Research paper

Identification of candidate diagnostic and prognostic biomarkers for pancreatic carcinoma

Yang Cheng [a,b,1], Kunyuan Wang [a,1], Lanlan Geng [b,1], Jingjing Sun [b], Wanfu Xu [b], Dingli Liu [a], Sitang Gong [b,*], Yun Zhu [a,**]

[a] Liver Tumor Center, Department of Infectious Diseases and Hepatology Unit, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong 510515, China
[b] Digestive Department, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, No.9 Jinsui Road, Guangzhou, Guangdong 510623, China

ARTICLE INFO

ABSTRACT

Background: Pancreatic carcinoma (PC) is one of the most aggressive cancers affecting human health. It is essential to identify candidate biomarkers for the diagnosis and prognosis of PC. The present study aimed to investigate the diagnosis and prognosis biomarkers of PC.

Methods: Differentially expressed genes (DEGs) were identified from the mRNA expression profiles of GSE62452, GSE28735 and GSE16515. Functional analysis and the protein-protein interaction network analysis was performed to explore the biological function of the identified DEGs. Diagnosis markers for PC were identified using ROC curve analysis. Prognosis markers were identified via survival analysis of TCGA data. The protein expression pattern of the identified genes was verified in clinical tissue samples. A retrospective clinical study was performed to evaluate the correlation between the expression of candidate proteins and survival time of patients. Moreover, comprehensive analysis of the combination of multiple genes/proteins for the prognosis prediction of PC was performed using both TCGA data and clinical data. In vitro studies were undertaken to elaborate the potential roles of these biomarkers in clonability and invasion of PC cells.

Findings: In total, 389 DEGs were identified. These genes were mainly associated with pancreatic secretion, protein digestion and absorption, cytochrome P450 drug metabolism, and energy metabolism pathway. The top 10 genes were filtered out following Fisher's exact test. ROC curve analysis demonstrated that TMPRSS4, SERPINB5, SLC6A14, SCEL, and TNS4 could be used as biomarkers for the diagnosis of PC. Survival analysis of TCGA data and clinical data suggested that TMC7, TMPRSS4, SCEL, SLC2A1, CENPF, SERPINB5 and SLC6A14 can be potential biomarkers for the prognosis of PC. Comprehensive analysis show that a combination of identified genes/proteins can predict the prognosis of PC. Mechanistically, the identified genes attributes to clonability and invasiveness of PC cells.

Interpretation: We synthesized several sets of public data and preliminarily clarified pathways and functions of PC. Candidate molecular markers were identified for diagnosis and prognosis prediction of PC including a novel gene, TMC7. Moreover, we found that the combination of TMC7, TMPRSS4, SCEL, SLC2A1, CENPF, SERPINB5 and SLC6A14 can serve as a promising indicator of the prognosis of PC patients. The candidate proteins may attribute to clonability and invasiveness of PC cells. This research provides a novel insight into molecular mechanisms as well as diagnostic and prognostic markers of PC.

Fund: National Natural Science Foundation of China [No. 81602646 & 81802339], Natural Science Foundation of Guangdong Province [No. 2016A030310254] and China Postdoctoral Science Foundation [No. 2016M600648].

1. Introduction

Pancreatic carcinoma (PC) is a highly malignant tumor that accounts for 216,000 new cancer cases annually and causes >200,000 deaths a year worldwide [1,2]. The 5-year survival rate for patients with PC is <5% due to early metastasis to regional lymph nodes and hematogenous spread to distant organs. The only potentially curable treatment of PC is surgical resection. However, approximately 80% of tumors are

* Corresponding author.
** Corresponding author.
E-mail addresses: sitangg@yahoo.com (S. Gong), zyfreemail@126.com (Y. Zhu).
[1] These authors contributed equally to the work.

Research in context

Evidence before this study

Early detection of pancreatic carcinoma (PC) is essential in order to provide patients with an optimal therapeutic approach. Carbohydrate antigen 19–9 is the only diagnostic marker approved by the FDA but its diagnostic potential is limited due to its restricted sensitivity and specificity. Integrated analysis is also required for accurate prognostic biomarkers that help to guide patients' therapy. Currently a variety of potential biomarkers in blood and tumors have been reported. However, many studies focused only on disparate genes, which are not sufficient for the diagnosis and prognosis of PC. Besides, few of these biomarkers have been validated for clinical use. Thus, the combination of different markers as diagnostic or prognostic indices appears promising. Microarrays run on high-throughput platforms have emerged as a promising and efficient tool for screening differentially expressed genes (DEGs) in cancers and identifying promising biomarkers for the diagnosis and prognosis of cancers. Many gene expression profiling microarrays have been conducted to find various DEGs in various cancers. However, inconsistent results are commonly obtained due to either sample heterogeneity in independent studies or studies conducted using only a single cohort.

Added value of this study

In this study, we synthesized several sets of public data and clinical data. Candidate molecular markers were identified for diagnosis and prognosis prediction of PC. Moreover, we found that the combination of the markers can serve as a promising prognostic indicator in patients with PC. A novel gene, TMC7, that is overexpressed in PC was identified. The role of the candidate differentially expressed proteins in PC cells including TMC7 was preliminary explored. This research provides a novel insight into diagnostic and prognostic markers as well as molecular mechanisms of PC.

Implications of all the available evidence

Our study suggested the identified seven genes can be potential target genes of PC. Inhibition of these genes may provide a potential therapeutic target for treatment of PC. We plan to explore whether the expression pattern of the potential diagnostic markers exist in circulating tumor cells from PC patients. We hope that these markers can help to make an early diagnosis and also help to predict the prognosis which might provide essential information regarding personalized treatment decisions for individual patients.

have been validated for clinical use [5,6]. Thus, the combination of different markers as diagnostic or prognostic indices appears promising.

Recently, microarrays run on high-throughput platforms have emerged as a promising and efficient tool for screening differentially expressed genes (DEGs) in cancers and identifying promising biomarkers for the diagnosis and prognosis of cancers [7]. Many gene expression profiling microarrays have been conducted to find various DEGs in various cancers [8–10]. However, inconsistent results are commonly obtained due to either sample heterogeneity in independent studies or studies conducted using only a single cohort.

The aim of the present study was to explore possible molecular mechanisms and potential diagnostic and prognostic biomarkers of PC. First, data from three Gene Expression Omnibus (GEO) database were combined and key genes involved in PC were picked out using bioinformatic methods. The diagnostic value of the genes was evaluated using ROC curve. And the prognostic value of the genes was evaluated using survival analysis of TCGA data. Following, the protein expression pattern of the candidate genes was detected in clinical tissue samples. We further performed a retrospective clinical study of the expression of candidate proteins and survival time of patients. Moreover, the combination of multiple proteins for the prognosis prediction of PC was evaluated. In vitro experiment was conducted to elaborate the potential roles of these biomarkers in clonability and invasion of PC cells.

2. Materials and methods

2.1. Differential expression analysis

Datasets were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo/). mRNA expression profiles of GSE62452 [11], GSE28735 [12] and GSE16515 [13] were downloaded from the GEO data repository. The dataset for GSE62452 includes 69 PC samples and 61 normal pancreatic tissue samples. GSE28735 incorporates 45 PC samples and 45 normal pancreatic tissue samples. GSE16515 consists of 36 PC samples and 16 normal samples. A total of 150 PC and 122 nonmalignant pancreatic tissue samples were included.

The original annotation files were downloaded, then quality control and normalization were performed using the robust multi-array average (RMA) method [14]. A *t*-test followed by Benjamini & Hochberg (BH) adjustment were applied to identify DEGs between tumor and normal tissues. The genes that met the cutoff criteria of a fold change > 2 and an adjusted *P*-value < .05 were considered DEGs. Integrative analysis of the three set of DEGs was done using Robust Rank Aggreg [15], and genes with a score of < 0.01 were selected. In total, 389 DEGs were selected. Clustering analyses were done to show expression patterns of the differentially expressed genes in tumor and normal tissues.

2.2. GO and pathway enrichment analyses

Online biological tools were used to investigate the functions and pathways of the candidate DEGs. Gene Ontology (GO) enrichment analysis (http://www.geneontology.org/) was used to explore the biological functions associated with the DEGs [16]. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis (http://www.genome.jp/kegg/pathway.html) was applied to further illuminate the pathways of the DEGs [17]. The Database for Annotation, Visualization and Integrated Discovery (DAVID) (https://david.abcc.Ncifcrf.gov/) can provide biological meaning for genes. GO and KEGG pathway enrichment analysis of identified DEGs was performed using DAVID [18,19]. P < 0.05 was set as the threshold. The false discovery rate (FDR) was controlled at the 0.01 threshold.

2.3. Protein-protein interaction (PPI) network construction

A PPI network of identified DEGs in all three datasets was constructed using the STRING online database (http://string-db.org) [20].

unresectable at the time of diagnosis [3,4]. For patients with advanced stage of PC, chemotherapy is the treatment of choice although the regimens have extensive side-effects, making them unsuitable for patients with a low performance status. For this reason, early detection of PC is essential in order to provide patients with an optimal therapeutic approach. In addition, integrated analysis is required for accurate prognostic biomarkers that help to guide patients' therapy.

Recent advances in genome-wide studies have revealed the diverse and complex genetic alterations of PC patients which may explain diverse disease behavior in a clinical setting. Currently, a variety of potential biomarkers in blood and tumors have been reported. However, many studies focused only on disparate genes, which are not sufficient for the diagnosis and prognosis of PC. Besides, few of these biomarkers

Cytoscape software (http://www.cytoscape.org/) was used to analyze the interaction relationships of the candidate DEGs encoding proteins in PC [21]. Functional modules in interaction networks were conducted using the Markov clustering algorithm. The most stringent protein interaction screening criteria (confidence > 0.9) were set as the threshold.

2.4. Data standardization

To identify strong candidate genes that can accurately recognize and represent changes in PC, standardized and comprehensive analysis of three sets of data from different platforms was performed. The YuGene transform was applied to eliminate the impact of the different platforms [22]. The YuGene software is implemented as an R package available from CRAN.

2.5. ROC analysis

Receiver operating characteristic (ROC) curve analysis was performed to evaluate the sensitivity and specificity of the DEGs for PC diagnosis using the R package. An area under the curve (AUC) value was calculated and used to designate the ROC effect.

2.6. Immunohistochemistry

Tissue microarray chips containing 99 samples of pancreatic cancer and 71 samples of paired normal pancreatic tissue were purchased from Outdo Biotech (Shanghai, China). The characteristics of the patients are shown in Supplementary File S1. Immunohistochemistry (IHC) staining was performed as previously described [23]. Microarray chips were stained with anti-CENPF (Abcam), anti-SCEL, anti- GLUT-1, anti-SLC6A14, anti-TMC7 (Thermofisher, Waltham, MA USA), anti-TMPRSS4 and anti- MASpin (Gene Tex, Irvine, CA, USA) antibodies. Control staining with only secondary antibodies was performed to ensure specificity. Rabbit IgG polyclonal-Isotype control (Abcam) was used as the negative control. The score for staining was independently assessed by two experienced pathologists based on the integrated staining intensity and the proportion of positive cells. The final score was determined by adding the staining intensity score and the average proportion of positive cells score; the final score ranged from 0 to 7. For the purpose of further analysis, the samples with a score of 0–3 were defined as low expression, while the samples with scores of 4–7 were defined as high expression.

2.7. The cox regression model

A multivariate Cox regression model was constructed on TCGA dataset using the seven DEGs (CENPF, SCEL, SLC2A1, SLC6A14, TMC7, TMPRSS4 and SERPINB5). The stimated baseline risk was calculated as the following formula: $h\hat{\ }(t) = h\hat{\ }_0 (t) \exp(x_i'\beta\hat{\ })$

Where x_i' indicates gene and $\beta\hat{\ }$ refers to the relative expression value of corresponding gene.

2.8. Survival analysis

A Kaplan-Meier survival analysis was performed for patients with differential gene/protein expression or of high−/low-risk group. For survival analysis of a single gene/protein, patients were assigned to high/low gene expression groups according to the median of the expression level of mRNA/protein. For survival analysis of combined seven genes, patients from TCGA dataset were divided into high−/ low-risk groups according to the Cox regression model. For survival analysis of combined seven proteins, clinical patients with high expression of more than four proteins were divided into the high-risk group and the remaining patients were divided into the low-risk group. Statistical significance was assessed using the log-rank test (P < 0.05).

2.9. Cell culture and transient transfection

Human pancreatic cancer cell lines PANC-1 and BxPC3 were obtained from Shanghai Advanced Research Institute, Chinese Academy of Sciences. Cells were cultured in DMEM supplemented with 10% fetal bovine serum. For transient transfection, cells were seeded in 6-well plates and transfected with plasmid using Lipofectamine 2000 (Invitrogen, USA) according to the manufacturer's instructions. Target sequences for shRNAs are summarized in Supplementary Table S1.

2.10. Soft agar assay

1×10^4 pre-treated PANC-1 and BxPC3 cells were used for soft agar assay as previously described [23]. Cells were cultured for 14 days. Viable cell colonies larger than 0.1 mm were counted using a dissecting microscope.

2.11. Transwell assay

PANC-1 and BxPC3 cells were treated as described above and seeded onto Matrigel-coated Transwell chambers for 24 h. Transwell assay was performed according to the method previously described [24]. Cells that had passed through the membrane were counted.

3. Results

3.1. Identification of differentially expressed genes in PC and data intergration

Three datasets of PC were downloaded from the GEO database. Gene expression profiles from GSE 16515 identified 1118 differentially expressed genes with 851 genes upregulated and 267 genes downregulated in PC samples when compared with normal pancreatic tissues. Gene expression profiles from GSE 62452 identified 325 differentially expressed genes with 204 genes upregulated and 121 genes downregulated in PC. Gene expression profiles from GSE 28735 identified 495 differentially expressed genes with 253 genes upregulated and 206 genes downregulated in PC (Fig. 1a). Heatmap analysis showed that these genes presented differential expression profiles between normal tissues and cancer tissues (Fig. 1b). For integrated analyses of the three GEO datasets, Robust Rank Aggreg was applied, and genes with a score < 0.01 were selected. A total of 389 consistently expressed DEGs were identified from the three datasets, and the cluster heatmap is shown in Fig. 1c.

3.2. Functional characterization of PC specific genes

To characterize the function of the DEGs and to identify key candidate pathways, GO functional analysis and KEGG pathway analysis were performed. GO analysis showed that the DEGs were mainly enriched in processes such as digestion, xenobiotic glucuronidation, proteolysis, and negative regulation of glucuronosyltransferase activity (Fig. 2a). KEGG pathway enrichment analysis demonstrated that the DEGs were mainly associated with pancreatic secretion, protein digestion and absorption, retinol metabolism, cytochrome P450 drug metabolism, pentose and glucuronate interaction, etc. (Fig. 2b).

Using the STRING online database, the PPI network complex of the DEGs was constructed to discover protein functions (Supplementary Fig. S1). The results of the PPI network were mostly consistent with the results of GO and KEGG enrichment, indicating that the differential proteins in the network play an important role in the development of PC. We also found that these proteins have a distinct multidirectional aggregation that forms many clusters. For example, genes in the cluster formed by the red nodes are mainly associated with the peroxidation of alcohol and aldehyde and in the transport of glucose. These results
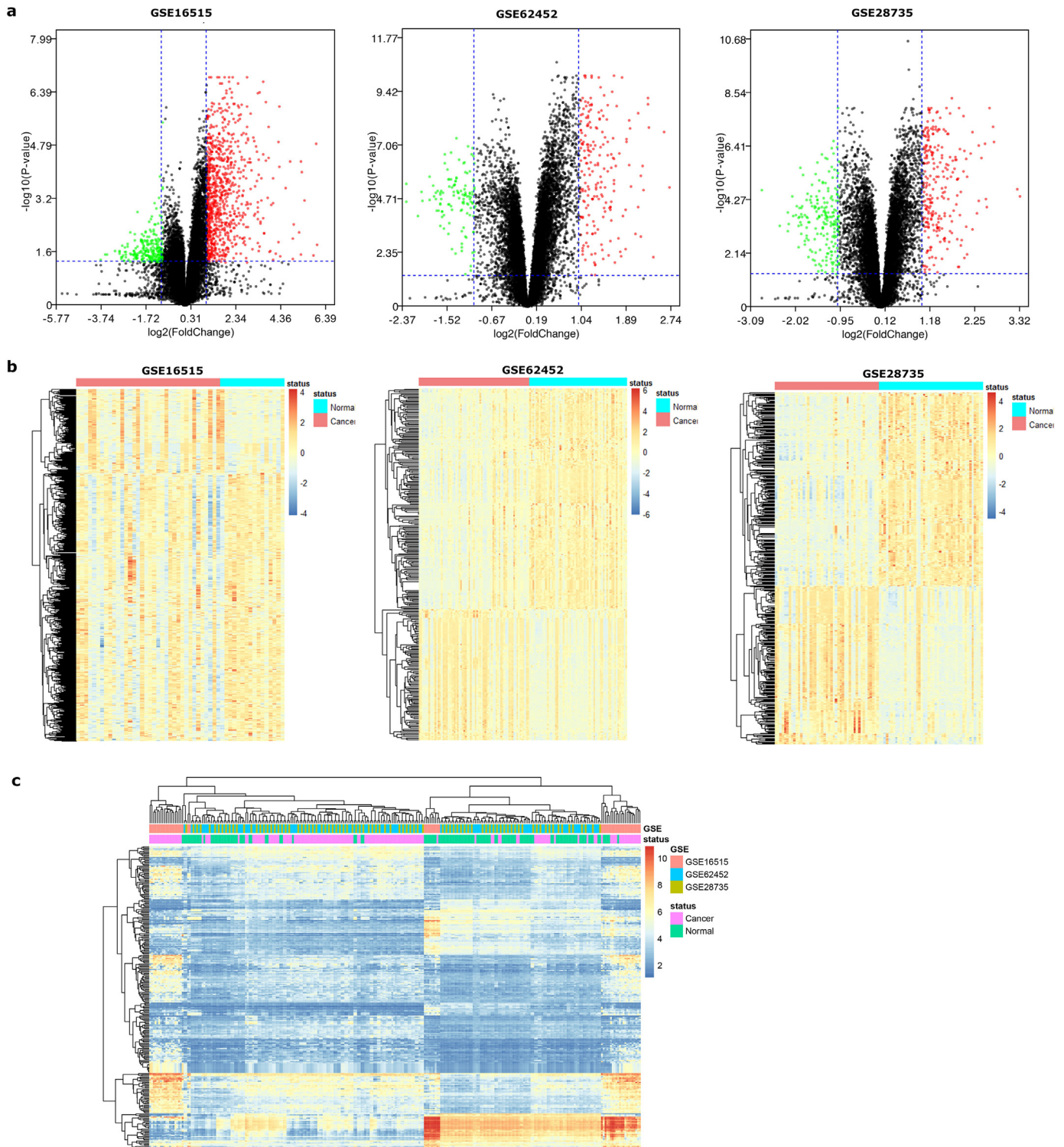
Fig. 1. Identification of differentially expressed genes in PC and data integration. (A) Volcano plot of genome-wide gene expression profiles in PC and adjacent normal tissues from GSE16515, GSE62452 and GSE28735. Red plots represent aberrantly expressed mRNAs with P < 0.05 and absolute $\log_2$FC > 1. Black plots represent normally expressed mRNAs. Green plots represent aberrantly expressed mRNAs with P < 0.05 and $\log_2$FC < −1. The abscissa shows the value of fold change in gene expression between tumors and normal tissues. The ordinate means the −log10 of the adjusted *P* value for each gene. (B) Heatmap analysis of differential expression profiles between normal tissues and cancer tissues from the three GEO databases. DEGs were defined with p < 0.05 and |$\log_2$FC| > 1. (C) The cluster heatmap of 389 consistently expressed DEGs from integrated analyses of the three GEO datasets. The normalized expression values are represented in shades of red and green, indicating expression above and below the median expression value across all tissues, respectively.
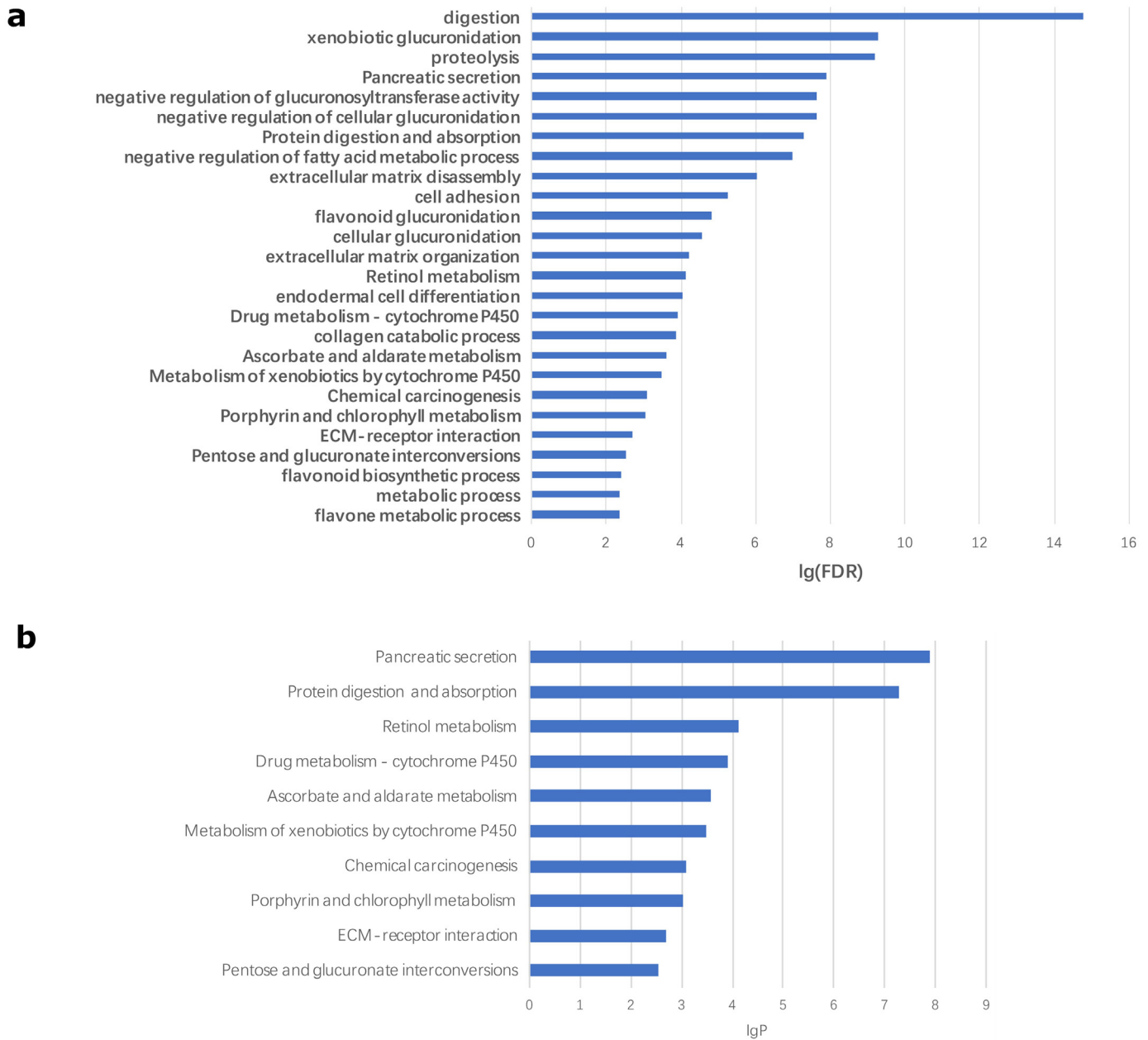
a

b

Fig. 2. Significantly enriched pathway terms of DEGs in pancreatic carcinoma. (A) DEG functional and signaling pathway enrichment were conducted using online websites for GO and KEGG pathways (B). Significantly enriched GO terms of DEGs in PC are shown based on their functions. P < 0.05 was set as the threshold. The false discovery rate (FDR) was controlled at the 0.01 threshold.

indicate that the energy metabolism pathway of PC may have changed. In addition, collagen secretion has also undergone major changes in PC.

Table 1
Top ten DEGs identified with Fisher's exact test.

Gene symbol	Official full name	P-value of fisher test
TMPRSS4	transmembrane protease, serine 4	8.58E-17
SERPINB5	serpin family B member 5	3.03E-16
LOC102725207	annexin A8-like	4.51E-16
SCEL	sciellin	4.51E-16
TNS4	tensin 4	4.51E-16
SLC6A14	solute carrier family 6 member 14	5.15E-16
TMC7	transmembrane channel like 7	7.07E-16
SLC2A1	solute carrier family 2 member 1	2.51E-15
CENPF	centromere protein F	3.05E-15
ITGB4	integrin subunit beta 4	3.05E-15

3.3. Identification of key genes for diagnosis of PC

First, key candidate genes were identified from the 389 DEGs with Fisher's exact test. The top 10 genes were selected (Table 1). To eliminate the influence of platform effect, YuGene was applied to standardize data. Hierarchical clustering of the integrated data shows that these 10 genes can distinguish PC from normal pancreatic tissue (Fig. 3a).

As these 10 genes are prominently expressed in PC, we performed a ROC curve analysis to evaluate their sensitivity and specificity for the diagnosis of PC. As shown in Fig. 3b, TMPRSS4, SERPINB5, SLC6A14, SCEL and TNS4 achieved an AUC value of >0.85, demonstrating that these genes have high sensitivity and specificity for PC diagnosis. The results suggested that TMPRSS4, SERPINB5, SLC6A14, SCEL and TNS4 can be used as biomarkers for the diagnosis of PC.
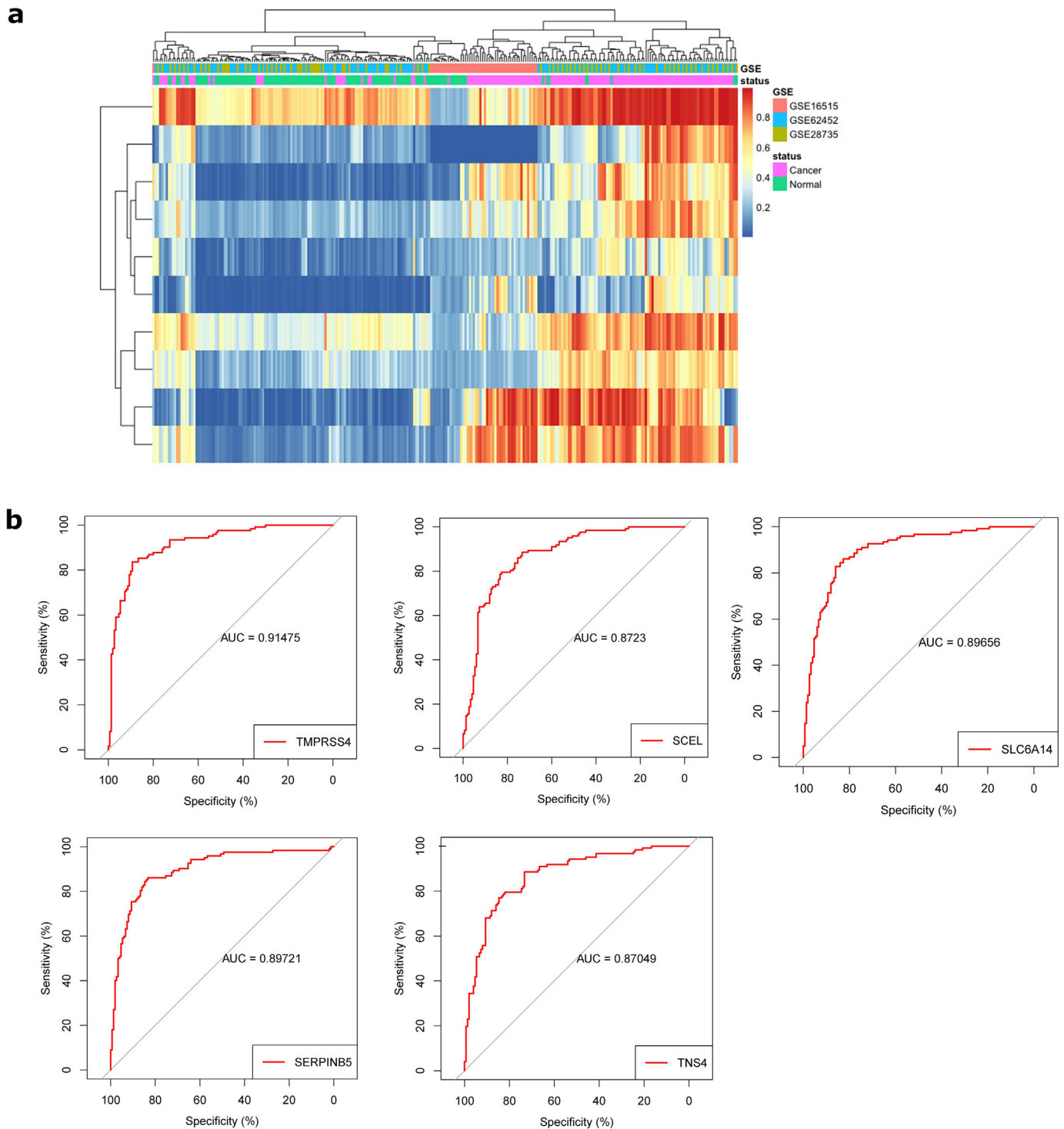
a

b

Fig. 3. Identification of key genes for diagnosis of PC. (A) Hierarchical clustering analysis of the top ten DEGs identified by Fisher's exact test. (B) ROC analysis of the top ten DEGs. Genes with an AUC value of >0.85 are shown.

3.4. Survival analysis of TCGA data

TCGA pancreatic cancer dataset contains transcriptome sequencing data and clinical data from 177 pancreatic cancer patients. Survival analysis of the top 10 genes was performed using TCGA data. As shown in Fig. 4a, TMC7, TMPRSS4, SCEL, SLC2A1, CENPF, SERPINB5 and SLC6A14 are closely related to the prognosis of patients. High expression of the seven genes is significantly associated with poor prognosis. This suggested that these seven genes may be potential

biomarkers for the prognosis of PC. As the combination of different markers as prognostic indices appears promising, we further evaluated the prognostic value of the combination of the seven genes for PC. A Univariate Cox regression model was implemented on TCGA dataset using the seven genes (Fig. 4b). The patients were classified into a high-risk group or low-risk group using the Cox regression model. The Kaplan-Meier survival analysis showed that patients in high-risk group have significantly shorter survival time compared with those in low-risk group (Fig. 4c). The comprehensive analysis
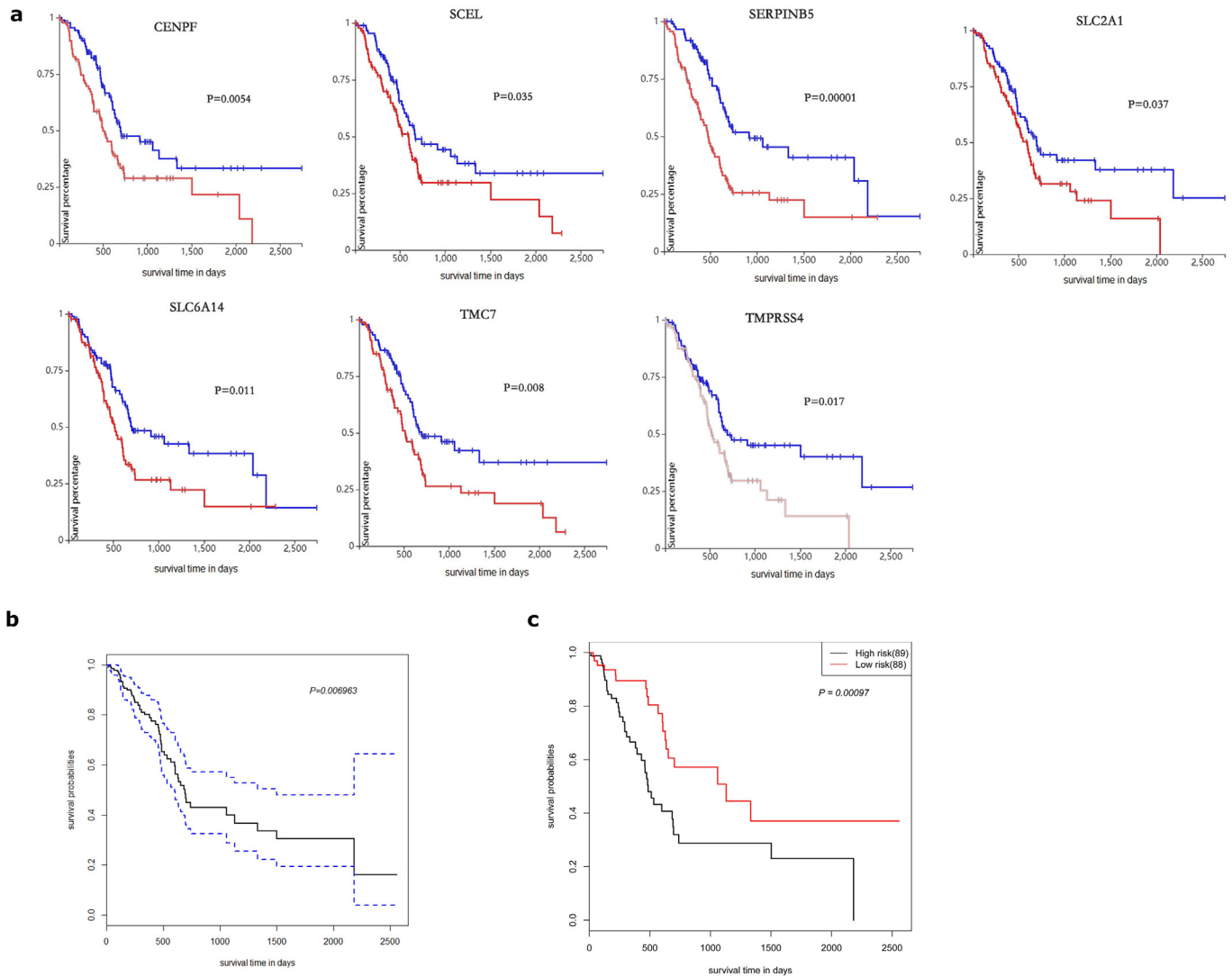
Fig. 4. Prognostic value of DEGs for PC in TCGA dataset. (A) Kaplan-Meier survival curves for patients of PC with high and low indicated gene expression in TCGA dataset. (B) The Univariate Cox regression model on TCGA dataset using the seven genes. (C) Kaplan-Meier survival curves for patients in the high-risk group and low-risk group using the Cox regression model based on TCGA data. *P*-Values were calculated using the log-rank test.

indicated that a combination of the seven genes can predict the prognosis of PC.

3.5. Validation of the DEGs in clinical tissue samples

To confirm the reliability of the identified DEGs (TMC7, TMPRSS4, SCEL, SLC2A1, CENPF, SERPINB5 and SLC6A14), we detected protein expression of the seven genes in 99 clinical PC tissues and 71 paired normal pancreatic tissues using IHC. The results showed that all seven proteins are significantly overexpressed in tumor tissues when compared with those in normal tissues (Fig. 5a & b, t-test, p < 0.05). We also noticed that the expression of SLC6A14, TMPRSS4, SLC2A1(GLUT-1) and SERPINB5 (MASPin) was significantly higher in patients of stage II than those of stage I (Fig. 5c, one-way ANOVA, p < .05). We concluded that the four genes may increase with PC progression during early stage.

3.6. Comprehensive analysis to evaluate the role of the seven proteins in PC prognosis prediction

We have proved that the mRNA level of TMC7, TMPRSS4, SCEL, SLC2A1, CENPF, SERPINB5 and SLC6A14 significantly associated with survival time in TCGA dataset. Following, we conducted a retrospective clinical study to determine whether the seven proteins can predict the prognosis of patients with PC. Survival analysis of the 99 clinical PC patients was performed. As shown in Fig. 6a, patients with high levels of protein expression of TMC7, TMPRSS4, SCEL, SLC2A1, CENPF, SERPINB5 and SLC6A14 in tumor tissue had significantly shorter survival times compared to those with low level protein expression. Consistent with the results of the genes, the results indicated that the seven proteins can also predict prognosis of PC.

Moreover, we evaluated the prognostic value of the combination of the seven proteins for PC. Clinical patients were grouped into either a high-risk group or low-risk group according to the expression levels of the seven proteins. Those with high expression of more than four of the seven proteins were assigned to the high-risk group, and the remaining patients were assigned to the low-risk group. Survival curve of the two groups showed that patients in the high-risk group had significantly shorter survival times than those in the low-risk group (Fig. 6b). The result indicated that the combination of the seven protein can also serve as a promising indicator for predicting the prognosis of patients with PC. The consistency of comprehensive analysis for the seven genes and proteins suggests that the combination of the identified candidate genes is reliable for prognosis evaluation of PC.
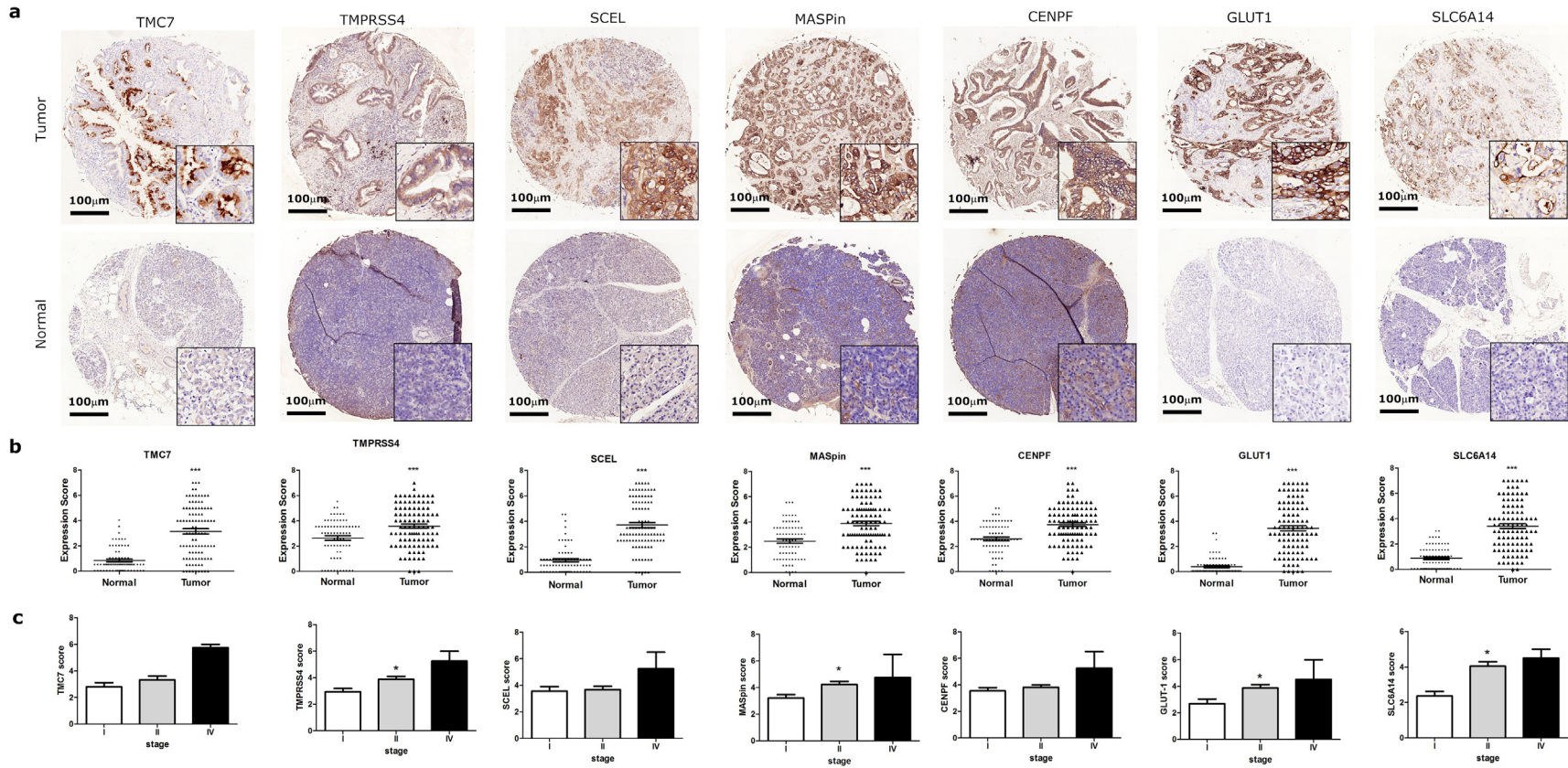
Fig. 5. Differentially expressed proteins in human PC tissue and normal pancreatic tissue. (A) The protein expression of CENPF, SCEL, SERPINB5 (MASPin), SLC2A1(GLUT-1), SLC6A14, TMC7 and TMRSS4 in clinical human PC tissue and normal tissue was detected by IHC. Representative photos are shown (100× and 400×). Scale bar = 100 μm. (B) Protein expression scores in PC tissue and normal pancreatic tissue are shown. Significance tested by t-test, ***p < 0.001 versus normal tissue. (C) The expression of indicated proteins in PC tissue from patients at different stages. Significance tested by one-way ANOVA, * p < 0.05 versus Stage I. All data are represented by mean ± SEM.
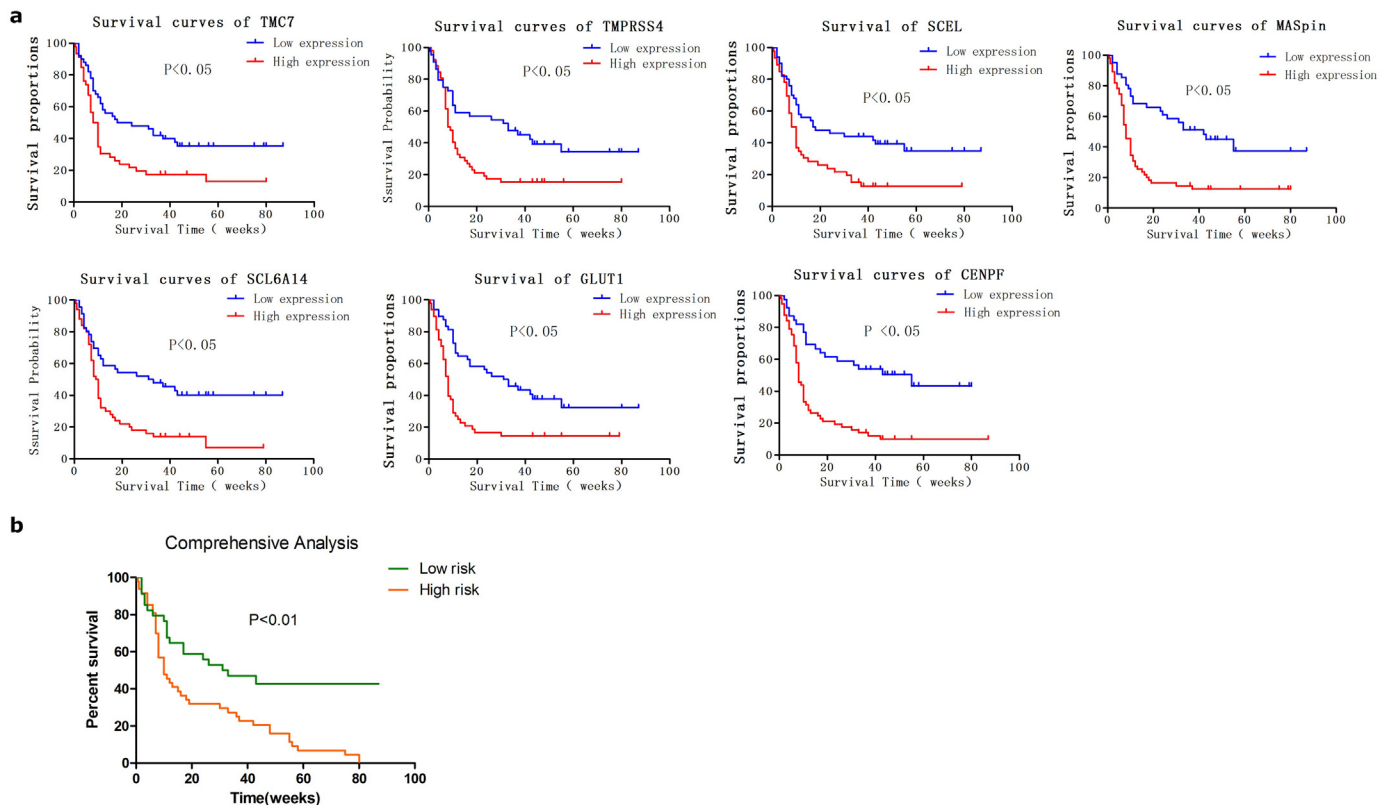
Fig. 6. Prognostic value of differentially expressed proteins for clinical pancreatic carcinoma. (A) Kaplan-Meier Survival curves for clinical PC patients with high and low expression of CENPF, SCEL, SERPINB5 (MASpin), SLC2A1(GLUT-1), SLC6A14, TMC7 and TMRSS4. (B) Survival curves for patients in the high-risk group and low-risk group based on expression scores of the differentially expressed proteins. Patients with high expression of more than four of the seven proteins were assigned to the high-risk group. The remaining patients were assigned to the low-risk group. P-Values were calculated using the log-rank test.

3.7. Knock down of the differentially expressed proteins inhibited clonability and invasion of PC cells

The potential roles of some of the identified proteins in PC biology are not clear yet. We performed soft agar colony formation assay and transwell assay to investigate the effect of the seven proteins on the clonability and invasiveness of the PC cells. Human pancreatic cancer cell lines were transient transfected with shRNA of the seven genes. The transfection efficiency was verified by western blot (Fig. 7a). Soft agar assay show that the number of cell clones was lower in the TMC7 knockdown group compared with that in the control group (Fig. 7b, one-way ANOVA, $p < 0.05$). Similar results have been observed in other proteins. As shown in Fig. 7c (one-way ANOVA, $p < 0·05$), more cells that had passed through the membrane cells were observed in control group while less cells were observed in the TMC7 knock down group. Similar results have been observed in CENPF, SCEL, SLC6A14, SLC2A1 and TMPRSS4. However, knockdown of SERPINB5 did not change the number of transmembrane cells (Data not shown). The results showed that knock down of the proteins can markedly inhibited clonability and invasive ability of PC cells. This suggested that these proteins could enhance the clonability and invasive ability of PC cells.

4. Discussion

PC is an aggressive disease that shows few symptoms until the cancer is advanced. Carbohydrate antigen 19–9 is the only diagnostic marker approved by the FDA but its diagnostic potential is limited due to its restricted sensitivity and specificity [25,26]. Since early diagnosis is still difficult, promising new biomarkers that help to allow an early diagnosis are needed. The high rate of positive lymph nodes and surgical margins at the time of resection has prompted investigators to evaluate preoperative chemoradiotherapy approaches. Prognostic biomarkers might provide essential information regarding personalized treatment decisions for individual patients. Novel biomarkers that help to estimate patients' prognosis are also needed.

In this study, genes that were consistently differentially expressed in normal pancreas tissues and tumor tissues across three GEO database were statistically analyzed. Hierarchical clustering analysis of the top ten DEGs showed that these genes can reflect the difference between normal and tumor tissues. ROC curve analysis indicated that five of these genes (TMPRSS4, SERPINB5, SLC6A14, SCEL, TNS4) may be potential molecular markers for the diagnosis of PC with relatively high sensitivity and specificity.

Survival analysis of TCGA data indicated that CENPF, SCEL, SERPINB5, SLC2A1, SLC6A14, TMC7 and TMPRSS4 is related to the prognosis of patients. The expression of these genes and the prognostic prediction value was further validated in clinical samples. All tested proteins were significantly overexpressed in tumor tissue, showing the same expression pattern as that obtained from database analysis. The expression of SLC6A14, TMPRSS4, SLC2A1(GLUT-1) and SERPINB5 (MASPin) was significantly higher in patients of stage II than those of stage I. We concluded that the four genes may increase with PC progression during early stage. However, due to limited number of patients of late stage in our study, it is uncertain whether the conclusion can be extrapolated to late stage. Survival analysis show that high expression of these proteins was related to poor prognosis.

Although single biomarker for prognosis has been widely reported, the robustness is a major concern. We further explored the role of the combination of the seven genes/proteins in prognosis prediction in PC. Comprehensive analysis of the seven genes and proteins showed that a combination of the seven genes/proteins can be used to predict the prognosis of patients. Altogether, the optimized seven genes/proteins-based model is a valuable and robust model in predicting the survival of PC.
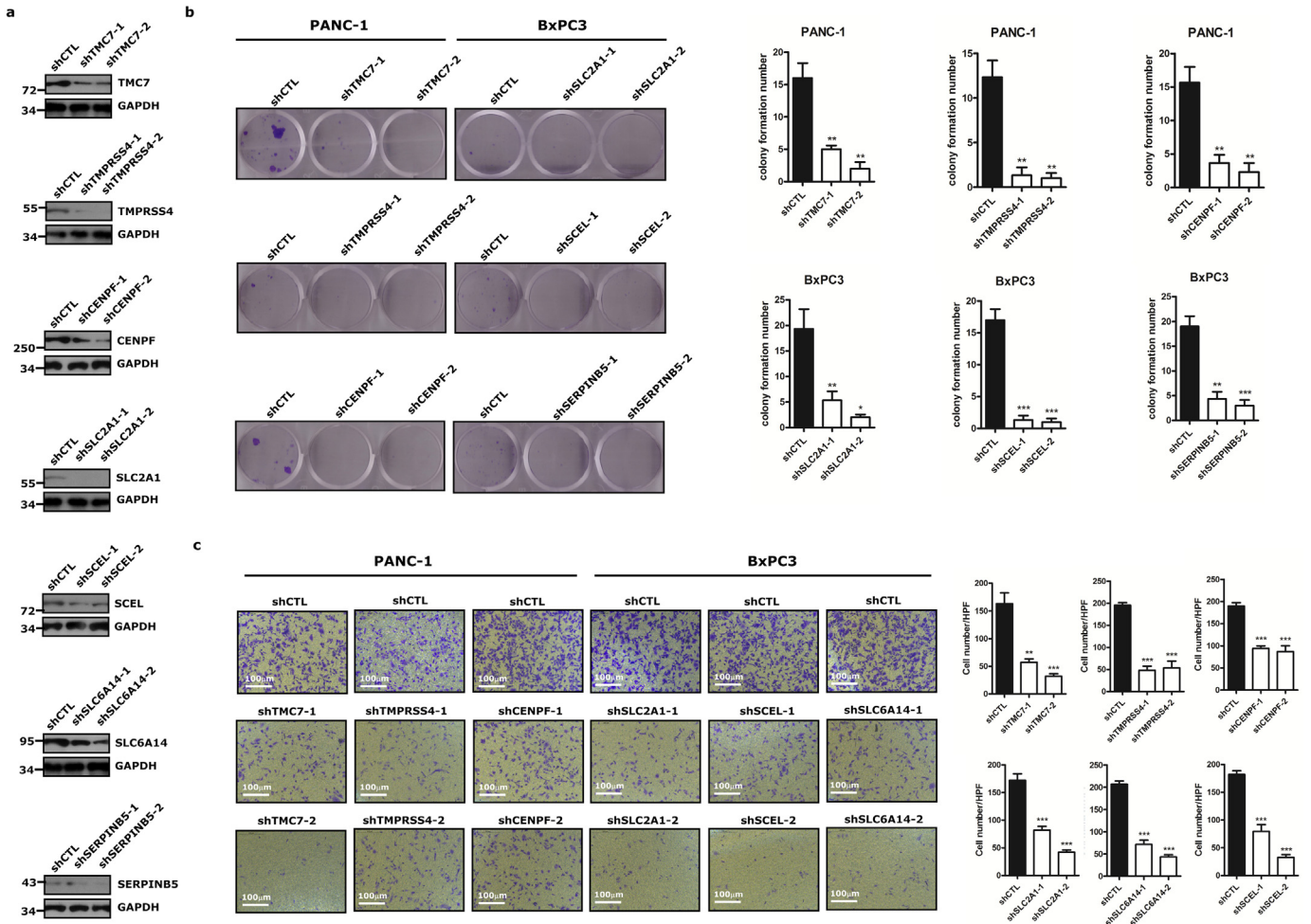
Fig. 7. Targeted genes/proteins contribute to clonability and invasiveness of PC cells in vitro. (A) PANC-1 cells were transient transfected with shRNA of TMC7, TMPRSS4, CENPF and control shRNA (shCTL). BxPC3 cells were transient transfected with shRNA of SLC2A1, SCEL, SERPINB5, SLC6A14 and shCTL. Western blot was used to detect the transfection efficiency. (B) Colony formation test of treated PANC-1 and BxPC3 cells was done. (C) Transwell assay in specific shRNA transfected or control PANC-1 and BxPC3 cells. Cells that had passed through the membrane were counted. Representative images are shown. Significance tested by one-way ANOVA, **$p < 0.01$; ***$p < 0.001$ versus shCTL. All data are represented by mean ± SEM.

The majority of PCs start in the exocrine cells, which produce the digestive juices. Altered metabolism is a critical hallmark of PC. Our study's functional enrichment and PPI network analyses show that the occurrence of PC is related to pathways of pancreatic secretion, protein digestion and absorption, and aldehyde peroxidation. These pathways mainly affect the biological processes of substance absorption and energy metabolism. In addition, PPI network analysis suggested that the changes in the secretion of collagen may also be closely related to the occurrence of pancreatic cancer.

Our preliminary in vitro study showed that TMC7, SCEL, SLC2A1, TMPRSS4, CENPF, SERPINB5 and SLC6A14 contributes to clonal formation and invasiveness in PC. This indicated that these genes could also be exploited as potential therapeutic targets for PC.

TMC7 is a member of a gene family predicted to encode transmembrane proteins [27]. The specific function of TMC7 in both normal cells and tumor cells is unknown at present. In the present study, we provided the first evidence that TMC7 is overexpressed in PC and contribute to tumor progression and metastasis.

SCEL and TMPRSS4 have been reported to overexpressed and may associated with poor prognosis in several kinds of tumors [28–31]. Studies have showed that SCEL is highly expressed in late stage primary colorectal cancer and mediates epithelial-mesenchymal transition (EMT) in colorectal cancer hepatic metastasis by activating wnt/β-catenin signal pathway [32]. Existing evidences have reported that TMPRSS4 promotes cell migration and metastasis by facilitating EMT in lung and

colon cancer cell lines [33]. High levels of TMPRSS4 have been found in PC. However, the specific role of SCEL and TMPRSS4 in the progression and prognosis of PC remains unknown. In the present study, we showed that the SCEL is overexpressed in PC. Our study also supports the notion of SCEL and TMPRSS4 as potential prognosis markers and promote tumor progression and metastasis in PC. It reminds us that SCEL and TMPRSS4 may promote the tumor progression and metastasis via inducing EMT in PC where further validation is needed.

SLC2A1 encodes for uniport protein GLUT-1, a member of glucose transporter family. GLUT-1 plays a significant role in glucose metabolism in malignant tumor. Down-regulation of GLUT-1 results in an inhibited Warburg effect in tumor cells [29]. Our study showed that GLUT-1 is a prognostic factor for overall survival of PC which is consistent with Yu M et al.'s report [34,35]. Moreover, we showed that knocking down of GLUT-1 inhibited the proliferation of PC cells. Given that PC differs greatly from normal pancreatic tissue in several glucose uptake and metabolism pathways, we speculate that GLUT-1 probably supports the growth of PC by regulating Warburg effect, although this inference requires further verification.

SLC6A14 is found to be overexpressed in PC [36]. Pharmacologic blockade of SLC6A14 can induce amino acid starvation in PC cells and reduce tumor cell growth and proliferation [37]. Similarly, our study showed that knockdown of SLC6A14 reduced cell colony and metastasis in PC. SERPINB5 encodes Maspin, which is an epithelial-specific member of the serpin superfamily. Maspin is involved in several biological

behaviors of tumor. Knocking down of Maspin elevates the gemcitabine sensitivity of hormone-independent prostate cancer [38]. Maspin is also involved in the angiogenesis of grem cell tumor via p53 pathway [39]. Aberrant SERPINB5 mRNA expression in pulmonary adenocarcinoma is associated with tumor metastasis and poor clinical outcome [40]. SERPINB5 also seems to assume an oncogenic role in PC [41]. Mardin WA et al. showed that SERPINB5 expression was correlated with increased metastasis and provide a diagnostic tool for PC [42] which consistent with the results obtained in our study.

CENPF, a structural protein of kinetochore, is also associated with tumor growth in a variety of cancers. CENPF is frequently overexpressed in hepatocellular carcinoma and silencing CENPF resulted in the cell cycle arrest [43]. CENPF is associated with prognosis of prostate cancer [44] and acts synergistically with FOXM1 to regulate target gene expression and activate key signaling pathways associated with tumor malignancy [45]. Computational analyses have shown that CENPF is tightly related to the progression of PC [46]. Survival analyses of clinical patients in our study further validate the result.

In this study, we synthesized several sets of public data and clinical data. Candidate molecular markers were identified for diagnosis and prognosis prediction of PC. A novel gene, TMC7, that is overexpressed in PC was identified. The role of the candidate differentially expressed proteins in PC cells including TMC7 was preliminary explored. Moreover, we found that the combination of the markers can serve as a promising prognostic indicator in patients with PC. This research provides a novel insight into molecular mechanisms, diagnostic and prognostic markers of PC.

Given the limited benefit of therapies, treatment decision-making in PC can be challenging [47]. Moreover, the accuracy of tumor staging assessment is limited by the sensitivity and specificity of imaging. In order to decide which treatment plan is appropriate for a patient, it is of utmost importance to weigh tumor staging and clinical circumstances such as life expectancy [48]. For example, neoadjuvant chemotherapy is recommended for resectable PC with high risk factors even without radiological evidence of metastases [49,50]. Our study suggested that a combination of the identified genes/proteins can be used as an indicator for risk stratification that help to predict the prognosis of PC. The identified prognostic biomarkers might provide essential information regarding personalized treatment decisions for individual patients and improve the therapeutic gain.

As these genes are overexpressed in tumor cells and tumor tissues, we speculate that the change of the expression of these genes may also exist in circulating tumor cells. Detection of the expression of these genes in circulating tumor cells is expected to make an early diagnosis of PC. We plan to carry out the relevant clinical research in this aspect in the following study. Moreover, a more detailed molecular mechanism analysis for these genes will be carried out to clarify their role in promoting PC progression and metastasis.

Research support

Declaration of interests

We declare no competing interests.

Author contributions

Conceived and designed the study: Yun Zhu, Sitang Gong. Performed the experiments: Yang Cheng, Kunyuan Wang, Lanlan Geng, Jingjing Sun. Wrote the paper: Yang Cheng and Yun Zhu. Analyzed the data: Wanfu Xu, Dingli liu. All authors read and approved the final manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.01.003.

References

[1] Abel EV, Simeone DM. Biology and clinical applications of pancreatic cancer stem cells. Gastroenterology 2013 Jun;144(6):1241–8.

[2] McGuire S. World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. Advances in nutrition (Bethesda, Md). 2016 Mar;7(2):418–9. .

[3] Wolfgang CL, Herman JM, Laheru DA, Klein AP, Erdek MA, Fishman EK, et al. Recent progress in pancreatic cancer. CA Cancer J Clin 2013 Sep;63(5):318–48.

[4] Maeda S, Shinchi H, Kurahara H, Mataki Y, Noma H, Maemura K, et al. Clinical significance of midkine expression in pancreatic head carcinoma. Br J Cancer 2007 Aug 6; 97(3):405–11.

[5] Harsha HC, Kandasamy K, Ranganathan P, Rani S, Ramabadran S, Gollapudi S, et al. A compendium of potential biomarkers of pancreatic cancer. PLoS Med 2009 Apr 7;6 (4):e1000046.

[6] Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature 2010 Oct 28;467(7319): 1114–7.

[7] Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. Nat Clin Pract Oncol 2008 Oct;5(10): 588–99.

[8] Bhalla S, Chaudhary K, Kumar R, Sehgal M, Kaur H, Sharma S, et al. Gene expression-based biomarkers for discriminating early and late stage of clear cell renal cancer. Sci Rep 2017 Mar 28;7:44997.

[9] Ho TH, Serie DJ, Parasramka M, Cheville JC, Bot BM, Tan W, et al. Differential gene expression profiling of matched primary renal cell carcinoma and metastases reveals upregulation of extracellular matrix genes. Anna Oncol 2017 Mar 1;28(3):604–10.

[10] Malouf GG, Ali SM, Wang K, Balasubramanian S, Ross JS, Miller VA, et al. Genomic characterization of renal cell carcinoma with sarcomatoid dedifferentiation pinpoints recurrent genomic alterations. Eur Urol 2016 Aug;70(2):348–57.

[11] Yang S, He P, Wang J, Schetter A, Tang W, Funamizu N, et al. A Novel MIF signaling pathway drives the malignant character of pancreatic cancer by targeting NR3C2. Cancer Res 2016 Jul 1;76(13):3838–50.

[12] Zhang G, He P, Tan H, Budhu A, Gaedcke J, Ghadimi BM, et al. Integration of metabolomics and transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer. Clin Cancer Res 2013 Sep 15;19(18):4983–93.

[13] Pei H, Li L, Fridley BL, Jenkins GD, Kalari KR, Lingle W, et al. FKBP51 affects cancer cell response to chemotherapy by negatively regulating Akt. Cancer Cell 2009 Sep 8;16 (3):259–66.

[14] Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics (Oxford, England) 2003 Jan 22;19(2):185–93.

[15] Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics (Oxford, England) 2012 Feb 15;28(4):573–80.

[16] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Gene Ontol Consortium Nat Genetics 2000 May;25 (1):25–9.

[17] Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000 Jan 1;28(1):27–30.

[18] Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4(1):44–57.

[19] Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009 Jan; 37(1):1–13.

[20] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2013 Jan;41(Database issue):D808–15.

[21] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003 Nov;13(11):2498–504.

[22] Le Cao KA, Rohart F, McHugh L, Korn O, Wells CA. YuGene: a simple approach to scale gene expression data derived from different platforms for integrated analyses. Genomics 2014 Apr;103(4):239–51.

[23] Cheng Y, Zhu Y, Xu W, Xu J, Yang M, Chen P, et al. PKCalpha in colon cancer cells promotes M1 macrophage polarization via MKK3/6-P38 MAPK pathway. Mol Carcinog 2018 Aug;57(8):1017–29.

[24] Zhu Y, Cheng Y, Guo Y, Chen J, Chen F, Luo R, et al. Protein kinase D2 contributes to TNF-alpha-induced epithelial mesenchymal transition and invasion via the PI3K/ GSK-3beta/beta-catenin pathway in hepatocellular carcinoma. Oncotarget 2016 Feb 2;7(5):5327–41.

[25] Slesak B, Harlozinska-Szmyrka A, Knast W, Sedlaczek P, van Dalen A, Einarsson R. Tissue polypeptide specific antigen (TPS), a marker for differentiation between pancreatic carcinoma and chronic pancreatitis. A comparative study with CA 19-9. Cancer 2000 Jul 1;89(1):83–8.

[26] Su SB, Qin SY, Chen W, Luo W, Jiang HX. Carbohydrate antigen 19-9 for differential diagnosis of pancreatic carcinoma and chronic pancreatitis. World J Gastroenterol 2015 Apr 14;21(14):4323–33.

[27] Hahn Y, Kim DS, Pastan IH, Lee B. Anoctamin and transmembrane channel-like proteins are evolutionarily related. Int J Mol Med 2009 Jul;24(1):51–5.

[28] Corona W, Karkera DJ, Patterson RH, Saini N, Trachiotis GD, Korman LY, et al. Analysis of Sciellin (SCEL) as a candidate gene in esophageal squamous cell carcinoma. Anticancer Res 2004 May-Jun;24(3a):1417–9.

[29] Szablewski L. Expression of glucose transporters in cancers. Biochim Biophys Acta 2013 Apr;1835(2):164–9.

[30] Takahashi T, Uehara H, Izumi K. Inhibitory effect of soluble EP2 receptor on ovarian tumor growth in nude mice and utility of TMPRSS4 as a combinatorial molecular target. Int J Oncol 2013 Aug;43(2):416–24.

[31] Larzabal L, Nguewa PA, Pio R, Blanco D, Sanchez B, Rodriguez MJ, et al. Overexpression of TMPRSS4 in non-small cell lung cancer is associated with poor prognosis in patients with squamous histology. Br J Cancer 2011 Nov 8;105(10): 1608–14.

[32] Chou CK, Fan CC, Lin PS, Liao PY, Tung JC, Hsieh CH, et al. Sciellin mediates mesenchymal-to-epithelial transition in colorectal cancer hepatic metastasis. Oncotarget 2016 May 3;7(18):25742–54.

[33] Jung H, Lee KP, Park SJ, Park JH, Jang YS, Choi SY, et al. TMPRSS4 promotes invasion, migration and metastasis of human tumor cells by facilitating an epithelial-mesenchymal transition. Oncogene 2008 Apr 17;27(18):2635–47.

[34] Nagarajan A, Dogra SK, Sun L, Gandotra N, Ho T, Cai G, et al. Paraoxonase 2 Facilitates Pancreatic Cancer Growth and Metastasis by Stimulating GLUT1-Mediated Glucose Transport. Mol Cell 2017 Aug 17;67(4):685–701 e6.

[35] Yu M, Zhou Q, Zhou Y, Fu Z, Tan L, Ye X, et al. Metabolic phenotypes in pancreatic cancer. PloS One 2015;10(2):e0115153.

[36] Karunakaran S, Umapathy NS, Thangaraju M, Hatanaka T, Itagaki S, Munn DH, et al. Interaction of tryptophan derivatives with SLC6A14 (ATB0,+) reveals the potential of the transporter as a drug target for cancer chemotherapy. Biochem J 2008 Sep 15;414(3):343–55.

[37] Coothankandaswamy V, Cao S, Xu Y, Prasad PD, Singh PK, Reynolds CP, et al. Amino acid transporter SLC6A14 is a novel and effective drug target for pancreatic cancer. Br J Pharmacol 2016 Dec;173(23):3292–306.

[38] Huang CY, Chang YJ, Luo SD, Uyanga B, Lin FY, Tai CJ, et al. Maspin mediates the gemcitabine sensitivity of hormone-independent prostate cancer. Tumour Biol 2016 Mar;37(3):4075–82.

[39] Celik H, Turunc T, Bal N, Hasirci E, Akay A, Peskircioglu CL. Expression of maspin in testis tumors with germ cells andits relation with angiogenesis factors. Turkish J Med Sci 2016 Jun 23;46(4):1197–202.

[40] Lu M, Li J, Huang Z, Du Y, Jin S, Wang J. Aberrant Maspin mRNA Expression is Associated with Clinical Outcome in patients with Pulmonary Adenocarcinoma. Med Sci Monitor 2016 Jan 13;22:134–9.

[41] Mardin WA, Ntalos D, Mees ST, Spieker T, Senninger N, Haier J, et al. SERPINB5 promoter hypomethylation differentiates pancreatic ductal adenocarcinoma from pancreatitis. Pancreas 2016 May-Jun;45(5):743–7.

[42] Mardin WA, Petrov KO, Enns A, Senninger N, Haier J, Mees ST. SERPINB5 and AKAP12 - expression and promoter methylation of metastasis suppressor genes in pancreatic ductal adenocarcinoma. BMC Cancer 2010 Oct 12;10:549.

[43] Dai Y, Liu L, Zeng T, Zhu YH, Li J, Chen L, et al. Characterization of the oncogenic function of centromere protein F in hepatocellular carcinoma. Biochem Biophys Res Commun 2013 Jul 12;436(4):711–8.

[44] Zhuo YJ, Xi M, Wan YP, Hua W, Liu YL, Wan S, et al. Enhanced expression of centromere protein F predicts clinical progression and prognosis in patients with prostate cancer. Int J Mol Med 2015 Apr;35(4):966–72.

[45] Aytes A, Mitrofanova A, Lefebvre C, Alvarez MJ, Castillo-Martin M, Zheng T, et al. Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. Cancer Cell 2014 May 12;25(5):638–51.

[46] Zhou Z, Cheng Y, Jiang Y, Liu S, Zhang M, Liu J, et al. Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis. Int J Biol Sci 2018;14(2):124–36.

[47] Clarke G, Johnston S, Corrie P, Kuhn I, Barclay S. Withdrawal of anticancer therapy in advanced disease: a systematic literature review. BMC Cancer 2015 Nov 11;15:892. PubMed PMID: 26559912.

[48] Zijlstra M, van der Geest LGM, van Laarhoven HWM, Lemmens V, van de Poll-Franse LV, Raijmakers NJH. Patient characteristics and treatment considerations in pancreatic cancer: a population based study in the Netherlands. Acta Oncol (Stockholm, Sweden) 2018 Sep;57(9):1185–91.

[49] Tempero MA, Malafa MP, Al-Hawary M, Asbun H, Bain A, Behrman SW, et al. Pancreatic Adenocarcinoma, Version 2.2017, NCCN Clinical Practice guidelines in Oncology. J Natl Comprehen Cancer Network 2017 Aug;15(8):1028–61.

[50] Casciani F, Marchegiani G, Malleo G, Bassi C, Salvia R. Pancreatic Cancer in the Era of Neoadjuvant Therapy: A Narrative Overview. Chirurgia (Bucharest, Romania : 1990) 2018 May-Jun;113(3):307–17.