTRGs were also significantly correlated with the immune cell levels (Figure S2).

3.7. Differences of immunotherapy response and drug sensitivity between two risk groups

A total of 30 immune checkpoint genes were significantly differential expressed between two risk groups (Fig. 7A). The TIDE score in the high-risk group was significantly higher than that in the low-risk group (Fig. 7B), indicating that the patients in the high-risk group exhibited a greater probability of tumor immune escape and a poor response to immune checkpoint blockade (ICB) treatment. Moreover, a total of 73 drugs showed significantly different IC50 between the two different risk groups (Fig. 7C). There was a

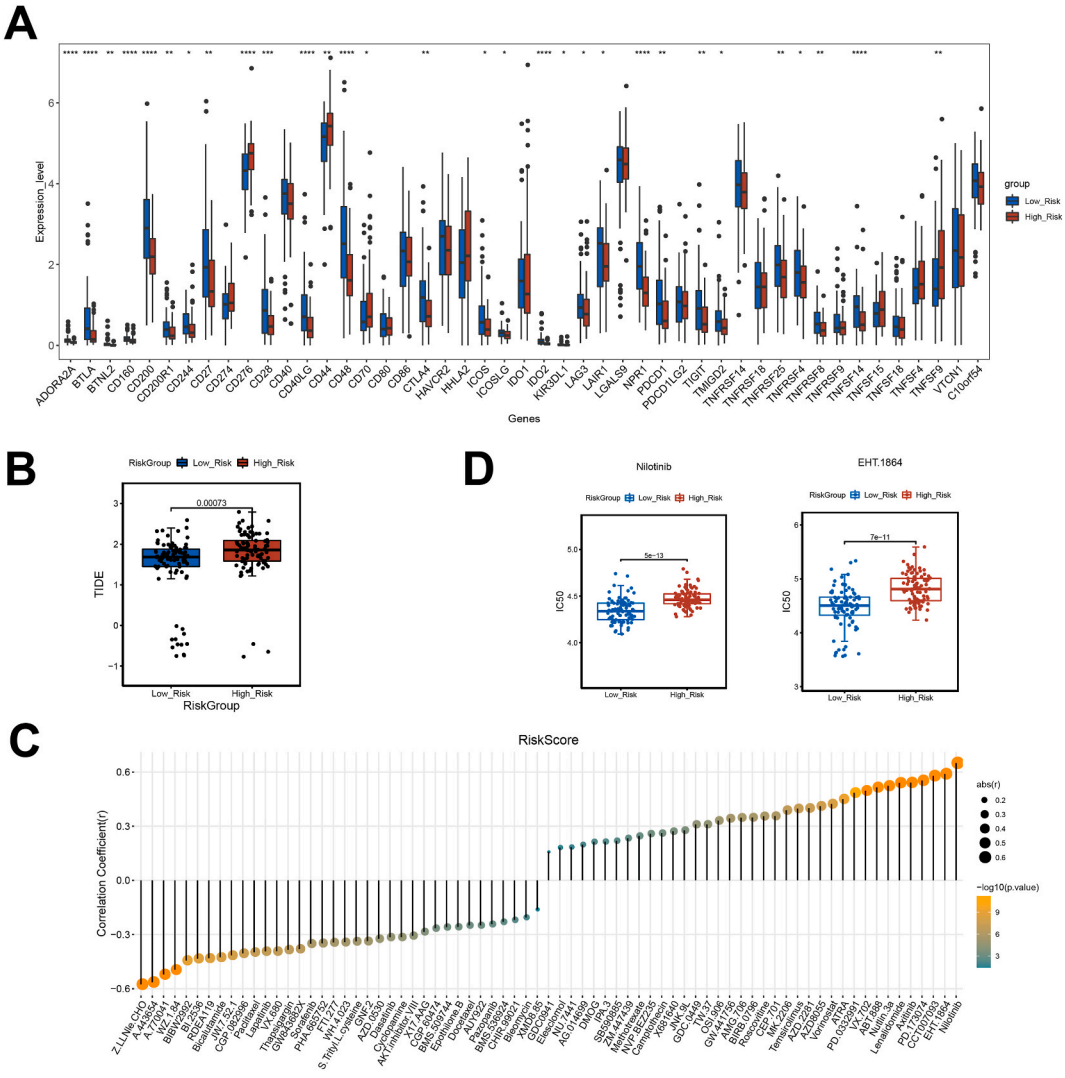


Fig. 7. Immunotherapy and drug sensitivity predicting. (A) Immune checkpoint genes expression differences between two risk groups. (B) TIDE score between two risk groups. (C) Correlation between risk score and drugs. (D) Differences of most significant drugs between two risk groups.

significant and strongest negative correlation between risk score and Z.LLNle.CHO drug, and a significant and strongest positive correlation between risk score and Nilotinib drug (Fig. 7C and D).

3.8. Mutation side and frequency of five biomarkers

The mutation frequency of *KRT18*, *ANLN*, *ECT2*, *RBM5*, and *RBM6* were 0 %, 2 %, 5 %, 0.7 %, and 2.7, respectively (Figure S3A). Moreover, there were three, five, and two mutation sites in *ANLN*, *ECT2*, and *RBM6*, respectively (Figure S3B). A previous study indicated that mutation of genes influences the prognosis of patients [22]. So, we further investigated whether mutation effected on the survival of PAAD patients. The results showed that among these five genes, only mutation of *ANLN* might be related to a poor survival (Figure S3C).

4. Discussion

Kbhb emerged to be related to cancer and metabolism progression. In the present study, a total of 63 DTRGs were identified in PAAD, and these genes were demonstrated to related to cell division and apoptosis biological functions. Through univariate cox regression and LASSO analysis, 30 DTRGs were selected to be related to prognosis and five (*KRT18*, *ANLN*, *ECT2*, *RBM5*, and *RBM6*) were identified as the optimal DTRGs in PAAD. Based on the five optimal DTRGs, a prognostic risk score model that exhibited promising predictive ability in PAAD survival was established. High-risk group showed lower survival rate. Moreover, based on the risk score, a nomogram model was also established, which can be used to predict the survival in clinic and the nomogram possessed perfect stability. Finally, high- and low-risk groups exert different immune infiltration levels, manifesting as lower risk score matching with higher immune cell levels. The results indicate an immune activation in low-risk status, which maybe the reason for the better survival in low-risk group. Furthermore, the immunotherapy response and drug sensitivity were all higher than that in low-risk groups.

For the selected optimal DTRGs, our analysis indicated that higher expression of *ANLN*, *KRT18*, and *ECT2* predicted worse overall survival; and higher expression of *RBM5* and *RBM6* predicted better overall survival. *ANLN*, full named Anillin, is a mitosis-related protein that is reported to promote contractile ring formation and cytokinesis. *ANLN*, is higher expressed, and is related to a poor prognosis in esophageal squamous cell carcinoma [23]. USP10 can regulate *ANLN* through ubiquitination to affect the tumor cell-cycle progression [23]. Similar expression and role of *ANLN* are emphasized in oral cancer and breast cancer [24,25]. In PAAD, higher expression of *ANLN* is related to lymph node metastasis and poor prognosis, which is consistent with our results. Mechanically, *ANLN* promotes cancer progression by mediating miR218-5p/LASP1 in PAAD [26].

Keratin 18 (*KRT18*) is one of the most abundant keratins in epithelial and endothelial cells, and has also been suggested highly expressed in various cancer [27–29]. *ECT2* is also an oncogene in various cancer. High expression of *ECT2* and is related to poor prognosis in breaster cancer [30]. *ECT2* involves in tumor cell growth and migration in colorectal cancer [31]. Moreover, *ECT2* is related to drug resistance. *ECT2* knockdown inhibits the proliferation and metastasis of tumor cells; promotes apoptosis, and enhances sensitivity to cisplatin in cervical cancer [32]. These results for *KRT18* and *ECT2* are also in accordance with our results. Importantly, *KRT18* is involved in the acetylation of non-histone substrate in human colon carcinoma cell line [29], demonstrating that *KRT18* might regulate tumor progression through acetylation.

RBM5 [33–36] and *RBM6* [37–39] belong to RNA-binding proteins, and are upregulated in tumor tissues in various cancers, which are also in agreement with our findings. *RBM5* exhibits pro-apoptosis activity and anti-tumor role in lung cancer [34]. Downregulation of *RBM5* can induce bladder cancer apoptosis through initiate miR-432-5p/β-catenin feedback loop [36]. *RBM6* also shows down-regulation in laryngocarcinoma, and overexpressed *RBM6* inhibits cancer incidence rate [38]. *RBM6* is downregulate in hepatocellular carcinoma tissue, and *RBM6* overexpression weaken cancer cell vitality and migration ability [39]. Moreover, a study indicates that RBM6-RBM5 chimeric expression may be a potential tumor differentiation marker [40]. Overall, the five genes might be powerful indication in PAAD, and exert robust predicting ability in PAAD, proving the promising ability of risk score model in predicting survival in PAAD again.

Nevertheless, limitation exists. First, the sample size in this analysis might be small, and the risk score model is necessary to be validated in larger cohorts. Second, although the roles of the five optimal genes have been reported in various cancers in previous studies, the regulation mechanism of them in PAAD is relatively absent, and the relationship between these genes and butyrylation is also needed to be clarified. Among the five identified genes, only *KRT18* is reported to regulate tumor progression through acetylation. More validation and exploration studies need to be performed in future. Anyway, our analysis provided a robust predicting tool for patients with PAAD, and also provided some indication for the following mechanism exploration in PAAD.

5. Conclusion

A five-gene (*KRT18*, *ANLN*, *ECT2*, *RBM5*, and *RBM6*) prognosis risk model is constructed in this study and the model is promising in predicting the survival of the PAAD. Based on the risk score model, low-risk group exhibits better survival rate, more active immune status, higher immunotherapy response and drug sensitivity. The study might provide effectively predictive and managemental indication in PAAD.

Data availability statement

All the data used in this analysis was downloaded from public online databases UCSC Xena database ((http://xena.ucsc.edu/) and

Gene Expression Omnibus (GEO https://www.ncbi.nlm.nih.gov/) database.

Ethics declarations

Review and approval by an ethics committee was not needed for this study because all the data used in the study was downloaded from public online databases and these databases were free for each researcher to use.

Funding

Not applicable.

CRediT authorship contribution statement

Fangfang Hu: Writing – original draft, Conceptualization. **Zhibin Bai:** Data curation. **Kai Yan:** Formal analysis. **Zheng Zhang:** Project administration. **Jiahua Zhou:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e34284.

References

[1] S.P. Pereira, L. Oldfield, A. Ney, P.A. Hart, M.G. Keane, S.J. Pandol, D. Li, W. Greenhalf, C.Y. Jeon, E.J. Koay, C.V. Almario, C. Halloran, A.M. Lennon, E. Costello, Early detection of pancreatic cancer, The lancet Gastroenterology & hepatology 5 (7) (2020) 698–710.
[2] E. Afghani, A.P. Klein, Pancreatic adenocarcinoma: trends in epidemiology, risk factors, and outcomes, Hematol. Oncol. Clin. N. Am. 36 (5) (2022) 879–895.
[3] T. Kamisawa, L.D. Wood, T. Itoi, K. Takaori, Pancreatic cancer. Lancet (London, England) 388 (10039) (2016) 73–85.
[4] A.D. Singhi, E.J. Koay, S.T. Chari, A. Maitra, Early detection of pancreatic cancer: opportunities and challenges, Gastroenterology 156 (7) (2019) 2024–2040.
[5] L. Sun, H. Zhang, P. Gao, Metabolic reprogramming and epigenetic modifications on the path to cancer, Protein & cell 13 (12) (2022) 877–919.
[6] S. Chang, S. Yim, H. Park, The cancer driver genes IDH1/2, JARID1C/KDM5C, and UTX/KDM6A: crosstalk between histone demethylation and hypoxic reprogramming in cancer metabolism, Experimental & molecular medicine 51 (6) (2019) 1–17.
[7] W. Liu, D. Hu, J. Li, Q. Xu, Z. Chen, B. Xu, CSTF2T facilitates pancreatic adenocarcinoma growth and metastasis by elevating H3K4Me1 methylation of CALB2 via ASH2L, Cancer Biol. Ther. 24 (1) (2023) 2216041.
[8] S. Yang, Y. Liu, C. Tang, A. Han, Z. Lin, J. Quan, Y. Yang, The CPT1A/Snail axis promotes pancreatic adenocarcinoma progression and metastasis by activating the glycolytic pathway, iScience 26 (10) (2023) 107869.
[9] Z. Xie, D. Zhang, D. Chung, Z. Tang, H. Huang, L. Dai, S. Qi, J. Li, G. Colak, Y. Chen, C. Xia, C. Peng, H. Ruan, M. Kirkey, D. Wang, L.M. Jensen, O.K. Kwon, S. Lee, S.D. Pletcher, M. Tan, D.B. Lombard, K.P. White, H. Zhao, J. Li, R.G. Roeder, X. Yang, Y. Zhao, Metabolic regulation of gene expression by histone lysine β-hydroxybutyrylation, Molecular cell 62 (2) (2016) 194–206.
[10] K.B. Koronowski, C.M. Greco, H. Huang, J.K. Kim, J.L. Fribourgh, P. Crosby, L. Mathur, X. Ren, C.L. Partch, C. Jang, F. Qiao, Y. Zhao, P. Sassone-Corsi, Ketogenesis impact on liver metabolism revealed by proteomics of lysine β-hydroxybutyrylation, Cell Rep. 36 (5) (2021) 109487.
[11] J. Huang, L. Liang, S. Jiang, Y. Liu, H. He, X. Sun, Y. Li, L. Xie, Y. Tao, L. Cong, Y. Jiang, BDH1-mediated LRRC31 regulation dependent on histone lysine β-hydroxybutyrylation to promote lung adenocarcinoma progression, MedComm 4 (6) (2023) e449.
[12] K. Liu, F. Li, Q. Sun, N. Lin, H. Han, K. You, F. Tian, Z. Mao, T. Li, T. Tong, M. Geng, Y. Zhao, W. Gu, W. Zhao, p53 β-hydroxybutyrylation attenuates p53 activity, Cell Death Dis. 10 (3) (2019) 243.
[13] S. Zhang, J. Yang, H. Wu, T. Cao, T. Ji, Establishment of a 7-gene prognostic signature based on oxidative stress genes for predicting chemotherapy resistance in pancreatic cancer, Front. Pharmacol. 14 (2023) 1091378.
[14] C. Yan, Y. Niu, F. Li, W. Zhao, L. Ma, System analysis based on the pyroptosis-related genes identifies GSDMC as a novel therapy target for pancreatic adenocarcinoma, J. Transl. Med. 20 (1) (2022) 455.
[15] H. Huang, D. Zhang, Y. Weng, K. Delaney, Z. Tang, C. Yan, S. Qi, C. Peng, P.A. Cole, R.G. Roeder, Y. Zhao, The regulatory enzymes and protein substrates for the lysine β-hydroxybutyrylation pathway, Sci. Adv. 7 (9) (2021).
[16] G.K. Smyth, Limma: linear models for microarray data, in: R. Gentleman, V.J. Carey, W. Huber, R.A. Irizarry, S. Dudoit (Eds.), Bioinformatics and Computational Biology Solutions Using R and Bioconductor, Springer New York, New York, NY, 2005, pp. 397–420.
[17] H. Chen, P.C. Boutros, VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R, BMC Bioinf. 12 (2011) 35.
[18] T. Mosmann, Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays, J. Immunol. Methods 65 (1–2) (1983) 55–63.
[19] W. Norbert, The wiener RMS (root mean square) error criterion in filter design and prediction, in: Extrapolation, Interpolation, and Smoothing of Stationary Time Series: with Engineering Applications, MIT Press, 1964, pp. 129–148.
[20] S. Guindon, J.F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, Syst. Biol. 59 (3) (2010) 307–321.

[21] P. Geeleher, N. Cox, R.S. Huang, pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels, PLoS One 9 (9) (2014) e107468.
[22] Y. Yu, Z. Wang, Q. Zheng, J. Li, FAM72 serves as a biomarker of poor prognosis in human lung adenocarcinoma, Aging (Albany NY) 13 (6) (2021) 8155–8176.
[23] Y.F. Cao, L. Xie, B.B. Tong, M.Y. Chu, W.Q. Shi, X. Li, J.Z. He, S.H. Wang, Z.Y. Wu, D.X. Deng, Y.Q. Zheng, Z.M. Li, X.E. Xu, L.D. Liao, Y.W. Cheng, L.Y. Li, L. Y. Xu, E.M. Li, Targeting USP10 induces degradation of oncogenic ANLN in esophageal squamous cell carcinoma, Cell Death Differ. 30 (2) (2023) 527–543.
[24] B. Wang, X.L. Zhang, C.X. Li, N.N. Liu, M. Hu, Z.C. Gong, ANLN promotes carcinogenesis in oral cancer by regulating the PI3K/mTOR signaling pathway, Head Face Med. 17 (1) (2021) 18.
[25] W. Zhou, Z. Wang, N. Shen, W. Pi, W. Jiang, J. Huang, Y. Hu, X. Li, L. Sun, Knockdown of ANLN by lentivirus inhibits cell growth and migration in human breast cancer, Mol. Cell. Biochem. 398 (1–2) (2015) 11–19.
[26] A. Wang, H. Dai, Y. Gong, C. Zhang, J. Shu, Y. Luo, Y. Jiang, W. Liu, P. Bie, ANLN-induced EZH2 upregulation promotes pancreatic cancer progression by mediating miR-218-5p/LASP1 signaling axis, Journal of experimental & clinical cancer research : CR 38 (1) (2019) 347.
[27] B. Chen, X. Xu, D.D. Lin, X. Chen, Y.T. Xu, X. Liu, W.G. Dong, KRT18 modulates alternative splicing of genes involved in proliferation and apoptosis processes in both gastric cancer cells and clinical samples, Front. Genet. 12 (2021) 635429.
[28] J. Zhang, S. Hu, Y. Li, KRT18 is correlated with the malignant status and acts as an oncogene in colorectal cancer, Biosci. Rep. 39 (8) (2019).
[29] C. Delouis, P. Prochasson, M. Laithier, O. Brison, Use of adenoviral E1A protein to analyze K18 promoter deregulation in colon carcinoma cells discloses a role for CtBP1 and BRCA1, BMC Mol. Biol. 6 (2005) 8.
[30] M. Yi, D. Zhang, B. Song, B. Zhao, M. Niu, Y. Wu, Z. Dai, K. Wu, Increased expression of ECT2 predicts the poor prognosis of breast cancer patients, Exp. Hematol. Oncol. 11 (1) (2022) 107.
[31] D.R. Cook, M. Kang, T.D. Martin, J.A. Galanko, G.H. Loeza, D.G. Trembath, V. Justilien, K.A. Pickering, D.F. Vincent, A. Jarosch, P. Jurmeister, A.M. Waters, P. S. Hibshman, A.D. Campbell, C.A. Ford, T.O. Keku, J.J. Yeh, M.S. Lee, A.D. Cox, A.P. Fields, R.S. Sandler, O.J. Sansom, C. Sers, A. Schaefer, C.J. Der, Aberrant expression and subcellular localization of ECT2 drives colorectal cancer progression and growth, Cancer Res. 82 (1) (2022) 90–104.
[32] X. Liu, J. Zhang, S. Ju, L. Liu, Y. Sun, L. Guo, Q. Zhen, S. Han, W. Lu, Y. Zhang, ECT2 promotes malignant phenotypes through the activation of the AKT/mTOR pathway and cisplatin resistance in cervical cancer, Cancer Gene Ther. 30 (1) (2023) 62–73.
[33] D. Jamsai, D.N. Watkins, A.E. O'Connor, D.J. Merriner, S. Gursoy, A.D. Bird, B. Kumar, A. Miller, T.J. Cole, B.J. Jenkins, M.K. O'Bryan, In vivo evidence that RBM5 is a tumour suppressor in the lung, Sci. Rep. 7 (1) (2017) 16323.
[34] L.C. Sutherland, K. Wang, A.G. Robinson, RBM5 as a putative tumor suppressor gene for lung cancer, J. Thorac. Oncol. : official publication of the International Association for the Study of Lung Cancer 5 (3) (2010) 294–298.
[35] Y. Wang, X.M. Jiang, Z.X. Feng, X.L. Li, W.L. Zhang, Long noncoding RNA PCAT-1 accelerates the metastasis of pancreatic cancer by repressing RBM5, Eur. Rev. Med. Pharmacol. Sci. 25 (3) (2021) 1164.
[36] Y.P. Zhang, K.L. Liu, Y.X. Wang, Z. Yang, Z.W. Han, B.S. Lu, J.C. Qi, Y.W. Yin, Z.H. Teng, X.L. Chang, J.D. Li, H. Xin, W. Li, Down-regulated RBM5 inhibits bladder cancer cell apoptosis by initiating an miR-432-5p/β-catenin feedback loop, Faseb. J. : official publication of the Federation of American Societies for Experimental Biology 33 (10) (2019) 10973–10985.
[37] F.E. Machour, E.R. Abu-Zhayia, S.W. Awwad, T. Bidany-Mizrahi, S. Meinke, L.A. Bishara, F. Heyd, R.I. Aqeilan, N. Ayoub, RBM6 splicing factor promotes homologous recombination repair of double-strand breaks and modulates sensitivity to chemotherapeutic drugs, Nucleic acids research 49 (20) (2021) 11708–11727.
[38] Q. Wang, F. Wang, W. Zhong, H. Ling, J. Wang, J. Cui, T. Xie, S. Wen, J. Chen, RNA-binding protein RBM6 as a tumor suppressor gene represses the growth and progression in laryngocarcinoma, Gene 697 (2019) 26–34.
[39] Z. Ye, J. Zhang, Y. Yang, Y. Wei, L. Li, X. Wang, RNA-binding protein RBM6 acts as a tumor suppressor gene to inhibit the progression of hepatocellular carcinoma, Journal of BUON : official journal of the Balkan Union of Oncology 26 (2) (2021) 402–408.
[40] K. Wang, G. Ubriaco, L.C. Sutherland, RBM6-RBM5 transcription-induced chimeras are differentially expressed in tumours, BMC Genom. 8 (2007) 348.