Integrated Bioinformatics Analysis of Gene Expression Profiles for Potential Biomarker Identification Towards Early Therapeutic Intervention in Pancreatitis and Pancreatic Ductal Adenocarcinoma

Manoj M Wagle
Manipal Academy of Higher Education School of Life Sciences
Ananya Rao Kedige
Manipal Academy of Higher Education School of Life Sciences
Shama P Kabekkodu
Manipal Academy of Higher Education School of Life Sciences
Sandeep Mallya (sandeep.mallya@manipal.edu)
Manipal Academy of Higher Education School of Life Sciences https://orcid.org/0000-0001-6318-7165

Research Article

Keywords: Pancreatic Ductal Adenocarcinoma, Pancreatitis, Inflammation, Pancreatic cancer, Microarray analysis, Differentially Expressed Genes

Posted Date: September 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-879048/v1

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is a malignancy associated with rapid progression and an abysmal prognosis. It has been reported that chronic pancreatitis can increase the risk of developing PDAC by 16-fold. Our study aims to identify the key genes and biochemical pathways mediating pancreatitis and PDAC. The gene expression datasets were retrieved from the EMBL-EBI ArrayExpress and NCBI GEO database. A total of 172 samples of normal pancreatic tissue, 68 samples of pancreatitis, and 306 samples of PDAC were used in this study. The differentially expressed genes (DEGs) identified were used to perform downstream analysis for ontology, interaction, and associated pathways.

Furthermore, hub gene expression was validated using the GEPIA2 tool and survival analysis using the Kaplan-Meier (KM) plotter. The potential druggability of the hub genes identified was determined using the Drug-Gene Interaction Database (DGIdb). Our study identified a total of 45 genes found to have altered expression levels in both PDAC and pancreatitis. Over-representation analysis revealed that protein digestion and absorption pathway, ECM-receptor interaction pathway, PI3k-Akt signaling pathway, and proteoglycans in cancer pathways as significantly enriched. Module analysis revealed 15 hub genes with 92 edges, of which 14 were found to be in the druggable genome category. Through bioinformatics analysis, we identified key genes and biochemical pathways disrupted in pancreatitis and PDAC. The results can provide new insights into targeted therapy and intervening therapeutically at an earlier stage can be used as an effective strategy to decrease the incidence and severity of PDAC.

1. Introduction

Pancreatic ductal adenocarcinoma (PDAC) is one of the most prevalent forms of pancreatic cancer worldwide, with an abysmal prognosis accounting for approximately 90% of neoplastic diseases of the pancreas [1]. Patients with the malignancy rarely present the symptoms resulting in a poor diagnosis and an increase in death rate. Despite advancements in the treatment options, the five-year overall survival rate is roughly around 8% making it the 4th common cause of cancer-related mortalities [2]. Although lifestyle factors such as age, alcohol consumption, tobacco use, and obesity play a vital role in the disease, family history, and genetic susceptibility also account for ~10% of pancreatic cancers [3, 4]. The incidence of PDAC in both males and females is higher in developed countries than in developing countries. Some studies have estimated that PDAC would rise to become the second most common cause of cancer-related mortality by 2030 [5, 6]. With a poorly understood etiology, potential treatment options for the management of PDAC include surgical resection (such as pancreaticoduodenectomy or Whipple procedure, total pancreatectomy), adjuvant chemotherapy, and radiation therapy [7, 8].

The aggressive biology and the complicated tumor microenvironment often promote metastasis microscopically, making it challenging to treat. In addition, gene instability, pre-existing cancer stem cells, and alterations in multiple signaling pathways result in an intrinsic chemoresistance that hinders the therapeutic drug delivery [9–11]. Dysregulation of molecular pathways such as K-Ras occur in 75-90% of pancreatic carcinomas, further stimulating downstream signaling cascades [12, 13]. Moreover, mutations

in the transcription factor P53 (TP53) gene can be seen in more than ~60% of pancreatic cancers [14]. Recent studies have shown how inflammation and an elevation in inflammatory cytokines often play a role in developing various cancers. PDAC is associated with significant peri and intra-tumoral inflammation and epithelial-mesenchymal transition (EMT) induction that serves as key mediators contributing to tumor initiation and its rapid progression. This is especially true in cases of chronic pancreatitis (an inflammatory pathophysiological disease of the pancreas) with the predisposing genes associated with a higher risk of developing pancreatic cancer. Hence, identifying the underlying mechanisms, cellular processes, and inflammatory pathways is crucial, which can further help us design drugs targeting tumor initiation and progression, thereby preventing cancer development [15, 16].

In the present study microarray data of PDAC and pancreatitis from publicly available datasets were analyzed to identify the impact of differentially expressed genes (DEGs) by performing downstream analysis applying various bioinformatics methods. The results of this study provide new biological insights which could be explored to identify novel therapeutic targets and biomarkers.

2. Material And Methods

2.1 Literature mining

NCBI Gene Expression Omnibus (GEO) and EMBL-EBI ArrayExpress were used to search for retrieve the microarray expression data for normal, PDAC, and pancreatitis. A total of 6 studies with the following IDs: GSE15471, GSE32676, GSE46234, E-MTAB-1791, E-GEOD-71989 and E-MEXP-2780 were retrieved and used in the study [17, 18]. Since data was generated using different platforms, all the datasets belonging to the respective platforms of Affymetrix GPL570 ([HG-U133_Plus_2]) and Illumina human WG6 BeadChip v3 were processed and analyzed independently. The results obtained were later pooled for a more comprehensive analysis. The detailed description of the methodology followed in the study is represented in (**Fig. 1** Flowchart describing the steps to obtain the hub genes and its analysis. It involved collection of data and preprocessing, screening and identification of overlapping DEGs, ontology and pathway enrichment analysis, protein-protein interaction, and analysis of hub genes).

2.2 Data pre-processing and DEG screening

The datasets were pre-processed, normalized, and analyzed for differential expression using BRB-Array tool 4.6.1 (Stable Version) [19]. The DEGs screened conformed to the following cutoff criteria: |logFC| > 2 and a high significance threshold of 0.001 of univariate tests. Overlapping DEGs between pancreatitis and the PDAC group of samples was identified using the Funrich software [20].

2.3 Ontology and Pathway enrichment analysis

Gene ontology (GO) terms describe non-overlapping information on biological process (BP), cellular component (CC), and molecular function (MF) of individual gene products [21]. Disease ontology provides standard ontologies and comprehensive information on human diseases [22]. KEGG is a

database resource encompassing the functional meaning of a biological system derived mainly from high-throughput experiments [23]. We used the R Bioconductor package, clusterProfiler, to conduct ontology and enrichment analysis to identify the biological significance of these overlapping DEGs [24–28].

2.4 Protein-Protein Interaction (PPI) and module analysis

PPIs are crucial for several biological functions in the body, and an abnormal interaction can often indicate diseases [29]. In this study, we used the STRING tool and Cytoscape software (version 3.8.2) for the construction and visualization of interaction networks [30, 31]. To identify the densely interconnected regions in the network and the hub genes, the Cytoscape application MCODE (Molecular Complex Detection) was used with the following analysis parameters: node score cutoff 0.2, k-score 2, degree cutoff 2, node density cutoff 0.1 and max depth 100 [32].

2.5 Analysis of 15 Hub genes

The key genes obtained were then reanalyzed to identify significant pathways, potential druggability using the Drug-Gene Interaction Database (DGIdb), and gene expression profiles using the GPEIA tool. DGIdb is a web-based resource providing information about druggable candidate genes and potential drug-gene interactions [33]. GEPIA2 is used for performing the gene expression analysis on the data from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTex) [34], which helps us to validate and correlate the expression profiles of hub genes between normal and PDAC tissues. The resultant data was used as an input in Cytoscape and the R package, circlize for visualization [35, 36]. We also performed the survival analysis on all the hub genes using the Kaplan-Meier plotter [37].

3. Results

3.1 Identification of Differentially Expressed Genes

A total of 6 datasets were retrieved that included gene expression profiles of 172 samples of normal pancreatic tissue, 68 samples of pancreatitis, and 306 samples of PDAC. The microarray analysis was performed to identify the differentially expressed genes in PDAC and pancreatitis. The results of these studies were then combined to identify 45 differential expressed genes common to both PDAC and pancreatitis.

3.2 Enrichment analysis and KEGG pathway analysis

Ontology analysis and KEGG pathway enrichment analysis for the 45 DEGs were conducted using the R Bioconductor package, clusterProfiler, with the criterion set at $p < 0.05$. Gene Ontology analysis showed that (i) The most enriched biological process functions for DEGs were extracellular matrix (ECM) organization, extracellular structure organization, ossification, cell-substrate adhesion, and collagen fibril organization (**Fig. 2a** Dot plot representing GO - biological process where the DEGs were mainly enriched in ECM organization, extracellular structure organization, ossification, cell-substrate adhesion, and

collagen fibril organization). (ii) In the molecular function group, the DEGs were enriched in collagen binding, growth factor binding, EMSC conferring tensile strength, and glycosaminoglycan binding (**Fig. 2b** Dot plot representing GO - molecular function where the DEGs were mainly enriched in collagen binding, growth factor binding, extracellular matrix structural constituent (EMSC), EMSC conferring tensile strength, and glycosaminoglycan binding). (iii) In the cellular component group, the DEGs were significantly connected with the collagen-containing ECM, collagen trimer, and its complex and endoplasmic reticulum lumen. (**Fig. 2c** Dot plot representing GO - cellular process where the DEGs were significantly associated with the collagen-containing ECM, collagen trimer and its complex, and endoplasmic reticulum lumen).

As shown in (**Fig. 3a** Bar plot representing disease ontology analysis for the 45 DEGs common to both pancreatitis and PDAC. The DEGs were significantly associated with lung disease, cell type benign neoplasm, and pancreatic cancer). Disease Ontology analysis indicated that the DEGs were significantly associated with lung disease, cell type benign neoplasm, and pancreatic cancer. As for the KEGG pathway analysis, protein digestion and absorption pathway, PI3k-Akt signaling pathway, ECM-receptor interaction pathway, proteoglycans in cancer pathways were significantly enriched (**Fig. 3b** Bar plot representing KEGG Pathway Analysis for the 45 DEGs for the 45 DEGs common to both pancreatitis and PDAC. The DEGs were mainly enriched in protein digestion and absorption pathway, PI3k-Akt signaling pathway, ECM-receptor interaction pathway, proteoglycans in cancer pathways).

3.3 PPI network and module analysis

The Protein-protein interaction network (PPI) for the 45 DEGs was constructed and visualized using Cytoscape and STRING, as shown in (**Fig. 4a** Represents protein-protein interaction (PPI) network constructed using the STRING database for the 45 DEGs common to both pancreatitis and PDAC). Module analysis using MCODE revealed 15 hub genes (nodes) with 92 edges (**Fig. 4b** Represents the module analysis of the PPI network constructed using the Cytoscape app, MCODE showing 15 nodes (hub genes) with 92 edges).

3.4 Analysis of hub genes

The 15 hub genes obtained were then reanalyzed for KEGG pathways, and the following five core genes were identified - COL1A1, THBS1, COL1A2, THBS2 and COL3A1 (**Fig. 5** Chord diagram representing the KEGG pathway reanalysis for the 15 hub genes. Total of five core genes were identified - COL1A1, THBS1, COL1A2, THBS2 and COL3A1. Among these, the genes COL1A1 and COL1A2 alone were significantly associated with 11 different pathways each). Among these, COL1A1 and COL1A2 were significantly associated with 11 different pathways each. Expression analysis was conducted using the GEPIA2 tool to compare expression levels in PDAC and Normal tissues. Results showed that all 15 hub genes were found to be significantly expressed (P-value < 0.001 and Log2FC > 2) (**Fig. 6** Box plots comparing the gene expression levels of PDAC and normal tissues for 15 hub genes P-value < 0.001 and Log2FC > 2). The samples from normal tissues are shown in grey, and PDAC tissues in red. Gene expression changes between groups in all key genes were found to be significant). The druggability of the hub genes was

determined using the DGIdb database. It was found that 14 out of 15 genes could be modulated and interact with small molecules and were found to be in the druggable genome category. The complete list of genes and their respective druggable gene category is shown in (**Table 1** Of the 15 hub genes analyzed for druggability using the DGIdb database, a total of 14 were found to be in the druggable genome category indicating their use as potential drug targets). The prognostic value associated with 15 key genes was analyzed using the KM plotter with $p < 0.05$ (**Fig. 7** Kaplan-Meier plots representing the survival analysis for 15 hub genes with respect to low expression (black color) and high expression (red) in PDAC tissue samples. Among these, survival analysis for COL6A1, COL6A3, COL8A1, LUM and THBS2 genes (marked with *) are statistically significant $p < 0.05$). The results showed that the gene COL6A1 was most significantly associated (log-rank $P = 0.0061$) with the overall survival of patients with PDAC.

We also compared the expression of these 15 genes based on age, gender, stage, race, and pancreatitis status using the online UALCAN tool [38]. The analysis did not reveal any notable gender specific difference in the expression of genes between males and females. Similarly, the expression values did not vary much with or without the presence of chronic pancreatitis at the first diagnosis. Although not statistically significant, samples from ‘occasional drinkers’ did show higher transcripts per million values compared to normal samples. Analysis of the expression data between different races showed that the gene Fibulin-1 (FBLN1) had an increased level of expression in African Americans when compared to Asians, Caucasians, and normal samples (Supplementary materials).

4. Discussion

Although there has been a massive improvement in the treatment for PDAC, mortality and incidence rates are still increasing at an unprecedented rate. Several studies have been carried out to uncover the molecular mechanisms involved in the onset, growth, and progression of PDAC. It has also been reported that chronic cases of pancreatitis can increase the risk of developing PDAC by 16-fold [39, 40]. Our study aims to identify essential genes, pathways, and interactions involved in both pancreatitis and PDAC. Analysis of the datasets revealed that the 45 DEGs were mainly enriched in collagen and growth factor binding, extracellular environment, and cell adhesion. Collagen is the crucial component of the extracellular matrix (ECM), and specific orientation and arrangements of ECM in a microscopic environment are thought to play crucial roles in tumor progression [41, 42]. This disruption in the ECM homeostasis can be caused by degradation and even deposition of collagen. Since tumor cells continuously interact with ECM, an increased disruption can accelerate tumor progression by negatively interfering with cell adhesion [43–46].

KEGG pathway analysis showed that protein digestion and absorption pathway, ECM-receptor interaction pathway, PI3k-Akt signaling pathway, and proteoglycans in cancer pathways might play essential roles in the progression of PDAC. Aberration of the PI3k-Akt signaling pathway can be seen in many different cancers. An increase in Akt activity is regularly seen in PDAC (~60% of cases) due to the loss of key regulators or mutations. K-Ras is an essential gene of the RAS/MAPK pathway (required for proliferation and maturation of cells), and activating mutations in this gene can be seen in ~95% of pancreatic

cancers, which further activates PI3K signaling [47–49]. These are the major reasons why targeting the PI3k-Akt pathway has been a significant interest in cancer drug discovery. Another important biomolecule of interest having multiple functions in angiogenesis and cancer are proteoglycans. Cell growth is influenced by proteoglycans through their interaction with growth factors and can sometimes cause deregulation of cell proliferation [50, 51]. Thus, their integration in tumor cell diagnostics can facilitate early diagnosis, as demonstrated in a few studies [52, 53].

We further constructed a protein-protein interaction network and performed a module analysis. The module consisted of 15 nodes with 92 edges. Based on the expression analysis using the GEPIA2 tool, all 15 hub genes were significantly expressed in PDAC. UALCAN analysis revealed that the hub genes were independent of the factors such as age, gender, stage, race, and pancreatitis status, which suggests that these genes can be used as biomarkers at a global scale for advancing PDAC treatment. Potential druggability determined using the DGIdb database showed that 14 out of 15 genes were in the druggable genome category and thus have a potential value for developing targeted drugs. Next, to understand which genes were significantly involved in the pathways analyzed before, we performed KEGG pathway reanalysis for the 15 hub genes. Five core genes were identified based on this - COL1A1, COL1A2, THBS1, THBS2, and COL3A1 were identified. Recently, few studies have reported the role of various collagen genes in tumorigenesis, which can lead to poor clinical outcomes [54, 55]. Further, differential expression of genes COL1A1, COL1A2, and THBS1 has been reported in several cancers, including colorectal cancer, hepatocellular carcinoma, and melanoma [56–58]. Some researchers have demonstrated the utilization of Thrombospondin-2 (THBS2) as a biomarker for risk prediction and early detection of PDAC, and as a strong prognostic indicator in colorectal cancer [59, 60]. All the above results suggest that these genes can provide valuable insights towards targeted therapy and personalized treatment in PDAC.

Our study presents useful findings linking inflammation to cancer, and the clinical significance of this study can be further verified by in-depth experimental research. However, there were a few limitations to this study. Firstly, this study compared pancreatitis and PDAC samples but did not consider the stage of individual samples. Secondly, the clinical data for the samples were not deeply analyzed due to inaccessibility

5. Conclusion

Here, we present a comprehensive analysis of the gene expression profiles by integrated bioinformatics approach to identify key genes, biochemical processes, and various pathways disrupted in both pancreatitis and PDAC. We majorly identified 15 hub genes that might play crucial roles in the above conditions. Literature studies indicate how inflammation can serve as a risk factor for developing cancer. Furthermore, our study suggests that treating pancreatitis can be used as an effective strategy to decrease the incidence and severity of PDAC. Hence, intervening therapeutically at an earlier stage by targeting these pathways and genes can significantly improve the patient's treatment response and survival rate.

Declarations

Funding:

No funding was received for conducting this study.

Conflicts of interest:

The authors declare that there are no conflicts of interest.

Availability of data and material

The authors declare that the data analyzed during the study are available in the NCBI Gene Expression Omnibus [GSE15471, GSE32676, GSE46234] and EMBL-EBI ArrayExpress [E-MTAB-1791, E-GEOD-71989 and E-MEXP-2780].

Code availability

Not applicable

Authors' contributions

Manoj M Wagle and Sandeep Mallya contributed to the study conception and design. Material preparation, data collection and analysis were performed by Manoj M Wagle and Ananya Rao Kedige. The first draft of the manuscript was written by Manoj M Wagle and all authors commented on previous versions of the manuscript. Sandeep Mallya and Shama P Kabekkodu was involved in supervision and validation of data. All authors read and approved the final manuscript.

Ethics approval:

Not applicable

Consent to participate:

Not applicable

Consent for publication:

Not applicable

Acknowledgement

The authors wish to thank the funding agencies DST-FIST, Government of India, TIFAC-CORE in pharmacogenomics and Manipal Academy of Higher Education, Manipal, India for providing the necessary facilities and infrastructure needed for the study. We would also like to thank Dr. K. Satyamoorthy (Director, Manipal School of Life Sciences, MAHE, Manipal) and Dr. Bobby Paul (Assistant

Professor and Head, Department of Bioinformatics, Manipal School of Life Sciences, MAHE, Manipal) for their support and suggestions towards valuable additions to the manuscript.

References

1. Kleeff J, Korc M, Apte M, et al (2016) Pancreatic cancer. Nat Rev Dis Primers 2:16022. https://doi.org/10.1038/nrdp.2016.22

2. Siegel RL, Miller KD, Jemal A (2018) Cancer statistics, 2018. CA Cancer J Clin 68:7–30. https://doi.org/10.3322/caac.21442

3. Shi C, Hruban RH, Klein AP (2009) Familial pancreatic cancer. Arch Pathol Lab Med 133:365–374. https://doi.org/10.5858/133.3.365

4. Permuth-Wey J, Egan KM (2009) Family history is a significant risk factor for pancreatic cancer: results from a systematic review and meta-analysis. Fam Cancer 8:109–117. https://doi.org/10.1007/s10689-008-9214-8

5. Rahib L, Smith BD, Aizenberg R, et al (2014) Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res 74:2913–2921. https://doi.org/10.1158/0008-5472.can-14-0155

6. Wong MCS, Jiang JY, Liang M, et al (2017) Global temporal patterns of pancreatic cancer and association with socioeconomic development. Sci Rep 7:3165. https://doi.org/10.1038/s41598-017-02997-2

7. Neoptolemos JP, Cunningham D, Friess H, et al (2003) Adjuvant therapy in pancreatic cancer: historical and current perspectives. Ann Oncol 14:675–692. https://doi.org/10.1093/annonc/mdg207

8. Alexakis N, Halloran C, Raraty M, et (2004) Current standards of surgery for pancreatic cancer. Br J Surg 91:1410–1427. https://doi.org/10.1002/bjs.4794

9. Lee CJ, Li C, Simeone DM (2008) Human Pancreatic Cancer Stem Cells: Implications for How We Treat Pancreatic Cancer. Transl Oncol 1:14. https://doi.org/10.1593/tlo.08013

10. Samuel N, Hudson TJ (2011) The molecular and cellular heterogeneity of pancreatic ductal adenocarcinoma. Nat Rev Gastroenterol Hepatol 9:77–87. https://doi.org/10.1038/nrgastro.2011.215

11. Oberstein PE, Olive KP (2013) Pancreatic cancer: why is it so hard to treat? Therap Adv Gastroenterol 6:321–337. https://doi.org/10.1177/1756283x13478680

12. Almoguera C, Shibata D, Forrester K, et al (1988) Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. Cell 53:549–554. https://doi.org/10.1016/0092-8674(88)90571-5

13. Malumbres M, Barbacid M (2003) RAS oncogenes: the first 30 years. Nat Rev Cancer 3:459–465. https://doi.org/10.1038/nrc1097

14. Rozenblum E, Schutte M, Goggins M, et al (1997) Tumor-suppressive Pathways in Pancreatic Carcinoma. Cancer Res 57.

15. Zheng L, Xue J, Jaffee EM, Habtezion A (2013) Role of immune cells and immune-based therapies in pancreatitis and pancreatic ductal adenocarcinoma. Gastroenterology 144:1230–1240. https://doi.org/10.1053/j.gastro.2012.12.042

16. Khalafalla FG, Khan MW (2017) Inflammation and Epithelial-Mesenchymal Transition in Pancreatic Ductal Adenocarcinoma: Fighting Against Multiple Opponents. Cancer Growth Metastasis 10:117906441770928. https://doi.org/10.1177/1179064417709287

17. Barrett T, Wilhite SE, Ledoux P, et al NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res 41. https://doi.org/10.1093/nar/gks1193

18. Athar A, Füllgrabe A, George N, et al (2019) ArrayExpress update - from bulk to single-cell expression data. Nucleic Acids Res 47:D711–D715. https://doi.org/10.1093/nar/gky964

19. Simon R, Lam A, Li M-C, et al (2007) Analysis of Gene Expression Data Using BRB-Array Tools. Cancer Inform 3:11-17. https://doi.org/10.1177/117693510700300022

20. Pathan M, Keerthikumar S, Ang CS, et al (2015) FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics 15:2597–2601. https://doi.org/10.1002/pmic.201400515

21. Hill DP, Smith B, McAndrews-Hill MS, et al (2008) Gene Ontology annotations: what they mean and where they come from. BMC Bioinformatics 2008 9:5 9:1–9. https://doi.org/10.1186/1471-2105-9-s5-s2

22. Schriml LM, Mitraka E, Munro J, et al (2019) Human Disease Ontology 2018 update: classification, content and workflow expansion. Nucleic Acids Res 47:D955–D962. https://doi.org/10.1093/nar/gky1032

23. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27

24. Yu G, Wang LG, Han Y, et al (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16:284–287. https://doi.org/10.1089/omi.2011.0118

25. Yu G, Wang LG, Yan GR, et al (2015) DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics 31:608–609. https://doi.org/10.1093/bioinformatics/btu684

26. Carlson M (2019) GO.db: A set of annotation maps describing the entire Gene Ontology. R package version 3.8.2.

27. Carlson M (2019) org.Hs.eg.db: Genome wide annotation for Human. R package version 3.8.2.

28. Yu G (2021) enrichplot: Visualization of Functional Enrichment Result. R package version 1.12.2.

29. Gonzalez MW, Kann MG (2012) Chapter 4: Protein Interactions and Disease. PLoS Comput Biol 8:e1002819. https://doi.org/10.1371/journal.pcbi.1002819

30. Shannon P, Markiel A, Ozier O, et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303

31. Szklarczyk D, Franceschini A, Wyder S, et al (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43:D447–D452. https://doi.org/10.1093/nar/gku1003

32. Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003 4:1 4:1–27. https://doi.org/10.1186/1471-2105-4-2

33. Freshour SL, Kiwala S, Cotto KC, et al (2021) Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. Nucleic Acids Res 49:D1144–D1151. https://doi.org/10.1093/nar/gkaa1084

34. Tang Z, Kang B, Li C, et al (2019) GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res 47:W556–W560. https://doi.org/10.1093/nar/gkz430

35. Harrington B (2004). Inkscape. http://www.inkscape.org/.

36. Gu Z, Gu L, Eils R, et al (2014) circlize Implements and enhances circular visualization in R. Bioinformatics 30:2811–2812. https://doi.org/10.1093/bioinformatics/btu393

37. Nagy Á, Munkácsy G, Győrffy B (2021) Pancancer survival analysis of cancer hallmark genes. Scientific Rep 11:6047. https://doi.org/10.1038/s41598-021-84787-5

38. Chandrashekar DS, Bashel B, Balasubramanya S, et al (2017) UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia 19:649–658. https://doi.org/10.1016/j.neo.2017.05.002

39. Carrière C, Young AL, Gunn JR, et al (2009) Acute pancreatitis markedly accelerates pancreatic cancer progression in mice expressing oncogenic Kras. Biochem Biophys Res Commun 382:561.

https://doi.org/10.1016/j.bbrc.2009.03.068

40. Kirkegård J, Mortensen FV, Cronin-Fenton D (2017) Chronic Pancreatitis and Pancreatic Cancer Risk: A Systematic Review and Meta-analysis. Am J Gastroenterol 112:1366–1372. https://doi.org/10.1038/ajg.2017.218

41. Friedl P, Wolf K (2008) Tube travel: the role of proteases in individual and collective cancer cell invasion. Cancer Res 68:7247–7249. https://doi.org/10.1158/0008-5472.can-08-0784

42. Cox TR, Erler JT (2011) Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. Dis Model Mech 4:165–178. https://doi.org/10.1242/dmm.004077

43. Paszek MJ, Zahir N, Johnson KR, et al (2005) Tensional homeostasis and the malignant phenotype. Cancer Cell 8:241–254. https://doi.org/10.1016/j.ccr.2005.08.010

44. Provenzano PP, Eliceiri KW, Campbell JM, et al (2006) Collagen reorganization at the tumor-stromal interface facilitates local invasion. BMC Med 4:1–15. https://doi.org/10.1186/1741-7015-4-38

45. Fang M, Yuan J, Peng C, et al (2014) Collagen as a double-edged sword in tumor progression. Tumour Biol 35:2871–2882. https://doi.org/10.1007/s13277-013-1511-7

46. Xu S, Xu H, Wang W, et al (2019) The role of collagen in cancer: from bench to bedside. J Transl Med 17:1–22. https://doi.org/10.1186/s12967-019-2058-1

47. Kennedy AL, Morton JP, Manoharan I, et al (2011) Activation of the PIK3CA/AKT pathway suppresses senescence induced by an activated RAS oncogene to promote tumorigenesis. Mol Cell 42:36–49. https://doi.org/10.1016/j.molcel.2011.02.020

48. Eser S, Reiff N, Messer M, et al (2013) Selective requirement of PI3K/PDK1 signaling for Kras oncogene-driven pancreatic cell plasticity and cancer. Cancer Cell 23:406–420. https://doi.org/10.1016/j.ccr.2013.01.023

49. Baer R, Cintas C, Dufresne M, et al (2014) Pancreatic cell plasticity and cancer initiation induced by oncogenic Kras is completely dependent on wild-type PI 3-kinase p110α. Genes Dev 28:2621–2635. https://doi.org/10.1101/gad.249409.114

50. Wang J, Svendsen A, Kmiecik J, et al (2011) Targeting the NG2/CSPG4 proteoglycan retards tumour growth and angiogenesis in preclinical models of GBM and melanoma. *PLoS One* 6: e23062. https://doi.org/10.1371/journal.pone.0023062

51. Knelson EH, Nee JC, Blobe GC (2014) Heparan sulfate signaling in cancer. Trends Biochem Sci 39:277–288. https://doi.org/10.1016/j.tibs.2014.03.001

52. Kurihara T, Itoi T, Sofuni A, et al (2008) Detection of circulating tumor cells in patients with pancreatic cancer: a preliminary result. J Hepatobiliary Pancreat Surg 15:189–195. https://doi.org/10.1007/s00534-007-1250-5

53. Effenberger KE, Schroeder C, Hanssen A, et al (2018) Improved Risk Stratification by Circulating Tumor Cell Counts in Pancreatic Cancer. Clin Cancer Res 24:2844–2850. https://doi.org/10.1158/1078-0432.ccr-18-0120

54. Kita Y, Mimori K, Tanaka F, et al (2009) Clinical significance of LAMB3 and COL7A1 mRNA in esophageal squamous cell carcinoma. Eur J Surg Oncol 35:52–58. https://doi.org/10.1016/j.ejso.2008.01.025

55. Wu YH, Chang TH, Huang YF, et al (2014) COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. Oncogene 33:3432–3440. https://doi.org/10.1038/onc.2013.307

56. Bonazzi VF, Nancarrow DJ, Stark MS, et al (2011) Cross-Platform Array Screening Identifies COL1A2, THBS1, TNFRSF10D and UCHL1 as Genes Frequently Silenced by Methylation in Melanoma. PLoS One 6:e26121. https://doi.org/10.1371/journal.pone.0026121

57. Hayashi M, Nomoto S, Hishida M, et al (2014) Identification of the collagen type 1 alpha 1 gene (COL1A1) as a candidate survival-related factor associated with hepatocellular carcinoma. BMC Cancer 2014 14:1 14:1–10. https://doi.org/10.1186/1471-2407-14-108

58. Zhang Z, Fang C, Wang Y, et al (2018) COL1A1: A potential therapeutic target for colorectal cancer expressing wild-type or mutant KRAS. Int J Oncol 53:1869. https://doi.org/10.3892/ijo.2018.4536

59. Kim J, Bamlet WR, Oberg AL, et al (2017) Detection of early pancreatic ductal adenocarcinoma with thrombospondin-2 and CA19-9 blood markers. Sci Transl Med 9:eaah5583. https://doi.org/10.1126/scitranslmed.aah5583

60. Tian Q, Liu Y, Zhang Y, et al (2018) THBS2 is a biomarker for AJCC stages and a strong prognostic indicator in colorectal cancer. J BUON 23:1331–1336.

Tables

Due to technical limitations, table 1 is only available as a download in the Supplemental Files section.

Figures

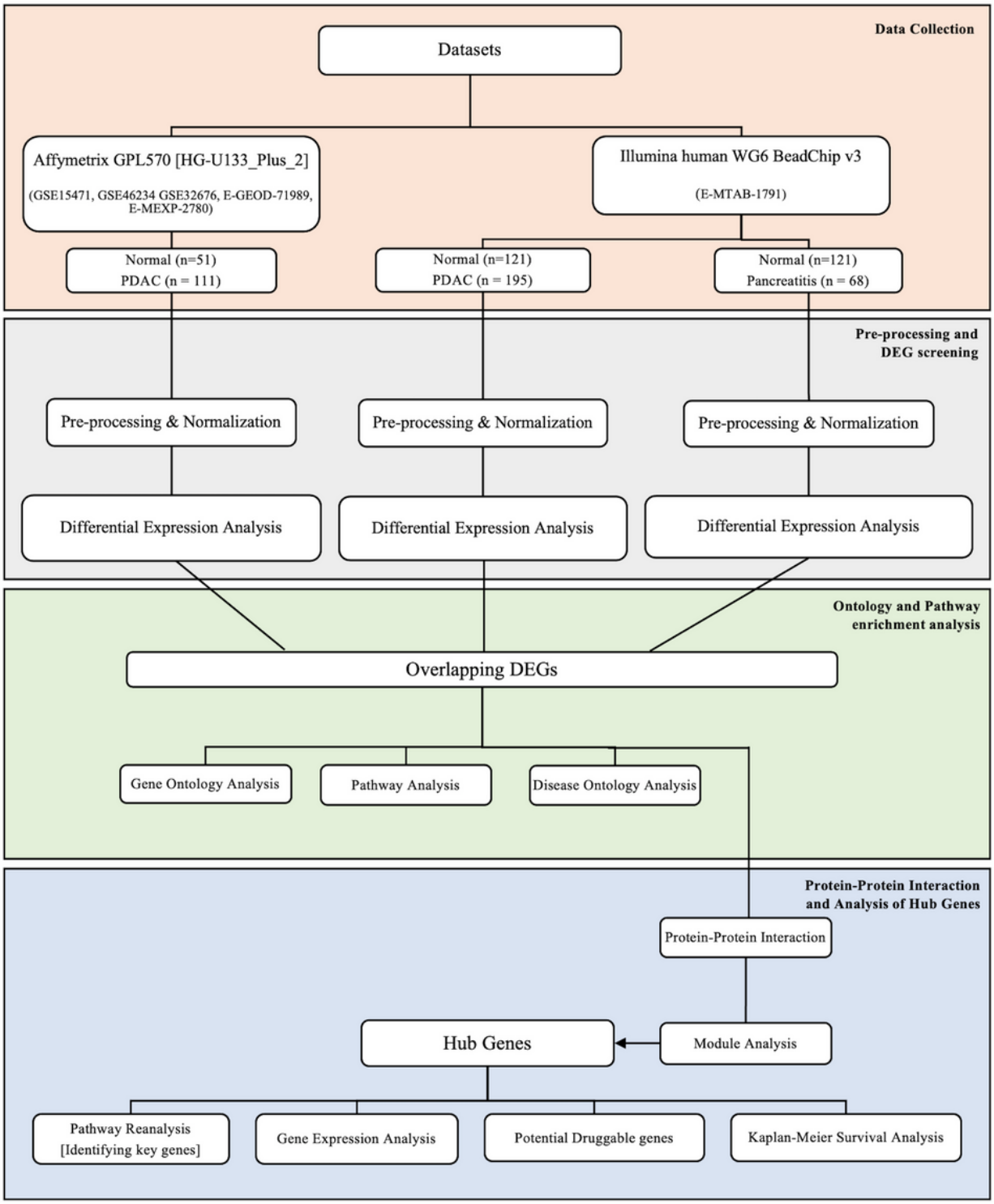

Figure 1

Flowchart describing the steps to obtain the hub genes and its analysis.

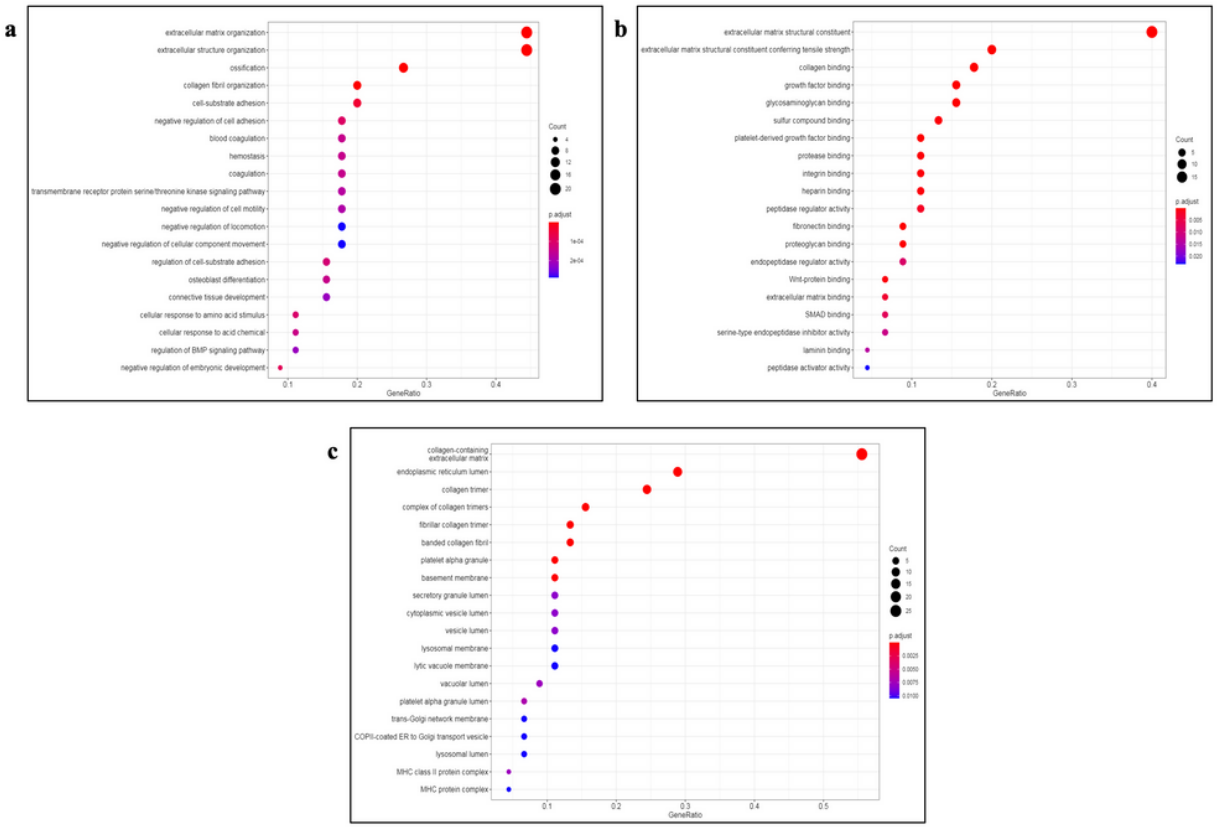

Figure 2

Dot plot representing GO - biological process where the DEGs were mainly enriched in ECM organization, extracellular structure organization, ossification, cell-substrate adhesion, and collagen fibril organization

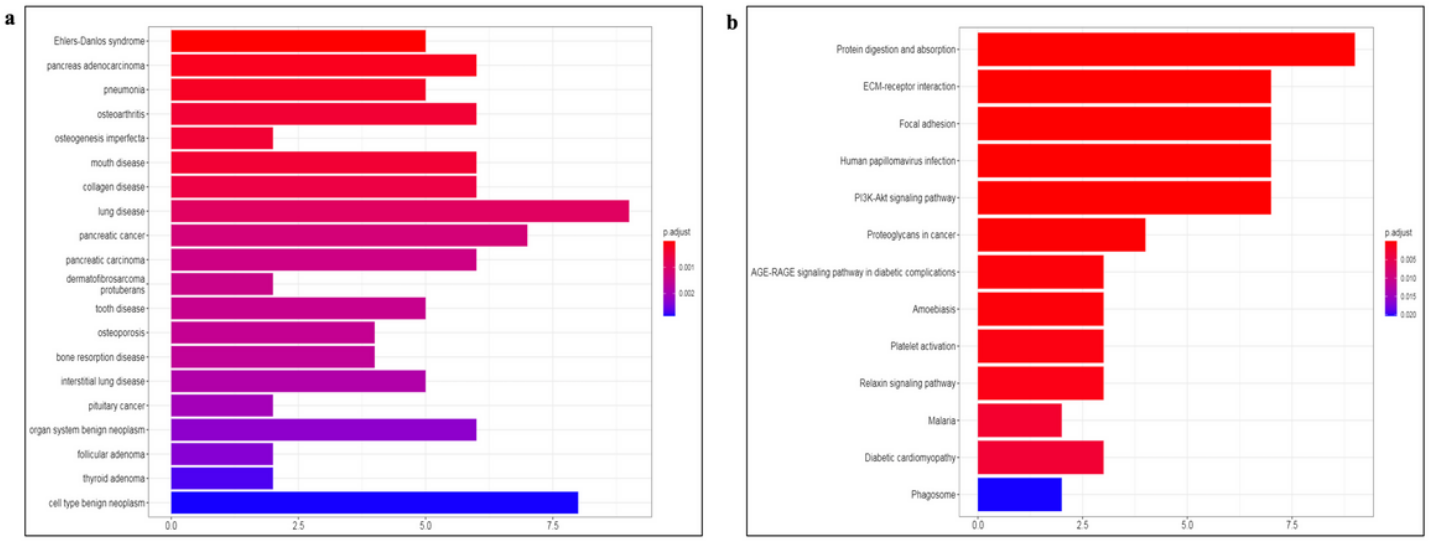

Figure 3

Bar plot representing disease ontology analysis for the 45 DEGs common to both pancreatitis and PDAC. The DEGs were significantly associated with lung disease, cell type benign neoplasm, and pancreatic cancer

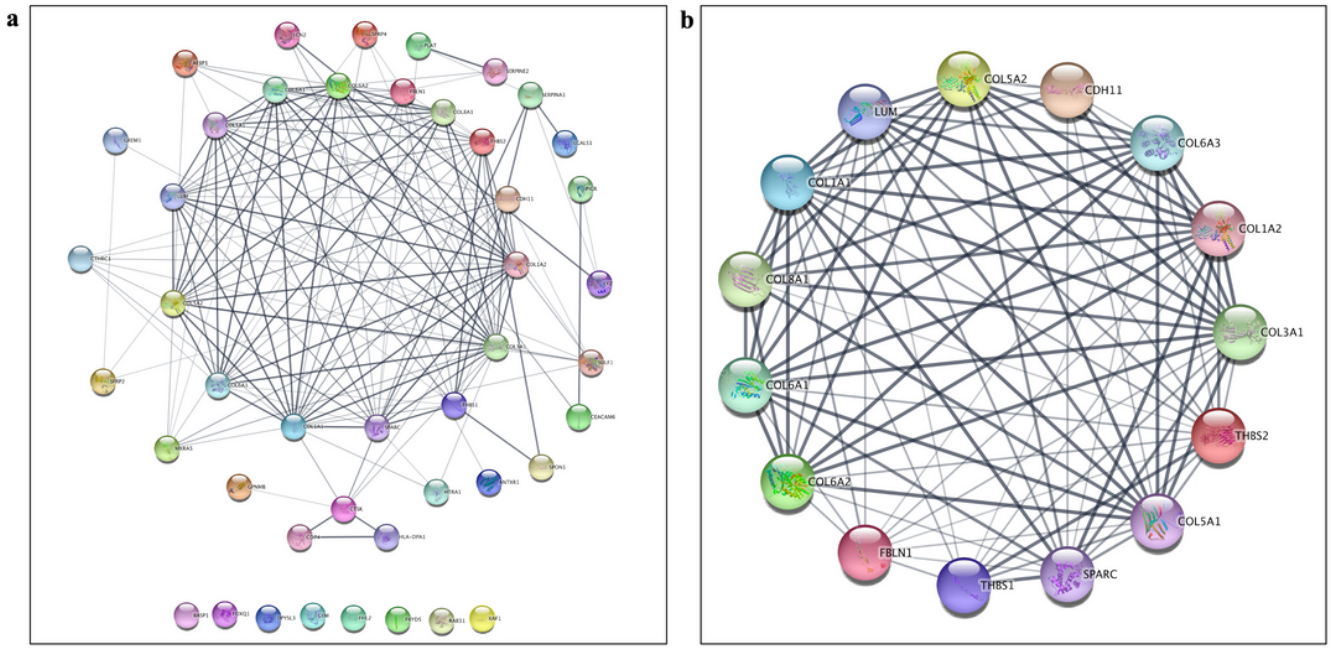

Figure 4

Represents protein-protein interaction (PPI) network constructed using the STRING database for the 45 DEGs common to both pancreatitis and PDAC

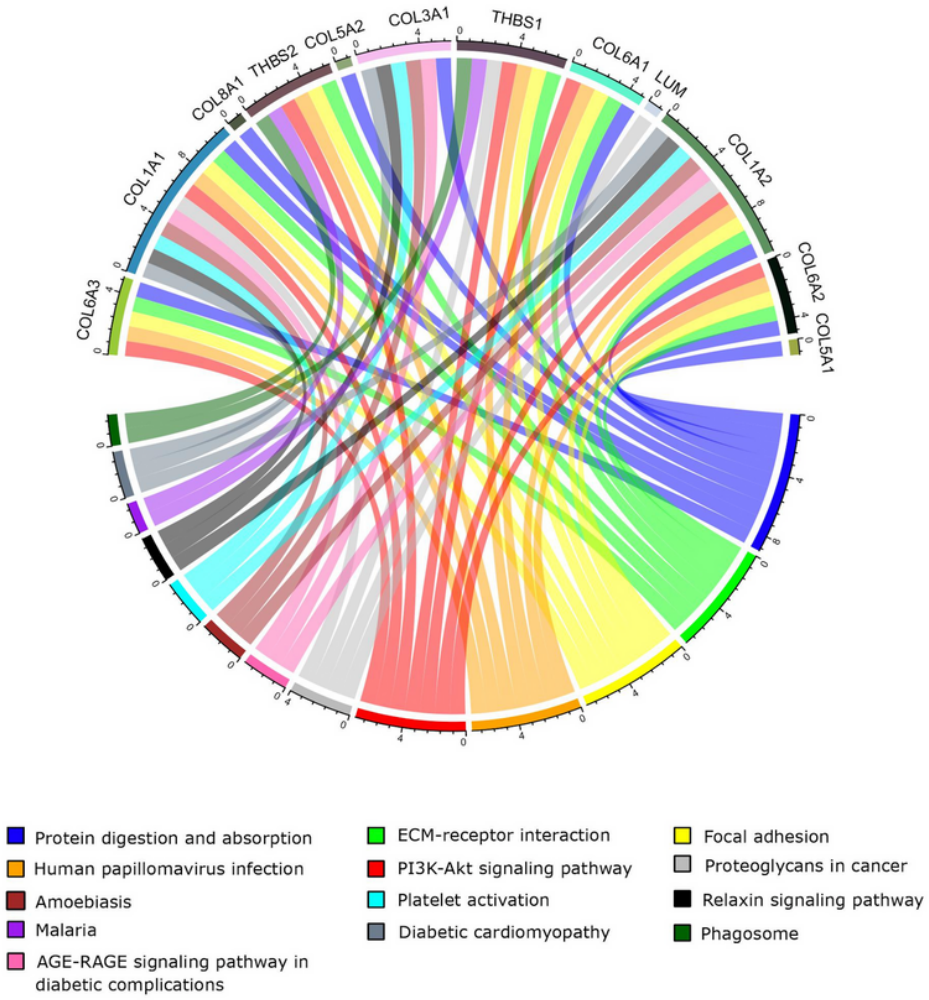

Figure 5

Chord diagram representing the KEGG pathway reanalysis for the 15 hub genes. Total of five core genes were identified - COL1A1, THBS1, COL1A2, THBS2 and COL3A1. Among these, the genes COL1A1 and COL1A2 alone were significantly associated with 11 different pathways each

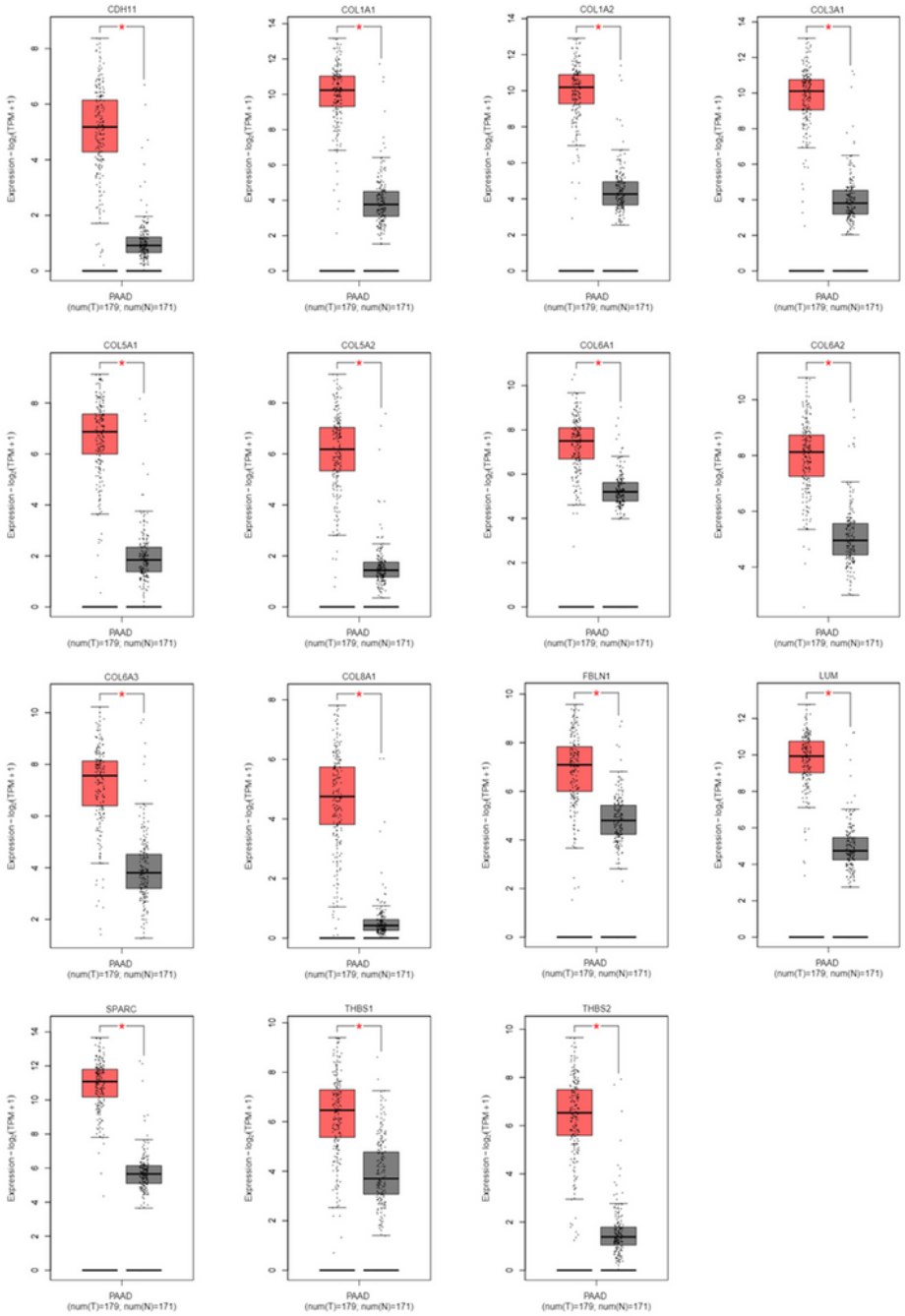

Figure 6

Box plots comparing the gene expression levels of PDAC and normal tissues for 15 hub genes P-value < 0.001 and Log2FC > 2

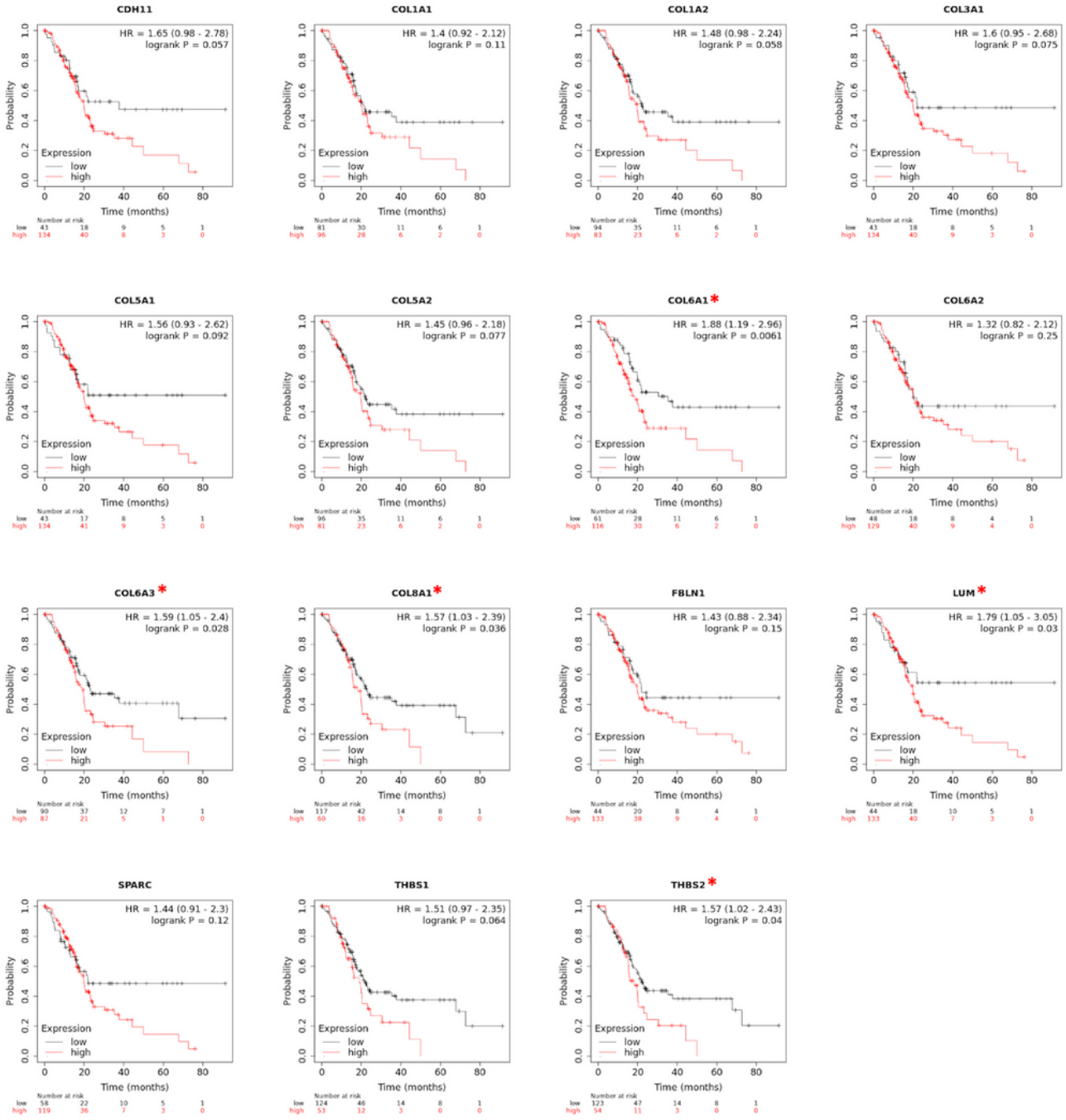

Figure 7

Kaplan-Meier plots representing the survival analysis for 15 hub genes with respect to low expression (black color) and high expression (red) in PDAC tissue samples. Among these, survival analysis for COL6A1, COL6A3, COL8A1, LUM and THBS2 genes (marked with *) are statistically significant $p < 0.05$

Supplementary Files

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

- Table1.xlsx