ITGA2, LAMB3, and LAMC2 May Be the Potential Therapeutic Targets in Pancreatic Ductal Adenocarcinoma: an Integrated Bioinformatics Analysis

Shajedul Islam
Health Sciences University of Hokkaido
Takao Kitagawa
Health Sciences University of Hokkaido
Byron Baron
University of Malta
Yoshihiro Abiko
Health Sciences University of Hokkaido
Itsuo Chiba
Health Sciences University of Hokkaido
Yasuhiro Kuramitsu (climates@hoku-iryo-u.ac.jp)
Health Sciences University of Hokkaido

Research Article

Keywords: Pancreatic ductal adenocarcinoma, ITGA2, LAMB3, LAMC2

Posted Date: April 7th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-379158/v1

ITGA2, LAMB3, and LAMC2 may be the potential therapeutic targets in pancreatic ductal adenocarcinoma: an integrated bioinformatics analysis

Shajedul Islam[1], Takao Kitagawa[1], Byron Baron[2], Yoshihiro Abiko[3], Itsuo Chiba[4], and Yasuhiro Kuramitsu*[1]

1. Advanced Research Promotion Center, Health Sciences University of Hokkaido, 1757 Kanazawa, Ishikari-Tobetsu, Hokkaido 061-0293, Japan.
2. Centre for Molecular Medicine and Biobanking, University of Malta, Msida, MSD 2080, Malta.
3. Division of Oral Medicine and Pathology, Department of Human Biology and Pathophysiology, School of Dentistry, Health Sciences University of Hokkaido, 1757 Kanazawa, Ishikari-Tobetsu, Hokkaido 061-0293, Japan.
4. Division of Disease Control and Molecular Epidemiology, Department of Oral Growth and Development, School of Dentistry, Health Sciences University of Hokkaido, 1757 Kanazawa, Ishikari-Tobetsu, Hokkaido 061-0293, Japan.

*To whom correspondence should be addressed: Professor Yasuhiro Kuramitsu
Advanced Research Promotion Center, Health Sciences University of Hokkaido, 1757 Kanazawa, Ishikari-Tobetsu, Hokkaido 061-0293, Japan. E-mail: climates@hoku-iryo-u.ac.jp

Keywords: Pancreatic ductal adenocarcinoma, ITGA2, LAMB3, LAMC2

ORCID ID: https://orcid.org/0000-0002-4939-1001 (Shajedul Islam)

Declarations

Funding

No funding was received

Disclosure of potential conflicts of Interest

The authors declare that they have no competing interests.

Availability of data and material

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability

The R scripts generated during the analysis of microarray datasets were submitted as a PDF file.

Author's contribution

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Shajedul Islam, Takao Kitagawa, Yoshihiro Abiko, Itsuo Chiba, and Yasuhiro Kuramitsu. The first draft of the manuscript was written by Shajedul Islam, Takao Kitagawa, Byron Baron, Yasuhiro Kuramitsu, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethics approval for human and animal experiments

Not applicable

Consent to participate

Not applicable

Consent for publication

Not applicable

Acknowledgments

Not applicable

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is the most common form of pancreatic cancer with an abysmal prognosis rate over the last few decades. Early diagnosis and prevention could effectively combat this malignancy. Therefore, it is crucial to discover potential biomarkers to identify asymptomatic premalignant or early malignant tumors of PDAC. Gene expression analysis is a powerful technique to identify candidate biomarkers involved in disease progression. In the present study, five independent gene expression datasets, including 321 PDAC tissues and 208 adjacent non-cancerous tissue samples, were subjected to statistical and bioinformatics analysis. A total of 20 differentially expressed genes (DEGs) were identified in PDAC tissues compared to non-cancerous tissue samples. Gene ontology and pathway enrichment analysis showed that DEGs were mainly enriched in extracellular matrix (ECM), cell adhesion, ECM-receptor interaction, and focal adhesion signaling. The protein-protein interaction network was constructed, and the hub genes were evaluated. Collagen type XII alpha 1 chain (COL12A1), fibronectin 1 (FN1), integrin subunit alpha 2 (ITGA2), laminin subunit beta 3 (LAMB3), laminin subunit gamma 2 (LAMC2), thrombospondin 2 (THBS2), and versican (VCAN) were identified as hub genes. The correlation analysis revealed that identified hub genes were significantly interconnected. Wherein COL12A1, FN1, ITGA2, LAMB3, LAMC2, and THBS2 were significantly associated with PDAC pathological stages. The Kaplan-Meier survival plots revealed that ITGA2, LAMB3, and LAMC2 expression were inversely correlated with a prolonged patient survival period. Furthermore, the Human Protein Atlas database was used to validate the expression and cellular origins of hub genes encoded proteins. The protein expression of hub genes was higher in pancreatic cancer tissue than in normal pancreatic tissue samples, wherein ITGA2, LAMB3, and LAMC2 were exclusively expressed in pancreatic cancer cells. Pancreatic cancer cell-specific expression of these three proteins may play pleiotropic roles in cancer progression. Our results collectively suggest that ITGA2, LAMB3, and LAMC2 could provide deep insights into pancreatic carcinogenesis molecular mechanisms and provide attractive therapeutic targets.

Key points

1. Regardless of ethnic differences, the expression of ITGA2, LAMB3, and LAMC2 is inversely linked to patient survival period with PDAC.
2. ITGA2, LAMB3, and LAMC2 individually might have high prognostic and diagnostic value and be potential therapeutic targets for PDAC.
3. The expression of ECM proteins by cancer cells may be critical for PDAC pathogenesis and progression.
4. Targeting ECM proteins that are only expressed by cancer cells could render them promising potential targets for diagnosing and treating PDAC.

Introduction

Pancreatic ductal adenocarcinoma (PDAC) is the most aggressive and common form of pancreatic cancer, accounting for 95% of all pancreatic malignant neoplasms [1]. The 5-year overall survival rate for patients with PDAC is less than 8% despite advances in medical oncology [2]. The poor prognosis of PDAC may be due to the lack of precise molecular biomarkers for early diagnosis and prognosis [3]. Therefore, there is an urgent need for more effective targeted therapies to improve the survival rate of patients with PDAC [4].

Gene expression microarrays and gene chips are extensively applied to reveal genetic aspects of diseases. These techniques are routinely used to monitor genome-wide expression levels of genes and are particularly suitable for screening differentially expressed genes (DEGs) between two samples [5]. The identification of DEGs may elucidate cancer pathogenesis, provide early diagnosis, and improve treatment. Hence, gene expression microarray analysis could be a promising approach to identify candidate biomarkers involved in disease progression.

The gene expression profiles from diverse microarray platforms are submitted to several public databases, including Gene Expression Omnibus (GEO: https://www.ncbi.nlm.nih.gov/gds/). Several previous studies used gene expression microarray technology to underpinning the DEGs of PDAC in recent years [6,7]. However, the results were inconsistent, and various aspects remain unclear due to sample heterogeneity. Moreover, those studies have not considered ethnic differences, and many studies have proven that ethnic differences may have relevance for disease gene expression profiles [8,9]. The present study aimed to improve DEGs accuracy and reliability in PDAC compared to adjacent non-cancerous tissue samples using several datasets from different ethnicities.

In the current study, gene expression datasets from PDAC were analyzed to identify DEGs. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment were performed using an online toolset. Then, the protein interaction networks were constructed and the hub genes were identified and further verified. The identified hub genes may serve as potential diagnostic and prognostic biomarkers and could be a promising approach for the treatment of PDAC. To the best of our knowledge, this analysis is the first to examine the gene expression microarray database in PDAC tissues and adjacent non-cancerous tissue samples, considering different ethnic groups.

Materials and Methods

1. Microarray datasets information

PDAC datasets were obtained from the Gene Expression Omnibus, a public functional genomic database containing high-throughput gene expression data, chips, and microarrays. The GEO database was searched using the following criteria: “human-derived pancreatic ductal adenocarcinoma tissues and adjacent non-cancerous tissue samples” (study keyword), “Homo sapiens” (organism), “expression profiling by array” (study type), “tissue” (attribute name), and “sample count” >50. After a systematic review, five independent PDAC microarray datasets were selected, including GSE62452 [10], GSE28735 [11], GSE15471 [12], GSE62165 [13], GSE102238 [14], with 321 primary tumor samples and 208 adjacent non-cancerous samples. The dataset GSE62452 was based on the GPL6244 platform (HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array) and included 69 tumor and 61 adjacent non-cancerous tissue samples. The dataset GSE28735 was based on the GPL6244 platform (HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array) and had 45 matched tumor and adjacent non-cancerous samples. The GSE15471 dataset was produced using the GPL570 Platform [(HG-U133_Plus_2) Affymetrix Human Genome U133 Plus 2.0 Array], including 39 matched tumors and adjacent non-cancerous samples. The GSE62165 dataset was based on the GPL13667 Platform [(HG-U219) Affymetrix Human Genome U219 Array], which

contained 118 tumors and 13 adjacent non-cancerous samples. The GSE102238 dataset was based on the GPL19072 Platform [Agilent-052909 CBC_lncRNAmRNA_V3], which included 50 matched tumor and adjacent non-cancerous samples. These 5 gene expression profiles were respectively from different regions, including North America, Europe, and Asia, thus averting the differences caused by sample heterogeneity of single profiles and revealing universal DEGs that apply to different ethnic groups, as it has been reported that ethnic difference may affect disease-associated gene expression profiles [8,9]. The clinical datasets included 321 tumors and 208 adjacent non-cancerous tissues diagnosed as PDAC (Table 1).

2. Identification of DEGs

DEGs between PDAC and adjacent non-cancerous tissue samples were screened by GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r) [15], an online tool that can be used to compare two or more datasets in a GEO series to identify DEGs according to the experimental conditions. Adjusted *p*-values (adj. *p*) and Benjamini and Hochberg false discovery rates were employed as criteria for statistically significant genes and to limit false positives. The data normalization was applied for the five datasets (supplementary Fig. 1). Probe sets with no corresponding gene symbols were removed, while genes with multiple gene probe sets were averaged. Log2 FC (fold change) ≥1.5 or ≥-1.5 and adj. $p<0.01$ was considered statistically significant. An online tool (http://www.interactivenn.net) was applied to draw Venn diagrams of the DEGs. Further, heatmap analysis was visualized with the Heatmapper web application [16]. A total of 20 DEGs were identified, which consisted of 19 upregulated genes and 1 downregulated gene.

3. External validation of the identified DEGs mRNA expression level

The external validation was done using the Gene Expression Profiling Interactive Analysis tool [17] (http://gepia2.cancer-pku.cn/#index; last access: 14th February 2021) by comparing transcriptomic data from The Cancer Genome Atlas (TCGA) (pancreatic adenocarcinoma), the TCGA normal and the Genotype-Tissue Expression (GTEx) database. $p<0.05$ was considered a statistically significant difference.

4. GO and KEGG pathway analysis of DEGs

To uncover the functional roles of DEGs, the GO was used to perform enrichment analysis, which covers the cellular component (CC), biological process (BP), and molecular function (MF) of the selected genes [18]. The KEGG is a database that illustrates the selected gene functions and pathways [19]. The Database for Annotation, Visualization, and Integrated Discovery (DAVID: https://david.ncifcrf.gov; last access: 14th February 2021) is a public online bioinformatics database that contains information on functional biological annotations for genes and proteins [20]. The cut-off criteria were selected based on $p<0.01$. Enrichment of the GO terms and KEGG pathways were performed for the candidate DEGs using DAVID.

5. Establishment of the PPI network and hub gene identification

To further explore the potential interplay among those DEGs, these were mapped to the STRING (https://string-db.org; version 11.0) database [21] and only interactions that enjoyed a minimum required combined score >0.4 were set as significant. Subsequently, the protein-protein interaction (PPI) networks were visualized using Cytoscape 3.8.2 (https://cytoscape.org/), an open-source bioinformatics software platform [22]. A combined score of 0.5 and a tissue-specific (pancreas) filter score of 1 was considered for the construction of the PPI network. Subsequently, the MCODE (Molecular Complex Detection) plugin was used to identify hub genes in the constructed network. The standard for selection was set as follows: MCODE scores ≥10, degree cut-off = 2, node score cut-off = 0.2, max depth = 100 and k-score = 2 [23].

6. Oncomine analysis of hub genes in pancreatic cancer

An independent database, namely Oncomine (https://www.oncomine.org/resource/login.html; last access: 14th February 2021), was used to validate hub gene expression. In the Oncomine database, the gene name "COL12A1", "FN1", "ITGA2", "LAMB3", "LAMC2", "THBS2" or "VCAN" was entered. The differential gene analysis module (cancer vs. normal analysis) was selected to retrieve the results. This analysis presented a series of pancreatic cancer studies and related COL12A1, FN1, ITGA2, LAMB3, LAMC2, THBS2, and VCAN mRNA expression in cancer and normal tissues. The filters were set as follows: i) Gene: COL12A1 or FN1 or ITGA2 or LAMB3 or LAMC2 or THBS2 or VCAN. ii) Analysis type: cancer vs. normal analysis. iii) Cancer type: pancreatic carcinoma. iv) Sample type: clinical specimen. v) Data type: mRNA. vi) Threshold settings: $p< 0.01$; FC >2; gene rank, top 10%.

7. **Finding prognostic genes for PDAC**

To explore the expression correlation of hub genes in PDAC, the Spearman coefficient correlation was analyzed using the GEPIA2 tool [17]. The interaction efficiency was represented as an R score. An R score of >0.8 was considered a significant correlation. Next, the expression levels of hub genes and pathological stages in PDAC tissues were assessed using the GEPIA2 platform. The GEPIA2 was also utilized for overall survival and disease-free survival analyses of the hub genes using the TCGA and GTEx databases. The plots were considered significant when showed in both overall and disease-free survival states. Beta-actin was used to normalize the expression of genes, and the median was selected for group cut-off criteria. $p<0.05$ was considered to indicate a statistically significant difference. Further, the expression of proteins encoded by hub genes in pancreatic cancer was validated using the Human Protein Atlas (HPA: https://www.proteinatlas.org) website based on spatial proteomics data and quantitative transcriptomics data (RNA-Seq) obtained from the immunohistochemical analysis of tissue microarrays.

Results

1. Identification of DEGs in PDAC

The five gene expression microarray datasets for PDAC, GSE62452, GSE28735, GSE15471, GSE62165, and GSE102238, were obtained from GEO. By screening the data with the GEO2R using $p<0.01$ and log2FC ≥1.5 or ≥-1.5 as cut-off criteria, 2636 upregulated and 1103 downregulated genes were obtained. In brief, 90 DEGs, including 45 upregulated and 45 downregulated genes, were obtained in the GSE62452 expression profile data (Fig. 1a). GSE28735, 127 DEGs, including 66 upregulated and 61 downregulated genes, were identified (Fig. 1b). In GSE15471, 706 DEGs, including 622 upregulated and 84 downregulated genes, were identified (Fig. 1c). 1984 DEGs, including 1380 upregulated and 604 downregulated genes, were identified from GSE62165 (Fig. 1d). In addition, 832 DEGs, including 523 upregulated and 309 downregulated genes, were identified from GSE102238 (Fig. 1e). The overview of the DEGs results was briefly presented in Figure 1f. After a comprehensive analysis of the five datasets, 20 DEGs were identified that were differentially expressed in all of them, with 19 genes up-regulated and 1 down-regulated in PDAC tissues compared to adjacent non-cancerous tissues (Fig. 2a). Figure 2b-c provides a heatmap of the 20 DEGs based on Log2FC. The functions and the involvement of identified DEGs on PDAC tissues are shown in Table 2.

2. The mRNA expression level of DEGs in PDAC

To confirm the mRNA expression levels of identified DEGs in PDAC tissues, TCGA datasets were analyzed using the GEPIA2 platform. Boxplots of the DEGs associated with PDAC were downloaded from the GEPIA2.

The results demonstrated that upregulated DEGs were significantly overexpressed in PDAC tissues in comparison to normal pancreatic tissues, while the downregulated DEG, PDK4 was significantly reduced in PDAC tissues in comparison to normal pancreatic tissues ($p<0.05$) (Fig. 3).

3. GO analysis and signaling pathway enrichment of DEGs in PDAC

To elucidate the functions of common DEGs, GO and KEGG pathway enrichment analysis was employed. In the CC category, the upregulated DEGs were mainly enriched in the ECM and extracellular space. In the BP category, the upregulated DEGs were mainly enriched in ECM organization and cell adhesion. While in MF category, upregulated DEGs were enriched with heparin and collagen binding functions. There was no enrichment showed for downregulated DEGs. The ECM-receptor interaction, focal-adhesion, and phosphoinositide-3-kinase-protein kinase B/Akt (PI3K-Akt) signaling were the most enriched pathways for upregulated DEGs. The results of the functional enrichment and KEGG pathway analyses for DEGs are exhibited in Table 3.

4. PPI network construction and identification of hub nodes

The PPI network of the DEGs was constructed using Cytoscape software and the STRING database. The PPI network of DEGs consisted of 58 nodes and 811 edges (Fig. 4a). The Cytoscape tool MCODE was used to screen hub genes in the network, with a cluster score of ≥10 as the inclusion criterion. The MCODE modules included 46 nodes and 432 edges with two clusters. Cluster-1 included 24 nodes and 260 edges with a combined score of 22.6. Wherein cluster-2 included 22 nodes and 172 edges with a cluster score of 16.4. After a comprehensive analysis, hub genes were identified from two clusters highlighted in red color (Fig. 4b-c). COL12A1, FN1, ITGA2, LAMB3, LAMC2, THBS2, and VCAN were finally selected as hub genes. The MCODE plugin scores are briefly shown in Table 4.

5. Oncomine analysis of hub genes in pancreatic cancer databases

As COL12A1, FN1, ITGA2, LAMB3, LAMC2, THBS2, and VCAN were selected from the other DEGs, further confirmation of the altered expressions was necessary. Oncomine analysis of cancer vs. normal tissue confirmed that COL12A1, FN1, ITGA2, LAMB3, LAMC2, THBS2, and VCAN were significantly overexpressed in pancreatic cancer from different datasets. A brief overview of those key genes expression in pancreatic cancer was shown by using a heatmap. The color intensity reflects the fold changes between different datasets. Moreover, in the Pei pancreas dataset, COL12A1, FN1, ITGA2, LAMB3, LAMC2, THBS2, and VCAN mRNA expression levels were higher in pancreatic cancer tissue than in normal pancreatic tissue samples (Fig. 5).

6. Expression correlation of hub genes in PDAC

To explore the correlation among the hub genes in PDAC, TCGA datasets were analyzed using the GEPIA2 platform. COL12A1, FN1, ITGA2, LAMB3, LAMC2, THBS2, and VCAN were observed to be significantly correlated (Fig. 6).

7. Association of hub genes in PDAC pathological stages

Further analysis of the TCGA PDAC data in GEPIA2 showed that the hub genes were significantly correlated with the pathological disease stages, underlying their prognostic value for PDAC. COL12A1, FN1, ITGA2, LAMB3, LAMC2, and THBS2 were observed to be significantly associated with PDAC stages (Fig. 7), wherein no significant association on PDAC tumor stages and VCAN was observed (data not shown).

8. Survival analysis of hub genes in PDAC

The Kaplan–Meier survival plots were used to observe the overall survival and disease free-survival status of the hub genes in PDAC. Elevated expression levels of ITGA2, LAMB3, and LAMC2 were found to be inversely

correlated with prolonged patient survival (Fig. 8), whereas no significant relationship was observed for other genes (data not shown).

9. Validation of expression of hub genes-encoded proteins

The expression levels of proteins encoded by the COL12A1, FN1, ITGA2, LAMB3, LAMC2, THBS2, and VCAN were obtained. The protein expression profiles in pancreatic cancer clinical specimens are shown in Figure 9. The antibody intensity for FN1, ITGA2, LAMB3, LAMC2, and VCAN was higher in PDAC tissues, while no staining was observed in corresponding normal tissues. COL12A1 had medium staining intensity with low intensity observed in normal pancreatic tissues. THBS2 had medium staining intensity in both pancreatic cancer and normal pancreatic tissues. Further observations revealed that COL12A1 and FN1 were predominantly expressed by stromal cells. THBS2 and VCAN were expressed in both stromal and pancreatic cancer cells, whereas ITGA2, LAMB3, and LAMC2 were solely expressed by pancreatic cancer cells.

Discussion

In the present study, 20 DEGs were identified (19 upregulated and 1 downregulated), which were differentially expressed in PDAC tissue compared to the adjacent non-cancerous pancreatic tissue samples. By using an online tool, the mRNA expression levels of DEGs in PDAC tissue samples were validated. The GO and KEGG pathway analysis revealed that DEGs were primarily enriched with ECM-organization, cell adhesion, ECM-receptor interaction, and focal adhesion, especially for the upregulated genes. The PPI network was constructed, and hub genes were selected. COL12A1, FN1, ITGA2, LAMB3, LAMC2, THBS2, and VCAN were identified as hub genes. To verify the expression level of hub genes, an independent database was then used. This confirmed that, compared to normal pancreatic tissues, identified hub genes were highly expressed in pancreatic cancer samples. The correlation analysis revealed that the hub genes in PDAC tissue samples are significantly interconnected. The interaction of hub genes with pathological stages in patients with PDAC showed that the expression of COL12A1, FN1, ITGA2, LAMB3, LAMC2, and THBS2 is negatively associated with disease progression. The survival plots of Kaplan-Meier showed that ITGA2, LAMB3, and LAMC2 expression are inversely correlated with prolonged patient survival. Using histopathological images from the Human Protein Atlas platform, the protein expression profiles of hub genes were validated. It was found that proteins encoded by hub genes are highly expressed in pancreatic cancer tissue compared to normal pancreatic tissue samples. It was also observed that ITGA2, LAMB3, and LAMC2 were the only proteins expressed in pancreatic cancer cells but not in stromal cells. The cancer cells specific expression of these three proteins might be crucial for PDAC pathogenesis and progression. Together, this data suggested that ITGA2, LAMB3, and LAMC2 individually might have high prognostic and diagnostic values, as well as the potential to be therapeutic targets for PDAC.

ITGA2 is a collagen receptor expressed on cell membranes and forms a heterodimer α2β1 with a β subunit, which mediates cell-to-ECM attachment [24]. The increased ITGA2 level was reported in pancreatic cancer and others, including gastric, liver, prostate, and breast cancer [25]. The increased ITGA2 expression promotes pancreatic cancer cell migration, invasion, metastasis, and chemoresistance [26,27]. In contrast, inhibition of ITGA2 abrogated these functions [25]. Although the exact mechanism by which ITGA2 is involved in pancreatic carcinogenesis remains unclear, it has been suggested that ITGA2 promotes pancreatic cancer progression through ECM remodeling [28,29]. The reconstituted ECM triggers pancreatic cancer progression by directly promoting cellular transformation and enhancing tumorigenic microenvironment formation by affecting stromal-cell behavior [30]. In this process, ITGA2

activates fibroblasts to cancer-associated fibroblasts (CAFs), resulting in extensive desmoplasia with ECM deposition [31], wherein desmoplasia is a characteristic feature of PDAC and constitutes up to 90% of the tumor volume. Mainly ECM and CAF, immune cells, and vascular components form the desmoplastic microenvironment [32,33]. ECM is a three-dimensional structural complex consisting of structural and non-structural proteins [34,35]. ECM-proteins can affect PDAC progression and patient survival by promoting cancer cell proliferation and metastatic spread [36]. Even though stromal cells produce over 90% of the ECM mass in PDAC, cancer cells produce elevated ECM-proteins, and cancer cell-derived ECM-proteins play important roles in PDAC carcinogenesis [37,38]. A previous report suggested that ECM proteins originating from cancer cells were the most strongly connected to poor patient survival. In contrast, ECM-proteins derived from stromal cells, include both proteins linked to good and poor patient outcomes [39]. Hence, using the Human Protein Atlas database, the protein expression profiles and cellular origins of hub genes encoded proteins in pancreatic cancer tissues were observed. ITGA2 is the transmembrane receptor for collagens and related proteins, as mentioned above [24], while COL12A1, FN1, LAMB3, LAMC2, THBS2, and VCAN are ECM-related proteins [39].

Our histopathological evidence has shown that COL12A1 and FN1 are expressed from stromal cells, THBS2, and VCAN from stromal and cancer cells, while ITGA2, LAMB3, and LAMC2 are expressed solely from the cancer cells. The Kaplan-Meier survival plots showed that ITGA2, among the ECM-proteins LAMB3 and LAMC2 expression, is inversely correlated with the overall and disease-free survival status in PDAC. Interestingly, a previous report confirmed that LAMB3 and LAMC2 were exclusively derived from pancreatic cancer cells [39]. This study reached a similar conclusion that increased levels of ECM-proteins originated from cancer cells, rather than being solely produced by stromal cells, correlate with poor patient outcomes. However, further studies are needed to clarify this phenomenon. Meanwhile, these results may explain why previous non-selective ECM depletion strategies led to poor patient outcomes and suggest more accurate ECM manipulations as PDAC treatments [40]. Together, the present data and the previous report suggested that cancer-cell-derived ECM-proteins may be potential therapeutic targets [39]. Therefore, sorting out the composition and changes of the ECM during PDAC progression would guide the development and application of more effective PDAC therapies.

It is worth noting that DEGs in PDAC have already been demonstrated in several studies [6,7]. However, the results were not consistent, which could be due to the differences in the selection of datasets and statistical procedures. In this study, we select the datasets from different regions, thus averting the differences caused by the heterogeneity of the samples and revealing universal DEGs that apply to different ethnic groups.

In conclusion, the present meta-analysis identified 20 DEGs. The hub genes are COL12A1, FN1, ITGA2, LAMB3, LAMC2, THBS2, and VCAN. The Kaplan-Meier survival plots indicate that ITGA2, LAMB3, and LAMC2 are inversely correlated with prolonged patient survival. Histopathological evidence shows that ITGA2, LAMB3, and LAMC2 are expressed exclusively from pancreatic cancer cells. The specific expression of these three proteins by cancer cells could make them promising potential targets for diagnosing and treating pancreatic cancer. However, a lack of adequate validation *in vitro* or *in vivo* is a limitation of this study. Therefore, future research will include experimental verification of our meta-analysis results using different laboratory approaches.

References

1. Bray, F. *et al.* Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries: Global Cancer Statistics 2018. *CA: A Cancer Journal for Clinicians* **68**, (2018).
2. Siegel, R., Miller, K. & Jemal, A. Cancer statistics, 2018. *CA: A Cancer Journal for Clinicians* **68**, (2018).
3. Ying, H. *et al.* Genetics and biology of pancreatic ductal adenocarcinoma. *Genes & Development* **30**, 355–385 (2016).
4. Baron, B. The unsuccessful hunt for pancreatic cancer biomarkers – time to search deeper in the proteome. **5**, 883–891 (2014).
5. Parkinson, H. *et al.* ArrayExpress update—an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic acids research* **39**, D1002-4 (2010).
6. Tang, Y., Zhang, Z., Tang, Y., Chen, X. & Zhou, J. Identification of potential target genes in pancreatic ductal adenocarcinoma by bioinformatics analysis. *Oncology Letters* **16**, (2018).
7. Jin, D. *et al.* Identification of prognostic risk factors for pancreatic cancer using bioinformatics analysis. *PeerJ* **8**, e9301 (2020).
8. Mitchell, K., Zingone, A., Toulabi, L., Boeckelman, J. & Ryan, B. Comparative Transcriptome Profiling Reveals Coding and Noncoding RNA Differences in NSCLC from African Americans and European Americans. *Clinical Cancer Research* **23**, 7412–7425 (2017).
9. Hardiman, G. *et al.* Systems analysis of the prostate transcriptome in African–American men compared with European–American men. *Pharmacogenomics* **17**, (2016).
10. Yang, S. *et al.* A Novel MIF Signaling Pathway Drives the Malignant Character of Pancreatic Cancer by Targeting NR3C2. *Cancer Research* **76**, (2016).
11. Zhang, G. *et al.* DPEP1 Inhibits Tumor Cell Invasiveness, Enhances Chemosensitivity and Predicts Clinical Outcome in Pancreatic Ductal Adenocarcinoma. *PloS one* **7**, e31507 (2012).
12. Badea, L., Herlea, V., Dima, S., Dumitrascu, T. & Popescu, I. Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. *Hepato-gastroenterology* **55**, 2016–2027 (2008).
13. Janky, R. *et al.* Prognostic relevance of molecular subtypes and master regulators in pancreatic ductal adenocarcinoma. *BMC Cancer* **16**, (2016).
14. Yang, M. *et al.* Perineural Invasion Reprograms the Immune Microenvironment through Cholinergic Signaling in Pancreatic Ductal Adenocarcinoma. *Cancer Research* **80**, canres.2689.2019 (2020).
15. Shao, N. *et al.* Identification of key candidate genes, pathways and related prognostic values in ER-negative/HER2-negative breast cancer by bioinformatics analysis. *Journal of B.U.ON. : official journal of the Balkan Union of Oncology* **23**, 891–901 (2018).
16. Babicki, S. *et al.* Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Research* **44**, gkw419 (2016).
17. Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic acids research* **47**, (2019).
18. Young, M., Wakefield, M., Smyth, G. & Oshlack, A. Gene ontology analysis for RNA-SEQ: accounting for selection bias. *Genome biology* **11**, R14 (2010).

19. Altermann, E. & Klaenhammer, T. Altermann, E. & Klaenhammer, T.R. PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. BMC Genomics 6, 60. *BMC genomics* **6**, 60 (2005).
20. Dennis, G. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome biology* **4**, P3 (2003).
21. Szklarczyk, D. *et al.* STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic acids research* **43**, (2014).
22. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome research* **13**, 2498–2504 (2003).
23. Li, L., Lei, Q., Zhang, S., Kong, L. & Qin, B. Screening and identification of key biomarkers in hepatocellular carcinoma: Evidence from bioinformatic analysis. *Oncology Reports* **38**, (2017).
24. Tuckwell, D. S., Calderwood, D. A., Green, L. J. & Humphries, M. *Integrin α2 I-domain is a binding site for collagens. *Journal of cell science* **108 (Pt 4**, 1629–1637 (1995).
25. Chuang, Y.-C. *et al.* Blockade of ITGA2 Induces Apoptosis and Inhibits Cell Migration in Gastric Cancer. *Biological Procedures Online* **20**, (2018).
26. Adorno, V. & Liu, H. Regulation and functions of integrin α2 in cell adhesion and disease. *Genes & Diseases* **6**, (2018).
27. Carnevale, I. *et al.* Abstract 4040: α 2 β 1 integrins are potential regulators of chemoresistance through modulation of biomechanical cues in pancreatic cancer. *Cancer Research* **78**, 4040 (2018).
28. Ren, D. *et al.* 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**, (2019).
29. Feig, C. *et al.* The Pancreas Cancer Microenvironment. *Clinical cancer research : an official journal of the American Association for Cancer Research* **18**, 4266–4276 (2012).
30. Venning, F., Wullkopf, L. & Erler, J. Targeting ECM Disrupts Cancer Progression. *Frontiers in Oncology* **5**, (2015).
31. Zeltz, C. *et al.* Cancer-associated fibroblasts in desmoplastic tumors: emerging role of integrins. *Seminars in Cancer Biology* **62**, (2019).
32. Whatcott, C., Posner, R., Von Hoff, D. & Han, H. Desmoplasia and chemoresistance in pancreatic cancer. in (2021).
33. Öhlund, D. *et al.* Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *The Journal of Experimental Medicine* **214**, jem.20162024 (2017).
34. Hynes, R. & Naba, A. Overview of the Matrisome--An Inventory of Extracellular Matrix Constituents and Functions. *Cold Spring Harbor perspectives in biology* **4**, a004903 (2011).
35. Humphrey, J., Dufresne, E. & Schwartz, M. Mechanotransduction and extracellular matrix homeostasis. *Nature reviews. Molecular cell biology* **15**, (2014).
36. Weniger, M., Honselmann, K. & Liss, A. The Extracellular Matrix and Pancreatic Cancer: A Complex Relationship. *Cancers* **10**, 316 (2018).
37. Xiong, G.-F. & Xu, R. Function of cancer cell-derived extracellular matrix in tumor progression. *Journal of Cancer Metastasis and Treatment* **2**, 357 (2016).

38. Tian, C. *et al.* Cancer Cell-Derived Matrisome Proteins Promote Metastasis in Pancreatic Ductal Adenocarcinoma. *Cancer Research* **80**, canres.2578.2019 (2020).
39. Tian, C. *et al.* Proteomic analyses of ECM during pancreatic ductal adenocarcinoma progression reveal different contributions by tumor and stromal cells. *Proceedings of the National Academy of Sciences* **116**, 201908626 (2019).
40. Amakye, D., Jagani, Z. & Dorsch, M. Unraveling the therapeutic potential of the Hedgehog pathway in cancer. *Nature medicine* **19**, 1410–1422 (2013).

Figure legends

Fig1: Differential expression of genes between PDAC tissue and adjacent non-cancerous tissue samples in the datasets. (a) GSE62452; (b) GSE28735; (c) GSE15471; (d) GSE62165; (e) GSE102238. The x-axis indicates the fold-change (log-scaled); the y-axis indicates the *p*-values (log-scaled). The red data-points represent upregulated genes, while blue data-points represent downregulated genes. The black data-points represent genes with no significant difference in expression. (f) The differential genes screened based on |Log2FC|≥ 1.5/-1.5 and a corrected *p*-value of *<0.01*. FC, fold change

Fig 2: Identification of DEGs. (a) Venn diagram of the DEGs in the five datasets. A total of 20 DEGs were included in all 5 datasets. (b) LogFC heatmap image of the 20 commonly changed DEGs of the five datasets. The 5 GEO datasets are denoted on the abscissa, and the gene names are displayed on the ordinate. (c) The average LogFC expression values of 20 DEGs. DEG, differentially expressed gene; FC, fold change

Fig 3: The mRNA expression level analysis of 20 DEGs in PDAC tissues. The boxplots were downloaded from the GEPIA2. The red boxes represent the expression levels in PDAC tissues. In contrast, the blue boxes represent the expression levels in normal tissues. *p<0.05* was regarded as statistically significant. DEGs, differentially expressed genes; PDAC, pancreatic ductal adenocarcinoma

Fig 4: PPI network construction of DEGs and identification of hub genes. (a) PPI network was constructed using Cytoscape. Red nodes represent upregulated genes, whereas green nodes represent downregulated genes. The line represents the interaction relationship between nodes. (b) Significant modules of cluster-1 were identified from the PPI network via the MCODE plug-in. This module consisted of 5 upregulated genes, which are represented by red color. (c) Significant modules of cluster-2 were identified from the PPI network via the MCODE plug-in. This module consisted of 2 upregulated genes, and red nodes represent key genes. PPI, protein-protein interaction; DEGs, differentially expressed genes

Fig 5: Oncomine analysis of key candidate genes in pancreatic cancer vs. normal tissue. Heat maps of key candidate gene expression in clinical pancreatic cancer samples vs. normal pancreatic tissue samples. [1. Pancreatic ductal adenocarcinoma epithelia vs. normal Badea pancreas; 2. Pancreatic ductal adenocarcinoma epithelia vs. normal Lacobuzio-Donahue pancreas; 3. Pancreatic carcinoma vs. normal Pei pancreas]. The color depth represents the intensity of fold changes. Box plots represent the mRNA expression level in pancreatic cancer and normal pancreatic tissues in the Pei pancreas dataset. *p<0.01* was considered statistically significant.

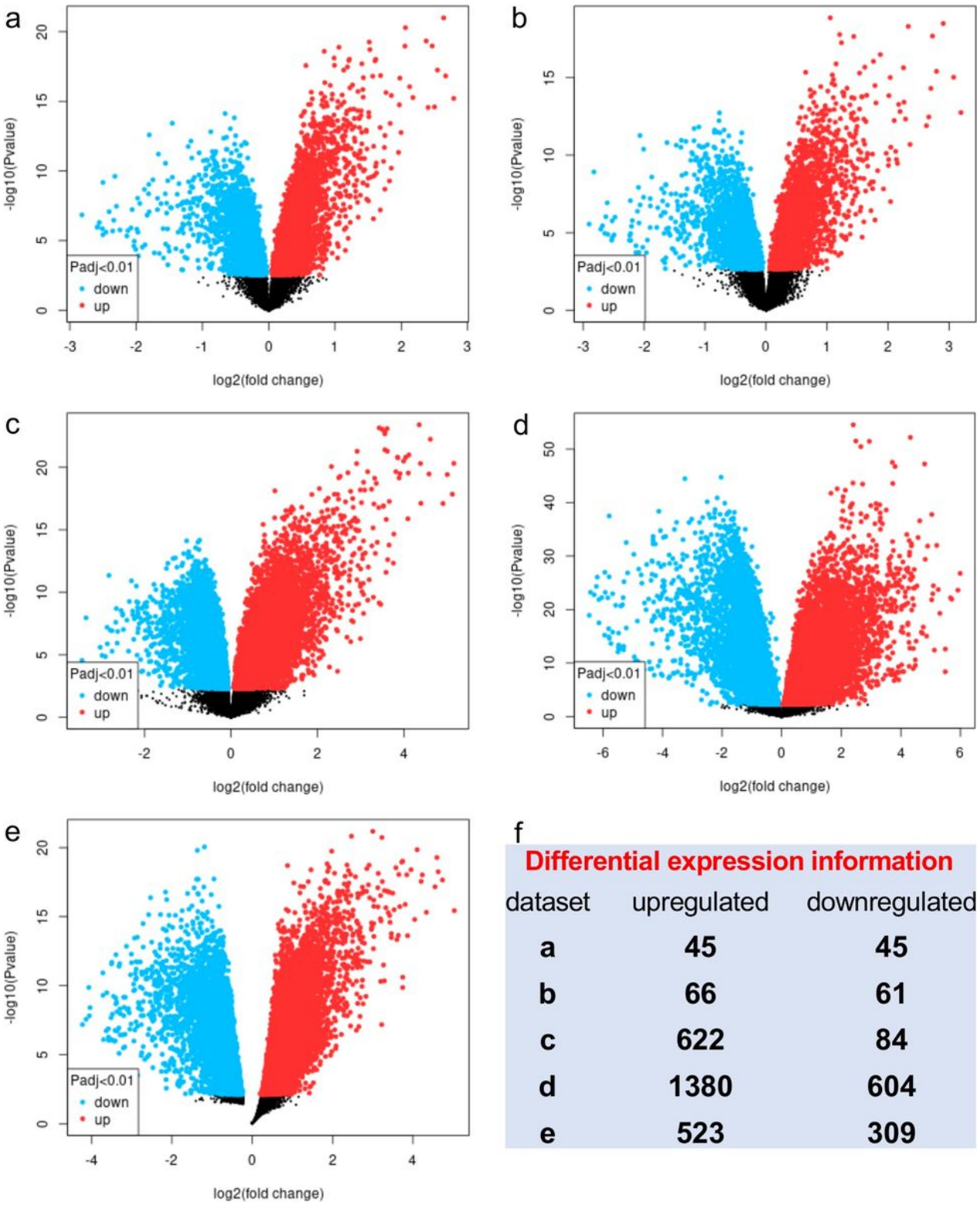
Fig 6: Expression correlation analysis of hub genes in PDAC tissues. The GEPIA2 platform analyzed the expression correlation levels. The Spearman correlation coefficient was used, and an R score of >0.8 was considered statistically significant. The light blue box represents the correlation coefficient based on R scores.

Fig 7: Pathological stages of hub genes in PDAC tissues. Association of mRNA expression and pathological tumor stages in patients with PDAC. Violin plots were created using the GEPIA2 platform based on the TCGA PDAC dataset. F-value indicates the statistical value of the F test; Pr (>F) indicates *p*-value. A *p*-value of *<0.05* was considered statistically significant.

Fig 8: Kaplan-Meier survival plots of hub genes in PDAC tissues. The Kaplan-Meier plots were generated by using the GEPIA2 platform. The overall survival and disease-free survival plots compared a high-risk group (in red) and a low-risk group (in blue) in PDAC tissues. *p<0.05* were regarded as statistically significant.

Fig 9: Immunohistochemical expression of hub genes in human pancreatic cancer specimens. The immunohistochemical data were obtained from the Human Protein Atlas. Staining demonstrated that the protein expression of hub genes was higher in pancreatic cancer tissue than in normal pancreatic tissue samples. The light blue box represents antibodies information. Image courtesy: Human Protein Atlas (http://www.proteinatlas.org).

Figures



Differential expression information

dataset	upregulated	downregulated
a	45	45
b	66	61
c	622	84
d	1380	604
e	523	309

Figure 1

Differential expression of genes between PDAC tissue and adjacent non-cancerous tissue samples in the datasets. (a) GSE62452; (b) GSE28735; (c) GSE15471; (d) GSE62165; (e) GSE102238. The x-axis indicates the fold-change (log-scaled); the y-axis indicates the p-values (log-scaled). The red data-points

represent upregulated genes, while blue data-points represent downregulated genes. The black data-points represent genes with no significant difference in expression. (f) The differential genes screened based on |Log2FC|≥ 1.5/-1.5 and a corrected p-value of <0.01. FC, fold change

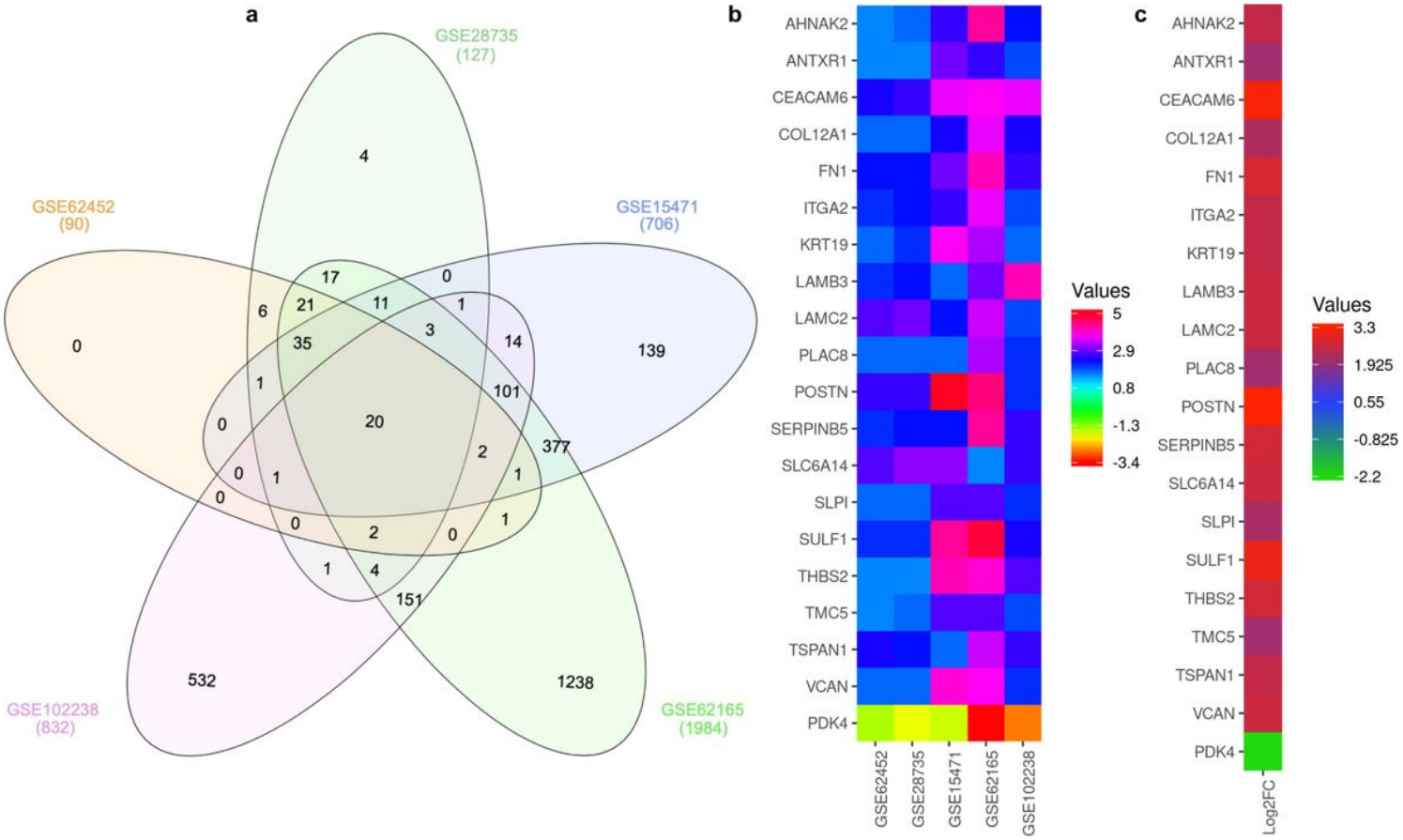


Figure 2

Identification of DEGs. (a) Venn diagram of the DEGs in the five datasets. A total of 20 DEGs were included in all 5 datasets. (b) LogFC heatmap image of the 20 commonly changed DEGs of the five datasets. The 5 GEO datasets are denoted on the abscissa, and the gene names are displayed on the ordinate. (c) The average LogFC expression values of 20 DEGs. DEG, differentially expressed gene; FC, fold change

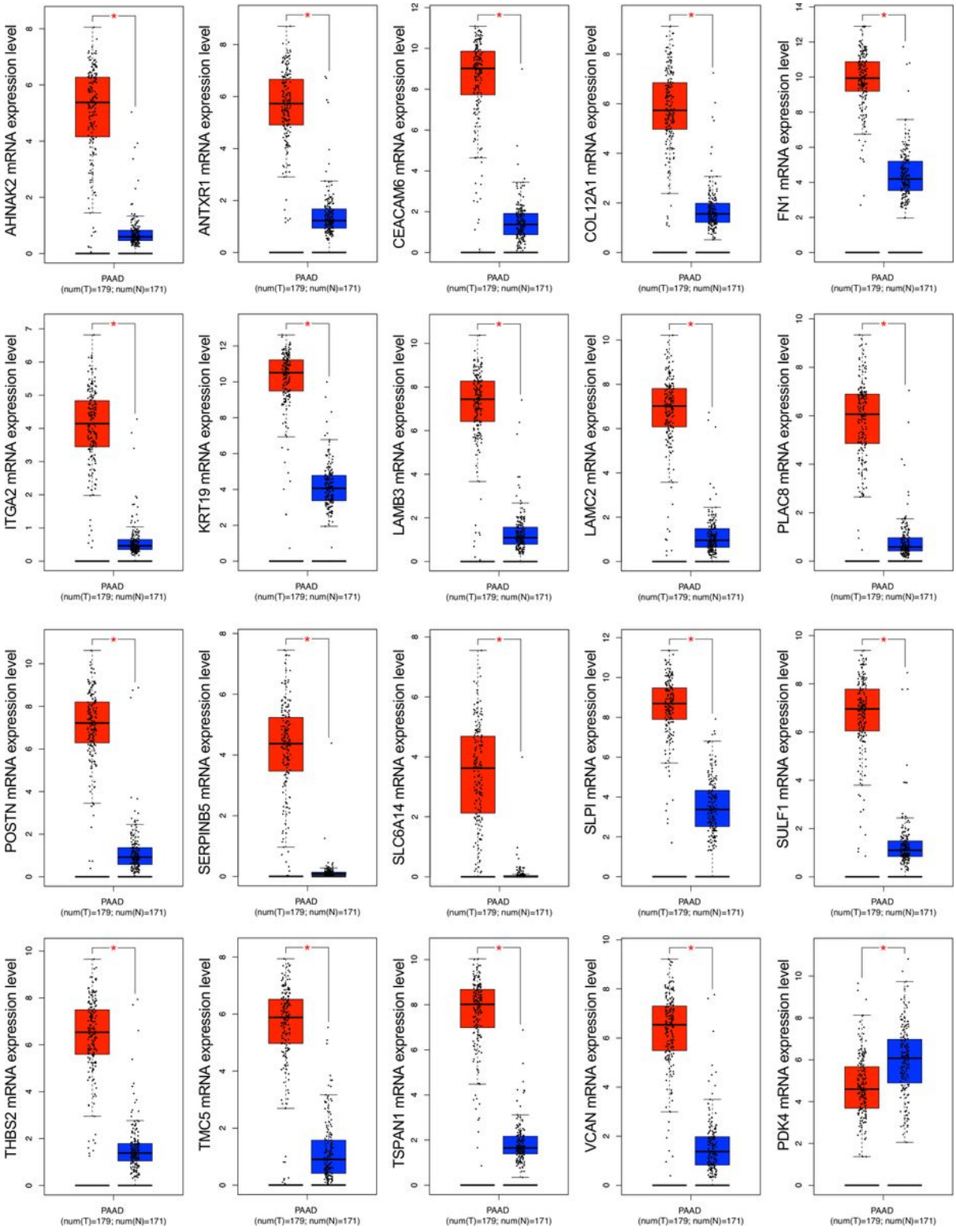


Figure 3

The mRNA expression level analysis of 20 DEGs in PDAC tissues. The boxplots were downloaded from the GEPIA2. The red boxes represent the expression levels in PDAC tissues. In contrast, the blue boxes represent the expression levels in normal tissues. $p<0.05$ was regarded as statistically significant. DEGs, differentially expressed genes; PDAC, pancreatic ductal adenocarcinoma

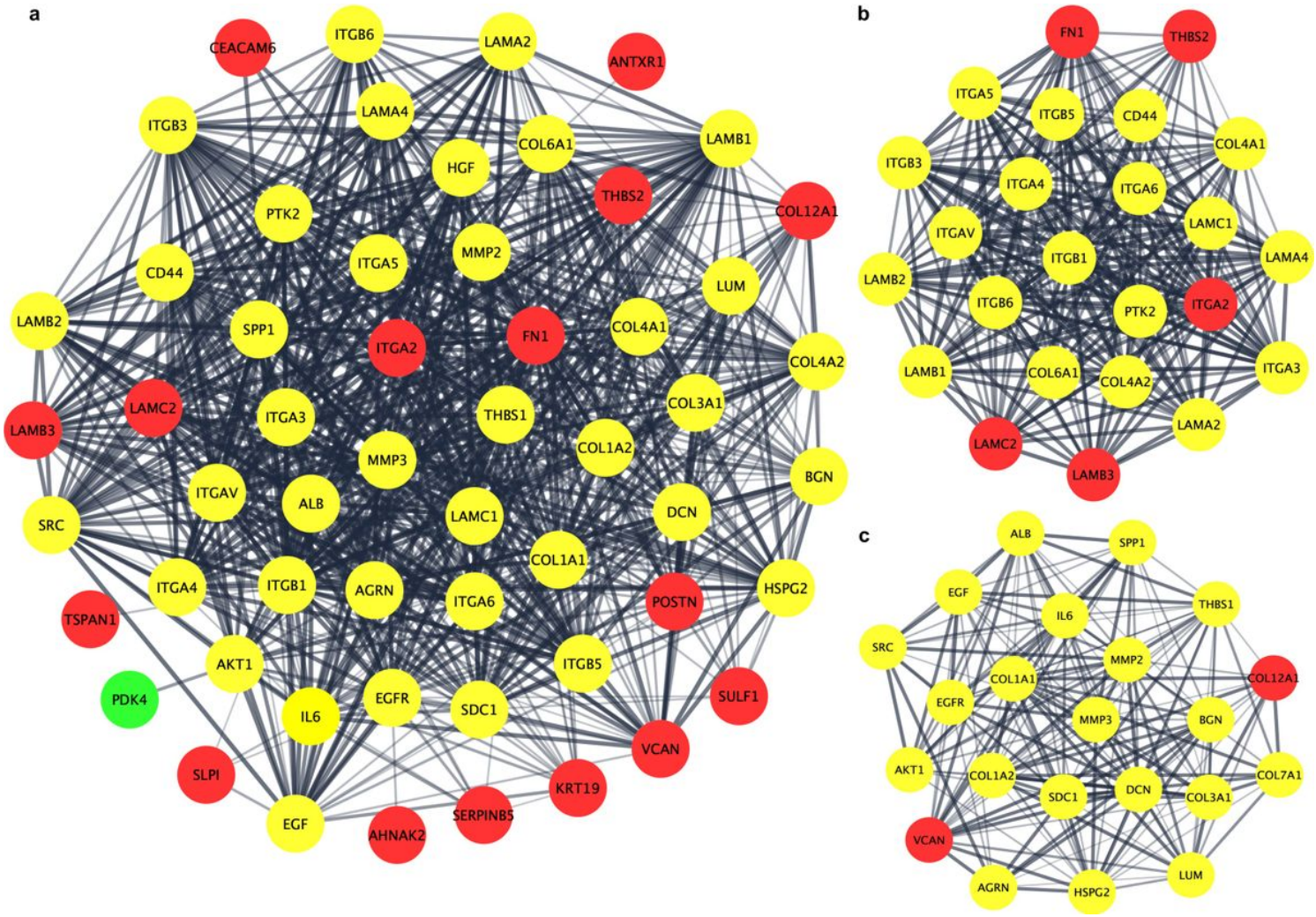


Figure 4

PPI network construction of DEGs and identification of hub genes. (a) PPI network was constructed using Cytoscape. Red nodes represent upregulated genes, whereas green nodes represent downregulated genes. The line represents the interaction relationship between nodes. (b) Significant modules of cluster-1 were identified from the PPI network via the MCODE plug-in. This module consisted of 5 upregulated genes, which are represented by red color. (c) Significant modules of cluster-2 were identified from the PPI network via the MCODE plug-in. This module consisted of 2 upregulated genes, and red nodes represent key genes. PPI, protein-protein interaction; DEGs, differentially expressed genes

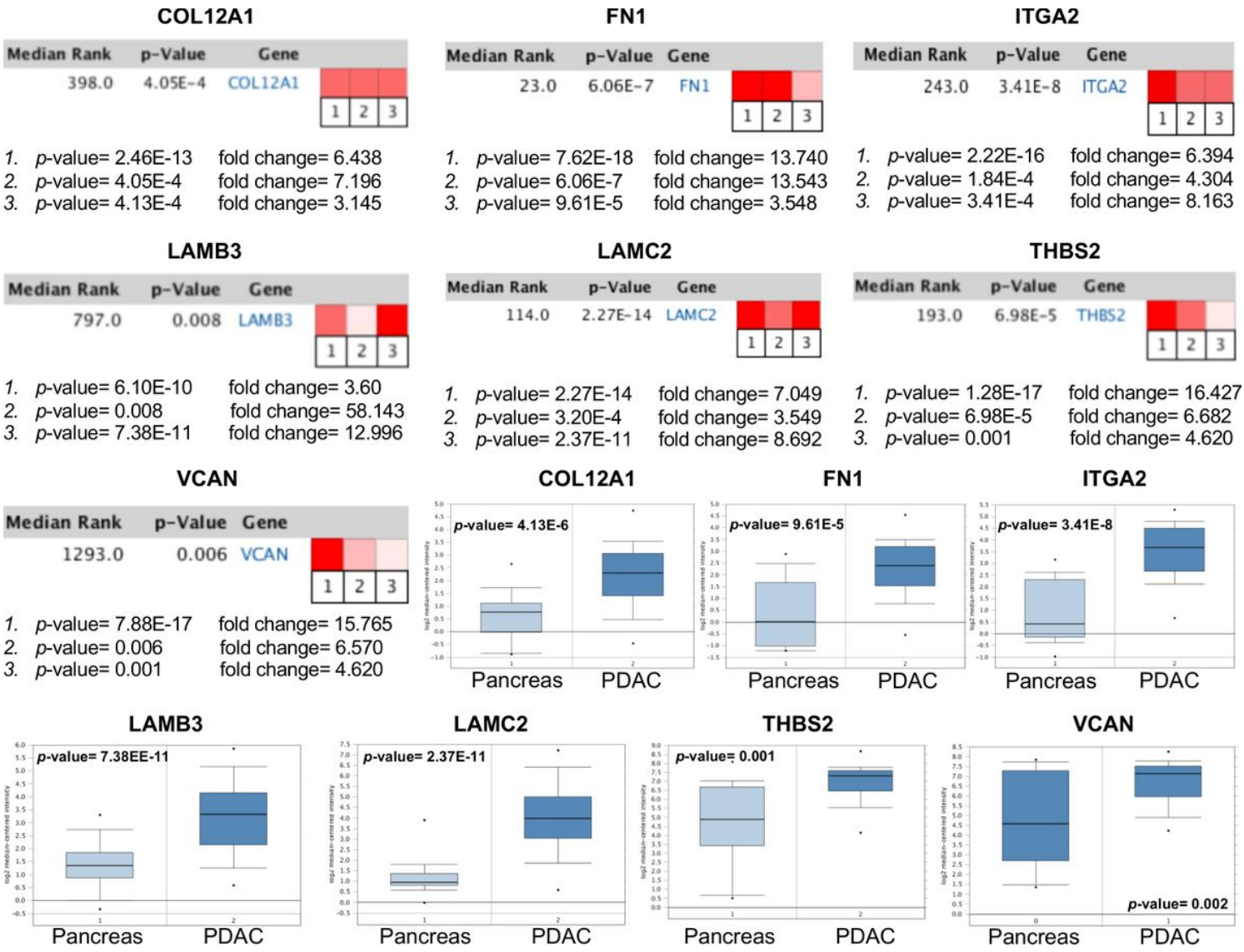
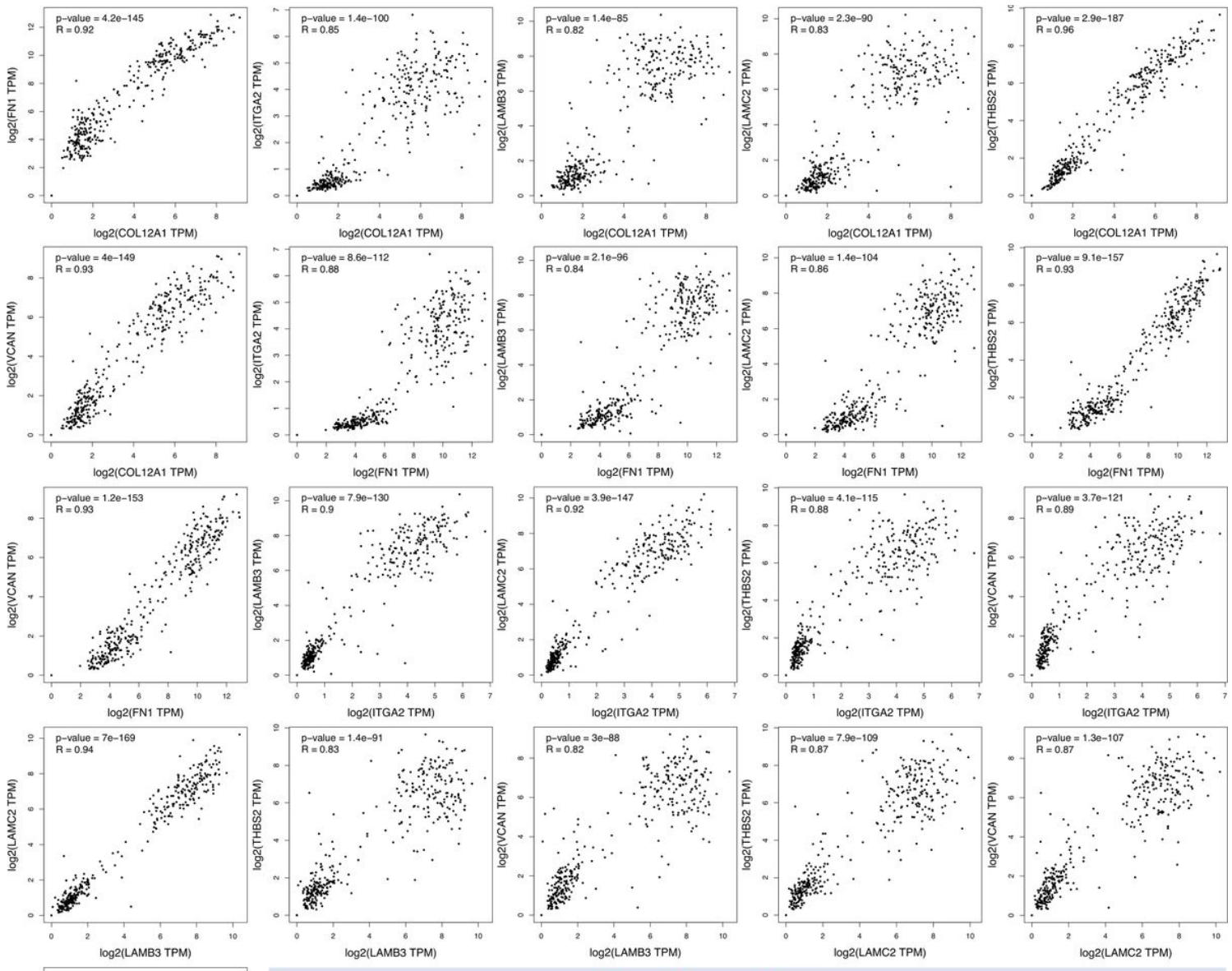


Figure 5

Oncomine analysis of key candidate genes in pancreatic cancer vs. normal tissue. Heat maps of key candidate gene expression in clinical pancreatic cancer samples vs. normal pancreatic tissue samples. [1. Pancreatic ductal adenocarcinoma epithelia vs. normal Badea pancreas; 2. Pancreatic ductal adenocarcinoma epithelia vs. normal Lacobuzio-Donahue pancreas; 3. Pancreatic carcinoma vs. normal Pei pancreas]. The color depth represents the intensity of fold changes. Box plots represent the mRNA expression level in pancreatic cancer and normal pancreatic tissues in the Pei pancreas dataset. $p<0.01$ was considered statistically significant.



Correlation coefficient scores

COL12A1	FN1	ITGA2	LAMB3	LAMC2	THBS2
FN1= 0.92	ITGA2= 0.88	LAMB3= 0.90	LAMC2= 0.94	THBS2= 0.87	VCAN = 0.96
ITGA2= 0.85	LAMB3= 0.84	LAMC2= 0.92	THBS2= 0.83	VCAN = 0.87	
LAMB3= 0.82	LAMC2= 0.88	THBS2= 0.88	VCAN = 0.82		
LAMC2= 0.83	THBS2= 0.93	VCAN = 0.89			
THBS2= 0.96	VCAN = 0.93				
VCAN = 0.93					

Figure 6

Expression correlation analysis of hub genes in PDAC tissues. The GEPIA2 platform analyzed the expression correlation levels. The Spearman correlation coefficient was used, and an R score of >0.8 was considered statistically significant. The light blue box represents the correlation coefficient based on R scores.

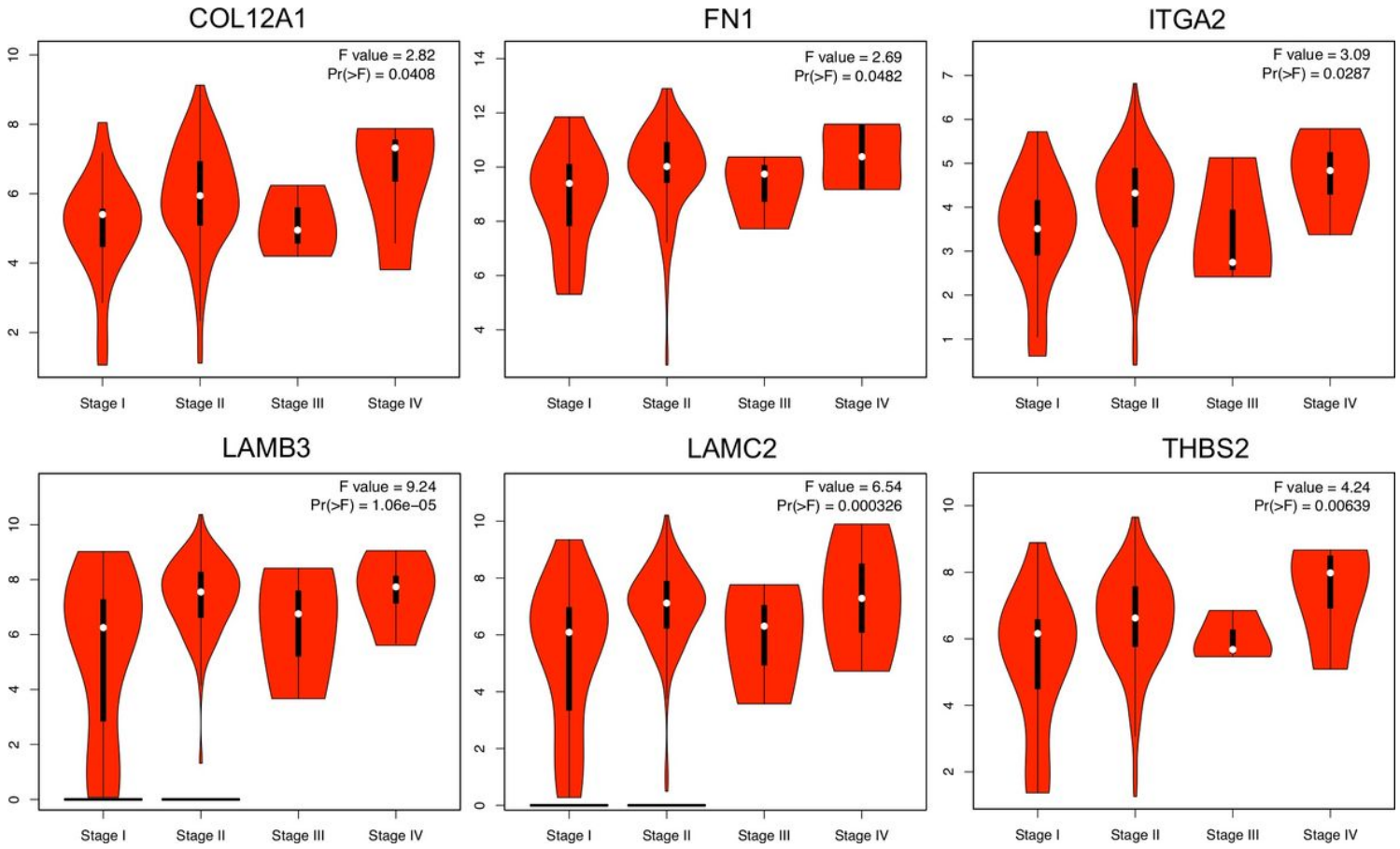


Figure 7

Pathological stages of hub genes in PDAC tissues. Association of mRNA expression 1 and pathological tumor stages in patients with PDAC. Violin plots were created using the GEPIA2 platform based on the TCGA PDAC dataset. F-value indicates the statistical value of the F test; Pr (>F) indicates p-value. A p-value of <0.05 was considered statistically significant.

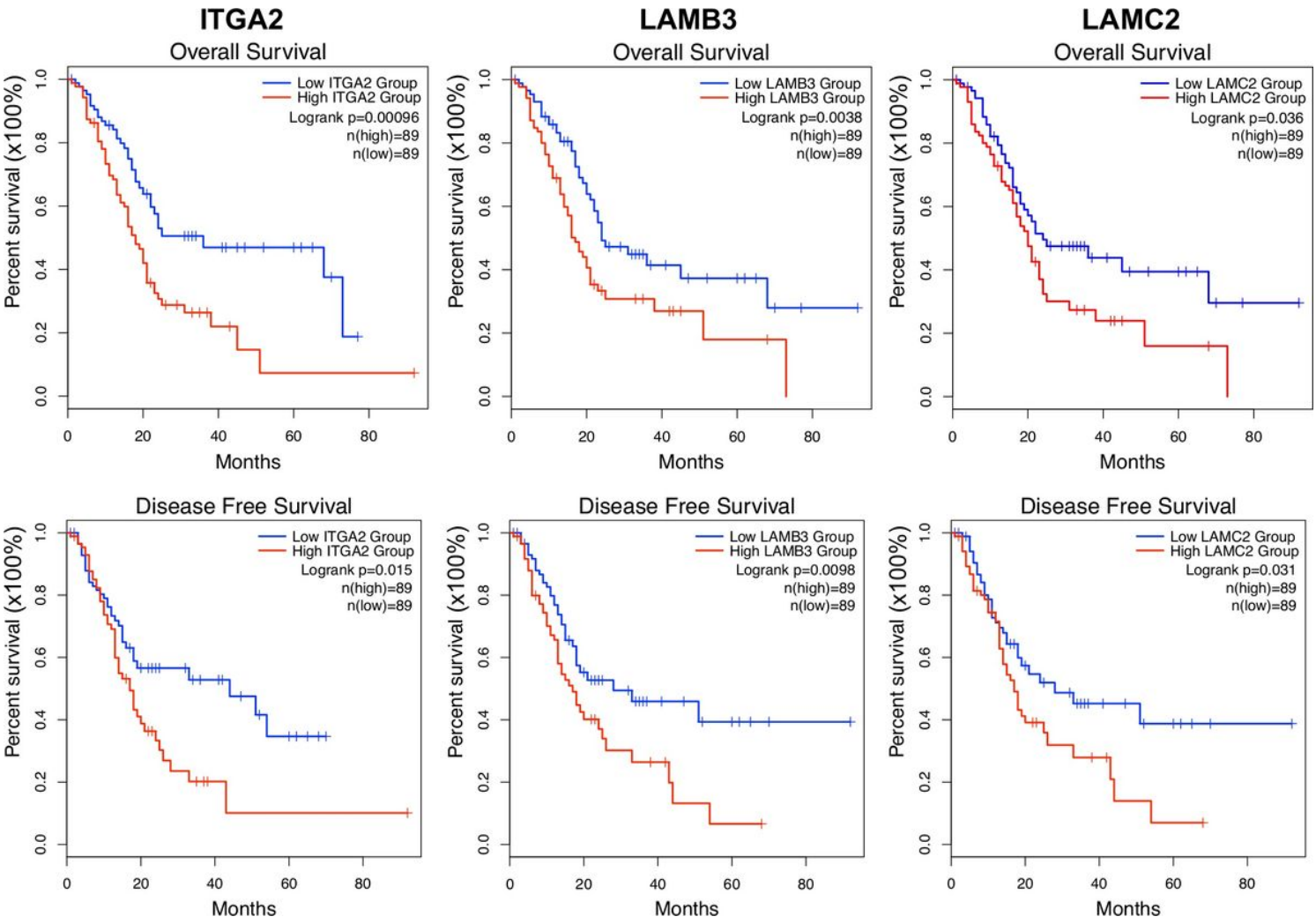


Figure 8

Kaplan-Meier survival plots of hub genes in PDAC tissues. The Kaplan-Meier plots were generated by using the GEPIA2 platform. The overall survival and disease-free survival plots compared a high-risk group (in red) and a low-risk group (in blue) in PDAC tissues. $p<0.05$ were regarded as statistically significant.

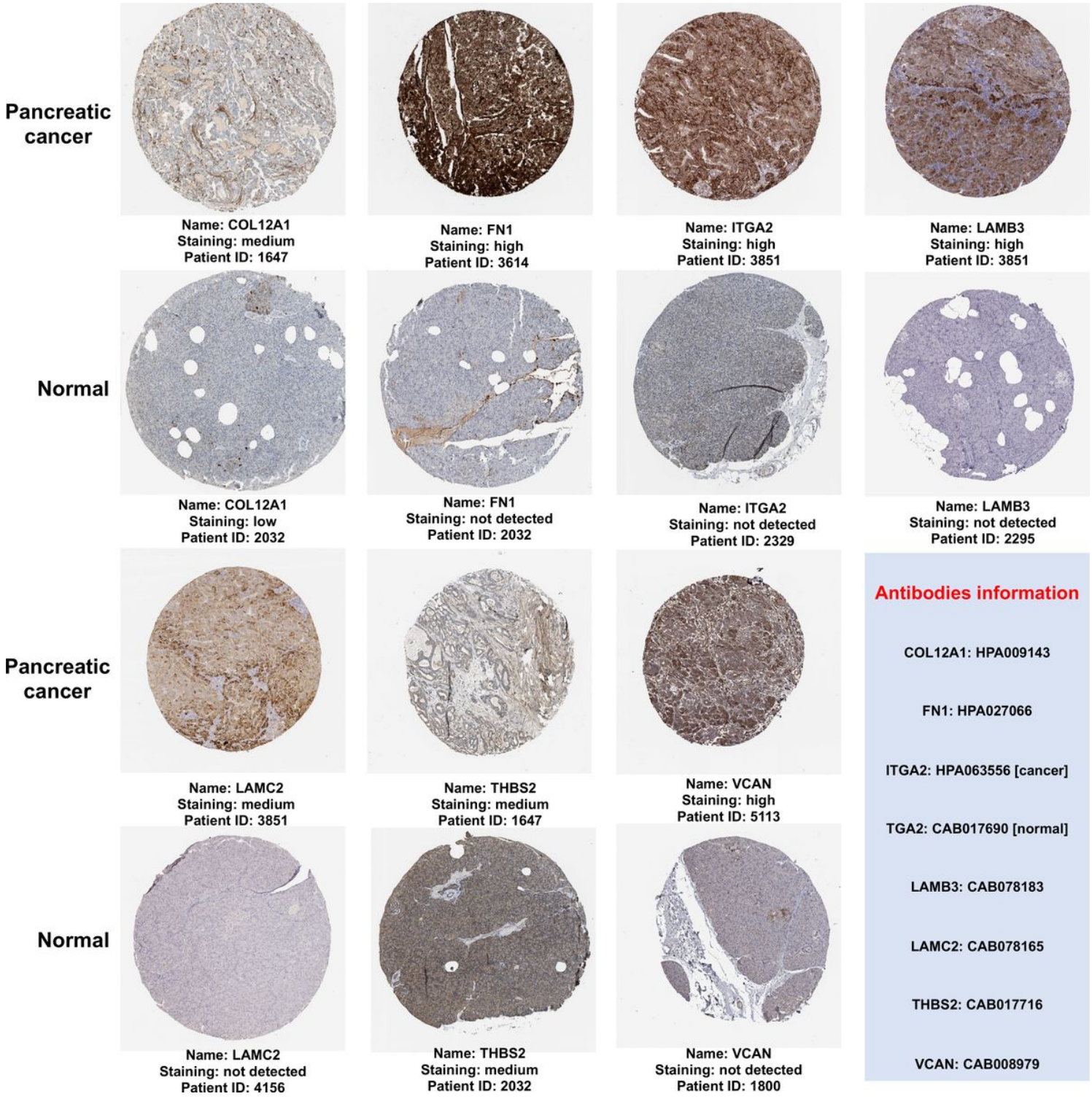


Figure 9

Immunohistochemical expression of hub genes in human pancreatic cancer specimens. The immunohistochemical data were obtained from the Human Protein Atlas. Staining demonstrated that the protein expression of hub genes was higher in pancreatic cancer tissue than in normal pancreatic tissue samples. The light blue box represents antibodies information. Image courtesy: Human Protein Atlas (http://www.proteinatlas.org).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplimentaryfile.pdf