Identification of potential biomarkers in Pancreatic Adenocarcinoma of micro-array gene expression data

Abstract: We uncover molecular biomarkers using GSE16515 data set publicly reachable at NIH/NCBI Gene Expression Omnibus database. Using Biobase, GEOquery, gplots packages in R software 3.6. Based on gene expression analysis, we detect 278 differentially expressed genes (DEGs) of up regulation, whereas we find 77 down-regulated gene. The gene ontology of pathway enrichments and KEGG enrichment analyses of DEGs were studied. 120 KEGG pathways associated with pancreatic adenocarcinoma (PAAD) were identified, among which the PI3K/AKT signaling pathway was observed to be significant. Other hub genes discussed in this study, may be used as potential targets for PAAD and related diseases diagnosis and treatment. The following 21 hub genes were detected through NetworkAnalyst on the basis of protein-protein interaction network by the STRING tool: CDK1, CCNB1, CDC20, PPARG, MET, ISG15, LEF1, SFN, DMD, FN1, RUNX2, UBC, TOP2A, ECT2, WNT2, EFNA5, PAK3, PKM, ITGB4, NEK2, and ALB. In the TCGA database, the quantification of expression of PPARG and SFN were examined and showed similarity with the previous results that both of the genes were significantly upregulated in pancreatic adenocarcinoma tumor cells in comparison to normal cells Additionally, the module study of protein-protein interactions revealed that 'amoebiasis', 'protein digestion and absorption', 'focal adhesion', and 'ECM-receptor interaction' had a close association with PAAD. In addition, PI3K/AKT signaling pathway in PAAD was observed to be significant. Other hub genes discussed in the study, might be utilized as promising targets for PAAD and related diseases diagnosis and drug therapies.

Commented [ei1]: adjust this according to pancancer

Keywords: biomarker; differentially expressed genes; pancreatic adenocarcinoma; gene ontology pathway enrichment; cancer

1 Introduction

One of the most deadly cancer is pancreatic ductal adenocarcinoma (PAAD) which has a 5-year overall survival rate as 3% due to the diagnosis at a distant stage (Siegel et al., 2018). There has been

significant improvements in terms of treatments such as pancreatectomy, radiotherapy, adjuvant and neo-adjuvant chemotherapies and palliative care in the previous decades (Gillen et al., 2010). However, pancreatectomy still stays the most efficient treatment, specifically for initial phase pancreatic cancer (Lambert et al., 2019). Thus, an updated knowledge of simple and basic mechanism of pancreatic cancer is necessary for more useful and curable therapies and the advancement of patient survival.

Microarray has become a significant tool in investigating pancreatic cancer genes and target therapeutic drugs. Recent studies suggest an extensive gene expression analysis of PAAD and related diseases by reviewing expressed gene data sets through a comparison between tumor and normal cells (Jones et al., 2008).

Moreover, relative investigation of the different level of expressed genes stays moderately constrained, and a dependable biomarker profile would be a need to develop new gene targets. The protein expression alterations in the advancement and growth of pancreatic adenocarcinoma cancer and related diseases require comprehensive analysis. Furthermore, the relations among the detected DEGs, specifically protein-protein interaction (PPI) networks and underlying signaling pathways should be clarified. Prospect molecular markers from such investigations can later be checked utilizing other methods to be used in early diagnosis of pancreatic cancers (Grønborg et al., 2006).

Pei *et al*., 2009, recently performed microarray experiments to study the differentially expressed levels of FKBP5 gene between the pancreatic tumor and normal samples (data obtainable at NCBI GEO database (Edgar et al., 2002), accession GSE16515)."

The DEGs of pancreatic tumor and normal samples were identified by comparing gene expression by fold change and t-test. Afterwards, the DEGs were filtered using online DAVID tool and the analysis of pathway enrichments (Sherman & Lempicki, 2009). By studying their hub nodes globally and between tumor and normal samples constructing PPI networks, the goal of this project is to study the pathway and genetic mechanisms of PAAD and related diseases growth and to come up with candidate biomarkers for diagnosis, therapeutic targets and predictions.

Earlier studies tackling pancreatic cancer and related diseases underlying biomarkers, due to the implicit evaluation of source and progenitor populations, need to support experimental studies with numerical analysis and statistical methods in addition to previous experiments on mice (Dineen et al., 2010)

Traditional therapeutic alternatives, particularly chemotherapy, are not efficient enough fighting PAAD, and notwithstanding advancements over the last 15 years, the rate of survival has not increased and becoming one of the most lethal cancer type (Rahib et al., 2014). Thus, constant efforts of the advancement of novel therapeutic alternatives is a need.

Much work has been done with microarray gene expression technology to reveal the central mechanism of pancreatic formation and progression and focus these methods for therapeutic approach. It still persists a request for more efficient treatments or methods that can improve curative responses to pancreatic cancer medication. In this project, we used microarray data sets of public transcriptome datasets of performed microarray experiments to detect the DEGs between the pancreatic tumor cell and normal cell specimens. Although results out of experimental studies should support our findings, our results will reveal potential biomarkers and bright therapeutic objectives for timely screening of pancreatic cancer. The present study also focused on the phosphoinositide-3/Akt (PI3K/AKT) signaling pathway and triggering receptor tyrosine kinases (RTKs) that takes a fundamental task in regulating downstream responses, involving cell viability, growth cycle, proliferation, cell migration and vascularization, by phosphorylating a variety of intracellular proteins (Franke, 2008; Hemmings & Restuccia, 2012). The pathway exists in all cells of

major eukaryotes and is extremely maintained. The current and future impersonal studies of suppressants fighting PI3K/AKT pathway in cancers should be clarified further. One of the goal is in this study to characterize PI3K/AKT pathway in PAAD to the practice. Second goal is to put forward the updated propitious to the PAAD patients for specialized therapies in PI3K/AKT.

2 Materials and Methods

2.1. Preprocessing of the data set

The publicly accessible data set of gene expression from pancreatic tumor cell and normal cell samples were pull out from the GEO database with GSE16515 (Pei et al., 2009). Genomic information ranging from gene sequences to protein structure predictions were obtained. As described by Pei et al., 2009, the data set contains a total of expression of 54,676 genes of in total 52 samples i.e., 36 tumor and 16 normal specimens.

Using the GEOquery package in Bioconductor following conventional procedures in R studio, the GSE16515 data set is studied (Davis & Meltzer, 2007). Along with GSE16515 gene expression data set, GSE28735 gene expression data set was used to confirm the findings. The DEGs are studied from the expression data set of 45 pancreatic tumor and 45 normal cells. The Cancer Genome Atlas (TCGA) database includes genomic sequencing data of 33 different cancer types from 11,328 patients.

The other packages we used in R studio are as the following; Biobase, biomaRT and gplots packages (Durinck et al., 2009; Huber et al., 2015; Warnes et al., 2009). To estimate the adjusted *p value* and avoid Type I errors, we used Bejamini-Hochberg Procedure to correct multiple testing. In order to adapt the statistical tests locally, hypergeometric model was performed for both of the down-regulated and up-regulated DEGs in the pathway and gene ontology (GO) enrichment analysis, and false discovery rate (FDR) were computed (Benjamini & Hochberg, 1995; Hochberg & Tamhane, 1987).

2.2. Experimental data and analysis codes

Analysis were conducted in the R statistical environment. Sample codes and analysis of GSE16515 data can be found https://github.com/mathbioGVN/GSE16515.HSC.microarray.project repository. We compared samples into two groups provided that pancreatic tumor and normal cells. The data set was normalized by computing the means of the samples of each group in R programming language. The process on separated samples which is grouped by categories was performed as computing fold-change (biological significance) difference between the means of the categories. A broadly performed statistical model is the t-distribution and its versions. A t-test compares the discrepancy of the average gene expression levels between the two samples or subgroups, given the noisiness of the data i.e. the difference in means between samples divided by the standard deviation. The genes are filtered in accordance with both fold change and *p value* criterion. Despite the fact that, methods to correct for multiple comparisons have been applicable for a long time such as Bonferroni correction, most of these methods are not appropriate to analyze gene expression data sets (Tarca et al., 2006). We highlight statistical significance performing *t-test* by taking *p value* cutoff 0.01 and fold cut-off > 2 to identify down and up-regulated DEGs between each category understudy.

2.3. Differentially expressed genes and clustering analysis

Using GEOquery package in Bioconductor, gene expression values were pull out for each sample and converted to base-2 logarithmic scale using R language. We used gplots package of R to create heatmaps of DEGs with heatmap.2 function. Clustering analysis of DEGs was done to match the expression pattern of DEGs in pancreatic tumor and normal cells.

2.4. GO terms and analysis of the pathway enrichments

Expression measurements annotations for up-regulated and down regulated DEGs for each group probes mapped to gene names using Ensemble Biomart package in R. All of the DEGs were characterized by their biological processes, molecular functions, and cellular components of GO and DAVID enrichments which stands for Visualization and Integrated Discovery (Sherman & Lempicki, 2009). All classified genes were cautiously examined and further parts like the Universal Protein resource, and physical properties GO and annotation types were taken using DAVID and KEGG Kyoto Encyclopedia of Genes and Genomes. We then compared the results of DAVID with NetworkAnalyst enrichments performed with KEGG (G. Zhou et al., 2019).

2.5. Screening of core genes using protein-protein interaction (PPI) network

NetworkAnalyst on the web accessible publicly, provides analysis of PPI networks for single gene lists using STRING Interactome (Szklarczyk et al., 2015). To comprehensively decipher the regulatory mechanisms in pancreatic ad and related diseases, DEGs from pancreatic tumors and normal cells samples were analyzed to form a PPI network and resulted core genes detected with previously reported GO classification and enrichment.

2.6. Core genes expression levels and survival study

In reference to the TCGA database, Ualcan (http://ualcan.path.uab. edu/index.html) (Chandrashekar et al., 2017) was utilized to conduct survival analysis. Kaplan-Meirer survival analysis was carried out using 21 identified hub genes relying on the hub gene expression values in PAAD. In contrast to normal cells, gene expression levels presents important individual differences in tumor cells. Low expression level shows the transcripts per million value (TPM) is equal or below the upper quartile whereas high expression level shows the TPM is above the upper quartile.

3 Results

3.1. Experimental data analysis

With gene expression result of the GSE16515 data set, we detect differentially expressed genes (**DEGs**) in total 355 genes from **pancreatic tumor** and **normal cells** which was demonstrated in volcano plot (Figure 1). We find the down-regulated and up-regulated DEGs of pancreatic tumor and normal cell comparison. The expression values were pull out, and a heatmap was created to show the tumor and normal cell discrepanc
> 2 and $p < 0.$
77 down-regul

(A) The boxplot shows gene expression of each sample of the raw data without normalization. (B) Plots displaying the gene expression discrepancy in pancreatic tumor and normal cells comparison. Black illustrates no change (NO), red illustrates down-regulated (Down), and blue illustrates up-regulated (Up) DEGs, FC, fold change.

Figure 1

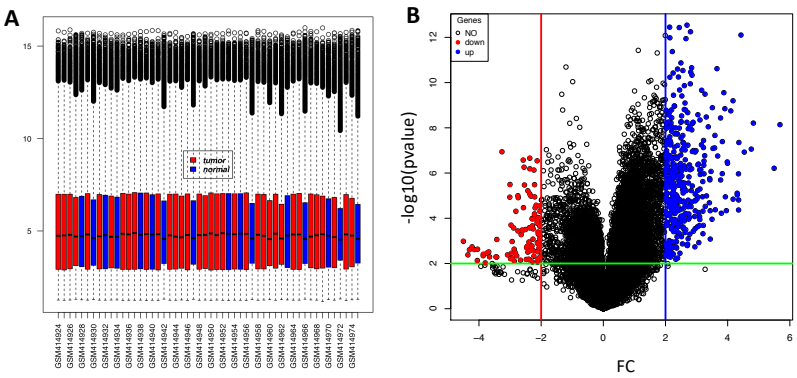
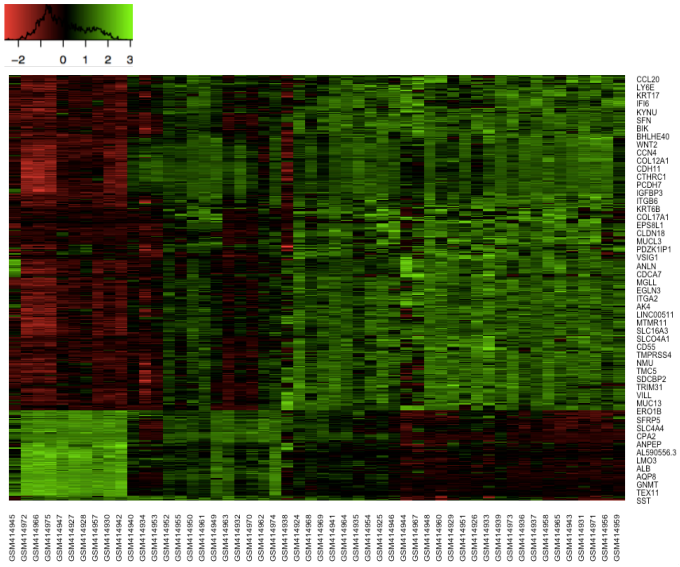


Figure 2 Heat map demonstrating DEGs in pancreatic tumor and normal cells aged samples. Each columns present samples, and rows present genes. Base-2 logarithmic values of the gene expression data are calculated. The progressive color changing from red to green represents the ranging from down to up-regulated DEGs.



3.2 Gene ontology and enrichment analysis

Table 2 shows the significant enrichment of DEGs using biological processes (BP) entailing collagen fibril

organization, collagen catabolic process, extracellular matrix organization, and type I interferon signaling pathway. The significant enrichment of DEGs in cellular component (CC) contains extracellular exosome, extracellular space, and extracellular region. Finally, the significant enrichments GO terms in molecular function (MF) is revealed metalloendopeptidase activity, extracellular matrix structural constituent, and heparin binding.

KEGG signaling pathway study outcomes demonstrated in which these DEGs were considerably enriched in ECM-receptor interaction, focal adhesin, PI3K/AKT signaling pathway, and central carbon metabolism in cancer. Among these pathways PI3K/AKT signaling pathway has vital influence on multiple cellular process which also has a role in cancer pathways almost in all cells also consistent previous findings (Jiang et al., 2020).

Table 2. Gene expression data set retrieved with top significant pathways GO enrichment analysis of DEGs in PAAD.

Category	Term	Count	*p*-value	Genes
BP	GO:0030574~collagen catabolic process	14	1.38E-10	MMP7, MMP1, COL11A1, COL12A1, KLK6, MMP12, COL1A1, MMP11, ADAMTS2, COL1A2, COL5A1, COL5A2, COL8A1, COL10A1
BP	GO:0030198~extracellular matrix organization	19	2.75E-08	ERO1A, ERO1B, POSTN, LAMB3, ITGB4, COL11A1, ITGA2, LAMA3, FN1, LAMC2, COL1A1, COMP, VCAN, COL1A2, COL5A1, COL5A2, COL8A1, COL10A1, ITGB6
BP	GO:0030199~collagen fibril organization	10	3.05E-08	GREM1, COL1A1, MMP11, ADAMTS2, COL1A2, COL5A1, COL11A1, COL12A1, COL5A2, LOXL2
BP	GO:0060337~type I interferon signaling pathway	10	2.64E-06	ISG20, IFI27, RSAD2, OAS1, OAS2, MX1, IFI6, ISG15, XAF1, OASL
CC	GO:0005615~extracellular space	77	6.52E-20	
CC	GO:0070062~extracellular exosome	105	4.85E-14	
CC	GO:0005576~extracellular region	72	6.01E-13	
CC	GO:0005578~proteinaceous extracellular matrix	22	1.67E-08	COL17A1, ECM1, POSTN, MMP7, MMP1, COL11A1, FN1, ADAMTS12, MMP12, COMP, MMP11, ADAMTS2, VCAN, COL1A2, COL5A1, COL5A2, MMP28, COL10A1, PI3, MUC4, WNT2, CTHRC1
MF	GO:0004222~metalloendopeptidase activity	12	6.95E-06	MMP12, MMP11, ADAMTS2, ADAM28, MMP7, FAP, MMP1, ADAM12, MMP28, ADAM9, ADAMTS12, KLK7
MF	GO:0005201~extracellular matrix structural constituent	9	2.87E-05	COL1A1, COMP, VCAN, COL1A2, COL5A1, COL11A1, COL5A2, MUC5B, MUC5AC
MF	GO:0008201~heparin binding	13	3.81E-05	POSTN, AOC1, MMP7, REG4, F11, FN1, LAMC2, CEL, THBS2, COMP, CXCL10, COL5A1, LIPH
KEGG	hsa04512:ECM-receptor interaction	15	3.15E-09	LAMB3, ITGB4, COL11A1, ITGA2, LAMA3, FN1, LAMC2, THBS2, COL1A1, COMP, COL1A2, COL5A1, COL5A2, SDC1, ITGB6
KEGG	hsa04510:Focal adhesin	17	3.15E-09	LAMB3, ITGB4, EGF, COL11A1, ITGA2, LAMA3, FN1, LAMC2, THBS2, COL1A1, COMP, COL1A2, COL5A1, COL5A2, ITGB6,

				PAK3, MET
KEGG	hsa04511:PI3K/AKT signaling pathway	17	02E-2	LAMB3, ITGB4, EGF, COL11A1, ITGA2, LAMA3, FN1, LAMC2, THBS2, EFNA5, COL1A1, COMP, COL1A2, COL5A1, COL5A2, ITGB6, MET
KEGG	hsa04 05230:Central carbon metabolism in cancer	6	0.1E-1	PKM, SLC2A1, SLC16A3, MET, HK2, PFKP

Abbreviations- gene ontology: GO; biological process: BP; cell component: CC; Kyoto Encyclopedia of Genes and Genomes: KEGG (by the p value)

3.4. The PPI network and KEGG pathway enrichment

Fig. 4 shows the PPI map between the set of input DEGs. The expressions of the nodes and their degree of connection were symbolized by yellow to red and fields, respectively in the visualized networks. The genes with the best three scores according to the *p value* is identified; these proteins also determines the functionality of the PPI network. Best scoring genes comprise with PPI network KEGG enrichment analysis represents involvement of ECM-receptor interaction (hsa04512), amoebiasis (hsa05146) , protein digestion and absorption (hsa04974), focal adhesion (hsa04510), and PI3K/AKT signaling pathway (hsa04151). The other significantly enriched pathways associated with pancreatic tumor and normal cells DEGs are ECM-receptor interaction (hsa04512) and Central carbon metabolism in cancer (hsa0405230). Furthermore, the enriched KEGG pathways also involved the 'PI3K/AKT signaling pathway' due to playing a vital function in tissue development which is also associated with the cancerous cells.

Figure 4

PPI network of DEGs was created and pictured using NetworkAnalyst. Progressive color change from yellow to the red represents relative gene expression value from down to up-regulated genes. CDK1 with the largest Betweenness Centrality (BC) was suggested to be central to the PPI network associated with pancreatic tumor and normal cells gene expression data. Moreover, CCNB1, CDC20, PPARG, MET, and ISG15 with the secondary highest degree and SFN with the secondary BC might be engaged in the progression

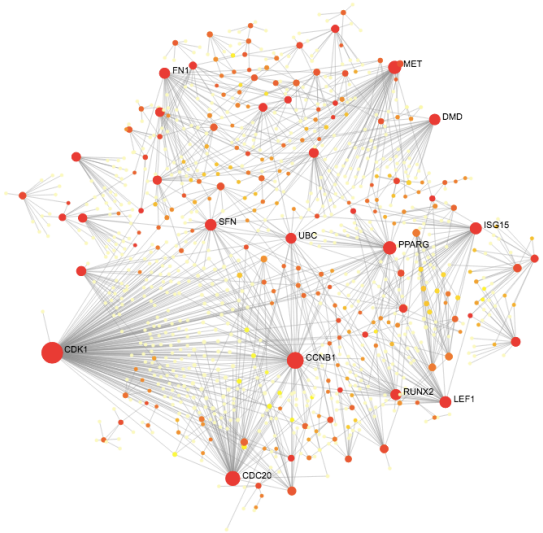


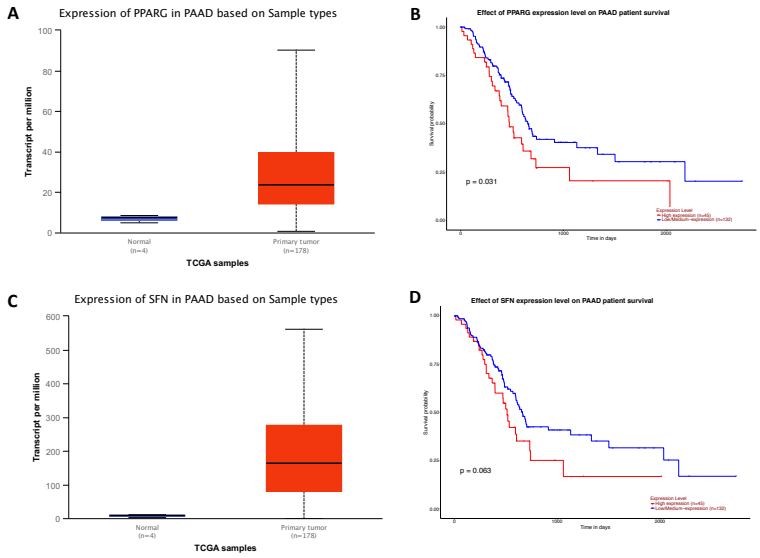
Fig.4 demonstrates PPI network of DEGs in pancreatic tumor and normal of all the DEGs . To conclude, PIP-network of pancreatic tumor and normal of all the DEGs (see Fig.4 and Table 3) in the expression data set. Hub genes in Fig.4 can be listed as CDK1, CCNB1, CDC20 , PPARG, MET, and ISG15. SFN gene is active in DNA repair, protein degradation, cell cycle regulation, endocytosis, control of cell signaling pathways, and kinase modification. UBC is a key gene that directly engages with other genes such as LC3, proposes that it might be a central component that heads to a bad prognosis of PAAD regulated by perineural invasion (Yang et al., 2019). ECT2 has a vital task in diseases comprising breast cancer and lung cancer. Related pathways of ECT2 are signaling pathways such as p75 NTR receptor-mediated signaling. GO annotations of molecular function related to this gene contain GTPase activator activity and protein homodimerization activity.

Table 3 Top 15 genes of PIP network of DEGs in pancreatic tumor and normal cells gene expression data.

Gene ID	Genes	Node Degree	Betweenness centrality	Expression
983	CDK1	191	170861.12	6.95
891	CCNB1	91	42477.27	6.05
991	CDC20	66	30778.32	5.08
5468	PPARG	48	43020.9	6.84
4233	MET	47	76573.47	9.87
9636	ISG15	37	31351.4	8.86
51176	LEF1	36	52887.37	6.81
2810	SFN	34	28856.95	8.26
1756	DMD	32	37877.67	7.59
2335	FN1	30	36750.92	12.61
860	RUNX2	30	26338.94	7.37
7316	UBC	28	156179.54	8.81
7153	TOP2A	24	18346.88	6.74
1894	ECT2	23	16342.09	7.29

3.5. Pancreatic tumor gene expression and survival analyses through UALCAN portal

Notably, PPARG and SFN proteins are related to mortality and/or survival. Once the detection of hub proteins of PPI network was revealed, survival analysis for PAAD was carried out through UALCAN. *PPARG*; peroxisome proliferator-activated receptor gamma and SFN; Stratifin were shown to be significantly related to survival (p-value < 0.05). Afterwards, the DEGs in PAAD were measured; PPARG and SFN were detected to be significantly differentially expressed. *PPARG* was upregulated significantly in pancreatic adenocarcinoma malignant cells in comparison to normal cells with p-value = 1.62×10^{-12} as a result of the differentially expressed values of both *PPARG* were studied using TCGA. More importantly, the outcomes match with those of the mentioned differential gene screening. In addition, SFN is also upregulated with in PAAD tumor and normal tissues with a p-value < 10^{-12}.



3.6. Validation of PPARG and SNF

45 pancreatic tumor and 45 normal samples of DEGs were studied. The expression level of PPARG and SFN was studied specifically. The outcomes of the analysis to validate the significance of PPARG and SFN are demonstrated in Table 4; it was noted that PPARG and SFN were expressed significantly upregulated in pancreatic adenocarcinoma cells in both GSE16515 and GSE28735.

Table 4 Base-2 logarithmic scale of differential expression of most significant hub genes in pancreatic cancer cells in two different datasets.

Datasets	Genes	FC	P-value
GSE16515	PPARG	6.85	1.62E-12
	SFN	5.04	
GSE28735	PPARG	6.17	<1E-12
	SFN	4.8	

3.7. The role of the PI3K/AKT signaling pathway

The core of this project due to it is close association with cancer within all of the significantly (*p value* < 0.01) enriched pathways of DEGs is the PI3K/AKT signaling pathway. There were 17 DEGs particularly engaged in this pathway, containing LAMB3, ITGB4, EGF, COL11A1, ITGA2, LAMA3, FN1, LAMC2, THBS2, EFNA5, COL1A1, COMP, COL1A2, COL5A1, COL5A2, ITGB6, MET (Fig. 5 and Table 2 and 4). We have performed the primary DEGs associated with PI3K/AKT signaling pathway in Fig.6. We implemented PI3K/AKT signaling pathway genes to construct Fig.5A showing Subnetwork 1 and Fig.5B showing Subnetwork 2 protein-protein interaction networks.

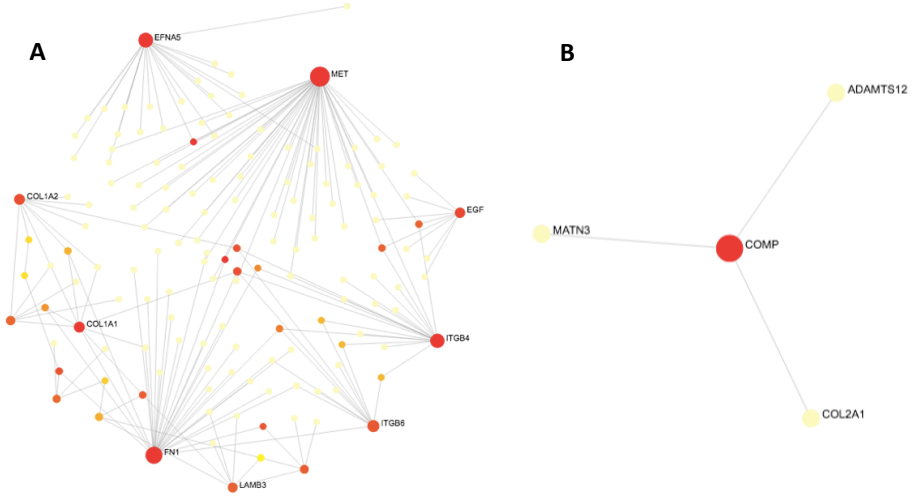
Associated genes with the DEGs of the data set enriched with the PI3K/AKT signaling pathway deciphered as new hub genes. In subnetwork 1 of PI3K/AKT pathway (Fig.5A), MET, FN1, and ITGB4 is the most significant genes in terms of BC. Moreover, collagen (COL) gene family also performs a key position which is demonstrated in the subnetwork of the PI3K/AKT signaling pathway. As shown in subnetwork 2, cartilage oligomeric matrix protein, COMP might be a gene that is associated with PI3K/AKT signaling pathway. We also found that the proteins associated with COMP gene in PI3K/AKT signaling pathway can be listed as ADAMTS12, MATN3, and COL2A1 in pancreatic tumor and normal cells gene expression data set. Even though the operational roles of other pathway members, such as ADAMTS12, stay unidentified downregulation or upregulation of particular ADAMTS proteins activated in serious human diseases, involving cancer (Llamazares et al., 2007) These outcomes confirm the vital task of the pathway of the PI3K/AKT signaling engaged in pancreatic tumor and related diseases treatment, offering new molecular therapeutic targets to improve fundamental drug agents.

Table 5 Top 10 most excessive KEGG pathway enrichment analysis of global DEGs in pancreatic tumor and normal cells micro-array gene expression data set.

Term	Count	p-value	Genes
hsa04512:ECM-receptor interaction	15	3.15E-09	LAMB3, ITGB4, COL11A1, ITGA2, LAMA3, FN1, LAMC2, THBS2, COL1A1, COMP, COL1A2, COL5A1, COL5A2, SDC1, ITGB6
hsa05146:Amoebiasis	13	2.27E-06	SERPINB3, CXCL8, SERPINB2, LAMB3, IL1R2, COL11A1, LAMA3, FN1, LAMC2, COL1A1, COL1A2, COL5A1, COL5A2
hsa04974:Protein digestion and absorption	12	2.28E-06	COL17A1, COL1A1, CPA2, COL1A2, CTRL, COL5A1, COL11A1, COL12A1, COL5A2, COL10A1, SLC16A10, KCNN4
hsa04510:Focal adhesion	17	7.13E-06	LAMB3, ITGB4, EGF, COL11A1, ITGA2, LAMA3, FN1, LAMC2, THBS2, COL1A1, COMP, COL1A2, COL5A1, COL5A2, ITGB6, PAK3, MET
hsa04151: PI3K/AKT signaling pathway	17	0.00271604	LAMB3, ITGB4, EGF, COL11A1, ITGA2, LAMA3, FN1, LAMC2, THBS2, EFNA5, COL1A1, COMP, COL1A2, COL5A1, COL5A2, ITGB6, MET
hsa05230:Central carbon metabolism in cancer	6	0.01154629	PKM, SLC2A1, SLC16A3, MET, HK2, PFKP
hsa00830:Retinol metabolism	6	0.01154629	SDR16C5, DHRS9, ADH1B, AOX1, UGT1A8, CYP2C18
hsa04610:Complement and coagulation cascades	6	0.01564074	C6, PLAU, F11, PLAUR, CD55, F5

hsa05144:Malaria	5	0.02015542	COMP, CXCL8, SDC1, THBS2, MET
hsa05200:Pathways in cancer	16	0.02027915	EGLN3, CXCL8, LAMB3, MMP1, EGF, ZBTB16, ITGA2, LAMA3, LEF1, SLC2A1, FN1, LAMC2, CKS2, PPARG, MET, WNT2

Figure 5 Hub genes associated with PI3K/AKT signaling pathway. (A) Subnetwork 1 and (B) Subnetwork 2 protein-protein interaction networks.



4 Discussions

The prevalence of pancreatic adenocarcinoma and the related survival rates have demonstrated a decreasing in tendency in the past years (Siegel et al., 2018). One study showed that PAAD patients survive for only 4 months typically without therapies. Moreover, patients who undergo surgery and take required therapies the survival is not significantly increased (Wang et al., 2015). Thus, precise quick identification of PAAD and the advancement of powerful specific remedy is of fundamental significance.

A recent research detected hub genes in PAAD that were stated to be of diagnostic significance (Lv et al., 2019). In this project, the microarray dataset GSE16515 gene expression was studied, holding gene

expression of 36 tumor and 16 normal cells. DEGs were only analyzed between tumor and normal cells. 278 upregulated and 77 downregulated genes were detected using R, and gene ontology enrichment and KEGG pathway analyses which showed locational and functional information of these DEGs. Upregulated DEGs were primarily positioned in the collagen trimers and extracellular exosome, and were engaged in cell adhesin, ECM organization and proteolysis, positive regulation of cell proliferation, collagen catabolic process and signal transduction. Contrary, downregulated DEGs were primarily enriched in proteolysis, digestion, and apoptotic process.

Finally, protein-protein interaction network construction show information about the core genes is used to identify function and disease associated with proteins. The highest interactive proteins CDK1, CCNB1, CDC20, PPARG, MET, LEF1, SFN, DMD, FN1, UBC, TOP2A, ECT2, WNT2, EFNA5, PAK3, PKM, ITGB4, NEK2, and ALB are predicted that are involved in several types of cancers like retinoblastoma, breast cancer, rhabdomyosarcoma, diabetes, childhood obesity, ovarian cancer, acute lymphoblastic leukemia, colorectal and pancreatic cancer (Santiago et al., 2017; Tovar & Graveel, 2017). ISG15 pathway takes an important part in the tumorigenesis and treatment of digestive system cancers (Zuo et al., 2016). RUNX2 gene can be employed as molecular marker for diagnosis of initial stage of pancreatic cancer can also operate as promising a pro-oncogenic potential (Ozaki et al., 2018).

Finally, 21 core genes were detected using a PPI network and the prognostic measures of the values of the core genes for PAAD patients were studied through UACLAN. Utilizing data from TCGA that both *PPARG* and *SFN* were upregulated significantly in PAAD cells in comparison to control cells, as per the outcomes of the differential gene analysis. The gene expression of GSE28735 data set was further studied for the verification of the upregulated expression of SFN and PPARG. Overall, it was revealed that *SFN* and *PPARG* were considerably related with poorer survival. Thus, these two genes may be associated with the carcinogenesis and angiogenesis of PAAD.

After the enrichment analyses of KEGG and GO pathways, the functional enrichment of peroxisome proliferator activated receptor gamma (PPARG) was studied. *PPARG* is primarily involved in the modulation of fatty acid storage and glucose metabolism. SFN plays a function in kinase modification, endocytosis, DNA repair, protein degradation, cell cycle regulation, and modulation of various cell signaling pathways.

Increased expression of PPARG might be the cause of the pathology of various diseases such as diabetes, obesity, atherosclerosis and pancreatic cancer. Therapies of inhibiting SFN downregulates Bcl-2 and XIAP expressions to stimulate apoptosis which directs to the downsizing in EMT, car, carcinogenic, angiogenic biomarkers with considerable suppression in tumor development (Robin et al., 2020).

This investigation further highlighted the PI3K/AKT signaling pathway involving differentially expressed genes in a broad various kinds of human cancer. AKT inhibitors have significant outlooks in the study of targeted cancer therapy since AKT is a vital transmitter of the PI3K/AKT signaling pathway (Jiang et al., 2020). The significant position of the PI3K/AKT signaling network in regulating cell growth, the cell cycle and apoptosis over the last 20 years was known (Rodon et al., 2013). Furthermore, mutations employ in PI3K/AKT signaling pathway genes and many related pathway genes provide pancreatic ductal adenocarcinoma carcinogenesis. In addition to their standard roles, PI3K/AKT further rules metabolism

characteristics of tumor cell and tumor microenvironment-intervened mechanisms of neoplasm development and survival.

The presented project has certain limitations. In studying the expression level both PPARG and SFN only two datasets were analyzed, and more gene expression data should be investigated with a higher number of control samples is a need to verify the presented results. Secondly, additional studies is required for clinical lab confirmation of predicted proteins that are expressed in pancreatic tumor and normal cells data set and express at the developmental stage of pancreatic adenocarcinoma. More research is needed in the field of cancer biology to detect pancreatic cancer and subset diseases at its early stage. This paper also emphasizes the importance of microarray experiment in comprehending pancreatic cancer and related diseases and approach to study several results of gene expression data, like differentially expressed genes analysis, pathway and process identification, and protein-protein interaction network study.

References

1. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.
2. Chandrashekar, D. S., Bashel, B., Balasubramanya, S. A. H., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B. V., & Varambally, S. (2017). UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia*, *19*(8), 649–658.
3. Davis, S., & Meltzer, P. (2007). GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics (Oxford, England)*, *23*, 1846–1847. https://doi.org/10.1093/bioinformatics/btm254
4. Dineen, S. P., Roland, C. L., Greer, R., Carbon, J. G., Toombs, J. E., Gupta, P., Bardeesy, N., Sun, H., Williams, N., & Minna, J. D. (2010). Smac mimetic increases chemotherapy response and improves survival in mice with pancreatic cancer. *Cancer Research*, *70*(7), 2852–2861.

5. Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, *4*(8), 1184.
6. Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, *30*(1), 207–210. https://doi.org/10.1093/nar/30.1.207
7. Franke, T. F. (2008). PI3K/Akt: Getting it right matters. *Oncogene*, *27*(50), 6473–6488. https://doi.org/10.1038/onc.2008.313
8. Gillen, S., Schuster, T., Büschenfelde, C. M. zum, Friess, H., & Kleeff, J. (2010). Preoperative/Neoadjuvant Therapy in Pancreatic Cancer: A Systematic Review and Meta-analysis of Response and Resection Percentages. *PLOS Medicine*, *7*(4), e1000267. https://doi.org/10.1371/journal.pmed.1000267
9. Grønborg, M., Kristiansen, T. Z., Iwahori, A., Chang, R., Reddy, R., Sato, N., Molina, H., Jensen, O. N., Hruban, R. H., & Goggins, M. G. (2006). Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. *Molecular & Cellular Proteomics*, *5*(1), 157–171.
10. Hemmings, B. A., & Restuccia, D. F. (2012). PI3K-PKB/Akt Pathway. *Cold Spring Harbor Perspectives in Biology*, *4*(9). https://doi.org/10.1101/cshperspect.a011189
11. Hochberg, Y., & Tamhane, A. C. (1987). *Multiple comparison procedures*. John Wiley & Sons, Inc.
12. Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., … Morgan, M.

(2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, *12*(2), 115–121. https://doi.org/10.1038/nmeth.3252

13. Jiang, N., Dai, Q., Su, X., Fu, J., Feng, X., & Peng, J. (2020). Role of PI3K/AKT pathway in cancer: The framework of malignant behavior. *Molecular Biology Reports*, *47*(6), 4587–4629. https://doi.org/10.1007/s11033-020-05435-1
14. Jones, S., Zhang, X., Parsons, D. W., Lin, J. C.-H., Leary, R. J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., & Jimeno, A. (2008). Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*, *321*(5897), 1801–1806.
15. Lambert, A., Schwarz, L., Borbath, I., Henry, A., Van Laethem, J.-L., Malka, D., Ducreux, M., & Conroy, T. (2019). An update on treatment options for pancreatic adenocarcinoma. *Therapeutic Advances in Medical Oncology*, *11*, 175883591987556. https://doi.org/10.1177/1758835919875568
16. Llamazares, M., Obaya, A. J., Moncada-Pazos, A., Heljasvaara, R., Espada, J., López-Otín, C., & Cal, S. (2007). The ADAMTS12 metalloproteinase exhibits anti-tumorigenic properties through modulation of the Ras-dependent ERK signalling pathway. *Journal of Cell Science*, *120*(20), 3544–3552. https://doi.org/10.1242/jcs.005751
17. Lv, K., Yang, J., Sun, J., & Guan, J. (2019). Identification of key candidate genes for pancreatic cancer by bioinformatics analysis. *Experimental and Therapeutic Medicine*, *18*(1), 451–458. https://doi.org/10.3892/etm.2019.7619

18. Ozaki, T., Yu, M., Yin, D., Sun, D., Zhu, Y., Bu, Y., & Sang, M. (2018). Impact of RUNX2 on drug-resistant human pancreatic cancer cells with p53 mutations. *BMC Cancer*, *18*(1), 309. https://doi.org/10.1186/s12885-018-4217-9

19. Pei, H., Li, L., Fridley, B. L., Jenkins, G. D., Kalari, K. R., Lingle, W., Petersen, G., Lou, Z., & Wang, L. (2009). FKBP51 Affects Cancer Cell Response to Chemotherapy by Negatively Regulating Akt. *Cancer Cell*, *16*(3), 259–266. https://doi.org/10.1016/j.ccr.2009.07.016

20. Rahib, L., Smith, B. D., Aizenberg, R., Rosenzweig, A. B., Fleshman, J. M., & Matrisian, L. M. (2014). Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States. *Cancer Research*, *74*(11), 2913–2921. https://doi.org/10.1158/0008-5472.CAN-14-0155

21. Robin, F., Angenard, G., Cano, L., Courtin-Tanguy, L., Gaignard, E., Khene, Z.-E., Bergeat, D., Clément, B., Boudjema, K., Coulouarn, C., & Sulpice, L. (2020). Molecular profiling of stroma highlights stratifin as a novel biomarker of poor prognosis in pancreatic ductal adenocarcinoma. *British Journal of Cancer*, *123*(1), 72–80. https://doi.org/10.1038/s41416-020-0863-1

22. Rodon, J., Dienstmann, R., Serra, V., & Tabernero, J. (2013). Development of PI3K inhibitors: Lessons learned from early clinical trials. *Nature Reviews Clinical Oncology*, *10*(3), 143–153.

23. Santiago, L., Daniels, G., Wang, D., Deng, F.-M., & Lee, P. (2017). Wnt signaling pathway protein LEF1 in cancer, as a biomarker for prognosis and a target for treatment. *American Journal of Cancer Research*, *7*(6), 1389–1406.

24. Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44.

25. Siegel, R. L., Miller, K. D., & Jemal, A. (2018). Cancer statistics, 2018. *CA: A Cancer Journal for Clinicians*, *68*(1), 7–30. https://doi.org/10.3322/caac.21442

26. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., & Tsafou, K. P. (2015). STRING v10: Protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, *43*(D1), D447–D452.

27. Tarca, A. L., Romero, R., & Draghici, S. (2006). Analysis of microarray experiments of gene expression profiling. *American Journal of Obstetrics and Gynecology*, *195*(2), 373–388. https://doi.org/10.1016/j.ajog.2006.07.001

28. Tovar, E. A., & Graveel, C. R. (2017). MET in human cancer: Germline and somatic mutations. *Annals of Translational Medicine*, *5*(10). https://doi.org/10.21037/atm.2017.03.64

29. Wang, X., Wang, L., Mo, Q., Dong, Y., Wang, G., & Jia, A. (2015). *Changes of Th17/Treg cell and related cytokines in pancreatic cancer patients*. 7.

30. Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., & Moeller, S. (2009). gplots: Various R programming tools for plotting data. *R Package Version*, *2*(4), 1.

31. Yang, Y.-H., Zhang, Y.-X., Gui, Y., Liu, J.-B., Sun, J.-J., & Fan, H. (2019). Analysis of the autophagy gene expression profile of pancreatic cancer based on autophagy-related protein microtubule-associated protein 1A/1B-light chain 3. *World Journal of Gastroenterology*, *25*(17), 2086–2098. https://doi.org/10.3748/wjg.v25.i17.2086

32. Zhou, G., Soufan, O., Ewald, J., Hancock, R. E. W., Basu, N., & Xia, J. (2019). NetworkAnalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Research*, *47*(W1), W234–W241.

33. Zuo, C., Sheng, X., Ma, M., Xia, M., & Ouyang, L. (2016). ISG15 in the tumorigenesis and treatment of cancer: An emerging role in malignancies of the digestive system. *Oncotarget*, *7*(45), 74393–74409. https://doi.org/10.18632/oncotarget.11911