



Research article

IGF2BP2-related modification patterns in pancreatic cancer: A machine learning-driven approach towards personalized treatment

Dongjie Chen^{a,b,c,d,1}, Longjun Zang^{e,1}, Yanling Zhou^f, Yongchao Yang^g, Xianlin Zhang^h, Zheng Li^h, Yufeng Shuⁱ, Wenzhe Gao^{a,**}, Hongwei Zhu^{a,*}, Xiao Yu^{a,***}

^a Department of Hepatopancreatobiliary Surgery, Third Xiangya Hospital, Central South University, Changsha, Hunan, PR China

^b Department of General Surgery, Pancreatic Disease Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, PR China

^c Research Institute of Pancreatic Diseases, Shanghai Key Laboratory of Translational Research for Pancreatic Neoplasms, Shanghai Jiao Tong University School of Medicine, Shanghai, PR China

^d State Key Laboratory of Oncogenes and Related Genes, Institute of Translational Medicine, Shanghai Jiao Tong University, Shanghai, PR China

^e Department of General Surgery, Taiyuan Central Hospital, Taiyuan, Shanxi, PR China

^f University of Shanghai for Science and Technology, Shanghai, PR China

^g Department of Burns and Plastic Surgery, Third Xiangya Hospital, Central South University, Changsha, Hunan, PR China

^h Department of General Surgery, Affiliated Renhe Hospital of China Three Gorges University, Yichang, Hubei, PR China

ⁱ Department of Gastroenterology, Third Xiangya Hospital, Central South University, Changsha, Hunan, PR China

ARTICLE INFO

Keywords:

IGF2BP2
m6A modification
Precision medicine
Pancreatic cancer

ABSTRACT

Pancreatic cancer (PC) is a malignant digestive system tumor with a very poor prognosis. N6-methyladenosine (m6A) is mediated by a variety of readers and participates in important regulatory roles in PC. Based on TCGA_PAAD, ICGC_AU_PAAD, ICGC_CA_PAAD, GSE28735 and GSE62452 datasets, We mapped the multi-omics changes of m6A readers in PC and found that m6A readers, especially IGF2BP family genes, had specific changes and were significantly associated with poor prognosis. An unsupervised consensus clustering algorithm was used to explore the correlation between specific expression patterns of m6A readers in PC and enrichment pathways, tumor immunity and clinical molecular subtypes. Then, the principal component analysis (PCA) algorithm was used to quantify specific expression patterns and screen core genes. Machine learning algorithms such as Bootstrapping and RSF were used to quantify the expression patterns of core genes and construct a prognostic scoring model for PC patients. What's more, pharmacogenomic databases were used to screen sensitive drug targets and small molecule compounds for high-risk PC patients in an all-around and multi-angle way. Our study has not only provided new insights into personalized prognostication approaches, but also thrown light on

* Corresponding author. Department of Hepatopancreatobiliary Surgery, The Third Xiangya Hospital, Central South University, Changsha, Hunan, PR China.

** Corresponding author. Department of Hepatopancreatobiliary Surgery, The Third Xiangya Hospital, Central South University, Changsha, Hunan, PR China.

*** Corresponding author. Department of Hepatopancreatobiliary Surgery, The Third Xiangya Hospital, Central South University, Changsha, Hunan, PR China.

E-mail addresses: gaowenzhexy@csu.edu.cn (W. Gao), zhw_0509@csu.edu.cn (H. Zhu), yuxiaoyx4@csu.edu.cn (X. Yu).

¹ These authors have contributed equally to this work and share first authorship.

<https://doi.org/10.1016/j.heliyon.2024.e28243>

Received 20 July 2023; Received in revised form 11 March 2024; Accepted 14 March 2024

Available online 16 March 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

integrating tailored risk stratification with precision therapy based on IGF2BP2-mediated m6A modification patterns.

1. Introduction

Pancreatic cancer (PC) is an insidious, rapidly progressing digestive tract malignant tumor with poor therapeutic effect and prognosis [1]. It is the seventh leading cause of cancer death, with a 5-year survival rate of only 10%. At present, surgical resection is considered to be the only cure for PC. However, clinical experience shows that, due to the absence of hallmark signs in the early stage of PC, patients usually have advanced stage by the time of diagnosis, and only about 20% of patients meet the surgical indications [2]. Unfortunately, the vast majority of patients will eventually experience tumor recurrence even after surgical resection, and current radiotherapy, chemotherapy, immunotherapy, and targeted therapy do not significantly improve the survival of patients with PC [3]. In the past decade, despite rapid advances in diagnostic methods, perioperative management, radiotherapy techniques, and systemic treatment of advanced disease, the clinical outcome of patients with PC has received little benefit. Therefore, there is an urgent need to classify tumors according to the molecular similarities and differences related to clinical and biology, improve the established morphological and imaging methods, so as to achieve a more accurate prognosis, better optimize individual patient management and risk stratification, and improve the selection of systematic treatment options [4,5].

Epigenetics refers to the regulation of transcription or translation processes to affect the characteristics and function of genes without changing the DNA sequence, which mainly includes DNA methylation, histone modification, genomic imprinting, chromosome remodeling, etc [6]. In recent years, with the development of high-throughput sequencing technology, the study of RNA modification has received more and more attention. To date, more than 100 types of modifications have been identified in various RNAs such as messenger RNAs (mRNAs), transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), microRNAs (miRNAs), and long noncoding RNAs (lncRNAs) [7]. Among them, the earliest discovered and most widely studied is N6-methyladenosine (m6A), which has been considered as the most common methylation modification in eukaryotic mRNAs [8]. It has been estimated that about 0.1–0.4% of adenine in mRNA is specifically modified by m6A, and m6A modification usually occurs in RRACH sequences (R = A, G, or U; R = A or

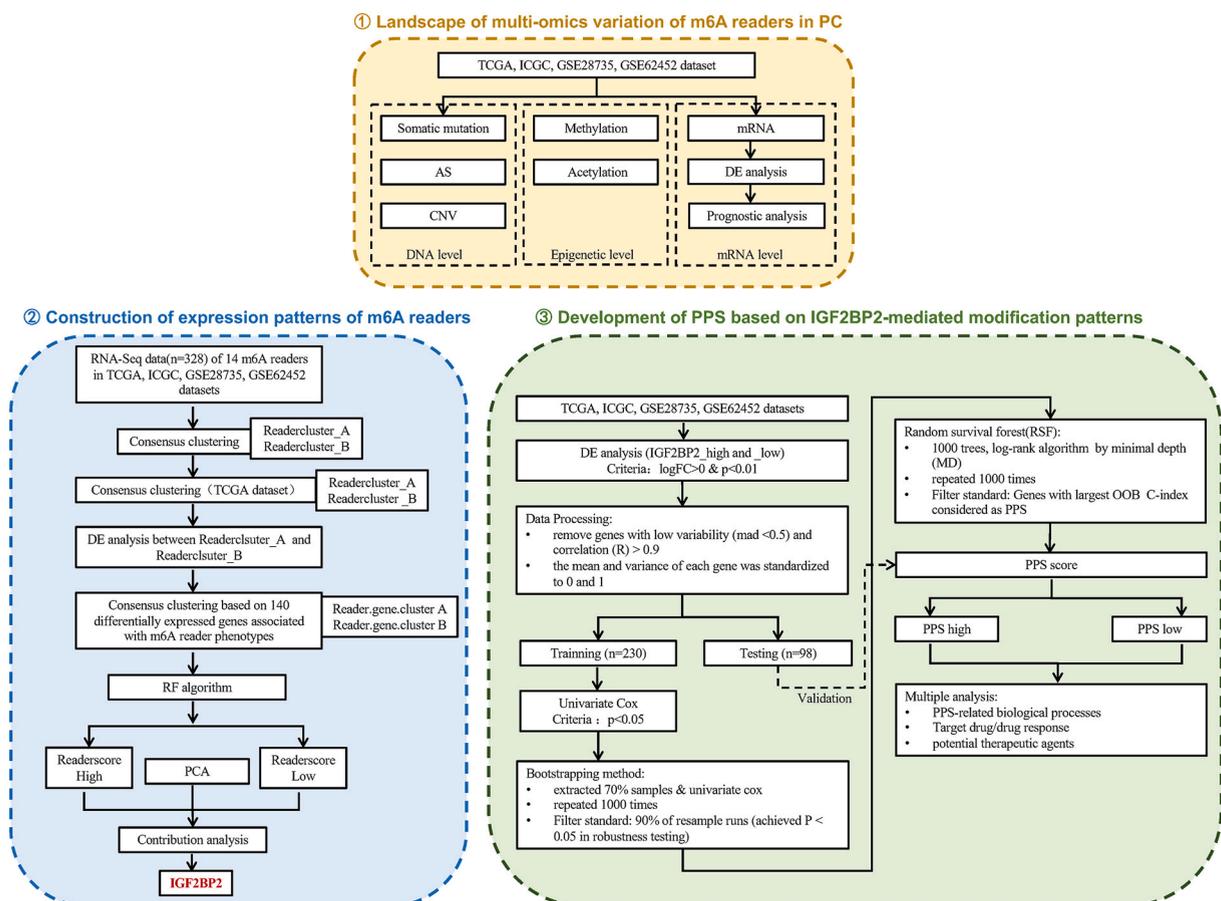


Fig. 1. Flowchat of this study.

G; H = A, C, or U), mainly enriched in the 3' non-coding region and near the stop codon. Regulation of m6A modification is dynamically reversible, mediated by m6A methyltransferases (also known as "Writers"), and can be removed by m6A demethylases (also known as "Erasers"). More importantly, the effect of m6A modification on RNA metabolism mainly depends on the recognition of different m6A-binding proteins (also known as "Readers") [9].

The m6A-modified transcript can be identified and bound by the m6A reader protein, which controls gene expression through diverse processes. These include the regulation of mRNA stability [10], mRNA splicing [11], mRNA structure [12], mRNA export [13], translation efficiency [14], and miRNA biogenesis [15], thereby exerting its regulatory effects. Emerging findings indicate a close association between m6A reader and the development and advancement of multiple human cancers, especially in PC [16]. For example, m6A reader IGF2BP2 could interact with LncRNA-PACERR and miR-671-3p, thereby promoting M2 polarization and pro-tumor function in TAMs of PC [17]. Liu et al. unveiled the pivotal tumor-inhibitory functions of the YTHDC1/miR-30d axis in the progression of PC, showcasing its significance across biological, mechanistic, and clinical realms in the context of human PC and glycolytic pathways [18].

Given that the majority of existing research concentrates on one individual reader, a thorough and profound comprehension of the overall traits of malign biological conduct in PC, orchestrated by the combined impact of comprehensive m6A readers, is currently lacking. Hence, in this study, starting from all m6A readers, we first drew the multi-omics level change map of m6A readers through the public database, and then explored the unique expression pattern of m6A readers in PC by unsupervised clustering and principal component analysis algorithm and screened the core genes. Finally, we used machine learning algorithms to construct core-gene expression pattern-based scoring models for pharmacogenomic studies, so as to tailor specialized management for PC patients. The overall workflow is displayed in Fig. 1.

2. Materials and Methods

2.1. Data acquisition and preprocessing

Public transcriptional profiling datasets, including the TCGA_PAAD dataset from The Cancer Genome Atlas Program (TCGA), GSE62452 and GSE28735 dataset from Gene Expression Omnibus (GEO), ICGC_AU_PAAD dataset for International Cancer Genome Consortium (ICGC) were included. For the TCGA_PAAD dataset, somatic mutation, copy number variation (CNV), RNA expression in transcript per million fragments mapped (FPKM) format data, methylation data, complete clinical information, alternative splicing (AS) data, histone modification data were obtained from UCSC Xena (<https://xenabrowser.net/datapages/>), TCGA Splice Seq dataset (<http://projects.insilico.us.com/TCGASpliceSeq/>), and Cistrome Data Browser database (<http://cistrome.org/db/>), respectively. RNA-seq data in fragments per kilobase of FPKM format was then transformed into transcripts per million (TPM). GEO datasets were downloaded from <https://www.ncbi.nlm.nih.gov/geo/> in normalized data format. R package "sva" was then utilized to combine different sourced GEO datasets and TCGA datasets. ICGC_AU_PAAD dataset was downloaded from <https://dcc.icgc.org/>, somatic mutation data, FPKM RNA-seq data and clinical information were included in our study. The somatic mutation data of ICGC_CA_PAAD was also included in our study and downloaded from <https://dcc.icgc.org/>. The expression data of normal tissue was downloaded from the Genotype-Tissue Expression (GTEx) database (<https://gtexportal.org/home/>). 16 m6A readers (YTHDF1, YTHDF2, YTHDF3, YTHDC1, YTHDC2, IGF2BP1, IGF2BP2, IGF2BP3, FMR1, HNRNPC, HNRNPA2B1, EIF3A, PRRC2A, RBMX, LRPPRC, SND1) were screened from literatures [14,19]. These genes with deleted expression profiles in TCGA_PAAD, ICGC_AU_PAAD, ICGC_CA_PAAD, GSE28735 and GSE62452 were excluded. Finally, 14 m6A readers were included in the subsequent study, and the deleted genes were HNRNPC and RBMX. For pharmacogenomic research, the CRISPR-Cas9 Essentiality Revealing Score (CERES) score was obtained from the Dependency Map (DepMap) dataset (<https://depmap.org/portal/>) to measure the dependence of a gene of interest in a PC cell line. A lower CERES score indicates that the gene is more important for cell growth and survival of a given PC. Expression profile data of human cancer cell lines (CCLs) were obtained from the Broad Institute Cancer Cell Line Encyclopedia (CCLE) project (<https://portals.broadinstitute.org/ccle/>). Drug sensitivity data of CCLs were achieved from the Cancer Therapeutics Response Portal (CTRP v.2.0, released October 2015, <https://portals.broadinstitute.org/ctrp>) and profiling relative inhibition simultaneously in mixtures (PRISM) dataset (19Q4, released December 2019, <https://depmap.org/portal/prism/>). The above two datasets used the Area Under the Curve (AUC) as a measure of drug sensitivity, and a lower AUC value indicated that the drug was more sensitive to treatment. The connective Map (CMap) database (<https://clue.io/>) was also used to explore the interaction network between drugs/small molecule compounds, genes and disease states. The correlation score was obtained according to the enrichment of differentially expressed genes in the reference gene set, and the value was between -100 and 100. A positive number indicated that genes were positively correlated with the reference gene set, while a negative number indicated that differentially expressed genes were negatively correlated with the reference gene set.

2.2. Unsupervised consensus clustering

After the expression matrix was standardized using `sweep()` function in R, the package "ConsensusClusterPlus" was applied for gene expression clustering. The parameters in this study were set as: `maxK = 4`, `reps = 100`, `pItem = 0.8`, `pFeature = 1`, `title = title`, `clusterAlg = "pam"`, `distance = "spearman"`. The cumulative distribution function (CDF) value and delta area of the CDF curve were used as evaluation criteria for every single clustering.

2.3. Differential analysis and pathway analysis

The “limma” package in R was employed to screen for differentially expressed genes (DEGs) between Readercluster_A and Readercluster_B groups. To rectify false positive results, adjusted P-values were applied using the default Benjamini-Hochberg false discovery rate (FDR) method. The cutoff values for identifying DEGs were set at $FDR < 0.01$ and $|\log_2\text{fold change (FC)}| > 2$. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses, were performed on the DEGs by using the “clusterProfiler” R package. Gene set enrichment analysis (GSEA) provided by MsigDB was adopted to determine the statistical significance of molecular pathways as well as the consistent heterogeneities between high and low-PPS groups based on the R package “fgsea”. $FDR < 0.05$ was considered statistically significant. Pathifier analysis is mainly used to identify specific signaling pathways at specific stages of cancer through correlation, variance stability and principal component analysis methods. For each PC sample, the Pathway Dereglulation Score (PDS) was calculated to estimate the extent to which the activity of a certain pathway in PC samples deviates from that in normal samples, and thus to aggregate the gene-level data Conversion to pathway-level data was used for subsequent analysis. The Pathifier algorithm relies on “Pathifier” package in R.

2.4. Dimensionality reduction analysis of random forest (RF)

To further extract the core genes for subsequent modeling analysis, the Boruta dimensionality reduction algorithm based on RF was used here. Specific analysis steps and codes are as follows: (1) findCorrelation () function in “Caret” R package was used to eliminate genes with high collinearity ($cor > 0.85$); (2) Dimensionality reduction analysis is carried out by using the “Boruta” R package and the Boruta () function in the “randomForest” R package, TentativeRoughFix () function for feature determination, and getSelectedAttributes () function for final feature selection. The Boruta gene obtained after the operation is the gene obtained after dimensionality reduction.

2.5. Principal component analysis (PCA) analysis and contribution screening

- (1) In the TCGA_PAAD dataset, based on the expression matrix of 14 m6A readers, the consensus clustering method was used to obtain the specific expression pattern of m6A readers, namely Readercluster;
- (2) DEGs between specific expression patterns of m6A readers were obtained using the differential analysis method, which was defined as m6A reader phenotypic-related differential genes;
- (3) The expression patterns of m6A reader phenotypic-related differential genes were obtained by consensus clustering analysis again, namely Reader. gene.cluster;
- (3) The univariate Cox regression analysis method was used to obtain prognostic genes with $p < 0.05$;
- (4) The aforementioned RF dimensionality reduction algorithm was used to further obtain the robust prognostic genes;
- (5) The prcomp () function in the R package “prcomp” was used for PCA analysis, and principal component 1 (PC1) and principal component 2 (PC2) were selected to construct the scoring model, which was named as Readerscore model, and the calculation formula was as follows:

$$\text{Readerscore} = \sum (\text{PC1}_i + \text{PC2}_i)$$

where i is the expression level of core genes in (4). The scoring system calculates a score based on genes in the gene set that have strong positive (or negative) correlations, while reducing the weight of unrelated genes. Then, the surv_cutpoint () function in the “Survminer” package was used to determine the optimal cut-off value, and the samples were divided into two subgroups: Readerscore_high and Readerscore_low for subsequent analysis.

(6) The factoextra () function in R package “prcomp” was used to draw PCA component diagram for the above PCA results, in which the value of coordinate axis PC1/2 was the explanation rate of the overall difference, the point in the figure represented the sample, the color represented the group, the arrow represented the original variable, and the direction represented the correlation between the original variable and the principal component. The length represents the contribution of the original data to the principal components. This step was used to screen the core genes with the highest contribution to principal components for subsequent analysis.

2.6. Gene set variation analysis (GSVA) analysis and single sample gene set enrichment analysis (ssGSEA) analysis

To explore diverse enrichment status in gene function for different clusters or subgroups, GSVA and ssGSEA were applied by R package “GSVA” and “GSEAbase”. Two gene sets were conducted for functional annotation from MsigDB (<http://www.gsea-msigdb.org/gsea/msigdb/>), which were c2. cp.kegg.v7.4. symbols and h. all.v7.4. symbols, respectively. The “limma” R package was then utilized to distinguish the enrichment differences between different subgroups. The pathways with $|\log_2FC| > 0.1$ and $FDR < 0.05$ were considered the most differentially enriched molecular pathways between the two subtypes.

2.7. Construction of poor prognosis signature (PPS)

In order to meet the standards of machine learning algorithms, RNA-Seq data and clinical data of 328 samples from TCGA_PAAD, ICGC_AU_PAAD and GSE62452 and GSE28735 were included. Data preprocessing methods are as follows:

- (1) According to the median expression of IGF2BP2, 328 PC samples were divided into two subgroups: IGF2BP2_high and IGF2BP2_low, and the “limma” R package was used for difference analysis. On the basis of $|\log_2FC| > 0$ and $\text{adj. } p < 0.01$, we screened out IGF2BP2 related differential genes;
- (2) apply () and findCorrelation () functions were utilized to remove genes that had strong collinearity ($R > 0.9$) and low variability ($MAD < 0.5$). The preprocess () function in R package “caret” was used to normalize and standardize the expression of DEGs. Finally, the createDataPartition () function was used to divide the treated differential gene expression matrix into training set and testing set according to the ratio of 7:3.
- (3) The coxph () function in the R package “survival” was used to perform univariate Cox regression analysis on the data of the training set, and statistically significant, prognostic differential genes ($p < 0.05$);
- (4) In order to re-evaluate the correlation between gene expression and prognosis, the “Bootstrapping” method described above was used to test the robustness of genes with prognostic differences. We performed univariate Cox regression analysis on 70% of the samples randomly selected from the training set. 1000 iterations were tuned to select the samples with significant prognostic value in 90% of the resampled samples, and were further included in the subsequent analysis;
- (5) To further screen the core genes, rfsrc () function in the R package “randomForestSRC” was used for random survival forest (RSF) analysis. In this RSF analysis, the function used log-rank and score-splitting algorithms to build 1000 survival numbers, and the included DEGs were ranked according to the minimal depth (MD) method. The smaller the variable value, the better the prediction efficiency. 10-fold cross-validation was adopted for tuning. Finally, the above RSF analysis was iterated 1000 times, and the Concordance index (C-index) of out-of-bag samples was used as the criterion to judge the quality of the model. The higher the c-index value, the best-fitting effect of the model constructed by this gene set was observed, and this gene set was named PPS.
- (6) Based on the gene set obtained in the previous step, coxph () function in R package “survival” was used for multivariate Cox regression analysis. The PPS scoring model was constructed with the correlation coefficients obtained from multivariate Cox regression, and the calculation formula was as follows:

$$\text{PPS score} = \sum_{i=1}^n \text{Coef}_i * x_i$$

where Coef_i is the multivariate Cox regression coefficient corresponding to the i_{th} gene, and x_i is the expression value of the i_{th} gene. In addition, the AUC value of the training set and testing set was used as the basis to judge the effectiveness of the PPS model and other models.

2.8. Screening for potential drug targets

In order to identify the target proteins (genes) with potential therapeutic significance in PC patients with high PPS scores, we first downloaded the detailed target information of 6125 compounds from The Drug Repurposing Hub database, and after deleting the duplicated target information, we performed the following analysis (Figure S1):

- (1) Spearman correlation analysis was used to analyze the correlation between the protein expression profile of target genes and PPS score in PC patients. $R > 0.4$ and $p < 0.05$ were used as the criteria to screen the drug targets related to the poor prognosis of PC.
- (2) PPS scores of PC cell lines in the CCLE database were calculated according to the above-mentioned scoring model construction formula, and Spearman correlation analysis was performed between CERES scores obtained from the DepMap database and PPS scores. Since a lower CERES score of a gene indicates a stronger dependence of the gene on the malignant biological phenotype of PC, $R < -0.3$ and $p < 0.05$ were used as the criteria to screen the drug targets related to the poor prognosis of PC.
- (3) The targets obtained in (1) and (2) were intersected, and the drug targets with high confidence were finally obtained.

2.9. Drug sensitivity analysis

In order to conduct drug sensitivity analysis of PC patients with high PPS scores, the CTRP and PRISM databases were used for the analysis. The detailed data acquisition and analysis process are as follows (Figure S2):

- (1) CTRP and PRISM database portals were used to download the gene expression matrix and drug sensitivity value (AUC) after drug intervention, and the cell lines with a missing value (NA value) of more than 20% or from hematopoietic and lymphoid tissues were excluded.
- (2) The pRRopheticPredict () function in the “pRRophetic” R package was used to predict the drug sensitivity of PC samples included in this study, and the predicted AUC value was obtained.
- (3) Wilcoxon rank-sum test was used to analyze the differential drug sensitivity between the high PPS score subgroup (top 10%) and the low PPS score subgroup (bottom 10%). To identify compounds with higher sensitivity (lower AUC values) in the subgroup with higher PPS scores. The screening criteria were: $\log_2FC > 0.02$ (CTRP), $\log_2FC > 0.03$ (PRISM);
- (4) Spearman correlation analysis was performed between drug susceptibility values and PPS scores to identify compounds with negative correlation coefficients. The screening criteria were: $R < -0.3$ (CTRP), $R < -0.45$ (PRISM);

- (5) By intersecting the compounds obtained in (3) and (4), highly sensitive drugs could be obtained based on CTRP and PRISM databases, respectively.

For complementary validation of the higher sensitivity drugs screened by the above methods, we also performed CMap analysis. Firstly, the differences between PC samples and normal pancreatic tissue samples in the GTEx database were analyzed, and the 300 genes with the most significant differences (including 150 significantly up-regulated genes and 150 significantly down-regulated genes) were screened out. Then, these DEGs were submitted to the CMap cloud platform (<https://clue.io/query/>) for analysis, and the CMap score of each compound was obtained. If the negative score of a compound is smaller, the gene expression pattern of the compound is opposite to the expression pattern of PC. It is suggested that this compound has a good therapeutic effect in PC.

2.10. Statistical analysis

All statistical analyses were performed in R (version 4.1.1). The comparison of count data was tested by Fisher's test and Chi-square test. For the measurement data that conformed to the normal distribution, the Student-t test was applied; besides, the Wilcoxon test was applied for non-normal distribution data between independent subgroups. Spearman analysis was applied to estimate the correlations between two variables that are not linearly related. The Kaplan-Meier (K-M) test was utilized to validate the fraction of PC patients living for a certain survival time and the log-rank test was conducted to compare the significance of the difference. R package "survival" was used for depicting K-M survival curve. A two-tailed p-value of less than 0.05 was deemed to be statistically significant unless specifically stated.

3. Results

3.1. Landscape of multi-omics variation of m6A readers in PC

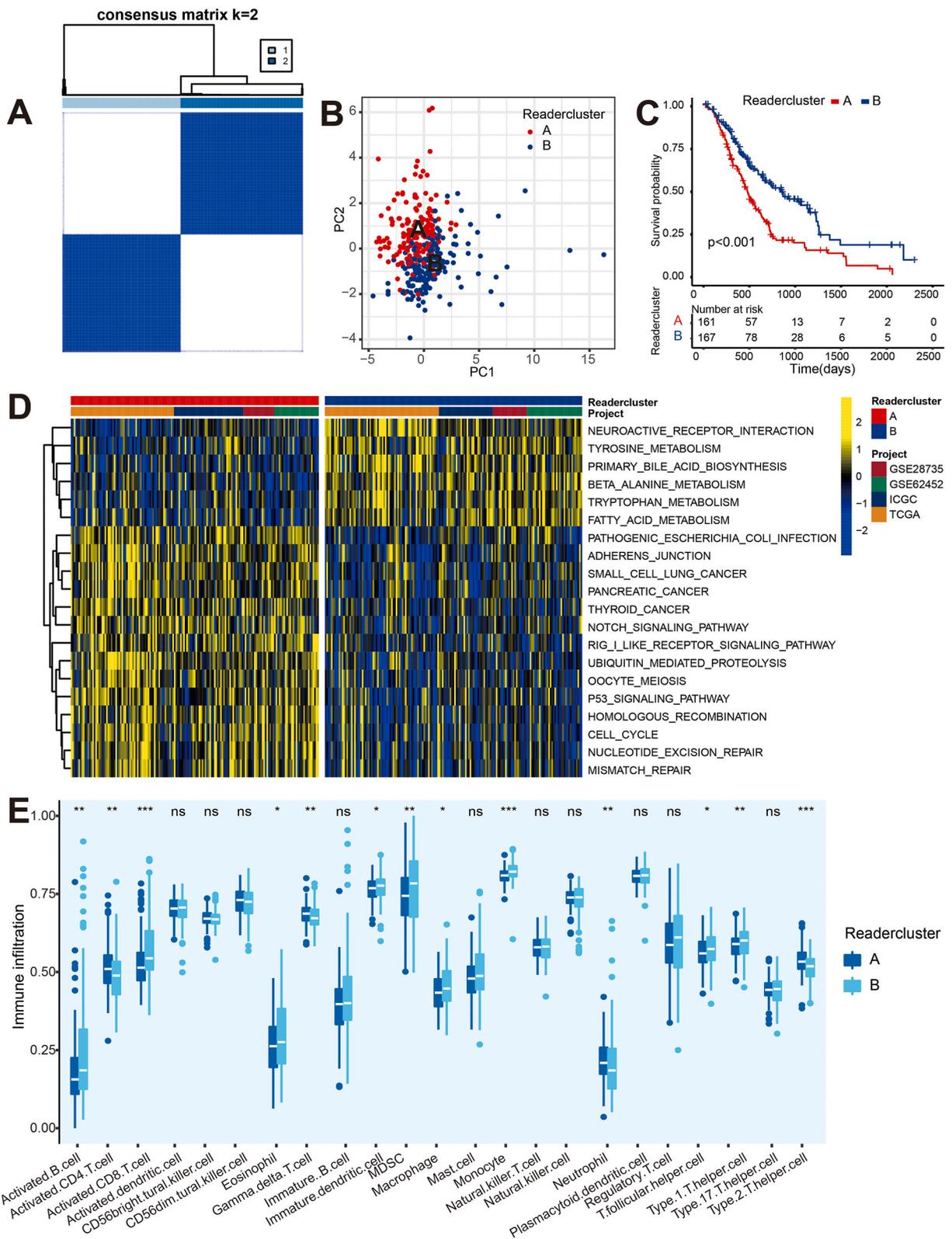
Firstly, we analyzed the mutation variation of m6A readers in PC, three datasets (namely, TCGA_PAAD, ICGC_AU_PAAD and ICGC_CA_PAAD) were selected, all of which have complete gene expression and clinical data. As shown in Figure S3, the overall mutation rate of 14 m6A readers in 335 samples was low, among which the top three genes with mutation rate were SND1, IGF2BP3 and IGF2BP2. From the perspective of mutation type, intronic mutation was the most common mutation in the readers, which indicated that the structural changes caused by intronic mutation might be the reason for the essential biological effects of the m6A readers.

With regards to the CNV, data from TCGA_PAAD datasets were included. Among 140 PC samples with entire CNV data, 57.14% (8/14) of the m6A readers had copy number amplification frequency greater than copy number deletion frequency, indicating that copy number amplification was a relatively common mutation event of m6A readers at the DNA level. In terms of single genes, IGF2BP1 had the highest frequency of CNV events (9.22%), followed by PRRC2A (8.51%) and YTHDF2 (7.09%) (Figure S4A). The copy number increase or decrease of large genome fragments with copy number variation of more than 1 kb is mainly manifested as submicroscopic deletions and duplications, which are important components of genome structural variation. Therefore, we also explored the relationship between CNV events and the expression of m6A readers. When the copy number of m6A readers increased, its mRNA expression also showed an up-regulated trend. Similarly, when the copy number of m6A readers was deleted, its expression was down-regulated accordingly (Figure S4B-O).

For the analysis of AS events of m6A readers in PC, only the TCGA_PAAD dataset with more comprehensive data information was included here. Through the TCGA Splice Seq database, we obtained 139 samples with AS events data, including exon skip (ES), alternate terminator (AT), alternate promoter (AP), alternate acceptor (AA), alternate donor (AD), retained intron (RI) and mutually exclusive (ME). From the level of whole transcriptome genes, ES events in PC samples ranked first (31.81%, 6750/21,218), AT and AP ranked second (17.98%, 3816/21,218) and third (17.55%, 3724/21,218), respectively (Figure S5A). However, for the 14 m6A readers, the overall number of AS events was small. Similarly, the most common type of AS was still ES (40%, 8/20) (Figure S5B). Therefore, exon deletion caused by ES is also an important link in the regulation of PC by m6A readers.

As the most common modification in biology, m6A modification under the epigenetic branch has been widely reported to be closely related to the occurrence and development of PC. To further verify the epigenetic expression of m6A readers, DNA methylation and histone acetylation were analyzed respectively. After downloading PC methylation data from the TCGA portal and conducting data quality control, we obtained the methylation signal value matrix (β value matrix) of m6A reader genes to quantify the methylation level of each gene. As shown in Figure S6, correlation analysis between the expression level of m6A readers and β value shows that, at $p < 0.05$ level, 71.43% of the expression of readers were negatively correlated with methylation levels. Among them, YTHDF3 ($R = -0.51$, $p = 1.9E-13$), IGF2BP2 ($R = -0.5$, $p = 6.3E-13$), The expression of SND1 ($R = -0.4$, $p = 2.5E-08$) had the most significant negative correlation with methylation level, which indicated that most readers were regulated by methylation in PC. For the analysis of the histone acetylation level of the m6A readers, the most common histone H3K27ac was selected as the research object here due to the limitation of the database. Through the aforementioned method, we paired the acetylation data in the Cistrome Data Browser database with the UCSC Browser for visualization. Compared with normal pancreatic cells, the histone H3K27ac acetylation modification peaks in the promoter regions of IGF2BP1, IGF2BP2, IGF2BP3, YTHDF2, YTHDF3 and PRRC2A in PC cell lines were significantly increased (Figure S7). This indicates that the above m6A readers may be regulated by H3K27ac acetylation, which also provides a theoretical basis for subsequent related studies in the future.

In the previous study, we systematically analyzed the m6A readers at the DNA level and epigenetic levels. On the whole, most of the



(caption on next page)

Fig. 2. Construction of specific expression patterns of m6A readers. (A) Heat map based on the expression matrix of 14 m6A readers, with the optimal grouping k value of 2. (B) PCA algorithm diagram shows that Readercluster_A and Readercluster_B groups have better classification effect. (C) Survival curve showed that the prognosis of patients in Readercluster_A group was poor. (D) Heat map of GSVA score based on KEGG background gene set. (E) ssGSEA analysis was performed based on the immune background gene set to judge the infiltration abundance of immune cells.

m6A readers were regulated by SNP, CNV, AS, methylation and acetylation. To further analyze the effect of these readers at the mRNA level. We selected 141 PC samples from the TCGA_PAAD database and 165 normal samples from the GTEx database for differential expression analysis, as shown in Figure S8 mRNA expression levels of IGF2BP2 ($\log_2FC = 5.45$, $FDR = 9.02E-89$), IGF2BP3 ($\log_2FC = 2.41$, $FDR = 4.47E-30$) and IGF2BP1 ($\log_2FC = 1.58$, $FDR = 2.28E-86$) in PC samples were significantly higher than those in normal tissues, which may be related to the high mutation frequency of IGF2BP2 and IGF2BP3 mentioned above. To further explore the relationship between m6A readers and the prognosis of PC patients, the 14 readers were divided into high expression group and low expression group according to their median expression level, and the overall survival (OS) between the two groups was compared. For YTHDC1, YTHDC2, YTHDF1, YTHDF2, YTHDF3 and PRRC2A, the high expression group of PC patients had better OS, while other readers showed opposite results (Figure S9A–N). At the same time, univariate Cox analysis was performed for 14 m6A readers, and the risk network diagram was drawn, which was consistent with the results of the survival analysis described above (Figure S9O). These results indicate that YTHDC1, YTHDC2, YTHDF1, YTHDF2, YTHDF3 and PRRC2A are protective factors in PC patients. IGF2BP1, IGF2BP2, IGF2BP3, FMR1, HNRNPA2B1, EIF3A, LRPPRC and SND1 were risk factors.

3.2. Construction of specific expression patterns of m6A readers

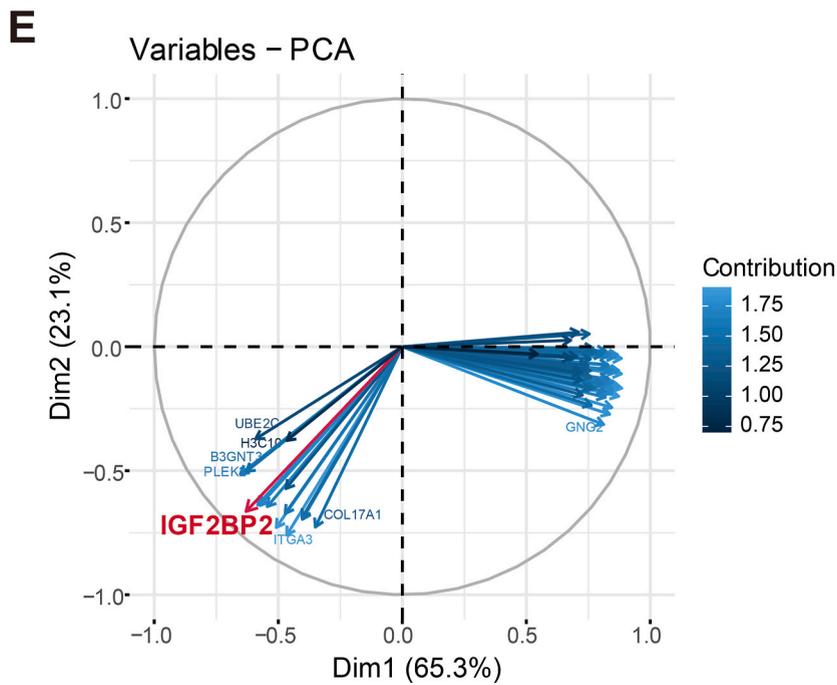
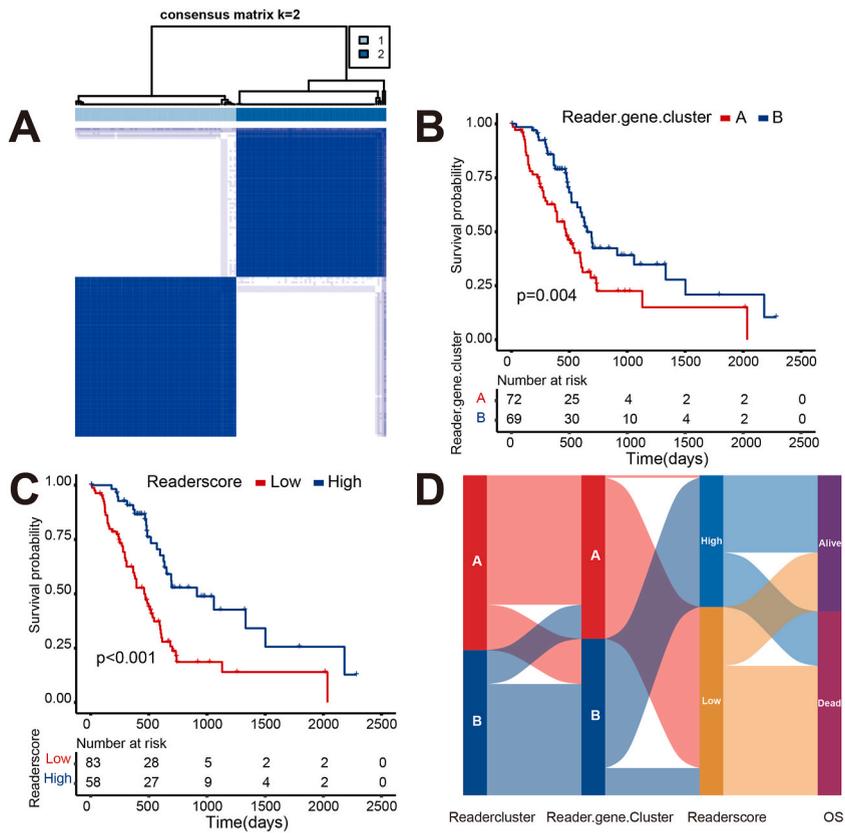
In order to explore the specific expression pattern of m6A readers, a total of 328 PC samples were included (TCGA_PAAD, ICG-C_AU_PAAD, GSE28735 and GSE62452 datasets). Based on the expression of 14 m6A readers, we conducted an unsupervised consensus clustering analysis. According to the heatmap of consensus clustering (Fig. 2A), the optimal $k = 2$ could be determined. A total of 328 PC patients were divided into two subgroups: Readercluster_A ($n = 161$, 49.09%) and Readercluster_B ($n = 167$, 50.91%). To verify the reliability of the grouping, PCA analysis was performed on the expression profile of 14 m6A readers. As shown in Fig. 2B, in the PCA diagram, the samples classified as Readercluster_A (red) and Readercluster_B (blue) have high discrimination, indicating that the grouping effect is robust. In addition, to explore the effect of different m6A reader gene expression patterns on the prognosis of PC patients, a survival analysis was performed. The samples in the Readercluster_A group were significantly lower than those of the Readercluster_B group on OS ($p < 0.001$), suggesting that the Readercluster_A expression pattern may lead to poor prognosis in PC patients (Fig. 2C).

To further detect the differences in biological behaviors among different m6A reader expression patterns, we performed GSVA analysis using the KEGG pathway as the background gene set. The expression pattern of the Readercluster_A subtype was mainly enriched in oncogenic pathways: Notch pathway, p53 pathway and pancreatic cancer pathway, while the expression pattern of Readercluster_B subtype was mainly enriched in metabolic pathways: fatty acid metabolism pathway, alanine metabolism pathway and tryptophan metabolism pathway (Fig. 2D). These results provide a direction for further research on the oncogenic mechanism downstream of the m6A readers.

We also preliminarily explored the difference in tumor immune microenvironment (TIME) between the specific expression patterns of the two m6A readers. ssGSEA analysis was performed using immune gene sets for the specific expression patterns of the two m6A readers. On the whole, the abundance of immune cell infiltration of the Readercluster_B expression pattern was significantly higher than that of Readercluster_A. From the perspective of the specific degree of cell infiltration, the Readercluster_B expression pattern has abundant innate immune cell infiltration, such as eosinophils, MDSC and dendritic cells (Fig. 2E). These results provided a basis for the better prognosis of the Readercluster_B expression pattern.

Next, to further explore the characteristics of the expression pattern of these m6A readers in different clinical features and biological behaviors, 141 PC samples in the TCGA_PAAD dataset with the most comprehensive clinical information were selected as the only study subjects for all subsequent analyses. Firstly, we used the expression matrix of 14 m6A readers in TCGA_PAAD dataset as the input object for consensus clustering analysis again. 141 PC patients were also divided into Readercluster_A ($n = 77$, 54.61%) and Readercluster_B ($n = 64$, 45.39%) subgroups (Figure S10A–C).

For checking the relationship between different m6A reader gene expression patterns and PC molecular subtypes, we first obtained the characteristic gene sets of three classical molecular classification (Moffitt classification, Collisson classification and Bailey classification) in the reference literature [4,20,21]. Then consensus clustering analysis was performed on the expression matrix of these characteristic gene sets in TCGA_PAAD. The characteristic gene sets of the three molecular types were well clustered in the TCGA_PAAD dataset, so as to obtain the data of the three molecular classifications in the TCGA_PAAD dataset: (1) Moffitt classification: Classical ($n = 66$, 46.81%), Basal ($n = 75$, 53.19%); (2) Collisson classification: Classical ($n = 36$, 25.53%), Exocrine ($n = 66$, 46.81%) and QM-PDA ($n = 39$, 27.66%). (3) Bailey classification: squamous ($n = 44$, 31.21%), progenitor ($n = 36$, 25.53%), immunogenic ($n = 21$, 14.89%) and Aberrantly Differentiated Endocrine Exocrine (ADEX) ($n = 40$, 28.37%) (Figure S10D–F). Finally, visual analyses were performed by adding column annotations (clinical information) to expression heatmaps. 84.1% (37/44) of the squamous type samples (with the worst prognosis) were found in the samples of the Readercluster_A group, and 85.71% (66/77) of the samples of the Readercluster_A group had mutations in the m6A readers (Figure S10G). These results may confirm that Readercluster_A was significantly associated with molecular subtypes with poor prognosis. In addition, among the 14 m6A readers, IGF2BP2 and IGF2BP3 were significantly differentially expressed between Readercluster_A group and Readercluster_B group (IGF2BP2, $p < 0.01$; IGF2BP3, p



(caption on next page)

Fig. 3. Generation of m6A readers phenotype-related expression patterns and screening of key gene. (A) Heat map of concordance cluster analysis based on 140 genes related to both m6A reader phenotype and prognosis showed that the optimal grouping k value was 2; (B) Survival curve showed that the Reader. gene.cluster_A expression pattern had a poor prognosis. (C) Survival curve showed that the survival rate of patients decreased significantly with the decrease of readerscore; (D) The correlation impact map of the three subtypes showed that the expression patterns of Readercluster type, Reader. gene.Cluster type and Readerscore type were highly consistent. (E) PCA plots of expression patterns associated with two m6A reader phenotypes. The results showed that PC1+PC2 could explain 88.14% of the data variation, and the variable IGF2BP2 had the highest contribution.

< 0.01), indicating that IGF2BP2 and IGF2BP3 may be key molecules with different expression patterns of m6A readers.

3.3. Generation of m6A readers phenotype-related expression patterns and screening of key gene

We found that the gene expression patterns of Readercluster_A and Readercluster_B in PC were significantly different, and affected tumor progression, metabolism and other pathways. To further explore and quantify the reasons for the differences in gene expression patterns of different m6A readers, we constructed a phenotype-related scoring model for m6A readers.

Firstly, in the TCGA_PAAD cohort, we performed differential analysis on the gene expression patterns of Readercluster_A and Readercluster_B in PC, and screened 1468 DEGs related to the phenotype of m6A readers. Univariate Cox regression analysis was performed on 1468 differentially expressed genes to screen out 140 differentially expressed genes that were significantly associated with prognosis. To explore the specific gene expression patterns associated with different m6A reader phenotypes, we performed consensus cluster analysis on 140 genes associated with both the phenotype and prognosis of m6A readers. We obtained unique gene expression patterns related to the phenotype of two groups of m6A readers. They were defined as Reader. gene.cluster_A (n = 72, 51.06%) and Reader. gene.Cluster_B (n = 69, 48.94%), respectively (Fig. 3A). Survival analysis showed that the expression pattern of Reader. gene.cluster_A had poor prognosis (p = 0.004) (Fig. 3B). So far, we found that the expression pattern of Reader. gene.cluster was highly consistent with that of Readercluster: (1) Among the samples classified as Reader. gene.cluster_A group, 79.17% of the samples were also classified as Readercluster_A group; Among the samples classified as Reader. gene.Cluster_B group, 71.01% of the samples were also classified as Readercluster_B group; (2) The survival time of PC patients classified as Reader. gene.cluster_A group and Readercluster_A group was significantly lower than that of patients classified as Reader. gene.cluster_B group and Readercluster_B group; (3) IGF2BP2 and IGF2BP3 were significantly differentially expressed in Reader. gene.cluster and Readercluster. These results indicated that there were indeed two specific modification patterns related to m6A readers in PC, and IGF2BP2 and IGF2BP3 may be the key regulatory genes.

To further quantify the expression patterns related to the phenotype of m6A readers, a scoring system was constructed by PCA method. The best cutoff value (0.74) was used to divide into two subgroups: Readerscore_high (n = 58, 41.13%) and Readerscore_low (n = 83, 58.87%). As shown in Fig. 3C, it can be found from the survival curve that with the reduction of Readerscore, the survival rate of patients significantly decreased. To explore the connection between these three subgroups mentioned above, we used the alluvial plot to more intuitively show the connection between sample attributes. The lower the Readerscore of PC patients, it is usually classified as Readercluster_A and Reader. gene.cluster_A group, and its survival time is also worse (Fig. 3D). These results confirmed that our constructed Readerscore system can effectively quantify expression patterns related to m6A readers' phenotypes.

Next, we performed contribution analysis on the expression data which used to construct the model to explore the core genes between the two different expression patterns. As shown in Fig. 3E, in this PCA plot, the explanation rate of principal component 1 (PC1) to the overall difference was 65.3%, the explanation rate of principal component 2 (PC2) to the overall difference was 23.1%, and the explanation rate of PC1+PC2 to the data difference reached 88.4%, indicating that the PCA had a good dimensionality reduction effect. Secondly, IGF2BP2 is the only selected m6A readers among the top 100 genes with the highest contribution to PC1, suggesting that IGF2BP2 can effectively explain the variation of the overall data, which further suggests that IGF2BP2 was the key gene leading to the change of expression pattern related to m6A readers' phenotypes.

3.4. Functional annotation of IGF2BP2-mediated modification patterns in PC

To further investigate the potential biological behavior of IGF2BP2-mediated patterns, the "clusterProfiler" R package was utilized to screen the differential genes in different IGF2BP2 expression patterns, we used the RNA-Seq data of PC in TCGA, ICGC and GEO databases to determine the median expression value of IGF2BP2. 328 PC samples were divided into IGF2BP2_high and IGF2BP2_low subgroups and 5044 differential genes were obtained, which were defined as the differential genes related to IGF2BP2-mediated phenotype. The top 2000 differential genes with significant differences were selected for heatmap (Figure S11A) to judge the efficiency of differential analysis. Then, we conducted GO and KEGG pathway enrichment analysis on the above 5044 differentially expressed genes. IGF2BP2_high subgroup was mainly enriched in cell cycle-related pathways ("mitotic cell cycle", "cell cycle"), tumor-related EMT and transcriptional dysregulation pathways ("epithelial cell differentiation", "Regulation of transcription by RNA polymerase", "transcriptional misregulation in cancer"; The IGF2BP2_low subgroup was mainly enriched in immune effector process ("immune effector process", "immune response", "regulation of immune system process", "primary immunodeficiency") (Figure S11B–C). All these results confirmed that the high expression of IGF2BP2 in PC patients may promote the malignant biological phenotype of PC by suppressing the immune system.

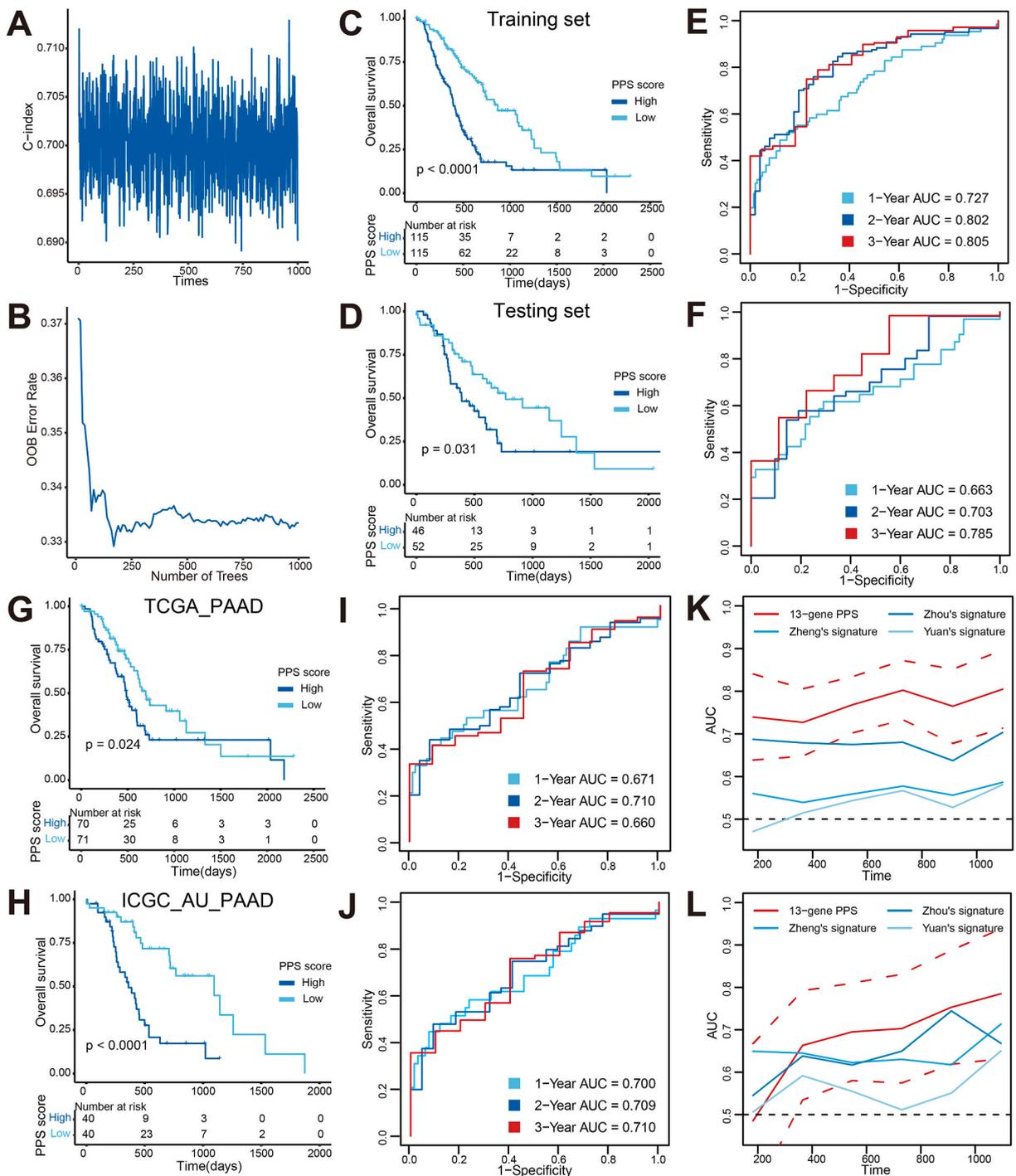


Fig. 4. Development of poor prognostic signatures based on IGF2BP2-mediated modification patterns via machine learning algorithms. (A) C-index change curve of RSF algorithm after 1000 iterations; (B) curve of out-of-pocket sample error rate of RSF algorithm; According to the median value of PPS score, the survival curve was drawn in the training set (C), testing set (D), TCGA_PAAD dataset (G), and ICGC_AU_PAAD dataset (H), the results showed that the higher the PPS score, the worse the prognosis of PDAC patients. The area under the 1-year, 2-year and 3-year survival curve predicted by PPS model in the training set (E), testing set (F), TCGA_PAAD dataset (I) and ICGC_AU_PAAD dataset (J). (K) Comparison of prediction performance between PPS model and other models in the training set; (L) Comparison of prediction performance between PPS model and other models in the testing set.

3.5. Development of poor prognostic signatures based on IGF2BP2-mediated modification patterns via machine learning algorithms

Based on the 5044 differential genes associated with the IGF2BP2 phenotype obtained by differential analysis, we explored prognostic gene markers for PC. 328 samples in TCGA, ICGC and GEO databases were divided into training set (n = 230) and testing set (n = 98) according to the ratio of 7:3. For the samples in the training set, we first performed univariate Cox regression analysis to obtain 889 genes associated with prognosis. The Bootstrapping algorithm was used to obtain 181 genes that were significantly associated with prognosis in 900 sampling sessions. Then, RSF analysis based on the minimum depth (MD-RSF) method was used to determine the final prognostic gene markers. In the RSF algorithm with 1000 iterations (Fig. 4A–B). The 13 gene sets (FNDC3B, L1CAM, PLXNA1, HMGA2, FAM110B, FAM83A, COX7A1, PMAIP1, KIF20B, SPDL1, SNCG, TGM2 and MUC16) included when C_index reached the maximum (0.7128982) were selected as poor prognostic signatures, defined as PPS associated with IGF2BP2 phenotype. Finally, the multivariate Cox regression coefficients of the above 13 PPS were calculated and the PPS scoring model was constructed. The specific PPS scoring formula was as follows:

$$\begin{aligned} \text{PPS score} = & \text{Expr}(\text{FNDC3B}) * 0.16009993 + \text{Expr}(\text{L1CAM}) * 0.28052790 + \text{Expr}(\text{PLXNA1}) * 0.22958423 + \text{Expr}(\text{HMGA2}) \\ & * 0.14046140 - \text{Expr}(\text{FAM110B}) * 0.18477060 + \text{Expr}(\text{FAM83A}) * 0.04911509 - \text{Expr}(\text{COX7A1}) * 0.23187426 \\ & - \text{Expr}(\text{PMAIP1}) * 0.08461359 + \text{Expr}(\text{KIF20B}) * 0.16976500 - \text{Expr}(\text{SPDL1}) * 0.09529266 + \text{Expr}(\text{SNCG}) * 0.06892878 \\ & - \text{Expr}(\text{TGM2}) * 0.01766628 + \text{Expr}(\text{MUC16}) * 0.04981059 \end{aligned}$$

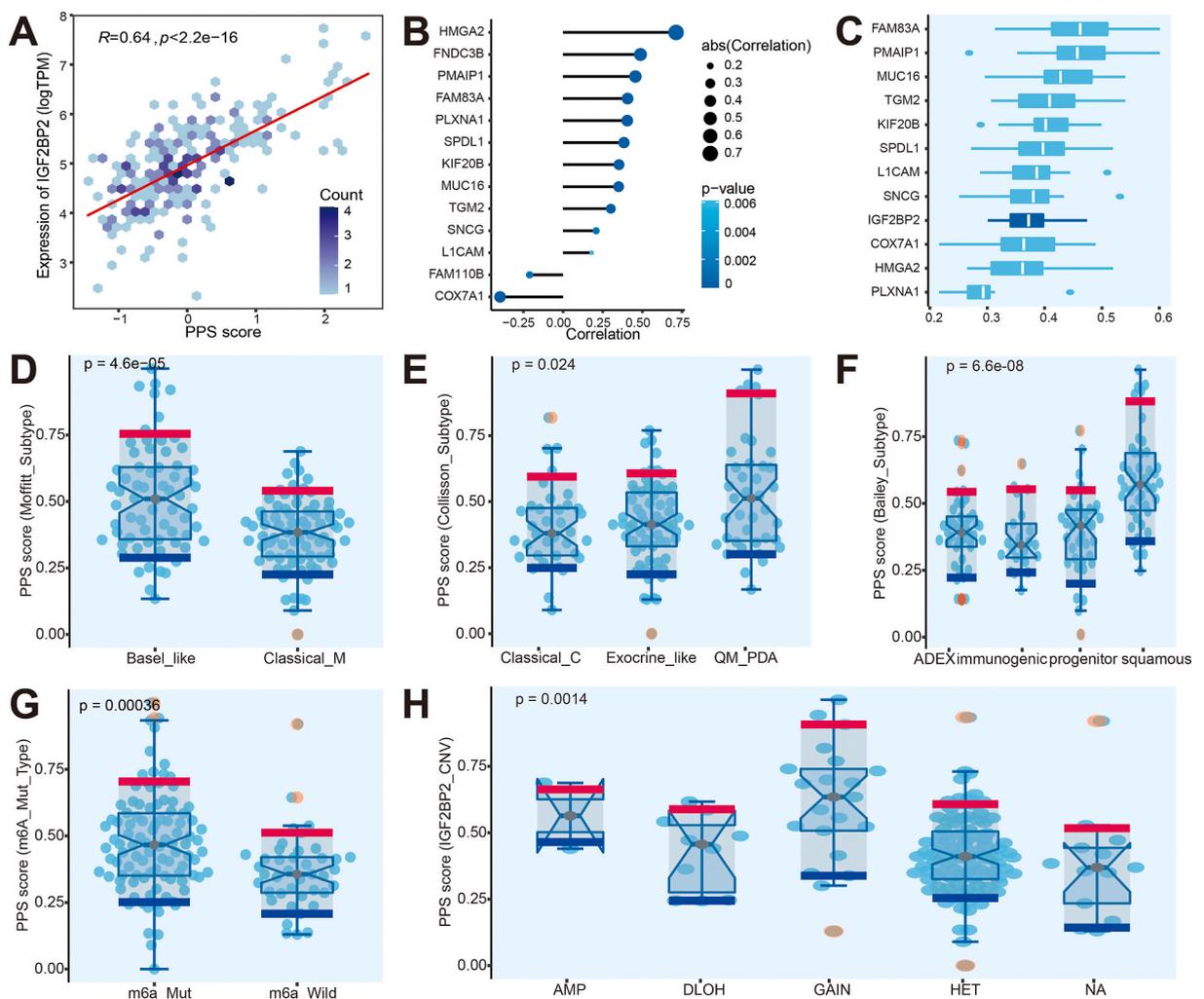


Fig. 5. Exploration of the PPS-related molecular subtype characteristics. (A) Correlation map between PPS score and IGF2BP2 expression; (B) correlation map between 13 PPS genes and IGF2BP2 expression; (C) Box plot of enrichment similarity analysis between 13 PPS genes and IGF2BP2 pathway. Correlation analysis between PPS score and Moffitt classification (D), Collisson classification (E), Bailey classification (F), m6A mutation status (G), IGF2BP2 copy number variation event (H).

To evaluate the prognostic value of the model, PC samples were divided into two subgroups according to the median PPS score, namely PPS_high and PPS_low. Then, survival analysis was performed on the training set, testing set, TCGA_PAAD dataset and ICGC_AU_PAAD dataset, and survival curves were drawn. In the four independent data sets mentioned above, higher PPS score was positively correlated with shorter overall survival (training set, $p < 0.0001$; testing set, $P = 0.031$; TCGA_PAAD, $p = 0.024$; ICGC_AU_PAAD, $p < 0.0001$) (Fig. 4C, D, G, H).

In order to further verify the prediction efficiency of this model, the receiver operating characteristic curve (ROC) was selected for visual analysis. In the training and testing sets, the PPS scoring model was effective for patients in 1 year (training set, AUC = 0.727; testing set, AUC = 0.663; TCGA_PAAD, AUC = 0.671; ICGC_AU_PAAD, AUC = 0.700), 2 years (training set, AUC = 0.802; testing set, AUC = 0.703; TCGA_PAAD, AUC = 0.710; ICGC_AU_PAAD, AUC = 0.709) and 3 years (training set, AUC = 0.805; testing set, AUC = 0.785; TCGA_PAAD, AUC = 0.660; ICGC_AU_PAAD, AUC = 0.710) had good efficacy in predicting survival (Fig. 4E, F, I, J). In addition, it has been reported that several scholars have also constructed prognostic models for PC patients, such as Zheng [22], Zhou [23] and Yuan [24]. In order to verify whether PPS score model has better predictive power compared with the prognostic model constructed by the above three teams, ROC analysis was also performed. PPS scoring model in the training set (PPS: 0.778 vs Zheng: 0.567, Zhou: 0.687, Yuan, 0.554) and testing set (PPS: 0.717 vs Zheng: 0.662, Zhou: The average AUC values of 0.651, Yuan, 0.584) were higher than those of the other three prognostic models (Fig. 4K-L). The above results fully verify the robustness and predictive effectiveness of our PPS scoring model.

Finally, in order to explore the relationship between 13 PPS genes and IGF2BP2, we conducted a correlation analysis. As shown in Fig. 5A–B, the expression of IGF2BP2 was highly correlated with PPS score ($R = 0.64$; $p = 2.2E-16$) and showed a significant correlation with the expression of 13 PPS genes. In addition, we also used the “GoseSim” R package to analyze the enrichment pathway similarity of 14 genes (IGF2BP2 and 13 PPS genes) and the functional similarity distribution map among these genes is shown in Fig. 5C. These results indicated that IGF2BP2 is not only highly correlated with the PPS gene at the transcriptome level, but also has high functional similarity with the PPS gene.

3.6. Exploration of the PPS-related molecular subtype characteristics, biological processes and drug targets

At present, the classical clinical molecular subtypes for PC include Moffitt subtype, Collisson subtype and Bailey subtype. To explore the correlation between PPS score and molecular subtypes in PC, we conducted a correlation analysis between them. The Basel_like subtype in the Moffitt classification, the QM-PDA subtype in the Collisson classification, and the Squamous subtype in the Bailey classification had higher PPS scores, which have been shown to have a poor prognosis (Fig. 5D–F). In addition, we also noted that patients with higher PPS scores had a higher frequency of m6A gene mutation events and IGF2BP2 copy number amplification

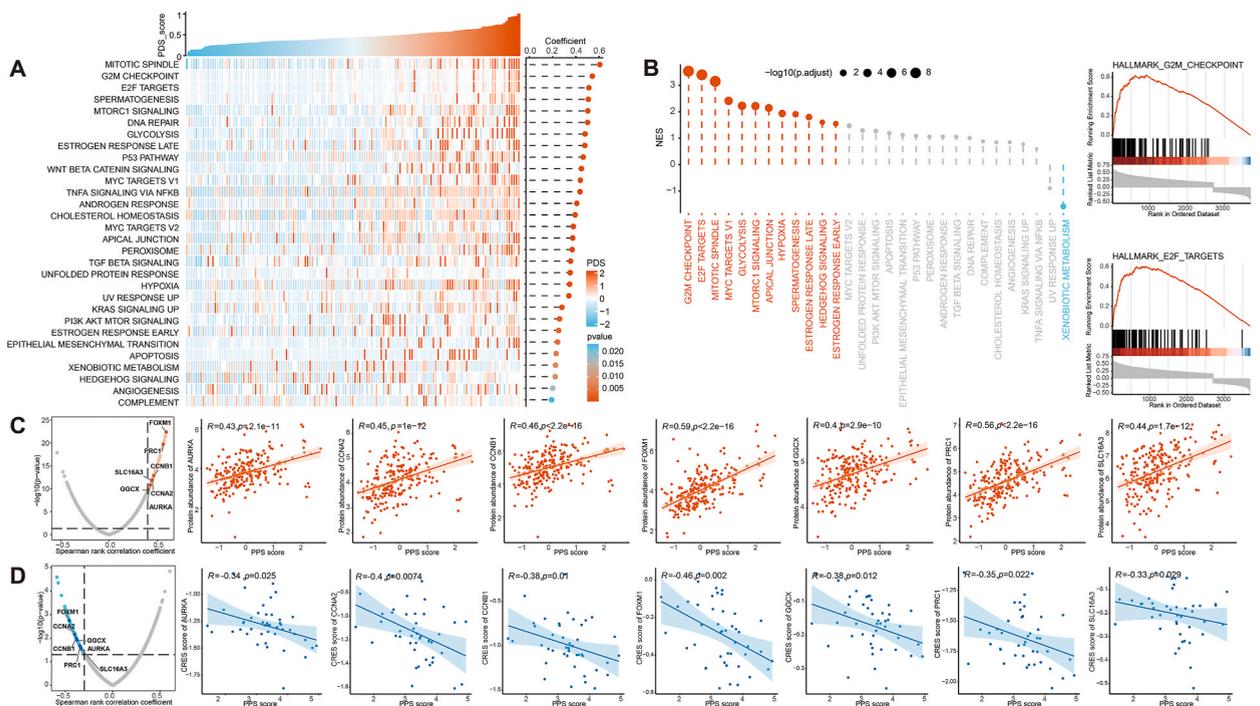


Fig. 6. Identification of PPS-related biological processes and drug targets. (A) The PDS score was calculated based on PPS model grouping, and the heat map of PDS score was drawn. (B) GSEA analysis was performed based on PPS model grouping to laterally verify the reliability of enriched pathways in patients with high PPS scores. (C) Volcano plot and scatter plot of correlation between PPS score and drug gene targets. (D) Correlation between PPS score and CERES score of drug gene target volcano plot and scatter plot.

events (Fig. 5G–H). All these results indicated that PC patients with higher PPS scores have worse prognosis, which is consistent with the conclusion confirmed in the previous part.

To investigate which biological processes contribute to the malignant phenotype in patients with high PPS scores, Pathifier analysis and GSEA analysis based on transcriptomics were performed. Firstly, the expression profiles of PPS_high and PPS_low subgroups were used for Pathifier analysis to obtain the PDS score of each sample, and the correlation between the PDS score and PPS score was calculated. If the correlation coefficient of a pathway (biological process) is positive, it can be considered that this biological process may be responsible for the poor prognosis of patients with high PPS score. Cell cycle and proliferation-related pathways (“mitotic spindle”, “G2M checkpoint”, “E2F targets”) had the highest correlation with poor prognosis (Fig. 6A). Next, to further verify the reliability of the above conclusions, we performed GSEA analysis using RNA-seq data from samples in the training set. First, the correlation coefficient between the PPS score of the sample and RNA-Seq data was calculated, and then the correlation coefficient was used as the input for GSEA analysis. As shown in Fig. 6B, genes with positive correlation coefficients were also enriched in cell cycle and proliferation-related pathways (“mitotic spindle”, “G2M checkpoint”, “E2F targets” and “Myc targets v1”). Through the above two methods, the obtained biological processes related to the PPS score were highly consistent, so we can infer that the dysregulation of the cell cycle and proliferation process may be the reason for the poor prognosis of patients with high PPS.

In patients with high PPS scores, proteins that are highly positively correlated with PPS scores may be potential therapeutic targets. However, most proteins fail to exert their targeting effects due to the lack of active sites to which small molecule compounds can bind. Therefore, to explore potential drug therapeutic targets for patients with high PPS scores and poor prognosis, we conducted the following analysis: First, the correlation between protein expression profile data and PPS score after drug intervention was analyzed. 448 targeted proteins were obtained by the standard of $R > 0.4$ and $p < 0.05$. Secondly, correlation analysis was performed between the CERES score in the DepMap database and the PPS score of PC cell lines in CCLE. 87 targeted proteins were screened out according to the standard of $R < -0.3$ and $p < 0.05$. 7 potential protein targets (FOXN1, PRC1, CCNB1, SLC16A3, CCNA2, GGCX and AURKA) were screened out (Fig. 6C–D). These results suggest that for patients with high PPS scores, effective therapeutic results may be achieved by inhibiting the above 7 protein targets.

3.7. Identification of potential therapeutic agents for high PPS score in PC

To screen potential therapeutic compounds related to the PPS model, this part of the analysis was based on the gene expression data and drug sensitivity data (AUC) of PC cell lines in the CTRP and PRISM databases. The information of 437 compounds in the CTRP database and 1438 compounds in the PRISM database were selected for subsequent analysis after removing the duplicated compound

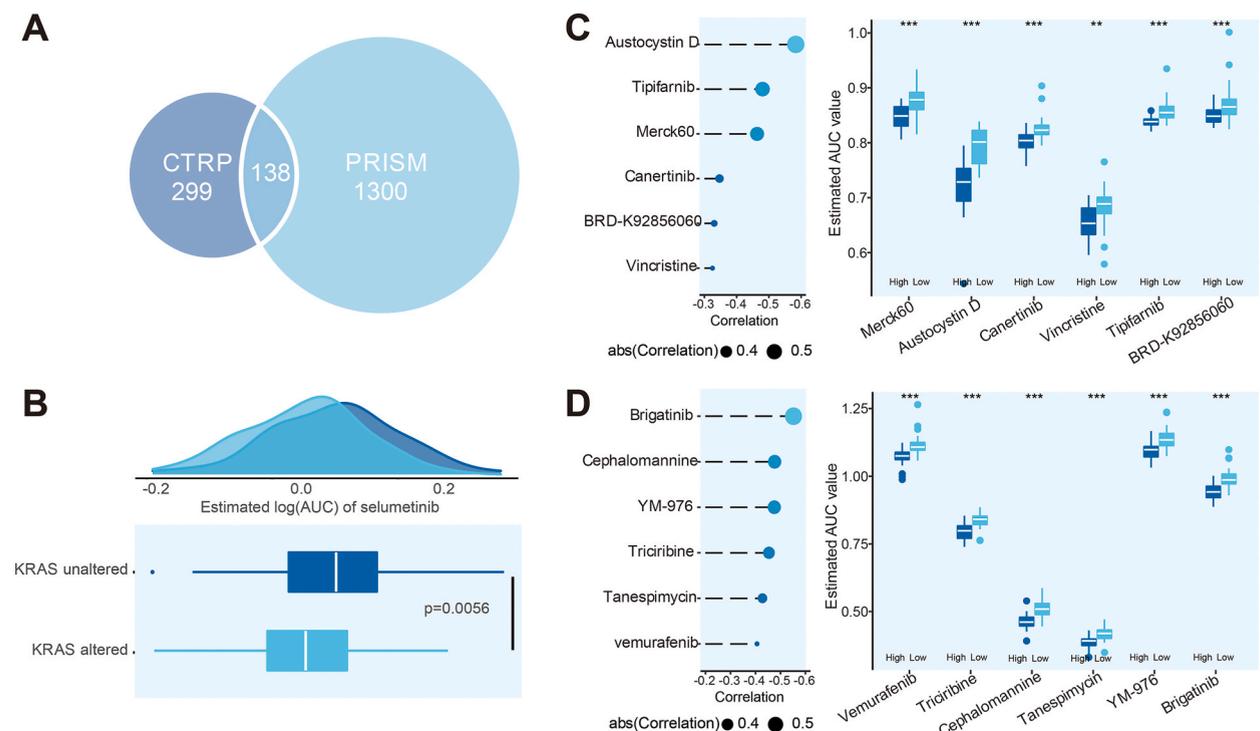


Fig. 7. Identification of candidate agents with higher drug sensitivity in high-PPS score patients. (A) Compound information summary Venn diagram from CTRP and PRISM database; (B) predictive value of drug susceptibility between Kras mutation and non-mutation groups. (C) Six potential compounds derived from CTRP were screened by Spearman correlation analysis and drug sensitivity analysis; (D) Six potential compounds derived from PRISM were identified by Spearman correlation analysis and differential susceptibility analysis.

information in the two databases and removing the PC cell lines with NA values over 20% or originating from hematopoietic and lymphoid tissues (Fig. 7A).

We first obtained the AUC values of each compound in PC samples using the “pRRophetic” package with a built-in ridge regression model based on the gene expression profile data of PC samples. To verify the credibility of the AUC value, we conducted the following analysis: Selumetinib is a PI3K pathway inhibitor used in the treatment of PC, and a recent study found that PC patients with KRAS pathway mutations had longer survival after treatment than PC patients without KRAS pathway mutations [25]. Therefore, according to the mutation status of the KRAS pathway, PC patients were divided into KRAS-altered and KRAS-unaltered subgroups. Wilcoxon rank-sum test was used to compare the sensitivity of the two subgroups to Selumetinib. As shown in Fig. 7B, the AUC of PC patients in the KRAS-altered group was significantly decreased ($p = 0.0056$), which was consistent with the clinical findings of Selumetinib above. The above results further confirmed the reliability of our predicted AUC. For patients with high PPS scores, potential compounds were screened from CTRP database and PRISM database. 6 compounds from the CTRP database (Austocystin D, Tipifarnib, Merck60, Canertinib, BRD-K92856060 and Vincristine) and 6 compounds from PRISM database (Brigatinib, Cephalomannine, YM-976, Triciribine, Tanespimycin and Vemurafenib) were obtained (Fig. 7C–D). The above 12 compounds had lower predictive AUC values in the samples with higher PPS scores, and their predictive AUC values were significantly negatively correlated with PPS scores.

Although the sensitivity of these 12 compounds to patients with high PPS scores was high, the analysis of these 12 compounds alone does not adequately indicate the therapeutic potential of these 12 compounds in patients with PC. Therefore, as shown in Fig. 8, we further conducted a multi-dimensional study. First, we used CMap analysis to explore compounds whose gene expression pattern was contrary to the specific expression pattern of PC (namely, gene expression was increased in PC tissues, but the expression was down-regulated after the intervention of some compounds). The CMap scores of three compounds (tipirifanib, vemurafenib and YM-976) were all less than -90 . These compounds may have potential therapeutic effects on PC. Second, the expression levels (including mRNA and protein levels) of the target of the candidate compounds were analyzed differently between PC tissues and normal tissues. The larger the difference, the greater the potential of the candidate compounds to treat PC. Third, we conducted a comprehensive literature search in the PubMed database (<https://www.ncbi.nlm.nih/>), in order to search for candidate compounds for the treatment of PC experimental and clinical evidence. Taken together, there is strong evidence that both tipifarnib and vemurafenib have the best therapeutic potential for patients with high PPS scores.

4. Discussion

More and more evidence has shown that m6A modification plays an indispensable role in inflammation, innate immunity and

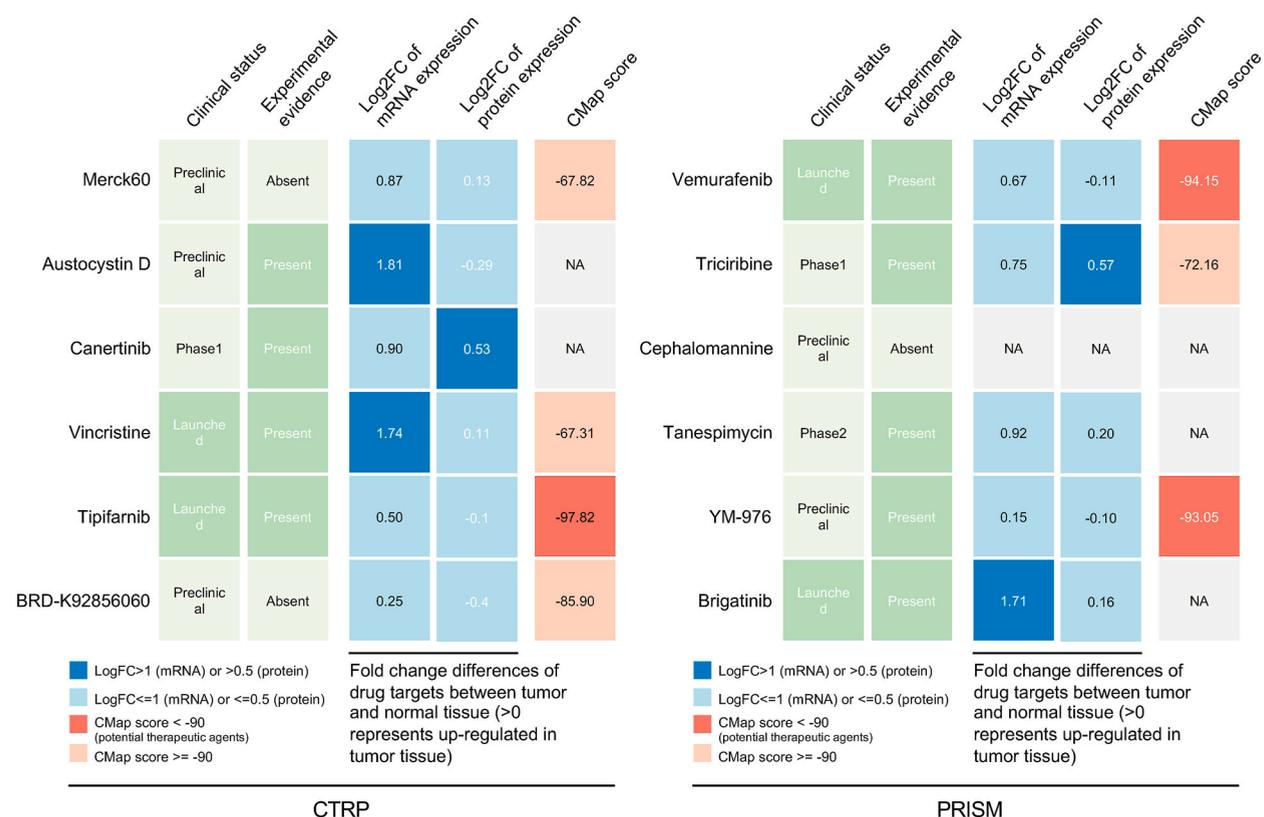


Fig. 8. Multiangle validation was performed to identify the most likely potential therapeutic compounds for patients with high PPS scores.

tumor regulation through the interaction of a variety of m6A regulatory factors [26]. Since most of the current studies focus on a single regulator, there is no comprehensive and in-depth understanding of the overall characteristics of malignant biological behavior of PC mediated by the comprehensive action of multiple m6A readers. To this end, based on the 14 m6A readers, we found two distinct expression patterns of m6A readers, and determined that the two expression patterns had significant heterogeneity in pathway enrichment, tumor immune microenvironment, and clinical classification. Finally, through contribution analysis of the constructed scoring model, IGF2BP2 was identified as the core gene affecting the related expression pattern of m6A readers. IGF2BP2 is an RNA-binding protein that regulates multiple biological processes. Previously, IGF2BP2 was thought to be a type 2 diabetes (T2D)-associated gene [27]. Indeed IGF2BP2 modulates cellular metabolism in human metabolic diseases such as diabetes, obesity and fatty liver through post-transcriptional regulation of numerous genes in multiple cell types. Emerging evidence shows that IGF2BP2 is an m6A reader that participates in the development and progression of PC by communicating with different RNAs such as miRNAs, mRNAs and lncRNAs [28]. Liu Y found that IGF2BP2 and miR-671-3p could interact with lncRNA-PACERR, inducing pro-tumor macrophages in PC [17]. Also, IGF2BP2 and lncRNA-DANCR work together to promote cancer stemness-like properties and pancreatic cancer pathogenesis [29].

In the past decade, the molecular biology of PC has been greatly enhanced by high-throughput sequencing analysis of large samples, which allows researchers to detect key events in the development of PC. With the recognition of PC-specific mismatch repair defects and homologous recombination defects, germline variants that increase the risk of PC have been identified, and these defects are conferring sensitivity to PC targeted therapy, which marks that the treatment strategy for PC has officially entered the era of precision medicine. In our study, we have confirmed that IGF2BP2 is a marker gene with certain molecular characteristics in PC, and higher expression of IGF2BP2 is closely related to worse prognosis. Considering the unique clinical features of PC patients with high IGF2BP2 expression, it is necessary to develop targeted treatment strategies for these patients. Therefore, an effective prognostic prediction model, the PPS score model, was developed based on the different expression patterns of IGF2BP2 in PC. Moreover, compared with other prognostic models, the PPS model based on different IGF2BP2 expression patterns has more significant and better predictive power in terms of prognosis. These results enable us to use precise prognostic prediction methods for patients with similar biological patterns to achieve a more reliable prognostic treatment effect.

PC is highly heterogeneous among individuals, and no study has found a treatment that is suitable for all patients with PC. In addition, due to the lack of corresponding biomarkers, patients with advanced PC have not achieved satisfactory therapeutic effects. Therefore, it is of great interest to find tailored treatment strategies for specific populations. The PPS model we constructed not only has good prognostic efficacy, but also can be used as a biomarker to guide targeted therapy. In this study, we identified seven potential therapeutic targets (FOXM1, PRC1, CCNB1, SLC16A3, CCNA2, GGCX, and AURKA) and two agents (tipirafenib and vemurafenib) for PC patients with high PPS scores.

The protein encoded by FOXM1 belongs to the forkhead box family of transcription factors, which is a transcription factor related to cell proliferation and is overexpressed in a variety of human malignant tumor cells [30]. Studies have shown that silencing FOXM1 can down-regulate the expression of ATX, thereby inhibiting the proliferation, invasion and metastasis of pancreatic cancer cells [31]. PRC1 is a member of the polycomb inhibitory complex, and inhibition of PRC1 can prevent immune escape and angiogenesis in double negative breast cancer cells [32]. The protein encoded by CCNB1 is a cyclin, which is mainly involved in the regulation of cell mitosis. Silencing CCNB1 can inhibit the proliferation of pancreatic cancer cells and promote cell senescence by activating p53 signaling pathway [33]. CCNA2-encoded proteins also belong to the highly conserved cyclin family, which binds to and activates cyclin-dependent kinase 2, thereby promoting the transition from G1/S to G2/M. The up-regulation of CCNA2 expression in KRAS mutant gastric cancer cell lines and primary tumors leads to increased sensitivity to PLK1 inhibitors, and it can predict the sensitivity of PLK1 inhibitors to the treatment of advanced gastric cancer [34]. SLC16A3 belongs to the solute carrier family, and the depletion of its encoded protein can promote the increase of reactive oxygen species and metabolism in pancreatic cancer cells, thereby promoting cell death [35]. The protein encoded by AURKA is a cell cycle regulating kinase, and inhibition of AURKA expression can reduce its phosphorylation level in pancreatic cancer cells and promote cell necrosis [36]. The above results suggest that the seven targets identified so far play a specific role in malignancies, especially PC, suggesting that the corresponding therapeutic targets developed for PC patients with high-risk IGF2BP2 expression patterns are feasible.

Tipirafenib, a farnesyltransferase inhibitor, has been widely used as a targeted therapy for the treatment of a variety of solid tumors. Considering that farnesylation is a key step in Ras protein membrane anchoring required for Ras activity, and KRAS mutations exist in 70%–90% of PC, inhibition of farnesyltransferase can inhibit KRAS gene function, thereby inhibiting the progression of PC [37]. The efficacy of tipirafenib has been demonstrated in a phase III clinical study in patients with advanced PC. The results showed that gemcitabine combined with tipirafenib can enhance anti-PC efficacy and is more tolerated by patients compared with gemcitabine alone [38]. In addition, tipirafenib combined with atorvastatin and celecoxib can synergistically reduce the sphere formation ability of PC cells, and can significantly inhibit cell proliferation and promote apoptosis of sphere-forming cells [39].

Vemurafenib is a low molecular weight compound used to inhibit the mutated serine-threonine kinase BRAF by selectively binding to the ATP-binding site of the BRAF V600E kinase, thereby inhibiting its activity. vemurafenib has been recommended as a first-line treatment for unresectable advanced BRAF V600E mutation-positive melanoma [40]. However, serine synthesis was significantly inhibited by vemurafenib in PC cell lines. Under serine depletion conditions, the death of PC cells treated with vemurafenib was significantly increased [41]. In a case report, after vemurafenib targeted therapy for a PC patient with BRAF V600E mutation, the primary tumor, liver metastases and lymphatic metastases of the patient were significantly reduced, and the skin toxicity was weak [42]. Unfortunately, no study has shown vemurafenib to prolong survival in patients with PC. This study provides new insights into improving the therapeutic effect of vemurafenib by selecting patients with vemurafenib-sensitive PC, thus providing new ideas for the precision treatment of PC.

The PPS scoring model constructed in our study specifically quantifies the expression pattern of IGF2BP2 and is significantly better than other methods in risk stratification and prediction of individualized treatment. However, this study still has several limitations. First, to obtain more reliable results with a larger sample size, we used a machine learning algorithm to estimate the IGF2BP2 expression pattern in a subset of patients. However, there may be slight differences between the estimated IGF2BP2 expression pattern and the actual IGF2BP2 expression pattern in these patients. Second, the results of drug target prediction and compound prediction cannot be verified against each other, which reduces the power of the conclusions in this part. Finally, all the conclusions of this study were based on public databases, and further experimental and clinical validation are needed to promote the clinical translation of the findings.

5. Conclusion

In conclusion, the current work developed a novel gene signature, PPS, for prognostic prediction of IGF2BP2-mediated m6A modification patterns in PC. The PPS model has important clinical significance in both low- and high-PPS patients. For patients with high PPS scores, based on multiple drug susceptibility and target databases, we have identified seven potential therapeutic targets and two compounds from multiple perspectives and all-around aspects, which opens a new direction for the application of precision medicine in PC. Overall, this study has not only provided new insights into personalized prognostication approaches but also thrown light on integrating tailored prognosis prediction with precision therapy.

Ethical statement

Not applicable.

Funding

This study is supported by the National Natural Science Foundation of China (No. 82000614; No. 81873589; No. 82203494), Natural Science Foundation of Hunan Province (No. 2020JJ5876; No. 2022JJ40741); Science and Technology Project of Changsha (No. kq2004146; No. kq2004143); Project of the Health Commission of Hunan Province (No. 202204015341; No. 202104011922); The Wisdom Accumulation and Talent Cultivation Project of The Third Xiangya Hospital of Central South University (No. YX202203) and The Shanxi Provincial Basic Research Program (Free Exploration) Youth Science Research Project. (No. 202303021212363 to L. Z.).

Data availability statement

All data used in this study are publicly available. The materials that support the conclusion of this article have been included within the Materials and Methods section “Data acquisition and preprocessing”.

CRedit authorship contribution statement

Dongjie Chen: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Longjun Zang:** Writing – review & editing, Writing – original draft, Methodology, Investigation. **Yanling Zhou:** Writing – review & editing, Writing – original draft. **Yongchao Yang:** Resources, Project administration. **Xianlin Zhang:** Resources. **Zheng Li:** Resources. **Yufeng Shu:** Supervision, Resources. **Wenzhe Gao:** Visualization, Validation, Supervision. **Hongwei Zhu:** Visualization, Validation, Supervision. **Xiao Yu:** Visualization, Validation, Supervision.

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.

Acknowledgments

We would like to exert compelling appreciation for the TCGA, ICGC and GEO projects.

Abbreviations

alternate acceptor AA

alternate donor AD

Aberrantly Differentiated Endocrine Exocrine ADEX

alternate promoter AP

alternative splicing AS

alternate terminator AT

Area Under the Curve AUC
 Cancer Cell Line Encyclopedia CCLC
 cancer cell lines CCLs
 cumulative distribution function CDF
 CRISPR-Cas9 Essentiality REvealing Score CERES
 Concordance index C-index
 connective Map CMap
 copy number variation CNV
 Cancer Therapeutics Response Portal CTRP
 differentially expressed genes DEGs
 Dependency Map DepMap
 exon skip ES
 fold change FC
 false discovery rate FDR
 transcript per million fragments mapped FPKM
 Gene Expression Omnibus GEO
 Gene Ontology GO
 Gene set enrichment analysis GSEA
 Gene set variation analysis GSVA
 Genotype-Tissue Expression GTEx
 International Cancer Genome Consortium ICGC
 Kyoto Encyclopedia of Genes and Genomes KEGG
 Kaplan-Meier K-M
 long noncoding RNAs lncRNAs
 N6-methyladenosine m6A
 minimal depth MD
 mutually exclusive ME
 messenger RNAs mRNAs
 microRNAs miRNAs
 overall survival OS
 Pancreatic cancer PC
 principal component analysis PCA
 Pathway Deregulation Score PDS
 Poor Prognosis Signature PPS
 profiling relative inhibition simultaneously in mixtures PRISM
 Random Forest RF
 retained intron RI
 ribosomal RNAs rRNAs
 single sample Gene Set Enrichment Analysis ssGSEA
 type 2 diabetes T2D
 The Cancer Genome Atlas Program TCGA
 tumor immune microenvironment TIME
 transformed into transcripts per million TPM
 transfer RNAs tRNAs

Appendix ASupplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e28243>.

References

- [1] R.L. Siegel, K.D. Miller, H.E. Fuchs, A. Jemal, Cancer statistics, 2022, *CA A Cancer J. Clin.* 72 (1) (2022) 7–33.
- [2] A. Vincent, J. Herman, R. Schulick, R.H. Hruban, M. Goggins, Pancreatic cancer, *Lancet* 378 (9791) (2011) 607–620.
- [3] C. Xia, X. Dong, H. Li, M. Cao, D. Sun, S. He, et al., Cancer statistics in China and United States, 2022: profiles, trends, and determinants, *Chin. Med. J.* 135 (5) (2022) 584–590.
- [4] P. Bailey, D.K. Chang, K. Nones, A.L. Johns, A.M. Patch, M.C. Gingras, et al., Genomic analyses identify molecular subtypes of pancreatic cancer, *Nature* 531 (7592) (2016) 47–52.
- [5] E.A. Collisson, P. Bailey, D.K. Chang, A.V. Biankin, Molecular subtypes of pancreatic cancer, *Nat. Rev. Gastroenterol. Hepatol.* 16 (4) (2019) 207–220.
- [6] M.A. Dawson, T. Kouzarides, Cancer epigenetics: from mechanism to therapy, *Cell* 150 (1) (2012) 12–27.
- [7] I.A. Roundtree, M.E. Evans, T. Pan, C. He, Dynamic RNA modifications in gene expression regulation, *Cell* 169 (7) (2017) 1187–1200.

- [8] R. Desrosiers, K. Friderici, F. Rottman, Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells, *Proc. Natl. Acad. Sci. U. S. A.* 71 (10) (1974) 3971–3975.
- [9] X. Jiang, B. Liu, Z. Nie, L. Duan, Q. Xiong, Z. Jin, et al., The role of m6A modification in the biological functions and diseases, *Signal Transduct. Targeted Ther.* 6 (1) (2021) 74.
- [10] X. Zhao, Y. Yang, B.F. Sun, Y. Shi, X. Yang, W. Xiao, et al., FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis, *Cell Res.* 24 (12) (2014) 1403–1419.
- [11] W. Xiao, S. Adhikari, U. Dahal, Y.S. Chen, Y.J. Hao, B.F. Sun, et al., Nuclear m(6)A reader YTHDC1 regulates mRNA splicing, *Mol. Cell* 61 (4) (2016) 507–519.
- [12] R.C. Spitale, R.A. Flynn, Q.C. Zhang, P. Crisalli, B. Lee, J.W. Jung, et al., Structural imprints in vivo decode RNA regulatory mechanisms, *Nature* 519 (7544) (2015) 486–490.
- [13] I.A. Roundtree, G.Z. Luo, Z. Zhang, X. Wang, T. Zhou, Y. Cui, et al., YTHDC1 mediates nuclear export of N(6)-methyladenosine methylated mRNAs, *Elife* 6 (2017).
- [14] X. Wang, B.S. Zhao, I.A. Roundtree, Z. Lu, D. Han, H. Ma, et al., N(6)-methyladenosine modulates messenger RNA translation efficiency, *Cell* 161 (6) (2015) 1388–1399.
- [15] C.R. Alarcón, H. Lee, H. Goodarzi, N. Halberg, S.F. Tavazoie, N6-methyladenosine marks primary microRNAs for processing, *Nature* 519 (7544) (2015) 482–485.
- [16] Y. Wang, Y. Wang, H. Patel, J. Chen, J. Wang, Z.S. Chen, et al., Epigenetic modification of m(6)A regulator proteins in cancer, *Mol. Cancer* 22 (1) (2023) 102.
- [17] Y. Liu, M. Shi, X. He, Y. Cao, P. Liu, F. Li, et al., LncRNA-PACERR induces pro-tumour macrophages via interacting with miR-671-3p and m6A-reader IGF2BP2 in pancreatic ductal adenocarcinoma, *J. Hematol. Oncol.* 15 (1) (2022) 52.
- [18] Y. Hou, Q. Zhang, W. Pang, L. Hou, Y. Liang, X. Han, et al., YTHDC1-mediated augmentation of miR-30d in repressing pancreatic tumorigenesis via attenuation of RUNX1-induced transcriptional activation of Warburg effect, *Cell Death Differ.* 28 (11) (2021) 3105–3124.
- [19] N. Qu, X. Bo, B. Li, L. Ma, F. Wang, Q. Zheng, et al., Role of N6-methyladenosine (m(6)A) methylation regulators in hepatocellular carcinoma, *Front. Oncol.* 11 (2021) 755206.
- [20] R.A. Moffitt, R. Marayati, E.L. Flate, K.E. Volmar, S.G. Loeza, K.A. Hoadley, et al., Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma, *Nat. Genet.* 47 (10) (2015) 1168–1178.
- [21] E.A. Collisson, A. Sadanandam, P. Olson, W.J. Gibb, M. Truitt, S. Gu, et al., Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy, *Nat. Med.* 17 (4) (2011) 500–503.
- [22] B. Zheng, J. Peng, A. Mollayup, A. Bakri, L. Guo, J. Zheng, et al., Construction of a prognostic prediction system for pancreatic ductal adenocarcinoma to investigate the key prognostic genes, *Mol. Med. Rep.* 17 (1) (2018) 216–224.
- [23] S. Zhou, Y. Yan, X. Chen, S. Zeng, J. Wei, X. Wang, et al., A two-gene-based prognostic signature for pancreatic cancer, *Aging* 12 (18) (2020) 18322–18342.
- [24] Q. Yuan, J. Ren, Z. Wang, L. Ji, D. Deng, D. Shang, Identification of the real Hub gene and construction of a novel prognostic signature for pancreatic adenocarcinoma based on the weighted gene Co-expression network analysis and least absolute shrinkage and selection operator algorithms, *Front. Genet.* 12 (2021) 692953.
- [25] B. Alagesan, G. Contino, A.R. Guimaraes, R.B. Corcoran, V. Deshpande, G.R. Wojtkiewicz, et al., Combined MEK and PI3K inhibition in a mouse model of pancreatic cancer, *Clin. Cancer Res.* 21 (2) (2015) 396–404.
- [26] X.Y. Chen, J. Zhang, J.S. Zhu, The role of m(6)A RNA methylation in human cancer, *Mol. Cancer* 18 (1) (2019) 103.
- [27] J. Wang, L. Chen, P. Qiang, The role of IGF2BP2, an m6A reader gene, in human metabolic diseases and cancers, *Cancer Cell Int.* 21 (1) (2021) 99.
- [28] N. Dai, The diverse functions of IMP2/IGF2BP2 in metabolism, *Trends Endocrinol. Metabol.* 31 (9) (2020) 670–679.
- [29] X. Hu, W.X. Peng, H. Zhou, J. Jiang, X. Zhou, D. Huang, et al., IGF2BP2 regulates DANCR by serving as an N6-methyladenosine reader, *Cell Death Differ.* 27 (6) (2020) 1782–1794.
- [30] C.J. Barger, C. Branick, L. Chee, A.R. Karpf, Pan-cancer analyses reveal genomic features of FOXM1 overexpression in cancer, *Cancers* 11 (2) (2019).
- [31] D. Xie, S. Yu, L. Li, M. Quan, Y. Gao, The FOXM1/ATX signaling contributes to pancreatic cancer development, *Am. J. Transl. Res.* 12 (8) (2020) 4478–4487.
- [32] W. Su, H.H. Han, Y. Wang, B. Zhang, B. Zhou, Y. Cheng, et al., The polycomb repressor complex 1 drives double-negative prostate cancer metastasis by coordinating stemness and immune suppression, *Cancer Cell* 36 (2) (2019), 139–55.e10.
- [33] H. Zhang, X. Zhang, X. Li, W.B. Meng, Z.T. Bai, S.Z. Rui, et al., Effect of CCNB1 silencing on cell cycle, senescence, and apoptosis through the p53 signaling pathway in pancreatic cancer, *J. Cell. Physiol.* 234 (1) (2018) 619–631.
- [34] Y. Lee, C.E. Lee, S. Oh, H. Kim, J. Lee, S.B. Kim, et al., Pharmacogenomic analysis reveals CCNA2 as a predictive biomarker of sensitivity to polo-like kinase 1 inhibitor in gastric cancer, *Cancers* 12 (6) (2020).
- [35] G. Baek, Y.F. Tse, Z. Hu, D. Cox, N. Buboltz, P. McCue, et al., MCT4 defines a glycolytic subtype of pancreatic cancer with poor prognosis and unique metabolic dependencies, *Cell Rep.* 9 (6) (2014) 2233–2249.
- [36] Y. Xie, S. Zhu, M. Zhong, M. Yang, X. Sun, J. Liu, et al., Inhibition of aurora kinase A induces necroptosis in pancreatic carcinoma, *Gastroenterology* 153 (5) (2017), 1429–43.e5.
- [37] E.K. Rowinsky, J.J. Windle, D.D. Von Hoff, Ras protein farnesyltransferase: a strategic target for anticancer therapeutic development, *J. Clin. Oncol.* 17 (11) (1999) 3631–3652.
- [38] E. Van Cutsem, H. van de Velde, P. Karasek, H. Oettle, W.L. Vervenne, A. Szawlowski, et al., Phase III trial of gemcitabine plus tipifarnib compared with gemcitabine plus placebo in advanced pancreatic cancer, *J. Clin. Oncol.* 22 (8) (2004) 1430–1438.
- [39] X.T. Xu, J. Chen, X. Ren, Y.R. Ma, X. Wang, Y.Y. Ma, et al., Effects of atorvastatin in combination with celecoxib and tipifarnib on proliferation and apoptosis in pancreatic cancer sphere-forming cells, *Eur. J. Pharmacol.* 893 (2021) 173840.
- [40] R. Gutzmer, D. Stroyakovskiy, H. Gogas, C. Robert, K. Lewis, S. Protsenko, et al., Atezolizumab, vemurafenib, and cobimetinib as first-line treatment for unresectable advanced BRAF(V600) mutation-positive melanoma (IMspire150): primary analysis of the randomised, double-blind, placebo-controlled, phase 3 trial, *Lancet* 395 (10240) (2020) 1835–1844.
- [41] K.C. Ross, A.J. Andrews, C.D. Marion, T.J. Yen, V. Bhattacharjee, Identification of the serine biosynthesis pathway as a critical component of BRAF inhibitor resistance of melanoma, pancreatic, and non-small cell lung cancer cells, *Mol. Cancer Therapeut.* 16 (8) (2017) 1596–1609.
- [42] A.K. Seghers, P.J. Cuyler, E. Van Cutsem, Molecular targeting of a BRAF mutation in pancreatic ductal adenocarcinoma: case report and literature review, *Targeted Oncol.* 15 (3) (2020) 407–410.