Biological and Clinical Sciences Research Journal
eISSN: 2708-2261; pISSN: 2958-4728
www.bcsrj.com
DOI: https://doi.org/10.54112/bcsrj.v6i8.1687
*Biol. Clin. Sci. Res. J., Volume 6(8), **2025**: 1687*
Original Research Article

Identification of Candidate Genes Involved in Causing Pancreatic Cancer Through an Integrated Bioinformatics Approach

Atiqa Nawab[1], Maliha Ghaffar*[1], Saman Zahra[1], Ayesha Zulifqar[1], Humaira Amin[2], Samavia Mustafa[1], Anam Abbas[1]

[1]Faculty of Life Sciences, Department of Biology, University of Okara, Pakistan
[2]Cholistan University of veterinary and animal sciences, Bahawalpur, Punjab, Pakistan 63100
**Corresponding author`s email address:* maliha.ghaffar@uo.edu.pk

(Received, 14th April 2025, Accepted 15th August 2025, Published 31st August 2025)

Abstract: *Pancreatic ductal adenocarcinoma (PDAC) is an aggressive malignancy with poor survival, underscoring the need for robust molecular markers that can refine diagnosis and guide therapy.* ***Objective:*** *The goal of this study is to enhance diagnostic and treatment approaches for Pancreatic Ductal Adenocarcinoma (PDAC), a highly aggressive malignancy. We utilized RNA sequencing (RNA-Seq) data to identify differentially expressed genes (DEGs), examine gene networks, and uncover critical molecular links that are involved in the development and progression of PDAC.* ***Methods:*** *Four microarray datasets (GSE15471, GSE16515, GSE62165, and GSE62452) were obtained from the Gene Expression Omnibus (GEO). Data normalization and differential expression analysis were conducted using GEO2R, with a p-value cut-off of < 0.001 to identify DGEs. Functional annotation of these genes was carried out using DAVID Bioinformatics Resources, focusing on KEGG pathways and Gene Ontology (GO) terms. Gene interaction networks were constructed using STRING, and hub genes were identified with the cytoHubba plugin in Cytoscape.* ***Results:*** *A total of 614 up-regulated and 314 down-regulated genes were identified. The functional annotation of these genes revealed notable enrichment in several KEGG pathways and GO categories associated with molecular functions, biological processes, and cellular constituents. Key biological pathways implicated in PDAC were identified by the gene interaction networks derived from the DEGs. The top ten hub genes were identified.* ***Conclusion:*** *This comprehensive analysis of RNA-Seq data from multiple PDAC datasets successfully identified hub genes. These findings provide a molecular foundation for developing novel therapeutic strategies and improved diagnostic markers for the treatment of PDAC.*

Keywords: Pancreatic Cancer, DEG, GEO2R, KEGG, STRING, Hub genes

[***How to Cite:*** Nawab A, Ghaffar M, Zahra S, Zulifqar A, Amin H, Mustafa S, Abbas A. Identification of candidate genes involved in causing pancreatic cancer through an integrated bioinformatics approach. *Biol. Clin. Sci. Res. J.*, **2025;** 6(8): 109-120. doi: https://doi.org/10.54112/bcsrj.v6i8.1687]

Introduction

Pancreatic Ductal Adenocarcinoma (PDAC) is considered a highly malignant cancer (1). The late disclosure and aggressiveness of PDAC present numerous challenges, often leading to treatment failure (2). This analysis aims to use the RNA-Seq data to identify variants and discover mutational signatures present in PDAC tumor samples (3). The irregular growth of cells in the pancreas leads to the development of cancer. These cells divide uncontrollably and eventually form a tumor (4). The scientific community is still investigating various studies on how new cells in the pancreas grow before the body requires them, or how old pancreatic cells may remain alive longer than they should (5).

This excess of cells results in the formation of a tumor or tissue mass. A tumor in the pancreas can be either malignant or benign. Recently, an enzyme in pancreatic cells was discovered that initiates metastatic disease. A method for detecting this enzyme is expected to undergo clinical trials soon. (6). Pancreatic cancer is considered an eminently malignant disease, with a survival rate of approximately 10% at five years in the USA, making it a leading cause of cancer-related mortality. Several risk factors are linked to the development of pancreatic tumors, including obesity, tobacco use, type 2 diabetes, and family history.(7). Computed tomography with high quality and intravenous contrast by utilizing dual state pancreatic protocol/code is commonly contemplated as the best procedure for pancreatic cancer analysis and to examine the surgicalability (8). The endoscopic ultrasound, when combined with fine needle aspiration, is a widely used integral staging modality that enables diagnostic confirmation. (9).

Many pancreatic tumors are of a complex nature at the epigenetic, metabolic, and genomic levels, with several crosstalk evident and activated pathways. (10)There exist multiple kinds of pancreatic tumors. However, this research will center on the ductal adenocarcinomas as they are considered for a massive bulk of the fatal pancreatic neoplasms. (9). In the United States, ductal adenocarcinoma accounts for fewer than 2% of contemporary cases, and it is considered the fifth leading cause of tumor-related death. (11).

The aggressiveness of tumor cells causes the poor diagnosis of pancreatic tumors in patients, the early metastasis, as well as the non-responsiveness towards chemotherapy regimens. (12). Hence, candidate gene access proposes huge potential for expanding the understanding of behavioral ecology. The variations in candidate genes' expression can expose their input towards behavioral changes and phenotypic plasticity. (13). Transcriptomic research on the complete genome can identify probable candidate genes related to specific organismal ideology. Nonetheless, the significance of these candidates is seldom verified through functional research or by correlating the findings across autonomous research. (14). The functional significance of putative candidate genes identified in transcriptome analysis should be validated through further experiments. (15).

Selecting candidate genes is similar to ranking and identifying multiple risk factors present in epidemiological studies. Researchers must select from a vast array of potential factors only those that are highly probable to be associated with the phenotype. (16)Functional genomics approaches, such as profiling of RNA expression through microarrays and high-throughput DNA sequencing, enable precise localization of the biological facts within the genome (17).

To identify potential genes involved in the development and proliferation of pancreatic tumors, the Gene Expression Omnibus (GEO) is utilized. This resource is also helpful for identifying differentially expressed genes (DEGs) that distinguish between samples with nonmalignant conditions and those with PDAC.(18). Using the Database for the Annotation, Visualization, and the Integrated Discovery (DAVID) web application, the DEGs were functionally annotated. (19). A PPI (protein-protein

interaction) system of DEGs was established by utilizing the Search Tool for Retrieval of the Interacting Genes (STRING) database. (20).

This study, most importantly, will utilize bioinformatics tools to identify differentially expressed genes (DEGs) in various microarray data sets, including GSE15471, GSE16515, GSE62165, and GSE62452. Additionally, it will employ a differential bioinformatics technique to identify hub genes involved in PDAC.

The main objective is to explore the underlying biological pathways and gene networks involved in the development of pancreatic cancer using bioinformatics tools, such as Gene expression analysis and comparison of DEGs using GEO2R across multiple pancreatic cancer datasets. KEGG pathway and Gene Ontology (GO) term analysis to ascertain the functions and pathways of common genes connected to pancreatic cancer. Visualization of gene networks in pancreatic cancer datasets using tools like STRING and Cytoscape, aiming to reveal key molecular interactions by identifying hub genes.

Methodology

-Datasets Retrieval

Gene Expression Omnibus (GEO) has retrieved four databases particularly for pancreatic cancer with GEO IDs as GSE15471, GSE16515, GSE62165, and GSE62452. GEO is considered a public data repository for functional genomics, leveraging advancements in next-generation sequencing, microarray technologies, and more. Moreover, the GEO consists of tools for the analysis, visualization, and identification of interest-specific data of users. (21).

-Normalization of Datasets

GEO2R

Datasets generated through microarray platforms were normalized through GEO2R. (22) which is a tool provided by GEO, and uses different packages from Bioconductor(23) To perform data analysis using the R language. GEO2R uses limma(24) and GEOQuery(25) packages of Bio conductor. To examine differential expression, GEO2R uses t-test (26) and default multiple-testing correction method is Benjamini and Hochberg FDR (27). It provides the option of choosing any other desired method from the "Options" tab's list on the console. It also applies log transformations on the data. To force the implementation of log2 transformation on the data "Yes" was checked in "Options" tab.

Selection Criteria for DEG

A DEG is a gene which has statistically significance difference in its expression levels between two experimental conditions (upregulated/downregulated) (28). The cut-off criteria for selection of DEGs from the datasets for comparison to find out the common DEGs was Adjusted p-value <0.001.

Identification of Common DEG between Datasets for Comparison

Common DEGs between all 4 datasets were identified though Venny, a web-based tool which shows the results in the form of graphically interactive venn diagrams. It can compare up to 4 lists of the genes. Input can be given both in the form of official symbols or any type of IDs of the genes. Input is provided one element per row (29).

- Functional Annotation of DEG

Gene Ontology (GO) Terms Enrichment Analysis

Using the built-in tools for functional annotation, DAVID (The Database for Annotation, Visualization and Integrated Discovery) Bioinformatics resource, researchers were able to independently perform enrichment analysis of common up- and down-regulated DEGs from each dataset using Gene Ontology (GO) terms (30). DAVID has four major tools. i) Functional annotation tool, ii) Gene functional classification tool, iii) Gene ID conversion tool and iv) Gene name batch viewer (31)**.**

GO term project aims to hierarchically classify the genes or their products into terms categorized into 3 classes comprising Molecular Function (MF), Biological Process (BP) and Cellular Component (CC) (32). Any biochemical activity in which the gene product is involved is known as MF. It describes only about the activity which happens and does not provide any information about the time and place of that specific activity. Examples include ATP binding, protein binding etc. CC category of GO terms provides information about the place where a certain gene product is performing its specific activity. Examples are Golgi apparatus, ribosomes etc. BP is any process which is occurring in living systems. It involves certain pathways and reactions in which the genes or gene products contribute. Examples include cAMP synthesis, signal transduction etc(33).

KEGG Pathways Analysis

KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways analysis was also carried out through the DAVID Bioinformatics database. KEGG is a knowledgebase that systematically analyzes genes/gene products and links information of these genes or gene products with high-level functional information (34). KEGG comprises three major databases. The GENES database contains genes from 24 complete and 12 partial genomes, along with annotations related to their functions. The PATHWAY database contains higher-order functional information in the form of graphical illustrations of different types of cellular processes, like signal transduction, cell cycle, etc., while the third database is LIGAND, which contains information about enzymatic reactions, chemical compounds, and enzyme molecules (34)The cut-off criteria for output of GO terms, which were in the form of enrichment of genes in 3 categories of GO terms (MF, CC, and BP) and KEGG pathway terms, were P-value<0.05 and counts (number of genes)>10.

Construction of Gene Interaction Networks

The STRING tool was utilized to create gene networks for the DEGs. (Search Tool for the Retrieval of Interacting Genes) (35) Which is available as a plugin of Cytoscape as stringAPP, and the network was analyzed through the NetworkAnalyzer tool of Cytoscape (36). It analyzes the networks and calculates various topological parameters, including the network's diameter, degrees of individual nodes, clustering, and topological coefficients. Results are displayed diagrammatically and can be saved as text files or images. (37).

Cytoscape is an open source package of bioinformatics software for the construction and visualization of gene networks, and its available plug-ins increase its capabilities to perform a dynamic range of functions. (36). STRING is a protein database that collects cores and integrates all the information on protein-protein interactions available from public resources, and also provides computational predictions about the stored information on the proteins. It provides information about both functional and physical interactions of the proteins. The latest version of STRING (11.0) covers PPI information of 5090 organisms, which is almost double the number it used to cover previously. STRING constructs networks on the basis of analysis of co-expression, text-mining of biological and biomedical literature, experimentally available data, databases of curated pathways and knowledge of protein-complex and evidence on co-occurrence of evolutionary and genomic signals across different organisms which happens due to presence of Orthologs (35).

Advantage of using StringAPP in Cytoscape over STRING database is that web interface of STRING database does not allow the creation of large networks while on the other hand, Cytoscape allows to generate large networks. Moreover, Cytoscape also provides other tools for the import, visualization and analysis of the additional data (38).

Identification of Hub Genes

Hub genes are the genes which have high correlation/connectivity in the network and plays central role in defining the particular function of that specific network or module (39).

Hub genes were identified through degree of the nodes by utilizing cytoHubba plugin of the Cytoscape. It helps to determine various hub genes using available network through Degree, Maximal Calique Centrality (MCC), Maximum Neighborhood Component (MNC) and some other topological analysis methods (40) It provides an interface for the analysis of the desired network. First, it calculates the score of each

node based on the 11 methods. Then the network consisting of desired number of top-ranked nodes based on method are retrieved in output console of Cytoscape (40)..

Results

Analysis of Datasets and Selection of DEG

Analyzed dataset of Pancreatic cancer with GSE IDs GSE1547, GSE16515, GSE62165 and GSE62452 contained 12358, 2482, 12382 and 4639 differentially expressed genes (DEGs) based on defined threshold of adjusted p-value <0.001.

Identification of Common DEG between Datasets

314 down-regulated and 614 up-regulated genes were found common among 4 datasets, following the cut-off criteria of adjusted p-value<0. 001.Se figures I and II

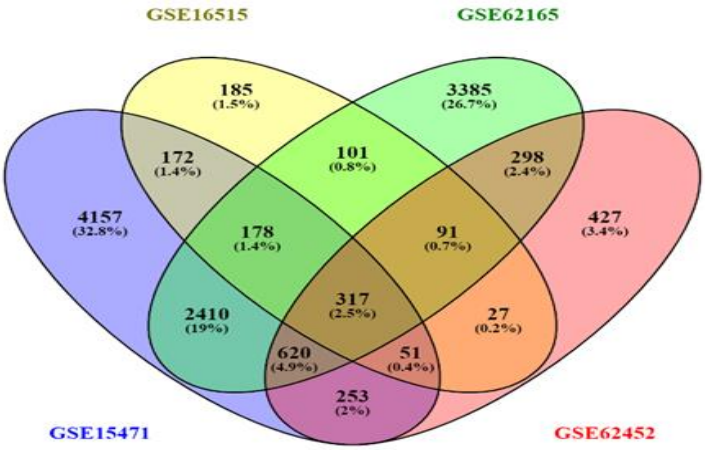

Fig 1: Common downregulated genes

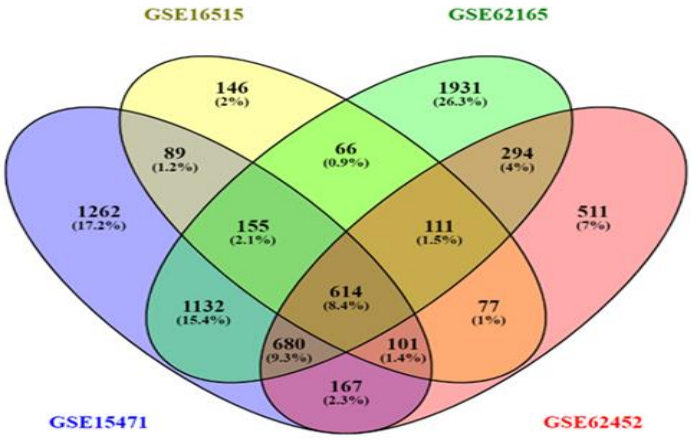

Fig 2: Common up-regulated genes

Functional Annotation

GO Terms Enrichment Analysis

A total of 317 common down regulated genes were found enriched in 77 GO terms of biological processes (BP) category. 4 terms which were selected by following cutoff mechanism of counts >10 and p-value<0.05 are shown in Figure III. 614 common up-regulated genes were found enriched in 269 GO terms of BP category. 41 terms were selected by following cutoff mechanism of counts >10 p-value<0.05 and 10/4 are plotted and shown in Figure IV. X-axis is representing the names of GO terms while counts and p-values are represented on the right side of the y-axis respectively. Size of every dot represents the number of genes enriched in each term. Color of the dots ranging from red to blue indicates p-value from smaller to larger

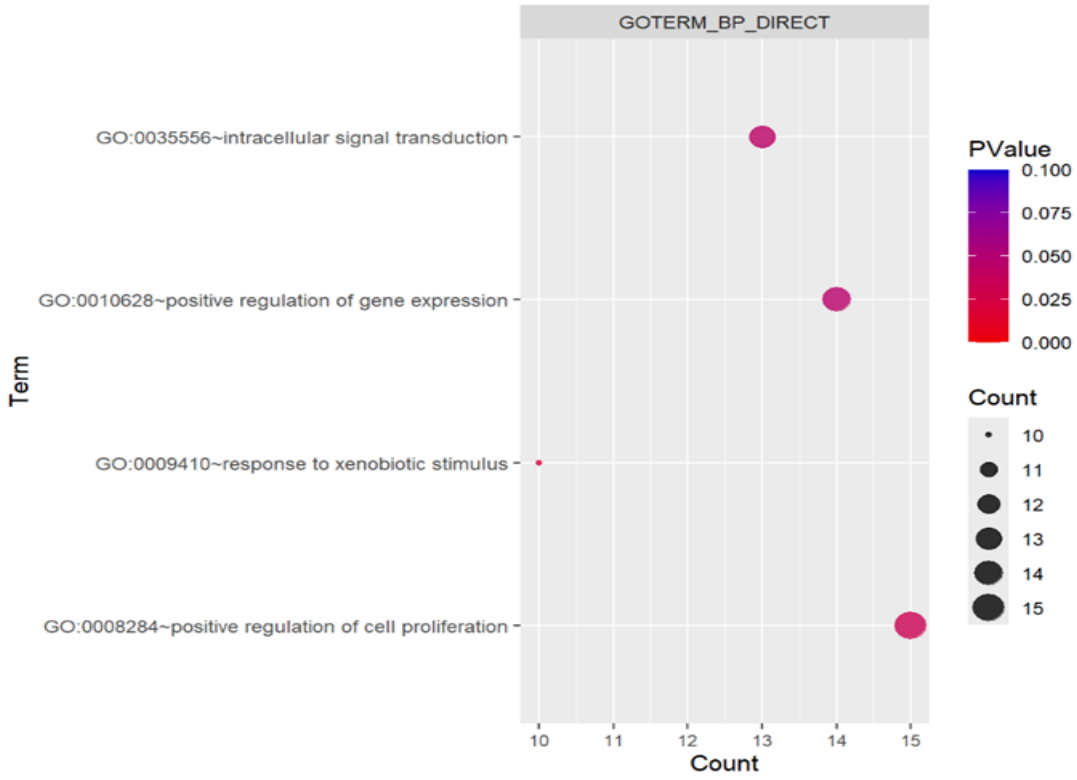

Fig 3: Common downregulated DEGs enriched in GO terms of BP category

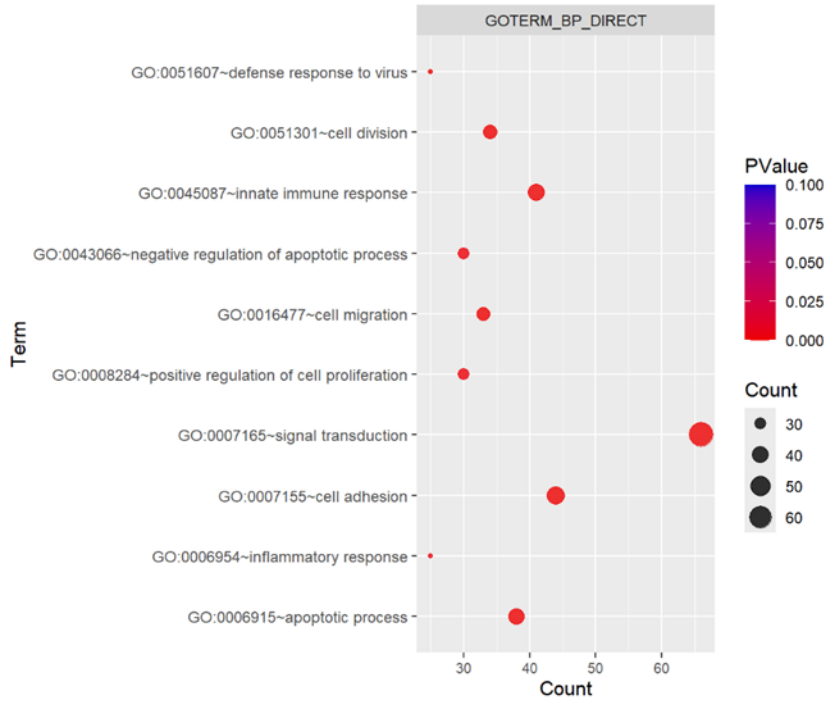

Fig 4: Common upregulated DEGs enriched in GO terms of BP category

A total of 317 common down-regulated genes were found enriched in 24 GO terms of the Cellular Components (CC) category. 10/11 terms were selected by following the cutoff mechanism of counts >10 and p-value <0.05, as shown in Figure V. A total of 614 common up-regulated genes were found enriched in 127 GO terms of the CC category. 10/54 terms were selected by following the cutoff mechanism of counts >10 and p-value<0.05, as shown in Figure VI. The X-axis represents the names of GO terms, while counts and p-values are represented on the right side of the y-axis, respectively. The size of each dot represents the number of genes enriched in each term. The color of the dots, ranging from red to blue, indicates the p-value from smallest to largest.

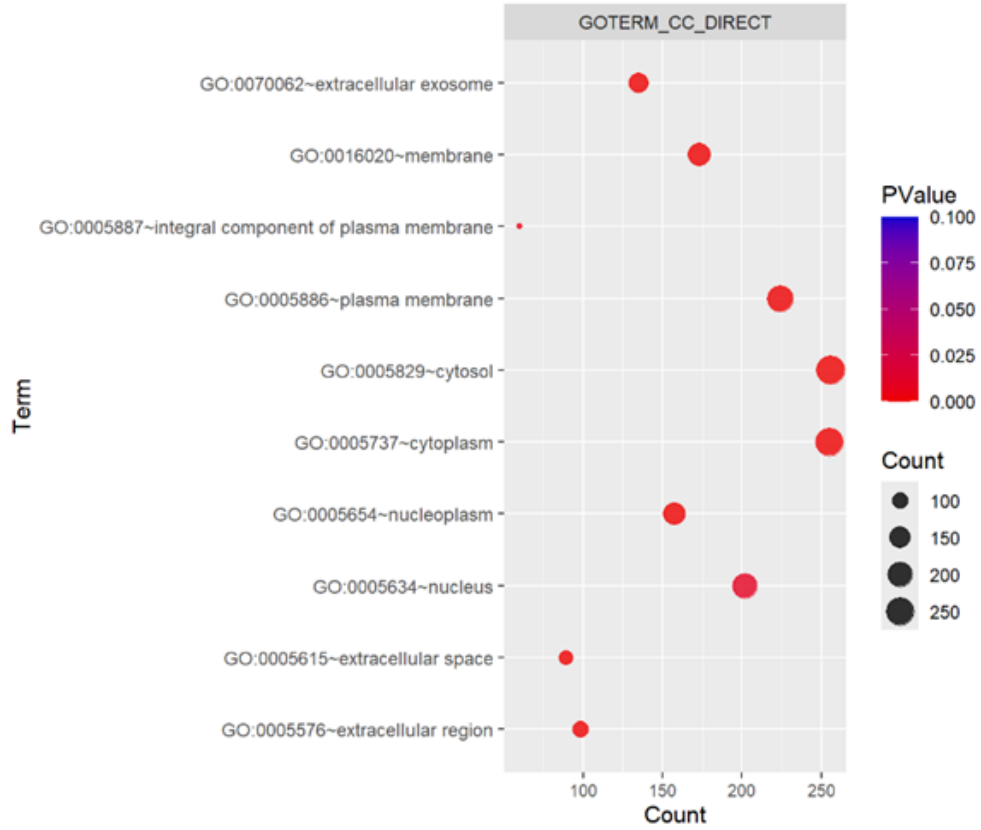

Fig 5: Common Downregulated DEGs enriched in GO terms of CC category

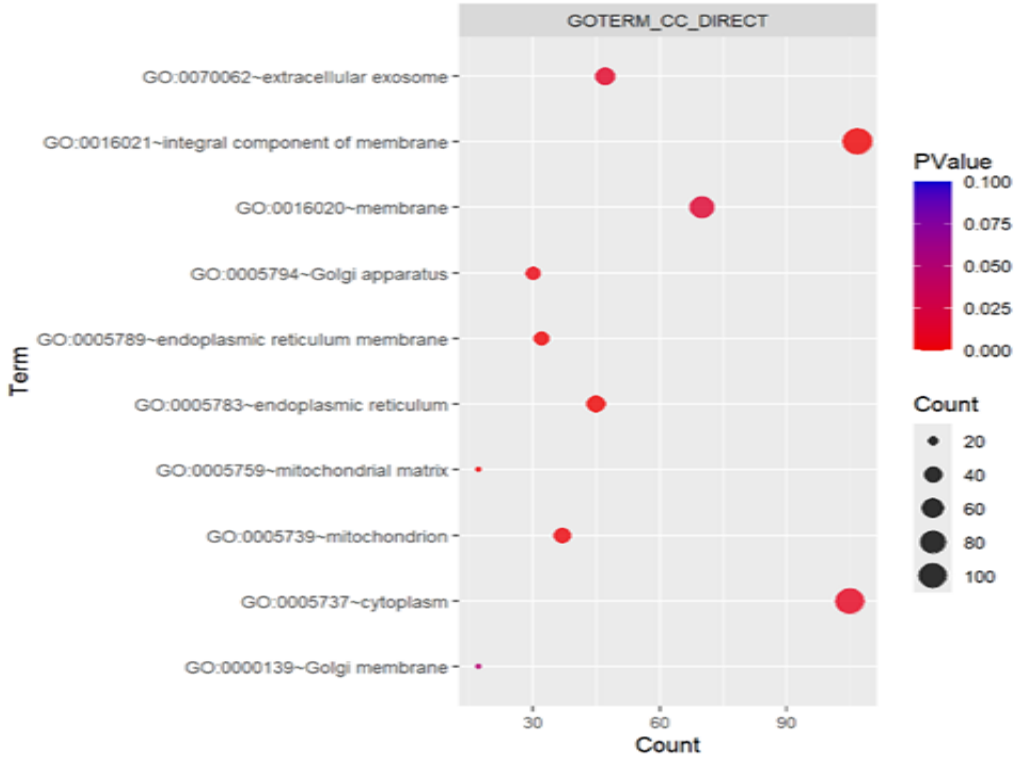

Fig 6: Common up-regulated DEGs enriched in GO terms of CC category

A total of 317 common down-coordinated genes were found enriched in 31 GO terms of the Molecular function (MF) category. 6 terms were selected by following the cutoff mechanism of counts >10 and p-value <0.05, as shown in Figure VII. A total of 614 common up-regulated genes were found enriched in 75 Gene Ontology terms of the MF category. 10/26 terms were selected by following the cutoff mechanism of counts >10 and p-value<0.05, as shown in Figure VIII. The X-axis represents the names of GO terms, while counts and p-values are represented on the right side of the y-axis, respectively. The size of every dot represents the number of genes enriched in each term. The color of the dots, ranging from red to blue, indicates the p-value from smallest to largest.

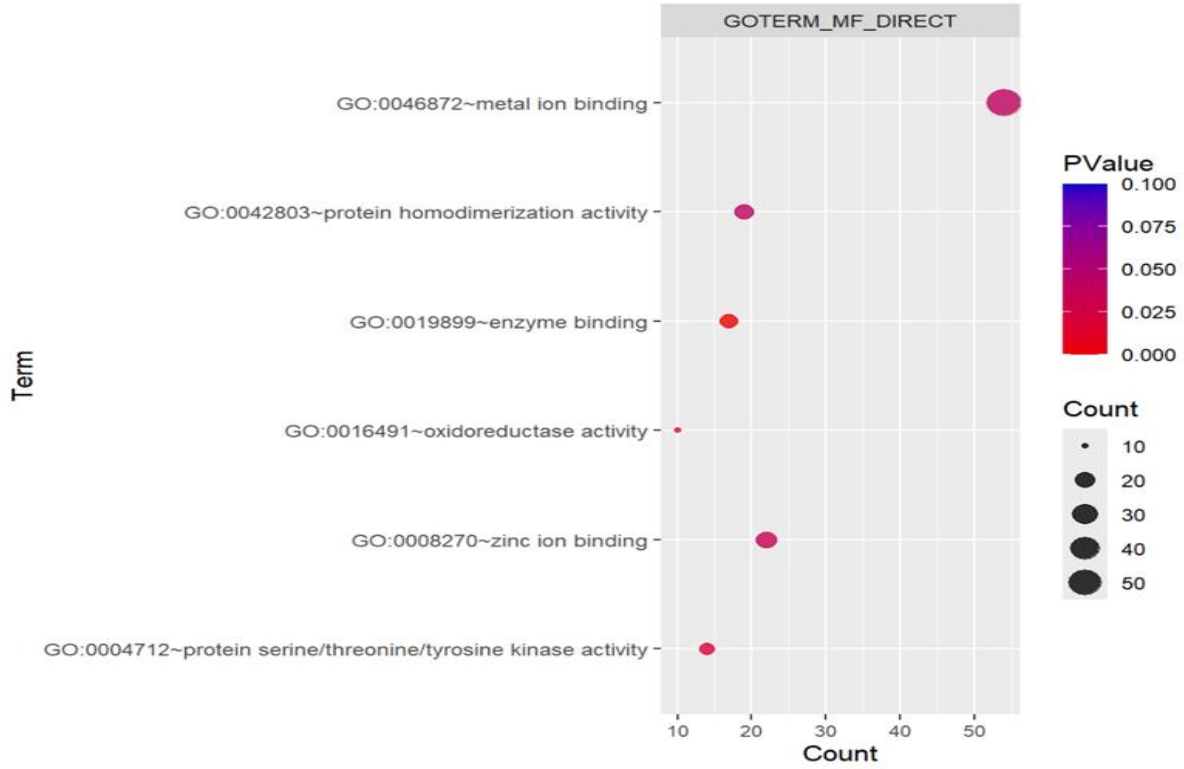

Fig 7: Common Down-regulated DEGs enriched in GO terms of MF category

Fig 8: Common up-regulated DEGs enriched in GO terms of MF category

A total of 317 were found enriched in 43 KEGG pathways. Four pathways were selected based on the cutoff mechanism, where counts >10 and p-value <0.05, as shown in Figure IX. Six hundred fourteen common upregulated genes were found enriched in 46 GO terms of the MF category. 10/35 pathways, which were selected by following the cutoff mechanism of counts >10 and p-value<0.05, are shown in Figure X. The X-axis represents the names of GO terms. In contrast, counts and p-values are represented on the right side of the y-axis, respectively. The size of each dot represents the number of genes enriched in each term. The color of the dots, ranging from red to blue, indicates the p-value from smallest to largest.

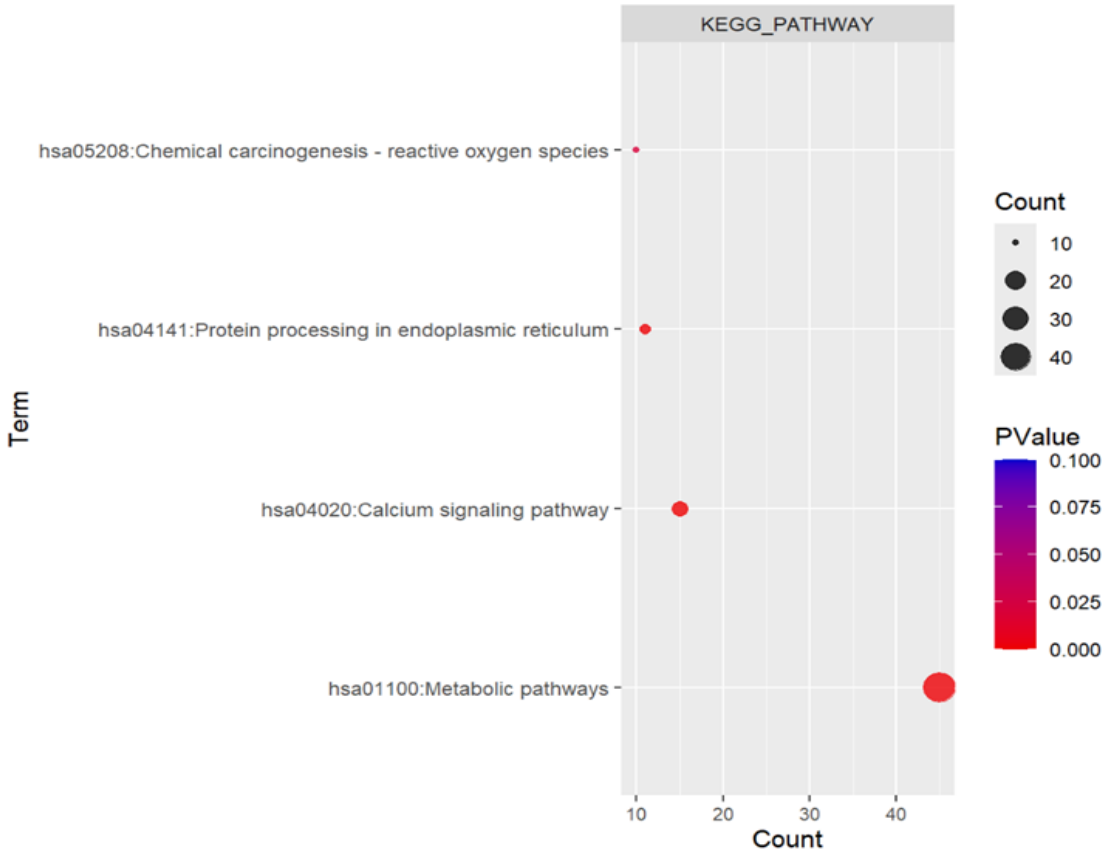

Fig 9: Common downregulated DEGs that are KEGG pathway enriched

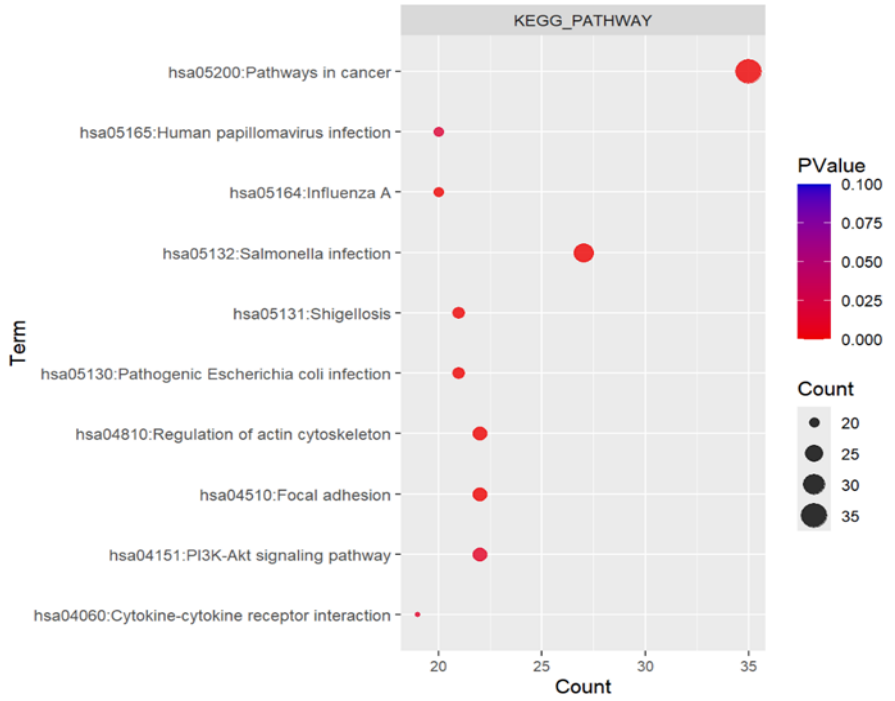

Fig 10: Common upregulated DEGs enriched in KEGG pathways

Gene interaction network construction

The initial interaction of the gene network of 317 down-regulated genes, which were found common among four datasets provided to stringAPP, contained 314 genes and 450 edges under a confidence score of ≥0.4. After removal of noninteracting genes, the final network consisted of 202 genes with 433 edges and is shown in Figure XI. The initial gene interaction network of the top 500 upregulated genes, which were found common among four datasets provided to StringApp, contained 500 genes and 3500 edges under a confidence score of ≥0.4. After removal of non-interacting genes, the final network consisted of 432 genes with 3470 edges and is shown in Figure XII.

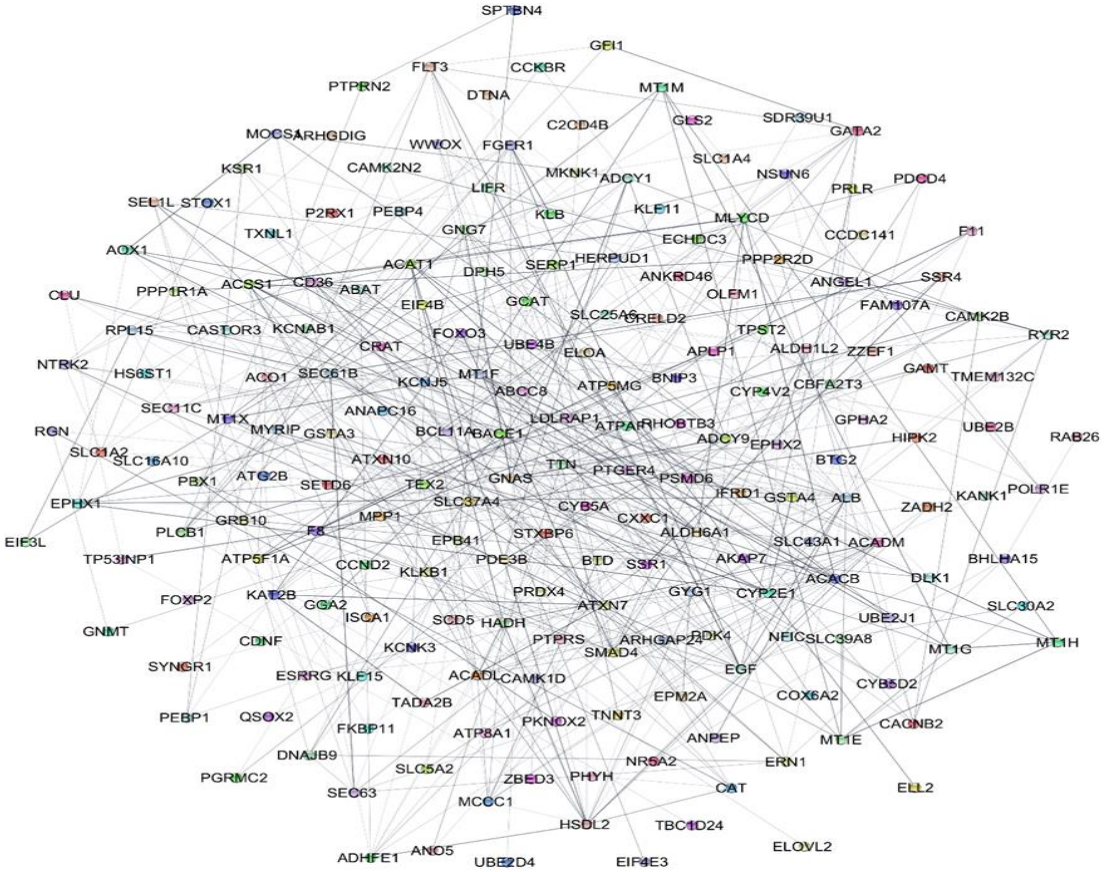

Fig 11: Gene interaction network of common downregulated DEGs among four datasets of pancreatic cancer

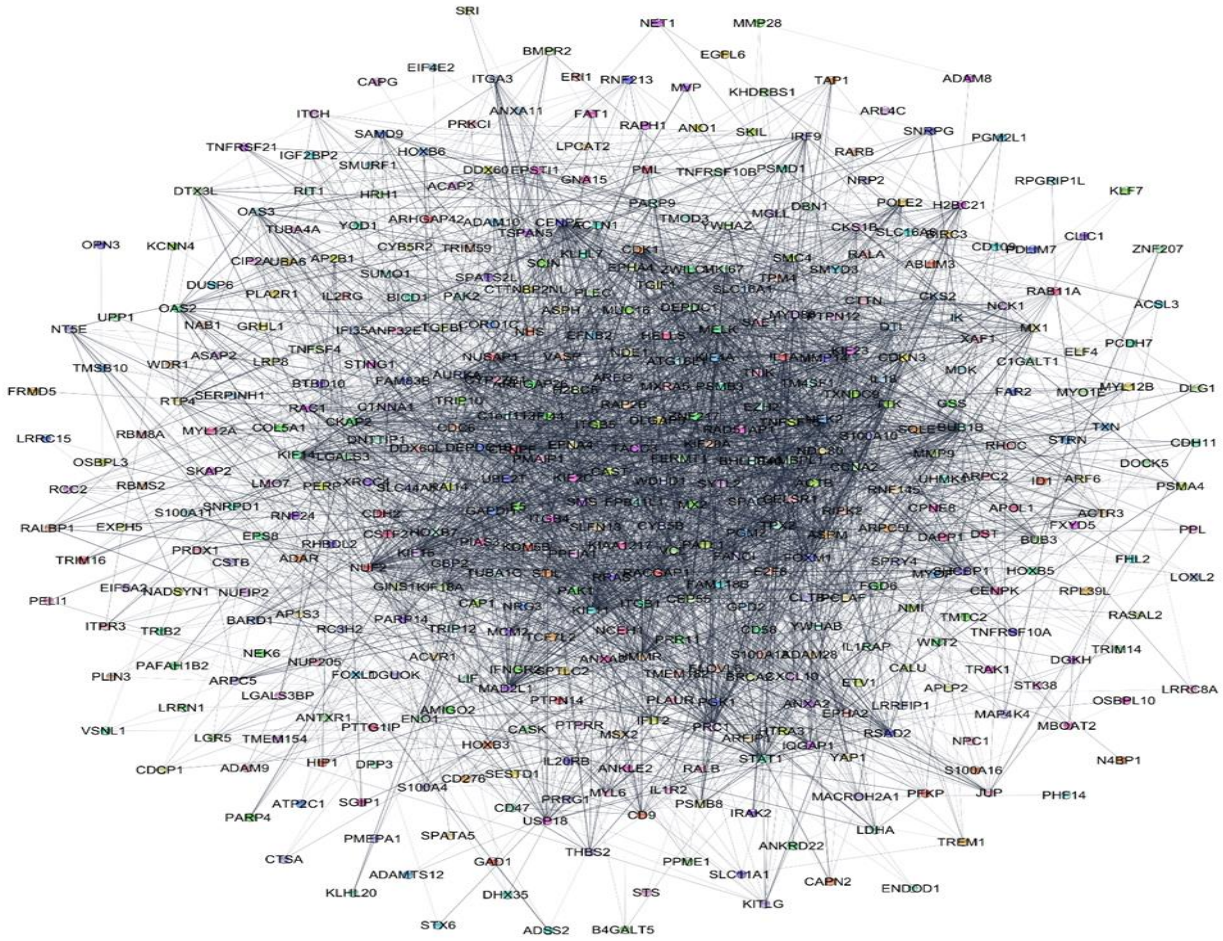

Fig. 12: Gene interaction network of common upregulated DEGs among four datasets of pancreatic cancer

Identification of Hub Genes

These hub genes were filtered from the interaction of gene networks of common down and up-regulated genes based on the degree, as shown in Figure XII and XIV.

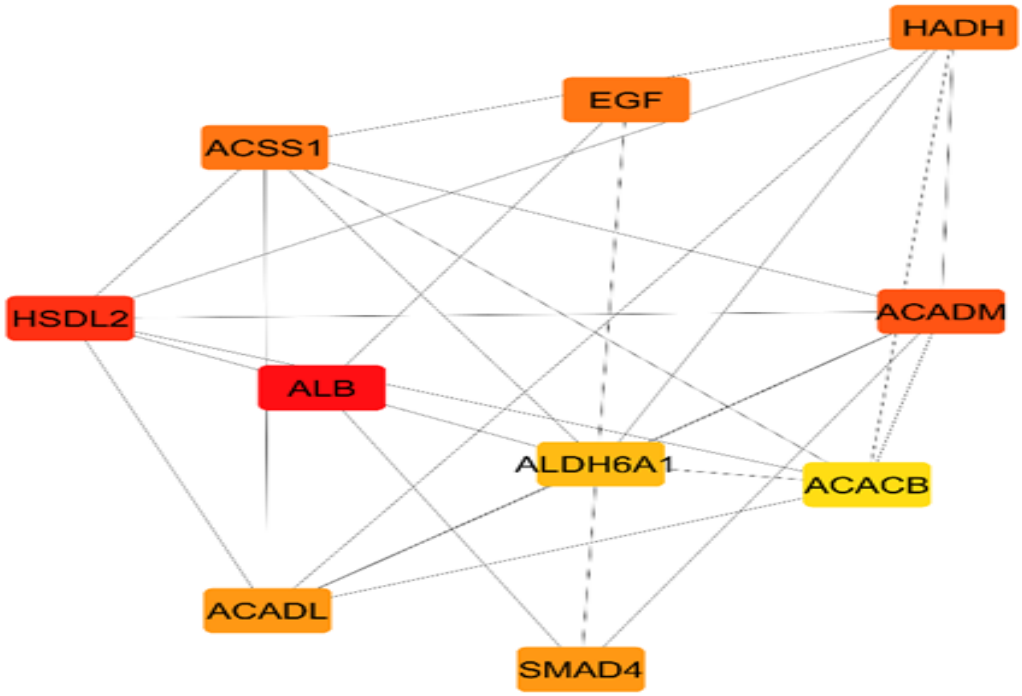

Fig. 13: Top 10 hub genes found in the interaction of the gene network of common down-regulated genes

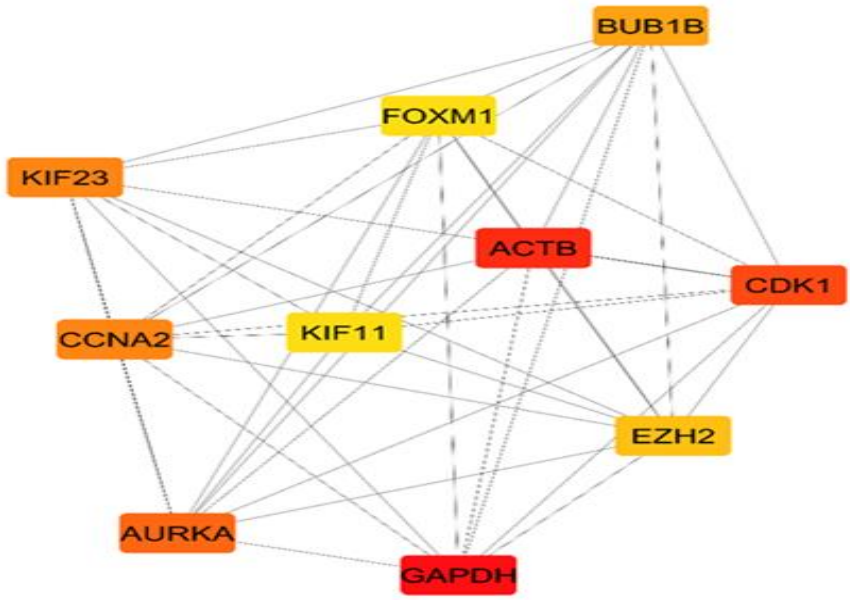

Fig. 14: Top 10 hub genes found in the gene interaction network of common up-regulated genes

The top 10 downregulated hub genes were ACACB, ALB, HADH, ALDH6A1, EGF, ACSS1, SMAD4, ACADM, ACAD, and HSDL2. And top 10 upregulated genes are following EZH2, ACTB, GAPDH, KIF23, AURKA, KIF11, CDK1, CCNA2, FOXM1 and BUB1B.As shown in Table I.

Table 1: downregulated and upregulated gene names and their network ID

Downregulated STRING Network ID	Name	Upregulated-STRING Network ID	Name
stringdb:9606.ENSP00000341044	ACACB	stringdb:9606.ENSP00000320147	EZH2
stringdb:9606.ENSP00000295897	ALB	stringdb:9606.ENSP00000494750	ACTB
stringdb:9606.ENSP00000474560	HADH	stringdb:9606.ENSP00000380070	GAPDH
stringdb:9606.ENSP00000450436	ALDH6A1	stringdb:9606.ENSP00000260363	KIF23
stringdb:9606.ENSP00000265171	EGF	stringdb:9606.ENSP00000216911	AURKA
stringdb:9606.ENSP00000316924	ACSS1	stringdb:9606.ENSP00000260731	KIF11
stringdb:9606.ENSP00000341551	SMAD4	stringdb:9606.ENSP00000378699	CDK1
stringdb:9606.ENSP00000359871	ACADM	stringdb:9606.ENSP00000481380	CCNA2
stringdb:9606.ENSP00000233710	ACADL	stringdb:9606.ENSP00000342307	FOXM1
stringdb:9606.ENSP00000381785	HSDL2	stringdb:9606.ENSP00000287598	BUB1B

Discussion

The GSE15471, GSE16515, GSE62165, and GSE62452 datasets were chosen to ensure a thorough examination of the gene expression profiles associated with pancreatic cancer. By integrating multiple pancreatic cancer transcriptome microarray datasets, we believe we were able to identify alterations at the molecular level with greater accuracy than studies based on a single dataset. This resulted in modifications of gene expression modules in pancreatic adenocarcinoma at the genome-wide scale. (41)The availability of DNA microarray datasets has sparked new machine learning and bioinformatics research directions. This kind of data is gathered from tissue and cell samples in order to identify variations in gene expression that may help diagnose diseases or differentiate between different kinds of tumors. Due to the numerous properties and small sample sizes of microarray data, machine learning researchers face a challenging problem in classifying the data. (42)Our analysis's findings also demonstrated that DEG was substantially concentrated in biological processes such as extracellular matrix deconstruction, cell migration, adhesion, and cell-adhesion (43) . Cell adhesion was the primary enriched process by genome-scale analysis in patients with pancreatic adenocarcinoma, according to a study on gene ontology analysis of the DEG.(41)

The microarray datasets were normalized using the GEO2R online tool. Software tools created in the R statistical language are widely used in research bioinformatics, including the tools in the Bioconductor repository. (44). For data analysis, GEO2R makes use of Bioconductor packages like limma and GEOQuery. The t-test is used for differential expression, and the Benjamini and Hochberg FDR is used for multiple-testing correction. (45). Venn diagram application tools are tools for examining the correlations between biological data and visualizing them in a Venn diagram. Additionally, we evaluate the picture beautification characteristics of the Venn diagram generators in terms of the graphical layout and give a brief overview of common genes. (46).Visualizing the overlap of DEGs was made possible by Venny's capacity to compare up to four lists and produce results in interactive Venn diagrams. This research identified 314 common down-regulated and 614 common up-regulated genes.

The DAVID Resources provide tools for functionalities that enable us to dynamically view genes from their lists on bio-pathways, condense large gene lists into gene functional groups, and convert between gene/protein identifiers, in addition to the standard gene-term enrichment analysis for any uploaded gene list. DAVID (http://david.niaid.nih.gov) gives a more profound understanding of the biological mechanisms associated with large gene lists (47).

The Database for Annotation, Visualization and Integrated Discovery was utilized for functionally annotating common DEGs. DAVID offers an extensive collection of functional annotation tools that enable the examination of KEGG pathways and GO concepts for enrichment. The BP category had a sizable number of genes that were enriched, including both down-regulated (77 terms) and up-regulated (269 terms). Prominent representation of important systems like signal transduction and metabolic activities suggests their crucial roles in the pathophysiology of pancreatic cancer. Significant enrichment in CC terms was observed for both down-regulated (24 terms) and up-regulated (127 terms) genes. These highlighted changes in cellular architecture and function in cancer cells included essential cellular components, such as the ribosome and Golgi apparatus. With 75 terms for up-regulated genes as well as 31 terms for down-regulated genes, the MF category likewise demonstrated a notable enrichment. Significant alterations were seen in functions like ATP-binding protein-binding, indicating disturbances in cellular energy metabolism and protein interactions. (48).

As previous studies have shown, DEGs are primarily associated with the reaction to testosterone and ketones, competitive promoter binding-mediated negative regulation of transcription, hypotonic response, and cellular stress response, as demonstrated by GO-BP analysis. In terms of CC terminology, the basal plasma membrane, microvilli, apical plasma membrane, apical portion of the cell, and actin-based cell projection were the primary areas of involvement for the DEGs. DEGs were primarily associated with oxidoreductase activity, which acts on NAD(P)H, heme protein as an acceptor, water channel activity, carbonate dehydratase activity, water transmembrane transporter activity, and NADH dehydrogenase (quinone) activity, according to MF terminology.(49)

An examination of KEGG pathways shed light on the molecular pathways that pancreatic cancer disrupts. Enriched in 46 pathways were common upregulated genes, and 43 pathways had common downregulated genes. This demonstrates how pancreatic cancer involves a significant rewiring of cellular circuits, including important signaling and metabolic pathways. The DEG was significantly abundant in the extracellular matrix (ECM)-receptor interaction, focal adhesion, mucin-type O-glycan production, proteoglycans in cancer, and phosphoinositide 3-kinase-protein kinase B (PI3K-Akt) signaling pathway, according to KEGG signaling pathway analysis (43).

Gene networks were built using STRING and examined in Cytoscape to comprehend intricate relationships among the DEGs. The cytoHubba plugin in Cytoscape is used to identify hub genes, which are essential to network operation. High connectivity genes, or hub genes, are frequently essential to the stability and proper operation of biological networks. Since these genes may be targets for therapeutic therapies, identifying them is essential. (50) . Gene network of 317 down-regulated genes which were found common among four datasets provided to stringAPP contained 314 genes and 450 edges and gene interaction network of top 614 upregulated genes which were found common among four datasets provided to stringAPP contained 500 genes and 3500 edges, Both networks provide valuable insights into the molecular interactions underpinning the distinct gene expression patterns observed in pancreatic cancer, with downregulated genes potentially indicating suppressed pathways, and upregulated genes highlighting activated oncogenic pathways. These findings could contribute to the identification of potential therapeutic targets and the development of more effective treatment strategies for pancreatic cancer, as previous studies show that PDACs commonly harbor activating mutations in the KRAS oncogene, which is a potent driver of tumor initiation and maintenance. Inactivating mutations in tumor suppressor genes such as *CDKN2A/p16*, *TP53*, and *SMAD4* cooperate with *KRAS* mutations to cause aggressive PDAC tumor growth (51).

The downregulated hub genes ACACA, which are involved in the initial stage of fatty acid production and metabolism, and acetyl-CoA carboxylase (ACC) are a crucial and rate-limiting switch that catalyzes the conversion of acetyl-CoA to malonyl-CoA, a process crucial for the emergence and growth of tumors. Acetyl-CoA carboxylase alpha (ACACA) and acetyl-CoA carboxylase beta (ACACB) are the two subtypes of ACC (52). According to the current study, the CRP/Alb ratio may be a valuable and encouraging inflammatory prognostic score for pancreatic cancer. With a cutoff value of 0.180, an increased CRP/Alb ratio is an independent predictor of a poor prognosis (53). Depending on the location of the tumor, HADH can either be a promoter or a suppressor, and this is directly related to prognostic evaluation. In known and similar cancers, the prognostic roles of HADHA and HADHB are similar (54). Among the 19 identified ALDH genes examined using microarray technology, ALDH1A1, ALDH1L2, and ALDH6A1 demonstrated markedly elevated expression at 20 weeks of fetal age in human fetal pancreas, in contrast to 8–10 weeks (54). A transmembrane glycoprotein called epidermal growth factor receptor (EGFR) is expressed in normal human tissues. Encoded by proto-oncogenes, it belongs to the tyrosine kinase family of growth factor receptors. Research has indicated that pancreatic cancer exhibits an overexpression of EGFR. The presence of metastases, poor survival, and more advanced disease are all correlated with over-expression. Thus, blocking the EGFR signaling pathway is a desirable therapeutic goal (55). Transforming growth factor-β (TGF-β) controls cellular processes and plays a significant part in the emergence of pancreatic cancer. As a member of the Smads family of TGF-β signal transducers, SMAD4 promotes both apoptosis and the proliferation of pancreatic cells. It is particularly inactivated in 50% of advanced pancreatic tumors (56).Exo-ACADM might be a valuable biomarker for determining how well patients will respond to chemotherapy (57).In PCs, HSDL2 was strongly expressed and associated with a worse overall survival rate. Cell growth was greatly hindered, and lipid metabolism was further impeded when HSDL2 was silenced. Elevated HSDL2 expression is linked to the advancement of PC and may represent a novel biomarker for a bad prognosis in the future, as well as a target for treatment (58).

The top upregulated hub genes are EZH2, whose activity was associated with PDAC dedifferentiation and tumor growth in PDAC models and human PDAC tissues. A mechanistic basis for EZH2-dependent PDAC progression was revealed by combining RNA- and chromatin immune precipitation sequencing studies, which identified EZH2 as a pivotal suppressor of differentiation programs in PDAC and revealed EZH2-dependent transcriptional repression of the classical subtype-defining transcription factor Gata6 (59)Most tumor cells and tissues are reported to have elevated levels of ACTB. Cancer metastasis and invasiveness have been linked to aberrant ACTB expression and polymerization, as well as ensuing cytoskeleton alterations.(60).Inhibiting the nuclear GAPDH-mediated mechanisms of cell death and promoting glycolysis in cancerous cells is the result of mutant p53's inability to stop the glycolytic enzyme Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) from translocating into the nucleus and maintaining its cytoplasmic position (61). A poor prognosis is linked to increased KIF23 expression, and KIF23 may be a possible target for therapy in pancreatic ductal adenocarcinoma. (62)Since AURKA is a direct target of the MAPK pathway, hyperactivation of the circuit—at least through ETS2—is what causes its overexpression in pancreatic cancer (63)Based on the results of the protein-based immunohistochemistry, KIF11 and KIF14 may be characterized as good prognostic indicators; however, the transcriptome data linked them to a poor prognosis (64)The overexpression of CDK1 mainly causes tumors through two methods. By promoting cell cycle progression at the G2/M phase checkpoint, cells undergo carcinogenic mutations that facilitate proliferation. Secondly, this process causes PDAC cells to acquire pluripotent properties, resulting in the formation of CSCs. These CSCs increase the heterogeneity, development, and initiation of tumors, making treatment more difficult. (65). As a member of the highly conserved cyclin family, CCNA2 is up-regulated in numerous cancer types, suggesting that it may play a role in the development and spread of cancer. High CCNA2 expression has been shown to stimulate hepatoma cell proliferation and may help track the effectiveness of chemotherapy for breast cancer. Zhou et al.'s bioinformatics research showed a strong correlation between the advancement of PDAC and CCNA2 overexpression. (66). Studies that

used immunohistochemical labeling with a particular anti-FOXM1 antibody to identify the expression of FOXM1 protein in a TMA of the primary pancreatic tumor and adjacent standard pancreatic tissue specimens revealed the overexpression of FOXM1 protein in pancreatic tumors and its clinical pathologic importance. (67).

In summary, this study's integration of several datasets and bioinformatics techniques yielded insightful knowledge on the molecular processes causing pancreatic cancer. Common DEGs, functional annotations, pathway enrichment, and hub gene discoveries provide a basis for additional experimental confirmation and possible therapeutic uses. These results may aid in the creation of innovative treatment and diagnostic approaches meant to enhance pancreatic cancer prognoses.

Conclusion

A thorough grasp of the common differentially expressed genes among various pancreatic cancer datasets is provided by this approach. The genesis and progression of pancreatic cancer may be significantly influenced by key hub genes and biological pathways that were identified by functional enrichment and network analysis. These findings provide insightful information for future investigations, possibly leading to the discovery of novel biomarkers and treatment targets for pancreatic cancer.

Declarations

Data Availability statement

All data generated or analysed during the study are included in the manuscript.

Ethics approval and consent to participate

Approved by the department concerned. (IRBEC-24)

Consent for publication

Approved

Funding

Not applicable

Conflict of interest

The authors declared the absence of a conflict of interest.

Author Contribution

AN, AG, SZ, AZ, HA,

Review of Literature, Data entry, Data analysis, Manuscript drafting and Development of Research Methodology Design

SM, AA

Study Design, manuscript review, critical input. Conception of Study.

All authors reviewed the results and approved the final version of the manuscript. They are also accountable for the integrity of the study.

References

1. Moore A, Donahue T. Pancreatic cancer. Jama. 2019;322(14):1426-.
2. Kukurba KR, Montgomery SB. RNA sequencing and analysis. Cold Spring Harbor Protocols. 2015;2015(11):pdb. top084970.
3. Pantlik J. Detection of somatic mutations and mutational signatures in pancreatic cancer using RNA-sequencing data: University of Zagreb. Faculty of Science. Department of Biology; 2023.
4. Singhi AD, Koay EJ, Chari ST, Maitra A. Early detection of pancreatic cancer: opportunities and challenges. Gastroenterology. 2019;156(7):2024-40.
5. Li D, Xie K, Wolff R, Abbruzzese JL. Pancreatic cancer. The Lancet. 2004;363(9414):1049-57.
6. Wei R, Qi G, Zeng Z, Shen N, Wang Z, Shen H, et al. IMUP and GPRC5A: two newly identified risk score indicators in pancreatic ductal adenocarcinoma. Cancer cell international. 2021;21:1-19.
7. Rong Y, Gao J, Kuang T, Chen J, Li Ja, Huang Y, et al. DIAPH3 promotes pancreatic cancer progression by activating selenoprotein TrxR1-mediated antioxidant effects. Journal of Cellular and Molecular Medicine. 2021;25(4):2163-75.
8. Zhang Mx, Wang H, Sun Gp. Tumor-suppressor Fbxw7 targets SIK2 for degradation to interfere with TORC2-AKT signaling in pancreatic cancer. Cell Biology International. 2020;44(9):1900-10.
9. Mizrahi JD, Surana R, Valle JW, Shroff RT. Pancreatic cancer. The Lancet. 2020;395(10242):2008-20.
10. Kleeff J, Korc M, Apte M, La Vecchia C, Johnson CD, Biankin AV, et al. Pancreatic cancer. Nature reviews Disease primers. 2016;2(1):1-22.
11. Adamska A, Domenichini A, Falasca M. Pancreatic ductal adenocarcinoma: current and evolving therapies. International journal of molecular sciences. 2017;18(7):1338.
12. Hidalgo M. Pancreatic cancer. New England Journal of Medicine. 2010;362(17):1605-17.
13. Fitzpatrick MJ, Ben-Shahar Y, Smid HM, Vet LE, Robinson GE, Sokolowski MB. Candidate genes for behavioural ecology. Trends in Ecology & Evolution. 2005;20(2):96-104.
14. Forrest AR, Carninci P. Whole genome transcriptome analysis. RNA biology. 2009;6(2):107-12.
15. Flint J, Munafò MR. Candidate and non-candidate genes in behavior genetics. Current opinion in neurobiology. 2013;23(1):57-61.
16. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nature Reviews Genetics. 2002;3(5):391-7.
17. Bargsten JW, Nap J-P, Sanchez-Perez GF, van Dijk AD. Prioritization of candidate genes in QTL regions based on associations between traits and biological processes. BMC plant biology. 2014;14:1-12.
18. Clough E, Barrett T. The gene expression omnibus database. Statistical Genomics: Methods and Protocols. 2016:93-110.
19. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009;4(1):44-57.
20. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic acids research. 2021;49(D1):D605-D12.
21. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207-10.
22. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Research. 2012;41(D1):D991-D5.
23. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome biology. 2004;5(10):R80.
24. Smyth G. LIMMA: Linear models for microarray data. 2005. p. 397-420.
25. Davis S, Meltzer P. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics (Oxford, England). 2007;23:1846-7.
26. Kalpić D, Hlupić N, Lovrić M. Student's t-Tests. In: Lovric M, editor. International Encyclopedia of Statistical Science. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 1559-63.
27. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.
28. Anjum A, Jaggi S, Varghese E, Lall S, Bhowmik A, Rai A. Identification of Differentially Expressed Genes in RNA-seq Data of Arabidopsis thaliana: A Compound Distribution Approach. J Comput Biol. 2016;23(4):239-47.
29. Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. 2007.
30. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 2003;4(5):P3.
31. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009;4(1):44-57.
32. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, et al. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 2004;32(Database issue):D258-61.
33. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-9.

34. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.
35. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research. 2019;47(D1):D607-D13.
36. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.
37. Mike Smoot MA, Yassen Assenov. NetworkAnalyzer 2020 [Available from: https://apps.cytoscape.org/apps/networkanalyzer#:~:text=NetworkAnalyzer%20performs%20analysis%20of%20biological,of%20connected%20pairs%20of%20nodes.
38. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. J Proteome Res. 2019;18(2):623-32.
39. Liu Y, Gu HY, Zhu J, Niu YM, Zhang C, Guo GL. Identification of Hub Genes and Key Pathways Associated With Bipolar Disorder Based on Weighted Gene Co-expression Network Analysis. Front Physiol. 2019;10:1081.
40. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8 Suppl 4:S11.
41. Zhu T, Gao Y-F, Chen Y-X, Wang Z-B, Yin J-Y, Mao X-Y, et al. Genome-scale analysis identifies GJB2 and ERO1LB as prognosis markers in patients with pancreatic cancer. Oncotarget. 2017;8(13):21281.
42. Alonso-Betanzos A, Bolón-Canedo V, Morán-Fernández L, Sánchez-Maroño N. A review of microarray datasets: where to find them and specific characteristics. Microarray Bioinformatics. 2019:65-85.
43. Lu W, Li N, Liao F. Identification of key genes and pathways in pancreatic cancer gene expression profile by integrative analysis. Genes. 2019;10(8):612.
44. Sepulveda JL. Using R and bioconductor in clinical genomics and transcriptomics. The Journal of Molecular Diagnostics. 2020;22(1):3-20.
45. Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. http://bioinfogp cnb csic es/tools/venny/index html. 2007.
46. Jia A, Xu L, Wang Y. Venn diagrams in bioinformatics. Briefings in bioinformatics. 2021;22(5):bbab108.
47. Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, et al. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic acids research. 2007;35(suppl_2):W169-W75.
48. Stormo GD, Hartzell 3rd G. Identifying protein-binding sites from unaligned DNA fragments. Proceedings of the National Academy of Sciences. 1989;86(4):1183-7.
49. Zhang X, Fu Q, Liu C, Luo Q, Yu P, Bian J, et al. Identification of the Genetic Association Between Ferroptosis and immune in Pancreatic Cancer. 2024.
50. Chen S-H, Chin C-H, Wu H-H, Ho C-W, Ko M-T, Lin C-Y, editors. cyto-Hubba: A Cytoscape plug-in for hub object analysis in network biology. 20th international conference on genome informatics; 2009: Citeseer.
51. Grant TJ, Hua K, Singh A. Molecular pathogenesis of pancreatic cancer. Progress in molecular biology and translational science. 2016;144:241-75.
52. Hong H-J, Shao Y, Zhang S, Yang G, Jia H, Yang X, et al. ACACB is a novel metabolism-related biomarker in the prediction of response to cetuximab therapy inmetastatic colorectal cancer: ACACB is a biomarker for CTX therapy in metastatic CRC. Acta Biochimica et Biophysica Sinica. 2022;54(11):1671.
53. Liu Z, Jin K, Guo M, Long J, Liu L, Liu C, et al. Prognostic value of the CRP/Alb ratio, a novel inflammation-based score in pancreatic cancer. Annals of surgical oncology. 2017;24:561-8.
54. Wang X, Song H, Liang J, Jia Y, Zhang Y. Abnormal expression of HADH, an enzyme of fatty acid oxidation, affects tumor development and prognosis. Molecular medicine reports. 2022;26(6):1-8.
55. Oliveira-Cunha M, Newman WG, Siriwardena AK. Epidermal growth factor receptor in pancreatic cancer. Cancers. 2011;3(2):1513-26.
56. Xia X, Wu W, Huang C, Cen G, Jiang T, Cao J, et al. SMAD4 and its role in pancreatic cancer. Tumor Biology. 2015;36:111-9.
57. Yang Y, Gu H, Zhang K, Guo Z, Wang X, Wei Q, et al. Exosomal ACADM sensitizes gemcitabine-resistance through modulating fatty acid metabolism and ferroptosis in pancreatic cancer. BMC cancer. 2023;23(1):789.
58. Han A, Xu R, Liu Y, Yin X, Lin Z, Yang W. HSDL2 acts as a promoter in pancreatic cancer by regulating cell proliferation and lipid metabolism. OncoTargets and therapy. 2021:435-44.
59. Patil S, Steuber B, Kopp W, Kari V, Urbach L, Wang X, et al. EZH2 regulates pancreatic cancer subtype identity and tumor progression via transcriptional repression of GATA6. Cancer research. 2020;80(21):4620-32.
60. Guo C, Liu S, Wang J, Sun M-Z, Greenaway FT. ACTB in cancer. Clinica chimica acta. 2013;417:39-44.
61. Butera G, Pacchiana R, Mullappilly N, Margiotta M, Bruno S, Conti P, et al. Mutant p53 prevents GAPDH nuclear translocation in pancreatic cancer cells favoring glycolysis and 2-deoxyglucose sensitivity. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research. 2018;1865(12):1914-23.
62. Gao C-T, Ren J, Yu J, Li S-N, Guo X-F, Zhou Y-Z. KIF23 enhances cell proliferation in pancreatic ductal adenocarcinoma and is a potent therapeutic target. Annals of translational medicine. 2020;8(21).
63. Furukawa T, Kanai N, Shiwaku H, Soga N, Uehara A, Horii A. AURKA is one of the downstream targets of MAPK1/ERK2 in pancreatic cancer. Oncogene. 2006;25(35):4831-9.
64. Klimaszewska-Wiśniewska A, Neska-Długosz I, Buchholz K, Durślewicz J, Grzanka D, Kasperska A, et al. Prognostic significance of KIF11 and KIF14 expression in pancreatic adenocarcinoma. Cancers. 2021;13(12):3017.
65. Wijnen R, Pecoraro C, Carbone D, Fiuji H, Avan A, Peters GJ, et al. Cyclin dependent kinase-1 (CDK-1) inhibition as a novel therapeutic strategy against pancreatic ductal adenocarcinoma (PDAC). Cancers. 2021;13(17):4389.
66. Dong S, Huang F, Zhang H, Chen Q. Overexpression of BUB1B, CCNA2, CDC20, and CDK1 in tumor tissues predicts poor survival in pancreatic ductal adenocarcinoma. Bioscience reports. 2019;39(2):BSR20182306.
67. Kong X, Li L, Li Z, Le X, Huang C, Jia Z, et al. Dysregulated expression of FOXM1 isoforms drives progression of pancreatic cancer. Cancer research. 2013;73(13):3987-96.