Identification of potential hub genes associated with the pathogenesis and prognosis of pancreatic duct adenocarcinoma using bioinformatics meta-analysis of multi-platform datasets

YUFAN MA[*], YINQUAN PU[*], LI PENG, XUJUAN LUO, JIN XU, YAN PENG and XIAOWEI TANG

Department of Gastroenterology, The Affiliated Hospital of Southwest Medical University,
Luzhou, Sichuan 646099, P.R. China

Abstract. Pancreatic duct adenocarcinoma (PDAC) is a highly malignant type of cancer with a low five-year survival rate. Gene alterations are crucial to the molecular pathogenesis of PDAC. Therefore, the present study analyzed gene expression profiles to reveal genes involved in the tumorigenesis of PDAC. A total of eight gene expression profiles (GSE15471, GSE16515, GSE41368, GSE62165, GSE62452, GSE71729, GSE71989 and GSE91035) and a PDAC dataset were acquired from the Gene Expression Omnibus and The Cancer Genome Atlas (TCGA) database, respectively. Differentially expressed genes (DEGs) were screened using functional annotation, Gene Ontology (GO) enrichment analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis and protein-protein interaction (PPI) network construction. A Cox proportional hazards model was then constructed and used to analyze the data. A total of 136 DEGs (67 up- and 69 downregulated genes) were identified between PDAC tissues and normal tissues. The 'extracellular matrix-related' genes were the most enriched in the GO term analysis. 'Pancreatic secretion', 'phosphoinositide-3-kinase–protein kinase B/Akt (PI3K-Akt) signaling pathway', 'protein digestion and absorption' and 'ECM-receptor interaction' were the most enriched categories in KEGG pathway analysis. Following PPI network construction, the 10 most significant genes [albumin, epidermal growth factor, matrix metalloproteinase (MMP) 9, epidermal growth factor receptor, fibronectin 1, MMP1, plasminogen activator inhibitor-1, tissue inhibitor of metalloproteinase 1, plasminogen activator urokinase (PLAU) and PLAU receptor) exhibiting a high degree of connectivity, were identified as the hub genes likely to be associated with the pathogenesis of PDAC. In addition, a prognostic predictive system for PDAC, composed of five genes (laminin subunit γ 2, laminin subunit β 3, serpin family B member 5, amphiregulin and secreted frizzled related protein 4), was constructed. This was validated in the GSE62452 dataset (using 66 PDAC samples with outcome data) and TCGA PDAC dataset (using 146 PDAC samples with outcome data). In conclusion, the present study revealed potential hub genes involved in PDAC progression, providing directive significance for individualized clinical decision-making and molecular-targeting therapy in patients with PDAC.

Introduction

Pancreatic ductal adenocarcinoma (PDAC) is a highly complex and malignant type of cancer in humans. It is the seventh leading cause of cancer-associated mortality, and is expected to rise to the third due to its increasing incidence and poor prognosis (1). Despite considerable improvements in surgical, radiation and chemotherapeutic treatments, 80% of patients with PDAC miss the optimal period for effective systemic therapy due to a lack of symptoms, anesis or disease regression at the time of diagnosis (2). Hence, the five-year overall survival rate for PDAC remains at 3-5%. There is thus an urgent need to identify new biomarkers to improve the understanding of the molecular mechanisms involved in PDAC pathogenesis (3). Potential prognostic biomarkers and novel therapeutic targets may help to improve current poor treatment outcomes.

Microarrays have been widely used to identify more sensitive and effective biomarkers for PDAC. Shen et al (4) reported that ribosomal protein genes nucleoporin 170, nucleoporin 160 and heterogeneous nuclear ribonucleoprotein U may be useful as molecular markers for early diagnosis of the disease. Ger et al (5) reported that the fms related tyrosine kinase 3 and poly(rC) binding protein 3 could potentially be used as prognostic biomarkers for pancreatic cancer. Another analysis considered dickkopf WNT signaling pathway inhibitor 1 and high mobility group AT-hook 2 to be hub genes that are strongly associated with Wnt family member 3A and tumor protein

Correspondence to: Dr Yan Peng or Dr Xiaowei Tang, Department of Gastroenterology, The Affiliated Hospital of Southwest Medical University, 25 Taiping Street, Jiangyang, Luzhou, Sichuan 646099, P.R. China
E-mail: 1806857826@qq.com
E-mail: solitude5834@hotmail.com

[*]Contributed equally

p53, respectively (6). In addition, KRAS, TP53, CDKN2A, SMAD4, RNF43, ARID1A, TGFbR2, GNAS, RREB1 and PBRM1 were identified as driver genes in PDAC (7). Variation in the significantly expressed genes associated with PDAC pathogenesis between different studies may be due to small sample sizes, the use of different microarray platforms and different statistical methods. To overcome these limitations, integrative meta-analysis using different microarray platforms with larger sample sizes may prove to be a powerful bioinformatics tool, improving the accuracy and reliability of data analysis.

In the present study, multiple Gene Expression Omnibus (GEO) datasets containing a large number of samples were acquired from two different microarray platforms (Affymetrix and Agilent). The RobustRankAggreg (RRA) 1.1 package (8) in R is based on a statistical model that allows for the evaluation of the significance of results. It can be used to identify differentially expressed genes (DEGs) across multiple datasets from different microarray platforms. By defining the rank vector for each gene, based only on the datasets where it is present, the results include DEGs that are not present in every dataset (9). Hence, the gene expression module was investigated to reveal the genes influencing PDAC tumorigenesis.

Materials and methods

Selection and retrieval of microarray datasets. A total of 8 gene expression profiles [GSE15471 (10), GSE16515 (11), GSE41368 (12), GSE62165 (13), GSE62452 (14), GSE71729 (15), GSE71989 (16) and GSE91035 (17)] from two platforms (Affymetrix and Agilent) were retrieved from the GEO (http://www.ncbi.nlm.nih.gov/geo/) database, using the keywords 'pancreatic ductal adenocarcinoma', 'Homo sapiens' and 'microarray'. The selected microarray datasets met the following inclusion criteria: i) Expression profiling by array; ii) samples included human PDAC and corresponding adjacent or normal pancreatic tissue; iii) n>10; and iv) the gene expression profile is complete. A total of 8 microarray datasets were retrieved from two different microarray platforms. The data included 452 PDAC samples and 204 normal pancreatic tissue samples (Table I). For further validation, normalized datasets (fragments per kilobase of transcript per million mapped reads upper quartile) of 146 PDAC samples (with complete expression profiles and clinical prognoses) were retrieved from The Cancer Genome Atlas (TCGA) database (7).

Data pre-processing and DEG analysis. Initially, $\log_2$ conversion and quantile normalization was performed on each individual GEO dataset. DEGs of each dataset were then screened using the limma package (http://bioinf.wehi.edu.au/limma) with R/Bioconductor 3.9 software (http://www.bioconductor.org/). The RRA package was used for gene integration analysis of the DEGs in the eight datasets. P<0.05 was considered to indicate a statistically significant result and a fold-change (log-scaled) of mean + 2SD was set as the threshold in the limma package. In the RRA package, an adjusted P<0.05 was considered to indicate a statistically significant difference.

Functional enrichment analysis. The functions of common DEGs were further analyzed using the clusterProfiler package

(https://guangchuangyu.github.io/software/clusterProfiler) with R/Bioconductor software for functional enrichment analysis. This included the following gene ontology (GO) categories: Molecular function (MF), biological process (BP) and cellular component (CC), as well as enrichment analysis of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. P<0.05 was set as the threshold value for MF, BP and CC; and P<0.01 was set as the threshold value for KEGG analysis.

Protein-protein interaction (PPI) network construction. In the present study, the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; string-db.org) database was used to construct the PPI network of common DEGs, which was then visualized using Cytoscape (3.7.1) software (18). The Cytoscape MCODE plug-in was used to search for clustered sub-networks, and the default parameters were as follows: Degree cutoff, ≥2; node score cutoff, ≥0.2; K-core, ≥2; max depth, 100.

Prediction system construction. As the GSE62452 dataset and TCGA data contain patient survival information, they can be used as training and validation datasets, respectively. Univariate Cox proportional hazard analysis was applied to identify the prognosis-associated genes in GEO datasets (training set), using survival analysis in R, with P<0.05 set as the significance threshold. Multivariate Cox regression analysis was then applied to further screen for factors associated with patient survival. Subsequently, a prediction system was constructed consisting of five signature prognostic genes [laminin subunit γ 2 (LAMC2), laminin subunit β 3 (LAMB2), Serpin family B member 5 (SERPINB5), amphiregulin (AREG) and secreted frizzled related protein 4 (SFRP4)], and was used to construct a risk score formula. Each patient's risk score was calculated and the median risk score was regarded as the cutoff point. Patients were divided into a high- and a low-risk group, in accordance with their prognostic risk scores. Kaplan-Meier (KM) analysis [with log-rank test (Mantel-Cox)] was then performed to calculate and compare the survival time between the two groups, with P<0.05 selected to indicate a significant difference. KM curves were constructed using the R 'survival' package. Finally, receiver operating characteristic (ROC) analysis was conducted using the R 'survivalROC' package to identify the sensitivity and specificity of the prediction system. The five aforementioned prognostic signature genes were applied to TCGA dataset (validation set) to verify whether they could effectively predict the prognosis of PDAC (19).

Results

Identification of DEGs. A total of 136 DEGs (67 up- and 69 downregulated genes) were identified between PDAC tissues and normal tissues by analyzing eight gene expression profiles retrieved from the GEO database. Volcano plots of the gene expression profile data, and a heat map and histogram of DEGs across the datasets are displayed in Figs. 1-3.

Functional enrichment analysis. To elucidate the functions of common DEGs, GO and KEGG pathway enrichment analyses were performed. The GO results determined that, in the MF

Table I. Gene expression profile data characteristics.

| Author, Year	Dataset	Count	Tumor	Normal	Platform	Region	(Refs.)
Badea et al, 2009	GSE15471	78	39	39	GPL570	Romania	10
Pei et al, 2009	GSE16515	52	36	16	GPL570	USA	11
Frampton et al, 2012	GSE41368	12	6	6	GPL6244	Italy	12
Janky et al, 2014	GSE62165	131	118	13	GPL13667	Belgium	13
Yang et al, 2014	GSE62452	130	69	61	GPL6244	USA	14
Moffitt et al, 2015	GSE71729	191	145	46	GPL20769	USA	15
Schmittgen, 2015	GSE71989	22	14	8	GPL570	USA	16
Schmittgen, 2016	GSE91035	40	25	15	GPL22763	USA	17
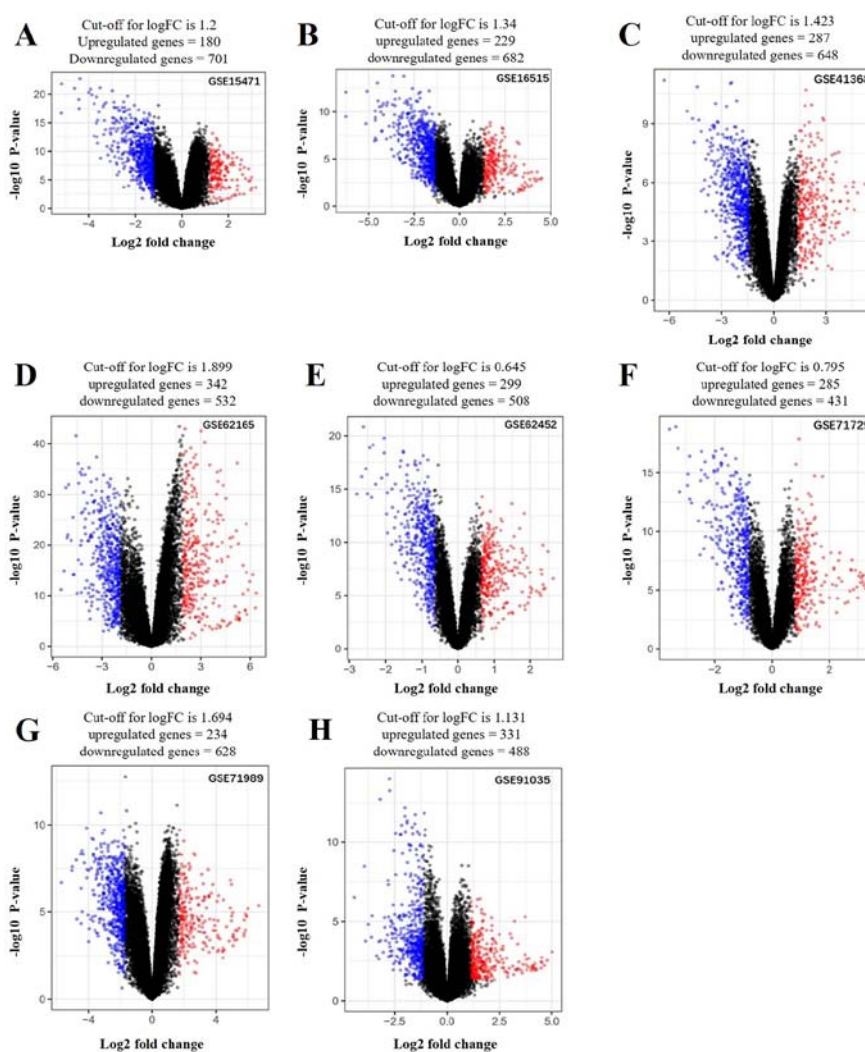
Figure 1. Volcano plots of genes that are significantly different between pancreatic tumor and adjacent non-tumor tissues. (A) GSE15471; (B) GSE16515; (C) GSE41368; (D) GSE62165; (E) GSE62452; (F) GSE71729; (G) GSE71898; (H) GSE91035. The x-axis indicates the fold-change (log-scaled); the y-axis indicates the P-values (log-scaled). Each symbol represents a different gene, and the red and blue symbols indicate upregulated and downregulated genes, respectively. $P<0.05$ was considered to be statistically significant and a fold change (log-scaled) of mean + 2SD was set as the threshold value.

category, the upregulated genes were predominantly enriched in 'extracellular matrix constituent' and 'glycosaminoglycan binding', whilst the downregulated genes were mainly enriched

in 'serine-type endopeptidase activity', 'serine-type peptidase activity' and 'serine hydrolase activity'. In the BP category, the upregulated genes were mainly enriched in 'ECM organization'
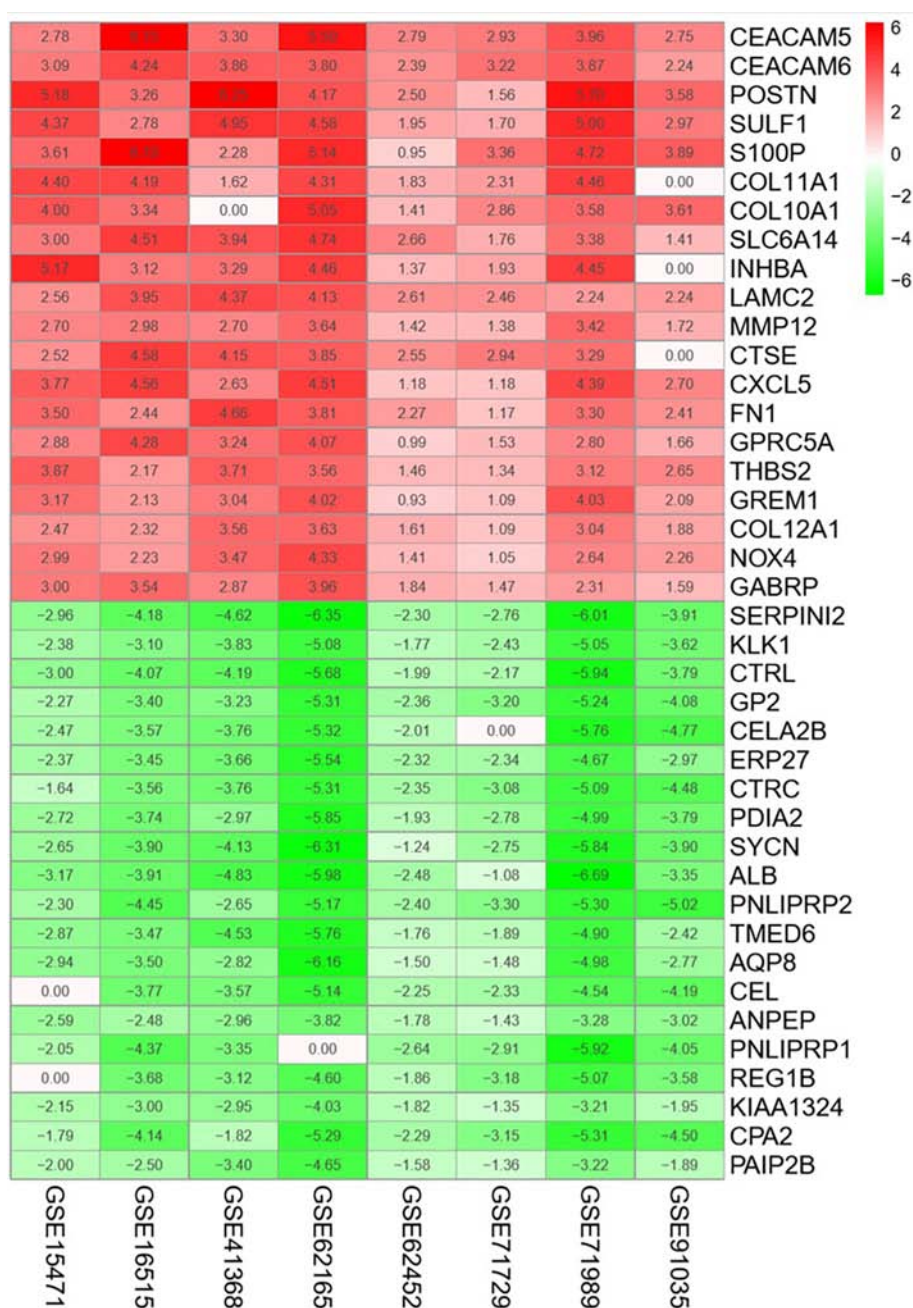
Figure 2. Heat map of the top 20 upregulated and downregulated DEGs. The gradient from green to red represents the change from down- to upregulation. DEGs, differently expressed genes.

and 'extracellular structure organization', whilst the downregulated genes were enriched in 'antimicrobial humoral response' and 'digestion'. In the CC category, the upregulated genes were mainly enriched in 'ECM' and 'collagen-containing ECM', whilst the downregulated genes were not enriched. 'Pancreatic secretion', 'phosphoinositide-3-kinase-protein kinase B/Akt (PI3K-Akt) signaling pathway', 'protein digestion and absorption' and 'ECM-receptor interaction' were the most enriched pathways in the KEGG pathway analysis. The results of the functional enrichment analysis are exhibited in Figs. 4-6 and support the results of previous studies (20,21).

Hub gene identification using PPI network construction and modular analysis. A PPI network based on the DEGs

was constructed using Cytoscape software and the STRING database (Fig. 7). The ten genes with the highest degree of connectivity [albumin (ALB), epidermal growth factor (EGF), MMP9, epidermal growth factor receptor (EGFR), fibronectin 1 (FN1), matrix metalloproteinase (MMP) 1, plasminogen activator inhibitor-1 (SERPINE1), tissue inhibitors of metallo-proteinases (TIMP1), plasminogen activator urokinase (PLAU) and PLAU receptor (PLAUR)] were selected as the hub genes; two modules with MCODE scores>5 were selected from the PPI network (Fig. 7B and C). Coincidentally, Module 1 was composed of the 10 hub genes previously selected (Table II).

Construction of the prediction system. The prognostic prediction system, composed of five signature genes (LAMC2,
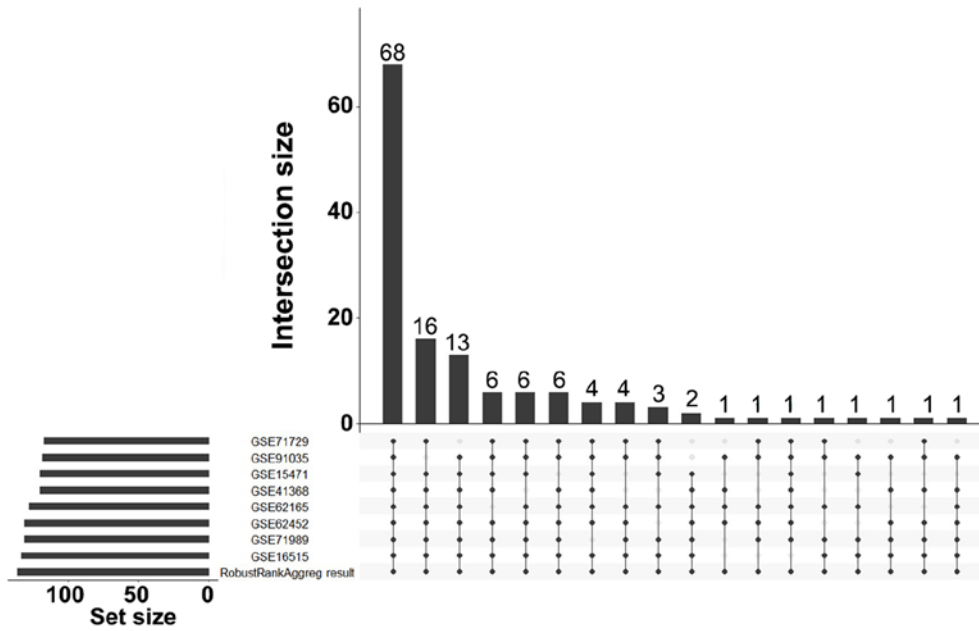
Figure 3. Histogram of the sources of the 136 DEGs. bar height represents the number of DEGs across multiple datasets linked by dots.
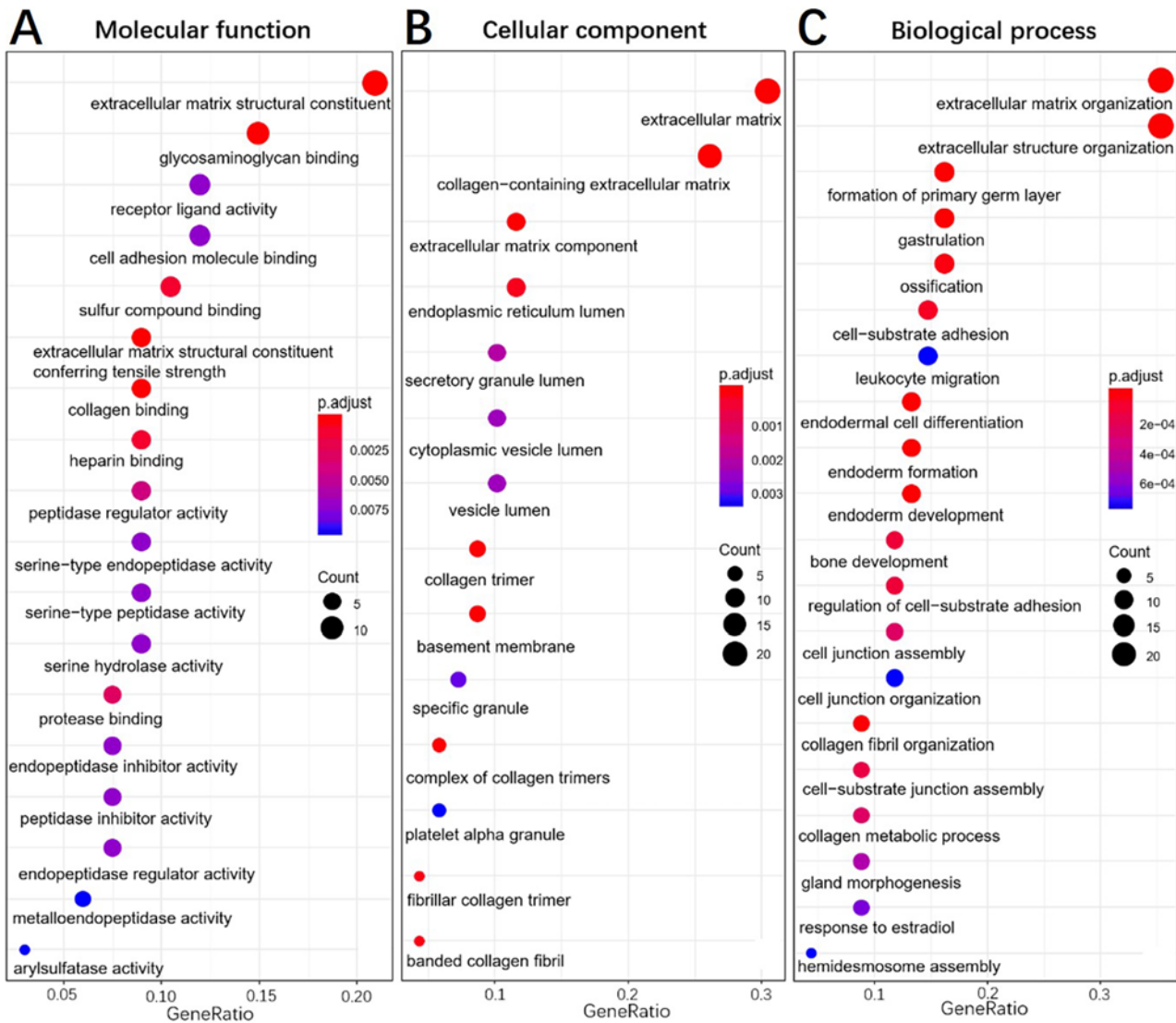

Figure 4. Gene ontology enrichment analysis of upregulated genes in pancreatic tumor compared with adjacent non-tumor tissues. (A), Molecular function, (B), Biological process, (C), Cellular component.
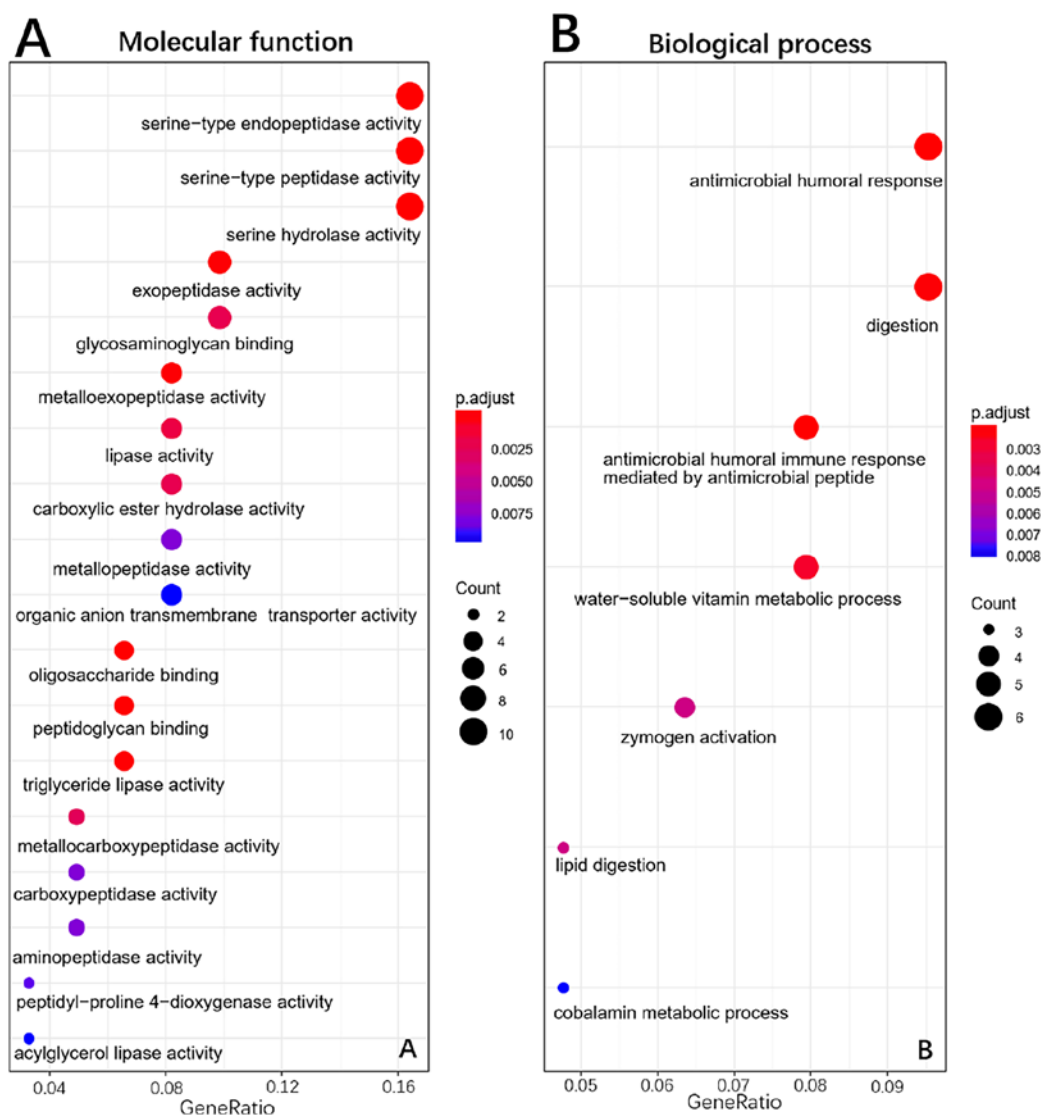
Figure 5. Gene ontology enrichment analysis of downregulated genes in pancreatic tumor compared with adjacent non-tumor tissues. (A), Molecular function, (B), Biological process.

LAMB3, SERPINB5, AREG and SFRP4), was constructed using the survival information of 66 patients in the training set (GSE62452). The risk score formula for each patient was calculated as shown below: Risk score=(1.0656) x LAMC2 + (-0.5804) x LAMB3 + (-0.4488) x SERPINB5 + (0.3060) x AREG + (-0.5294) x SFRP4. The area under the ROC curve was found to be 0.862, and consequently, specificity and sensitivity were both determined to be highest when the risk score was 0.960 (Fig. 8). The PDAC patients of the GSE62452 dataset were divided into a high-risk group (risk score, ≥0.960; n=33) and a low-risk group (risk score, <0.960; n=33). The patients in the low-risk group (45.5%; 95% CI, 30.1-68.7%) had a significantly higher survival rate than those in the high-risk group (48.0%; 95% CI, 33.5-68.7%; P=4x10^{-6}; Fig. 9A).

Validation of the prediction system. A dataset retrieved from TCGA, consisting of the clinical prognostic information of 146 patients with PDAC, was used as an independent validation dataset for the prognostic prediction system. The individual risk score of each patient was calculated using the aforementioned formula. A risk score of 0.960 was used as the threshold and the patient samples were divided into high- and low-risk groups (n=73 each). KM survival analysis showed that the high-risk group (48.7%; 95% CI, 37.8-62.7%) had significantly poorer OS scores than the low-risk group (48.1%; 95% CI, 36.0-64.4%; P=0.008; Fig. 9B).

Discussion

The pathogenesis of PDAC is extremely complex. Whilst there have been many studies on the biological mechanisms underpinning PDAC, the results are inconsistent and various aspects remain unclear. This could be attributable to the three following aspects: i) Small study sample sizes; ii) datasets retrieved from different platforms; and iii) different statistical analysis methods. However, using a combination of multiple datasets from different platforms, the present study aimed to improve the accuracy and reliability of these results. In the present study, eight PDAC datasets (*GSE15471*, *GSE16515*, *GSE41368*, *GSE62165*, *GSE62452*, *GSE71729*, *GSE71989* and *GSE91035*)
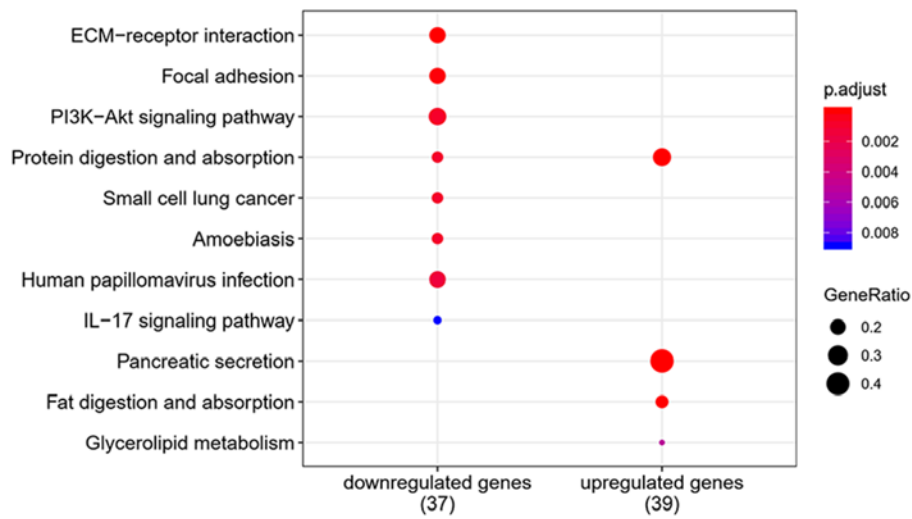
Figure 6. Pathway analysis for differentially expressed genes between pancreatic tumor and adjacent non-tumor tissues.
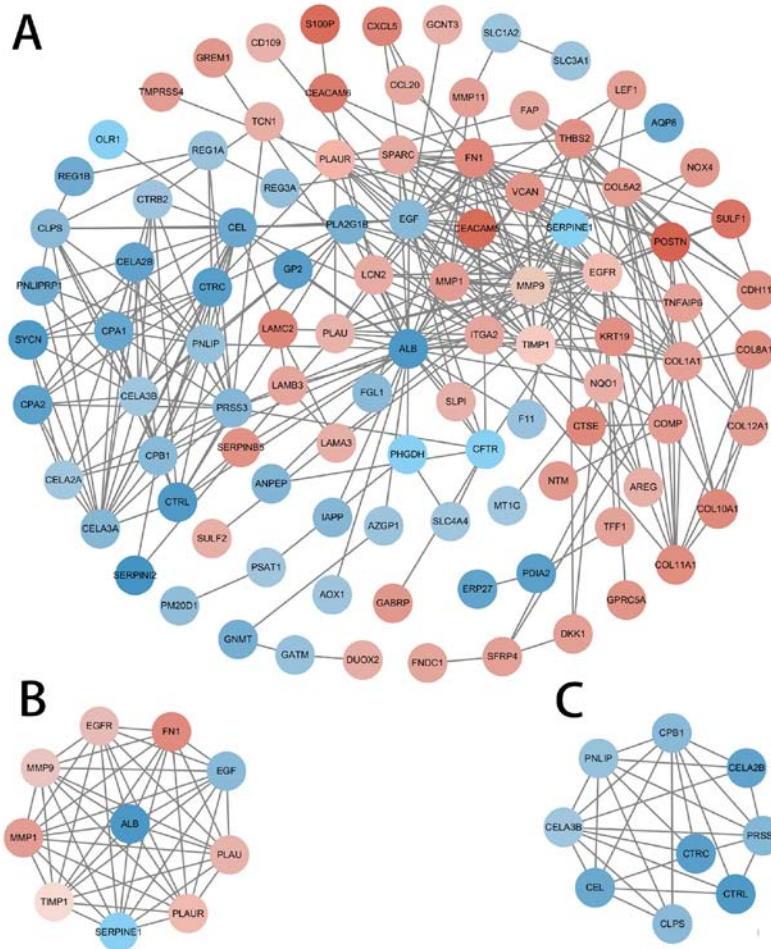

Figure 7. PPI network of DEGs. The MCODE algorithm was applied to this network to identify neighborhoods where proteins were densely connected. the gradient from green to red represents the change from down to upregulation; two significant modules with a score>5.0 were selected. (A) PPI network of DEGs; (B) Module 1, MCODE score=8.182; and (C) Module 2, MCODE score=5.2. PPI, protein-protein interaction; DEGs, differentially expressed genes.

from two platforms (Affymetrix and Agilent) were analyzed, and the RRA package was used to account for the differences between the platforms. Finally, a dataset from TCGA was used to verify the data and to further improve reliability.

A total of 136 DEGs were identified from the GEO datasets. These included 67 up- and 69 downregulated genes, which were differentially expressed in PDAC samples, compared with the normal controls.

Table II. Hub genes with high degree of connectivity.

Name	MCODE cluster	MCODE score	Degree
FN1	Cluster 1	7.363636	20
ALB	Cluster 1	7.363636	33
MMP1	Cluster 1	7.363636	16
TIMP1	Cluster 1	7.363636	14
MMP9	Cluster 1	7.363636	23
SERPINE1	Cluster 1	7.363636	13
PLAUR	Cluster 1	7.363636	12
PLAU	Cluster 1	7.363636	11
EGF	Cluster 1	7.363636	26
EGFR	Cluster 1	7.363636	23

FN1, fibronectin 1; ALB, albumin; MMP, matrix metallopeptidase; TIMP1, tissue inhibitor of metalloproteinases 1; SERPINE1, serpin family E member 1; PLAUR, plasminogen activator urokinase receptor; PLAU, plasminogen activator urokinase; EGF, epidermal growth factor; EGFR, epidermal growth factor receptor.

Figure 8. AUC for the GSE62452 dataset. ROC, receiver operating characteristic; AUC, area under the curve.

GO enrichment analysis determined that the most significant enrichments in MF, BP and CC were related to the extracellular matrix (ECM), especially for the upregulated genes. For KEGG pathways analysis, 'pancreatic secretion', 'PI3K-Akt signaling pathway' and 'ECM-receptor interaction' were highly enriched. High enrichment of ECM related genes and pathways suggests that ECM regulation is closely associated with PDAC progression. The ECM is a complex, three-dimensional structure composed of both structural and non-structural proteins (22,23). It plays a fundamental role in facilitating cell differentiation, apoptosis, proliferation and migration (24). The ECM is also the most abundant component in the tumor microenvironment (TME), which profoundly influences the behavior of cancer cells (25). Furthermore, certain studies have suggested that alterations in the ECM can induce cell transformation and metastasis, promoting the development and progression of tumors (26,27). In the TME of PDAC, some ECM proteins, such as collagen, fibronectin and laminin, are significantly upregulated (28). Each of these proteins promotes the growth and invasion of PDAC cells (29-32). Moreover, the results of the present study indicated LAMC2 and LAMB2 are significantly associated with PDAC prognosis.

Based on PPI network analysis of the common DEGs in the selected studies, the 10 most highly connected genes (ALB, EGF, MMP9, EGFR, FN1, MMP1, SERPINE1, TIMP1, PLAU and PLAUR) were screened and Module 1 is, coincidentally, composed of the 10 hub genes. Furthermore, FN1 and PLAU have previously been identified as hub genes in a similar study (10). Multiple studies have determined that these genes are associated with tumorigenesis and progression. For example, the FN1 gene encodes fibronectin, which is a major constituent of the ECM within the TME. The binding of FN1 to its receptor activates the FN1 signaling pathway in pancreatic cancer cells, and promotes tumor cell survival, invasion, metastasis and angiogenesis (33). The expression of fibronectin in pancreatic cancer cells is also associated with a low survival rate (34,35). In addition, FN1 participates in the progression of ovarian cancer
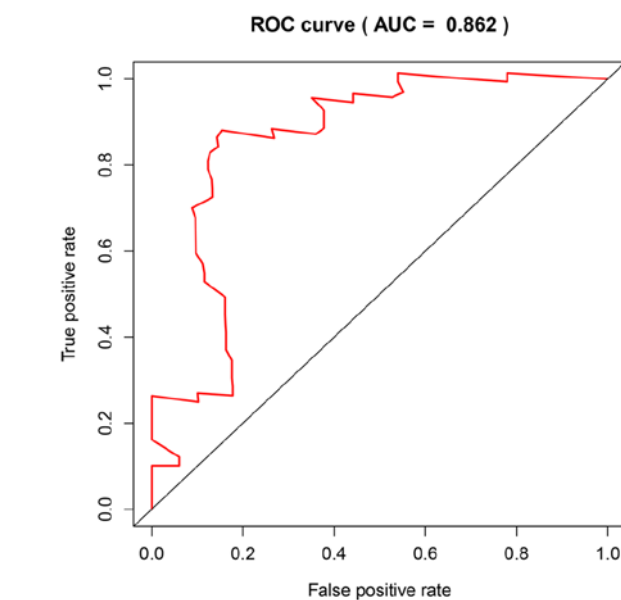
by increasing the expression level of matrix metalloproteinase MMP-9 (36). MMPs and TIMPs are important enzymes in the process of ECM degradation. Expression of the MMP-9 protein is upregulated during the progression of various cancer types, including pancreatic cancer, and is directly involved in tumor cell migration, invasion, metastasis, tumor-related inflammation and angiogenesis (37,38). Therefore, its expression is associated with malignant progression (39).

The MMP-1 protein is one of the most widely expressed MMPs. It activates the G protein-coupled receptor protease-activated receptor-1 (PAR1) and induces secretion of bioactive proteins (most prominently interleukin-8, growth-regulated oncogene-α and C-C motif chemokine ligand 2) to regulate tumor migration (40). TIMP-1 enzyme is a natural inhibitor of MMPs, but it can also stimulate cell proliferation and prevent apoptosis, promoting cancer progression (41). The binding of TIMP1 to receptors on the cell surface activates the PI3K/Akt signaling pathway via Ras, and the EGF signal then induces TIMP1 expression (42,43). High plasma TIMP1 levels may interact with EGFR signaling, and thereby reduce the anti-tumor effects of EGFR inhibitors (44). The EGF and EGFR proteins are widely recognized for their role in numerous cancer types, including PDAC (45,46). Upon binding to EGF, EGFR forms homologous dimers, auto-phosphorylates and interacts with downstream factors to activate genes involved in cell proliferation, differentiation, survival and migration (47,48). As such, inhibitors of the EGFR signaling pathway have received attention as potential therapeutic agents (49). Moreover, EGF-induced activation of EGFR increases MMP-9 expression levels through the activation of the PI3K/Akt pathway in patients with glioblastoma (50).

In the present study, the ALB gene showed the highest connectivity to other genes in Module 1A, and this gene encodes the most abundant protein found in human blood (51). Cachexia is expressed in <80% of patients with PDAC, and plasma albumin is often reduced (52,53). In the current study, it was speculated that besides digestive and absorption disor-
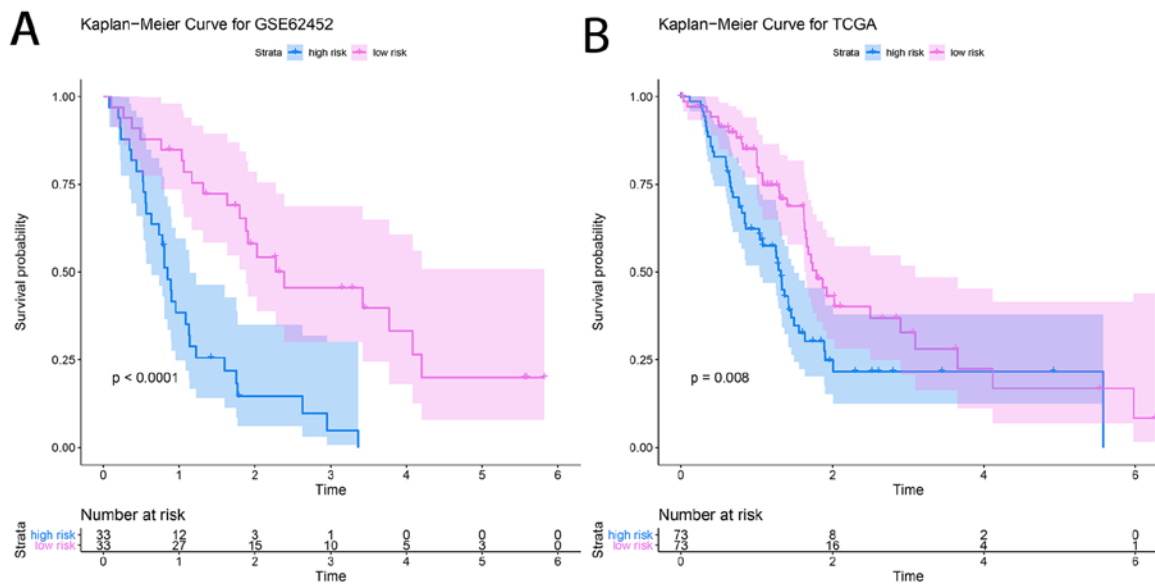
Figure 9. Kaplan-Meier survival curves of (A) GSE62452 and (B) The Cancer Genome Atlas.

ders, cachexia may also be related to the low expression of ALB. Nevertheless, investigations into the regulatory genetic mechanism of ALB in PDAC have rarely been reported. PLAU, PLAUR and SERPINE1 represent the main components of the system. This system not only participates in cell signaling pathways (such as angiogenesis, cell growth, cell adhesion and migration) to influence cancer-related processes, but also in the disruption of the ECM through various pathways, such as the downregulation of the tumor suppressor gene p53, interference of Hoxa5 function, inhibition of the p38 pathway and activation of mitogen-activated protein kinase 1 (54-57).

An accurate and credible prognostic prediction system may help to better evaluate patient prognosis, provide a basis for clinical decision-making and indicate new therapeutic targets. Of the five genes associated with prognosis in the present study, LAMC2 and SER5 have also been reported to be prognosis-associated genes in PDAC (6,10). The LAMC2 and laminin subunit α 3 proteins are important components of laminin 5. Through interaction with cell surface receptors, they participate in a variety of biological processes, including cell adhesion, differentiation, tumor angiogenesis and metastasis. The upregulation of LAMC2 is considered an indicator of adverse prognosis and the high metastatic potential of multiple cancer types, including colorectal cancer and lung adenocarcinoma (58). LAMC2 gene-silencing has been shown to significantly inhibit cell migration and invasion in head and neck squamous cell carcinoma cells (59). The LAMB3 protein regulates epithelial mesenchymal transition-related proteins and MMP-9 via the activation of the Akt signaling pathway, to promote the invasion and metastasis of tumors in papillary thyroid cancer (60). The important role of LAMA3 in the metastasis of lung adenocarcinoma has also been reported (61).

SERPINB5 is regarded as a tumor suppressor and an important senescence-associated biomarker. Notably, it is expressed in PDAC, but not in healthy pancreatic tissue (62). There is abundant evidence that SERPINB5 plays an oncogenic and pro-metastatic role (63,64), and is associated with the prognosis of patients with PDAC (62).

AREG is a ligand of EGFR and is overexpressed in various cancer tissues. It has been demonstrated that high AREG expression can be used as an independent prognostic indicator of poor overall survival in patients with PDAC, and there is a significant association between AREG/EGFR co-expression and poor tumor differentiation (65). The AREG protein has been clinically indicated as a prognostic and predictive biomarker, and numerous novel strategies have been developed to disrupt AREG-mediated oncogenic pathways (66).

SFRP4 is a Wnt-signaling antagonist. Due to its pro-apoptotic properties, SFRP4 provides axon-guidance information and functions as a tumor suppressor in multiple tissue types. Highly methylated SFRP4 induces transcriptional silencing, which activates abnormal Wnt signaling, leading to tumorigenesis and tumor progression (67). Moreover, SFRP4 monotherapy (or in combination with chemotherapy) can inhibit proliferation, reduce cell survival and initiate apoptosis in cancer stem cells in breast, prostate and ovarian cancer cell lines. This increases cell sensitivity to chemotherapy, amplifying the treatment effect (68). However, the mechanism of SFRP4 in PDAC remains poorly characterized.

Certain limitations of the present study should be noted. Firstly, the effects of certain patient and disease characteristics (tumor grade/stage, sex, age and race) on gene expression were not accounted for. Secondly, the GEO and TCGA data were obtained from public databases, and consequently, evaluation of data quality was challenging. Thirdly, variation in the accuracy of the prediction system for different molecular subtypes of PDAC has not been investigated due to a lack of consensus on the clinical classification of different molecular subtypes. Fourthly, TCGA dataset was used to validate 136 DEGs identified in the meta-analysis of the GEO datasets, but the results are not as similar as anticipated. This could be due to the small sample size of normal tissues in TCGA dataset. Finally, analysis was performed on gene expression databases; further molecular experimentation is required to validate the results.

In conclusion, the present study confirmed certain DEGs that were previously indicated in similar studies, and has also

identified a set of genes that were not discovered in individual analyses. Therefore, this technique should be considered useful for the further analysis of heterogeneous expression datasets, improvement of DEG recognition techniques and the discovery of new biomarkers for PDAC diagnosis or therapeutic targeting. Furthermore, the present study identified 10 hub genes that may be involved in the pathogenesis of PDAC using multi-platform, multi-gene expression profile datasets and bioinformatics meta-analysis. In addition, a reliable predictive system, composed of five genes, for determining the prognosis of patients with PDAC has been constructed. The results of the current study may help to guide individualized clinical decision-making and future molecular-targeted therapies.

Availability of data and materials

The datasets used and/or analyzed during the present study are available from the GEO and TCGA databases.

Authors' contributions

XT and YP conceived and designed the study. YQP and YM drafted the manuscript. YM and YQP collected and analyzed the data. LP critically revised the manuscript. XL, JX and LP were responsible for interpretation of the data. The manuscript was revised by YP XT, with final approval from YP.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68: 394-424, 2018.
2. Jemal A, Siegel R, Xu J and Ward E: Cancer statistics, 2010. CA Cancer J Clin 60: 277-300, 2010.
3. Hidalgo M: Pancreatic cancer. N Engl J Med 362: 1605-1617, 2010.
4. Shen Q, Yu M, Jia JK, Li WX, Tian YW and Xue HZ: Possible Molecular markers for the diagnosis of pancreatic ductal adenocarcinoma. Med Sci Monit 24: 2368-2376, 2018.
5. Ger M, Kaupinis A, Petrulionis M, Kurlinkus B, Cicenas J, Sileikis A, Valius M and Strupas K: Proteomic identification of FLT3 and PCBP3 as potential prognostic biomarkers for pancreatic cancer. Anticancer Res 38: 5759-5765, 2018.
6. Tang Y, Zhang Z, Tang Y, Chen X and Zhou J: Identification of potential target genes in pancreatic ductal adenocarcinoma by bioinformatics analysis. Oncol Lett 16: 2453-2461, 2018.
7. Cancer Genome Atlas Research Network:; and Cancer Genome Atlas Research Network: Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. Cancer Cell 32: 185-203 e13, 2017.
8. Kolde R and Laur S. RobustRankAggreg: Methods for robust rank aggregation. 2013.
9. Kolde R, Laur S, Adler P and Vilo J: Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics 28: 573-580, 2012.
10. Badea L, Herlea V, Dima SO, Dumitrascu T and Popescu I: Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. Hepatogastroenterology 55: 2016-2027, 2008.
11. Pei H, Li L, Fridley BL, Jenkins GD, Kalari KR, Lingle W, Petersen G, Lou Z and Wang L: FKBP51 affects cancer cell response to chemotherapy by negatively regulating Akt. Cancer Cell 16: 259-266, 2009.
12. Frampton AE, Castellano L, Colombo T, Giovannetti E, Krell J, Jacob J, Pellegrino L, Roca-Alonso L, Funel N, Gall TM, *et al*: MicroRNAs cooperatively inhibit a network of tumor suppressor genes to promote pancreatic tumor growth and progression. Gastroenterology 146: 268-277 e18, 2014.
13. Janky R, Binda MM, Allemeersch J, Van den Broeck A, Govaere O, Swinnen JV, Roskams T, Aerts S and Topal B: Prognostic relevance of molecular subtypes and master regulators in pancreatic ductal adenocarcinoma. BMC Cancer 16: 632, 2016.
14. Yang S, He P, Wang J, Schetter A, Tang W, Funamizu N, Yanaga K, Uwagawa T, Satoskar AR, Gaedcke J, *et al*: A Novel MIF signaling pathway drives the malignant character of pancreatic cancer by targeting NR3C2. Cancer Res 76: 3838-3850, 2016.
15. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, Rashid NU, Williams LA, Eaton SC, Chung AH, *et al*: Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet 47: 1168-1178, 2015.
16. Jiang J, Azevedo-Pouly AC, Redis RS, Lee EJ, Gusev Y, Allard D, Sutaria DS, Badawi M, Elgamal OA, Lerner MR, Brackett DJ, *et al*: Globally increased ultraconserved noncoding RNA expression in pancreatic adenocarcinoma. Oncotarget 7: 53165-53177, 2016;
17. Sutaria DS, Jiang J, Azevedo-Pouly ACP, Lee EJ, Lerner MR, Brackett DJ, Vandesompele J, Mestdagh P and Schmittgen TD: Expression Profiling Identifies the Noncoding Processed Transcript of HNRNPU with Proliferative Properties in Pancreatic Ductal Adenocarcinoma. Noncoding RNA 3: ncrna3030024, 2017
18. Kohl M, Wiese S and Warscheid B. Cytoscape: Software for visualization and analysis of biological networks. Methods Mol Biol 696: 291-303, 2011.
19. Zhao X, Sun S, Zeng X and Cui L: Expression profiles analysis identifies a novel three-mRNA signature to predict overall survival in oral squamous cell carcinoma. Am J Cancer Res 8: 450-461, 2018.
20. Zhu T, Gao YF, Chen YX, Wang ZB, Yin JY, Mao XY, Li X, Zhang W, Zhou HH and Liu ZQ: Genome-scale analysis identifies GJB2 and ERO1LB as prognosis markers in patients with pancreatic cancer. Oncotarget 8: 21281-21289, 2017.
21. Li H, Wang X, Fang Y, Huo Z, Lu X, Zhan X, Deng X, Peng C and Shen B: Integrated expression profiles analysis reveals novel predictive biomarker in pancreatic ductal adenocarcinoma. Oncotarget 8: 52571-52583, 2017.
22. Hynes RO and Naba A: Overview of the matrisome-an inventory of extracellular matrix constituents and functions. Cold Spring Harb Perspect Biol 4: a004903, 2012.
23. Humphrey JD, Dufresne ER and Schwartz MA: Mechanotransduction and extracellular matrix homeostasis. Nat Rev Mol Cell Biol 15: 802-812, 2014.
24. Venning FA, Wullkopf L and Erler JT: Targeting ECM disrupts cancer progression. Front Oncol 5: 224, 2015.
25. Fang M, Yuan J, Peng C and Li Y: Collagen as a double-edged sword in tumor progression. Tumour Biol 35: 2871-2882, 2014.

26. Miles FL and Sikes RA: Insidious changes in stromal matrix fuel cancer progression. Mol Cancer Res 12: 297-312, 2014.
27. Pickup MW, Mouw JK and Weaver VM: The extracellular matrix modulates the hallmarks of cancer. EMBO Rep 15: 1243-1253, 2014.
28. Mollenhauer J, Roether I and Kern HF: Distribution of extracellular matrix proteins in pancreatic ductal adenocarcinoma and its influence on tumor cell proliferation in vitro. Pancreas 2: 14-24, 1987.
29. Armstrong T, Packham G, Murphy LB, Bateman AC, Conti JA, Fine DR, Johnson CD, Benyon RC and Iredale JP: Type I collagen promotes the malignant phenotype of pancreatic ductal adenocarcinoma. Clin Cancer Res 10: 7427-7437, 2004.
30. Vaquero EC, Edderkaoui M, Nam KJ, Gukovsky I, Pandol SJ and Gukovskaya AS: Extracellular matrix proteins protect pancreatic cancer cells from death via mitochondrial and nonmitochondrial pathways. Gastroenterology 125: 1188-1202, 2003.
31. Koenig A, Mueller C, Hasel C, Adler G and Menke A: Collagen type I induces disruption of E-cadherin-mediated cell-cell contacts and promotes proliferation of pancreatic carcinoma cells. Cancer Res 66: 4662-4671, 2006.
32. Grzesiak JJ and Bouvet M: The alpha2beta1 integrin mediates the malignant phenotype on type I collagen in pancreatic cancer cell lines. Br J Cancer 94: 1311-1319, 2006.
33. Topalovski M and Brekken RA: Matrix control of pancreatic cancer: New insights into fibronectin signaling. Cancer Lett 381: 252-258, 2016.
34. Glasner A, Levi A, Enk J, Isaacson B, Viukov S, Orlanski S, Scope A, Neuman T, Enk CD, Hanna JH, et al: NKp46 receptor-mediated interferon-γ production by natural killer cells increases fibronectin 1 to alter tumor architecture and control metastasis. Immunity 48: 107-119 e4, 2018.
35. Hu D, Ansari D, Zhou Q, Sasor A, Said Hilmersson K and Andersson R: Stromal fibronectin expression in patients with resected pancreatic ductal adenocarcinoma. World J Surg Oncol 17: 29, 2019.
36. Shibata K, Kikkawa F, Nawa A, Thant AA, Naruse K, Mizutani S and Hamaguchi M: Both focal adhesion kinase and c-Ras are required for the enhanced matrix metalloproteinase 9 secretion by fibronectin in ovarian cancer cells. Cancer Res 58: 900-903, 1998.
37. Himelstein BP, Canete-Soler R, Bernhard EJ, Dilks DW and Muschel RJ: Metalloproteinases in tumor progression: The contribution of MMP-9. Invasion Metastasis 14: 246-258, 1994-1995.
38. Bloomston M, Zervos EE and Rosemurgy AS II: Matrix metalloproteinases and their role in pancreatic cancer: A review of preclinical studies and clinical trials. Ann Surg Oncol 9: 668-674, 2002.
39. Grunwald B, Vandooren J, Locatelli E, Fiten P, Opdenakker G, Proost P, Kruger A, Lellouche JP, Israel LL, Shenkman L and Comes Franchini M: Matrix metalloproteinase-9 (MMP-9) as an activator of nanosystems for targeted drug delivery in pancreatic cancer. J Control Release 239: 39-48, 2016.
40. Agarwal A, Tressel SL, Kaimal R, Balla M, Lam FH, Covic L and Kuliopulos A: Identification of a metalloprotease-chemokine signaling system in the ovarian cancer microenvironment: Implications for antiangiogenic therapy. Cancer Res 70: 5880-5890, 2010.
41. Grunwald B, Schoeps B and Kruger A: Recognizing the molecular multifunctionality and interactome of TIMP-1. Trends Cell Biol 29: 6-19, 2019.
42. Hirsch FR, Varella-Garcia M and Cappuzzo F: Predictive value of EGFR and HER2 overexpression in advanced non-small-cell lung cancer. Oncogene 28(Suppl 1): S32-S37, 2009.
43. Normanno N, De Luca A, Bianco C, Strizzi L, Mancino M, Maiello MR, Carotenuto A, De Feo G, Caponigro F and Salomon DS: Epidermal growth factor receptor (EGFR) signaling in cancer. Gene 366: 2-16, 2006.
44. Olayioye MA, Neve RM, Lane HA and Hynes NE: The ErbB signaling network: Receptor heterodimerization in development and cancer. EMBO J 19: 3159-3167, 2000.
45. Chong CR and Janne PA: The quest to overcome resistance to EGFR-targeted therapies in cancer. Nat Med 19: 1389-1400, 2013.
46. Rivera F, Lopez-Tarruella S, Vega-Villegas ME and Salcedo M: Treatment of advanced pancreatic cancer: From gemcitabine single agent to combinations and targeted therapy. Cancer Treat Rev 35: 335-339, 2009.
47. Wang F, Xiao W, Sun J, Han D and Zhu Y: MiRNA-181c inhibits EGFR-signaling-dependent MMP9 activation via suppressing Akt phosphorylation in glioblastoma. Tumour Biol 35: 8653-8658, 2014.
48. Qiu Q, Yang M, Tsang BK and Gruslin A: EGF-induced trophoblast secretion of MMP-9 and TIMP-1 involves activation of both PI3K and MAPK signalling pathways. Reproduction 128: 355-363, 2004.
49. Wang T, Yamashita K, Iwata K and Hayakawa T: Both tissue inhibitors of metalloproteinases-1 (TIMP-1) and TIMP-2 activate Ras but through different pathways. Biochem Biophys Res Commun 296: 201-205, 2002.
50. Tarpgaard LS, Orum-Madsen MS, Christensen IJ, Nordgaard C, Noer J, Guren TK, Glimelius B, Sorbye H, Ikdahl T, Kure EH, et al: TIMP-1 is under regulation of the EGF signaling axis and promotes an aggressive phenotype in KRAS-mutated colorectal cancer cells: A potential novel approach to the treatment of metastatic colorectal cancer. Oncotarget 7: 59441-59457, 2016.
51. Curry S, Mandelkow H, Brick P, and Franks N: Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites. Nat Struct Biol 5: 827-835, 1998.
52. Mueller TC, Burmeister MA, Bachmann J and Martignoni ME: Cachexia and pancreatic cancer: Are there treatment options? World J Gastroenterol 20: 9361-9373, 2014.
53. Bachmann J, Buchler MW, Friess H and Martignoni ME: Cachexia in patients with chronic pancreatitis and pancreatic cancer: Impact on survival and outcome. Nutr Cancer 65: 827-833, 2013.
54. Mengele K, Napieralski R, Magdolen V, Reuning U, Gkazepis A, Sweep F, Brunner N, Foekens J, Harbeck N and Schmitt M: Characteristics of the level-of-evidence-1 disease forecast cancer biomarkers uPA and its inhibitor PAI-1. Expert Rev Mol Diagn 10: 947-962, 2010.
55. Asuthkar S, Stepanova V, Lebedeva T, Holterman AL, Estes N, Cines DB, Rao JS and Gondi CS: Multifunctional roles of urokinase plasminogen activator (uPA) in cancer stemness and chemoresistance of pancreatic cancer. Mol Biol Cell 24: 2620-2632, 2013.
56. Xue A, Xue M, Jackson C and Smith RC: Suppression of urokinase plasminogen activator receptor inhibits proliferation and migration of pancreatic adenocarcinoma cells via regulation of ERK/p38 signaling. Int J Biochem Cell Biol 41: 1731-1738, 2009.
57. Botla SK, Savant S, Jandaghi P, Bauer AS, Mucke O, Moskalev EA, Neoptolemos JP, Costello E, Greenhalf W, Scarpa A, et al: Early epigenetic downregulation of microRNA-192 expression promotes pancreatic cancer progression. Cancer Res 76: 4149-4159, 2016.
58. Ding J, Yang C and Yang S: LINC00511 interacts with miR-765 and modulates tongue squamous cell carcinoma progression by targeting LAMC2. J Oral Pathol Med 47: 468-476, 2018.
59. Kinoshita T, Nohata N, Hanazawa T, Kikkawa N, Yamamoto N, Yoshino H, Itesako T, Enokida H, Nakagawa M, Okamoto Y and Seki N: Tumour-suppressive microRNA-29s inhibit cancer cell migration and invasion by targeting laminin-integrin signalling in head and neck squamous cell carcinoma. Br J Cancer 109: 2636-2645, 2013.
60. Jung SN, Lim HS, Liu L, Chang JW, Lim YC, Rha KS and Koo BS: LAMB3 mediates metastatic tumor behavior in papillary thyroid cancer by regulating c-MET/Akt signals. Sci Rep 8: 2718, 2018.
61. Wang XM, Li J, Yan MX, Liu L, Jia DS, Geng Q, Lin HC, He XH, Li JJ and Yao M: Integrative analyses identify osteopontin, LAMB3 and ITGB1 as critical pro-metastatic genes for lung cancer. PLoS One 8: e55714, 2013.
62. Ohike N, Maass N, Mundhenke C, Biallek M, Zhang M, Jonat W, Luttges J, Morohoshi T, Kloppel G and Nagasaki K: Clinicopathological significance and molecular regulation of maspin expression in ductal adenocarcinoma of the pancreas. Cancer Lett 199: 193-200, 2003.
63. Kashima K, Ohike N, Mukai S, Sato M, Takahashi M and Morohoshi T: Expression of the tumor suppressor gene maspin and its significance in intraductal papillary mucinous neoplasms of the pancreas. Hepatobiliary Pancreat Dis Int 7: 86-90, 2008.
64. Cao D, Zhang Q, Wu LS, Salaria SN, Winter JW, Hruban RH, Goggins MS, Abbruzzese JL, Maitra A and Ho L: Prognostic significance of maspin in pancreatic ductal adenocarcinoma: Tissue microarray analysis of 223 surgically resected cases. Mod Pathol 20: 570-578, 2007.
65. Wang L, Wu H, Wang L, Lu J, Duan H, Liu X and Liang Z: Expression of amphiregulin predicts poor outcome in patients with pancreatic ductal adenocarcinoma. Diagn Pathol 11: 60, 2016.
66. Busser B, Sancey L, Brambilla E, Coll JL and Hurbin A: The multiple roles of amphiregulin in human cancer. Biochim Biophys Acta 1816: 119-131, 2011.
67. Pawar NM and Rao P: Secreted frizzled related protein 4 (sFRP4) update: A brief review. Cell Signal 45: 63-70, 2018.
68. Deshmukh A, Kumar S, Arfuso F, Newsholme P and Dharmarajan A: Secreted Frizzled-related protein 4 (sFRP4) chemo-sensitizes cancer stem cells derived from human breast, prostate, and ovary tumor cell lines. Sci Rep 7: 2256, 2017.