Comprehensive analysis of abnormal expression, prognostic value and oncogenic role of the hub gene FN1 in pancreatic ductal adenocarcinoma *via* bioinformatic analysis and *in vitro* experiments

Xiaohua Lei[1], Guodong Chen[1], Jiangtao Li[1], Wu Wen[1], Jian Gong[2] and Jie Fu[3]

[1] The First Affiliated Hospital, Department of Hepato-Biliary-Pancreatic Surgery, Hengyang Medical School, University of South China, Hengyang, Hunan, China
[2] Department of Gastroenterology, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China
[3] Department of General Surgery, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China

Corresponding authors
Jian Gong, jiangong001@csu.edu.cn
Jie Fu, fujie2016@csu.edu.cn

ABSTRACT

Background: Pancreatic ductal adenocarcinoma (PDAC) is one of the most commonly diagnosed cancers with a poor prognosis worldwide. Although the treatment of PDAC has made great progress in recent years, the therapeutic effects are still unsatisfactory. Methods. In this study, we identified differentially expressed genes (DEGs) between PDAC and normal pancreatic tissues based on four Gene Expression Omnibus (GEO) datasets (GSE15471, GSE16515, GSE28735 and GSE71729). A protein–protein interaction (PPI) network was established to evaluate the relationship between the DEGs and to screen hub genes. The expression levels of the hub genes were further validated through the Gene Expression Profiling Interactive Analysis (GEPIA), ONCOMINE and Human Protein Atlas (HPA) databases, as well as the validation GEO dataset GSE62452. Additionally, the prognostic values of the hub genes were evaluated by Kaplan–Meier plotter and the validation GEO dataset GSE62452. Finally, the mechanistic roles of the most remarkable hub genes in PDAC were examined through *in vitro* experiments.
Results: We identified the following nine hub genes by performing an integrated bioinformatics analysis: COL1A1, COL1A2, FN1, ITGA2, KRT19, LCN2, MMP9, MUC1 and VCAN. All of the hub genes were significantly upregulated in PDAC tissues compared with normal pancreatic tissues. Two hub genes (FN1 and ITGA2) were associated with poor overall survival (OS) rates in PDAC patients. Finally, *in vitro* experiments indicated that FN1 plays vital roles in PDAC cell proliferation, colony formation, apoptosis and the cell cycle.
Conclusions: In summary, we identified two hub genes that are associated with the expression and prognosis of PDAC. The oncogenic role of FN1 in PDAC was first illustrated by performing an integrated bioinformatic analysis and *in vitro* experiments. Our results provide a fundamental contribution for further research aimed finding novel therapeutic targets for overcoming PDAC.

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) is one of the cancers with worst prognosis, with a 5-year survival rate of less than 5% and a median survival time of less than 6 months after diagnosis (*Jemal et al., 2011*; *Zhang et al., 2019*; *Zhou et al., 2017*). Because of the non-specificity of symptoms, most PDAC patients are diagnosed at an advanced stage. Even patients whose disease is detected early, and who receive radical surgery followed by chemotherapy, have a median survival rate of 2 years and a 5-year survival rate of 15–20% (*Zhou et al., 2017*).

Although we are constantly updating and deepening our understanding of the complex genomic, epigenetic, and metabolomic alterations, and complex interplay between tumor cells and immune cells, mesenchymal cells, and epithelial cells in PDAC, unfortunately, patient quality of life and survival time have not improved significantly (*Neoptolemos et al., 2018*). It is currently believed that the development of PDAC is caused by multiple stages of gene mutations, including KRAS, TP53, SMAD4 and CDKN2A (*Biankin et al., 2012*; *Kamisawa et al., 2016*; *Zhang et al., 2019*). However, no effective therapeutic drugs have been developed for these targets. To explore effective biomarkers and therapeutic drugs for cancers, high-throughput sequencing and chips are widely used worldwide (*Kristensen et al., 2019*; *Lusito et al., 2019*; *Massard et al., 2017*). At the same time, combined with bioinformatics analysis and experimental study, valuable biological information was discovered that was useful for cancer treatment (*Jiang et al., 2019*; *Li et al., 2020*; *Li et al., 2018*; *Xu et al., 2020*). In recent years, some genes have been reported to play oncogenic roles in pancreatic cancer, but the clinical sample size enrolled in these studies is not enough, and the biological role of these genes in PDAC has not been fully studied (*Okada et al., 2021*; *Wang et al., 2019*; *Zhou et al., 2019*).

In this study, differentially expression genes (DEGs) between pancreatic ductal adenocarcinoma and normal pancreatic tissues were explored using four Gene Expression Omnibus (GEO) datasets, and protein–protein interaction (PPI) network was constructed through the Search Tool for the Retrieval of Interacting Genes (STRING) database based on these DEGs. Next, enrichment analysis was applied to these DEGs to explore the underlying mechanisms of PDAC. Furthermore, hub genes were screened by integrated analysis of the correlation between these DEGs, and their expression levels and prognostic values in PDAC were detected by online databases including Gene Expression Profiling Interactive Analysis (GEPIA), ONCOMINE, the Human Protein Atlas (HPA), the Kaplan–Meier Plotter and the validation GEO dataset. Finally, the biological roles of the most remarkable hub gene in PDAC, FN1, were further investigated through *in vitro* experiments.

MATERIALS & METHODS

Gene expression profile data

Microarray data of PDAC patients were obtained from the GEO database (http://www.ncbi.nlm.nih.gov/geo/). In total, 262 pancreatic ductal adenocarcinoma samples and 143

Table 1 The gene expression profile data characteristics.

Reference	PMID	Record	Tissue	Platform	Tumor	Normal
Badea et al.	19260470	GSE15471	PDAC	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	36	36
Pei et al.	19732725	GSE16515	PDAC	[HG133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	36	16
Zhang et al.	22363658	GSE28735	PDAC	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	45	45
Moffitt et al.	26343385	GSE71729	PDAC	Agilent-014850 Whole Human Genome Microarray 4 × 44 K	145	46
Yang et al.	27197190	GSE62452	PDAC	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	69	61

normal samples were enrolled in our study. The GSE15471 dataset includes 36 PDAC samples and matching noncancerous samples. GSE16515 consists of 36 PDAC samples and 16 noncancerous samples. GSE28375 includes 45 PDAC samples and matching noncancerous samples. GSE71729 consists of 145 PDAC samples and 46 noncancerous samples. GSE62452 contains 69 PDAC samples and 61 noncancerous samples. In this study, GSE15471, GSE16515, GSE28375 and GSE71729 were used as discovery datasets, and GSE62452, which contains survival data was used as a validation dataset. The gene expression profile data characteristics are listed in Table 1.

DEGs extraction

The DEGs between PDAC and corresponding normal samples in each microarray were screened using the limma package (*Ritchie et al., 2015*) in R software. DEGs with a $|log2FC| > 1$ and adjusted *P*-value < 0.05 in each GEO dataset were considered statistically significant. Gene integration for the DEGs identified from the four discovery datasets was conducted by Robust Rank Aggregation (RRA) (*Kolde et al., 2012*) method.

PPI network construction and hub gene identification

The STRING online database, which is a database of known and predicted protein interactions, was applied to construct the PPI network of DEGs, and only interactions with a combined score >0.4 were considered statistically significant (*Szklarczyk et al., 2015*). Moreover, Cytoscape software was used to visualize the PPI network (*Shannon et al., 2003*). Five methods in the plugin cytoHubba of Cytoscape software were used to evaluate the relationship between genes in the PPI network, including Betweenness, Closeness, Degree, EcCentrity and Radiality. The combination of these five methods is a common screening method of hub genes, which helps to improve the stability (*Chin et al., 2014*; *Wu et al., 2019*). The top 40 genes found by each method were intersected to find hub genes.

Functional enrichment analysis

The ClusterProfiler package in R software was used for functional enrichment analysis. The functional enrichment analysis for the upregulated and downregulated DEGs, included molecular function (MF), biological process (BP), cellular component (CC) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. A Q-value < 0.05 was considered statistically significant.

Validation of the expression levels of the hub genes

To validate the mRNA expression levels of the screened hub genes between PDAC and normal pancreatic tissues, the online tool GEPIA (http://gepia.cancer-pku.cn/) (*Tang et al., 2017*), ONCOMINE database (https://www.oncomine.org/) (*Rhodes et al., 2007*) and validation dataset GSE62452 were used. The protein expression levels of the hub genes were determined though the HPA database (https://www.proteinatlas.org/) (*Ouyang et al., 2019*).

Survival analysis of the hub genes

The effects of the hub genes on overall survival (OS) were determined by the web-based Kaplan–Meier (KM) plotter tool (https://kmplot.com/analysis/) and the survival data of the validation dataset GSE62452. PDAC patients were divided into two groups ("High expression" and "Low expression") based on the best cutoff calculated automatically. The hazard ratios (HR) with 95% confidence intervals and log rank P-values were calculated and are shown on the plot (*Buddhan et al., 2007*).

Cell culture and transfection

The human PDAC cell lines PANC1 and SW1990 were purchased from the Type Culture Collection of the Chinese Academy of Sciences (Shanghai, China). All cell lines were maintained in Dulbecco's modified Eagle medium (HyClone, Logan, UT, USA) supplemented with 10% fetal bovine serum, 1% penicillin and streptomycin (Gibco, New York, NY, US). Cells were cultured at 37 °C in an atmosphere containing 5% CO_2. Small interfering RNAs (siRNAs) targeting FN1 and a scrambled siRNA used as a normal control (NC) were purchased from RiboBio (Guangzhou, China). Transfection was conducted by Lipofectamine 2,000 Reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. Briefly, PDAC cells were seeded into six-well plates (100,000 cells/well) in 1 mL of complete medium. After culturing for 12 h, siRNA targeting FN1 mixed with the same volume of Lipofectamine 2,000 was added to the wells. The final concentration of siFN1 is 100 nM. After 72 h, the transfected cells were harvested for further study.

Colony formation assay

PDAC cells were seeded into six-well plates (1,000 cells/well) in 1.5 mL of complete medium. After culturing for 12 days, the colonies were stained with 0.5% crystal violet, and the number of colonies was counted.

Cell counting kit-8 (CCK-8) assay

Cell proliferation viability was assessed by CCK-8 assay (Dojindo, Japan). Briefly, PDAC cells were seeded into 96-well plates at a density of 2,000 cells/well with 100 µl of culture medium. A total of 10 µl of CCK-8 reagent was added to each well at the indicated time points (0, 24, 48 and 72 h). The plates were incubated in the dark at 37 °C for 2 h, and then the optical density was measured at 450 nm. The experiments were performed in triplicate.

5-Ethynyl-2′-deoxyuridine (EdU) assay

EdU immunofluorescence staining was conducted using an EdU Kit provided by RiboBio (Guangzhou, China) according to the manufacturer's instructions. In brief, suspended PDAC cells with different treatments were seeded into 96-well plates (2,000 cells/well) overnight. Then the cells were incubated with EdU for 2 h at 37 °C. After fixation in 4% paraformaldehyde for 10 min, the cells in plates were permeabilized with 0.5% Triton X-100. Next, the cells were stained with 1X Apollo reaction solution for 30 min. Finally, the nuclei were stained with Hoechst for 30 min and then imaged on a fluorescence microscope.

Quantitative real-time PCR (qRT-PCR)

Total RNA was extracted from PDAC cells using RNAiso Plus Reagent (Takara, Japan). qRT-PCR and data analysis were performed as previously described (*Fu et al., 2021*). Specifically, GAPDH was used as an internal control. Relative expression levels of FN1 were calculated according to the $2^{-\Delta\Delta Ct}$ method. All of the primers in this study were synthesized by ShenGong (Shanghai, China). The FN1 primer sequences were as follows: (forward) 5′- TGGGCAACTCTGTCAACGAA-3′ and (reverse) 5′- GCCATTTTC TCCCTGACGG -3′. The GAPDH primer sequences were as follows: (forward) 5′-GC GACTTCAACAGCAACTCCC-3′ and (reverse) 5′-CACCCTGTTGCTGTAGCCGTA-3′.

Flow cytometry

Cell cycle were measured by flow cytometry. Flow cytometry was performed as previously described (*Fu et al., 2021*). Briefly, PDAC cells after treatments were fixed in 70% ethanol at −20 °C overnight. The next day, the cells were stained with PI (Sigma–Aldrich, St. Louis, MO, USA) for 20 min at room temperature in the dark and analyzed by flow cytometry.

Western blotting

Total cell lysates were extracted by radioimmunoprecipitation assay (RIPA) buffer containing 1 mM phenylmethanesulfonyl fluoride (PMSF). The protein concentration was detected by BCA method, and 25 μg of protein from cell lysates were loaded onto the gel in each group. Western blotting was performed as previously described(*Zhu et al., 2020*). The antibodies used in this study are listed: BAX (proteintech, 1:5,000), Bcl2 (proteintech, 1:1,000), caspase9 (proteintech, 1:500), cyclinD1 (proteintech, 1:1,000). Actin (proteintech, 1:5,000) was used as an internal reference.

Gene set enrichment analysis (GSEA)

The ClusterProfiler package in R software was used for GSEA. Both of discovery datasets (GSE15471, GSE16515, GSE28375 and GSE71729) and validation dataset GSE62452 were included in GSEA (gene sets "KEGG_CELL_CYCLE" and "KEGG_APOPTOSIS"), which was performed between tumors and normal tissues.

Statistical analysis

SPSS was used to analyze the data. All data were expressed as the mean ± SD. Student's t-test was used in the two-group comparisons, and one-way ANOVA was used for more than two groups. Differences were considered significant at a P value < 0.05.

RESULTS

Identification of DEGs in PDAC microarray datasets and construction of the PPI network

A total of 922, 1,359, 415 and 460 DEGs from the GSE15471, GSE16515, GSE28735 and GSE71729 datasets were identified after standardization of the microarray results (Fig. 1A). A total of 233 genes overlapped among the four datasets consisting of 102 downregulated genes and 131 upregulated genes identified by the RRA method. The PPI network of the 233 DEGs was constructed by the STRING database and visualized by Cytoscape (Fig. 1B). According to the cut-off criteria, the top 20 differentially expressed upregulated and downregulated genes are shown in Fig. 1C.

Functional and pathway enrichment analyses

For the downregulated DEGs, GO analysis results (Fig. 2A) revealed that the biological processes (BPs) of DEGs were significantly changed in intestinal cholesterol absorption, lipid digestion, organic hydroxy compound transport, cellular modified amino acid metabolic process, etc. The DEGs were mainly enriched in cellular components (CCs) including pore complexes, membrane attack complexes, brush border membranes, cell projection membranes, etc. The DEGs were mainly enriched in molecular functions (MFs) including triglyceride lipase activity, serine-type endopeptidase activity, exopeptidase activity, serine hydrolase activity, etc. Enriched KEGG pathways (Fig. 2C) included pancreatic secretion, protein digestion and absorption, fat digestion and absorption, complement and coagulation cascades, glycerolipid metabolism, etc. For the upregulated genes, GO analysis results revealed (Fig. 2B) that the biological processes (BPs) of DEGs were significantly changed in collagen fibril organization, endoderm formation, endodermal cell differentiation, extracellular structure organization, etc. The DEGs were mainly enriched in cellular components (CCs) including complexes of collagen trimers, collagen trimers, extracellular matrix components and, etc. The DEGs were mainly enriched in molecular functions (MFs) including protease binding, integrin binding, extracellular matrix structural constituent conferring tensile strength, glycosaminoglycan binding, etc. KEGG pathways (Fig. 2D) included the PI3K-AKT signaling pathway, human papillomavirus infection, focal adhesion, ECM-receptor interaction, protein digestion and absorption, etc.

Identification of the hub genes

We then evaluated the relationship between DEGs in the PPI network to select hub genes using five methods including Betweenness, Closeness, Degree, EcCentricity and Radiality.
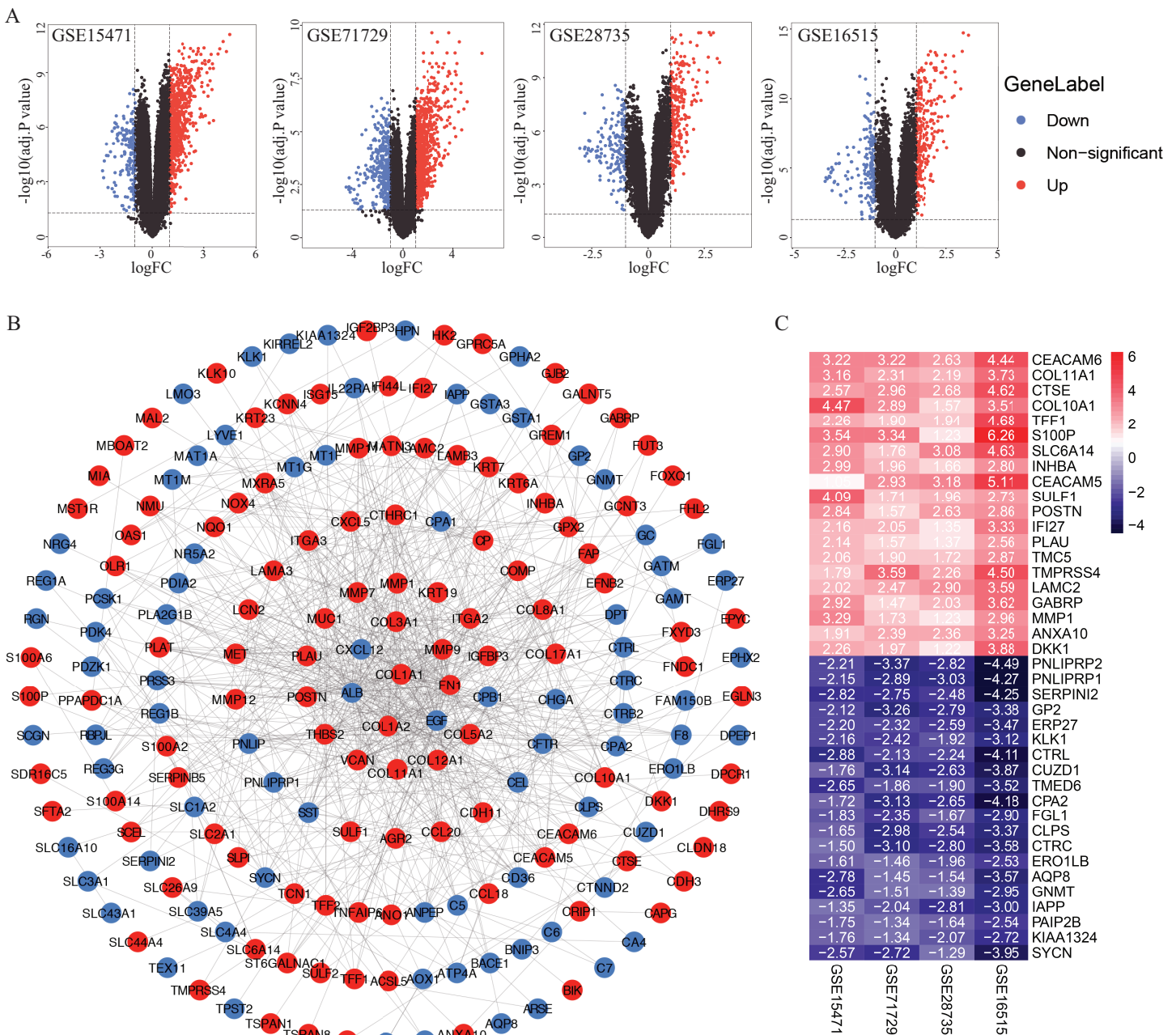
Figure 1 Identification of DEGs in PDAC microarray datasets and PPI network construction. (A) Volcano plot of the DEGs in GSE15471, GSE16515, GSE28735 and GSE71729 datasets. (B) PPI network of the 233 intersected DEGs. (C) Heat map of the top 20 upregulated and downregulated DEGs. Upregulated genes were marked in red, downregulated genes were marked in blue. DEGs, differentially expressed genes; PDAC, pancreatic ductal adenocarcinoma; PPI, protein-protein interaction.

The top 40 genes found by each method were intersected to identify the hub genes. Thereafter, five downregulated genes (ALB, EGF, CFTR, CXCL12 and SST) and nine upregulated genes (MUC1, MMP9, KRT19, COL1A1, COL1A2, LCN2, ITGA2 and VCAN) in PDAC tissues were identified.
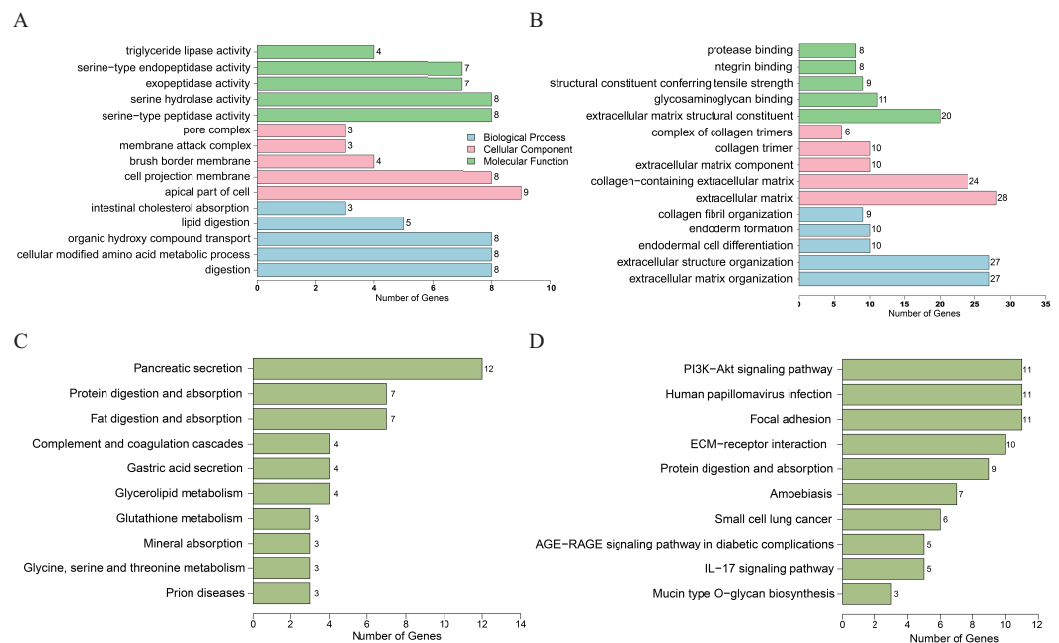
Figure 2 DEGs analyzed by GO enrichment (A, B) and KEGG enrichment (C, D) according to downregulation (A, C) or upregulation (B, D) of genes. DEGs, differentially expressed genes; GO, gene ontology. Y axis represents GO or KEGG pathway items; X axis represents the number of genes.

Validation of the expression levels of the hub genes

To validate the mRNA expression levels of the 14 screened hub genes between PDAC and normal pancreatic tissues, the online tool GEPIA based on the TCGA (https://portal.gdc.cancer.gov/) (*Sidaway, 2017*) was used firstly. The results showed that 11 hub genes were significantly differentially expressed in PDAC tissues compared with normal pancreatic tissues (Fig. 3). Next, the validation dataset GSE62452 further confirmed that all 14 hub genes were differentially expressed in PDAC (Fig. 4). To verify the expression levels of the remaining 11 hub genes screened by GEPIA in a wider range of samples, a meta-analysis was performed based on the ONCOMINE database. As shown in Fig. 5, nine hub genes were significantly highly expressed in PDAC tissues including COL1A1, COL1A2, FN1, ITGA2, KRT19, LCN2, MMP9, MUC1 and VCAN. Finally, the protein expression levels of these nine hub genes were determined using the Human Protein Atlas (HPA) database, and all of them seemed to be highly expressed in PDAC tissues compared with normal pancreatic tissues (Fig. 6).

Survival analysis

To further explore the prognostic values of these nine hub genes, overall survival analyses were conducted by Kaplan–Meier plotter and the survival data of the validation dataset GSE62452. As shown in Fig. 7A, seven hub genes overexpressed in PDAC patients predicted poor prognosis including FN1 (Hazard ratio (HR) = 1.58, p = 0.029), ITGA2 (HR =2.8, p = 1.2E−05), KRT19 (HR = 3.23, p = 7.9E−05), LCN2 (HR = 1.59, p = 0.034), MMP9 (HR = 1.65, p = 0.046), MUC1 (HR = 2.56, p = 0.00081) and
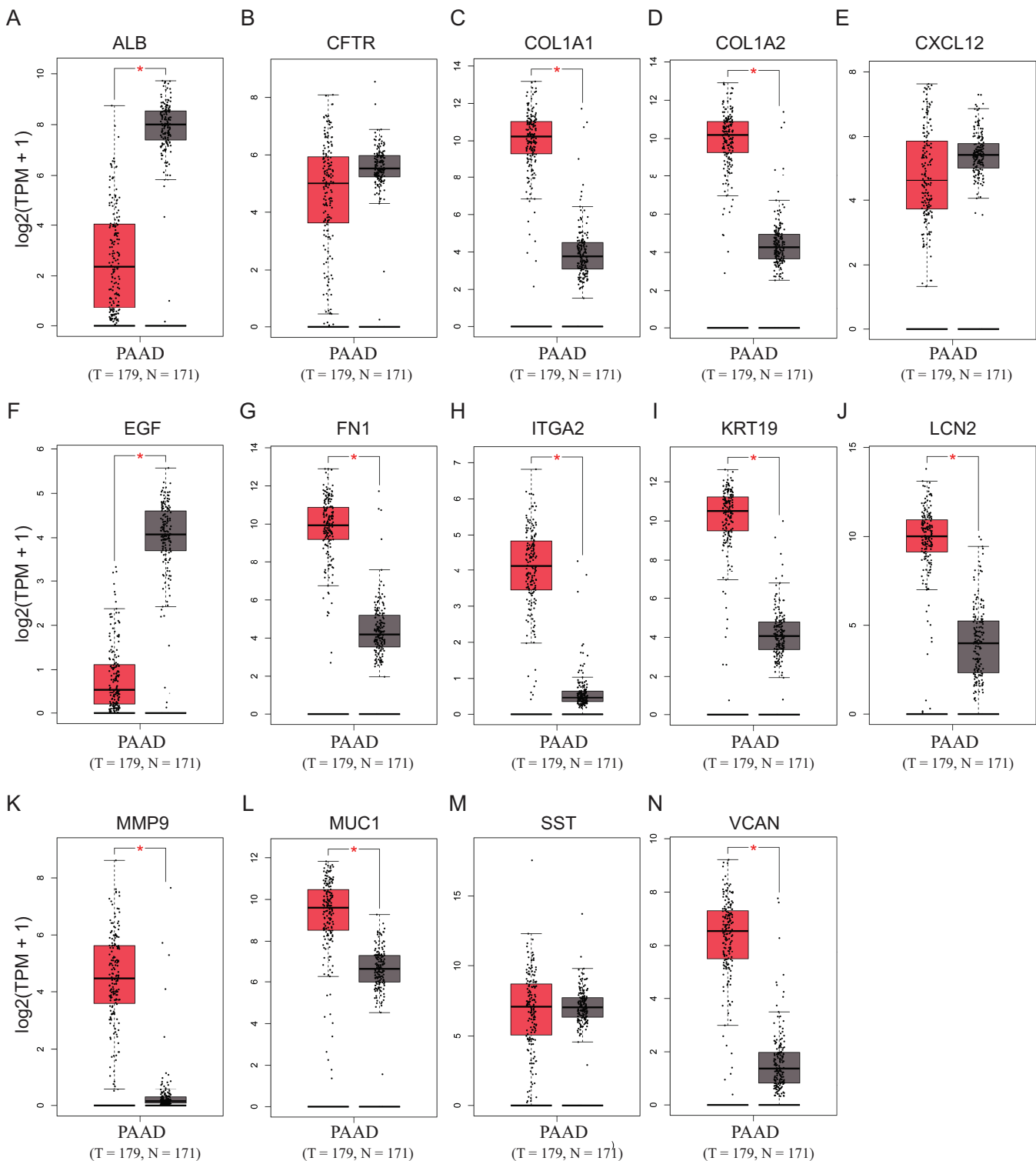
Figure 3 Validation of the mRNA expression levels of (A) ALB, (B) CFTR, (C) COL1A1, (D) COL1A2, (E) CXCL12, (F) EGF, (G) FN1, (H) ITGA2, (I) KRT19, (J) LCN2, (K) MMP9, (L) MUC1, (M) SST and (N) VCAN in pancreatic cancer tissues compared with normal pancreatic tissue. $^{*}P < 0.05$ was considered statistically significant. The expression level is described by log2(TPM+1). Red represents tumor ($n = 179$), brown represents nontumor ($n = 171$).
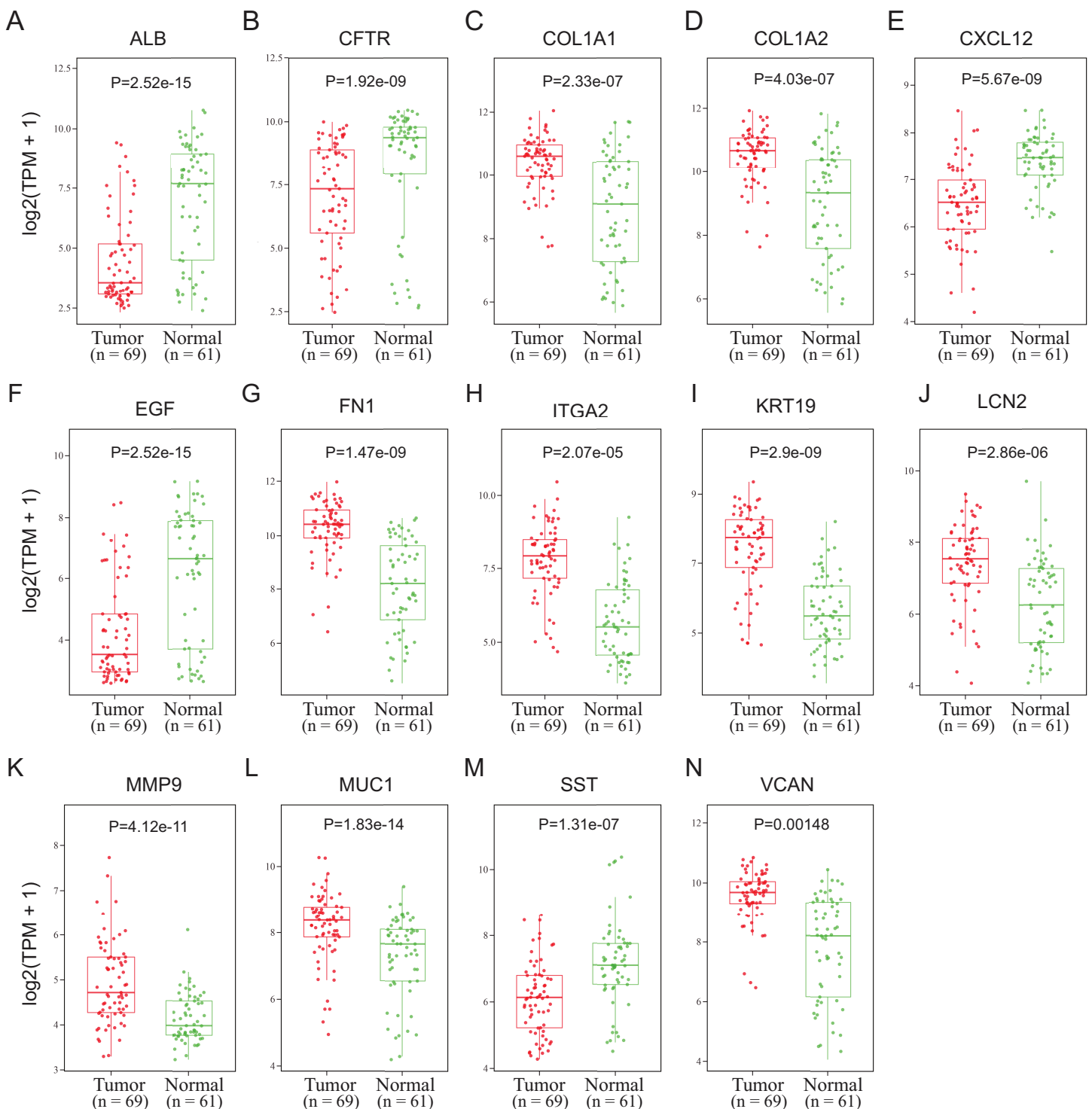
Figure 4 Validation of the mRNA expression levels of (A) ALB, (B) CFTR, (C) COL1A1, (D) COL1A2, (E) CXCL12, (F) EGF, (G) FN1, (H) ITGA2, (I) KRT19, (J) LCN2, (K) MMP9, (L) MUC1, (M) SST and (N) VCAN in pancreatic cancer tissues compared with normal pancreatic tissue. The expression level is described by log2(TPM+1). Red represents tumor (n = 69), green represents nontumor (n = 61).
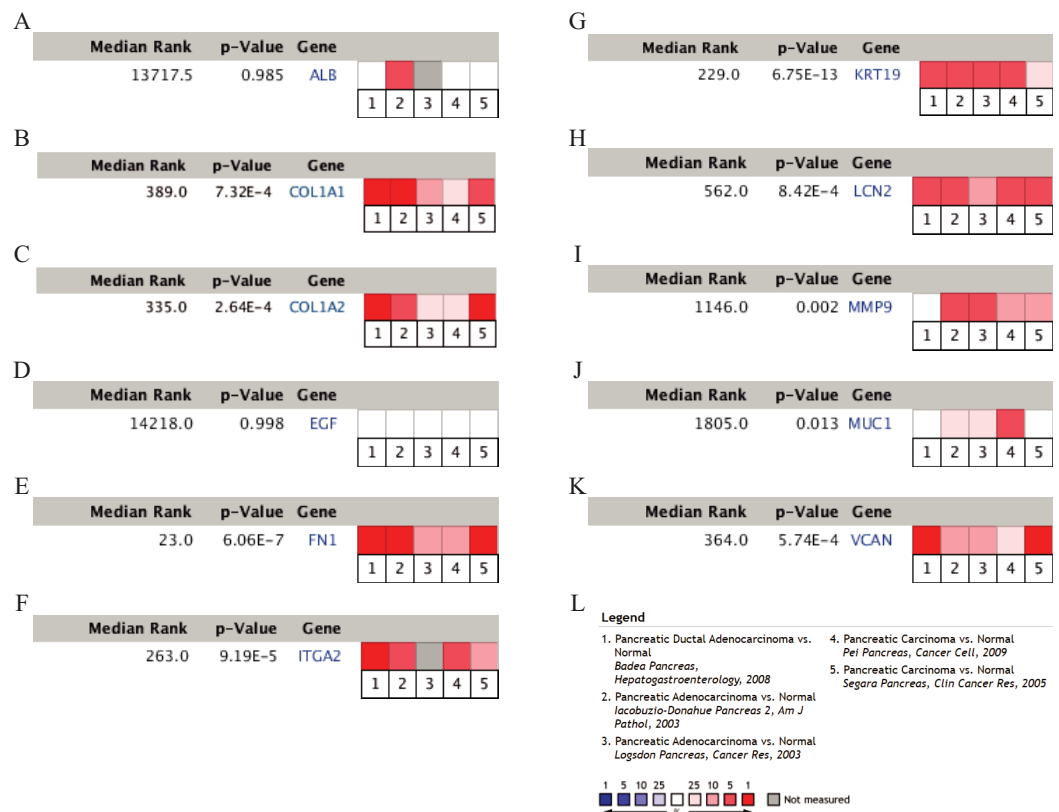
Figure 5 A meta-analysis of the mRNA expression levels of (A) ALB, (B) COL1A1, (C) COL1A2, (D) EGF, (E) FN1, (F) ITGA2, (G) KRT19, (H) LCN2, (I) MMP9, (J) MUC1 and (K) VCAN in PDAC tissues compared with normal pancreatic tissues by ONCOMINE. The colored squares represent the median rank of these genes across five datasets in ONCOMINE. *P*-value < 0.05 was considered as statistically significant. PDAC, pancreatic ductal adenocarcinoma.

VCAN (HR = 1.94, p = 0.014), while there were no significant differences in COL1A1 (HR = 1.4, p = 0.11) and COL1A2 (HR = 1.48, p = 0.058). However, in the survival analysis results of the validation GEO dataset, only two genes (FN1 (HR = 2.009, p = 0.018) and ITGA2 (HR = 1.806, p = 0.046)) indicated significantly poor prognosis, while there were no significant differences in other seven genes (COL1A1 (HR = 1.935, p = 0.05), COL1A2 (HR = 1.666, p = 0.113), KRT19 (HR = 1.578, p = 0.134), LCN2 (HR = 1.197, p = 0.557), MMP9 (HR = 1.476, p = 0.196), MUC1 (HR = 1.405, p = 0.289) and VCAN (HR = 1.416, p = 0.231)) (Fig. 7B). To our knowledge, the mechanistic role of FN1 in pancreatic cancer has not been studied, so FN1 was chosen for further analysis in PDAC cells.

Depletion of FN1 inhibited cell proliferation, blocked the cell cycle and induced apoptosis in PDAC cells

To further elucidate the role of FN1 in pancreatic carcinogenesis, we knocked down FN1 using small interfering RNA (siRNA) in PANC1 and SW1990 cell lines. The knockdown efficiency of FN1 was confirmed using qRT-PCR (Fig. 8A). The results showed that
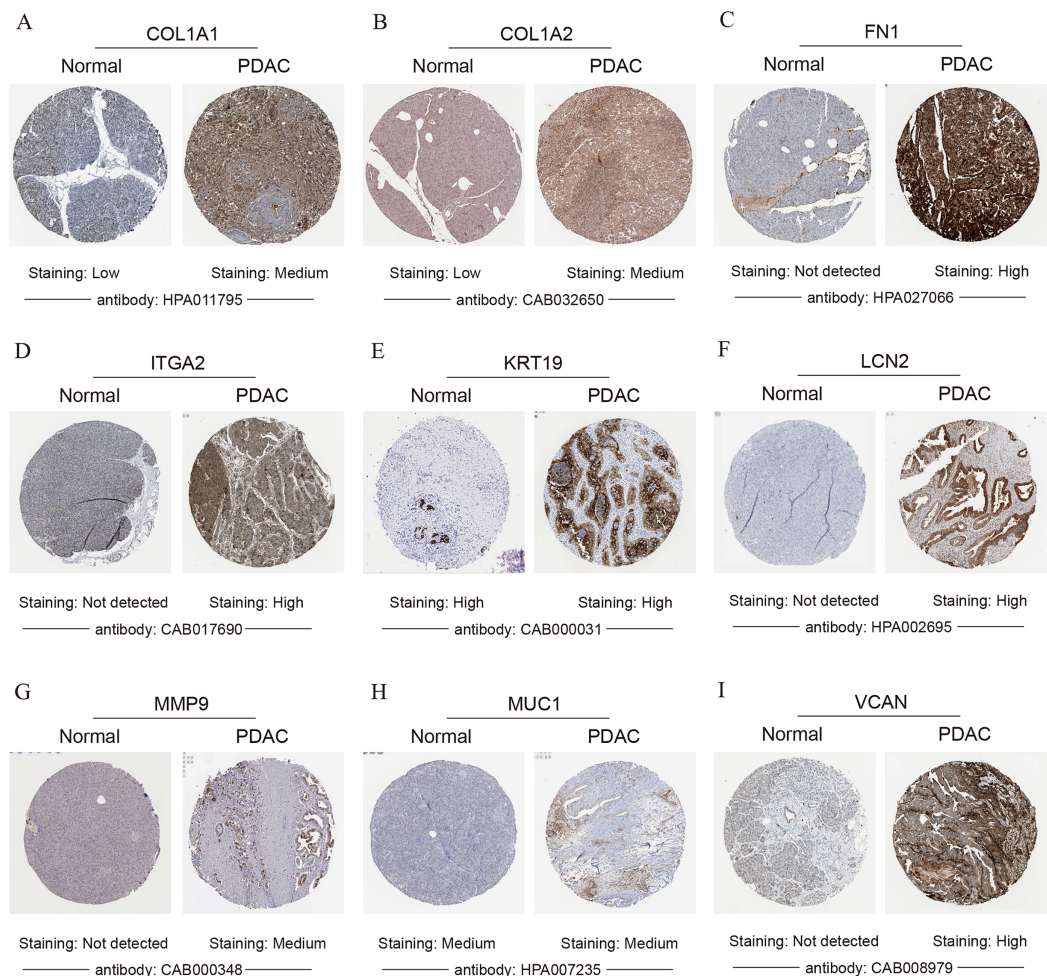
Figure 6 Validation of the protein expression levels of (A) COL1A1, (B) COL1A2, (C) FN1, (D) ITGA2, (E) KRT19, (F) LCN2, (G) MMP9, (H) MUC1 and (I) VCAN in PDAC tissues compared with normal pancreatic tissues by HPA database. PDAC, pancreatic ductal adenocarcinoma; HPA, human protein atlas.

siFN1-1 and siFN1-2 had better knockdown effects and were selected for further studies. Colony formation assays and CCK-8 assays suggested that FN1 downregulation significantly inhibited the proliferation abilities of PANC1 and SW1990 cells (Figs. 8B, 8C). Consistent with these results, EdU assays showed that cell viability was inhibited in FN1-downregulated PDAC cells (Figs. 8D, 8E). We next explored the effects of FN1 on the cell cycle using flow cytometry and western blotting. The cell cycle was obviously arrested in G0/G1 phase, and the expression of CyclinD1, which promotes cell cycle progression, was significantly decreased after FN1 knockdown (Figs. 8F, 8G). Finally, apoptosis-related proteins were detected by western blot. As illustrated by the results, FN1 knockdown significantly increased the expression of the proapoptotic proteins BAX and clevaead-caspase9, and decreased the expression level of the antiapoptotic protein Bcl2 (Fig. 8H). These results indicate that FN1 downregulation significantly inhibited PANC1
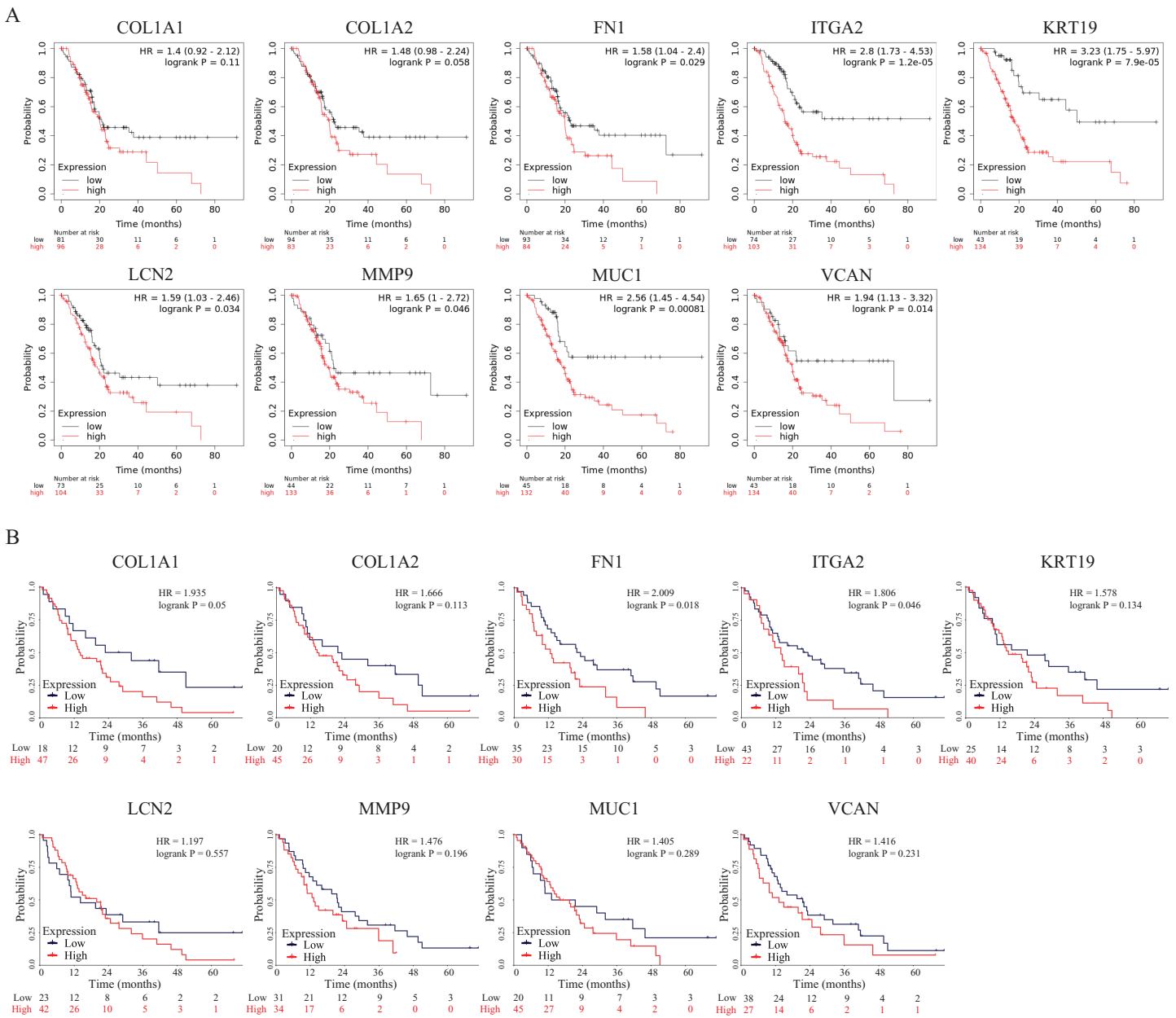
Figure 7 OS of the nine hub genes (COL1A1, COL1A2, FN1, ITGA2, KRT19, LCN2, MMP9, MUC1 and VCAN) in PDAC patients analyzed by (A) Kaplan–Meier plotter and the validation; (B) GEO dataset GSE62452. Log-rank $P < 0.05$ was considered as statistically significant. OS, overall survival; PDAC, pancreatic ductal adenocarcinoma.

and SW1990 cell growth, blocked the cell cycle and induced apoptosis *in vitro*, which was further confirmed by GSEA (Fig. 9).

DISCUSSION

PDAC is a malignant tumor with poor prognosis, due to the lack of specific early diagnostic markers and effective therapeutic targets. To further clarify the pathogenesis of pancreatic cancer and identify effective therapeutic targets, integrated computational
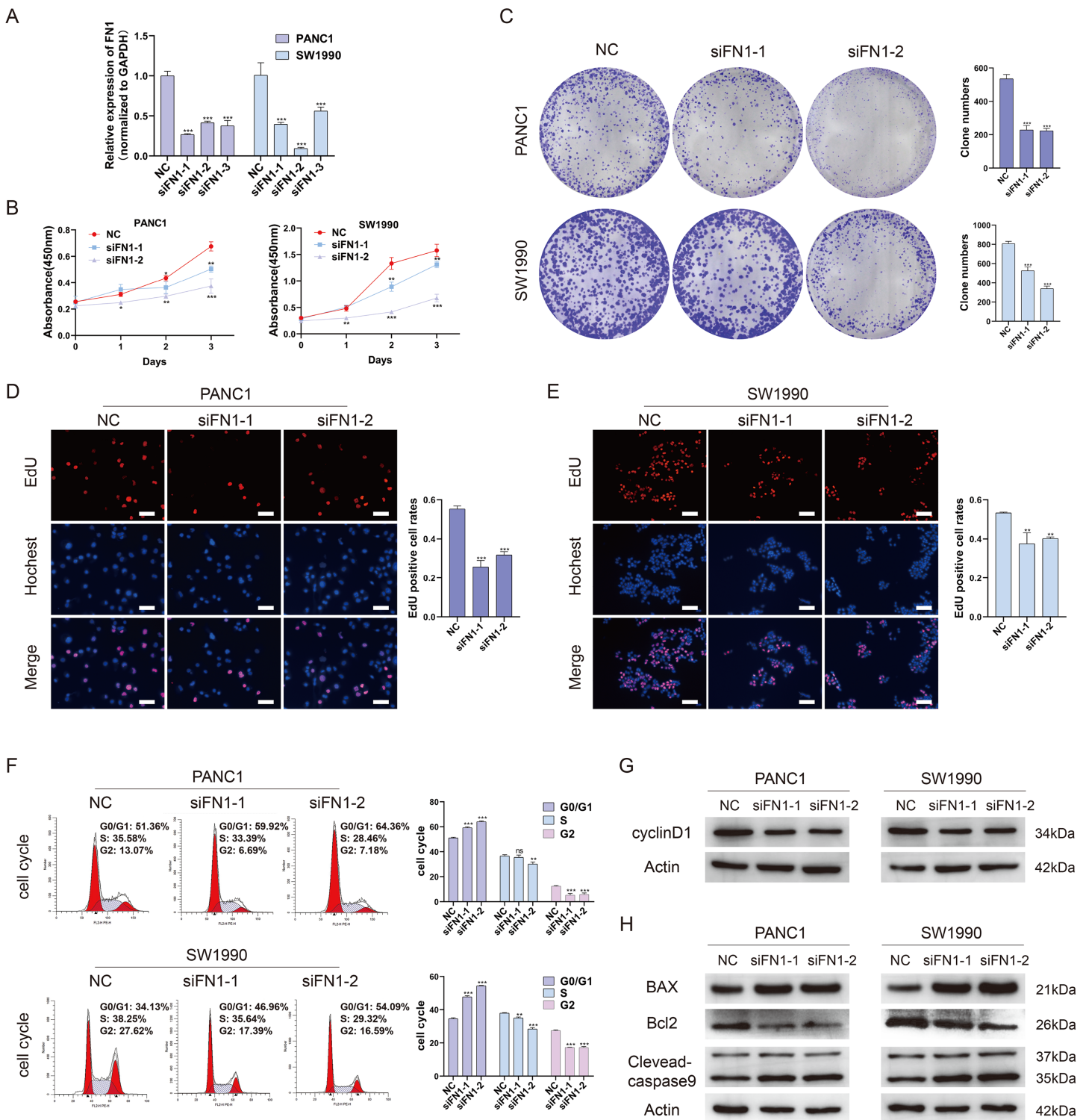
Figure 8 Depletion of FN1 inhibited cell proliferation, blocked cell cycle and induced apoptosis of PDAC cells. (A) The knockdown efficiency of the siRNAs against FN1 was confirmed by qRT-PCR. (B–E) CCK-8, colony formation and EdU assays were performed to evaluate the proliferation of the PDAC cell lines. (F, G) Cell cycle was analyzed by flow cytometry and western blot. (H) Cell apoptosis marker Bax, clevaead-caspase9 and Bcl2 were analyzed by western blot. PDAC, pancreatic carcinoma; CCK-8, Cell Counting Cit-8; EdU, 5-Ethynyl-2′-deoxyuridine. Scale bar = 50 μm. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.
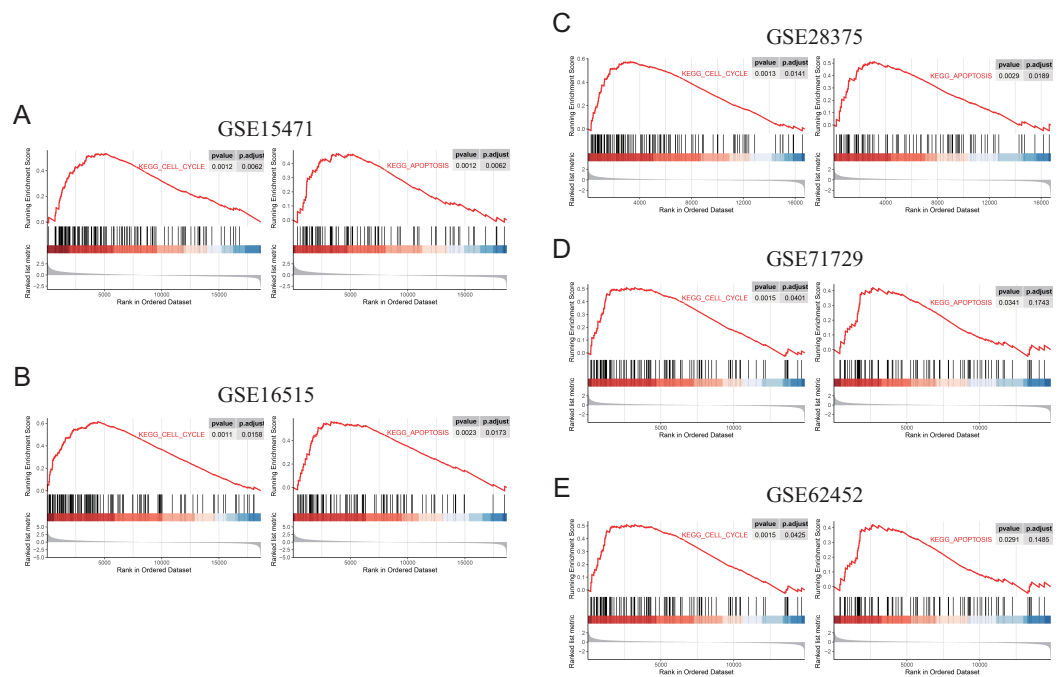
Figure 9 GSEA (gene sets "KEGG_CELL_CYCLE" and "KEGG_APOPTOSIS") results of discovery GEO datasets and the validation GEO dataset. (A) GSEA results of GSE15471. (B) GSEA results of GSE16515. (C) GSEA results of GSE28375. (D) GSEA results of GSE71729. (E) GSEA results of GSE62452.

analyses based on a large number of clinical samples from different sources were conducted. In this study, four GEO datasets (GSE15471, GSE16515, GSE28735 and GSE71729) were analyzed as discovery datasets to screen DEGs that may play important roles in pancreatic carcinogenesis. Next, a PPI network was constructed, and GO annotation and KEGG pathway enrichment analyses of the DEGs were performed *via* the STRING database and R software. Then, hub genes were screened through an integrated analysis of the correlation between these DEGs, and their expression levels in PDAC were subsequently identified *via* the GEPIA, ONCOMINE, HPA databases and validation GEO dataset GSE62452. Nine genes (COL1A1, COL1A2, FN1, ITGA2, KRT19, LCN2, MMP9, MUC1 and VCAN) were considered hub genes, and all of them were significantly highly expressed in PDAC tissues at both the mRNA and protein levels. Furthermore, the prognostic values of these nine hub genes were examined by Kaplan-Meier plotter and the validation GEO dataset GSE62452. In the prognostic analysis of the two groups, only ITGA2 and FN1 had significant predictive values.

ITGA2 is the alpha subunit of a transmembrane receptor for collagens and related proteins. Overexpression of ITGA2 promotes the proliferation and invasion of tumor cells, while its knockdown decreases the degree of these processes (*Ren et al., 2019*). ITGA2 has been reported to play a critical role in modulating the pancreatic cancer immune response by transcriptionally increasing the expression of PD-L1 in cancer cells (*Ren et al., 2019*). At the same time, it has been reported that ITGA2 is associated with unfavorable survival of PDAC patients by affecting the malignant biological behavior of pancreatic

cancer cells (*Rozengurt, Sinnett-Smith & Eibl, 2018*; *Shimomura et al., 2020*; *Yan et al., 2020*). Considering that the mechanism of ITGA2 in PDAC has been studied partly, we focused on FN1 which has no mechanism study in PDAC for further study.

FN1 is a glycoprotein present in a soluble dimeric form in plasma and in a dimeric or multimeric form at the cell surface and in extracellular matrix. Fibronectin is involved in cell adhesion and migration processes, including embryogenesis, wound healing, blood coagulation, host defense, and metastasis (*Glasner et al., 2018*; *Griffith et al., 2006*). Recent studies have shown that FN1 plays an important role in the occurrence and development of a variety of tumors (*Ding, Liu & Lai, 2020*; *Lee et al., 2015*; *Yoshihara et al., 2020*; *Zhan et al., 2018*; *Zhou et al., 2019*). For example, FN1 was highly expressed in sporadic medullary thyroid cancer tissues, and the immunohistochemical score of FN1 in tumor tissues was negatively correlated with American Joint Committee on Cancer stage, tumor histopathological typing and lymph node classification (*Zhan et al., 2018*). Peritoneal dissemination of ovarian cancer induces intercellular interactions through the FN1/AKT signaling pathway and induces the reduction in platinum sensitivity in ovarian cancer cells (*Yoshihara et al., 2020*). The FN1 gene may provide an oligomerization domain of its constituent active promoter and encoding protein to overexpress and promote the activation of the FGFR1 kinase domain in mesenchymal tumors (*Lee et al., 2015*). Furthermore, there are also some studies reported that FN1 is highly expressed in PDAC, but its mechanistic role of FN1 in PDAC remains fully unknown (*Ding, Liu & Lai, 2020*; *Zhou et al., 2019*).

Based on this knowledge, we focused on FN1, which has both significant differential expression and remarkably significant prognostic value in PDAC. In our study, the biological functions of FN1 in PDAC growth were validated in the pancreatic cancer cell lines PANC1 and SW1990. Knockdown of FN1 in PDAC cell lines through siRNA inhibited cell proliferation and colony formation rates. The results provide strong support for our computational analyses, which suggests that FN1 may be a potential biomarker for the selection of therapeutic strategies for PDAC. Regarding the mechanism, knockdown of FN1 promotes cell apoptosis and induces cell cycle arrest by reducing the expression of cyclin D1, indicating that FN1 contributes to the development of pancreatic cancer by affecting cell apoptosis and the cell cycle in a certain manner.

CONCLUSIONS

In summary, two remarkable hub genes (FN1 and ITGA2), that are abnormally overexpressed and predict a poor prognosis in PDAC patients were identified in our study based on a large number of clinical samples. In addition, for the first time, we highlighted the association between FN1 and cell viability, apoptosis and the cell cycle in pancreatic cancer cells. FN1 might be a novel biomarker for therapy and prognostic prediction in patients with PDAC. More importance should be attached to FN1 and other hub genes in further studies of PDAC, and further explorations into relative molecular mechanisms should be performed.

ACKNOWLEDGEMENTS

The authors want to thank the TCGA, GEO, ONCOMINE and HPA databases for free use.

Funding

This work was supported by the Scientific Research Fund Project of Hunan Provincial Health Commission (20200284 and 20201919). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Scientific Research Fund Project of Hunan Provincial Health Commission: 20200284 and 20201919.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Xiaohua Lei conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Guodong Chen performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Jiangtao Li performed the experiments, prepared figures and/or tables, and approved the final draft.
- Wu Wen performed the experiments, prepared figures and/or tables, and approved the final draft.
- Jian Gong conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Jie Fu conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:
The data are available at NCBI: GSE15471, GSE16515, GSE28735, GSE71729 and GSE62452.

Data Availability

Data are available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.12141#supplemental-information.

REFERENCES

Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, Miller DK, Wilson PJ, Patch AM, Wu J, Chang DK, Cowley MJ, Gardiner BB, Song S, Harliwong I, Idrisoglu S, Nourse C, Nourbakhsh E, Manning S, Wani S, Gongora M, Pajic M, Scarlett CJ, Gill AJ, Pinho AV, Rooman I, Anderson M, Holmes O, Leonard C, Taylor D, Wood S, Xu Q, Nones K, Fink JL, Christ A, Bruxner T, Cloonan N, Kolle G, Newell F, Pinese M, Mead RS, Humphris JL, Kaplan W, Jones MD, Colvin EK, Nagrial AM, Humphrey ES, Chou A, Chin VT, Chantrill LA, Mawson A, Samra JS, Kench JG, Lovell JA, Daly RJ, Merrett ND, Toon C, Epari K, Nguyen NQ, Barbour A, Zeps N, Australian Pancreatic Cancer Genome I, Kakkar N, Zhao F, Wu YQ, Wang M, Muzny DM, Fisher WE, Brunicardi FC, Hodges SE, Reid JG, Drummond J, Chang K, Han Y, Lewis LR, Dinh H, Buhay CJ, Beck T, Timms L, Sam M, Begley K, Brown A, Pai D, Panchal A, Buchner N, De Borja R, Denroche RE, Yung CK, Serra S, Onetto N, Mukhopadhyay D, Tsao MS, Shaw PA, Petersen GM, Gallinger S, Hruban RH, Maitra A, Iacobuzio-Donahue CA, Schulick RD, Wolfgang CL, Morgan RA, Lawlor RT, Capelli P, Corbo V, Scardoni M, Tortora G, Tempero MA, Mann KM, Jenkins NA, Perez-Mancera PA, Adams DJ, Largaespada DA, Wessels LF, Rust AG, Stein LD, Tuveson DA, Copeland NG, Musgrove EA, Scarpa A, Eshleman JR, Hudson TJ, Sutherland RL, Wheeler DA, Pearson JV, McPherson JD, Gibbs RA, Grimmond SM. 2012. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491(7424)**:399–405 DOI 10.1038/nature11547.

Buddhan S, Sivakumar R, Dhandapani N, Ganesan B, Anandan R. 2007. Protective effect of dietary squalene supplementation on mitochondrial function in liver of aged rats. *Prostaglandins, Leukotrienes and Essential Fatty Acids* **76(6)**:349–355 DOI 10.1016/j.plefa.2007.05.001.

Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. 2014. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology* **8(Suppl 4)**:S11 DOI 10.1186/1752-0509-8-s4-s11.

Ding J, Liu Y, Lai Y. 2020. Identifying *MMP14* and *COL12A1* as a potential combination of prognostic biomarkers in pancreatic ductal adenocarcinoma using integrated bioinformatics analysis. *PeerJ* **8(3)**:e10419 DOI 10.7717/peerj.10419.

Fu J, Zhang X, Yan L, Shao Y, Liu X, Chu Y, Xu G, Xu X. 2021. Identification of the hub gene BUB1B in hepatocellular carcinoma via bioinformatic analysis and in vitro experiments. *PeerJ* **9(6447)**:e10943 DOI 10.7717/peerj.10943.

Glasner A, Levi A, Enk J, Isaacson B, Viukov S, Orlanski S, Scope A, Neuman T, Enk CD, Hanna JH, Sexl V, Jonjic S, Seliger B, Zitvogel L, Mandelboim O. 2018. NKp46 receptor-mediated interferon-γ production by natural killer cells increases fibronectin 1 to alter tumor architecture and control metastasis. *Immunity* **48(1)**:107–119.e104 DOI 10.1016/j.immuni.2017.12.007.

Griffith OL, Melck A, Jones SJ, Wiseman SM. 2006. Meta-analysis and meta-review of thyroid cancer gene expression profiling studies identifies important diagnostic biomarkers. *Journal of Clinical Oncology* **24(31)**:5043–5051 DOI 10.1200/JCO.2006.06.7330.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. 2011. Global cancer statistics. *CA: A Cancer Journal for Clinicians* **61**:69–90 DOI 10.3322/caac.20107.

Jiang Y, Sun A, Zhao Y, Ying W, Sun H, Yang X, Xing B, Sun W, Ren L, Hu B, Li C, Zhang L, Qin G, Zhang M, Chen N, Zhang M, Huang Y, Zhou J, Zhao Y, Liu M, Zhu X, Qiu Y, Sun Y, Huang C, Yan M, Wang M, Liu W, Tian F, Xu H, Zhou J, Wu Z, Shi T, Zhu W, Qin J, Xie L, Fan J, Qian X, He F. 2019. Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* **567(7747)**:257–261 DOI 10.1038/s41586-019-0987-8.

Kamisawa T, Wood LD, Itoi T, Takaori K. 2016. Pancreatic cancer. *Lancet* **388(10039)**:73–85 DOI 10.1016/S0140-6736(16)00141-0.

Kolde R, Laur S, Adler P, Vilo J. 2012. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* **28(4)**:573–580 DOI 10.1093/bioinformatics/btr709.

Kristensen LS, Andersen MS, Stagsted LVW, Ebbesen KK, Hansen TB, Kjems J. 2019. The biogenesis, biology and characterization of circular RNAs. *Nature Reviews Genetics* **20(11)**:675–691 DOI 10.1038/s41576-019-0158-7.

Lee JC, Jeng YM, Su SY, Wu CT, Tsai KS, Lee CH, Lin CY, Carter JM, Huang JW, Chen SH, Shih SR, Marino-Enriquez A, Chen CC, Folpe AL, Chang YL, Liang CW. 2015. Identification of a novel FN1-FGFR1 genetic fusion as a frequent event in phosphaturic mesenchymal tumour. *Journal of Pathology* **235(4)**:539–545 DOI 10.1002/path.4465.

Li C, Sun YD, Yu GY, Cui JR, Lou Z, Zhang H, Huang Y, Bai CG, Deng LL, Liu P, Zheng K, Wang YH, Wang QQ, Li QR, Wu QQ, Liu Q, Shyr Y, Li YX, Chen LN, Wu JR, Zhang W, Zeng R. 2020. Integrated omics of metastatic colorectal cancer. *Cancer Cell* **38(5)**:734–747.e739 DOI 10.1016/j.ccell.2020.08.002.

Li J, Wang W, Xia P, Wan L, Zhang L, Yu L, Wang L, Chen X, Xiao Y, Xu C. 2018. Identification of a five-lncRNA signature for predicting the risk of tumor recurrence in patients with breast cancer. *International Journal of Cancer* **143(9)**:2150–2160 DOI 10.1002/ijc.31573.

Lusito E, Felice B, D'Ario G, Ogier A, Montani F, Di Fiore PP, Bianchi F. 2019. Unraveling the role of low-frequency mutated genes in breast cancer. *Bioinformatics* **35(1)**:36–46 DOI 10.1093/bioinformatics/bty520.

Massard C, Michiels S, Ferte C, Le Deley MC, Lacroix L, Hollebecque A, Verlingue L, Ileana E, Rosellini S, Ammari S, Ngo-Camus M, Bahleda R, Gazzah A, Varga A, Postel-Vinay S, Loriot Y, Even C, Breuskin I, Auger N, Job B, De Baere T, Deschamps F, Vielh P, Scoazec JY, Lazar V, Richon C, Ribrag V, Deutsch E, Angevin E, Vassal G, Eggermont A, Andre F, Soria JC. 2017. High-throughput genomics and clinical outcome in hard-to-treat advanced cancers: results of the MOSCATO 01 trial. *Cancer Discovery* **7**:586–595 DOI 10.1158/2159-8290.CD-16-1396.

Neoptolemos JP, Kleeff J, Michl P, Costello E, Greenhalf W, Palmer DH. 2018. Therapeutic developments in pancreatic cancer: current and future perspectives. *Nature Reviews Gastroenterology & Hepatology* **15(6)**:333–348 DOI 10.1038/s41575-018-0005-x.

Okada Y, Nishiwada S, Yamamura K, Sho M, Baba H, Takayama T, Goel A. 2021. Identification of laminin γ2 as a prognostic and predictive biomarker for determining response to gemcitabine-based therapy in pancreatic ductal adenocarcinoma. *European Journal of Cancer* **146**:125–134 DOI 10.1016/j.ejca.2020.12.031.

Ouyang W, Winsnes CF, Hjelmare M, Cesnik AJ, Åkesson L, Xu H, Sullivan DP, Dai S, Lan J, Jinmo P, Galib SM, Henkel C, Hwang K, Poplavskiy D, Tunguz B, Wolfinger RD, Gu Y, Li C, Xie J, Buslov D, Fironov S, Kiselev A, Panchenko D, Cao X, Wei R, Wu Y, Zhu X, Tseng KL, Gao Z, Ju C, Yi X, Zheng H, Kappel C, Lundberg E. 2019. Analysis of the human protein atlas image classification competition. *Nature Methods* **16(12)**:1254–1261 DOI 10.1038/s41592-019-0658-6.

Ren D, Zhao J, Sun Y, Li D, Meng Z, Wang B, Fan P, Liu Z, Jin X, Wu H. 2019. Overexpressed ITGA2 promotes malignant tumor aggression by up-regulating PD-L1 expression through the activation of the STAT3 signaling pathway. *Journal of Experimental & Clinical Cancer Research* **38(1)**:485 DOI 10.1186/s13046-019-1496-1.

Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P, Varambally S, Ghosh D, Chinnaiyan AM. 2007. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9(2)**:166–180 DOI 10.1593/neo.07112.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43(7)**:e47 DOI 10.1093/nar/gkv007.

Rozengurt E, Sinnett-Smith J, Eibl G. 2018. Yes-associated protein (YAP) in pancreatic cancer: at the epicenter of a targetable signaling network associated with patient survival. *Signal Transduction and Targeted Therapy* **3(1)**:11 DOI 10.1038/s41392-017-0005-2.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13(11)**:2498–2504 DOI 10.1101/gr.1239303.

Shimomura H, Okada R, Tanaka T, Hozaka Y, Wada M, Moriya S, Idichi T, Kita Y, Kurahara H, Ohtsuka T, Seki N. 2020. Role of miR-30a-3p regulation of oncogenic targets in pancreatic ductal adenocarcinoma pathogenesis. *International Journal of Molecular Sciences* **21(18)**:6459 DOI 10.3390/ijms21186459.

Sidaway P. 2017. Pancreatic cancer: TCGA data reveal a highly heterogeneous disease. *Nature Reviews Clinical Oncology* **14(11)**:648 DOI 10.1038/nrclinonc.2017.146.

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. 2015. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* **43(D1)**:D447–D452 DOI 10.1093/nar/gku1003.

Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. 2017. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research* **45(W1)**:W98–W102 DOI 10.1093/nar/gkx247.

Wang W, Lou W, Ding B, Yang B, Lu H, Kong Q, Fan W. 2019. A novel mRNA-miRNA-lncRNA competing endogenous RNA triple sub-network associated with prognosis of pancreatic cancer. *Aging* **11(9)**:2610–2627 DOI 10.18632/aging.101933.

Wu K, Yin X, Jin Y, Liu F, Gao J. 2019. Identification of aberrantly methylated differentially expressed genes in prostate carcinoma using integrated bioinformatics. *Cancer Cell International* **19(1)**:51 DOI 10.1186/s12935-019-0763-8.

Xu JY, Zhang C, Wang X, Zhai L, Ma Y, Mao Y, Qian K, Sun C, Liu Z, Jiang S, Wang M, Feng L, Zhao L, Liu P, Wang B, Zhao X, Xie H, Yang X, Zhao L, Chang Y, Jia J, Wang X, Zhang Y, Wang Y, Yang Y, Wu Z, Yang L, Liu B, Zhao T, Ren S, Sun A, Zhao Y, Ying W, Wang F, Wang G, Zhang Y, Cheng S, Qin J, Qian X, Wang Y, Li J, He F, Xiao T, Tan M. 2020. Integrative proteomic characterization of human lung adenocarcinoma. *Cell* **182(1)**:245–261.e217 DOI 10.1016/j.cell.2020.05.043.

Yan W, Liu X, Wang Y, Han S, Wang F, Liu X, Xiao F, Hu G. 2020. Identifying drug targets in pancreatic ductal adenocarcinoma through machine learning, analyzing biomolecular networks, and structural modeling. *Frontiers in Pharmacology* **11**:534 DOI 10.3389/fphar.2020.00534.

Yoshihara M, Kajiyama H, Yokoi A, Sugiyama M, Koya Y, Yamakita Y, Liu W, Nakamura K, Moriyama Y, Yasui H, Suzuki S, Yamamoto Y, Ricciardelli C, Nawa A, Shibata K,

Kikkawa F. 2020. Ovarian cancer-associated mesothelial cells induce acquired platinum-resistance in peritoneal metastasis via the FN1/Akt signaling pathway. *International Journal of Cancer* 146(8):2268–2280 DOI 10.1002/ijc.32854.

Zhan S, Li J, Wang T, Ge W. 2018. Quantitative proteomics analysis of sporadic medullary thyroid cancer reveals FN1 as a potential novel candidate prognostic biomarker. *Oncologist* 23(12):1415–1425 DOI 10.1634/theoncologist.2017-0399.

Zhang J, Bai R, Li M, Ye H, Wu C, Wang C, Li S, Tan L, Mai D, Li G, Pan L, Zheng Y, Su J, Ye Y, Fu Z, Zheng S, Zuo Z, Liu Z, Zhao Q, Che X, Xie D, Jia W, Zeng MS, Tan W, Chen R, Xu RH, Zheng J, Lin D. 2019. Excessive miR-25-3p maturation via N(6)-methyladenosine stimulated by cigarette smoke promotes pancreatic cancer progression. *Nature Communications* 10(1):1858 DOI 10.1038/s41467-019-09712-x.

Zhou YY, Chen LP, Zhang Y, Hu SK, Dong ZJ, Wu M, Chen QX, Zhuang ZZ, Du XJ. 2019. Integrated transcriptomic analysis reveals hub genes involved in diagnosis and prognosis of pancreatic cancer. *Molecular Medicine* 25(1):47 DOI 10.1186/s10020-019-0113-2.

Zhou P, Li B, Liu F, Zhang M, Wang Q, Liu Y, Yao Y, Li D. 2017. The epithelial to mesenchymal transition (EMT) and cancer stem cells: implication for treatment resistance in pancreatic cancer. *Molecular Cancer* 16(1):52 DOI 10.1186/s12943-017-0624-9.

Zhu Z, Song J, Guo Y, Huang Z, Chen X, Dang X, Huang Y, Wang Y, Ou W, Yang Y, Yu W, Liu CY, Cui L. 2020. LAMB3 promotes tumour progression through the AKT-FOXO3/4 axis and is transcriptionally regulated by the BRD2/acetylated ELK4 complex in colorectal cancer. *Oncogene* 39(24):4666–4680 DOI 10.1038/s41388-020-1321-5.