RESEARCH ARTICLE

Editorial Process: Submission:10/30/2022 Acceptance:05/15/2023

Integrated Bioinformatics Analysis Identifies Crucial Biochemical Processes Shared between Pancreatitis and Pancreatic Ductal Adenocarcinoma

Manoj M Wagle[1,2], Ananya Rao Kedige[2,3], Shama P Kabekkodu[4], Sandeep Mallya[2]*

Abstract

Background: Pancreatic ductal adenocarcinoma (PDAC) is a highly aggressive malignancy associated with rapid progression and an abysmal prognosis. Previous research has shown that chronic pancreatitis can significantly increase the risk of developing PDAC. The overarching hypothesis is that some of the biological processes disrupted during the inflammatory stage tend to show significant dysregulation, even in cancer. This might explain why chronic inflammation increases the risk of carcinogenesis and uncontrolled proliferation. Here, we try to pinpoint such complex processes by comparing the expression profiles of pancreatitis and PDAC tissues. **Methods:** We analyzed a total of six gene expression datasets retrieved from the EMBL-EBI ArrayExpress and NCBI GEO databases, which included 306 PDAC, 68 pancreatitis and 172 normal pancreatic samples. The disrupted genes identified were used to perform downstream analysis for ontology, interaction, enriched pathways, potential druggability, promoter methylation, and the associated prognostic value. Further, we performed expression analysis based on gender, patient's drinking habit, race, and pancreatitis status. **Results:** Our study identified 45 genes with altered expression levels shared between PDAC and pancreatitis. Over-representation analysis revealed that protein digestion and absorption, ECM-receptor interaction, PI3k-Akt signaling, and proteoglycans in cancer pathways as significantly enriched. Module analysis identified 15 hub genes, of which 14 were found to be in the druggable genome category. **Conclusion:** In summary, we have identified critical genes and various biochemical processes disrupted at a molecular level. These results can provide valuable insights into certain events leading to carcinogenesis, and therefore help identify novel therapeutic targets to improve PDAC treatment in the future.

Keywords: Pancreatic cancer- PDAC- Inflammation- Differentially Expressed Genes

Asian Pac J Cancer Prev, **24 (5)**, 1601-1610

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 (Kleeff et al., 2016). Patients with the malignancy rarely present the symptoms resulting in a poor diagnosis and increased mortality rates. 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 deaths (Siegel et al., 2018). 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 (Permuth-Wey and Egan, 2009; Shi et al., 2009). The incidence of PDAC in both males and females is higher in developed countries than in developing countries. Some studies have estimated that PDAC will become the second most common cause of cancer-related mortality by 2030 (Rahib et al., 2014). 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 (Alexakis et al., 2004; Neoptolemos et al., 2003).

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

*[1]School of Mathematics and Statistics and Computational Systems Biology Group, Children's Medical Research Institute, University of Sydney, New South Wales, Australia. [2]Department of Bioinformatics, Manipal School of Life Sciences, Manipal Academy of Higher Education, Karnataka, India. [3]Université Grenoble Alpes, CNRS, CEA, Grenoble, France. [4]Department of Cell and Molecular Biology, Manipal School of Life Sciences, Manipal Academy of Higher Education, Karnataka, India. *For Correspondence: psandeepmallya@gmail.com*

alterations in multiple signaling pathways result in an intrinsic chemoresistance that hinders the therapeutic drug delivery (Lee et al., 2008; Oberstein and Olive, 2013; Samuel and Hudson, 2011). Dysregulation of molecular pathways such as K-Ras occurs in 75-90% of pancreatic carcinomas, further stimulating downstream signaling cascades (Almoguera et al., 1988; Malumbres and Barbacid, 2003). Moreover, mutations in the transcription factor P53 (*TP53*) gene can be seen in more than ~60% of pancreatic cancers (Rozenblum et al., 1997). 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, it is crucial to identify the underlying mechanisms, cellular processes, and inflammatory pathways, which can further help us design drugs targeting these biomarkers (Khalafalla and Khan, 2017; Zheng et al., 2013).

The present study analyzes the microarray data of PDAC and pancreatitis tissues from publicly available datasets to derive the biological meaning of differentially expressed genes (DEGs) using bioinformatics methods. The results of this study provide valuable biological insights which could be further explored to identify novel therapeutic targets in PDAC.

Materials and Methods

Retrieval of gene expression datasets

The microarray data for normal, PDAC, and pancreatitis tissues were obtained from NCBI Gene Expression Omnibus (GEO) and EMBL-EBI ArrayExpress. A total of 6 datasets (GSE15471, GSE32676, GSE46234, E-MTAB-1791, E-GEOD-71989, and E-MEXP-2780) were analyzed in this study (Athar et al., 2019; Barrett et al., 2013). Since data was generated using different platforms, all the datasets belonging to the respective platforms (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 Figure 1.

Data pre-processing and differentially expressed genes (DEGs) screening

The datasets were pre-processed, normalized, and analyzed for differential expression using BRB-Array tool 4.6.1 (Stable Version) (Simon et al., 2008). The pre-processing and normalization criteria included - (i) If the spot intensity is below the minimum value i.e., 10, then threshold the intensity at the minimum value (ii) Average the replicate spots within an array (iii) Exclude a gene if the 50th percentile of intensities < 500 or the percentage of data filtered/missing > 50% (iv) Each array was normalized using quantile normalization. The DEGs screened conformed to the following cutoff criteria: $|\log FC| > 2$ and a high significance threshold of 0.001 of univariate tests. Only DEGs with a false discovery rate (FDR) < 0.05 were considered for further analysis. Overlapping DEGs between pancreatitis and the PDAC samples was identified using the Funrich software (Pathan et al., 2015).

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 (Hill et al., 2008). In contrast, the ontologies and comprehensive information on human diseases are described in the Disease Ontology (DO) (Schriml et al., 2019). KEGG is a database resource encompassing the functional meaning of a biological system derived mainly from high-throughput experiments (Kanehisa and Goto, 2000). We used the R Bioconductor package, clusterProfiler, which integrates the data from the above resources to perform ontology and enrichment analysis (Yu et al., 2012).

Protein-Protein Interaction (PPI) and module analysis

PPIs are crucial for several biological functions in the body, and any dysregulation can often indicate diseases (Gonzalez and Kann, 2012). In this study, we used the Search Tool for the Retrieval of Interacting Genes (STRING) database and the Cytoscape software (version 3.8.2) for the construction and visualization of interaction networks (Shannon et al., 2003; Szklarczyk et al., 2015). MCODE (Molecular Complex Detection) was used to identify the densely interconnected regions in the network 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 (Bader and Hogue, 2003).

Pathway reanalysis, potential druggability, and gene expression analysis

The hub genes obtained were then reanalyzed to identify core genes. The potential druggability was determined using the Drug-Gene Interaction Database (DGIdb). DGIdb is a web-based resource providing information about druggable candidate genes and potential drug-gene interactions (Freshour et al., 2021). The expression of hub genes was evaluated using the GPEIA2 tool (Tang et al., 2019). GEPIA2 performs gene expression analysis using the data from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTex), which helps us to validate and correlate the expression profiles between normal and PDAC tissues. The results of the KEGG pathway reanalysis were visualized using the R package, circlize (Gu et al., 2014).

Survival Analysis, tumor subgroup expression analysis, and promoter methylation

Survival analysis of the hub genes was performed using the Kaplan-Meier (KM) plotter (Nagy et al.,

2021). We also performed hub gene expression analysis based on factors such as gender, patient's drinking habit, race, and pancreatitis status using the UALCAN tool (Chandrashekar et al., 2017). Further, the promoter methylation profile of individual hub genes was compared between normal and PDAC tissues.

Results

Identification of differentially expressed genes

A total of 6 datasets were retrieved that included 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 genes that were differentially expressed in both PDAC and pancreatitis.

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 processes were extracellular matrix (ECM) organization, extracellular structure organization, ossification, cell-substrate adhesion, and collagen fibril organization (Figure 2a). (ii) In the molecular function group, the DEGs were mainly enriched in collagen binding, growth factor binding, EMSC

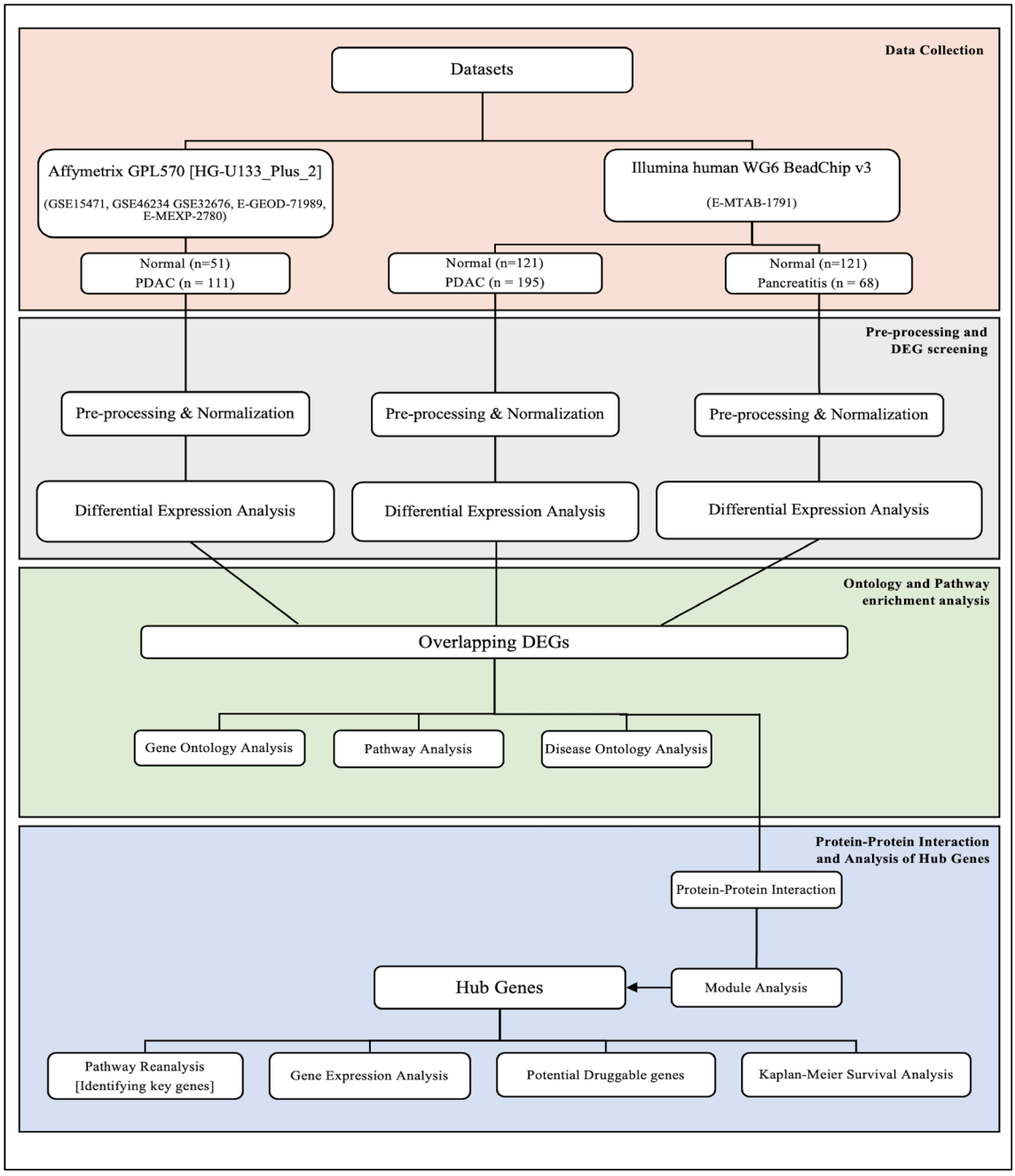

Figure 1. Flowchart Describing the Overview of the Methodology Followed in the Present Study. This involved collection of raw data & preprocessing, screening and identification of overlapping DEGs, ontology and pathway enrichment analysis, protein-protein interaction, and analysis of hub genes.

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

Druggable gene category	Gene Count	Gene (s)
Druggable genome	14	COL6A3, COL1A1, FBLN1, COL8A1, THBS2, COL5A2, SPARC, COL3A1, THBS1, COL6A1, LUM, COL1A2, COL6A2, COL5A1
Cell surface	2	SPARC, THBS1
Clinically actionable	2	COL1A1, CDH11
Drug resistance	2	COL1A1, THBS1
External side of plasma membrane	1	THBS1
Protease inhibitor	1	COL6A3
Transcription factor	1	THBS1

conferring tensile strength, and glycosaminoglycan binding (Figure 2b). (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 (Figure 2c). As shown in Figure 3a, 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,

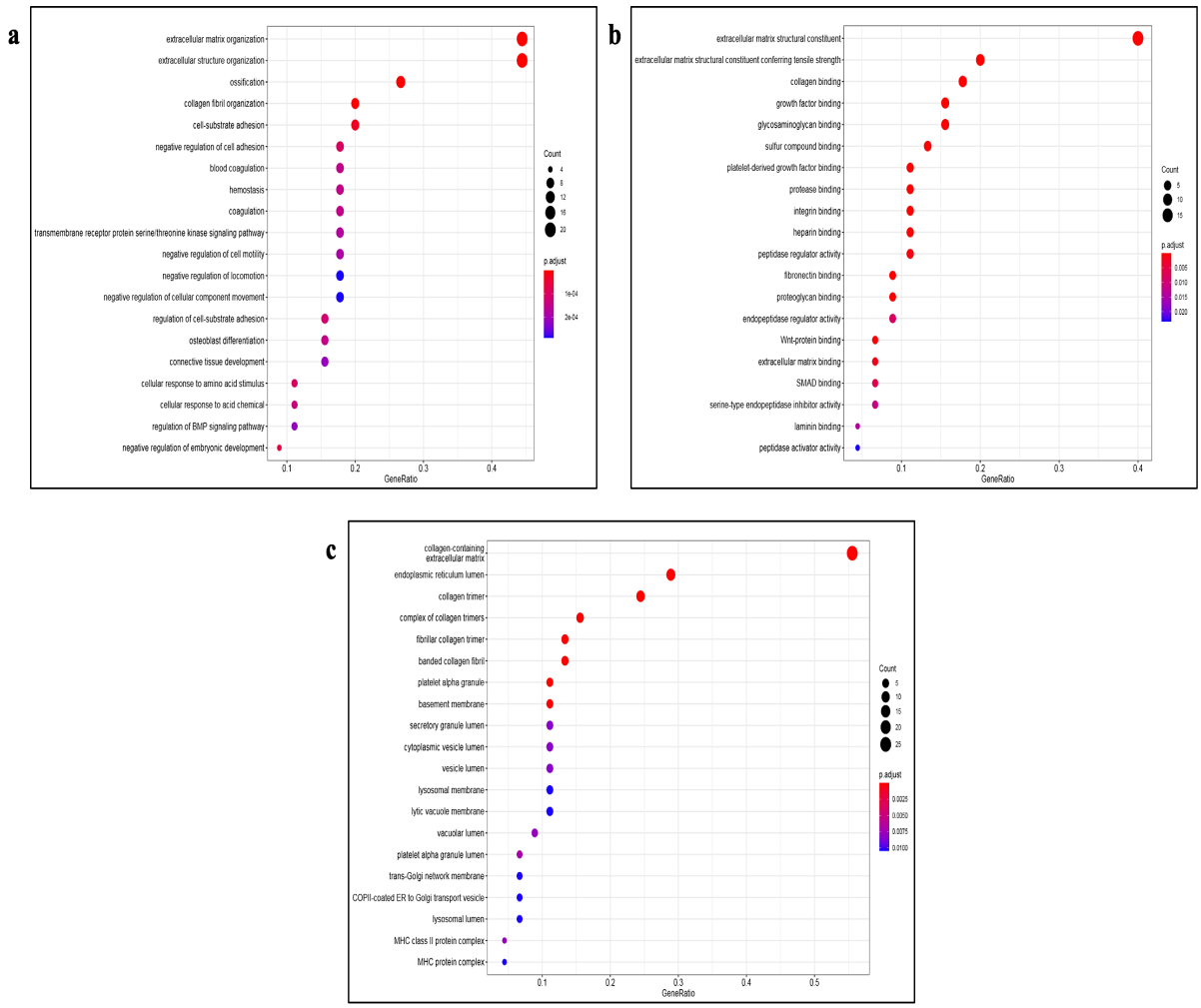

Figure 2, a Dot Plot Representing GO - Biological Process, Where the DEGs Were Mainly Enriched in ECM Organization, Extracellular Structure Organization, Ossification, Cell-substrate Adhesion, and Collagen Fibril Organization. b, Dot plot representing GO - molecular function. The DEGs were mainly enriched in collagen binding, growth factor binding, extracellular matrix structural constituent (EMSC), EMSC conferring tensile strength, and glycosaminoglycan binding. c, 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.

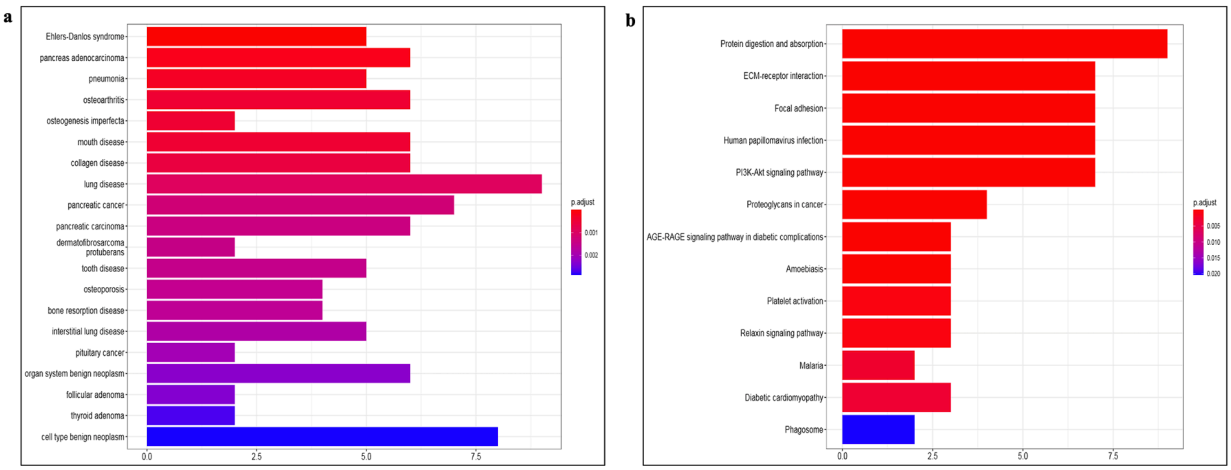

Figure 3. a, Bar Plot Representing Disease Ontology Analysis for the 45 DEGs Common to Both Pancreatitis and PDAC. The DEGs were significantly associated with lung disease, cell type benign neoplasm, and pancreatic cancer. b, Bar plot representing KEGG Pathway Analysis for the 45 DEGs common to both 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.

ECM-receptor interaction pathway, and proteoglycans in cancer pathways were significantly enriched (Figure 3b).

PPI network and identification of hub genes

The Protein-protein interaction network (PPI) constructed for the overlapping DEGs using the STRING database with a combined score > 0.4 (default threshold) showed 45 nodes with a total of 152 edges, representing a densely interconnected network (Figure 4a). Module analysis using MCODE revealed 15 hub genes (with 92 edges), including - COL6A3, COL1A1, FBLN1, COL8A1, THBS2, CDH11, COL5A2, SPARC, COL3A1, THBS1, COL6A1, LUM, COL1A2, COL6A2, and COL5A1 (Figure 4b).

Pathway reanalysis, druggability, and gene expression analysis

The 15 hub genes identified were then reanalyzed for KEGG pathways, and the following five core genes were identified - COL1A1, THBS1, COL1A2, THBS2, and COL3A1 (Figure 5). Among these, COL1A1 and COL1A2 were associated with 11 different pathways each. Expression analysis between normal and PDAC tissues showed that all 15 hub genes were found to be significantly expressed (P-value < 0.001 and Log2FC > 2) (Figure 6). It was found that 14 out of 15 genes were in the druggable genome category, suggesting that they could be modulated and interact with small molecules. The complete list of genes and their corresponding druggable gene category is shown in Table 1.

KM Survival Analysis, tumor subgroup & promoter methylation

The prognostic value associated with hub genes was analyzed using the KM plotter at a P-value threshold of < 0.05 (Figure 7). The results showed that

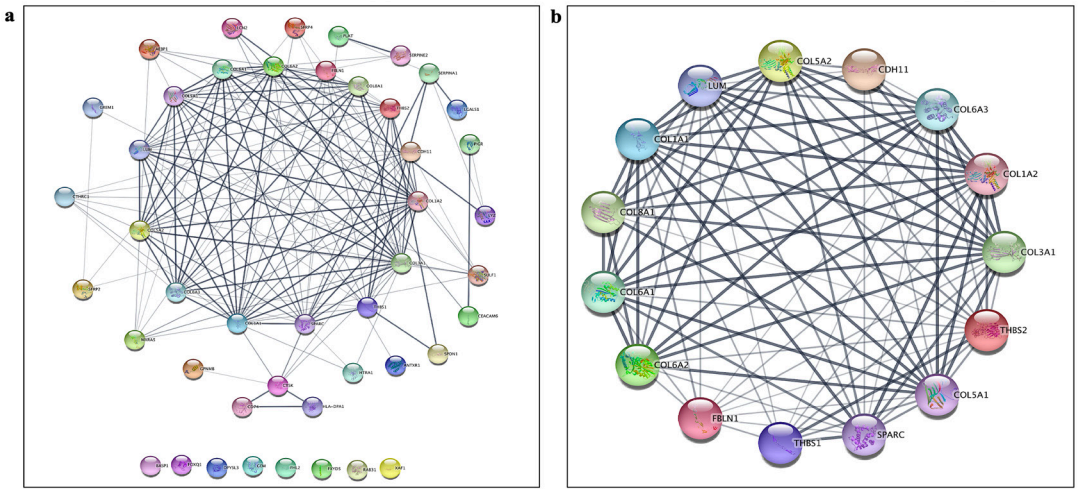

Figure 4. a, Represents the Protein-protein Interaction (PPI) Network Constructed Using the STRING Database for the 45 DEGs Common to Both Pancreatitis and PDAC. b, Represents the module analysis of the PPI network constructed using the Cytoscape app, MCODE showing 15 nodes (hub genes) with 92 edges.

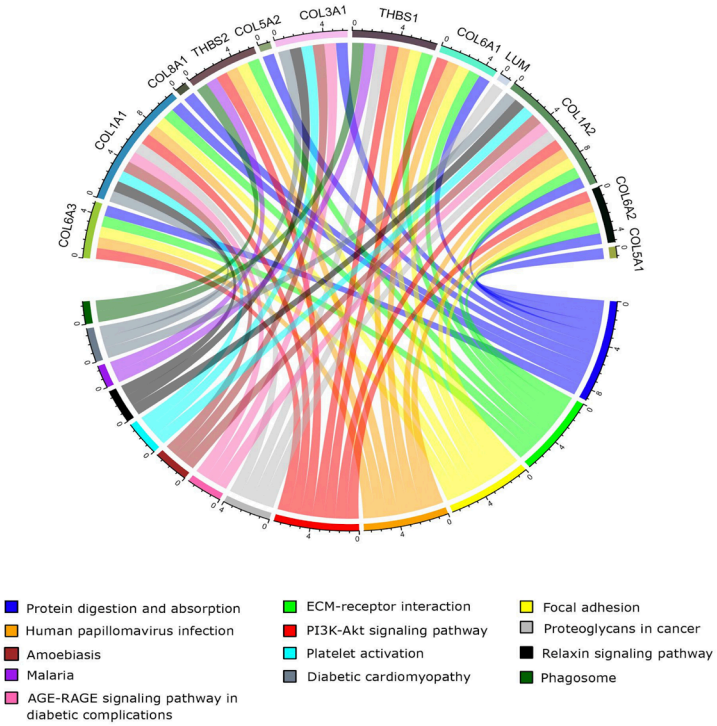

Figure 5. Chord Diagram Representing the KEGG Pathway Reanalysis for the 15 Hub Genes. A total of five core genes were identified - COL1A1, THBS1, COL1A2, THBS2 and COL3A1. Among these, the genes COL1A1, and COL1A2 alone were significantly associated with 11 different pathways each.

the genes - COL6A1, COL6A3, COL8A1, LUM & THBS2 caused a significant reduction in the overall survival rate of PDAC patients, with COL6A1 being the most statistically significant (log-rank P = 0.0061). Next, the tumor subgroup analysis of the hub genes was performed. The analysis based on gender did not reveal any notable differences in the gene expression between male and female PDAC patients (Suppl. Figure. 1). Similarly, the expression values did not vary much with or without the presence of chronic pancreatitis for most of the hub genes. However, higher transcript per million (TPM) values were observed for the gene - COL6A1 in patients with pancreatitis than in normal and non-pancreatitis patients (Suppl. Figure. 2). Although not statistically significant, samples from 'occasional drinkers' showed higher TPM values compared to other groups. But this can be overlooked based on the observation that the data from occasional drinkers cover a greater range indicating highly probable values making it less reliable to conclude (Suppl. Figure. 3). Expression analysis based on different races showed that African Americans exhibited higher median TPM values, with gene - FBLN1 being most statistically significant compared to Asians, Caucasians, and normal samples. The observations and comparisons between the races are biased due to differences in sample count and thus have low statistical significance (Suppl. Figure. 4). Evaluation of regulation of gene expression by promoter methylation revealed no significant change in the methylation profiles for most of the hub genes, except gene - COL3A1, which showed a deviation compared to normal samples. No methylation profile data was available for the gene - LUM (Suppl Figure 5).

Discussion

Cancer is one of the complex diseases resulting from various phenomena, including significant gene-environment interactions that result in disordered cell proliferation. 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 (Carrière et al., 2009; Kirkegård et al., 2017). The present study aims to identify dysregulated genes, pathways, and biochemical processes shared between pancreatitis and PDAC. Analysis of six different datasets obtained from publicly available databases revealed a total of 45 overlapping DEGs between pancreatitis and PDAC. These DEGs were mainly enriched in collagen and growth factor binding, extracellular environment, and cell adhesion. Collagen is a crucial component of the extracellular matrix (ECM), and specific orientation and arrangements of ECM in a microscopic environment are thought to play essential roles in tumor progression (Cox and Erler, 2011; Friedl and Wolf, 2008). 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 (Fang et al., 2014; Paszek et al., 2005; Xu et al., 2019).

KEGG pathway analysis showed that the protein

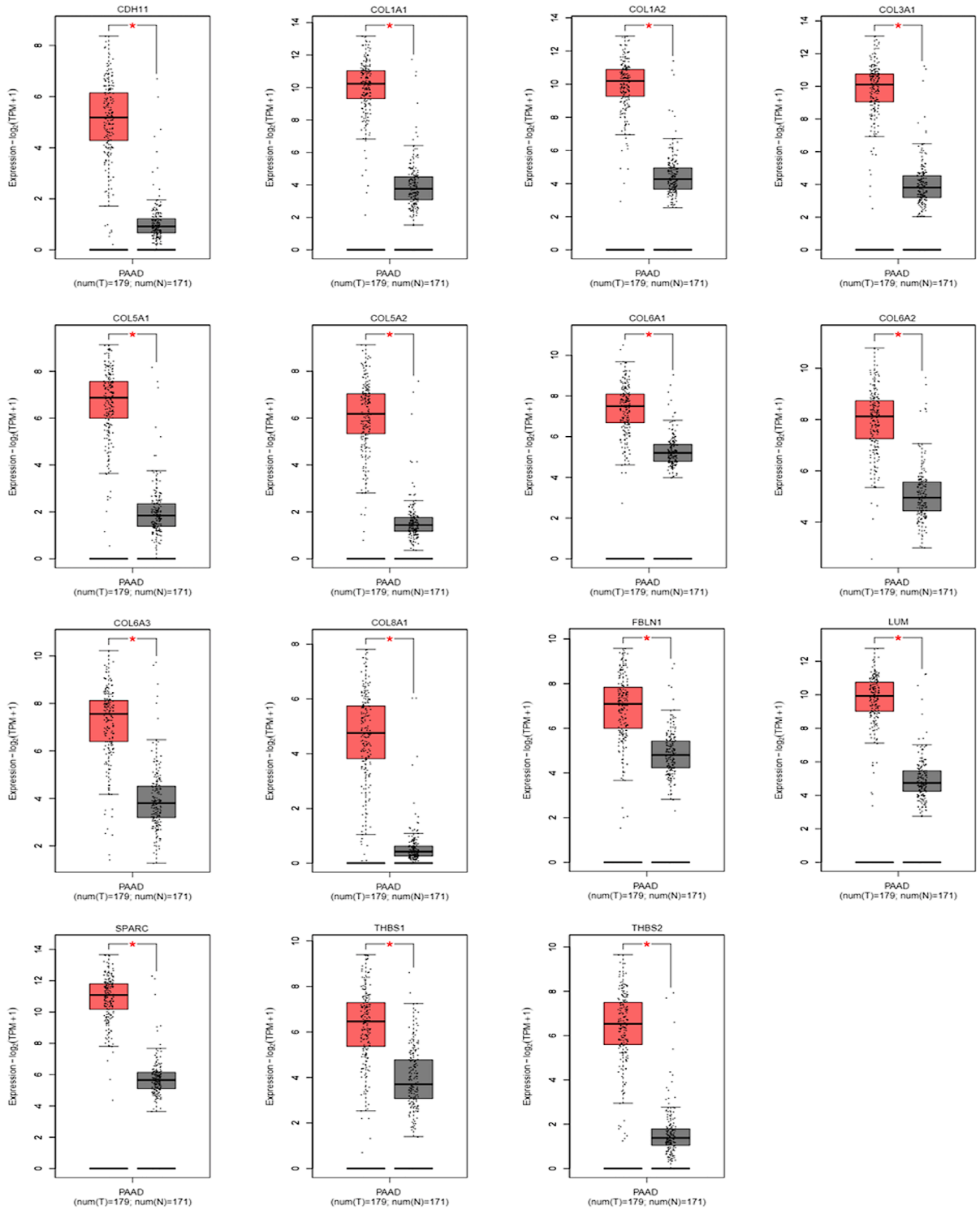

Figure 6. Box Plots Comparing the Gene Expression Levels of PDAC and Normal Tissues for 15 Hub Genes P-value < 0.001 and Log2FC > 2. The samples from normal tissues are shown in grey and PDAC tissues in red. Gene expression changes between groups in all hub genes were found to be significant (marked with an *).

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. Moreover, 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 (Baer et al., 2014; Eser et al., 2013; Kennedy et al., 2011). These are the major reasons why targeting the PI3k-Akt pathway has been a significant interest in cancer drug discovery. Proteoglycans are another important biomolecule of interest, having multiple

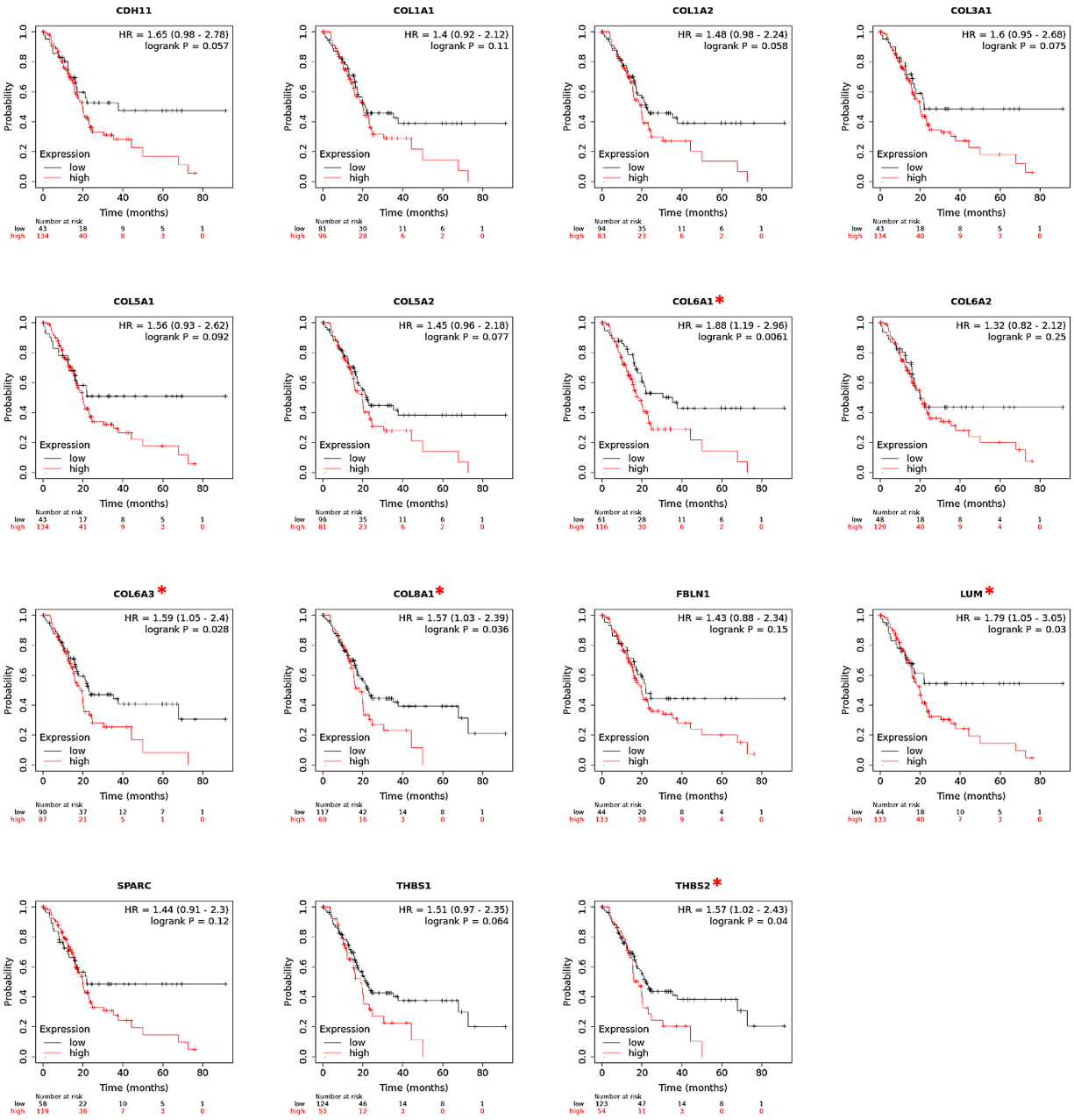

Figure 7. Kaplan-Meier Plots Representing the Survival Analysis for 15 Hub Genes with Respect to Low Expression (black color) and High Expression (red) in PDAC Tissue Samples. Among these, genes - COL6A1, COL6A3, COL8A1, LUM, and THBS2 (marked with an *) are statistically significant P-Value < 0.05.

functions in angiogenesis and cancer. They often influence cell growth through their interaction with growth factors and can sometimes cause deregulation of cell proliferation (Knelson et al., 2014; Wang et al., 2011). Thus, their integration in tumor cell diagnostics can facilitate early diagnosis, as demonstrated in a few studies (Effenberger et al., 2018; Kurihara et al., 2008).

We further constructed a protein-protein interaction network and performed a module analysis. The module consisted of 15 nodes with 92 edges. Expression analysis using the GEPIA2 tool revealed all 15 hub genes to be significantly expressed in PDAC. Interestingly, the expression of most of the hub genes was independent of factors such as gender, drinking habits, race, and pancreatitis status, which suggests that these genes can be used as biomarkers on 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. Based on this, we identified five core genes - COL1A1, COL1A2, THBS1, THBS2, and COL3A1. Among these, three were protein-coding collagen genes. Recently, a few studies have demonstrated how various collagen genes can play a role in tumorigenesis leading to poor clinical outcomes

(Kita et al., 2009; Wu et al., 2013). Further, differential expression of genes COL1A1, COL1A2, and THBS1 has been reported in several cancers, including colorectal cancer, hepatocellular carcinoma, and melanoma (Bonazzi et al., 2011; Hayashi et al., 2014; Zhang et al., 2018). Researchers have also demonstrated the utilization of Thrombospondin-2 (THBS2) as a biomarker for risk prediction and early detection of PDAC and as a robust prognostic indicator in colorectal cancer (Kim et al., 2017; Tian et al., 2018).

Taken together, this study analyzed a total of six different datasets by comprehensive bioinformatics analysis to identify critical genes and various biochemical processes thought to play crucial roles in certain events leading to carcinogenesis and its progression in PDAC. 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 of samples was not analyzed due to inaccessibility. Despite these drawbacks, the integrative approach used in the present study offers more precise findings compared to other studies that analyzed only a single dataset. Majorly, we identified 45 DEGs shared between PDAC and Pancreatitis, including 15 hub genes, namely, COL6A3, COL1A1, FBLN1, COL8A1, THBS2, CDH11, COL5A2, SPARC, COL3A1, THBS1, COL6A1, LUM, COL1A2, COL6A2, and COL5A1. In-depth experimental research is needed to elucidate the role and exact molecular mechanisms of these genes. This can further help identify novel therapeutic targets to improve PDAC treatment. Such personalized therapies can significantly reduce the sequelae of cancer treatment while improving patients' quality of life and overall survival rate.

Author Contribution Statement

All authors contributed to the present study. MMW and SM conceptualized the work. MMW and ARK acquired, analyzed, and interpreted the data. SPK and SM supervised and validated the work. MMW wrote the original draft of the manuscript. MMW and SM were responsible for subsequent revisions and editing of the manuscript. All authors approved the final version of the manuscript.

Acknowledgments

We would like to thank Dr. K. Satyamoorthy (Former Director, Manipal School of Life Sciences, MAHE, Manipal) and Dr. Bobby Paul (Associate Professor and Head, Department of Bioinformatics, Manipal School of Life Sciences, MAHE, Manipal) for their support.

Funding statement

The authors wish to thank the Department of Science and Technology - Fund for Improvement of Science & Technology Infrastructure (DST-FIST), Government of India, Technology Information Forecasting Assessment Council - Centre of Relevance and Excellence (TIFAC-CORE) in Pharmacogenomics, and the Manipal Academy of Higher Education (MAHE), Manipal, India, for providing the necessary facilities and infrastructure for the study.

Ethical/ Scientific body approval

The present study included secondary data analysis, therefore, does not require approval from the Ethical committee/Scientific body.

Data availability

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

Conflict of Interest

The authors declare no competing interests.

References

Alexakis N, Halloran C, Raraty M, et al (2004). Current standards of surgery for pancreatic cancer. *Br J Surg*, **91**, 1410–27.

Almoguera C, Shibata D, Forrester K, et al (1988). Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. *Cell*, **53**, 549–54.

Athar A, Füllgrabe A, George N, et al (2019). ArrayExpress update – from bulk to single-cell expression data. *Nucleic Acids Res*, **47**, 711–5.

Bader GD, Hogue CWV (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, **4**, 1–27.

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**, 2621–35.

Barrett T, Wilhite SE, Ledoux P, et al (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*, **41**, 991–5.

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.

Carrière C, Young AL, Gunn JR, Longnecker DS, Korc M (2009). Acute pancreatitis markedly accelerates pancreatic cancer progression in mice expressing oncogenic Kras. *Biochem Biophys Res Commun*, **382**, 561–5.

Chandrashekar DS, Bashel B, Balasubramanya SAH, et al (2017). UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia*, **19**, 649–58.

Cox TR, Erler JT (2011). Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. *Dis Model Mech*, **4**, 165–78.

Effenberger KE, Schroeder C, Hanssen A, et al (2018). Improved risk stratification by circulating tumor cell counts in pancreatic cancer. *Clin Cancer Res*, **24**, 2844–50.

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**, 406–20.

Fang M, Yuan J, Peng C, Li Y (2014). Collagen as a double-edged sword in tumor progression. *Tumor Biol*, **35**, 2871–82.

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**, 1144–51.

Friedl P, Wolf K (2008). Tube Travel: The Role of Proteases in Individual and Collective Cancer Cell Invasion. *Cancer*

Res, **68**, 7247–9.
Gonzalez MW, Kann MG (2012). Chapter 4: Protein Interactions and Disease. *PLoS Comput Biol*, **8**, e1002819.
Gu Z, Gu L, Eils R, Schlesner M, Brors B (2014). circlize implements and enhances circular visualization in R. *Bioinformatics*, **30**, 2811–2.
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*, **14**, 1–10.
Hill DP, Smith B, McAndrews-Hill MS, Blake JA (2008). Gene Ontology annotations: What they mean and where they come from. *BMC Bioinformatics*, **9**, 1–9.
Kanehisa M, Goto S (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*, **28**, 27–30.
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.
Khalafalla FG, Khan MW (2017). Inflammation and Epithelial-Mesenchymal Transition in Pancreatic Ductal Adenocarcinoma: Fighting Against Multiple Opponents. *Cancer Growth Metastasis*, **10**, 1179064417709287.
Kim J, Bamlet WR, Oberg AL, et al (2017). Detection of early pancreatic ductal adenocarcinoma with thrombospondin-2 & CA19-9 blood markers. *Sci Transl Med*, **9**, 398.
Kirkegård J, Mortensen FV, Cronin-Fenton D (2017). Chronic Pancreatitis and Pancreatic Cancer Risk: A Systematic Review and Meta-analysis. *Am J Gastroenterol*, **112**, 1366–72.
KitaY, Mimori K, Tanaka F, et al (2009). Clinical significance of LAMB3 and COL7A1 mRNA in esophageal squamous cell carcinoma. *Eur J Surg Oncol*, **35**, 52–8.
Kleeff J, Korc M, Apte M, et al (2016). Pancreatic cancer. *Nat Rev Dis Primers*, **2**, 16022.
Knelson EH, Nee JC, Blobe GC (2014). Heparan sulfate signaling in cancer. *Trends Biochem Sci*, **39**, 277–88.
Kurihara T, Itoi T, Sofuni A, et al (2008). Detection of circulating tumor cells in patients with pancreatic cancer: a preliminary result. *J Hepato-Biliary-Pancreatic Surg*, **15**, 189–95.
Lee CJ, Li C, Simeone DM (2008). Human pancreatic cancer stem cells: Implications for how we treat pancreatic cancer. *Transl Oncol*, **1**, 14–8.
Malumbres M, Barbacid M (2003). RAS oncogenes: the first 30 years. *Nat Rev Cancer*, **3**, 459–65.
Nagy Á, Munkácsy G, Győrffy B (2021). Pancancer survival analysis of cancer hallmark genes. *Sci Rep*, **11**, 1–10.
Neoptolemos JP, Cunningham D, Friess H, et al (2003). Adjuvant therapy in pancreatic cancer: historical and current perspectives. *Ann Oncol*, **14**, 675–92.
Oberstein PE, Olive KP (2013). Pancreatic cancer: why is it so hard to treat?. *Therap Adv Gastroenterol*, **6**, 321–37.
Paszek MJ, Zahir N, Johnson KR, et al (2005). Tensional homeostasis and the malignant phenotype. *Cancer Cell*, **8**, 241–54.
Pathan M, Keerthikumar S, Ang CS, et al (2015). FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics*, **15**, 2597–601.
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**, 109–17.
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**, 2913–21.
Rozenblum E, Schutte M, Goggins M, et al (1997). Tumor-suppressive Pathways in Pancreatic Carcinoma. *Cancer Res*, **57**, 1731–4.
Samuel N, Hudson TJ (2011). The molecular and cellular heterogeneity of pancreatic ductal adenocarcinoma. *Nat Rev Gastroenterol Hepatol*, **9**, 77–87.
Schriml LM, Mitraka E, Munro J, et al (2019). Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res*, **47**, D955–62.
Shannon P, Markiel A, Ozier O, et al (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res*, **13**, 2498–504.
Shi C, Hruban RH, Klein AP, Hruban O, Sol Goldman Pancreatic T (2009). Familial Pancreatic Cancer. *Arch Pathol Lab Med*, **133**, 365–74.
Siegel RL, Miller KD, Jemal A (2018). Cancer statistics, 2018. *CA Cancer J Clin*, **68**, 7–30.
Simon R, Lam A, Li M-C, et al (2008). Analysis of Gene Expression Data Using BRB-Array Tools. *Cancer Inform*, **6**, 9–15.
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**, 447–52.
Tang Z, Kang B, Li C, Chen T, Zhang Z (2019). GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res*, **47**, 556–60.
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**, 1331–6.
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.
Wu YH, Chang TH, Huang YF, Huang HD, Chou CY (2013). COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. *Oncogene*, **33**, 3432–40.
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.
Yu G, Wang LG, Han Y, He QY (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS*, **16**, 284–7.
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**, 1869–80.
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**, 1230–40.

This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.