Exosome and lipid metabolism-related genes in pancreatic adenocarcinoma: a prognosis analysis

Jia Wu[1,*], Yajun Li[1,*], Ghulam Nabi[2], Xin Huang[3], Xu Zhang[1], Yuanzhen Wang[1], Liya Huang[1]

[1]Department of Gastroenterology, General Hospital of Ningxia Medical University, Yinchuan, Ningxia, China
[2]Institute of Nature Conservation, Polish Academy of Sciences, Krakow, Poland
[3]Department of Gastroenterology, Traditional Chinese Medicine Hospital of Yinchuan, Yinchuan, Ningxia, China
*Share first authorship

Correspondence to: Liya Huang; **email:** txmbw@126.com, https://orcid.org/0000-0002-7541-1248

ABSTRACT

Objective: The purpose of the study was to investigate the role of exosome and lipid metabolism-related genes (EALMRGs) mRNA levels in the diagnosis and prognosis of Pancreatic Adenocarcinoma (PAAD).
Methods: The mRNA expression pattern of PAAD and pan-cancers with prognostic data were obtained from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) database. EALMRGs were acquired from GeneCards and MSigDB database after merging and deduplication. Prognostic EALMRGs were screened through univariate COX regression analysis, and a prognostic model was constructed based on these genes by least absolute shrinkage and selection operator (LASSO) regression. The prognostic value of EALMRGs was then validated in pan-cancer data. The time characteristics ROC curve analysis was performed to evaluate the effectiveness of the prognostic genes.
Results: We identified 5 hub genes (ABCB1, CAP1, EGFR, PPARG, SNCA) according to high and low-risk groups of prognoses. The risk formula was verified in three other cohort of pancreatic cancer patients and was explored in pan-cancer data. Additionally, T cell and dendritic cell infiltration was significantly increased in low-risk group. The expression of the 5 hub genes was also identified in single-cell sequencing data of pancreatic cancer with pivotal pathways. Additionally, functional enrichment analysis based on pancreatic cancer data in pancreatic cancer showed that protein serine/threonine kinase activity, focal adhesion, actin binding, cell-substrate junction, organic acid transport, and regulation of transporter activity were significant related to the expression of genes in EALMRGs.
Conclusions: Our risk formula shows potential prognostic value in multiple cancers and manifest pivotal alterations in immune infiltration and biological pathway in pancreatic cancer.

INTRODUCTION

Pancreatic adenocarcinoma (PAAD) is a serious kind of cancer with high mortality and increasing morbidity worldwide [1]. Likewise, the incidence and mortality rate of PAAD remains on the rise in China but relatively stable in Japan and South Korea [2]. It is not easy to detect PAAD at an early stage until late in progression, thus losing effective treatment opportunities, which is an important reason for the high mortality in pancreatic cancer. Although carbohydrate antigen 19-9 (CA19-9) and carcinoembryonic antigen (CEA) are commonly used in clinical practice for tumor markers, they have limited specificity and sensitivity to screen patients with early pancreatic ductal adenocarcinoma (PDAC) [3]. Therefore, it is essential to identify effective prognostic biomarkers and to establish a valid prognostic model in PAAD.

Exosomes are nanosized (30-150nm), physiologically released extracellular vesicles of endosomal origin, and carry substances such as nucleic acids, proteins, and lipids [4, 5]. Cancer derived exosomes are involved in multiple biological processes, including epithelial-to mesenchymal transition, cell proliferation, and angiogenesis [6–9]. Furthermore, there is a considerable amount of literature on deregulation of lipid metabolism in cancer cells. Considering the increasing evidence that lipid metabolism disorder plays a key role in PAAD patients, it is necessary to comprehensively evaluate the prognostic significance of exosome and lipid metabolism-related genes, which might provide potential prognostic biomarkers. However, studies on this problem have received scant attention in research literature.

This study acquired the transcriptome profiling data of PAAD with clinical information from the TCGA database and GEO. A total of 52 exosome and lipid metabolism-related genes (EALMRGs) were obtained from the GeneCards and MSigDB databases after merging and deduplication. We constructed a prognostic model and implemented external and internal validation to evaluate the effectiveness and availability of the model. Finally, 5 significant prognoses related EALMRGs were identified and verified in GEO. To further explore the potential mechanism and relationship of these genes, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and protein-protein interaction (PPI) analysis were conducted. We also performed Gene Set Enrichment Analysis (GSEA) and Gene Set Variation Analysis (GSVA) to understand the functional enrichment and differences between high and low-risk groups based on the prognostic model. We constructed a potential prognostic model and identified 5 significant prognosis EALMRGs and explored the potential mechanism. In this way, our study reveals the value of EALMRGs in predicting the prognosis of PAAD patients and provides clues for the specific mechanisms associated with lipid metabolism and cancer.

MATERIALS AND METHODS

Data collection and pre-processing

We downloaded high-throughput sequencing RNA data and clinical information on PAAD from the TCGA database (https://portal.gdc.cancer.gov/). Excluding 4 samples of para-carcinoma tissues, a total of 179 PAAD patients were enrolled. RNA sequencing data were transformed from fragments per kilobase per million (FPKM) formats to transcripts per million (TPM) reads for this study. Corresponding clinical information was downloaded from the UCSC Xena browser (http://genome.ucsc.edu) [10]. The count

sequencing and clinical data in the TCGA-PAAD dataset were standardized using R packet limma [11]. In addition, two raw gene expression datasets; GSE62452 [12] and GSE57495 [13] downloaded from the GEO database [14] were used to validate further analysis using GEOquery.

Human genes with comprehensive information were provided by the GeneCards database [15]. Based on the term "Exosome" as a keyword and "Protein Coding", "Relevance score>2" as a screening criterion, we obtained 589 exosome-related genes (ERGs) from the GeneCards database. Furthermore, we also collected 77 ERGs on the Molecular Signatures Database (MSigDB database) [16] website using "Exosome" as the keyword. A total of 628 ERGs were obtained after merging and removing duplicates. Likewise, we used "lipid metabolism" as the keyword on the GeneCards database, and 680 lipid metabolism-related genes (LMRGs) were collected. Meanwhile, 102 LMRGs were obtained on the MSigDB database using keywords "lipid metabolism". A total of 747 LMRGs were collected after merging and removing duplicates. Finally, we obtained 52 EALMRGs by intersecting 628 ERGs and 747 LMRGs (Supplementary Table 1).

Construction and evaluation of EALMRGs prognostic models

We selected TCGA-PAAD tumor samples (Table 1) and performed univariate COX regression analysis on overall survival (OS) related EALMRGs by survival packet. A total of 10 prognosis-related EALMRGs were acquired with P<0.1. Based on these 10 EALMRGs, the software package "glmnet" [17] was used to perform the Least absolute shrinkage and selection operator (LASSO) [18] logistic regression analysis with tenfold cross-validation. LASSO regression is a common machine learning technique for constructing prognostic diagnostic models. It is used to solve the problem of overfitting by adding a penalty term (lambda × absolute value of slope), using regularization based on linear regression, and additionally, improve the model's generalization ability.

$$riskScore = \sum_i Coefficient(Hub\,gene_i) *$$

$$mRNA\,Expression(Hub\,gene_i)$$

Subsequently, prognostic-related genes were identified as key genes. We drew KM curves and AUC to evaluate the prognosis value of these key genes. A risk factor graph was drawn to show the groups according to each sample's risk score and survival outcomes in this prognostic model. We also used calibration analysis to evaluate the performance of the prognostic model based on EALMRGs using the rms R package. Decision curve

Table 1. Baseline characteristics in TCGA-PAAD.

Characteristics	Overall
Pathologic T stage, n (%)	
T1	7 (4)
T2	24 (13.6)
T3	143 (80.8)
T4	3 (1.7)
Pathologic N stage, n (%)	
N0	50 (28.7)
N1	124 (71.3)
Pathologic M stage, n (%)	
M0	80 (94.1)
M1	5 (5.9)
Gender, n (%)	
Female	80 (44.7)
Male	99 (55.3)
Age, median (IQR)	65 (57, 73)
OS event, n (%)	
Alive	86 (48)
Dead	93 (52)
DSS event, n (%)	
Yes	73 (42.2)
No	100 (57.8)
PFI event, n (%)	
Yes	105 (58.7)
No	74 (41.3)

analysis (DCA) is a simple method for evaluating clinical predictive models, diagnostic trials, and molecular markers. We also drew DCA maps to evaluate the predictive effect of the model on the 1-, 3-, and 5-year survival state of PAAD patients using the ggDCA R package.

Identification of DEGs based on EALMRGs prognostic models

For data analysis, the dataset TCGA-PAAD was corrected and standardized by limma R packet. We calculated the risk score of the TCGA-PAAD sample based on the prognostic model and then divided the samples into high and low-risk groups according to the risk score. Differentially expressed genes (DEGs) between two groups in the TCGA-PAAD dataset were obtained and presented by plotting volcanic maps using ggplot2.package. By selecting DEGs with |logFC| >0 and P.adj<0.05, we drew heat maps and group comparison maps to display the key gene expression in the TCGA-PAAD dataset using pheatmap R packet. In addition, we standardized GSE62452 and GSE57495 from the GEO dataset and selected samples of pancreatic cancer patients to batch merge them into a Dataset-PAAD using the sva package. According to the risk score, we divided Dataset-PAAD into high and low-risk groups and drew a group comparison map. Genes with consistent trends of TCGA-PAAD and Dataset-PAAD were selected as hub genes for subsequent analysis.

Bulk data download and pre-processing

The clinical phenotype data of PAAD was downloaded from the TCGA database. The samples lacking survival time and survival state were removed, and all those with survival time greater than 0 days were saved. Meanwhile, TCGA expression profile data (log2 (TPM)) was downloaded. Finally, 176 tumor samples were obtained.

The clinical phenotype data of three external validation datasets: PACA_AU, PACA_CA, and PAEN-AU were downloaded from the ICGC database. The samples lacking survival time and survival state were removed, and all those with survival time greater than 0 days were saved. Meanwhile, the corresponding expression profile data (log2 (TPM)) was downloaded.

GDC Pan Cancer (PANCAN) 33 pan-cancer sequencing data and corresponding survival information were downloaded from UCSC Xena to analyze the pan-cancer section.

scRNA data download

Single-cell sequencing data in the GEO database, GSE155698, was downloaded. It contained 16 pancreatic cancer tissue samples and 3 adjacent normal pancreas samples. The procedure of single-cell sequencing is exerted using the Seurat package in R software.

Functional similarity analysis

In this study, GO and KEGG enrichment was performed by the clusterProfiler [19–22] R package with a screening criterion of P.adj < 0.1 and a P-value correction method of Benjamin Hochberg (BH). Gene Set Enrichment Analysis (GSEA) and GSVA [23–25] were performed to determine differentially enriched signaling pathways between high and low-risk groups. A function or pathway term with adjusted P.adj < 0.05, and false discovery rate (FDR) < 0.25 was considered statistically significant. In our study, the PPI network was constructed using the STRING database [26] to obtain the interactions among hub genes in EALMRGs, and the least required interaction score was 0.4. Furthermore, we used the GeneMANIA website [27] to predict the functionally similar genes of the screened hub genes and constructed an interaction network. We used the miRDB database [28] to predict miRNAs that interacted with Hub genes and then plotted the mRNA-miRNA interaction network using data with Target Score>90. Besides, we searched for TF that binds to hub genes through the CHIPBase and hTFtarget database, and the common parts were retained in both databases [29, 30].

Cell culture

Pancreatic carcinoma cell lines of Homo sapiens, including PANC-1, SW1990, BXPC-3, CFPAC1, ASPC1, and normal pancreatic cell line HPDE6-C7 were obtained as a gift from Dr. Deyu Zhang in Changhai Hospital. Cells were cultured in DMEM

with 10% fetal bovine serum (Gibco, USA), 100 µg/mL streptomycin, and 100 U/mL penicillin (Invitrogen, USA) at 37° C under 5% CO2 and 1% O^2. All the following experiments were independently repeated three times.

RNA extraction and qRT-PCR of the hub genes

Trizol reagent (Invitrogen, Carlsbad, CA, U.S.A.) was used to extract total RNA from the cells according to the manufacturer's protocol. Superscript II reverse transcriptase and random primers were used to synthesize cDNA. Quantitative real-time PCR (qRT-PCR) was conducted on the ABI 7900HT Sequence Detection System with SYBR-Green dye (Applied Biosystems, Foster City, CA, U.S.A.). All primers were as following: GAPDH Forward: 5′-GGACCTGACCTGCCGTCTAG-3′, Reverse: 5′-GTAGCCCAGGATGCCCTTGA-3′, ABCB1: Forward: 5′-CCCATCATTGCAATAGCAGG-3′and Reverse: 5′-GTTCAAACTTCTGCTCCTGA-3′, CAP1: Forward: 5′-GAAGGTGAAGGT CGGAGTC-3′and Reverse: 5′-CCCGAATCACATTCTCCAAGA A-3′, EGFR: Forward: 5′-GGACGACGTGGTGGATGC-3′and Reverse: 5′-GGCGCCTGTGGGGTCTGAGC-3′, PPARG: Forward: 5′- ACCAAAGTGCAATCAAAGT GGA-3′and Reverse: 5′- ATGAGGGAGTTGGAAGGC TCT-3′, SNCA: Forward: 5′- AAGAGGGTGTTCTCTA TGTAGGC-3′ and Reverse: 5′- GCTCCTCCAACATT TGTCACTT-3′. The reaction parameters included a denaturation program (10 min at 95° C), followed by an amplification and quantification program over 45 cycles (15 s at 95° C and 34 s at 60° C). Each sample was tested in triplicates, and each sample underwent a melting curve analysis to check for the specificity of amplification. The expression level was determined as a ratio between the hub genes and the internal control GAPDH in the same mRNA sample and calculated by the comparative CT method. Expression levels of hub genes were calculated by the $2^{-\delta\delta}Ct$ method.

Statistical analysis

All data analyses in this study were performed and visualized in R software (Version 4.1.2) or GraphPad Prism (Version 8). The Wilcoxon rank sum test was used to compare the two groups of continuous variables, and the statistical significance of normal distribution variables was estimated through an independent Student t-test. LASSO regression analysis was based on the R package glmnet and ROC [31] using the R package pROC. If not particularly specified, the results were calculated using Spearman correlation analysis to calculate the correlation coefficients between different molecules. All P-statistics were bilateral, with a P-value<0.05 as a statistically significant criterion.

RESULTS

Merge and correction of dataset in GEO

The gene expression matrix of GSE62452 and GSE57495 datasets from GEO was firstly background corrected and data normalized. We obtained Dataset-PAAD using the sva R package to remove batch effects of the combined two datasets. We also compared the datasets before and after the removal of batch effects through distribution Boxplot (Figure 1A, 1B) and Principal Component Analysis (PCA) (Figure 1C, 1D).

The results indicated that the batch effect of Dataset-PAAD samples was eliminated after processing.

Construction of EALMRGs prognostic models

We obtained 52 EALMRGs by intersecting 628 ERGs and 747 LMRGs (Figure 2A). A total of 10 prognosis-related EALMRGs were obtained using univariate Cox regression analysis with P < 0.1 in EALMRGs from TCGA-PAAD, including *ABCB1*, *CAP1*, *EGFR*, *ITGB1*, *MAPK1*, *MTREX*, *PPARG*, *RAB7A*, *SNCA*, *VDAC1* (Table 2), and forest plot was drawn to show
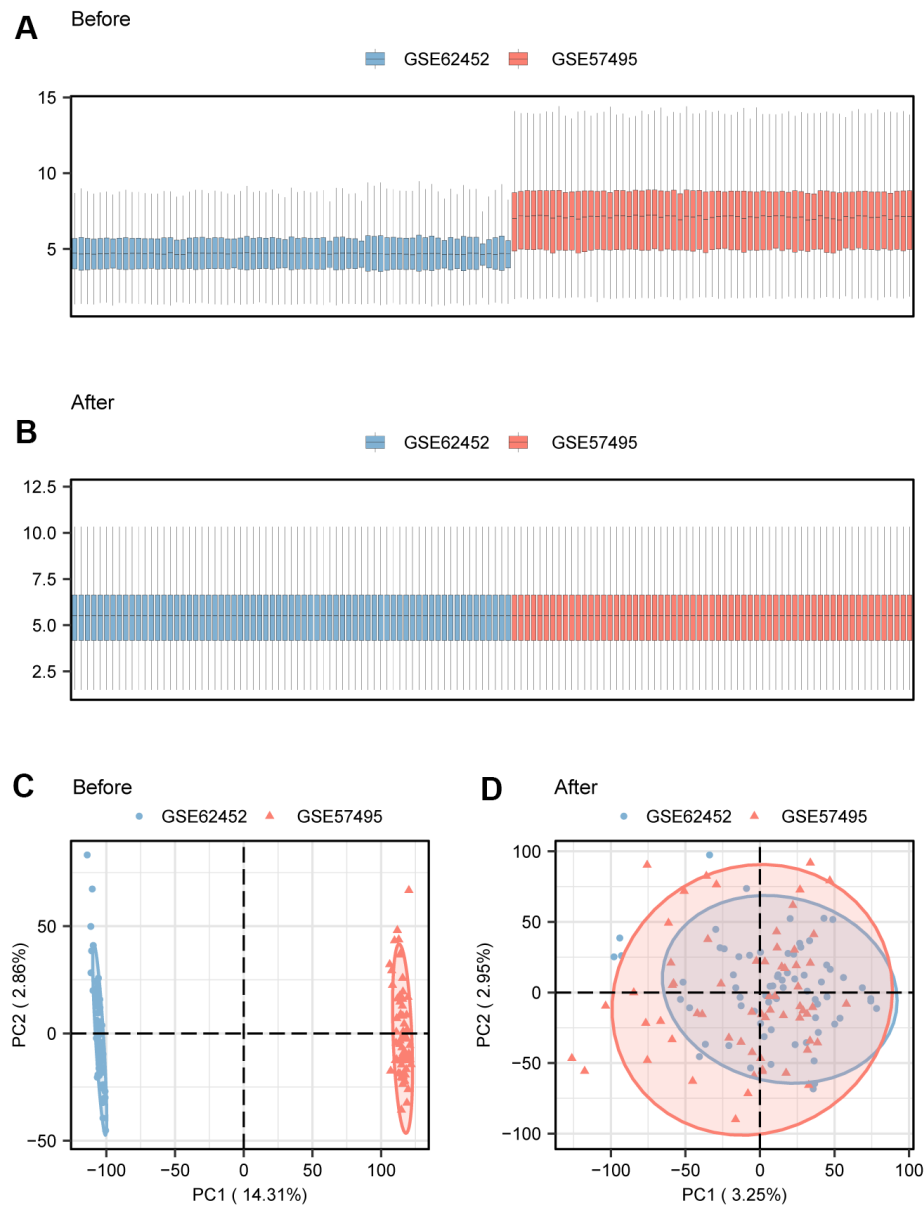

Figure 1. Merge and correction of dataset. (**A**) Boxplot to show merged dataset before batch processing. (**B**) Boxplot to show merged dataset after batch processing. (**C**) PCA to show merged dataset before batch processing. (**D**) PCA to show merged dataset after batch processing. PCA, Principal Component Analysis.
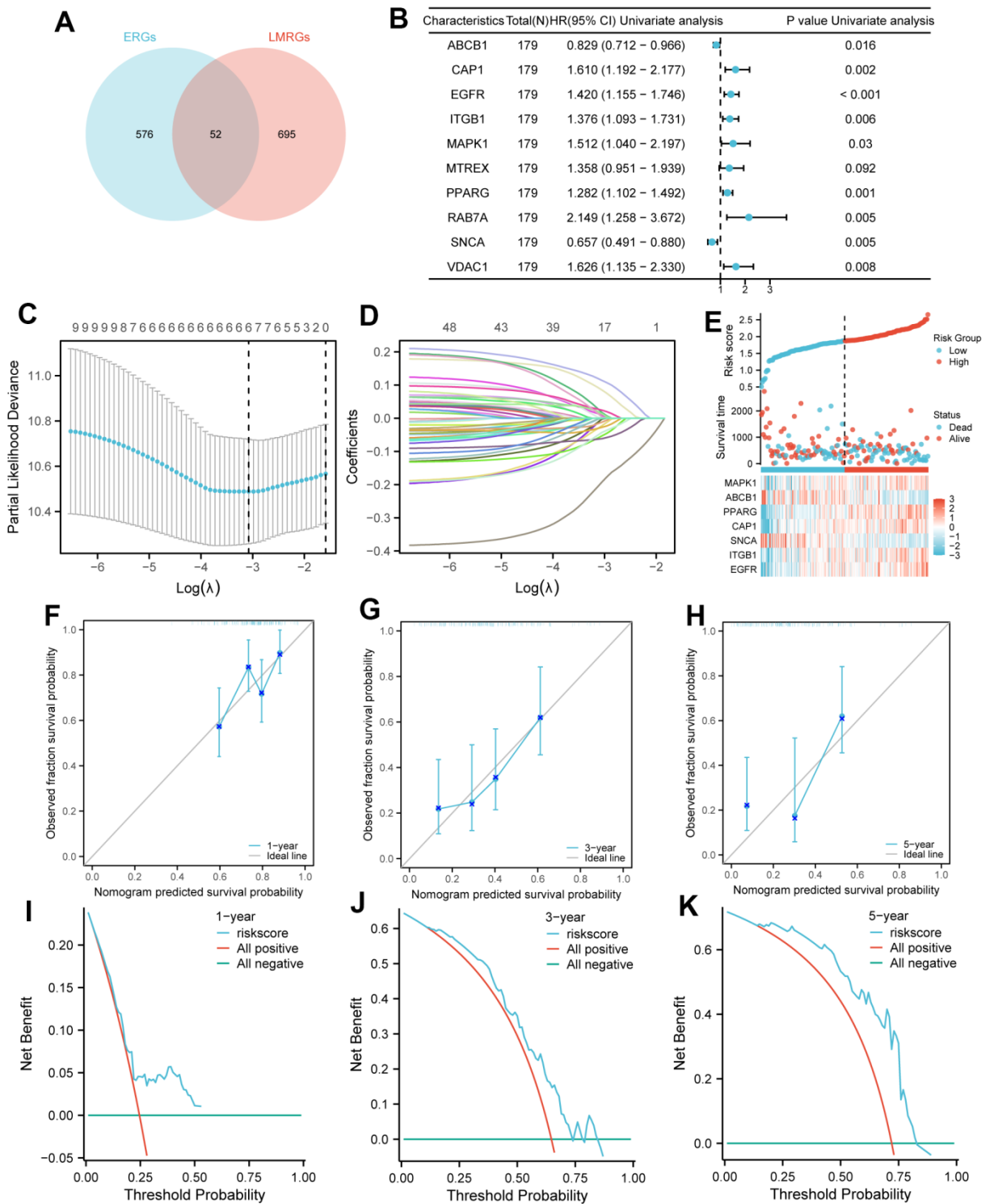
Figure 2. Model establishment for prognosis. (**A**) Venn diagram: the blue circle on the left includes the 576 ERGs, the red circle on the right includes the 695 LMRGs, and the intersection of the two circles includes the 52 EALMRGs. (**B**) Forest plot to show the result of univariate Cox regression analysis. (**C**) The confidence interval under each lambda. (**D**) The changing trajectory of each independent variable. (**E**) Risk score nomogram consists of 3 parts, including risk group, survival outcomes, and heatmap. (**F–H**) The calibration curves of the nomogram are at 1-year (**F**), 3-year (**G**), and 5-year (**H**), respectively. The X-axis in curves represented nomogram predicted survival probability and Y-axis represented observed fraction survival probability. (**I–K**) DCA diagrams of the models for 1-year (**I**), 3-year (**J**), and 5-year (**K**), respectively. The X-axis in DCA diagrams represents Threshold Probability and the Y-axis represents Net Benefit. LASSO, Least absolute shrinkage and selection operator; TCGA, The cancer genome atlas; OS, Overall survival; DCA, decision curve analysis; LASSO, least absolute shrinkage and selection operator.

Table 2. Univariate and multivariate Cox regression.

Characteristics	Total (N)	Univariate analysis		Multivariate analysis	
		HR (95% CI)	P-value	HR (95% CI)	P-value
ABCB1	179	0.829 (0.712 - 0.966)	0.016	0.783 (0.632 - 0.970)	0.025
CAP1	179	1.610 (1.192 - 2.177)	0.002	0.932 (0.543 - 1.602)	0.800
EGFR	179	1.420 (1.155 - 1.746)	< 0.001	1.161 (0.859 - 1.568)	0.331
ITGB1	179	1.376 (1.093 - 1.731)	0.006	1.054 (0.655 - 1.697)	0.829
MAPK1	179	1.512 (1.040 - 2.197)	0.030	1.925 (0.924 - 4.008)	0.080
MTREX	179	1.358 (0.951 - 1.939)	0.092	0.998 (0.581 - 1.716)	0.995
PPARG	179	1.282 (1.102 - 1.492)	0.001	1.052 (0.864 - 1.281)	0.614
RAB7A	179	2.149 (1.258 - 3.672)	0.005	0.984 (0.428 - 2.260)	0.969
SNCA	179	0.657 (0.491 - 0.880)	0.005	0.695 (0.489 - 0.990)	0.044

the results (Figure 2B). We used LASSO regression analysis to construct a LASSO prognostic model based on the expression of these 10 genes. The optimal lambda value for the evaluation index corresponds to 7 genes with non-zero coefficients (Figure 2C), and a LASSO prognostic variable trajectory map was drawn (Figure 2D). These 7 genes were identified as key genes, including *ABCB1*, *CAP1*, *EGFR*, *ITGB1*, *MAPK1*, *PPARG*, and *SNCA*. We also visualized the high and low expression in the LASSO prognostic model with risk score, consisting of 3 parts, including risk group, survival outcomes, and heatmap (Figure 2E).

Evaluation of prognostic EALMRGs genes

In our study, we conducted a 1-, 3- and 5-year prognostic calibration analysis based on the LASSO model and plotted a calibration curve. The results revealed that the 1- and 3-year prediction effect was better than that of the 5-year (Figure 2F–2H). Meanwhile, we used DCA diagrams to evaluate the clinical utility of the LASSO prognostic model at 1-, 3- and 5-year, and the results presented that the predictive value of the model at 5-year was better than that at 1- and 3-year (Figure 2I–2K).

Furthermore, according to these 7 key genes (*ABCB1*, *CAP1*, *EGFR*, *ITGB1*, *MAPK1*, *PPARG*, *SNCA*), we plotted the KM curves to show the survival state of high and low expression groups in tumor samples from the TCGA-PAAD dataset. The result of KM curves suggested that the survival rate of PAAD patients in the high-expression group of genes *ABCB1* and *SNCA* was higher than that in the low-expression group over time, while the survival rate of PAAD patients in the high-expression group of *CAP1*, *EGFR*, *ITGB1*, *MAPK1*, *PPARG* was lower than that in the low-expression group over time (Figure 3A–3G). We also plotted time ROC curves of high and low gene expression groups

(Figure 3H–3N) and the trend of AUC of 7 key genes and prognostic models over time (Figure 4A–4G). All 7 genes performed well in predicting the prognosis of PAAD patients at 5-year, with EGFR and ITGB1 having better overall predictive effects than other genes with certain accuracy in LASSO models.

Validation of risk models using external data sets

We used these 5 key genes to construct a risk model on the TCGA-PAAD dataset, calculated the risk score for each patient, and divided into high and low-risk groups by the median. According to the KM curve, it could be seen that the survival of patients in the high-risk group was poor, and the risk model had good value in predicting the prognosis survival time (Figure 5A). To verify the generalization of the model, we evaluated the model on three external independent datasets. Through the prediction of the model, there were significant differences in survival between high and low-risk groups, and it has good predictive performance for patient prognosis and survival time (Figure 5B–5D). Moreover, the mRNA level of ABCB1 and SNCA were identified with significant downregulation in pancreatic cancer cell lines compared to normal pancreatic cells. Additionally, the mRNA level of CAP1, EGFR and PPARG were identified with significant downregulation in pancreatic cancer cell lines compared to normal pancreatic cells (Supplementary Figure 1A).

Immune microenvironment analysis in high and low-risk groups

To further analyze the differences in the immune microenvironment in patients with high and low-risk, we compared the immune scores predicted by ESTIMATE. We found lower levels of immune infiltration in high-risk patients (Figure 6A). We then used the MCP-counter algorithm to examine the
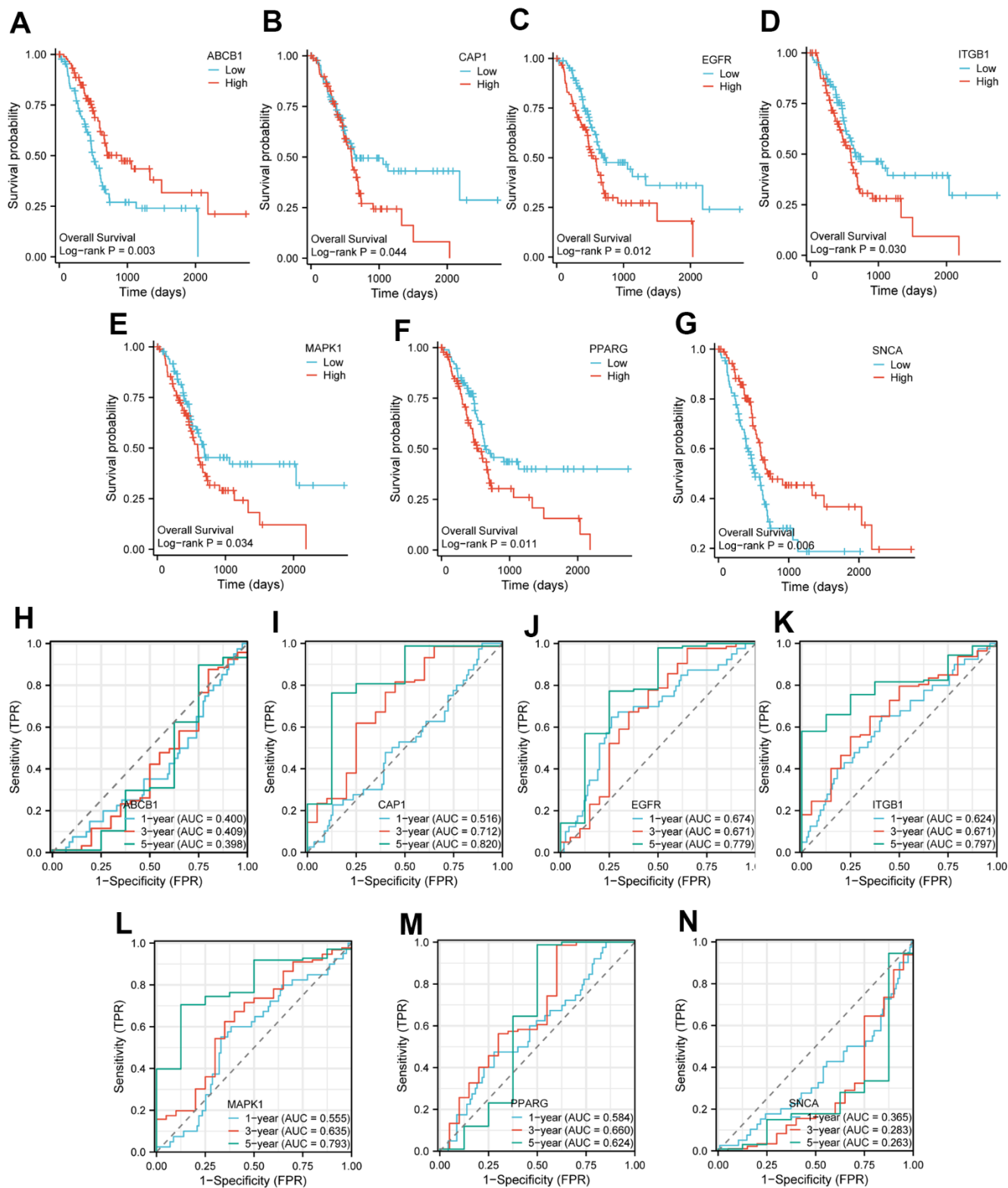
Figure 3. KM curve and time ROC of key genes in TCGA-PAAD. (**A**–**G**) The KM curve of the high and low expression groups of Key genes in TCGA-PAAD. Patients with high expression of *ABCB1* (**A**) and *SNCA* (**G**) had significantly longer overall survival; patients with high expression of *CAP1* (**B**), *EGFR* (**C**), *ITGB1* (**D**), *MAPK1* (**E**) and *PPARG* (**F**) had significantly shorter overall survival. (**H**–**N**) The time ROC curve of the high and low expression groups of Key genes in TCGA-PAAD. ROC curve showed the efficiency of 7 key gene expression levels to predict the prognosis over time. The X-axis represents a false positive rate, and the Y-axis represents a true positive rate. OS, Overall survival; KM, Kaplan–Meier; ROC, receiver operating characteristic curve.

approximate infiltration of cell types, and the results showed that multiple immune cell types had less infiltration in high-risk patients (Figure 6B). The ssgsea algorithm was then used to analyze the relationship between the high-low risk group and 28 kinds of immune cell scores, and we found that the low-risk group had higher immune cell scores (Figure 6C). The infiltration of immune cells is directly related to patient survival, and some studies have shown a certain correlation between

chemokines and immune cell infiltration. Therefore, we analyzed the correlation between risk scores and chemokines and their receptors (Figure 6D).

The function of immune cells is affected by immune checkpoints. We then analyzed the correlation between risk score and immune checkpoints (Figure 6E) and found a positive correlation with most immune checkpoints.
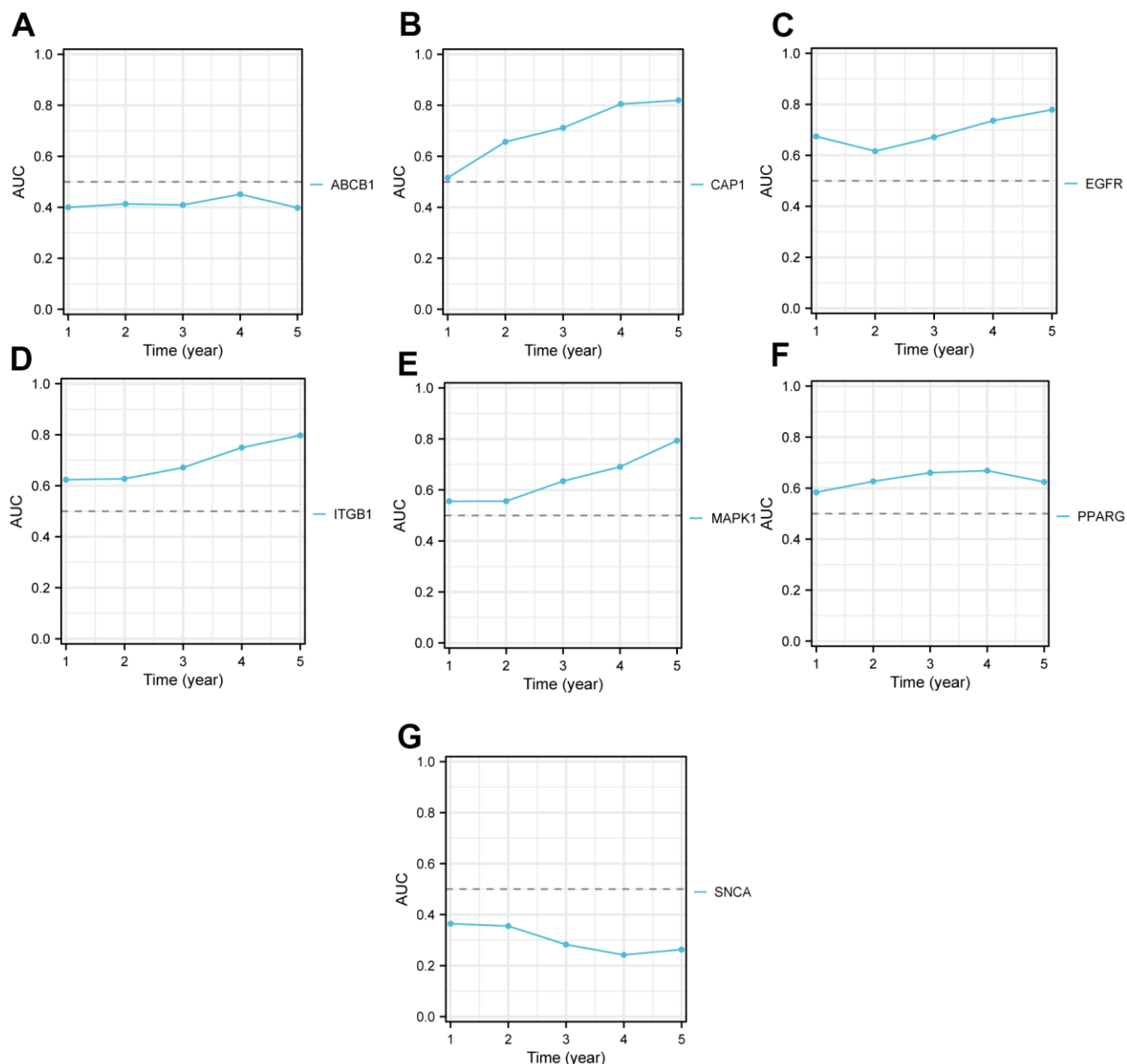
Figure 4. The time AUC curve of the high and low expression groups of Key genes in TCGA-PAAD. The AUC of *ABCB1* (**A**) is less than 0.5 over time, which indicates they were protective factors, and the closer the AUC value is to 0, the better the prediction performance. The AUC of *CAP1* (**B**), *EGFR* (**C**), *ITGB1* (**D**), *MAPK1* (**E**), and *PPARG* (**F**) are all over than 0.5 over time, which indicates they were risk factors, and the closer the AUC value is to 1, the better the prediction performance. The AUC of *SNCA* (**G**) is also less than 0.5 over time. The X-axis represents time, and Y-axis represents the AUC value. TCGA, The cancer genome atlas; AUC, Area Under Curve.

Subsequently, we analyzed the potential clinical effects of immunotherapy evaluated by using TIDE (http://tide.dfci.harvard.edu/) software in our defined high-risk groups. The higher the TIDE prediction score, the higher the likelihood of immune escape, indicating a lower likelihood of patients benefiting from immuno-therapy. It was found that the higher the risk score, the higher the TIDE score (Figure 6F).

More profoundly, to examine the differences in immune stages, we used the TIP analysis tumor immune phenotype (TIP) database (http://biocc.hrbmu.edu.cn/TIP), a web-based tool that can evaluate the immune microenvironment based on the cancer immune cycle. We calculated the scores for each of the seven steps of the immune cycle in each sample (one step for each color from left to right) and found a higher immune step score (Figure 6G) in the low-risk group.

Analysis of prognostic value of risk model in pan-cancer

We evaluated the model on 33 pan-cancer data, and through univariate cox analysis, we found that it was significant in 5 datasets other than PAAD, and as a risk prognostic factor in 4 datasets (Figure 7A).

Simultaneously, a favorable C-index (Figure 7B) was shown in multiple datasets. Subsequently, K-M plot was plotted for these four datasets, and patients in the high-risk group also have poorer prognoses in multiple cancers (Figure 7C). Moreover, we used the GSVA package to perform a pathway enrichment score on the pan-cancer dataset for the pathways in KEGG. Subsequently, correlation analysis was conducted based on their respective risk scores to screen for pathways (P-value < 0.05), and then pathways that were retained for more than 10 cancer species were screened and shown in Supplementary Figure 1B.

Discovery of DEGs based on EALMRGs prognostic models

We calculated the risk scores of the dataset samples from TCGA-PAAD based on the LASSO regression results. According to the risk scores, samples from TCGA-PAAD and Dataset-PAAD were divided into low and high-risk groups. Subsequently, we performed differential analysis on the processed TCGA-PAAD dataset. A total of 13270 DEGs were identified with |logFC|>0 and P.adj < 0.05 and included 8016 upregulated and 5254 downregulated genes. The DEGs were visualized by the volcano plot (Figure 8A).
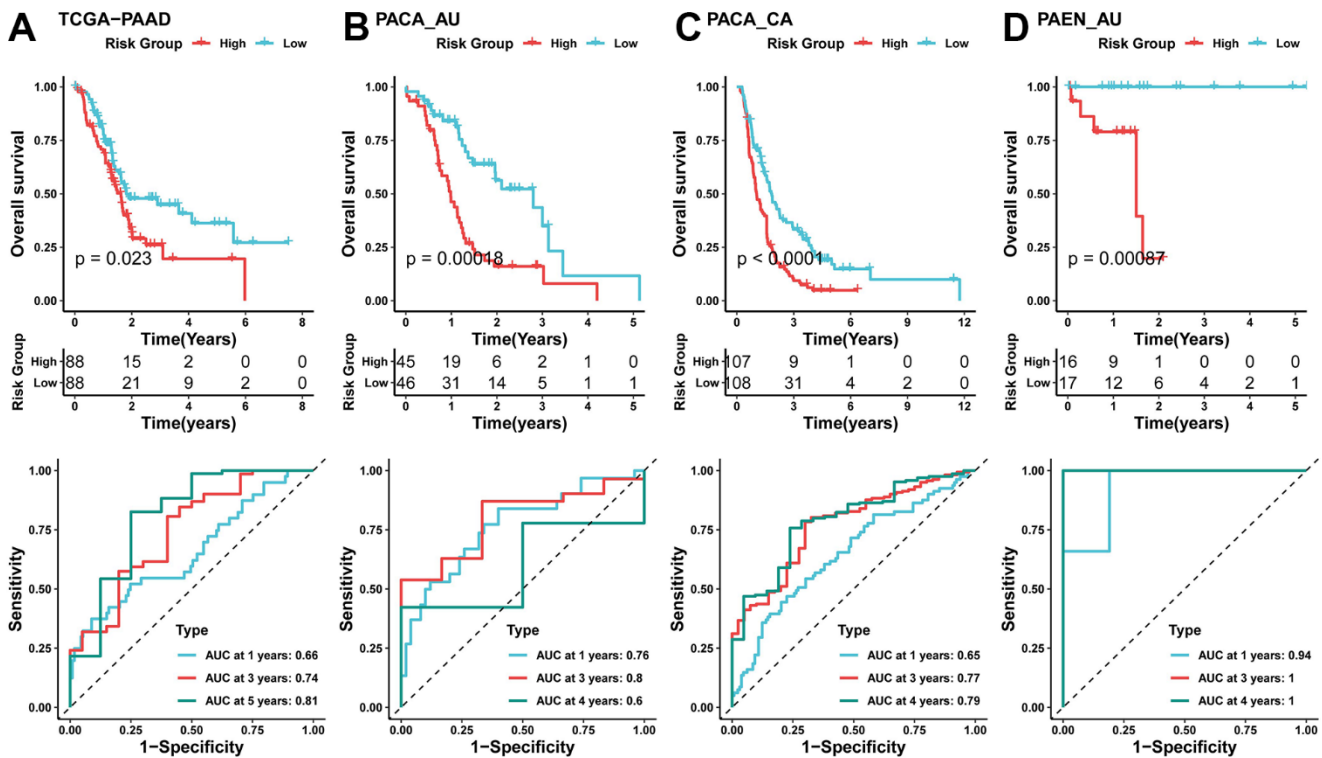

Figure 5. Validation of risk models using external data sets. (**A**) KM curve (top) of TCGA-PAAD dataset, Time-dependent area under the receiver operating characteristic curve (AUC) at 1-, 3-, and 5-year in the A TCGA-PAAD (bottom); (**B–D**) KM curve (top), Time-dependent area under the receiver operating characteristic curve (AUC) at 1-, 3-, and 4-year in the B PACA_AU, C PACA_CA, D PAEN_AU (bottom).
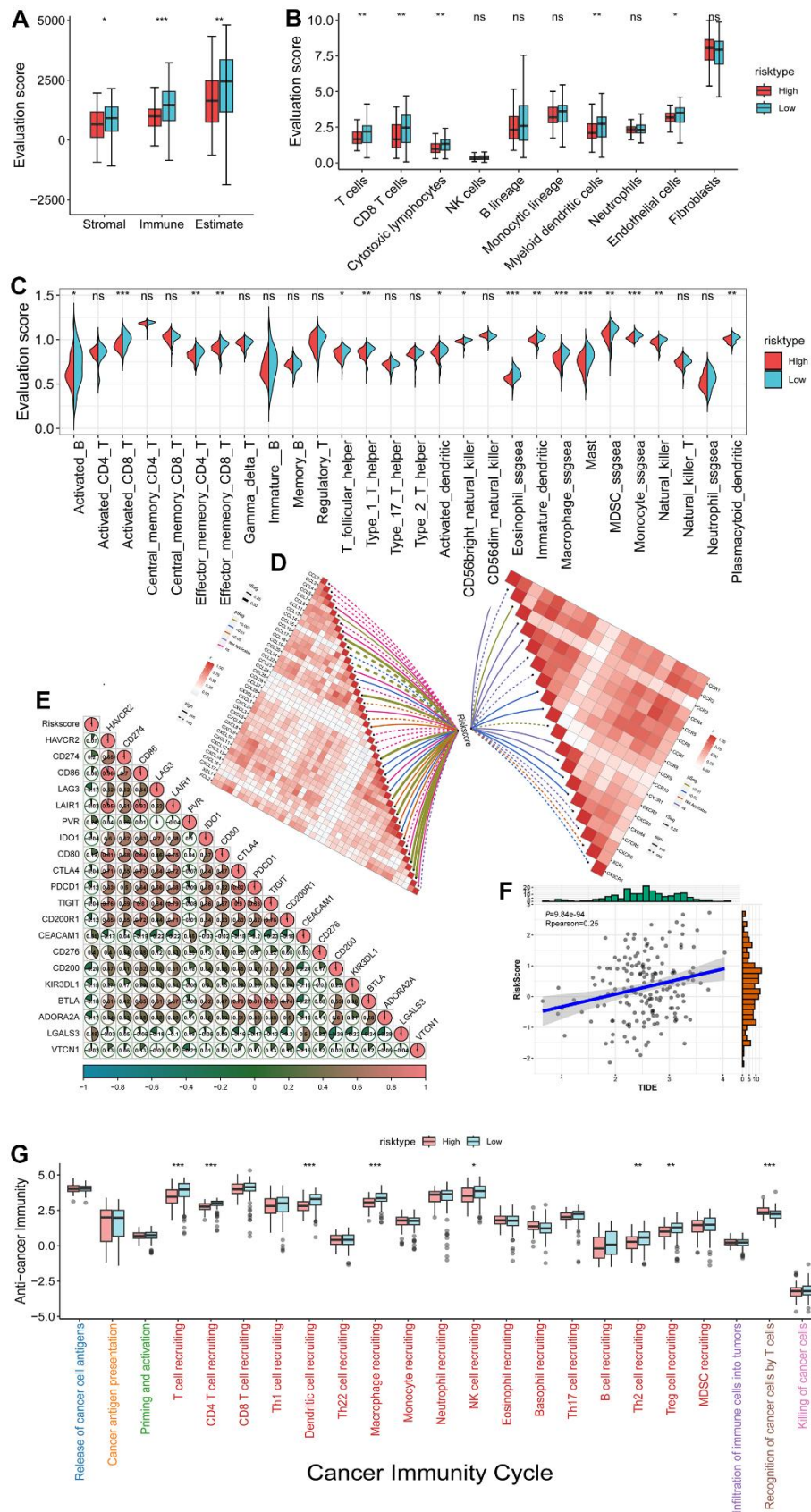
Figure 6. Immune microenvironment analysis in high and low risk groups. (**A**) Comparison of the Estimate score in high and low-risk groups; (**B**) MCP-counter immune cell infiltration score was compared between high and low-risk groups; (**C**) The ssgsea calculated 28

kinds of immune cell infiltration scores in high and low-risk groups. (**D**) Correlation analysis between risk score and chemokines and receptors; (**E**) Correlation analysis between risk score and 20 kinds of immune checkpoints; (**F**) Correlation analysis between risk score and TIDE score (immune escape); (**G**) TIP calculated comparison of tumor immune cycle step scores in high-low risk groups. *<0.05 **<0.01 ***<0.001.
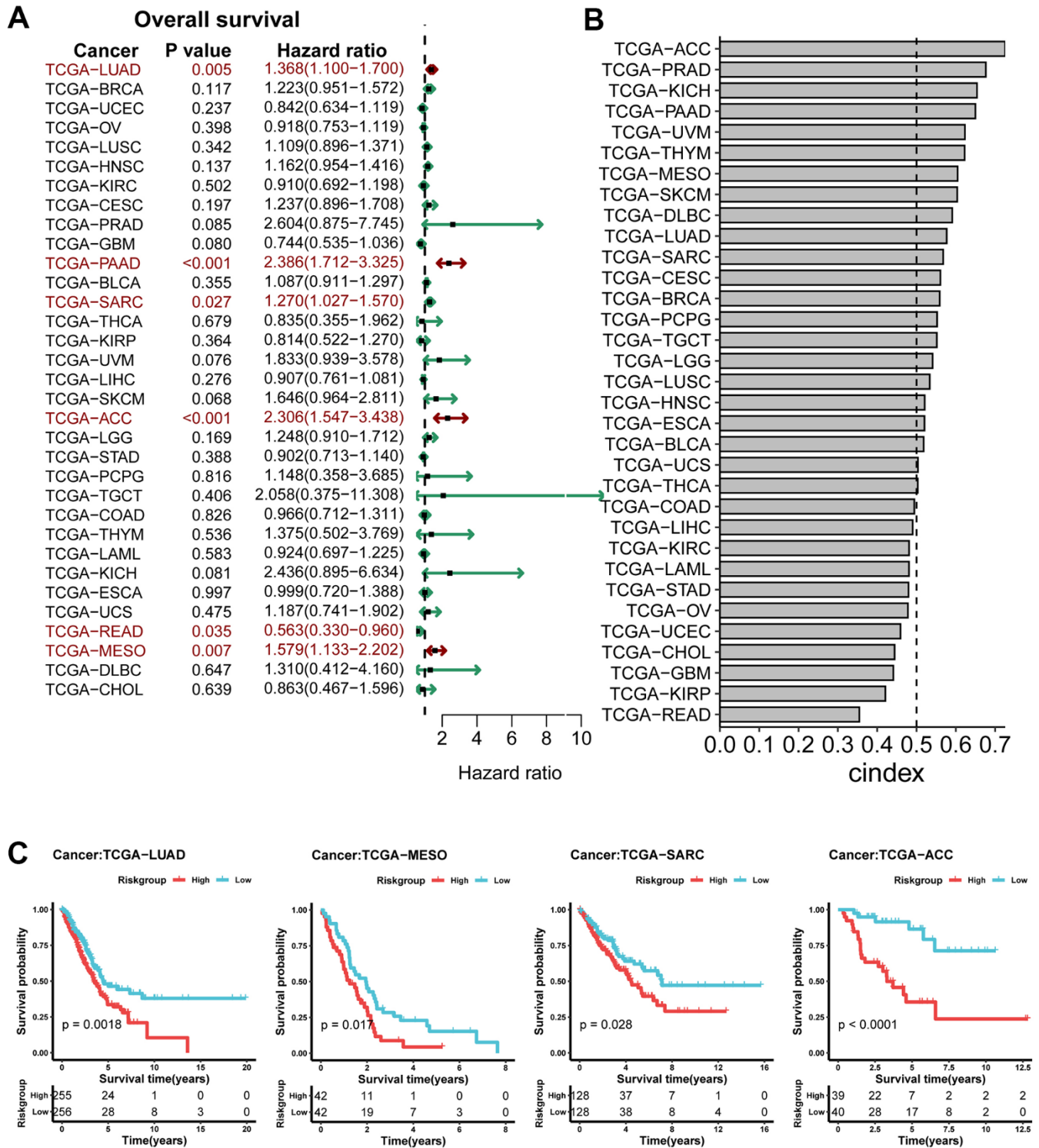
Figure 7. Analysis of prognostic value of risk model in pan-cancer. (**A**) Analysis of the relationship between survival model and OS in pan-cancer; (**B**) Cindex of survival model on pan-cancer; (**C**) KM curves of TCGA-LUAD, TCGA-MESO, TCGA-SARC, TCGA-ACC.
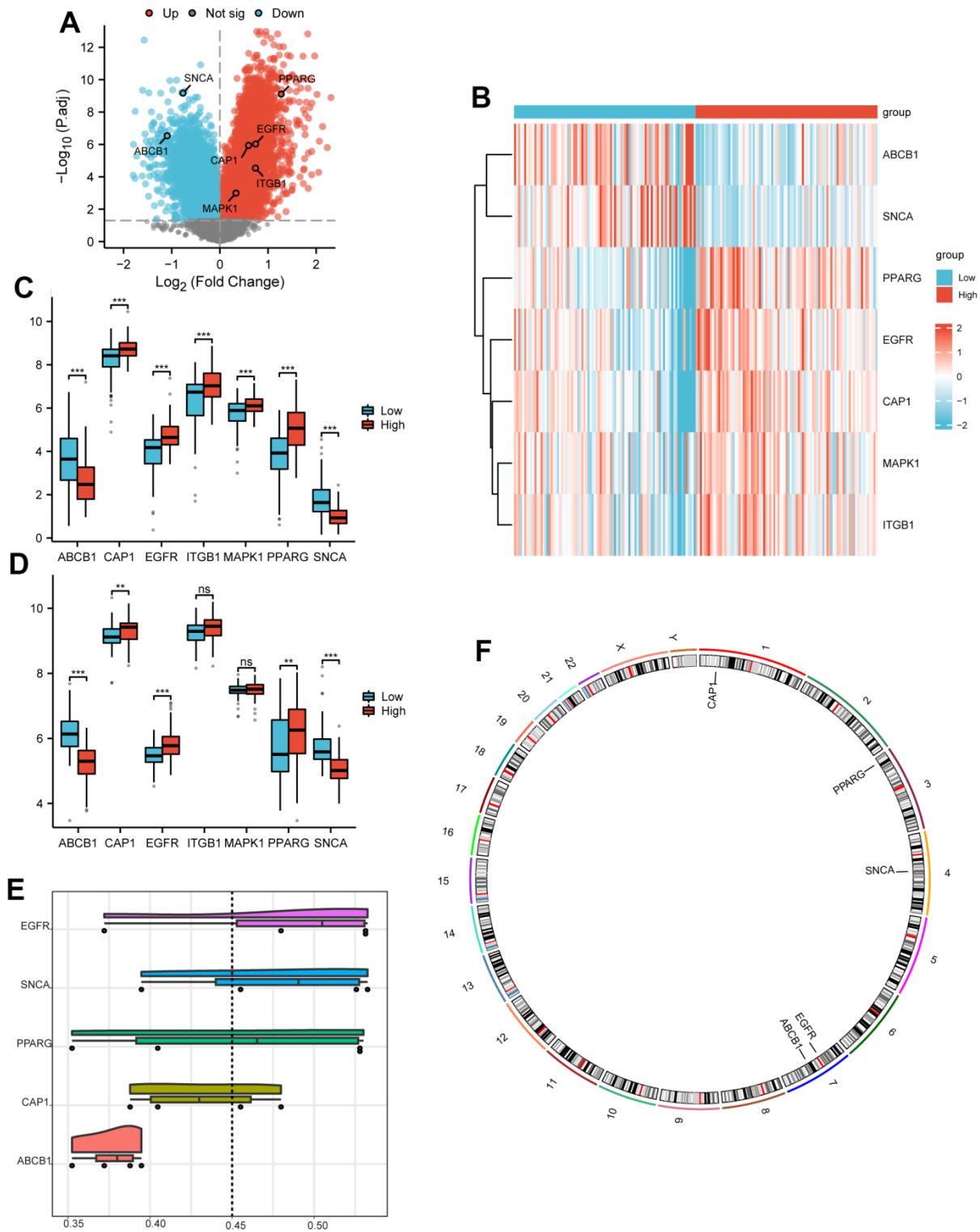
Figure 8. Differential expression and functional similarity analysis. (**A**) Volcano plot of DEGs between high and low-risk groups in TCGA-PAAD, mapping 8016 upregulated genes (red dots) and 5254 downregulated genes (blue dots). No significantly changed genes are marked as gray dots. (**B**) The heatmap to show high and low-risk groups and 7 key gene expressions. (**C**, **D**) Boxplots to show the expression difference of *ABCB1*, *CAP1*, *EGFR*, *ITGB1*, *MAPK1*, *PPARG*, and *SNCA* between high and low-risk groups in TCGA-PAAD dataset (**C**) and Dataset PAAD (**D**). (**E**) Raincloud plots to show the functional similarity analysis of 5 hub genes; X-axis represents a similarity score, and the larger the value, the higher the functional similarity with other genes. (**F**) Chromosomal mapping of 5 Hub genes. Ns represents P-value ≥ 0.05, with no statistically significant difference; * represents P-value < 0.05, with a statistically significant difference; ** represents P-value < 0.01; *** represents P-value < 0.001.

We also analyzed the differential expression of 7 key genes (*ABCB1*, *CAP1*, *EGFR*, *ITGB1*, *MAPK1*, *PPARG*, *SNCA*) between low and high-risk groups in the TCGA-PAAD dataset and drew a heatmap using R packet (Figure 8B). The differential expression of key genes can be seen in Table 3. Boxplots were drawn for key genes in the TCGA-PAAD dataset and Dataset-PAAD (Figure 8C, 8D). The expression trends of all key genes in both datasets were consistent between high and low-risk groups.

According to Figure 8D, 5 genes (*ABCB1*, *CAP1*, *EGFR*, *PPARG*, *SNCA*) showed statistical significance in Dataset-PAAD, selected as hub genes for subsequent analysis.

We calculated semantic similarity among GO terms, sets of GO terms, gene products, and gene clusters using the GOSemSim.R package to analyze the functional similarity of these 5 hub genes. We then visualized the functional similarity analysis results through Raincloud plots (Figure 8E). The results revealed that EGFR had better functional similarity than other hub genes. We also used the RCircos package to annotate the location of 5 hub genes on human chromosomes (Figure 8F). As shown in the graph, these hub genes were mainly distributed on chromosomes 1, 3, 4, and 7, among which chromosome 7 was distributed with 2 hub genes. This indicated that these hub genes had close positions on human chromosomes and were also closely related at the genomic level.

Pancreatic cancer single cell dimension reduction annotation

Firstly, by setting each gene to be expressed in at least 3 cells and at least 200 genes per cell, the single-cell data was filtered to obtain 55339 cells. Moreover, PercentageFeatureSet function was used to calculate the proportion of mitochondria and rRNA to ensure that the expressed genes of each cell are greater than 200 and less than 10,000, and Supplementary Table 1 shows the count of cells before and after filtration. The Supplementary Figure 2A indicates that UMI is significantly correlated with the amount of mRNA, while UMI/mRNA is not significantly correlated with the content of mitochondrial genes. Supplementary Figure 2B indicates the violin diagram before (TOP) and after (Bottom) quality control.

Further, we standardized the data for each of the three samples by log-normalization, searched for highly variable genes (based on variance stabilization transformation ("vst") to identify variable features) by FindVariableFeatures function, scaled all genes by ScaleData function and found anchor points by PCA

dimensionality reduction with RunPCA (Supplementary Figure 2C). Dim=35 was selected, and finally, the Find Neighbors and FindClusters functions were used to cluster cells (Resolution=0.3), and a total of 20 subgroups (Supplementary Figure 3) were obtained. At the same time, the RunTSNE function was used to conduct TSNE dimension reduction analysis on 51492 cells. the classical immune cell marker (The markers used were shown in Supplementary Figure 4) reported in the literature was combined with the SigleR algorithm to annotate 20 subgroups of cells, and finally 17 cell types were obtained and visualized according to sample source and cell type (Figure 9A–9D).

Subsequently, we used the FindAllMarkers function to screen marker genes for 17 cell types using logfc=0.35 (multiple of differences) and Minpct=0.35 (smallest expression ratio of differential genes) and performed the screening with a corrected $p<0.05$. Here, we only show the expression of the top 5 marker genes with the most outstanding contributions in each subpopulation (Figure 9E), and the results of marker genes are shown in the table scRNA_marker_gene.txt. Further, we analyzed the proportion of these 17 types of cells in each sample (Figure 9F).

To further identify malignant cells, we used copykat algorithm to predict malignant cells (Figure 9G, 9H).

Hub gene was analyzed at the single-cell level

We first examined the expression of hub gene in different cell types (Figure 10A), and then based on the 5 BP results enriched by hub gene on Bulk data, we used the ssgsea algorithm to score these 5 BP in each cell type (Figure 10B–10F).

Functional enrichment analysis of Hub genes and DEGs

The function of these 5 hub genes was predicted by analyzing GO and KEGG. The GO annotations consisted of three parts, including BP, CC, and MF with criterion P.adj < 0.1, which were used to analyze the functional enrichment of genes. These 5 hub genes were mainly related to BP of regulation of JNK cascade (GO:0046328), JNK cascade (GO:0007254), regulation of transporter activity (GO:0032409), organic acid transport (GO:0015849), regulation of protein serine/threonine kinase activity (GO:0071900) in GO annotation analysis. These genes were also mainly associated with cell cortex (GO:0005938), vesicle lumen (GO:0031983), platelet alpha granule membrane (GO:0031092), focal adhesion (GO:0005925), cell-substrate junction (GO:0030055) in CC and actin binding (GO:0003779), ubiquitin protein ligase binding

Table 3. Key genes DEA in TCGA-PAAD.

GENE	logFC	AveExpr	t	P-value	P.adj	B
SNCA	-0.76663	1.396627692	-7.459	3.71E-12	6.79E-10	17.23201
PPARG	1.272794	4.420288301	7.424397	4.53E-12	7.81E-10	17.03936
ABCB1	-1.09405	3.098367518	-6.05659	8.08E-09	2.91E-07	9.838242
EGFR	0.741565	4.344547598	5.763118	3.58E-08	9.45E-07	8.413015
CAP1	0.595678	8.441607674	5.713291	4.59E-08	1.15E-06	8.175759
ITGB1	0.738944	6.717817488	4.87041	2.45E-06	2.90E-05	4.386726
MAPK1	0.334392	5.940672572	3.822517	0.000183	0.001014	0.336601

DEA, Differential Expression Analysis; TCGA, The cancer genome atlas.

(GO:0031625), prostaglandin receptor activity (GO:0004955), ubiquitin-like protein ligase binding (GO:0044389), and prostanoid receptor activity (GO:0004954) in MF. The results of GO annotations can be seen in Table 4. KEGG analysis was conducted to explore the relationship between hub genes and signaling pathways. The results indicated that hub genes were not enriched in the KEGG pathway.

We presented the results of GO annotations using bar charts (Figure 11A) and network charts (Figure 11B–11D). Subsequently, the GO annotations of a joint logFC in these 5 hub genes were conducted. Based on the enrichment analysis, we calculated the z-score corresponding to each molecule by providing the logFC values of the key genes from the difference analysis results of the TCGA-PAAD high and low-risk groups. We showed the GO enrichment analysis results of a joint logFC using a bubble plot (Figure 11E).

We further performed GSEA between the high and low-risk groups based on the prognostic model to uncover enrichment signaling pathways in PAAD with P.adj < 0.05 and FDR value (qvalue) < 0.25 (Supplementary Table 2). The results showed that genes in the high and low-risk groups were significantly associated with NO2IL12 pathway (Supplementary Figure 5B), peptide hormone metabolism (Figure 8C), ADORA2B mediated anti-inflammatory cytokines production (Supplementary Figure 5D), metabolic reprogramming in pancreatic cancer (Supplementary Figure 5E), MET promotes cell motility (Supplementary Figure 5F), assembly of collagen fibrils, and other multimeric structures (Supplementary Figure 5G). We drew a ridge plot (Supplementary Figure 5A) to display the above results.

We also conducted a GSVA to explore the differences of Hallmark genes in the prognosis model between the high and low-risk groups. The results showed significant differences in the high and low-risk groups of the TCGA-PAAD prognostic model among 36 hallmark genes sets, including adipogenesis, androgen

response, apical junction, apical surface, and apoptosis (Supplementary Table 3). We also analyzed the differential expression of these 36 HALLMARK pathways between high and low-risk groups in the TCGA-PAAD dataset and drew a heatmap to display the results using pheatmap R packet (Supplementary Figure 6A). According to the results of GSVA, we investigated the differences of 36 HALLMARK pathways between different groups in the TCGA-PAAD dataset using the Mann Whitney U test and displayed the results through grouped bar chart (Supplementary Figure 6B), which revealed that the expression of these 36 HALLMARK pathways were significantly different between high and low-risk groups.

Construction of PPI network and identification of prognostic-related gene

To better understand the potential interactions between the 5 hub genes (*ABCB1*, *CAP1*, *EGFR*, *PPARG*, *SNCA*) in the PAAD group, a PPI network was constructed using the STRING database, and the interaction threshold was set at 0.400. The PPI analysis is visualized in Figure 12A. The results showed that other hub genes interacted with at least one hub gene except for *CAP1* under the least required interaction score of 0.400. Among them, *EGFR* had interactions with three hub genes, a gene with many interactions with other genes. In addition, we also predicted and constructed an interaction network of functionally similar genes among these five hub genes using GeneMANIA to observe their interactions, co-expression, co-localization, and other information (Figure 12B).

Besides, we used mRNA-miRNA data from the miRDB database to predict miRNAs that interact with five hub genes. We then screened with the criteria of target score>90, and the interaction pairs were visualized by Cytoscape to show mRNA-miRNA interaction network (Figure 12C). A total of 65 pairs mRNA-miRNA interactions were constituted by 5 hub genes (*ABCB1*,
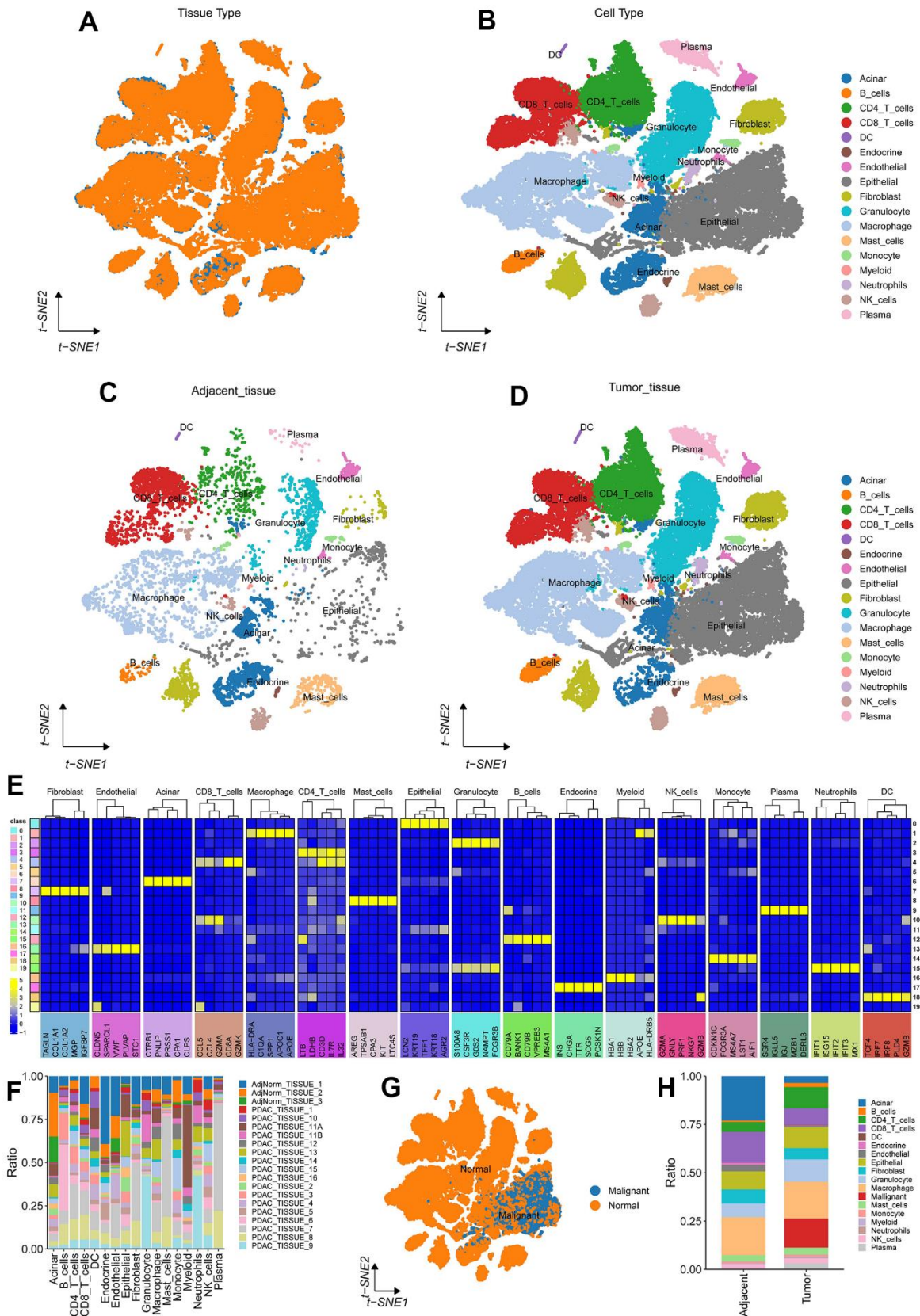
Figure 9. Pancreatic cancer single cell dimension reduction annotation. (**A**, **B**) tSNE on Tumor(yellow)/Adjacent(blue) and all patient tissues. (**C**, **D**) tSNE on 3 adjacent/normal pancreas (left) and 16 PDA patient (right) tissues. (**E**) Top 5 marker basis for each cell type; (**F**) Cell count statistics for each sample; (**G**) Copykat prediction results; (**H**) The proportion of various cell types adjacent to the tumor.

CAP1, *EGFR*, *PPARG*, *SNCA*) and 62 miRNAs (Supplementary Table 4).

TF control gene expression through interactions with target genes (mRNA) during the post transcriptional stage. We searched for TFs combined with hub genes through CHIPBase and hTFtarget database, downloaded the interaction relationships found in the two databases, and performed the intersection. Finally, a total of 67 pairs of mRNA-TFs interaction relationships were constituted by 5 hub genes (*ABCB1*, *CAP1*, *EGFR*, *PPARG*, *SNCA*) and 51 TFs, which were visualized using Cytoscape (Figure 12C). The specific mRNA-TFs interaction relationships are shown in the Supplementary Table 5.

DISCUSSION

Due to the presence of atypical symptoms, insidious location, rapid disease progression, and poor prognosis, PAAD has emerged as one of the most malignant and aggressive solid tumors [32]. Although surgical and adjuvant treatment for pancreatic cancer has been extensively developed, it is still insufficient to improve the prognosis of PAAD patients. However, exploration and mining of effective biomarkers in pancreatic cancer have become increasingly vital and beneficial for diagnosing, treating, and supervising PAAD patients, which are still relatively scarce. Serum CA19-9 is the only serum biomarker for treatment approved by the Food and Drug Administration [33], and it also has a certain suggestive effect on the evaluation of prognosis and overall survival rate in PAAD patients [34]. Nevertheless, CA19-9 is significantly elevated in a variety of benign diseases of the digestive system, such as cirrhosis, pancreatitis, and cholangitis. Furthermore, there are non-specific expressions of false negative results caused by the Lewis negative genotype and false positive results caused by obstructive jaundice, which severely limits the application of serum CA19-9 in the PAAD prognosis prediction [35]. Thus, it is crucial to identify novel biomarkers to predict prognosis and enhance individualized therapies in PAAD patients. Exosomes, as novel biomarkers for tumors, have been developed for constructing prognostic models in esophageal carcinoma [36], hepatocellular carcinoma [37], and pancreatic ductal adenocarcinoma [38]. Nowadays, increasing studies have found that abnormal lipid metabolism is associated with tumors, and researchers also constructed a prognostic model based
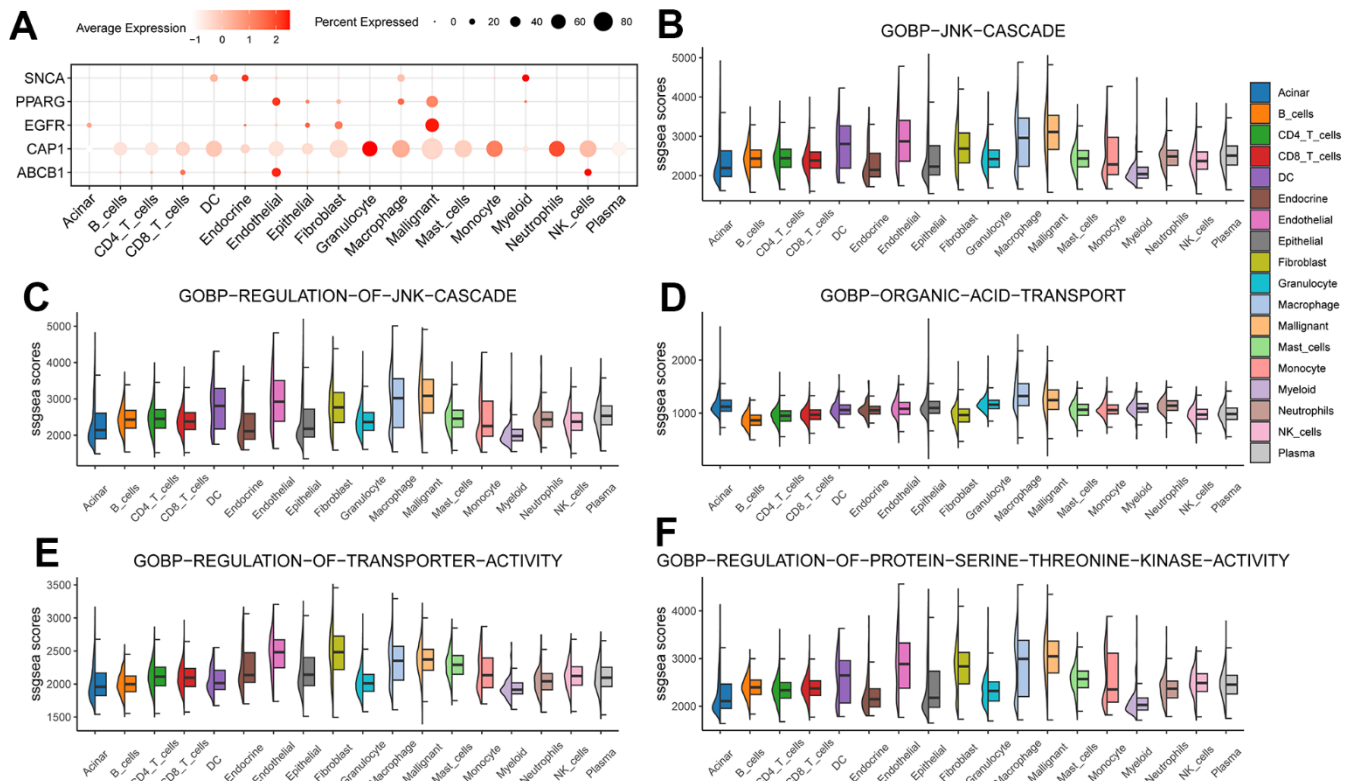

Figure 10. Hub gene was analyzed at the single-cell level. (**A**) Hub gene expression in individual cell types; (**B**–**F**) 5 BP enrichment scores in each cell type.

Table 4. GO enrichment analysis results of hub genes.

ONTOLOGY	ID	Description	GeneRatio	BgRatio	P.adj	q-value
	GO:0046328	regulation of JNK cascade	2023/1/5	141/18800	0.06017	0.01692
	GO:0007254	JNK cascade	2023/1/5	175/18800	0.06518	0.01833
	GO:0032409	regulation of transporter activity	2023/3/5	305/18800	0.00793	0.00223
BP	GO:0015849	organic acid transport	2023/3/5	318/18800	0.00793	0.00223
	GO:0071900	regulation of protein serine/threonine kinase activity	2023/3/5	372/18800	0.0087	0.00245
	GO:0032768	regulation of monooxygenase activity	2023/2/5	57/18800	0.00916	0.00258
	GO:0010517	regulation of phospholipase activity	2023/2/5	66/18800	0.01093	0.00307
	GO:0005938	cell cortex	2023/2/5	310/19594	0.05377	0.0259
	GO:0031983	vesicle lumen	2023/2/5	327/19594	0.05377	0.0259
	GO:0031092	platelet alpha granule membrane	2023/1/5	17/19594	0.05377	0.0259
CC	GO:0005925	focal adhesion	2023/2/5	419/19594	0.05377	0.0259
	GO:0030055	cell-substrate junction	2023/2/5	428/19594	0.05377	0.0259
	GO:0015629	actin cytoskeleton	2023/2/5	499/19594	0.06048	0.02914
	GO:0005771	multivesicular body	2023/1/5	63/19594	0.08852	0.04264
	GO:0003779	actin binding	2023/3/5	439/18410	0.01196	0.00287
	GO:0031625	ubiquitin protein ligase binding	2023/2/5	298/18410	0.02641	0.00635
	GO:0004955	prostaglandin receptor activity	2023/1/5	10/18410	0.02641	0.00635
MF	GO:0044389	ubiquitin-like protein ligase binding	2023/2/5	317/18410	0.02641	0.00635
	GO:0004954	prostanoid receptor activity	2023/1/5	11/18410	0.02641	0.00635
	GO:0008179	adenylate cyclase binding	2023/1/5	12/18410	0.02641	0.00635
	GO:0097677	STAT family protein binding	2023/1/5	12/18410	0.02641	0.00635

GO, Gene Ontology; BP, biological process; CC, cellular component; MF, molecular function.

on lipid metabolism-related genes to predict patients with gastric [39] and hepatocellular carcinoma [40]. Even though exosomes or lipid metabolism genes have been shown to be accurate predictors of tumor prognosis in previous studies, there is still lack of research on EALMRGs as a predictive biomarker for PAAD.

In this study, we constructed a 7-gene (*ABCB1*, *CAP1*, *EGFR*, *ITGB1*, *MAPK1*, *PPARG*, *SNCA*) prognostic model based on EALMRGs using LASSO logistic regression analysis. Calibration analysis results suggested that the prediction performance in 1- and 3-year were better than that in 5-year; however, the DCA analysis results were the opposite. Several previous studies also reported some specific prognostic models in PAAD. For example, Yang et al. constructed a 3-gene (*ALOX5*, *ALOX12*, *CISD1*) prognostic model based on ferroptosis-related genes, and they found that the model had a high AUC (=0.976) at 5-year in TCGA-training set, as well as in the GSE62452 (AUC=0.743) [41]. Moreover, a novel 9-gene prognostic model based on multi-omics in PAAD was developed and proven to have a good performance (average AUC>0.8) in predicting the prognosis [42]. Furthermore, Su et al. reported the high prognostic performance of pyroptosis-

related genes (*GSDMC*, *IRF1*, and *PLCG1*). They found that this model performed well in predicting the 1-, 3- and 5-year survival of TCGA-PAAD patients [43]. Despite this, our model demonstrates good predictive value in general, and we are the first to examine EALMRG signatures in PAAD prognosis prediction.

We also obtained 13270 DEGs between high and low-risk groups based on the EALMRGs related prognostic model from TCGA-PAAD, with 8016 up-regulated genes and 5254 down-regulated genes. The GSEA results showed that these DEGs were closely related to metabolic reprogramming in pancreatic cancer. Metabolic reprogramming in pancreatic cancer, including glucose metabolism, lipid metabolism, tricarboxylic acid cycle, and amino acid metabolism, not only provides nutrition for tumor occurrence and development, but also affects the function of anti-tumor immune cells and immunosuppressive cells in the microenvironment [44]. Besides, mesenchymal-epithelial transition (MET) promoting cell motility plays an important role in various tumors [45]. However, the GSVA results suggested that these DEGs differed in 36 Hallmarks, such as androgen response and apical junction. A previous study reported that androgen receptors and related responses were

associated with human carcinogenesis in hepatocellular and pancreatic cancer [46]. The androgen response has also been proven to play a vital role in the occurrence and development of tumors [47]. The results of our study were consistent with this literature, but the specific mechanism in PAAD still needs further exploration.

We identified 5 significant EALMRGs, including *ABCB1*, *CAP1*, *EGFR*, *PPARG*, and *SNCA*. The results suggested that all of them have good predictive performance in the fifth year, with *ABCB1* and *SNCA* as protective factors having better overall predictive ability at 1-, 3- and 5-year. These 5 genes are closely distributed on chromosomes, indicating a close functional connection at the genomic level. Thus, we further explored the biological role of these 5 genes and PAAD. Consistent with the published data, these 5

genes were enriched in several GO terms, such as protein serine/threonine kinase activity [48], focal adhesion [49], and actin binding [50], which have suggested the possible relationship with their regulation and mediation in PAAD patients. The enrichment analysis result also indicated that cell-substrate junction is involved in the progress of several cancers, including gastric [51], nasopharyngeal [52], and lung cancers [53], but pancreatic cancer has not yet been reported. However, disease and disorder research on cancers has not been conducted in relation to the organic acid transport and regulation of transporter activity, which is likely a direction for our research on tumor mechanisms.

Through PPI construction, we identified several genes, including mRNA (*CAP2*, *SLC22A3*), miRNA (*miR-9985*, *miR-27*, *miR-548*), and TFs (*CEBPA*), involved in
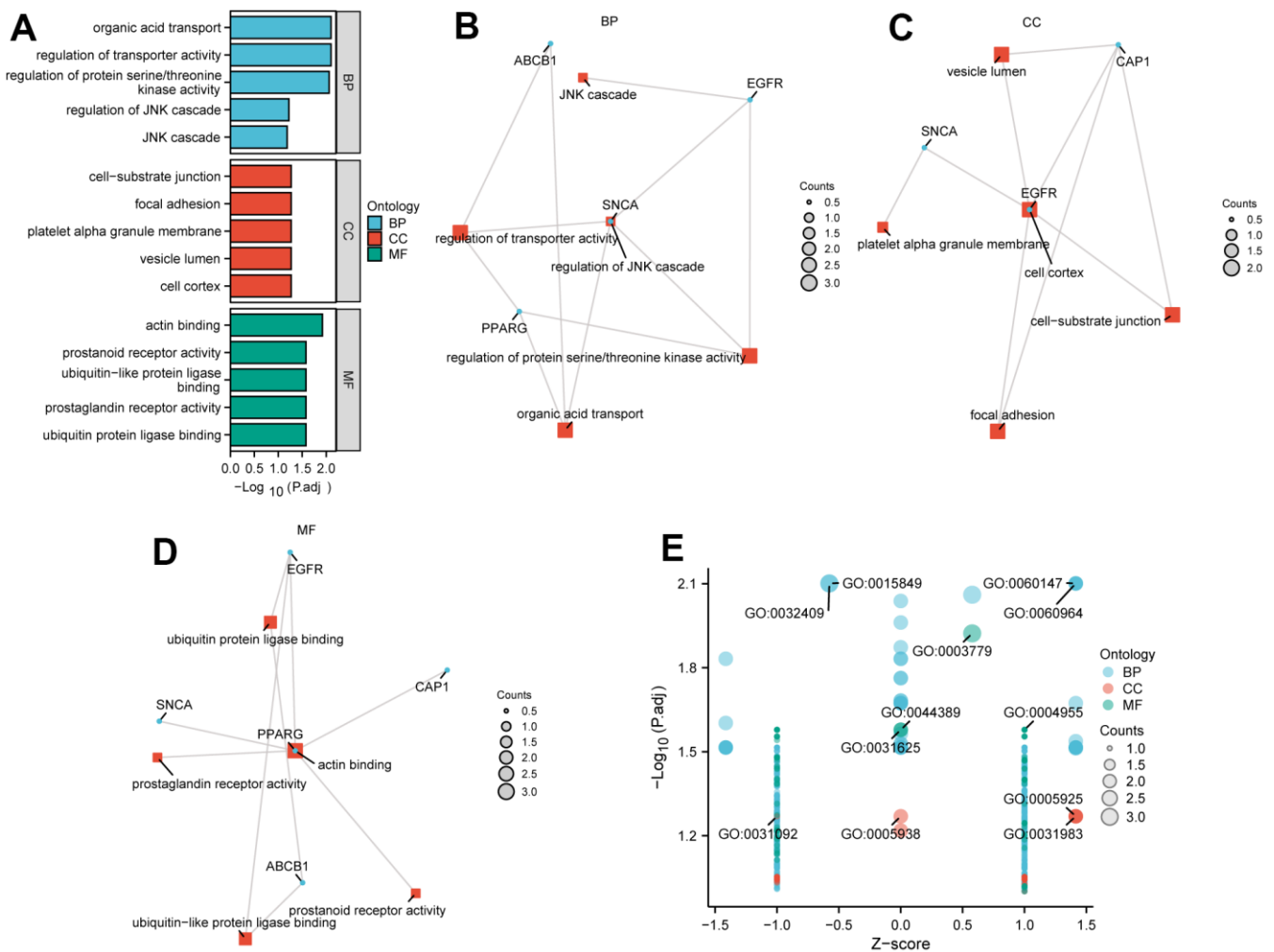
Figure 11. GO enrichment analysis. (**A**)The GO enrichment analyses of DEGs in Hub genes. (**B–D**) Chordal graph of GO enrichment for 5 Hub genes: BP pathway (**B**), CC pathway (**C**), and MF pathway (**D**). In the network diagram, blue dots represent specific genes, and red blocks represent specific pathways. (**E**) A bubble plot shows GO enrichment of a joint logFC. GO, Gene Ontology.
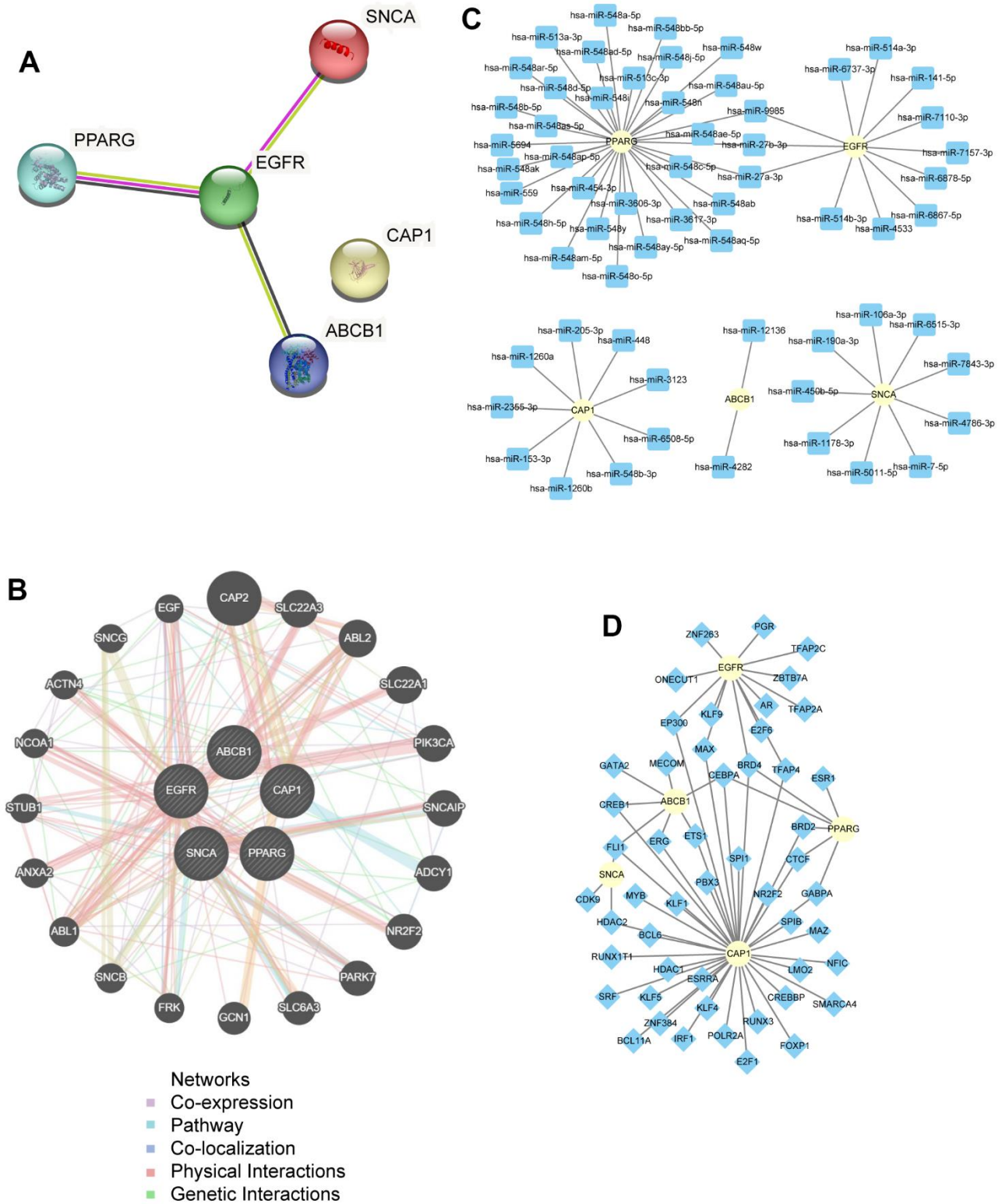
Figure 12. PPI network analysis. (**A**) PPI network of hub genes. (**B**) PPI network of functionally similar genes analysis in Hub genes. Black circles with white slashes represent the input hub genes, and other black circles without white slashes represent predicted functionally similar genes; red lines represent physical interactions among genes, purple lines represent co-expression relationships among genes, yellow lines represent shared protein domains among genes, blue lines represent co-localization relationships among genes, and green lines represent genetic interactions among genes. (**C**) An interaction network of mRNA-miRNA in hub genes. The yellow circle represents mRNA; the blue square represents miRNA. (**D**) An interaction network of mRNA-TFs in hub genes. The yellow circle represents mRNA; the blue diamond shape represents TFs. PPI, protein-protein interaction; TF, Transcription Factor.

the prognosis mechanism of pancreatic cancer. The *SLC22A3* and *CAP2* might become specific predictive signatures for diagnosing pancreatic cancer [54–56]. *Has-miR-9985-584p* could regulate most identified ferroptosis-related genes to participate in the developing type 2 diabetic islets [57], which is still an important pathogenic factor for pancreatic cancer. MicroRNAs such as *hsa-miR-27a-3p* and *hsa-miR-27b-3p* were found to correlate with *EGFR* and *PPAG* in PAAD, and these 2 microRNAs were also correlated with gastric and esophageal cancers [58, 59]. However, *hsa-miR-548-family* seems to embrace a very close relationship with *PPARG* from our study, and it belongs to the top down-regulated colorectal cancer from Josef Horak's research [60]. Dysregulation of *CEBPA* expression is widely reported in human cancer, for which various mechanisms have been described [61]. A study reported that *CEBPA* could promote the proliferation, invasion and migration of pancreatic cancer cells, and upregulation of *CEBPA* can be induced by *KDM6B* knockdown [62]. The above results not only indicate that these 5 EALMRGs have great potential to predict the prognosis of PAAD forebode the possible interaction of genes and molecules, and further provide a theoretical basis for the EALMRGs prognosis mechanism.

Currently, an increasing body of evidence suggests that these 5 hub genes are associated with the malignancy of tumors, especially PAAD. ATP-binding cassette transporter B1 (*ABCB1*) is called MDR1 P-glycoprotein and has been reported to play an important role in the chemotherapy resistance of pancreatic cells through upregulation of drug efflux [63]. *ABCB1* is a pivotal transcriptional factor of WNT/β-catenin signaling, and the upregulation of ABCB1 expression was modulated by the specific gain-of-function *CTNNB1* mutations [64]. Notably, transcriptional level of *ABCB1* is also increased through leptin activation from tumor-related microenvironment [65]. Another research reported that patients with high protein levels of *ABCB1* are correlated with worse prognosis [66]. However, our study found that an up-regulated level of *ABCB1* could improve the survival rate of pancreatic cancer. The opposite result might relate to the association of *ABCB1* with molecular status, tumor characteristics, demographics, and genetic variants [67], which still needs more laboratory and clinical data to be explored and validated.

Initially, *SNCA* was firstly identified as a pivotal promoter in the development of Parkinson's disease [68] and is important in maintaining mitochondrial homeostasis, proteasome function, and molecular chaperone activity [69]. Several studies reported that *SNCA* is related to carcinogenesis. In a meta-analysis of genetic parkinsonism and cancer, SNCA was predominantly associated with gastrointestinal cancers [70], such as colorectal cancer [71]. Matteo Bianchini and co-workers discovered that α-syn was significantly upregulated in PDAC [72]. Nevertheless, our study found that SNCA was positively associated with the survival rate of PAAD patients, suggesting that an intricate molecular mechanism exists between SNCA and PAAD that needs to be further explored.

CAP1 has been reported with upregulation in multiple gastrointestinal, breast, and lung cancers [73–84]. Peroxisome proliferator-activated receptor gamma (*PPARG*) is a member of the nuclear receptor superfamily of ligand-activated transcription factors [85], which is involved in lipid and glucose homeostasis and adipocyte inflammation and differentiation, among other activities [86–88]. Furthermore, many studies have shown that *PPARG* plays a vital role in regulating the growth of several different cancers, including prostate [89], bladder [90], breast [91], and colorectal cancer [92]. The *PPARG* and *DNMTs* appear interrelated in pancreatic cancer, and this interaction might influence cell phenotype and disease behavior [93]. Together with our study, these results suggest that *PPARG* embraces potential value in predicting the prognosis of PAAD patients, remaining to be further explored.

EGFR belongs to a family of receptor tyrosine kinases that comprises three other members [94–96]. Several studies have revealed that the increased expression of *EGFR* is widely identified in multiple cancers [97–102]. *EGFR* was reported with overexpression in most proportions of PDAC and correlation with multiple clinical characteristics [103, 104]. These studies indicated that *EGFR* might be a potential prognostic gene in PAAD.

These 5 EALMRGs have a role in the biological behavior of pancreatic tumors; in addition, they have exhibited good performance in prognosis efficacy for PAAD patients. To our knowledge, the clinical value of exosome and lipid metabolism-related genes has not been fully elucidated in multiple cancers. Before this study, Ye et al. developed the 4-genes risk formula based on lipid metabolism-related genes [105]. Zhu et al. identified 11-genes risk signature formula related to lipid metabolism [106]. However, these formulas did not extend the potential clinical impact of exosome-related genes. Our current study identified the change of exosome and lipid metabolism-related genes with clinical value, and the 7-gene (ABCB1, CAP1, EGFR, ITGB1, MAPK1, PPARG, SNCA) prognostic formula is a novel and reliable predictive index in pancreatic cancer and other multiple cancers.

Nonetheless, there are inevitable limitations to our research, which will be rectified and improved in our following study. Firstly, our data originated from previously published datasets, and we performed second data mining and analysis through our own procedure. More thorough and comprehensive investigations are needed prior to clinical application. Secondly, we validated our model with some other datasets in GEO datasets, this method cannot substitute for validation using in-house clinical data. Consequently, further rigorous prospective studies are needed to evaluate the feasibility and authenticity of the model in clinical applications. Thirdly, despite the 5 genes showing good performance in predicting the prognosis of PAAD patients, further verification of these genes *in vivo* and *in vitro* is still needed.

CONCLUSIONS

In conclusion, the EALMRGs prognostic model of PAAD was constructed using LASSO regression. The model was evaluated by internal and external validation and showed good prognostic performance. For the first time, we further identified 5 EALMRGs (*ABCB1*, *CAP1*, *EGFR*, *PPARG,* and *SNCA*) that might become prognostic biomarkers of PAAD, of which *EGFR* exhibited better prognostic efficiency than other genes. Additionally, the risk formula not only showed potential prognostic value in multiple cancers, but also manifested pivotal alterations in immune infiltration and biological pathway in pancreatic cancer. These findings could provide an effective and reliable method of predicting the prognosis of PAAD patients, which might be a potential and specific direction for further research.

AUTHOR CONTRIBUTIONS

JW, YL: Data collection, analysis. GN: Polishing and critical revision. XH, XZ, YW, LH: drafting manuscript and analysis.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

REFERENCES

1. Rahib L, Wehner MR, Matrisian LM, Nead KT. Estimated Projection of US Cancer Incidence and Death to 2040. JAMA Netw Open. 2021; 4:e214708. https://doi.org/10.1001/jamanetworkopen.2021.4708 PMID:33825840

2. Ma JF, He HR, Yan AT, Ding JC, Zhu ZE, Wu Z. Epidemiological trends and major risk attribution analysis of pancreatic cancer in China, Japan and South Korea from 1990 to 2019. Chinese Journal of Digestive Surgery. 2022; 21: 507–19. https://doi.org/10.3760/cma.j.cn115610-20220310-00126

3. Yang Y. Current status and future prospect of surgical treatment for pancreatic cancer. Hepatobiliary Surg Nutr. 2020; 9:89–91. https://doi.org/10.21037/hbsn.2019.12.04 PMID:32140487

4. Sun W, Ren Y, Lu Z, Zhao X. The potential roles of exosomes in pancreatic cancer initiation and metastasis. Mol Cancer. 2020; 19:135. https://doi.org/10.1186/s12943-020-01255-w PMID:32878635

5. Sorop A, Constantinescu D, Cojocaru F, Dinischiotu A, Cucu D, Dima SO. Exosomal microRNAs as Biomarkers and Therapeutic Targets for Hepatocellular Carcinoma. Int J Mol Sci. 2021; 22:4997. https://doi.org/10.3390/ijms22094997 PMID:34066780

6. Ariston Gabriel AN, Wang F, Jiao Q, Yvette U, Yang X, Al-Ameri SA, Du L, Wang YS, Wang C. The involvement of exosomes in the diagnosis and treatment of pancreatic cancer. Mol Cancer. 2020; 19:132. https://doi.org/10.1186/s12943-020-01245-y PMID:32854710

7. Costa-Silva B, Aiello NM, Ocean AJ, Singh S, Zhang H, Thakur BK, Becker A, Hoshino A, Mark MT, Molina H, Xiang J, Zhang T, Theilen TM, et al. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. Nat Cell Biol. 2015; 17:816–26. https://doi.org/10.1038/ncb3169 PMID:25985394

8. Snaebjornsson MT, Janaki-Raman S, Schulze A. Greasing the Wheels of the Cancer Machine: The Role of Lipid Metabolism in Cancer. Cell Metab. 2020; 31:62–76. https://doi.org/10.1016/j.cmet.2019.11.010 PMID:31813823

9. Liu Q, Luo Q, Halim A, Song G. Targeting lipid metabolism of cancer cells: A promising therapeutic strategy for cancer. Cancer Lett. 2017; 401:39–45. https://doi.org/10.1016/j.canlet.2017.05.002 PMID:28527945

10. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN, Zhu J, Haussler D. Visualizing and interpreting cancer

genomics data via the Xena platform. Nat Biotechnol. 2020; 38:675–8.
https://doi.org/10.1038/s41587-020-0546-8
PMID:32444850

11. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43:e47.
https://doi.org/10.1093/nar/gkv007 PMID:25605792

12. Yang S, He P, Wang J, Schetter A, Tang W, Funamizu N, Yanaga K, Uwagawa T, Satoskar AR, Gaedcke J, Bernhardt M, Ghadimi BM, Gaida MM, et al. A Novel MIF Signaling Pathway Drives the Malignant Character of Pancreatic Cancer by Targeting NR3C2. Cancer Res. 2016; 76:3838–50.
https://doi.org/10.1158/0008-5472.CAN-15-2841
PMID:27197190

13. Chen DT, Davis-Yadley AH, Huang PY, Husain K, Centeno BA, Permuth-Wey J, Pimiento JM, Malafa M. Prognostic Fifteen-Gene Signature for Early Stage Pancreatic Ductal Adenocarcinoma. PLoS One. 2015; 10:e0133562.
https://doi.org/10.1371/journal.pone.0133562
PMID:26247463

14. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, et al. NCBI GEO: archive for functional genomics data sets-- update. Nucleic Acids Res. 2013; 41:D991–5.
https://doi.org/10.1093/nar/gks1193 PMID:23193258

15. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016; 54:1.30.
https://doi.org/10.1002/cpbi.5
PMID:27322403

16. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011; 27:1739–40. Epub 2011 May 5.
https://doi.org/10.1093/bioinformatics/btr260
PMID: 21546393

17. Engebretsen S, Bohlin J. Statistical predictions with glmnet. Clin Epigenetics. 2019; 11:123.
https://doi.org/10.1186/s13148-019-0730-1
PMID:31443682

18. Cai W, van der Laan M. Nonparametric bootstrap inference for the targeted highly adaptive least absolute shrinkage and selection operator (LASSO) estimator. Int J Biostat. 2020. [Epub ahead of print].

https://doi.org/10.1515/ijb-2017-0070
PMID:32772002

19. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics. 2010; 26:976–8.
https://doi.org/10.1093/bioinformatics/btq064
PMID:20179076

20. Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015; 43:D1049–56.
https://doi.org/10.1093/nar/gku1179 PMID:25428369

21. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16:284–7.
https://doi.org/10.1089/omi.2011.0118
PMID:22455463

22. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28:27–30.
https://doi.org/10.1093/nar/28.1.27 PMID:10592173

23. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005; 102:15545–50.
https://doi.org/10.1073/pnas.0506580102
PMID:16199517

24. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015; 1:417–25.
https://doi.org/10.1016/j.cels.2015.12.004
PMID:26771021

25. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013; 14:7.
https://doi.org/10.1186/1471-2105-14-7
PMID:23323831

26. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019; 47:D607–13.
https://doi.org/10.1093/nar/gky1131 PMID:30476243

27. Franz M, Rodriguez H, Lopes C, Zuberi K, Montojo J, Bader GD, Morris Q. GeneMANIA update 2018. Nucleic Acids Res. 2018; 46:W60–4.
https://doi.org/10.1093/nar/gky311 PMID:29912392

28. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res. 2020; 48:D127–31.
https://doi.org/10.1093/nar/gkz757 PMID:31504780

29. Zhou KR, Liu S, Sun WJ, Zheng LL, Zhou H, Yang JH, Qu LH. ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. Nucleic Acids Res. 2017; 45:D43–50.
https://doi.org/10.1093/nar/gkw965 PMID:27924033

30. Zhang Q, Liu W, Zhang HM, Xie GY, Miao YR, Xia M, Guo AY. hTFtarget: A Comprehensive Database for Regulations of Human Transcription Factors and Their Targets. Genomics Proteomics Bioinformatics. 2020; 18:120–8.
https://doi.org/10.1016/j.gpb.2019.09.006 PMID:32858223

31. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol. 2010; 5:1315–6.
https://doi.org/10.1097/JTO.0b013e3181ec173d PMID:20736804

32. Vincent A, Herman J, Schulick R, Hruban RH, Goggins M. Pancreatic cancer. Lancet. 2011; 378:607–20.
https://doi.org/10.1016/S0140-6736(10)62307-0 PMID:21620466

33. Kim JE, Lee KT, Lee JK, Paik SW, Rhee JC, Choi KW. Clinical usefulness of carbohydrate antigen 19-9 as a screening test for pancreatic cancer in an asymptomatic population. J Gastroenterol Hepatol. 2004; 19:182–6.
https://doi.org/10.1111/j.1440-1746.2004.03219.x PMID:14731128

34. Unger K, Mehta KY, Kaur P, Wang Y, Menon SS, Jain SK, Moonjelly RA, Suman S, Datta K, Singh R, Fogel P, Cheema AK. Metabolomics based predictive classifier for early detection of pancreatic ductal adenocarcinoma. Oncotarget. 2018; 9:23078–90.
https://doi.org/10.18632/oncotarget.25212 PMID:29796173

35. Fahrmann JF, Bantis LE, Capello M, Scelo G, Dennison JB, Patel N, Murage E, Vykoukal J, Kundnani DL, Foretova L, Fabianova E, Holcatova I, Janout V, et al. A Plasma-Derived Protein-Metabolite Multiplexed Panel for Early-Stage Pancreatic Cancer. J Natl Cancer Inst. 2019; 111:372–9.
https://doi.org/10.1093/jnci/djy126 PMID:30137376

36. Zhao F, Li Z, Dong Z, Wang Z, Guo P, Zhang D, Li S. Exploring the Potential of Exosome-Related LncRNA Pairs as Predictors for Immune Microenvironment, Survival Outcome, and Microbiotain Landscape in Esophageal Squamous Cell Carcinoma. Front Immunol. 2022; 13:918154.
https://doi.org/10.3389/fimmu.2022.918154 PMID:35880180

37. Chen W, Zhang F, Xu H, Hou X, Tang D. Prospective Analysis of Proteins Carried in Extracellular Vesicles with Clinical Outcome in Hepatocellular Carcinoma. Curr Genomics. 2022; 23:109–17.
https://doi.org/10.2174/1389202923666622030412545 8 PMID:36778976

38. Wang J, Wu X, Xu J, Liao Y, Deng M, Wang X, Li J. Differential expression and bioinformatics analysis of exosome circRNAs in pancreatic ductal adenocarcinoma. Transl Oncol. 2023; 33:101686.
https://doi.org/10.1016/j.tranon.2023.101686 PMID:37182508

39. Zeng J, Tan H, Huang B, Zhou Q, Ke Q, Dai Y, Tang J, Xu B, Feng J, Yu L. Lipid metabolism characterization in gastric cancer identifies signatures to predict prognostic and therapeutic responses. Front Genet. 2022; 13:959170.
https://doi.org/10.3389/fgene.2022.959170 PMID:36406121

40. Wang W, Zhang C, Yu Q, Zheng X, Yin C, Yan X, Liu G, Song Z. Development of a novel lipid metabolism-based risk score model in hepatocellular carcinoma patients. BMC Gastroenterol. 2021; 21:68.
https://doi.org/10.1186/s12876-021-01638-3 PMID:33579192

41. Yang J, Wei X, Hu F, Dong W, Sun L. Development and validation of a novel 3-gene prognostic model for pancreatic adenocarcinoma based on ferroptosis-related genes. Cancer Cell Int. 2022; 22:21.
https://doi.org/10.1186/s12935-021-02431-8 PMID:35033072

42. Xu D, Wang Y, Liu X, Zhou K, Wu J, Chen J, Chen C, Chen L, Zheng J. Development and clinical validation of a novel 9-gene prognostic model based on multi-omics in pancreatic adenocarcinoma. Pharmacol Res. 2021; 164:105370.
https://doi.org/10.1016/j.phrs.2020.105370 PMID:33316381

43. Su K, Peng Y, Yu H. Development of a Prognostic Model Based on Pyroptosis-Related Genes in Pancreatic Adenocarcinoma. Dis Markers. 2022; 2022:9141117.
https://doi.org/10.1155/2022/9141117 PMID:35677632

44. Xiang H, Yang R, Tu J, Xi Y, Yang S, Lv L, Zhai X, Zhu Y, Dong D, Tao X. Metabolic reprogramming of immune cells in pancreatic cancer progression. Biomed Pharmacother. 2023; 157:113992.

https://doi.org/10.1016/j.biopha.2022.113992
PMID:36395610

45. Fioroni I, Dell'Aquila E, Pantano F, Intagliata S, Caricato M, Vincenzi B, Coppola R, Santini D, Tonini G. Role of c-mesenchymal-epithelial transition pathway in gastric cancer. Expert Opin Pharmacother. 2015; 16:1195–207.
https://doi.org/10.1517/14656566.2015.1037739
PMID:25881479

46. Kanda T, Jiang X, Yokosuka O. Androgen receptor signaling in hepatocellular carcinoma and pancreatic cancers. World J Gastroenterol. 2014; 20:9229–36.
https://doi.org/10.3748/wjg.v20.i28.9229
PMID:25071315

47. González-Mariscal L, Miranda J, Gallego-Gutiérrez H, Cano-Cortina M, Amaya E. Relationship between apical junction proteins, gene expression and cancer. Biochim Biophys Acta Biomembr. 2020; 1862:183278.
https://doi.org/10.1016/j.bbamem.2020.183278
PMID:32240623

48. Zhao Z, Ju Q, Ji J, Li Y, Zhao Y. N6-Methyladenosine Methylation Regulator RBM15 is a Potential Prognostic Biomarker and Promotes Cell Proliferation in Pancreatic Adenocarcinoma. Front Mol Biosci. 2022; 9:842833.
https://doi.org/10.3389/fmolb.2022.842833
PMID:35223996

49. Murphy KJ, Zhu J, Trpceski M, Pereira BA, Timpson P, Herrmann D. Focal adhesion kinase priming in pancreatic cancer, altering biomechanics to improve chemotherapy. Biochem Soc Trans. 2022; 50:1129–41.
https://doi.org/10.1042/BST20220162 PMID:35929603

50. Klose T, Abiatari I, Samkharadze T, De Oliveira T, Jäger C, Kiladze M, Valkovskaya N, Friess H, Michalski CW, Kleeff J. The actin binding protein destrin is associated with growth and perineural invasion of pancreatic cancer. Pancreatology. 2012; 12:350–7.
https://doi.org/10.1016/j.pan.2012.05.012
PMID:22898637

51. Lee Y, Finch-Edmondson M, Cognart H, Zhu B, Song H, Low BC, Sudol M. Common and Unique Transcription Signatures of YAP and TAZ in Gastric Cancer Cells. Cancers (Basel). 2020; 12:3667.
https://doi.org/10.3390/cancers12123667
PMID:33297432

52. Xiong D, Chen D, Liu D, Wu W, Dou X, Ji X, Li J, Zhang X. The Overexpression of NMHC IIA Promoted Invasion and Metastasis of Nasopharyngeal Carcinoma Cells. J Cancer. 2021; 12:4218–28.
https://doi.org/10.7150/jca.47506 PMID:34093822

53. Zhao Y, Li X, Zhang H, Yan M, Jia M, Zhou Q. A Transcriptome Sequencing Study on Genome-Wide Gene Expression Differences of Lung Cancer Cells Modulated by Fucoidan. Front Bioeng Biotechnol. 2022; 10:844924.
https://doi.org/10.3389/fbioe.2022.844924
PMID:35299642

54. Cervenkova L, Vycital O, Bruha J, Rosendorf J, Palek R, Liska V, Daum O, Mohelnikova-Duchonova B, Soucek P. Protein expression of ABCC2 and SLC22A3 associates with prognosis of pancreatic adenocarcinoma. Sci Rep. 2019; 9:19782.
https://doi.org/10.1038/s41598-019-56059-w
PMID:31874997

55. Mohelnikova-Duchonova B, Strouhal O, Hughes DJ, Holcatova I, Oliverius M, Kala Z, Campa D, Rizzato C, Canzian F, Pezzilli R, Talar-Wojnarowska R, Malecka-Panas E, Sperti C, et al. SLC22A3 polymorphisms do not modify pancreatic cancer risk, but may influence overall patient survival. Sci Rep. 2017; 7:43812.
https://doi.org/10.1038/srep43812
PMID:28272475

56. Zhou YY, Chen LP, Zhang Y, Hu SK, Dong ZJ, Wu M, Chen QX, Zhuang ZZ, Du XJ. Integrated transcriptomic analysis reveals hub genes involved in diagnosis and prognosis of pancreatic cancer. Mol Med. 2019; 25:47.
https://doi.org/10.1186/s10020-019-0113-2
PMID:31706267

57. Yin M, Zhou L, Ji Y, Lu R, Ji W, Jiang G, Ma J, Song X. In silico identification and verification of ferroptosis-related genes in type 2 diabetic islets. Front Endocrinol (Lausanne). 2022; 13:946492.
https://doi.org/10.3389/fendo.2022.946492
PMID:35992146

58. Bibi F, Naseer MI, Alvi SA, Yasir M, Jiman-Fatani AA, Sawan A, Abuzenadah AM, Al-Qahtani MH, Azhar EI. microRNA analysis of gastric cancer patients from Saudi Arabian population. BMC Genomics. 2016 (Suppl 9); 17:751.
https://doi.org/10.1186/s12864-016-3090-7
PMID:27766962

59. Lu M, Li J, Fan X, Xie F, Fan J, Xiong Y. Novel Immune-Related Ferroptosis Signature in Esophageal Cancer: An Informatics Exploration of Biological Processes Related to the TMEM161B-AS1/hsa-miR-27a-3p/GCH1 Regulatory Network. Front Genet. 2022; 13:829384.
https://doi.org/10.3389/fgene.2022.829384
PMID:35281840

60. Horak J, Kubecek O, Siskova A, Honkova K, Chvojkova I, Krupova M, Manethova M, Vodenkova S, García-Mulero S, John S, Cecka F, Vodickova L, Petera J, et al. Differences in genome, transcriptome, miRNAome, and methylome in synchronous and metachronous liver metastasis of colorectal cancer. Front Oncol. 2023; 13:1133598.

https://doi.org/10.3389/fonc.2023.1133598
PMID:37182133

61. Koschmieder S, Halmos B, Levantini E, Tenen DG. Dysregulation of the C/EBPalpha differentiation pathway in human cancer. J Clin Oncol. 2009; 27:619–28.
https://doi.org/10.1200/JCO.2008.17.9812
PMID:19075268

62. Yamamoto K, Tateishi K, Kudo Y, Sato T, Yamamoto S, Miyabayashi K, Matsusaka K, Asaoka Y, Ijichi H, Hirata Y, Otsuka M, Nakai Y, Isayama H, et al. Loss of histone demethylase KDM6B enhances aggressiveness of pancreatic cancer through downregulation of C/EBPα. Carcinogenesis. 2014; 35:2404–14.
https://doi.org/10.1093/carcin/bgu136
PMID:24947179

63. Ambudkar SV, Kimchi-Sarfaty C, Sauna ZE, Gottesman MM. P-glycoprotein: from genomics to mechanism. Oncogene. 2003; 22:7468–85.
https://doi.org/10.1038/sj.onc.1206948
PMID:14576852

64. Katoh M. Multi-layered prevention and treatment of chronic inflammation, organ fibrosis and cancer associated with canonical WNT/β-catenin signaling activation (Review). Int J Mol Med. 2018; 42:713–25.
https://doi.org/10.3892/ijmm.2018.3689
PMID:29786110

65. Harbuzariu A, Rampoldi A, Daley-Brown DS, Candelaria P, Harmon TL, Lipsey CC, Beech DJ, Quarshie A, Ilies GO, Gonzalez-Perez RR. Leptin-Notch signaling axis is involved in pancreatic cancer progression. Oncotarget. 2017; 8:7740–52.
https://doi.org/10.18632/oncotarget.13946
PMID:27999190

66. Lu Z, Kleeff J, Shrikhande S, Zimmermann T, Korc M, Friess H, Büchler MW. Expression of the multidrug-resistance 1 (MDR1) gene and prognosis in human pancreatic cancer. Pancreas. 2000; 21:240–7.
https://doi.org/10.1097/00006676-200010000-00004
PMID:11039467

67. Modi A, Roy D, Sharma S, Vishnoi JR, Pareek P, Elhence P, Sharma P, Purohit P. ABC transporters in breast cancer: their roles in multidrug resistance and beyond. J Drug Target. 2022; 30:927–47.
https://doi.org/10.1080/1061186X.2022.2091578
PMID:35758271

68. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R, Stenroos ES, Chandrasekharappa S, Athanassiadou A, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science. 1997; 276:2045–7.

https://doi.org/10.1126/science.276.5321.2045
PMID:9197268

69. Burré J, Sharma M, Südhof TC. Cell Biology and Pathophysiology of α-Synuclein. Cold Spring Harb Perspect Med. 2018; 8:a024091.
https://doi.org/10.1101/cshperspect.a024091
PMID:28108534

70. Sturchio A, Dwivedi AK, Vizcarra JA, Chirra M, Keeling EG, Mata IF, Kauffman MA, Pandey MK, Roviello G, Comi C, Versino M, Marsili L, Espay AJ. Genetic parkinsonisms and cancer: a systematic review and meta-analysis. Rev Neurosci. 2020; 32:159–67.
https://doi.org/10.1515/revneuro-2020-0083
PMID:33151182

71. Alizadeh-Sedigh M, Fazeli MS, Mahmoodzadeh H, Sharif SB, Teimoori-Toolabi L. Methylation of FBN1, SPG20, ITF2, RUNX3, SNCA, MLH1, and SEPT9 genes in circulating cell-free DNA as biomarkers of colorectal cancer. Cancer Biomark. 2022; 34:221–50.
https://doi.org/10.3233/CBM-210315 PMID:34957998

72. Bianchini M, Giambelluca M, Scavuzzo MC, Di Franco G, Guadagni S, Palmeri M, Furbetta N, Gianardi D, Costa A, Gentiluomo M, Gaeta R, Pollina LE, Falcone A, et al. In Pancreatic Adenocarcinoma Alpha-Synuclein Increases and Marks Peri-Neural Infiltration. Int J Mol Sci. 2022; 23:3775.
https://doi.org/10.3390/ijms23073775
PMID:35409135

73. Field J, Vojtek A, Ballester R, Bolger G, Colicelli J, Ferguson K, Gerst J, Kataoka T, Michaeli T, Powers S. Cloning and characterization of CAP, the S. cerevisiae gene encoding the 70 kd adenylyl cyclase-associated protein. Cell. 1990; 61:319–27.
https://doi.org/10.1016/0092-8674(90)90812-s
PMID:2184942

74. Ono S. The role of cyclase-associated protein in regulating actin filament dynamics - more than a monomer-sequestration factor. J Cell Sci. 2013; 126:3249–58.
https://doi.org/10.1242/jcs.128231 PMID:23908377

75. Yu G, Swiston J, Young D. Comparison of human CAP and CAP2, homologs of the yeast adenylyl cyclase-associated proteins. J Cell Sci. 1994; 107:1671–8.
https://doi.org/10.1242/jcs.107.6.1671
PMID:7962207

76. Yamazaki K, Takamura M, Masugi Y, Mori T, Du W, Hibi T, Hiraoka N, Ohta T, Ohki M, Hirohashi S, Sakamoto M. Adenylate cyclase-associated protein 1 overexpressed in pancreatic cancers is involved in cancer cell motility. Lab Invest. 2009; 89:425–32.
https://doi.org/10.1038/labinvest.2009.5
PMID:19188911

77. Liu Y, Cui X, Hu B, Lu C, Huang X, Cai J, He S, Lv L, Cong X, Liu G, Zhang Y, Ni R. Upregulated expression of CAP1 is associated with tumor migration and metastasis in hepatocellular carcinoma. Pathol Res Pract. 2014; 210:169–75.
https://doi.org/10.1016/j.prp.2013.11.011
PMID:24359721

78. Li M, Yang X, Shi H, Ren H, Chen X, Zhang S, Zhu J, Zhang J. Downregulated expression of the cyclase-associated protein 1 (CAP1) reduces migration in esophageal squamous cell carcinoma. Jpn J Clin Oncol. 2013; 43:856–64.
https://doi.org/10.1093/jjco/hyt093 PMID:23904342

79. Yu XF, Ni QC, Chen JP, Xu JF, Jiang Y, Yang SY, Ma J, Gu XL, Wang H, Wang YY. Knocking down the expression of adenylate cyclase-associated protein 1 inhibits the proliferation and migration of breast cancer cells. Exp Mol Pathol. 2014; 96:188–94.
https://doi.org/10.1016/j.yexmp.2014.02.002
PMID:24509166

80. Tan M, Song X, Zhang G, Peng A, Li X, Li M, Liu Y, Wang C. Overexpression of adenylate cyclase-associated protein 1 is associated with metastasis of lung cancer. Oncol Rep. 2013; 30:1639–44.
https://doi.org/10.3892/or.2013.2607 PMID:23842884

81. Wu H, Hasan R, Zhang H, Gray J, Williams D, Miller M, Allen F, Lee V, Kelly T, Zhou GL. Phosphorylation Regulates CAP1 (Cyclase-Associated Protein 1) Functions in the Motility and Invasion of Pancreatic Cancer Cells. Sci Rep. 2019; 9:4925.
https://doi.org/10.1038/s41598-019-41346-3
PMID:30894654

82. Zhou GL, Zhang H, Field J. Mammalian CAP (Cyclase-associated protein) in the world of cell migration: Roles in actin filament dynamics and beyond. Cell Adh Migr. 2014; 8:55–9.
https://doi.org/10.4161/cam.27479 PMID:24429384

83. Zhang H, Ghai P, Wu H, Wang C, Field J, Zhou GL. Mammalian adenylyl cyclase-associated protein 1 (CAP1) regulates cofilin function, the actin cytoskeleton, and cell adhesion. J Biol Chem. 2013; 288:20966–77.
https://doi.org/10.1074/jbc.M113.484535
PMID:23737525

84. Zhang H, Zhou GL. CAP1 (Cyclase-Associated Protein 1) Exerts Distinct Functions in the Proliferation and Metastatic Potential of Breast Cancer Cells Mediated by ERK. Sci Rep. 2016; 6:25933.
https://doi.org/10.1038/srep25933 PMID:27173014

85. Tontonoz P, Spiegelman BM. Fat and beyond: the diverse biology of PPARgamma. Annu Rev Biochem. 2008; 77:289–312.
https://doi.org/10.1146/annurev.biochem.77.061307.091829 PMID:18518822

86. Fujisawa T, Sugiyama M, Tomimoto A, Wada K, Endo H, Takahashi H, Yoneda K, Yoneda M, Inamori M, Saito S, Terauchi Y, Kadowaki T, Tsuchiya N, et al. Inhibition of peroxisome proliferator-activated receptor gamma promotes tumorigenesis through activation of the beta-catenin / T cell factor (TCF) pathway in the mouse intestine. J Pharmacol Sci. 2008; 108:535–44.
https://doi.org/10.1254/jphs.08193fp PMID:19075513

87. McAlpine CA, Barak Y, Matise I, Cormier RT. Intestinal-specific PPARgamma deficiency enhances tumorigenesis in ApcMin/+ mice. Int J Cancer. 2006; 119:2339–46.
https://doi.org/10.1002/ijc.22115 PMID:16858678

88. Mirza AZ, Althagafi II, Shamshad H. Role of PPAR receptor in different diseases and their ligands: Physiological importance and clinical implications. Eur J Med Chem. 2019; 166:502–13.
https://doi.org/10.1016/j.ejmech.2019.01.067
PMID:30739829

89. Galbraith L, Leung HY, Ahmad I. Lipid pathway deregulation in advanced prostate cancer. Pharmacol Res. 2018; 131:177–84.
https://doi.org/10.1016/j.phrs.2018.02.022
PMID:29466694

90. Tate T, Xiang T, Wobker SE, Zhou M, Chen X, Kim H, Batourina E, Lin CS, Kim WY, Lu C, Mckiernan JM, Mendelsohn CL. Pparg signaling controls bladder cancer subtype and immune exclusion. Nat Commun. 2021; 12:6160.
https://doi.org/10.1038/s41467-021-26421-6
PMID:34697317

91. Wilson HE, Stanton DA, Rellick S, Geldenhuys W, Pistilli EE. Breast cancer-associated skeletal muscle mitochondrial dysfunction and lipid accumulation is reversed by PPARG. Am J Physiol Cell Physiol. 2021; 320:C577–90.
https://doi.org/10.1152/ajpcell.00264.2020
PMID:33439777

92. Villa ALP, Parra RS, Feitosa MR, Camargo HP, Machado VF, Tirapelli DPDC, Rocha JJ, Feres O. PPARG expression in colorectal cancer and its association with staging and clinical evolution. Acta Cir Bras. 2020; 35:e202000708.
https://doi.org/10.1590/s0102-865020200070000008
PMID:32813759

93. Pazienza V, Tavano F, Benegiamo G, Vinciguerra M, Burbaci FP, Copetti M, di Mola FF, Andriulli A, di Sebastiano P. Correlations among PPARγ, DNMT1, and DNMT3B Expression Levels and Pancreatic Cancer. PPAR Res. 2012; 2012:461784.
https://doi.org/10.1155/2012/461784 PMID:22919364

94. Hynes NE, Lane HA. ERBB receptors and cancer: the complexity of targeted inhibitors. Nat Rev Cancer. 2005; 5:341–54.
https://doi.org/10.1038/nrc1609 PMID:15864276

95. Scaltriti M, Baselga J. The epidermal growth factor receptor pathway: a model for targeted therapy. Clin Cancer Res. 2006; 12:5268–72.
https://doi.org/10.1158/1078-0432.CCR-05-1554 PMID:17000658

96. Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. Nat Rev Mol Cell Biol. 2001; 2:127–37.
https://doi.org/10.1038/35052073 PMID:11252954

97. Butti R, Das S, Gunasekaran VP, Yadav AS, Kumar D, Kundu GC. Receptor tyrosine kinases (RTKs) in breast cancer: signaling, therapeutic implications and challenges. Mol Cancer. 2018; 17:34.
https://doi.org/10.1186/s12943-018-0797-x PMID:29455658

98. Liao BC, Lee JH, Lin CC, Chen YF, Chang CH, Ho CC, Shih JY, Yu CJ, Yang JC. Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors for Non-Small-Cell Lung Cancer Patients with Leptomeningeal Carcinomatosis. J Thorac Oncol. 2015; 10:1754–61.
https://doi.org/10.1097/JTO.0000000000000669 PMID:26334749

99. Evangelopoulos ME, Weis J, Krüttgen A. Signalling pathways leading to neuroblastoma differentiation after serum withdrawal: HDL blocks neuroblastoma differentiation by inhibition of EGFR. Oncogene. 2005; 24:3309–18.
https://doi.org/10.1038/sj.onc.1208494 PMID:15735700

100. Lee H, Hwang Y, Radermacher R, Chun HH. Performance investigation of multi-stage saturation cycle with natural working fluids and low GWP working fluids. International Journal of Refrigeration. 2015; 51:103–11.
https://doi.org/10.1016/j.ijrefrig.2014.12.018

101. Nicholson RI, Gee JM, Harper ME. EGFR and cancer prognosis. Eur J Cancer. 2001 (Suppl 4); 37:S9–15.
https://doi.org/10.1016/s0959-8049(01)00231-3 PMID:11597399

102. Talukdar S, Emdad L, Das SK, Fisher PB. EGFR: An essential receptor tyrosine kinase-regulator of cancer stem cells. Adv Cancer Res. 2020; 147:161–88.
https://doi.org/10.1016/bs.acr.2020.04.003 PMID:32593400

103. Yamanaka Y, Friess H, Kobrin MS, Buchler M, Beger HG, Korc M. Coexpression of epidermal growth factor receptor and ligands in human pancreatic cancer is associated with enhanced tumor aggressiveness. Anticancer Res. 1993; 13:565–9.
PMID:8317885

104. Uegaki K, Nio Y, Inoue Y, Minari Y, Sato Y, Song MM, Dong M, Tamura K. Clinicopathological significance of epidermal growth factor and its receptor in human pancreatic cancer. Anticancer Res. 1997; 17:3841–7.
PMID:9427790

105. Ye Y, Chen Z, Shen Y, Qin Y, Wang H. Development and validation of a four-lipid metabolism gene signature for diagnosis of pancreatic cancer. FEBS Open Bio. 2021; 11:3153–70.
https://doi.org/10.1002/2211-5463.13074 PMID:33386701

106. Zhu K, Xiaoqiang L, Deng W, Wang G, Fu B. Development and validation of a novel lipid metabolism-related gene prognostic signature and candidate drugs for patients with bladder cancer. Lipids Health Dis. 2021; 20:146.
https://doi.org/10.1186/s12944-021-01554-1 PMID:34706720
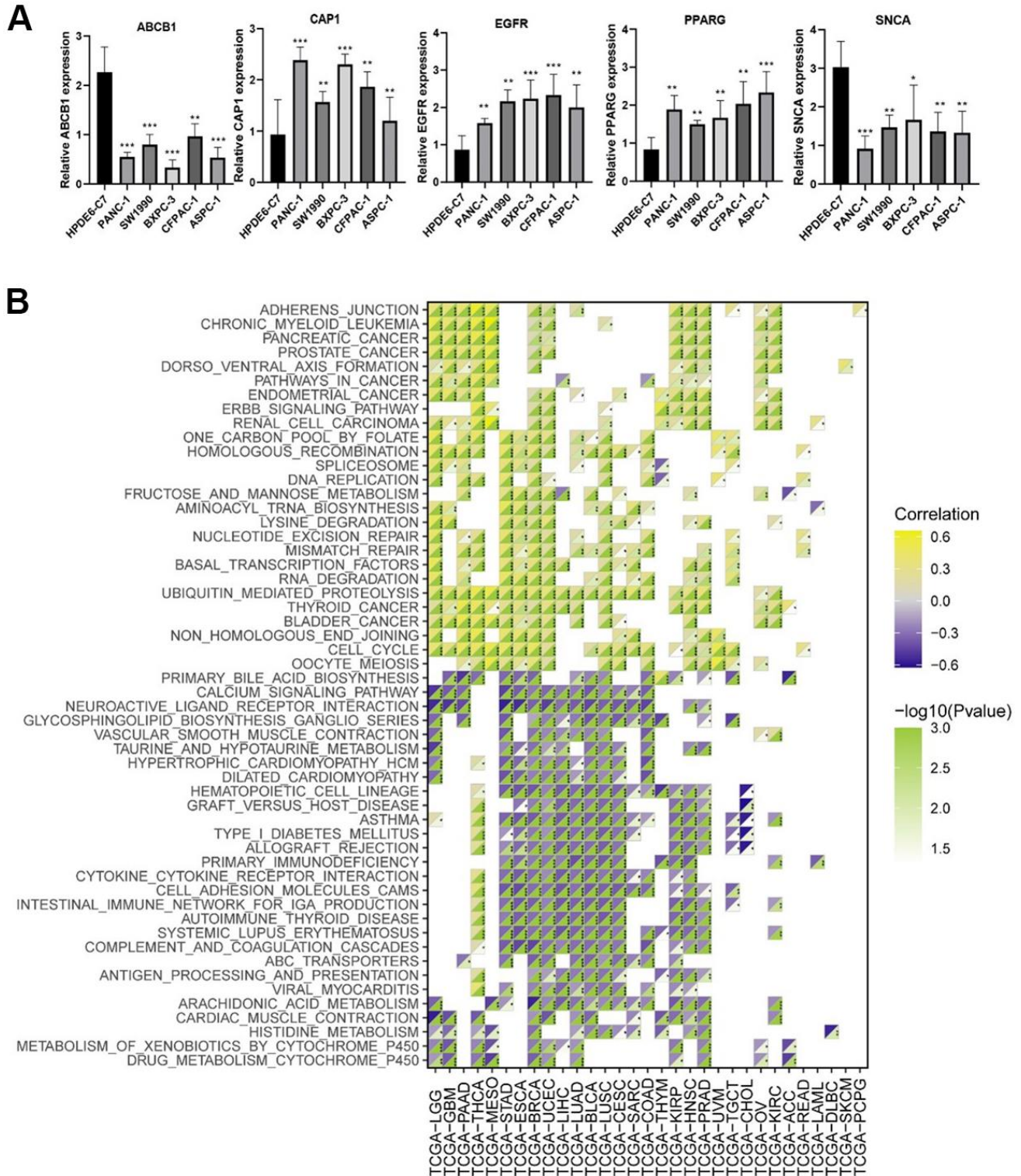
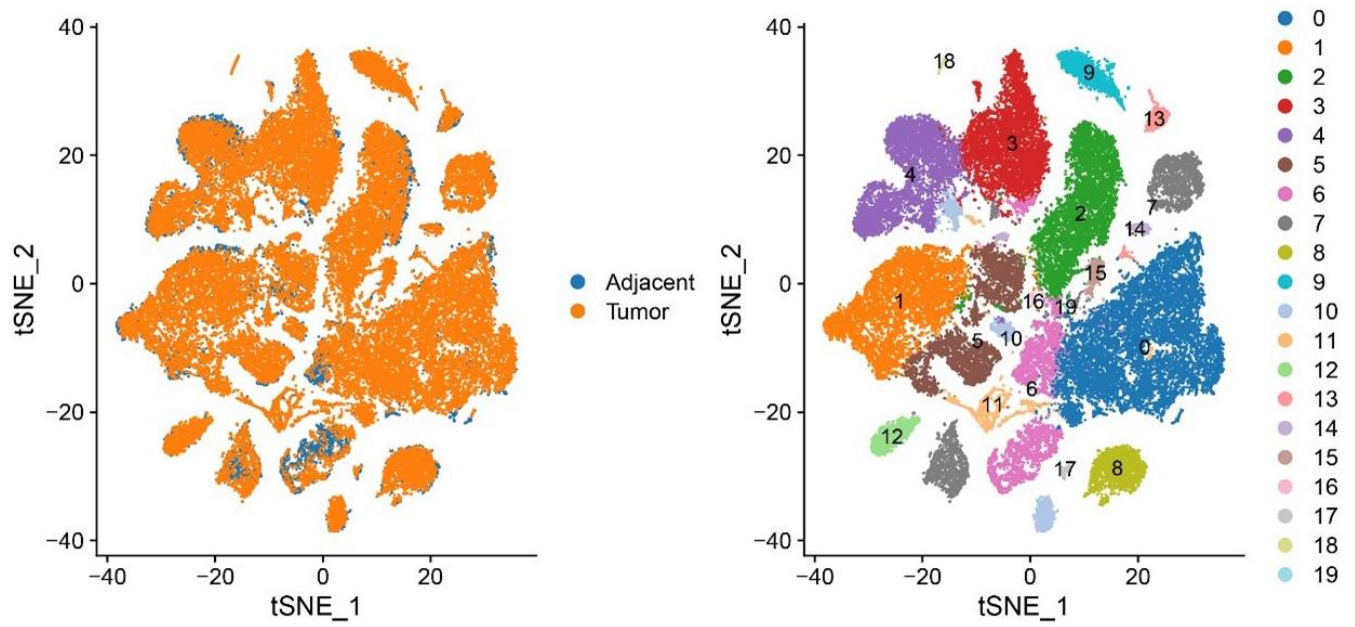**Supplementary Figure 1.** (**A**) The expression of the 5 hub genes in pancreatic cancer cell lines comparing to pancreatic normal cell lines. ** means P-value <0.01, *** means P-value <0.001. (**B**) Correlation analysis of GSVA enrichment score and identified risk score in pan-cancer.
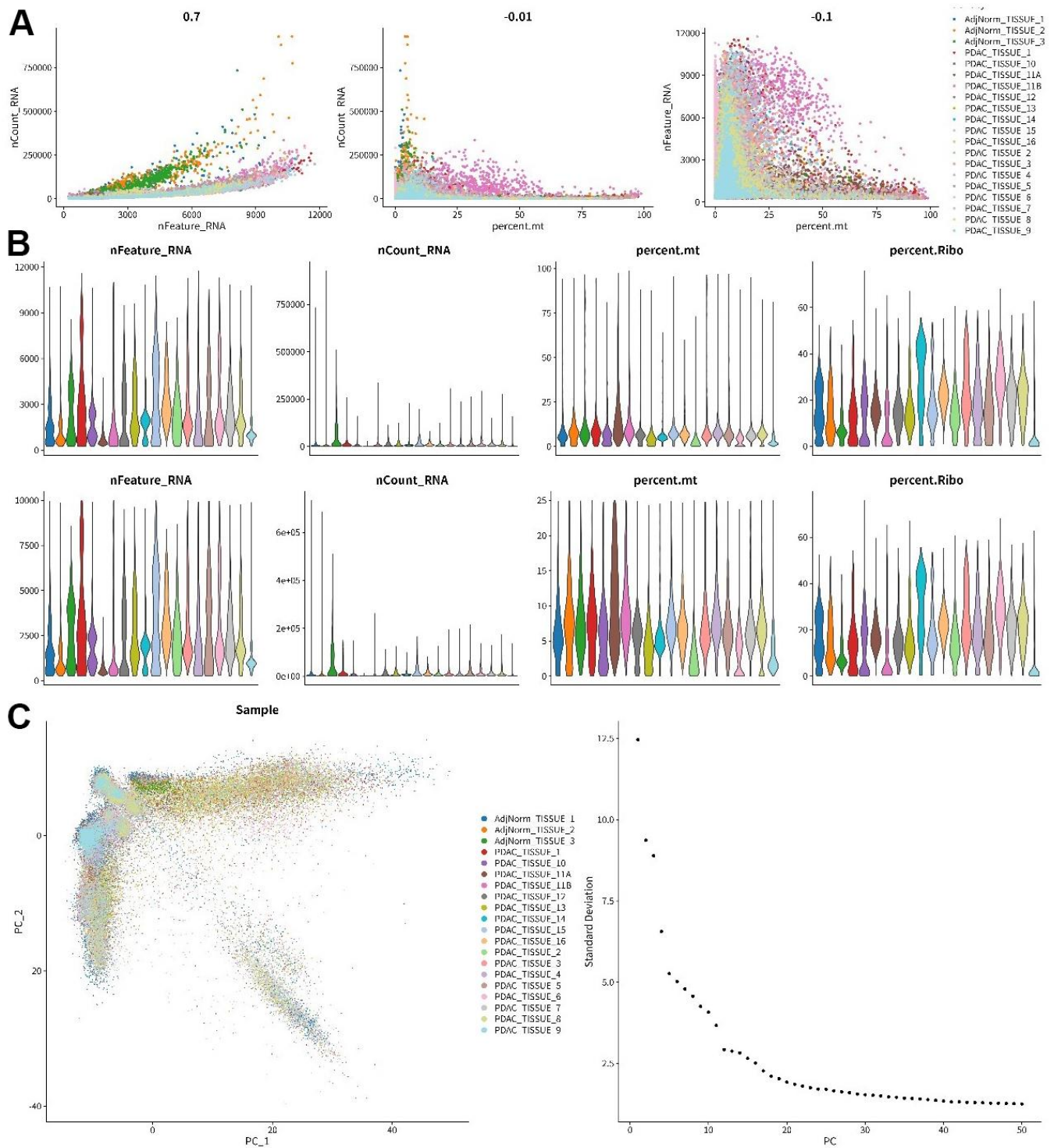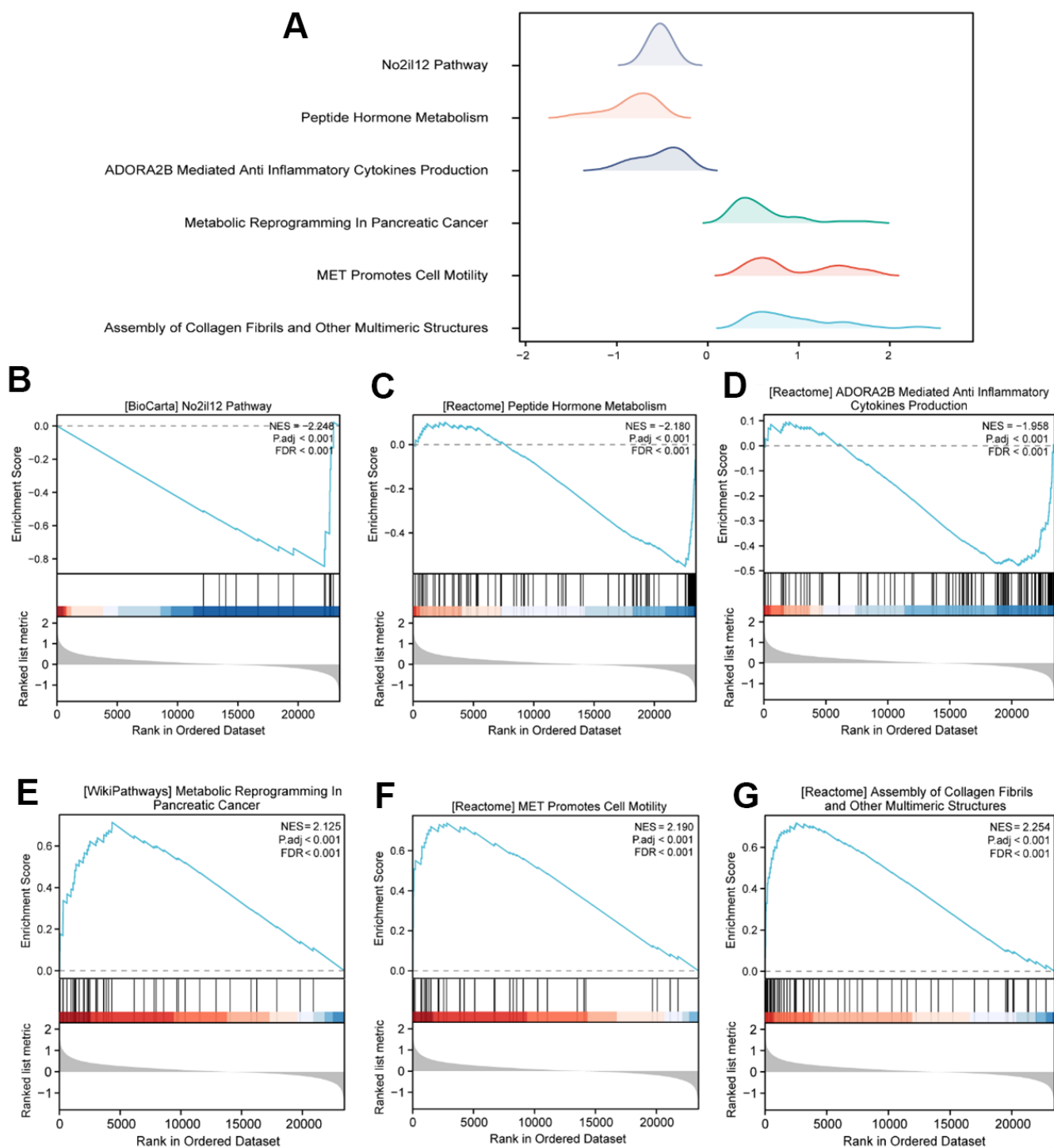
**Supplementary Figure 2.** (**A**–**C**) The quality of the enrolled single-cell sequencing data.

Supplementary Figure 3. Left: The tSNE plot between tumor and adjacent tumor. Right: The tSNE plot of 19 group after clustered.
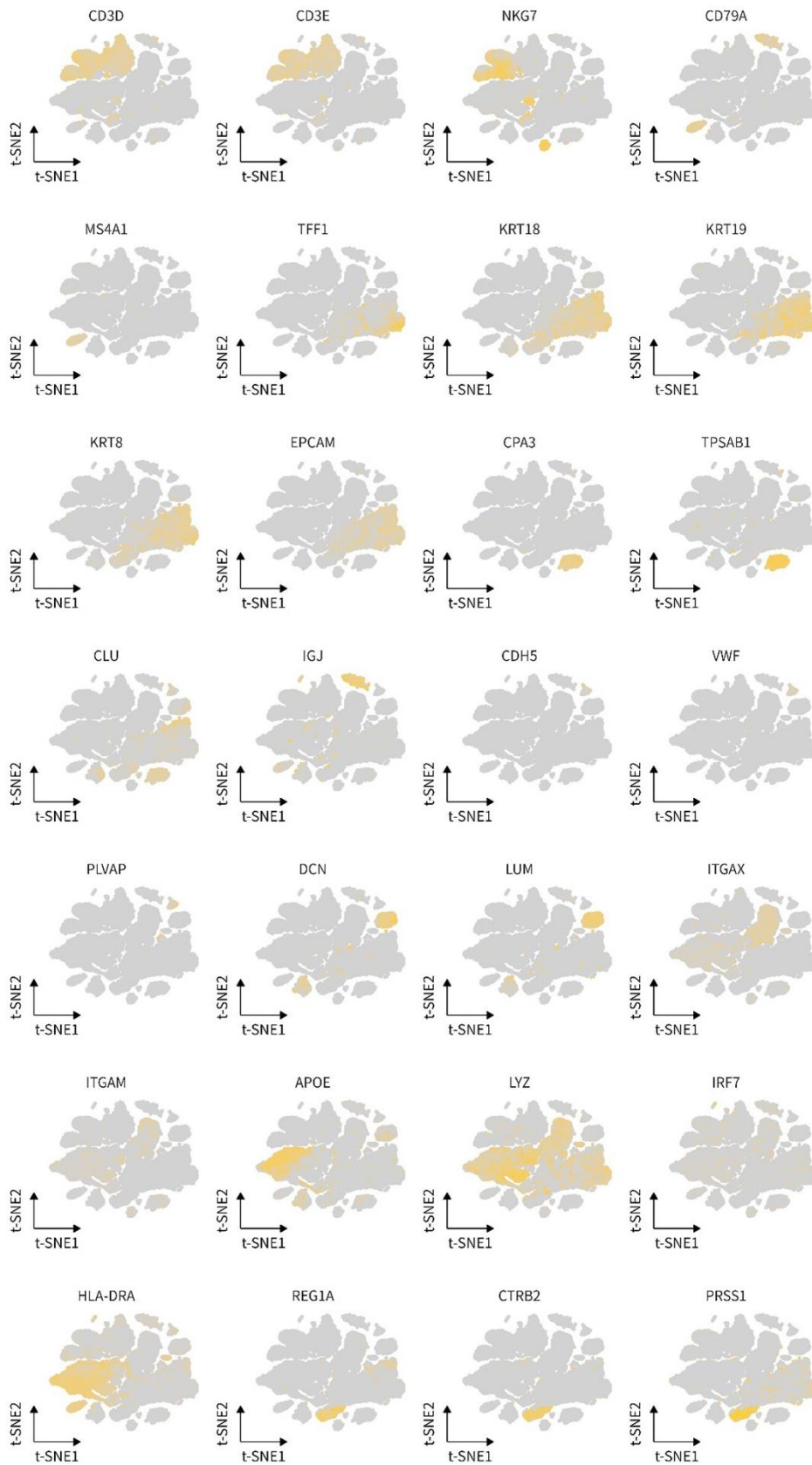
Supplementary Figure 4. The t-SNE plot of multiple markers in single-cell sequencing data.

Supplementary Figure 5. (**A**) The identified GSEA pathways between the high and low-risk group. (**B**–**G**) The enrichment score of identified GSEA pathway.
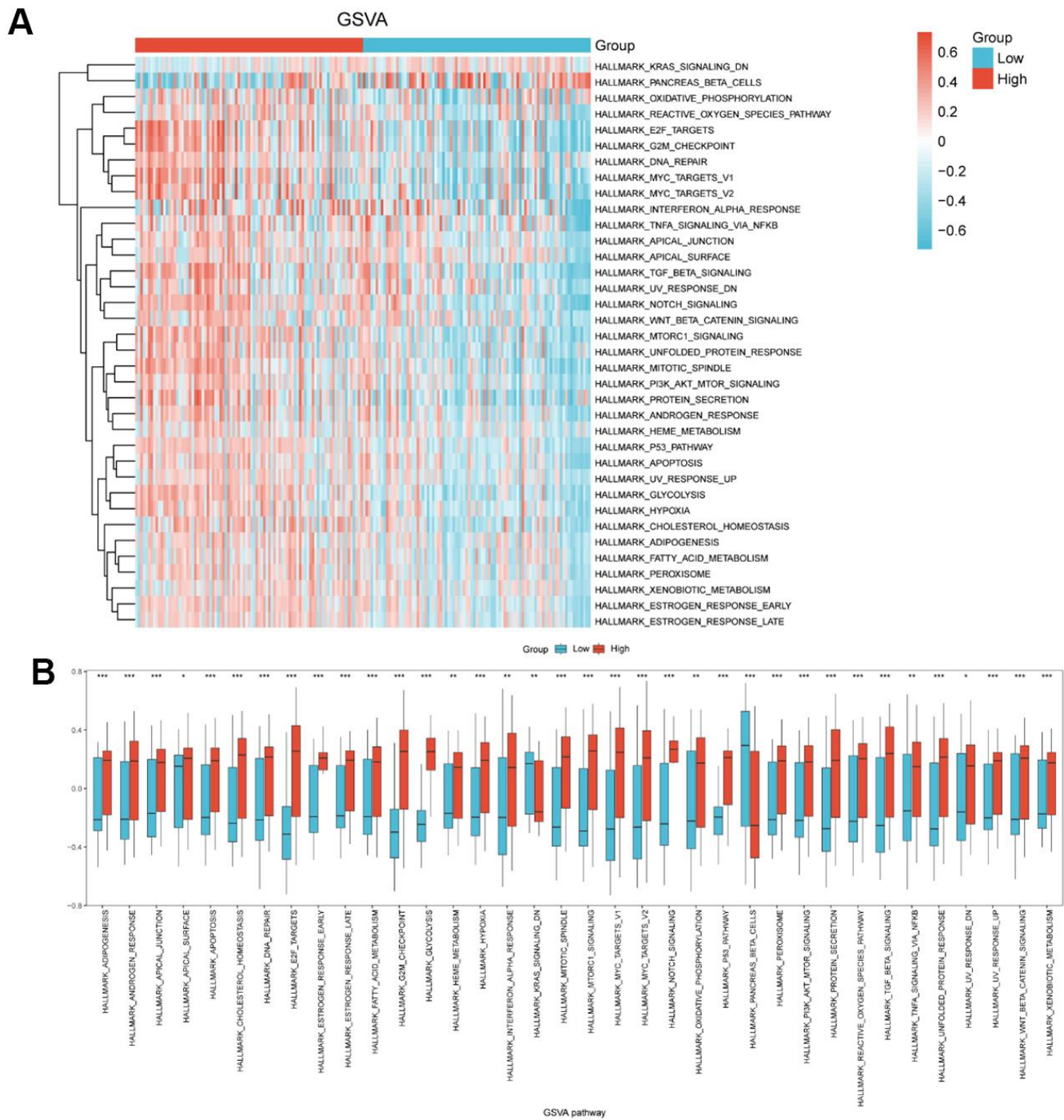
Supplementary Figure 6. (**A**) The identified GSVA pathways in the high and low-risk groups. (**B**) The identified GSEA pathways in the high and low-risk groups.

Supplementary Tables

Supplementary Table 1. Exosome and lipid metabolism related genes, EALMRGs.

EALMRGs					
HRAS	AKT1	ACTB	ALB	CLU	FASN
RAF1	EGFR	TP53	GGT1	CCL2	ABCB1
RAB7A	STAT3	IL1B	SOD1	PPARG	ITGB3
BAG6	ICAM1	TGFB1	PRNP	TNF	ACE
PTEN	APP	GRB2	CTNNB1	SIRT1	IDH1
TLR4	ASAH2	SNCA	HSPA5	HSPA1B	APOA1
DPP4	ITGB1	HSPG2	VDAC1	HIF1A	F2
APOE	CDC42	HSPA1A	CAP1	VIM	NAMPT
APOA2	MTREX	MAPK1	MAPK3		

Supplementary Table 2. GSEA enrichment analysis results of hub genes.

ID	SetSize	Enrichment score	NES	P-value	P.adj	qvalue
REACTOME_ASSEMBLY_OF_COLLAGEN_FIBRILS_AND_OTHER_MULTIMERIC_STRUCTURES	61	0.718364	2.253987	2.31E-10	1.75E-08	9.89E-09
REACTOME_MET_PROMOTES_CELL_MOTILITY	41	0.738372	2.189964	6.95E-08	2.45E-06	1.38E-06
WP_METABOLIC_REPROGRAMMING_IN_PANCREATIC_CANCER	42	0.714381	2.125378	1.14E-06	2.55E-05	1.44E-05
PID_ECADHERIN_STABILIZATION_PATHWAY	41	0.711503	2.110273	7.96E-07	1.86E-05	1.05E-05
REACTOME_SIGNALING_BY_TGF_BETA_RECEPTOR_COMPLEX	92	0.606184	2.026585	1.52E-07	4.70E-06	2.66E-06
WP_AEROBIC_GLYCOLYSIS	11	0.892959	1.968318	7.83E-06	0.000105	5.94E-05
REACTOME_CELLULAR_RESPONSE_TO_HYPOXIA	74	0.599001	1.937214	7.64E-06	0.000104	5.89E-05
REACTOME_TP53_REGULATES_TRANSCRIPTION_OF_CELL_DEATH_GENES	44	0.643168	1.934143	4.23E-05	0.000416	0.000235
REACTOME_REGULATION_OF_MRNA_STABILITY_BY_PROTEINS_THAT_BIND_AU_RICH_ELEMENTS	86	0.586933	1.933052	1.52E-06	3.19E-05	1.80E-05
REACTOME_AUF1_HNRNP_D0_BINDS_AND_DESTABILIZES_MRNA	54	0.622465	1.926354	1.03E-05	0.000129	7.27E-05
REACTOME_METABOLISM_OF_POLYAMINES	58	0.609515	1.905315	1.49E-05	0.000172	9.73E-05
REACTOME_CLASS_B_2_SECRETIN_FAMILY_RECEPTORS	89	-0.36262	-1.43144	0.006947	0.023977	0.013554
BIOCARTA_IL12_PATHWAY	19	-0.56542	-1.59812	0.007225	0.024532	0.013867
WP_SUDDEN_INFANT_DEATH_SYNDROME_SIDS_SUSCEPTIBILITY_PATHWAYS	154	-0.37325	-1.60073	0.000166	0.00126	0.000712
REACTOME_CLASS_C_3_METABOTROPIC_GLUTAMATE_PHEROMONE_RECEPTORS	23	-0.55704	-1.63056	0.007794	0.026038	0.014719
BIOCARTA_IL17_PATHWAY	13	-0.65349	-1.68422	0.013016	0.038538	0.021785
REACTOME_ANTI_INFLAMMATORY_RESPONSE_FAVOURING_LEISHMANIA_PARASITE_INFECTION	204	-0.38274	-1.69503	3.68E-06	5.94E-05	3.36E-05

REACTOME_GLUCAGON_SIGNALING_IN_METABOLIC_REGULATION	32	-0.56229	-1.81894	0.001201	0.006002	0.003393
REACTOME_METABOLISM_OF_ANGIOTENSINOGEN_TO_ANGIOTENSINS	18	-0.68755	-1.90406	0.001362	0.006673	0.003772
REACTOME_INTEGRATION_OF_ENERGY_METABOLISM	107	-0.47311	-1.92199	1.56E-06	3.19E-05	1.80E-05
WP_FATTY_ACID_OMEGAOXIDATION	13	-0.7473	-1.92599	0.000916	0.004851	0.002742
REACTOME_ADORA2B_MEDIATED_ANTI_INFLAMMATORY_CYTOKINES_PRODUCTION	114	-0.4792	-1.95779	5.92E-07	1.48E-05	8.36E-06
REACTOME_PEPTIDE_HORMONE_METABOLISM	89	-0.55222	-2.17985	4.94E-08	1.84E-06	1.04E-06
BIOCARTA_NO2IL12_PATHWAY	15	-0.84594	-2.24776	1.62E-06	3.28E-05	1.86E-05

GSEA, Gene Set Enrichment Analysis.

Supplementary Table 3. Gene set variation analysis.

logFC	AveExpr	t	P-value	P.adj	B
GLYCOLYSIS	0.36444594	-0.01523	9.863187	5.07E-19	2.53E-17
MTORC1_SIGNALING	0.357089084	-0.01789	8.325416	1.19E-14	2.25E-13
NOTCH_SIGNALING	0.347591157	0.004941	8.305719	1.35E-14	2.25E-13
G2M_CHECKPOINT	0.385941475	-0.0264	8.018213	8.12E-14	1.02E-12
MITOTIC_SPINDLE	0.328281017	-0.01179	7.940877	1.31E-13	1.31E-12
P53_PATHWAY	0.262456174	-0.01229	7.614802	9.57E-13	7.97E-12
MYC_TARGETS_V1	0.367066553	-0.02282	7.18631	1.22E-11	8.71E-11
E2F_TARGETS	0.37467454	-0.0379	7.129268	1.70E-11	1.06E-10
ESTROGEN_RESPONSE_EARLY	0.234150989	0.007557	6.531819	5.05E-10	2.80E-09
UNFOLDED_PROTEIN_RESPONSE	0.286734015	-0.02842	6.427516	8.95E-10	4.48E-09
TGF_BETA_SIGNALING	0.294572698	-0.00919	6.17579	3.49E-09	1.59E-08
HYPOXIA	0.241525103	-0.00795	6.123298	4.61E-09	1.92E-08
ESTROGEN_RESPONSE_LATE	0.21123882	0.001622	5.921306	1.33E-08	5.10E-08
CHOLESTEROL_HOMEOSTASIS	0.256458022	-0.0232	5.907857	1.43E-08	5.10E-08
MYC_TARGETS_V2	0.313551213	-0.03854	5.832338	2.11E-08	7.03E-08
PROTEIN_SECRETION	0.285422891	-0.03426	5.619062	6.22E-08	1.87E-07
DNA_REPAIR	0.237067433	-0.01739	5.614501	6.37E-08	1.87E-07
PI3K_AKT_MTOR_SIGNALING	0.223725184	-0.01771	5.53459	9.49E-08	2.63E-07
PEROXISOME	0.200951847	-0.0095	5.209335	4.61E-07	1.21E-06
APOPTOSIS	0.203206043	-0.00695	5.133401	6.60E-07	1.65E-06
UV_RESPONSE_UP	0.182326528	-0.00785	5.016017	1.14E-06	2.72E-06
ANDROGEN_RESPONSE	0.208896672	-0.0156	4.817842	2.82E-06	6.41E-06
APICAL_JUNCTION	0.172333537	0.005362	4.373193	1.95E-05	4.24E-05
ADIPOGENESIS	0.16377501	-0.00324	4.248083	3.27E-05	6.82E-05
FATTY_ACID_METABOLISM	0.167904487	-0.01793	4.233237	3.48E-05	6.96E-05
PANCREAS_BETA_CELLS	-0.254195748	0.008054	-4.19493	4.07E-05	7.82E-05
REACTIVE_OXYGEN_SPECIES_PATHWAY	0.181776847	-0.01654	4.106457	5.81E-05	0.000108
WNT_BETA_CATENIN_SIGNALING	0.163912122	-0.00018	3.825307	0.000173	0.00031
XENOBIOTIC_METABOLISM	0.120265442	0.003998	3.231334	0.001436	0.002476
INTERFERON_ALPHA_RESPONSE	0.178409452	-0.00718	3.183146	0.001684	0.002807
OXIDATIVE_PHOSPHORYLATION	0.152020094	-0.02721	2.923599	0.00385	0.006209
TNFA_SIGNALING_VIA_NFKB	0.13788571	0.004791	2.859715	0.00468	0.007312

HEME_METABOLISM	0.102568627	-0.01779	2.756254	0.006375	0.009659
UV_RESPONSE_DN	0.120433127	-0.00831	2.615587	0.009571	0.014075
KRAS_SIGNALING_DN	-0.088942579	0.016323	-2.53184	0.012098	0.017282
APICAL_SURFACE	0.09332276	0.029442	2.227793	0.026983	0.037476

GSVA, Gene Set Variation Analysis.

Supplementary Table 4. mRNA-miRNA interaction network nodes.

mRNA	miRNA	mRNA	miRNA
ABCB1	*hsa-miR-12136*	*PPARG*	*hsa-miR-548h-5p*
ABCB1	*hsa-miR-4282*	*PPARG*	*hsa-miR-548au-5p*
CAP1	*hsa-miR-1260a*	*PPARG*	*hsa-miR-548d-5p*
CAP1	*hsa-miR-1260b*	*PPARG*	*hsa-miR-548y*
CAP1	*hsa-miR-448*	*PPARG*	*hsa-miR-548ae-5p*
CAP1	*hsa-miR-205-3p*	*PPARG*	*hsa-miR-548ad-5p*
CAP1	*hsa-miR-153-3p*	*PPARG*	*hsa-miR-548aq-5p*
CAP1	*hsa-miR-6508-5p*	*PPARG*	*hsa-miR-548b-5p*
CAP1	*hsa-miR-548b-3p*	*PPARG*	*hsa-miR-548c-5p*
CAP1	*hsa-miR-2355-3p*	*PPARG*	*hsa-miR-548n*
CAP1	*hsa-miR-3123*	*PPARG*	*hsa-miR-559*
EGFR	*hsa-miR-141-5p*	*PPARG*	*hsa-miR-548ak*
EGFR	*hsa-miR-6878-5p*	*PPARG*	*hsa-miR-548bb-5p*
EGFR	*hsa-miR-514a-3p*	*PPARG*	*hsa-miR-548j-5p*
EGFR	*hsa-miR-4533*	*PPARG*	*hsa-miR-3606-3p*
EGFR	*hsa-miR-514b-3p*	*PPARG*	*hsa-miR-9985*
EGFR	*hsa-miR-6867-5p*	*PPARG*	*hsa-miR-27a-3p*
EGFR	*hsa-miR-7157-3p*	*PPARG*	*hsa-miR-513c-3p*
EGFR	*hsa-miR-6737-3p*	*PPARG*	*hsa-miR-27b-3p*
EGFR	*hsa-miR-7110-3p*	*PPARG*	*hsa-miR-513a-3p*
EGFR	*hsa-miR-9985*	*PPARG*	*hsa-miR-3617-3p*
EGFR	*hsa-miR-27a-3p*	*PPARG*	*hsa-miR-5694*
EGFR	*hsa-miR-27b-3p*	*PPARG*	*hsa-miR-454-3p*
PPARG	*hsa-miR-548am-5p*	*PPARG*	*hsa-miR-190a-3p*
PPARG	*hsa-miR-548ay-5p*	*SNCA*	*hsa-miR-5011-5p*
PPARG	*hsa-miR-548o-5p*	*SNCA*	*hsa-miR-6515-3p*
PPARG	*hsa-miR-548ab*	*SNCA*	*hsa-miR-106a-3p*
PPARG	*hsa-miR-548ap-5p*	*SNCA*	*hsa-miR-1178-3p*
PPARG	*hsa-miR-548w*	*SNCA*	*hsa-miR-4786-3p*
PPARG	*hsa-miR-548as-5p*	*SNCA*	*hsa-miR-7-5p*
PPARG	*hsa-miR-548i*	*SNCA*	*hsa-miR-450b-5p*
PPARG	*hsa-miR-548a-5p*	*SNCA*	*hsa-miR-7843-3p*
PPARG	*hsa-miR-548ar-5p*	*SNCA*	

Supplementary Table 5. mRNA-TF interaction network nodes.

mRNA	TF	mRNA	TF
ABCB1	*CEBPA*	*CAP1*	*RUNX3*
ABCB1	*CREB1*	*CAP1*	*SMARCA4*
ABCB1	*ERG*	*CAP1*	*SPI1*
ABCB1	*ETS1*	*CAP1*	*SPIB*
ABCB1	*FLI1*	*CAP1*	*SRF*
ABCB1	*GATA2*	*CAP1*	*TFAP4*
ABCB1	*MECOM*	*CAP1*	*CEBPA*
CAP1	*E2F1*	*CAP1*	*ZNF384*
CAP1	*EP300*	*CAP1*	*CREB1*
CAP1	*ERG*	*CAP1*	*CREBBP*
CAP1	*ESRRA*	*CAP1*	*CTCF*
CAP1	*ETS1*	*EGFR*	*EP300*
CAP1	*FLI1*	*EGFR*	*KLF9*
CAP1	*FOXP1*	*EGFR*	*MAX*
CAP1	*GABPA*	*EGFR*	*ONECUT1*
CAP1	*HDAC1*	*EGFR*	*PGR*
CAP1	*HDAC2*	*EGFR*	*BRD4*
CAP1	*IRF1*	*EGFR*	*TFAP2A*
CAP1	*BCL11A*	*EGFR*	*TFAP2C*
CAP1	*KLF1*	*EGFR*	*TFAP4*
CAP1	*KLF4*	*EGFR*	*ZBTB7A*
CAP1	*KLF5*	*EGFR*	*ZNF263*
CAP1	*BCL6*	*EGFR*	*AR*
CAP1	*LMO2*	*EGFR*	*E2F6*
CAP1	*MAX*	*PPARG*	*GABPA*
CAP1	*MAZ*	*PPARG*	*BRD2*
CAP1	*MYB*	*PPARG*	*BRD4*
CAP1	*NFIC*	*PPARG*	*CEBPA*
CAP1	*BRD2*	*PPARG*	*CTCF*
CAP1	*NR2F2*	*PPARG*	*ESR1*
CAP1	*BRD4*	*SNCA*	*CDK9*
CAP1	*PBX3*	*SNCA*	*FLI1*
CAP1	*POLR2A*	*SNCA*	*HDAC2*
CAP1	*RUNX1T1*		

TF, Transcription Factor.