

The value of a metabolic reprogramming-related gene signature for pancreatic adenocarcinoma prognosis prediction

Zhen Tan^{1,2,3,4,*}, Yubin Lei^{1,2,3,4,*}, Jin Xu^{1,2,3,4,*}, Si Shi^{1,2,3,4}, Jie Hua^{1,2,3,4}, Bo Zhang^{1,2,3,4}, Qingcai Meng^{1,2,3,4}, Jiang Liu^{1,2,3,4}, Yiyin Zhang^{1,2,3,4}, Miaoyan Wei^{1,2,3,4}, Xianjun Yu^{1,2,3,4}, Chen Liang^{1,2,3,4}

¹Department of Pancreatic Surgery, Fudan University Shanghai Cancer Center, Shanghai 200032, China

²Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China

³Shanghai Pancreatic Cancer Institute, Shanghai 200032, China

⁴Pancreatic Cancer Institute, Fudan University, Shanghai 200032, China

*Equal contribution

Correspondence to: Xianjun Yu, Chen Liang; email: yuxianjun@fudanpci.org, liangchen@fudanpci.org

Keywords: pancreatic adenocarcinoma, metabolic reprogramming

Received: June 3, 2020

Accepted: September 9, 2020

Published: November 20, 2020

Copyright: © 2020 Tan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Pancreatic ductal adenocarcinoma (PDAC) is one of the most fatal malignancies worldwide. Extensive enhancement of glycolysis and reprogramming of lipid metabolism are both associated with the development and progression of PDAC. Previous studies have suggested that various gene signatures could convey prognostic information about PDAC. However, the use of these signatures has some limitations, perhaps because of a lack of knowledge regarding the genetic and energy supply backgrounds of PDAC. Therefore, we conducted multi-mRNA analysis based on metabolic reprogramming to identify novel signatures for accurate prognosis prediction in PDAC patients. In this study, a three-gene signature comprising MET, ENO3 and CD36 was established to predict the overall survival of PDAC patients. The three-gene signature could divide patients into high- and low-risk groups by disparities in overall survival verified by log-rank test in two independent validation cohorts and could differentiate tumors from normal tissues with excellent accuracy in four Gene Expression Omnibus (GEO) cohorts. We also found a positive correlation between the risk score of the gene signature and inherited germline mutations in PDAC predisposition genes. A glycolysis and lipid metabolism-based gene nomogram and corresponding calibration curves showed significant performance for survival prediction in the TCGA-PDAC dataset. The high-risk designation was closely connected with oncological signatures and multiple aggressiveness-related pathways, as determined by gene set enrichment analysis (GSEA). In summary, our study developed a three-gene signature and established a prognostic nomogram that objectively predicted overall survival in PDAC. The findings could provide a reference for the prediction of overall survival and could aid in individualized management for PDAC patients.

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) is one of the most invasive solid malignancies and could become the second leading source of cancer-related deaths in the United States. Surgical resection, which is the only established therapy, significantly enhances the five-year survival rate to 20–30%. However, fewer than 20% of

all PDAC patients are eligible for resection since most patients are diagnosed with advanced-stage disease featuring metastasis. However, patients with PDAC in the same TNM stage can differ in overall survival (OS), perhaps because of the complex desmoplastic microenvironment and the different molecular characteristics of PDAC. Thus, new strategies and better signatures are needed to help predict prognosis

and subsequently improve individualized treatment for PDAC patients.

Recent advances in gene chips and high-throughput next-generation sequencing technology have modified the transcriptomic research landscape and demonstrated that mRNA prognostic signatures can help to predict OS in PDAC [1]. Numerous bioinformatics analyses have been conducted to elucidate molecular mechanisms and direct clinical practice [2]. Raman et al. developed a five-gene prognostic model that accurately predicted OS from a PDAC dataset in The Cancer Genome Atlas (TCGA) (the TCGA-PDAC dataset) using Cox proportional hazards regression analysis [3]. Similarly, Yan et al. revealed a four-gene signature (with LYRM1, KNTC1, IGF2BP2, and CDC6) that is significantly related to the progression of pancreatic cancer through the same method [4]. However, this method is not suitable for high-dimensional microarray data because of the limitation of overfitting. The least absolute shrinkage and selection operator (LASSO) regression method does not have this class of limitation and has been broadly used for optimized selection of genes [5]. Notably, PDAC is characterized by an incredibly nutrient-deficient and hypoxic environment caused by vascular disturbances, desmoplastic reactions and unrestrained growth. Oncogenic activation of Kirsten rat sarcoma 2 viral oncogene homolog (KRAS) and mutations in other tumor suppressor genes promote abnormal mitochondrial metabolism and enhance glycolysis and lipid metabolism since these genes are intrinsic components of metabolic plasticity [6]. Reprogramming of metabolic pathways, including enhancement of lipid metabolism and glycolysis, has emerged as a hallmark of cancer [7, 8]. The most common transformation is enhancement of glycolysis, which enables vigorous growth of cells by generating a large variety of substrates and facilitates invasion and migration by affecting glycolytic enzymes to improve the supply of ATP [9, 10]. Biosynthesis, glycosylation and redox homeostasis are also linked to intermediates related to glycolysis. Overexpression of glycolytic enzymes and increased lactate production in PDAC cells promote metastatic colonization of distant organs and tumor angiogenesis by regulating the invasion-metastasis cascade [11]. Fatty acids (FAs) and glycerol are key molecules released by adipocytes and are the main precursors of the lipids used as energy sources by cancer cells. As a result, cancer-associated adipocytes play a role in providing energy for cancer cells since they supply sufficient FAs and lipids to affect the metabolism of tumor cells and promote the growth of malignant tumors. FAs are also used by cells to synthesize membranes and to generate signaling molecules that stimulate the proliferation and invasion of cancer cells. Previous studies have sought to delineate correlations between lipid remodeling and the

progression of invasive malignant biological behavior in PDAC. Lipid breakdown and fibrotic changes in the tumor microenvironment enhance the levels of FAs.

In this study, we performed a systematic and comprehensive gene signature discovery and validation effort, with consideration of various components of metabolic reprogramming, to develop a multi-gene model for the robust prediction of the prognoses of PDAC patients. A prognostic nomogram consisting of the identified gene signature and clinical prognostic factors was also created for prediction of OS. The molecular mechanism and inherited germline mutations of the gene signature were also investigated. All of these findings could provide a reference for improved prediction of OS and could aid in the individualized management of PDAC patients.

RESULTS

Identification of DEGs from public datasets

The results of this study are summarized in a flow chart (Figure 1). The detailed information in three eligible PDAC datasets in the GEO database (GSE15471, GSE16515, and GSE32676) met our criteria. Analysis of these PDAC datasets revealed that 237 differentially expressed genes (DEGs) were shared among the 3 series of comparisons between tumor and adjacent paracancerous tissues in the Venn and UpSet diagrams; these DEGs were regarded as credible DEGs (Figure 2A, 2B). In addition, KEGG pathway enrichment analyses were performed for these overlapping up- or down-regulated genes. As shown in Figure 2C, the DEGs were most enriched in the ECM-receptor interaction pathway, the PI3K-Akt signaling pathway and numerous energy metabolism-related pathways, such as the proteoglycans in cancer pathway and the central carbon metabolism in cancer pathway.

Prognostic signature construction

Based on these 237 DEGs and clinical features from the TCGA database, 100 genes significantly related to OS time ($P < 0.05$) were identified by the log-rank test, as shown in Figure 3A. Twelve glycolysis and lipid metabolism-related genes significantly associated with survival of PDAC were screened (Table 1). To avoid overfitting during the subsequent model construction, LASSO regression was performed to build a prognostic signature, which included twelve genes from the 237 previously identified DEGs. The LASSO coefficient profiles of the twelve genes are presented in Figure 3B. A formula to calculate the risk score for PDAC was derived based on individual three-gene expression levels weighted by regression coefficients, as follows:

$$\text{Risk score} = (1.0167070 \times \text{expression level of MET}) \\ - (0.4313794 \times \text{expression level of ENO3}) \\ - (0.1764747 \times \text{expression level of CD36}).$$

With the median risk score as the cutoff, PDAC patients were classified into a low-risk group (n = 89) and a high-risk group (n = 88) (Figure 3C). Survival analysis confirmed that the survival times of the low risk patients were distinctly longer than those of the high-risk patients in the TCGA discovery cohort (P < 0.0001; Figure 3D). The finding was subsequently validated in ICGC validation datasets; the results obtained for these datasets agreed with those obtained for the discovery

cohort (P = 0.038; Figure 3F). The same trend was also verified in the FUSCC validation cohort (P = 0.017; Figure 3H). Furthermore, to evaluate the performance of the three-gene signature in predicting the prognoses of PDAC patients, time-dependent ROC curves with respect to the 1-year, 3-year and 5-year survival rates were constructed for the TCGA, ICGC and FUSCC datasets, and the area under the curve (AUC) values for these curves were evaluated (Figure 3E, 3G, 3I). The AUC at 1, 3, and 5 years was 0.709, 0.749, and 0.836 in the TCGA set, respectively. The AUC at 1, 3, and 5 years was 0.644, 0.621, and 0.955 in the ICGC set, respectively. The AUC at 1 and 3 years was 0.601, 0.745 in the FUSCC set, respectively.

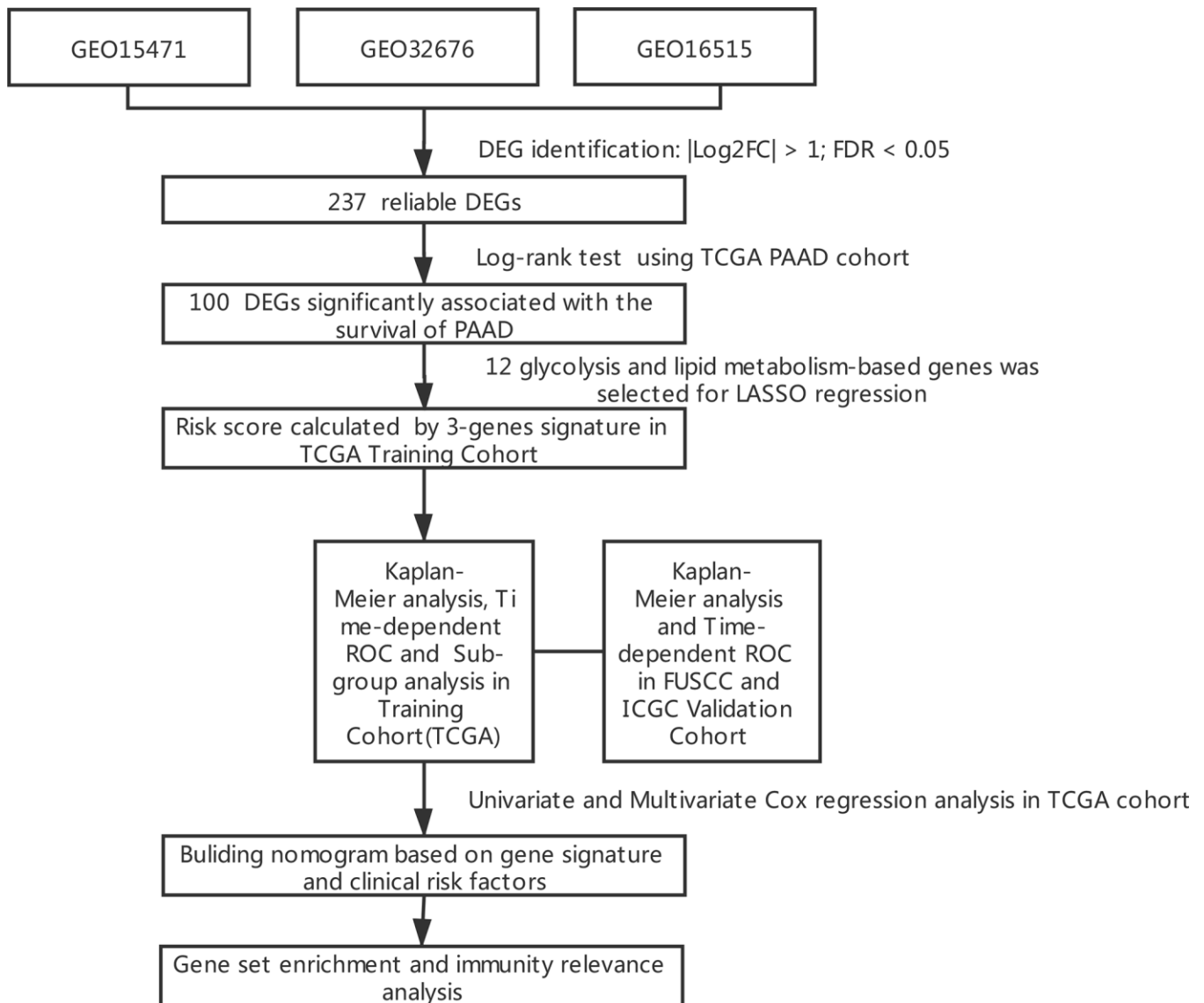


Figure 1. Flowchart presenting the process of establishing the gene signature in this study.

Validation of the expression of and alterations in the three genes

The mRNA expression levels of the three genes were validated using the Gene Expression Profiling Interactive Analysis (GEPIA) database and GEO. In the GEPIA dataset, only MET mRNA expression was significantly elevated in tumor samples (Figure 4A–4C). In the GEO dataset, we found that MET expression levels were significantly higher in PDAC tumor tissue than in paracancerous tissue in the GEO32676 and GEO15471 datasets (Figure 4D). In contrast, the mRNA expression of ENO3 and CD36 was markedly down-regulated in tumor tissues (Figure 4E, 4F). Protein expression of the DEGs in pancreatic tumor and non-tumor tissues was evaluated by the Human Protein Atlas (<https://www.proteinatlas.org/>) [12]. Immunohistochemical (IHC) staining information for the three genes in PDAC from the Human Protein Atlas database is shown in Figure 4G–4I. Moreover, the OS times were also obtained from the GEPIA database (Figure 4J–4L). A total of 848 patients in four cohorts (the ICGC, Queensland Centre for Medical Genomics [QCMG], TCGA and UT Southwestern Medical Center [UTSW] cohorts) were included in this study for analysis of the mutation information of the three glycolysis- and lipid metabolism-related genes. Overall, amplification and deep deletion were the most

common types of mutations in MET, CD36 and ENO3 (Figure 4M).

Validation of the risk score formula in the GEO and TCGA cohorts

ROC curve analysis was performed with AUCs to evaluate the usefulness of the risk score of the gene signature in distinguishing PDAC tissues from control samples in four GEO cohorts (Figure 5A). Tumor tissues could be reliably identified based on this risk score. Subgroup analyses were performed to assess the three-gene signature and clinical characteristics of pancreatic cancer (including American Joint Committee on Cancer [AJCC] stage and grade) from appropriate datasets. In terms of AJCC stage and grade, stage T3-4 and grade 3-4 patients had higher risk scores than stage T1-2 and grade 1-2 patients. However, the same trend was not noted for N stages (Figure 5B–5D). The Kaplan-Meier curves of these clinical features are shown in Figure 5E–5H. The survival time of the high-risk score group was significantly shorter than that of the low-risk score group for all pathological T stages and for grades 1-2. In terms of mutations, risk scores were observed to be highly correlated with the mutation states of key genes in PDAC. The risk scores for the CDKN2A, KRAS and TP53 mutant groups were significantly higher than that of the wild-type group (Figure 5I–5L).

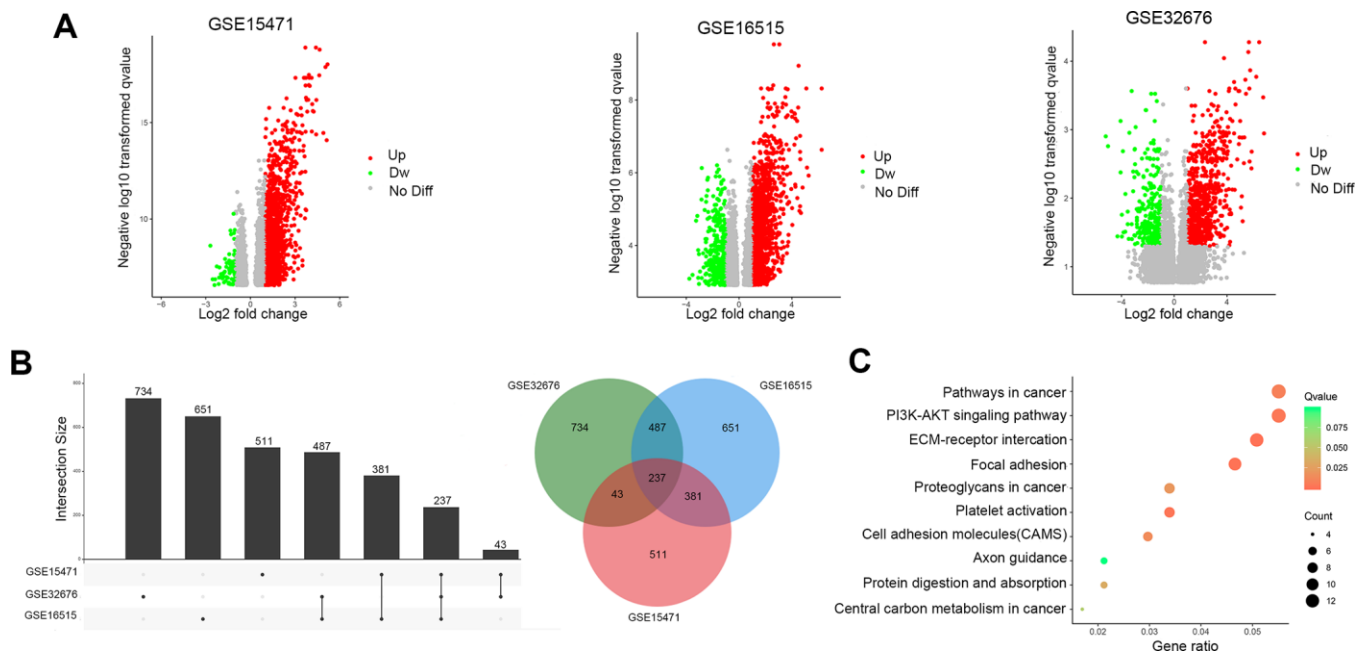


Figure 2. Identification of DEGs in pancreatic cancer between tumor and paracancerous tissues. (A) Volcano plots of DEGs in the 3 indicated datasets. (X-axis: log₂(FC); Y-axis: -log₁₀(FDR) for each gene. Genes with FDR < 0.01 and FC > 1 or < -1 were considered as DEGs in each series. Blue: down-regulated genes; Gray: non-differential genes; Red: up-regulated genes). (B) Upset Venn diagrams of the DEGs identified in 3 GEO datasets. (C) Top 10 enriched KEGG pathways of the DEGs.

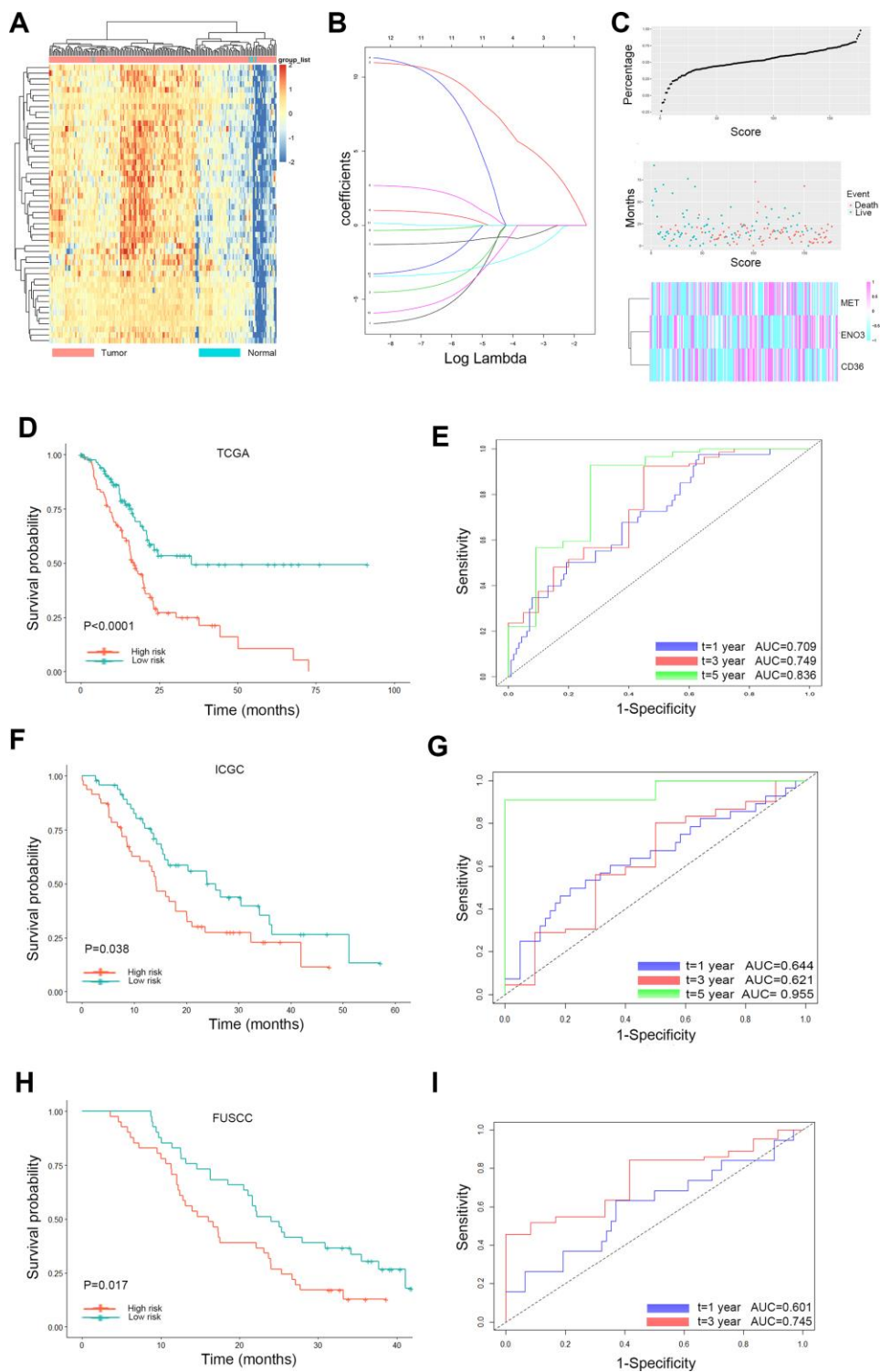


Figure 3. Lasso analysis and Kaplan-Meier curve for the patients in the TCGA, ICGC and FUSCC cohorts. (A) Representative heatmap of the DEGs significantly related to OS time identified by the log-rank test in the TCGA cohort. **(B)** LASSO coefficient profiles of the 12 glycolysis and lipid metabolism-based genes. LASSO, least absolute shrinkage and selection operator method. **(C)** Risk score analysis of the differentially expressed DEG signatures of PDAC. Risk scores of DEG signatures (top); survival status and duration of cases (middle); low-score and high-score groups for the three genes (bottom). **(D)** The Kaplan-Meier plot (low risk vs. high risk PDAC cases) of 5-year overall survival in the TCGA cohort. **(E)** Time-dependent ROC analyses at 1, 3, and 5 years in the TCGA cohort. **(F)** The Kaplan-Meier plot (low risk vs. high risk PDAC cases) of 5-year overall survival in the ICGC cohort. **(G)** Time dependent ROC analyses at 1, 3, and 5 years in the ICGC cohort. **(H)** The Kaplan-Meier plot (low risk vs. high risk PDAC cases) of 5-year overall survival in the FUSCC cohort. **(I)** Time dependent ROC analyses at 1, 3, and 5 years in the FUSCC cohort.

Table 1. The survival information of the 12 glycolysis and lipid metabolism-based genes.

gene	Overall Survival		Disease Free Survival	
	HR	Log-rank p	HR	Log-rank p
PKM	1.7	0.0097	1.6	0.046
MET	2.2	0.00023	2.2	0.00044
HK2	1.5	0.053	1.5	0.084
LDHA	2.2	0.0002	2.1	0.00064
ENO3	0.65	0.045	0.64	0.042
LDLR	1.7	0.012	1.4	0.15
CD36	0.56	0.0053	0.64	0.046
FAS	1.7	0.013	1.2	0.48
SDC1	1.6	0.018	1.8	0.011
SDC4	1.7	0.015	1.8	0.0083
ITPR3	1.3	0.21	1.6	0.042
ITGB1	1.6	0.029	1.8	0.013

Correlations between the three-gene signature and clinical characteristics

We examined the association of the three-gene signature (risk score) with clinical features in PDAC patients using univariate and multivariate Cox proportional hazard regression analyses. Univariate Cox proportional hazard regression showed that N stage, age, and risk score could predict poor survival of PDAC patients, as shown in Table 2 ($P < 0.05$). In addition, multivariate Cox proportional hazard regression showed that the risk score ($P < 0.001$) was an independent prognostic indicator of PDAC (Figure 6A). Based on the results acquired from multivariate Cox regression of OS in the TCGA dataset, we developed a nomogram to predict 1-, 3- and 5-year survival probability in PDAC (Figure 6B). The C-index of our nomogram in the TCGA cohort was 0.689 (95% confidence interval, 0.628-0.749), and the calibration curves for this nomogram presented good agreement between the possibility estimated by the nomogram and the actual proportion (Figure 6C–6E).

Analysis of gene set enrichment relevance of the three-gene signature

To elucidate the molecular mechanism of the three-gene signature, patients from the TCGA PDAC dataset were divided into high- and low-risk groups according to the cutoff value. GSEA compared the high- and low-risk groups. Gene ontology (GO) gathers information about molecular function, biological processes, and cellular components in a number of different organisms. These enriched gene ontology (GO) data revealed that molecular alteration in the high-risk group was closely related to the nuclear periphery, apical junction and

nuclear matrix. Oncological signatures were significantly enriched, including HOXC6, SMAD2, SMAD3 and MTORC (Figure 7).

DISCUSSION

Pancreatic ductal adenocarcinoma (PDAC) is one of the most invasive solid malignancies and could become the second leading source of cancer-related deaths in the United States by 2030 [13]. Therefore, updated models with effective gene signatures are needed that can better identify patients with PDAC in high- and low-risk groups. Such models could aid in the selection of appropriate treatments and improve clinical management. Currently, clinical indices and tumor biomarkers are used to predict prognosis and relapse in patients with PDAC [14]. However, the value of some clinical features, such as tumor dimensions and lymph node metastases, in relapse and survival prediction remains controversial. Carbohydrate antigen 19-9 (CA19-9), CA50 and carcinoembryonic antigen (CEA) are classic serum biomarkers used for PDAC prognosis prediction. However, these tumor markers are not specific to PDAC; thus, elevations in their levels can be related to other gastrointestinal carcinomas. Compared with most solid cancers, PDAC is recognized as a more heterogeneous disease influenced by genetic polymorphisms and hypoxic environments, and metabolic reprogramming is a novel feature of PDAC. It has become increasingly recognized that extensive enhancement of glycolysis and reprogramming of lipid metabolism are both associated with the development and progression of PDAC [15]. Previous studies have suggested that various gene signatures could convey prognostic information about PDAC. However, the use of these signatures has some limitations, perhaps because of a lack of knowledge

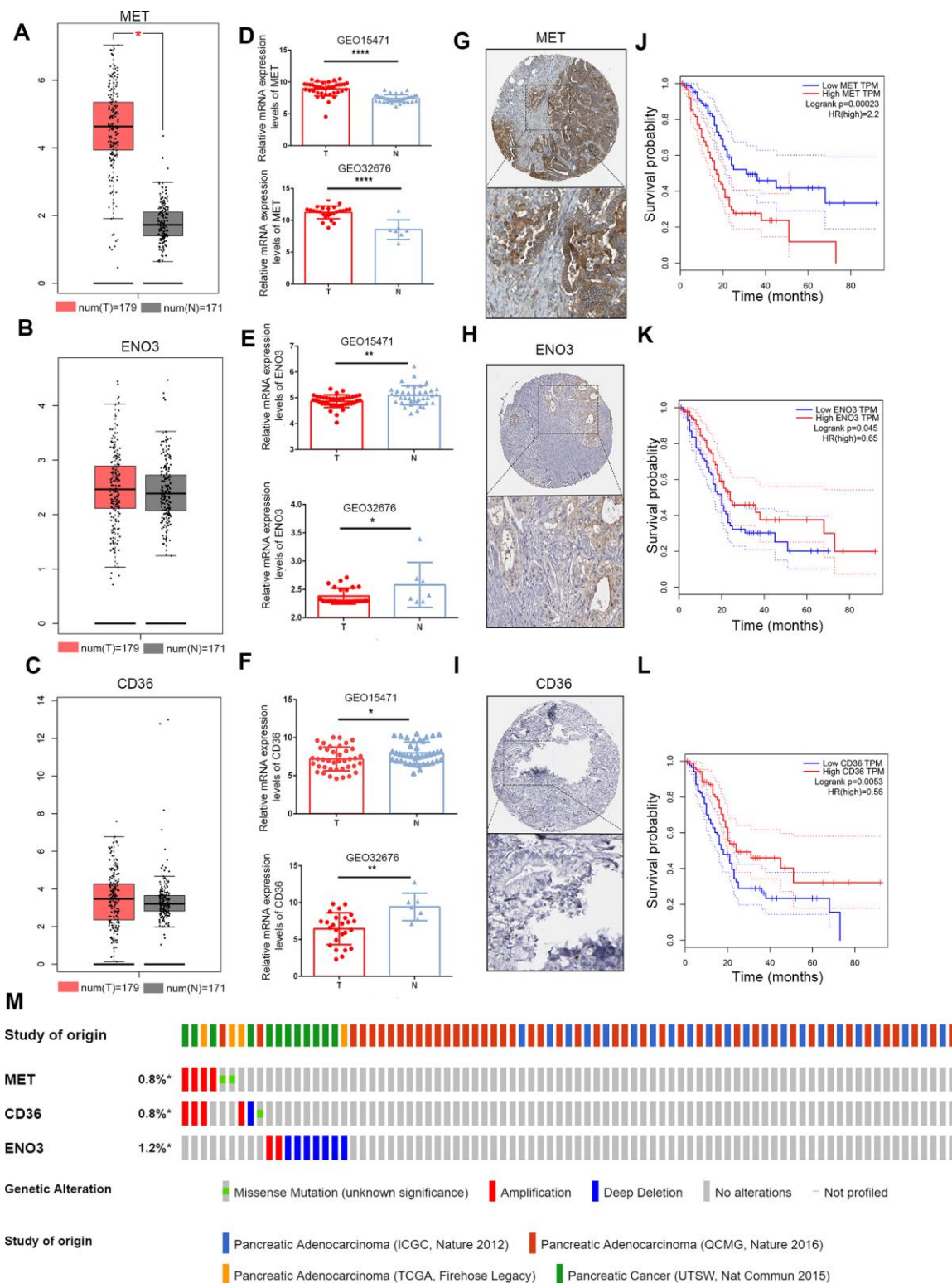


Figure 4. Validation of expression and alteration of the three genes in pancreatic cancer. (A–C) The MET, CD36 and ENO3 mRNA expression levels in TCGA pancreatic cancer tumor tissue and matching normal tissue from data on TCGA and GTEx. Data were obtained from the GEPIA (<http://gepia.cancer-pku.cn/>). (D–F) The MET, CD36 and ENO3 mRNA expression levels in GEO32676 and GEO15471 pancreatic cancer tumor tissue compared with non-tumor tissues. (G–I) The representative protein expression of the 3 glycolysis and lipid metabolism-based genes in pancreatic cancer tumor tissue. Data were obtained from the human protein atlas (<https://www.proteinatlas.org/>). (J–L) Survival analysis of patients with PAAD in terms of MET, CD36 and ENO3 in TCGA patients. (M) Genetic alterations of the three genes in the ICGC, QCMG, TCGA and UTSW pancreatic cancer datasets. Data were obtained from the cBioportal (<https://www.cbioportal.org/>).

regarding the genetic and energy supply backgrounds of PDAC. Therefore, we conducted multi-mRNA analysis based on metabolic reprogramming to identify novel signatures for accurate prognosis prediction in PDAC patients.

In the current study, a three-gene signature based on metabolic reprogramming was identified using the LASSO Cox regression model. This signature offered excellent accuracy in identifying patients with poor survival. The signature also showed significant ability to distinguish between PDAC and non-PDAC in GEO datasets and a strong, positive correlation with the mutation states of key genes in PDAC. Univariable and multivariate Cox analyses indicated that the three-gene signature was a powerful and independent prognostic indicator. Additionally, we developed a nomogram that included clinical factors. GSEA was used to further explore the underlying mechanism of the signature in high-risk PDAC patients.

PDAC cells have a powerful ability to survive in harsh environments by transforming their energy metabolism

processes. MET, also known as c-MET, is a receptor tyrosine kinase that builds connections between the extracellular matrix and cytoplasm by binding with its ligand hepatocyte growth factor (HGF). The MET receptor incorporates its ligand, HGF, through MET dimerization, and this process activates MET. In cancer cells, aberrant HGF/c-Met axis triggering, which is strongly linked to c-Met gene mutations and amplification, promotes tumor development/progression by inducing the PI3K/AKT, Ras/MAPK, JAK/STAT, SRC, Wnt/ β -catenin, and other signaling pathways. Up-regulated MET expression is also related to a shorter time to distant metastasis in PDAC patients receiving adjuvant and neoadjuvant chemoradiation therapy [16]. Furthermore, Yan et al. revealed a novel link between paracrine HGF/c-MET signaling and enhancement of stem cell-like potential and glycolysis in PDAC, involving activation of yes-associated protein [17]. Currently, drugs targeting c-Met that are in clinical trials can be divided into monoclonal antibodies (e.g., onartuzumab) and small molecule inhibitors, including ATP-competitive inhibitors (e.g., crizotinib) and non-ATP competitive inhibitors (e.g., tivantinib). Targeting

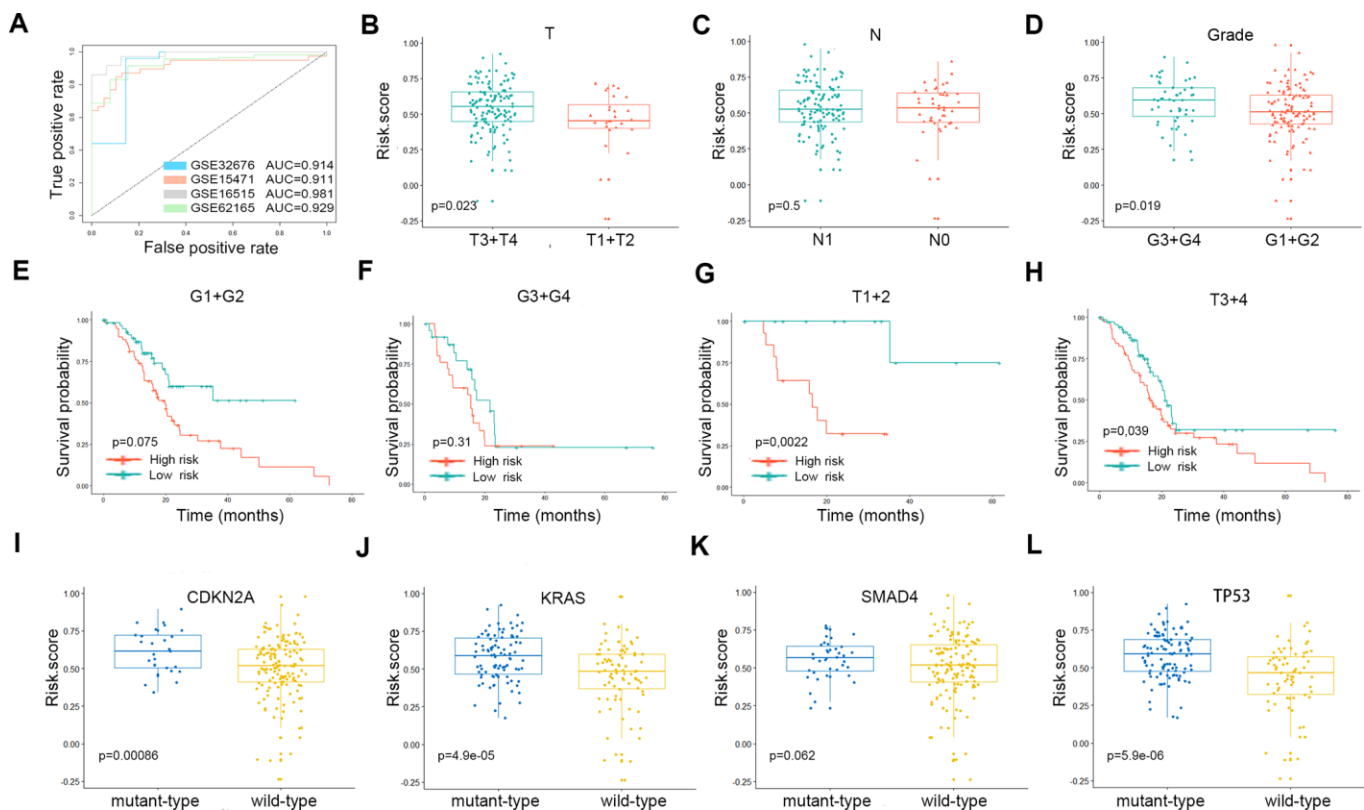


Figure 5. (A) The ROC curves of the risk scores differentiating pancreatic cancer from normal tissues in the four validation GEO datasets. The clinical and tumor mutation relevance of the three gene signatures. (B–D) The distribution of the risk scores in different AJCC stages in the TCGA cohort. (E–H) The Kaplan-Meier plot (low risk- score vs. high risk- score) of 5-year overall survival in patients in the TCGA cohort. (I–L) The expression level of the risk score in different mutation statuses of KRAS, TP53, CDKN2A, and SMAD4 in the TCGA dataset.

Table 2. Univariable Cox regression analyses of the risk score, age, gender, grade and tumor stage TCGA.

	HR (95% CI for HR)	Wald. Test	P. value
Age	1 (1-1.1)	5.4	0.02
Gender	0.8 (0.53-1.2)	1	0.31
Risk. Score	31 (8-120)	24	8.60E ⁻⁰⁷
Grade	1.5 (0.93-2.3)	2.8	0.097
Stage	0.74 (0.23-2.4)	0.26	0.61
N	2.1 (1.3-3.6)	7.9	0.0049
T	1.8 (0.93-3.5)	3	0.081

of c-Met for the treatment of tumorigenesis and metastasis is still believed to have broad clinical significance [18]. Enolase (ENO), also known as phosphopyruvate hydratase, is a metalloenzyme that catalyzes the conversion of 2-phosphoglycerate (2-PG) into phosphoenolpyruvate (PEP) during glycolysis and reverses the transformation of phosphoric acid-pyruvate into 2-phosphate-d-glycerate through glycogen synthesis. The ENOs in mammalian cells are composed of 3 subunits: α , or nonneuronal ENO (NNE); β , or muscle-specific ENO (MSE); and γ , or neuron-specific ENO (NSE) [19]. These molecules are cytoplasmic enzymes

associated with glycolysis and gluconeogenesis. ENO1 is activated by various glucose transporters and glycolytic enzymes associated with cell cycle progression and the Warburg effect in tumor cells. Some studies have shown that ENO1 acts as a potent promoter in tumor cells by regulating AMPK/mTOR and PI3K/AKT signaling and inactivating a downstream signaling pathway [20]. Up-regulated expression of the ENO2 gene has also been observed in non-small cell lung carcinomas (NSCLCs). The β -subunit of ENO is encoded by the ENO3 gene. Choa Park et al. suggested that up-regulated ENO3 gene expression is directly positively correlated with the

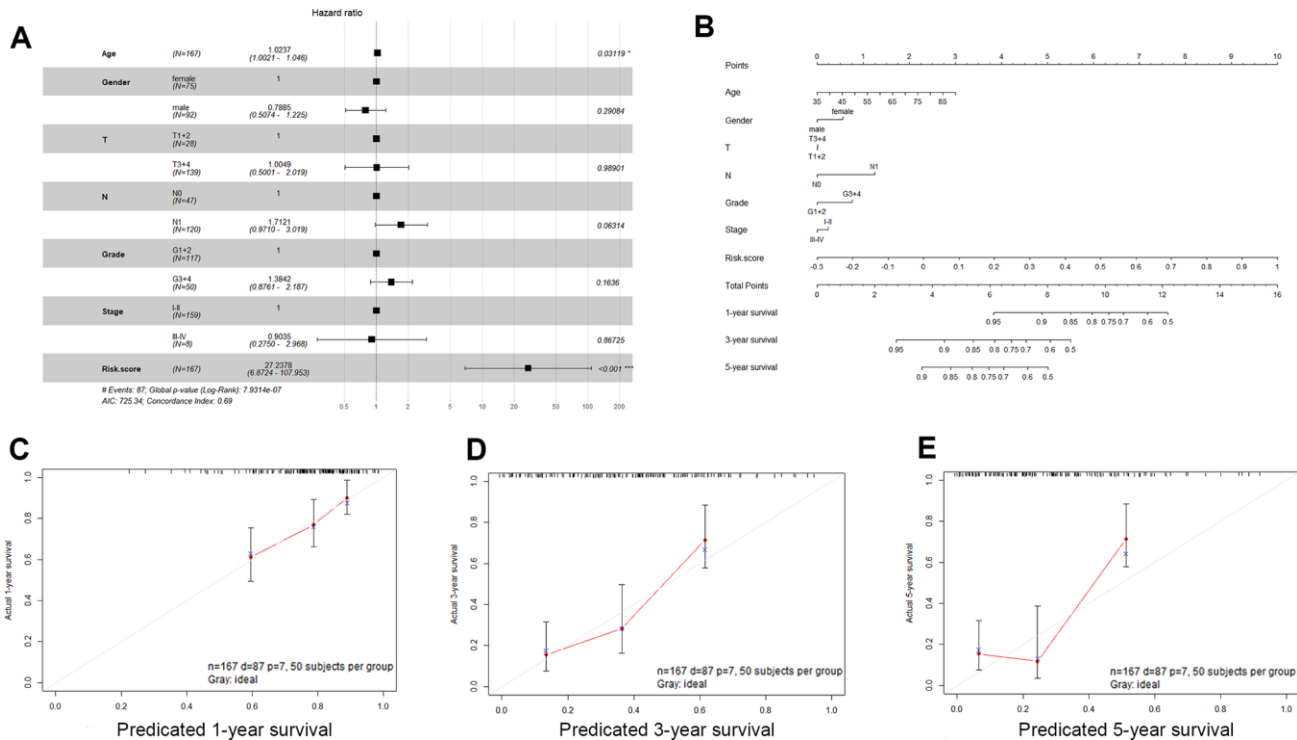


Figure 6. Validation of the nomogram in predicting overall survival of pancreatic cancer in the TCGA dataset. (A) Forest plot summary of multivariable Cox regression analyses of the risk score, age, sex, grade and tumor stage in the TCGA cohort. The squares represent the hazard ratio (HR), and the transverse lines represent 95% CIs. CI, confidence interval. **(B)** A nomogram to predict survival probability at 1, 3 and 5 years for PDAC patients based on the results derived from the TCGA cohort. **(C-E)** Calibration curve for the nomogram when predicting 1, 3- and 5-year overall survival.

consequence of STK11 loss of function in lung cancer. Additionally, down-regulation of ENO3 expression has been found to exert a selective anticancer effect in STK11 mutant cells compared with STK11 wild-type cells. However, the role of ENO3 in PDAC remains unknown [21].

Fatty acid (FA) derivatives consist of sterols, phospholipids, and sphingolipids, along with signaling

molecules that mediate the invasion, migration, and chemoresistance of cancer cells [7]. CD36, a transmembrane glycoprotein that belongs to the scavenger receptor class B family, is an essential member of this group [22]. In addition to FA uptake, CD36 promotes cholesterol uptake and transduces intracellular signaling, mediating the metabolic targeting of FAs. CD36 has also been revealed to play roles in other important cellular processes related to tumor

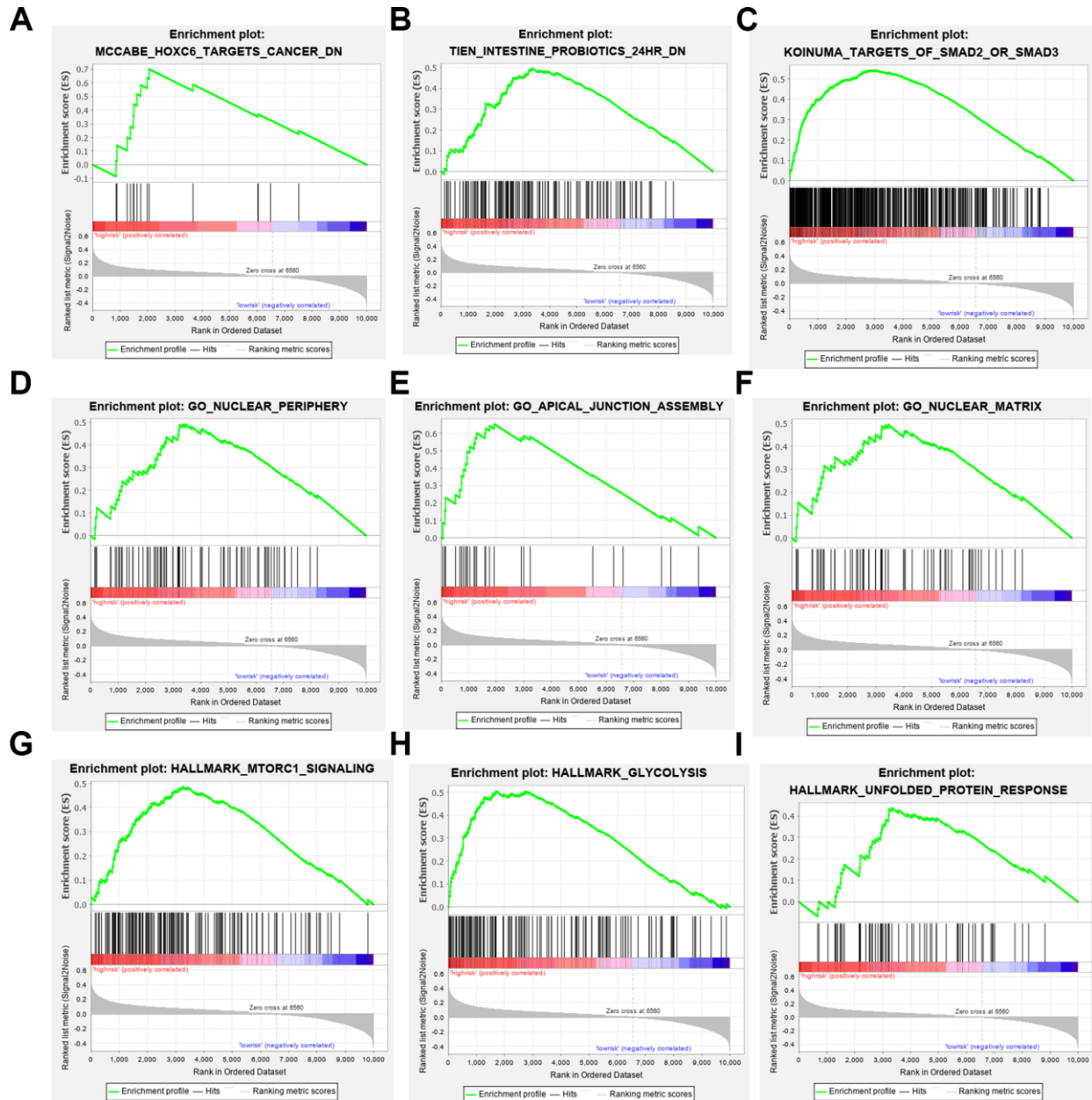


Figure 7. Gene set enrichment analyses. (A–C) Top 3 (HOXC6 target cancer, intestine probiotics, target of SMAD2 or SMAD3.) oncological signatures significantly C2 (hallmark gene sets) enriched in the high-risk group identified by gene set enrichment analysis. (D–F) Top 3 (nuclear periphery, apical junction assembly, nuclear matrix.) oncological signatures significantly C5 (biological process) enriched in the high-risk group identified by gene set enrichment analysis. (G–I) Top 3 (MTORC1 signaling, glycolysis, unfolded protein response.) oncological signatures significantly H (hallmark gene sets) enriched in the high-risk group identified by gene set enrichment analysis.

biology, including angiogenesis, antigen presentation and cell adhesion [23]. Recent reports have revealed that low expression of CD36 is associated with low TNM stages and CA19-9 levels but poor survival prognosis in PDAC patients [24]. However, Masahiko Kubo et al. suggested that increased expression of CD36 could increase gemcitabine resistance by regulating anti-apoptotic proteins and the endothelial-mesenchymal transition of PDAC cells [25]. These controversial results rendered the role of CD36 in PDAC unclear and could be partly due to limited awareness of FA components and their contributions to the biological behavior of PDAC.

Like other studies, the present study inevitably had some limitations. First, the main sources of our research data were public datasets, and our findings were not validated in a prospective cohort. Moreover, more experiments incorporating multiplatform analyses with genomics, transcriptomics and proteomics approaches should be performed to investigate the molecular mechanism of the identified signature to illuminate the association between the three-gene signature and PDAC prognosis.

In conclusion, our study successfully identified a three-gene signature and a prognostic nomogram for the prediction of OS in pancreatic cancer patients. The three-gene signature was found to be closely associated with the progression and prognosis of pancreatic cancer. This classifier could have clinical applications in guiding the use of metabolic reprogramming and could be useful for the personalized management of PDAC patients. Future investigations of the related molecular mechanisms and prospective, randomized clinical trials regarding this signature will be of great clinical significance and could provide a roadmap for precision medicine.

MATERIALS AND METHODS

Acquisition of PDAC gene expression data and clinical data

Microarray gene expression profiles for PDAC were downloaded from the NCBI Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>), TCGA (<http://cancergenome.nih.gov/>) and the International Cancer Genome Consortium (ICGC) database (<https://dcc.icgc.org/>). The gene expression microarray datasets GSE15471, GSE16515 and GSE32676 were chosen and downloaded for differentially expressed gene (DEG) analysis. The datasets met the following criterion: they contained data on human pancreatic tissue samples from tumors and adjacent paracancerous areas. Datasets that were unanalyzable, that included

few DEGs (fewer than 100), or that included defectively annotated genes were excluded. The probe names were transformed into gene symbols using the annotation files supplied by the manufacturer. The median ranking value was used as the expression value if several probes matched a single gene. Robust multiarray average (RMA)-normalized data were log₂-transformed for further analysis. Normalized RNA-sequencing data in transcripts per million (TPM) format and the associated clinical information of the PDAC samples were downloaded from the TCGA dataset. For validation, we used transcriptomic data from the ICGC and recruited 82 patients who underwent surgical resection at Fudan University Shanghai Cancer Center (FUSCC) from 2010 to 2012. These patients were diagnosed according to strict pathological criteria, and total RNA was extracted from their tumor tissues. Strictly controlled postoperative follow-ups were conducted for all of the patients. All of the procedures were implemented after obtaining approval from the Clinical Research Ethics Committee of FUSCC, and informed consent was acquired from each patient.

DEG identification and bioinformatic analysis

The “limma” package of R software (version 3.6.2) was used to establish a prognostic gene signature by identifying the DEGs from the mRNA data. A $|\log_2(\text{fold change [FC]})| > 1$, a P-value < 0.05 , and a false discovery rate (FDR) < 0.05 were considered the cutoffs for DEGs between tumor and nontumor samples in the GEO database. Next, the DEGs that overlapped among 3 series, as determined with Venn and UpSet diagrams, were regarded as credible DEGs. Gene ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were applied to explore the potential biological roles of the DEGs. Significantly relevant signaling pathways were identified with the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (<https://david.ncifcrf.gov/>) [26]. P < 0.05 was considered the threshold for statistical significance.

Establishment of the LASSO regression model and calculation of risk score

We used log-rank tests to examine the associations between gene expression and OS in the TCGA dataset. DEGs with P-values < 0.05 were considered statistically significant and were included in the following analyses. Genes obtained from the Molecular Signature Database (MSigDB) gene sets “REACTOME_GLYCOLYSIS” and “REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS” were used as glycolytic and lipid metabolism-related genes, respectively [7, 8, 27].

Table 3. The sequences of the primers for the three genes signatures.

Primer	Sequence (5'to 3')	Base number
m-MET-F	AGCAATGGGGAGTGTAAGAGG	22
m-MET-R	CCCAGTCTTGTACTCAGCAAC	21
m-ENO3-F	GGCTGGTTACCCAGACAAGG	20
m-ENO3-R	TCGTACTTCCCATTGCGATAGAA	23
m-CD36-F	GGCTGTGACCGGAACTGTG	19
m-CD36-R	AGGTCTCCAACCTGGCATTAGAA	22

Twelve glycolysis and lipid metabolism-related genes were selected to construct LASSO Cox regression analysis models with the R package “glmnet”. LASSO analysis is the most popular method for evaluating survival data and is especially appropriate for analyzing gene expression profiles. The “glmnet” package returned a series of models. For each model, the tuning parameter λ was inversely correlated with the complexity of the model and the value of deviance. Ten-fold cross-validation was performed to screen the best model from the series (the model with the minimum mean cross-validation error). The patients were divided into high-risk and low-risk groups according to a cutoff median risk score. Kaplan-Meier analysis and time-dependent receiver-operating characteristic (ROC) curve analysis were used to test predictive differences between the high- and low-risk groups in the discovery and two validation datasets. Cox regression analysis was employed to test whether the risk score was an independent prognostic factor. A nomogram and related calibration curves were created based on the TCGA cohort for further clinical application.

Gene set enrichment analysis (GSEA) and mutation analyses

GSEA was applied to elucidate the molecular mechanisms of the prognostic gene sets [28]. The TCGA dataset was divided into high- and low-risk groups according to the cutoffs. GSEA was performed in the Java program GSEA, version 3.0. The MSigDB v. 6.2. C2 collection (curated gene sets), C5 collection (GO gene sets), and H collection (hallmark gene sets) were searched to identify enriched KEGG pathways and biological information and dysfunctional oncogenic signatures associated with the high-risk group. Results with a |normalized enrichment score (NES)| > 1 and an FDR < 0.05 were considered statistically significant. Somatic mutations (SNPs and small indels) in TCGA were obtained with MuTect2 in the University of California, Santa Cruz (UCSC), with the Xena browser (<https://xenabrowser.net/datapages/>). Genetic alterations in the three genes associated related to pancreatic cancer were obtained from cBioPortal (<https://www.cbioportal.org/>) [29].

RNA extraction, reverse transcription, and qRT-PCR analysis

RNA was obtained from patients' samples and clinically diagnosed with PDAC at the Fudan University Shanghai Cancer Center (FUSCC) from 2010 to 2012. Informed consent was obtained from each patient, and all of the experiments were performed with the approval of the Clinical Research Ethics Committee of FUSCC (ethical code: 050432-4-1212B). In the FUSCC validation set, total RNA was extracted from 82 patient samples using a MiniBEST Universal RNA Extraction Kit (9767, TaKaRa). A PrimeScript RT Reagent Kit (K1622, Thermo Scientific) was used to synthesize first-strand cDNA isolated from total RNA. Then, SYBR Green qRT-PCR was conducted on an ABI 7900HT platform (Applied Biosystems, USA). We used GAPDH mRNA as an internal reference. The primers used for the mRNA molecules tested in this study were synthesized by Sangon (Shanghai, China), and the sequences are listed in Table 3.

Statistical analysis

Statistical analysis was performed in R software (version 3.6.2, <https://www.r-project.org/>) and GraphPad Prism software, version 7.01 (GraphPad Software, La Jolla, CA, USA). Categorical variables were analyzed by the χ^2 test or Fisher's exact test. Continuous variables were analyzed using Student's t-test for paired samples. Kaplan-Meier curves and log-rank tests were used. Calibration plots were produced to assess whether actual outcomes matched predicted outcomes for the nomogram. All of the statistical tests were 2 sided, and a P-value < 0.05 was considered to indicate statistical significance.

Abbreviations

AUC: area under the curve; DEG: differentially expressed genes; FAs : fatty acids; FUSCC: Fudan University Shanghai Cancer Center; FDR: false discovery rate; GEO: Gene Expression Omnibus; GSEA: gene set enrichment analysis; ICGC: International Cancer

Genome Consortium; LASSO: least absolute shrinkage and selection operator; MSigDB: Molecular Signatures Database; OS: overall survival; PDAC: pancreatic ductal adenocarcinoma; TCGA: The Cancer Genome Atlas.

AUTHOR CONTRIBUTIONS

CL and XY designed this study; ZT, YBL, JX and SS developed the methodology and acquired the related data; JH, QM, BZ and JL performed the data analysis and interpretation; ZT, BL, MW and YZ drafted and revised the manuscript; CL and XY supervised the study.

CONFLICTS OF INTEREST

The authors confirm that the contents of this article have no conflicts of interest.

FUNDING

This study was jointly funded by the National Natural Science Foundation of China (Nos. 81902428, 81802352 and 81772555), the Shanghai Sailing Program (No. 19YF1409400), the National Science Foundation for Distinguished Young Scholars of China (No. 81625016), the Clinical and Scientific Innovation Project of Shanghai Hospital Development Center (SHDC12018109) and the Scientific Innovation Project of Shanghai Education Committee (2019-01-07-00-07-E00057).

REFERENCES

- Adams CR, Htwe HH, Marsh T, Wang AL, Montoya ML, Subbaraj L, Tward AD, Bardeesy N, Perera RM. Transcriptional control of subtype switching ensures adaptation and growth of pancreatic cancer. *Elife*. 2019; 8:e45313. <https://doi.org/10.7554/elife.45313> PMID:31134896
- Anslan S, Bahram M, Hiiesalu I, Tedersoo L. PipeCraft: flexible open-source toolkit for bioinformatics analysis of custom high-throughput amplicon sequencing data. *Mol Ecol Resour*. 2017; 17:e234–40. <https://doi.org/10.1111/1755-0998.12692> PMID:28544559
- Raman P, Maddipati R, Lim KH, Tozeren A. Pancreatic cancer survival analysis defines a signature that predicts outcome. *PLoS One*. 2018; 13:e0201751. <https://doi.org/10.1371/journal.pone.0201751> PMID:30092011
- Yan X, Wan H, Hao X, Lan T, Li W, Xu L, Yuan K, Wu H. Importance of gene expression signatures in pancreatic cancer prognosis and the establishment of a prediction model. *Cancer Manag Res*. 2018; 11:273–83. <https://doi.org/10.2147/CMAR.S185205> PMID:30643453
- Frost HR, Amos CI. Gene set selection via LASSO penalized regression (SLPR). *Nucleic Acids Res*. 2017; 45:e114. <https://doi.org/10.1093/nar/gkx291> PMID:28472344
- Ying H, Kimmelman AC, Lyssiotis CA, Hua S, Chu GC, Fletcher-Sananikone E, Locasale JW, Son J, Zhang H, Colloff JL, Yan H, Wang W, Chen S, et al. Oncogenic kras maintains pancreatic tumors through regulation of anabolic glucose metabolism. *Cell*. 2012; 149:656–70. <https://doi.org/10.1016/j.cell.2012.01.058> PMID:22541435
- Sunami Y, Rebelo A, Kleeff J. Lipid metabolism and lipid droplets in pancreatic cancer and stellate cells. *Cancers (Basel)*. 2017; 10:3. <https://doi.org/10.3390/cancers10010003> PMID:29295482
- Yang J, Ren B, Yang G, Wang H, Chen G, You L, Zhang T, Zhao Y. The enhancement of glycolysis regulates pancreatic cancer metastasis. *Cell Mol Life Sci*. 2020; 77:305–21. <https://doi.org/10.1007/s00018-019-03278-z> PMID:31432232
- McDonald OG, Li X, Saunders T, Tryggvadottir R, Mentch SJ, Warmoes MO, Word AE, Carrer A, Salz TH, Natsume S, Stauffer KM, Makohon-Moore A, Zhong Y, et al. Epigenomic reprogramming during pancreatic cancer progression links anabolic glucose metabolism to distant metastasis. *Nat Genet*. 2017; 49:367–76. <https://doi.org/10.1038/ng.3753> PMID:28092686
- Liang C, Qin Y, Zhang B, Ji S, Shi S, Xu W, Liu J, Xiang J, Liang D, Hu Q, Liu L, Liu C, Luo G, et al. Energy sources identify metabolic phenotypes in pancreatic cancer. *Acta Biochim Biophys Sin (Shanghai)*. 2016; 48:969–79. <https://doi.org/10.1093/abbs/gmw097> PMID:27649892
- Zhao H, Duan Q, Zhang Z, Li H, Wu H, Shen Q, Wang C, Yin T. Up-regulation of glycolysis promotes the stemness and EMT phenotypes in gemcitabine-resistant pancreatic cancer cells. *J Cell Mol Med*. 2017; 21:2055–67. <https://doi.org/10.1111/jcmm.13126> PMID:28244691
- Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015; 347:1260419. <https://doi.org/10.1126/science.1260419> PMID:25613900
- Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer

- incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* 2014; 74:2913–21.
<https://doi.org/10.1158/0008-5472.CAN-14-0155>
 PMID:[24840647](https://pubmed.ncbi.nlm.nih.gov/24840647/)
14. Ray K. Pancreatic cancer: biomarkers for the early detection of PDAC. *Nat Rev Gastroenterol Hepatol.* 2017; 14:504–05.
<https://doi.org/10.1038/nrgastro.2017.111>
 PMID:[28765582](https://pubmed.ncbi.nlm.nih.gov/28765582/)
 15. Sousa CM, Kimmelman AC. The complex landscape of pancreatic cancer metabolism. *Carcinogenesis.* 2014; 35:1441–50.
<https://doi.org/10.1093/carcin/bgu097> PMID:[24743516](https://pubmed.ncbi.nlm.nih.gov/24743516/)
 16. Fu YT, Zheng HB, Zhou L, Zhang DQ, Liu XL, Sun H. Valproic acid, targets papillary thyroid cancer through inhibition of c-met signalling pathway. *Am J Transl Res.* 2017; 9:3138–47.
 PMID:[28670399](https://pubmed.ncbi.nlm.nih.gov/28670399/)
 17. Yan B, Jiang Z, Cheng L, Chen K, Zhou C, Sun L, Qian W, Li J, Cao J, Xu Q, Ma Q, Lei J. Paracrine HGF/c-MET enhances the stem cell-like potential and glycolysis of pancreatic cancer cells via activation of YAP/HIF-1 α . *Exp Cell Res.* 2018; 371:63–71.
<https://doi.org/10.1016/j.yexcr.2018.07.041>
 PMID:[30056064](https://pubmed.ncbi.nlm.nih.gov/30056064/)
 18. Zhang Y, Xia M, Jin K, Wang S, Wei H, Fan C, Wu Y, Li X, Li X, Li G, Zeng Z, Xiong W. Function of the c-met receptor tyrosine kinase in carcinogenesis and associated therapeutic opportunities. *Mol Cancer.* 2018; 17:45.
<https://doi.org/10.1186/s12943-018-0796-y>
 PMID:[29455668](https://pubmed.ncbi.nlm.nih.gov/29455668/)
 19. Merkulova T, Dehaupas M, Nevers MC, Créminon C, Alameddine H, Keller A. Differential modulation of alpha, beta and gamma enolase isoforms in regenerating mouse skeletal muscle. *Eur J Biochem.* 2000; 267:3735–43.
<https://doi.org/10.1046/j.1432-1327.2000.01408.x>
 PMID:[10848992](https://pubmed.ncbi.nlm.nih.gov/10848992/)
 20. Zhao M, Fang W, Wang Y, Guo S, Shu L, Wang L, Chen Y, Fu Q, Liu Y, Hua S, Fan Y, Liu Y, Deng X, et al. Enolase-1 is a therapeutic target in endometrial carcinoma. *Oncotarget.* 2015; 6:15610–27.
<https://doi.org/10.18632/oncotarget.3639>
 PMID:[25951350](https://pubmed.ncbi.nlm.nih.gov/25951350/)
 21. Park C, Lee Y, Je S, Chang S, Kim N, Jeong E, Yoon S. Overexpression and selective anticancer efficacy of ENO3 in STK11 mutant lung cancers. *Mol Cells.* 2019; 42:804–09.
<https://doi.org/10.14348/molcells.2019.0099>
 PMID:[31697874](https://pubmed.ncbi.nlm.nih.gov/31697874/)
 22. Pepino MY, Kuda O, Samovski D, Abumrad NA. Structure-function of CD36 and importance of fatty acid signal transduction in fat metabolism. *Annu Rev Nutr.* 2014; 34:281–303.
<https://doi.org/10.1146/annurev-nutr-071812-161220>
 PMID:[24850384](https://pubmed.ncbi.nlm.nih.gov/24850384/)
 23. Ladanyi A, Mukherjee A, Kenny HA, Johnson A, Mitra AK, Sundaresan S, Nieman KM, Pascual G, Benitah SA, Montag A, Yamada SD, Abumrad NA, Lengyel E. Adipocyte-induced CD36 expression drives ovarian cancer progression and metastasis. *Oncogene.* 2018; 37:2285–301.
<https://doi.org/10.1038/s41388-017-0093-z>
 PMID:[29398710](https://pubmed.ncbi.nlm.nih.gov/29398710/)
 24. Jia S, Zhou L, Shen T, Zhou S, Ding G, Cao L. Down-expression of CD36 in pancreatic adenocarcinoma and its correlation with clinicopathological features and prognosis. *J Cancer.* 2018; 9:578–83.
<https://doi.org/10.7150/jca.21046> PMID:[29483963](https://pubmed.ncbi.nlm.nih.gov/29483963/)
 25. Kubo M, Gotoh K, Eguchi H, Kobayashi S, Iwagami Y, Tomimaru Y, Akita H, Asaoka T, Noda T, Takeda Y, Tanemura M, Mori M, Doki Y. Impact of CD36 on chemoresistance in pancreatic ductal adenocarcinoma. *Ann Surg Oncol.* 2020; 27:610–19.
<https://doi.org/10.1245/s10434-019-07927-2>
 PMID:[31605325](https://pubmed.ncbi.nlm.nih.gov/31605325/)
 26. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009; 4:44–57.
<https://doi.org/10.1038/nprot.2008.211>
 PMID:[19131956](https://pubmed.ncbi.nlm.nih.gov/19131956/)
 27. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics.* 2011; 27:1739–40.
<https://doi.org/10.1093/bioinformatics/btr260>
 PMID:[21546393](https://pubmed.ncbi.nlm.nih.gov/21546393/)
 28. 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](https://pubmed.ncbi.nlm.nih.gov/16199517/)
 29. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013; 6:pl1.
<https://doi.org/10.1126/scisignal.2004088>
 PMID:[23550210](https://pubmed.ncbi.nlm.nih.gov/23550210/)