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

Address

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 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 pancretic 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 identified using NetworkAnalyst based on the protein-interaction network of DEGs determined using the online tool STRING: CDK1, CCNB1, CDC20, PPARG, MET, ISG15, LEF1, SFN, DMD, FN1, RUNX2, UBC, TOP2A, ECT2, WNT2, EFNA5, PAK3, PKM, ITGB4, NEK2, and ALB. Among these genes, the expression levels of PPARG and SFN were analyzed in TCGA database, and the results were consistent with those of the previously noted differential gene analysis; was significantly upregulated in PAAD tumor tissues compared with normal tissues. Additionally, the module analysis of protein-protein interactions revealed that 'ECM-receptor interaction', 'amoebiasis', 'protein digestion and absorption', and 'focal adhesion' 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 used as potential targets for pancreatic cancer and related diseases diagnosis and treatment.

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

Reference to this paper should be made as follows: Author. (xxxx) 'Title', *Int. J. xxxxxxxxxxx xxxxxxxxxxx*,

Biographical notes:

This paper is a revised and expanded version of a paper entitled ***[title]*** *presented at [****name, location and date of conference****].*

1 Introduction

Pancreatic Adenocarcinoma (PAAD) is one of the most deadly malignancies due to the diagnosis at a distant stage which has a 5-year overall survival rate as 3% (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; Tempero et al., 2014). However, pancreatectomy still remains the most effective treatment, especially for early stage pancreatic cancer (Lambert et al., 2019). Thus, an updated understanding of the fundamental mechanism of pancreatic cancer is necessary for more effective therapies and the advancement of patient survival.

Microarray has become a significant tool in investigating pancreatic cancer genes and target therapeutic drugs (W. Zhou et al., 1998). Recent studies suggest an extensive gene expression analysis of pancreatic cancer by reviewing publicly available gene expression data set by comparing tumor and normal cells (Alldinger et al., 2005; Jones et al., 2008; W. Zhou et al., 1998).

Moreover, relative study of the differentially expressed genes (DEGs) stays moderately constrained, and a dependable biomarker profile would be a need to develop new gene targets (Grützmann et al., 2005). The protein expression alterations in the development and growth of pancreatic cancer and related diseases require comprehensive analysis. Candidate biomarkers from such studies can subsequently be tested using other techniques for use in early detection of cancers (Grønborg et al., 2006). Furthermore, the relations among the detected DEGs, specifically protein-protein interaction (PPI) networks and underlying signaling pathways should be clarified.

Pei *et al*., 2009, recently performed microarray experiments to identify the expression differences of FKBP5 gene between the pancreatic tumor and normal samples (data accessible 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 screened using DAVID Gene Ontology (GO) 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 objective of this project is to investigate the molecular mechanisms of pancreatic cancer 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)

Conventional therapeutic options, including chemotherapy, are not effective against pancreatic cancer, and despite improvements in the last decade, the mortality rate has not declined and is estimated to become the second cause of cancer-related deaths by 2030 (Rahib et al., 2014). Therefore, continuous efforts focus on the development of novel therapeutic options (Kamimura et al., 2018).

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 identify the expression differences of filtered genes between the pancreatic tumor and normal samples. Although experimental studies are needed to confirm our findings, our results

will reveal potential biomarkers and bright therapeutic objectives for early detection of pancreatic cancer. The present study also focused on the phosphoinositide-3-kinase–protein kinase B/Akt (PI3K-PKB/Akt) signaling pathway and activating receptor tyrosine kinases (RTKs) that plays a key task in regulating downstream responses, including cell survival, growth, proliferation, cell migration and angiogenesis, by phosphorylating a range of intracellular proteins (Franke, 2008; Hemmings & Restuccia, 2012). The pathway exists in all cells of major eukaryotes and is extremely maintained (Manning & Cantley, 2007). The current and future clinical trials of inhibitors against PI3K/AKT pathway in cancers should be investigated further. Our goal is to feature PI3K/AKT pathway in cancers to the clinic and bring the new promising to patients for targeted therapies.

2 Materials and Methods

2.1. Preprocessing of the data set

The publicly available gene expression data set from pancreatic tumor and normal 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 36 tumor samples and 16 normal samples; a total of 52 samples.

The GSE16515 data set is analyzed by using the GEOquery package in Bioconductor following standard procedures in R studio (Durinck et al., 2005, 2009; Tarca et al., 2006; Warnes et al., 2009). The other packages we used in R studio are as the following; Biobase, biomaRT and gplots packages (Davis & Meltzer, 2007; Durinck et al., 2005; 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 functional GO and pathway enrichment analysis, and false discovery rate (FDR) were computed (Benjamini & Hochberg, 1995; Dudoit et al., 2003; 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 performed to compare the expression pattern of DEGs in 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 gene ontology (GO) enrichment of the database for Annotation and DAVID which stands for Visualization and Integrated Discovery (Huang et al., 2007; Sherman & Lempicki, 2009). All classified genes were cautiously examined and further parts like the Universal Protein resource, and physical properties Gene Ontology (GO) and annotation types were taken using DAVID and KEGG Kyoto Encyclopedia of Genes and Genomes (Eden et al., 2009; Supek et al., 2011). We then compared the results of DAVID with NetworkAnalyst enrichments performed with KEGG (G. Zhou et al., 2019).

2.5. Hub genes screening from the protein-protein interaction (PPI) network

NetworkAnalyst, publicly accessible on the web, 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 with previously reported GO classification and enrichment.

2.6. Expression levels and survival analysis of hub genes in PAAD

Based on the information of 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 showed the TPM value was equal or below the upper quartile value whereas high expression level showed the TPM value was above the upper quartile value.

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 discrepancy (Figure 2). DEGs were selected with common *t* test, and labelled with fold change cut-off > 2 and $p < 0.01$. Here, we detected 278 differentially expressed genes of up regulation, whereas we find 77 down-regulated gene.

Figure 1 (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.

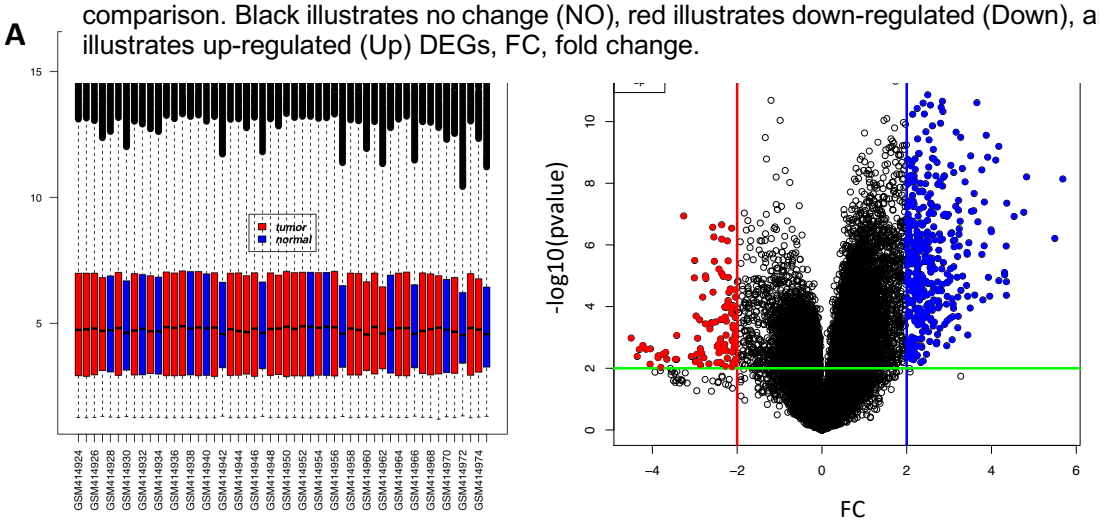
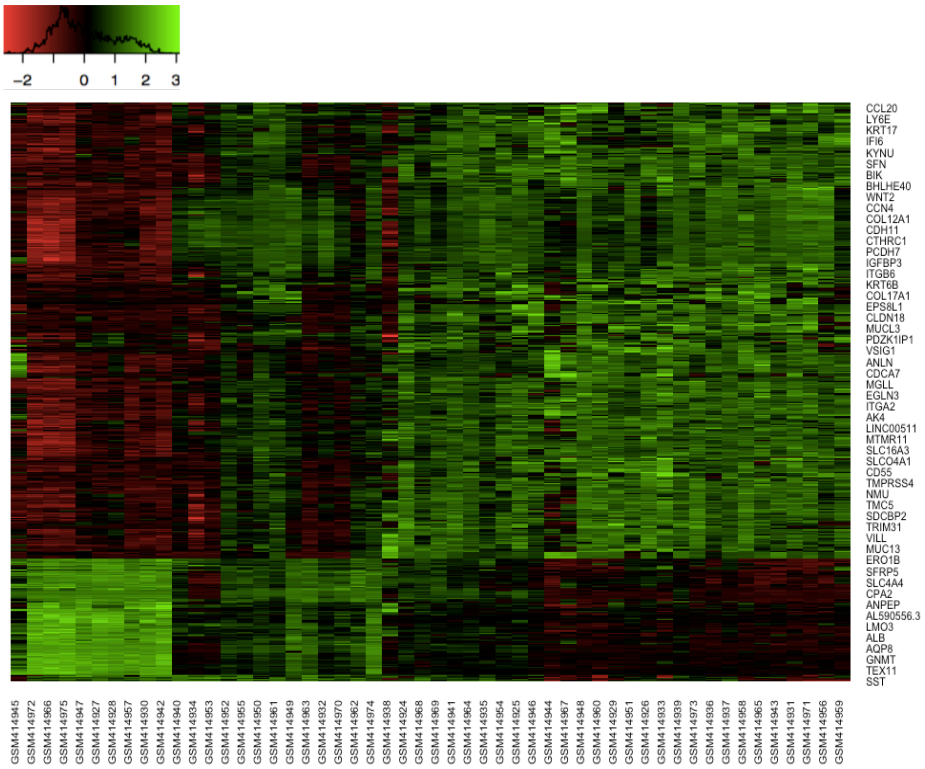


Figure 2 Heat map demonstrating DEGs in pancreatic tumor and normal cells aged samples. Each columns present samples, and rows present genes. Base-2 logarithmic values of the gene expression data are calculated. The progressive color changing from red to green represents the ranging from down to up-regulated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



3.2 Gene ontology and enrichment analysis

Table 2 shows the significant enrichment of DEGs using biological processes (BP) involving collagen catabolic process, extracellular matrix organization, collagen fibril organization and type I interferon signaling pathway. The significant enrichment of DEGs in cellular component (CC) contains extracellular space, extracellular exosome, and extracellular region. Finally, the significant enrichments GO terms in molecular function (MF) is revealed metalloendopeptidaseactivity, extracellular matrix structural constituent, and heparin binding.

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

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

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 key function in cell and tissue growth 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 pancreatic adenocarcinoma and other substance diseases.

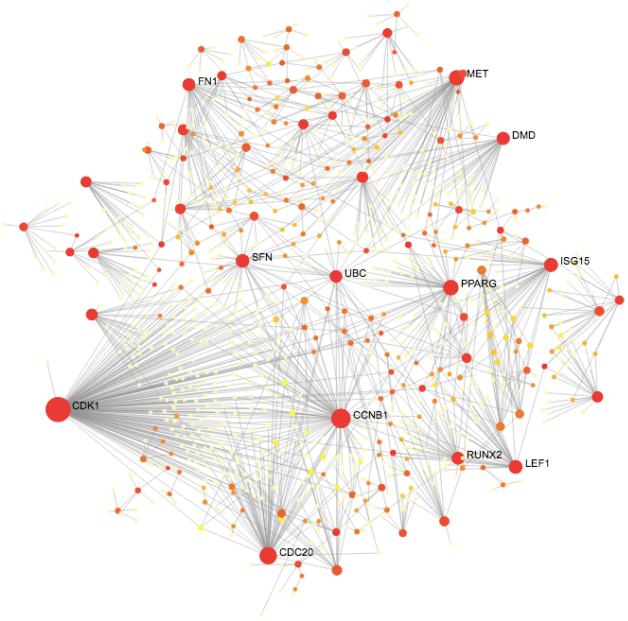


Fig.4 demonstrates PPI network of DEGs in pancreatic tumor and normal of all the DEGs . To conclude, PIP-network of pancreatic tumor and normal of all the DEGs (see Fig.4 and Table 3) in the expression data set. Hub genes in Fig.4 can be listed as CDK1, CCNB1, CDC20 , PPARG, MET, and ISG15. SFN gene is active in protein degradation, DNA repair, cell cycle regulation, kinase modification, endocytosis, and regulation of other cell signaling pathways. UBC is a key gene that directly engages with other genes such as LC3, suggest that it may be a key factor that leads to a poor prognosis of pancreatic cancer mediated by perineural invasion and indicates a new direction for further research (Yang et al., 2019). ECT2 has a key role in diseases including breast cancer and lung

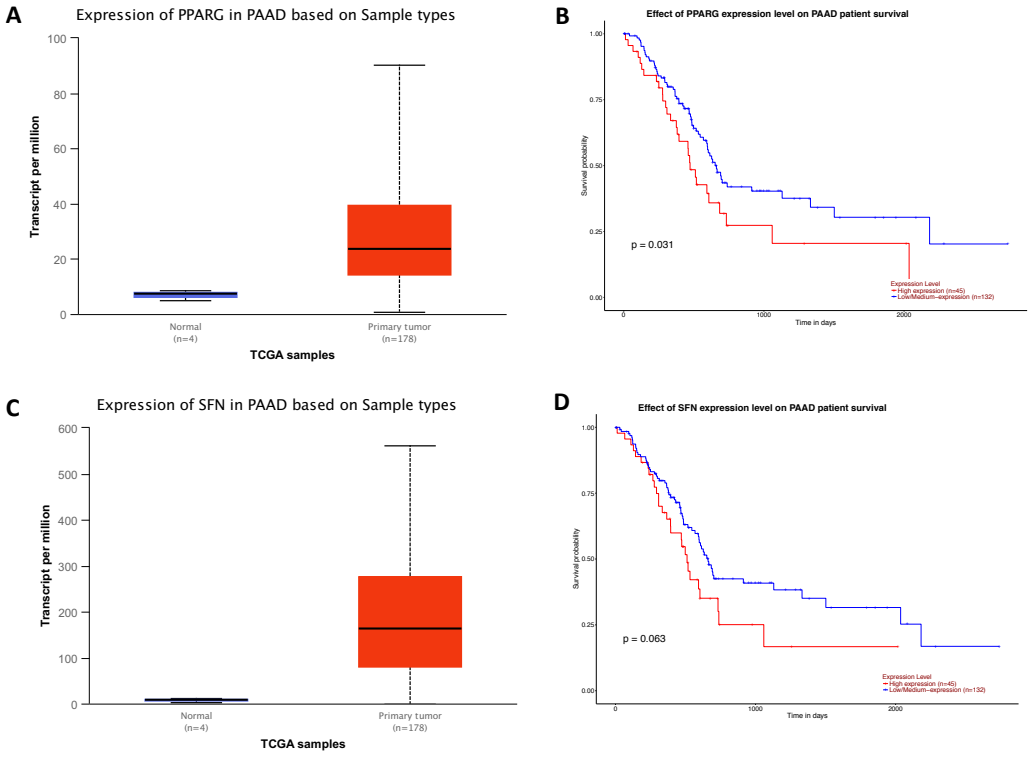
cancer. Related pathways of ECT2 are signaling pathways such as p75 NTR receptor-mediated signaling. Gene Ontology (GO) annotations of molecular function related to this gene include *protein homodimerization activity* and *GTPase activator activity*.

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 Gene expression level and survival analysis

Notably, PPARG and SFN genes were the only genes associated with survival. Following the identification of hub genes of PPI network revealed, survival analysis for PAAD was performed using UALCAN. *PPARG* ; Peroxisome proliferator-activated receptor gamma and SFN; Stratifin were demonstrated to be significantly associated with survival ($P<0.05$). Subsequently, the expression levels of genes in primary pancreatic cancer were compared; there were only one gene was identified to be significantly differentially expressed, *PPARG*, SFN was not significantly differentially expressed. The expression levels of *PPARG* were analyzed in TCGA database, and the results were consistent with those of the mentioned differential gene analysis; *PPARG* was significantly upregulated in PAAD tumor tissues compared with normal tissues $P=1.62\times10^{-12}$.



3.6. Validation of PPARG and SNF

45 tumor samples and 45 normal samples of DEGs were analyzed. The expression level of PPARG and SFN was studied specifically. The outcomes of the analysis to validate the significance of PPARG and SFN are presented in Table 4; it was observed that PPARG and SFN was significantly upregulated in pancreatic tumor tissue in the two GEO datasets.

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. Analysis 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.

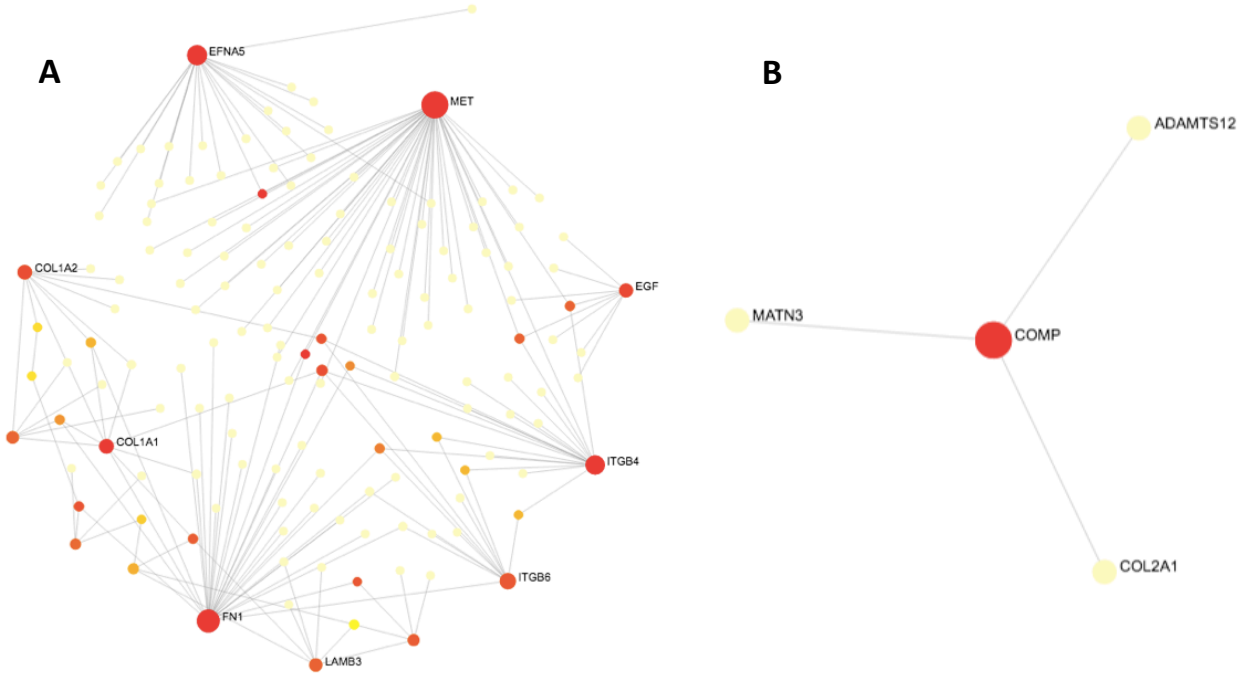
Associated genes with the DEGs of the data set enriched with the PI3K-Akt signaling pathway deciphered as new hub genes. In subnetwork 1 of PI3K-Akt pathway (Fig.5A), MET, FN1, and ITGB4 is the most significant genes in terms of BC. Moreover, collagen (COL) gene family also performs a key position which 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 functional roles of other family members, such as ADAMTS12, remain unknown, deficiency or overexpression of certain ADAMTS proteins is directly involved in serious human diseases, including cancer. (Llamazares et al., 2007). These outcomes confirm the key task of the PI3K-Akt signaling pathway engaged in pancreatic cancer 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 mortality rates have demonstrated an increasing tendency in previous years (Siegel et al., 2018). Research performed with pancreatic adenocarcinoma patients survive for only 4 months typically without therapies; even in patients who undergo surgery and take required therapies, the survival is not significantly increased (Wang et al., 2015).

As a result, precise quick identification of pancreatic cancer and the advancement of powerful specific therapy is of fundamental significance.

A recent study detected hub genes in pancreatic cancer that were reported to be of diagnostic relevance (Lv et al., 2019). In this study, the microarray GSE16515 from the GEO was conducted, holding data of 36 tumor and 16 normal cells (Pei et al., 2009). Differentially expressed genes were only analyzed between normal tissues and tumor tissues. A total of 355 DEGs (278 upregulated and 77 downregulated) were identified using R, and GO (Gene Ontology Consortium, 2004) and KEGG pathway analyses of DEGs revealed the locations and functions of DEGs. Upregulated genes were mainly located in the extracelluar exosome and collagen trimers, and were involved in cell adhesin, ECM organization and proteolysis, positive regulation of cell proliferation, collagen catabolic process and signal transduction. Conversely, downregulated genes were mainly enriched in proteolysis, digestion, and apoptotic process . A PPI network was built, and 21 core genes were identified; the prognostic value of these genes for patients with pancreatic cancer was analyzed using UACLAN. *SFN* and *PPARG* were significantly associated with poorer survival; it was then revealed using data from TCGA that *PPARG* was significantly upregulated in pancreatic cancer tissues compared with control tissues, consistent with the results of the differential gene analysis. It was predicted that these two genes may be associated with the carcinogenesis and angiogenesis of pancreatic cancer.

Following GO and KEGG analyses, the functional enrichment of peroxisome proliferator activated receptor-γ (PPARG) was investigated. *PPARG* is mainly involved in the regulation of fatty acid storage and glucose metabolism. SFN plays a key role in in protein degradation, DNA repair, cell cycle regulation, kinase modification, endocytosis, and regulation of other cell signaling pathways. Increased expression of PPARG might be the cause of the pathology of various diseases such as obesity, diabetes, atherosclerosis and pancreatic cancer. Therapies of inhibiting SFN downregulates the expression of Bcl-2 and XIAP to induce apoptosis that directs to the downsizing in EMT, car, carcinogenic, angiogenic markers with considerable inhibition in tumor development in mice (Li et al., 2013).

Protein-protein interaction network construction reveal information about the hub 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 , oral squamous cell carcinoma, colorectal and pancreatic cancer (Santiago et al., 2017; Sun et al., 2020; Tovar & Graveel, 2017; Tyagi et al., 2011). ISG15 pathway plays an important role in the tumorigenesis and treatment of cancers of the digestive system. (Zuo et al., 2016, p. 15). RUNX2 gene can be employed as biomarkers for diagnosis of early stage of pancreatic cancer can also operate as promising a pro-oncogenic potential (Ozaki et al., 2018).

This investigation further highlighted the PI3K-Akt signaling pathway involving differentially expressed genes in a broad various kinds of human cancer (Jiang et al., 2020). The significant role of the PI3K-Akt signaling network in regulating cell proliferation, the cell cycle and apoptosis over the last 20 years was known (Hennessy et al., 2005; Rodon et al., 2013). Furthermore, mutations employ in PI3K pathway genes and other-associated regulated pathway genes contribute to pancreatic ductal adenocarcinoma tumorigenesis. In addition their standard roles, PI3K also regulates

metabolic attributes of cancer cell and tumor microenvironment-mediated regulation of tumor growth and survival (Murthy et al., 2018).

The present study presented certain limitations. In examining the expression level both PPARG and SFN only two datasets were investigated, and further studies examining a high number of control samples are required to confirm the present results. Secondly, additional studies is required for clinical lab confirmation of predicted genes 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

Alldinger, I., Dittert, D., Peiper, M., Fusco, A., Chiappetta, G., Staub, E., Löhr, M., Jesenofsky, R., Baretton, G., & Ockert, D. (2005). Gene expression analysis of pancreatic cell lines reveals genes overexpressed in pancreatic cancer. *Pancreatology*, *5*(4–5), 370–379.

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.

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.

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

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.

Dudoit, S., Shaffer, J. P., & Boldrick, J. C. (2003). Multiple hypothesis testing in microarray experiments. *Statistical Science*, 71–103.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., & Huber, W. (2005). BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics*, *21*(16), 3439–3440.

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.

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

Franke, T. F. (2008). PI3K/Akt: Getting it right matters. *Oncogene*, *27*(50), 6473–6488. https://doi.org/10.1038/onc.2008.313

Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, *32*(suppl_1), D258–D261.

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

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.

Grützmann, R., Boriss, H., Ammerpohl, O., Lüttges, J., Kalthoff, H., Schackert, H. K., Klöppel, G., Saeger, H. D., & Pilarsky, C. (2005). Meta-analysis of microarray data on pancreatic cancer defines a set of commonly dysregulated genes. *Oncogene*, *24*(32), 5079–5088.

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

Hennessy, B. T., Smith, D. L., Ram, P. T., Lu, Y., & Mills, G. B. (2005). Exploiting the PI3K/AKT pathway for cancer drug discovery. *Nature Reviews Drug Discovery*, *4*(12), 988–1004.

Hochberg, Y., & Tamhane, A. C. (1987). *Multiple comparison procedures*. John Wiley & Sons, Inc.

Huang, D. W., Sherman, B. T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M. W., & Lane, H. C. (2007). DAVID Bioinformatics Resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Research*, *35*(suppl_2), W169–W175.

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

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.

Kamimura, K., Yokoo, T., & Terai, S. (2018). Gene Therapy for Pancreatic Diseases: Current Status. *International Journal of Molecular Sciences*, *19*(11), 3415. https://doi.org/10.3390/ijms19113415

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

Li, S.-H., Fu, J., Watkins, D. N., Srivastava, R. K., & Shankar, S. (2013). Sulforaphane regulates self-renewal of pancreatic cancer stem cells through the modulation of Sonic hedgehog–GLI pathway. *Molecular and Cellular Biochemistry*, *373*(1), 217–227. https://doi.org/10.1007/s11010-012-1493-6

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

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

Manning, B. D., & Cantley, L. C. (2007). AKT/PKB signaling: Navigating downstream. *Cell*, *129*(7), 1261–1274. https://doi.org/10.1016/j.cell.2007.06.009

Murthy, D., Attri, K. S., & Singh, P. K. (2018). Phosphoinositide 3-Kinase Signaling Pathway in Pancreatic Ductal Adenocarcinoma Progression, Pathogenesis, and Therapeutics. *Frontiers in Physiology*, *9*. https://doi.org/10.3389/fphys.2018.00335

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

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

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

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.

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.

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

Shi, X., Wang, J., Lei, Y., Cong, C., Tan, D., & Zhou, X. (2019). Research progress on the PI3K/AKT signaling pathway in gynecological cancer (Review). *Molecular Medicine Reports*, *19*(6), 4529–4535. https://doi.org/10.3892/mmr.2019.10121

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

Sun, Y., Zhao, C., Ye, Y., Wang, Z., He, Y., Li, Y., & Mao, H. (2020). High expression of fibronectin 1 indicates poor prognosis in gastric cancer. *Oncology Letters*, *19*(1), 93–102. https://doi.org/10.3892/ol.2019.11088

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.

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

Tempero, M. A., Malafa, M. P., Behrman, S. W., Benson, A. B., Casper, E. S., Chiorean, E. G., Chung, V., Cohen, S. J., Czito, B., & Engebretson, A. (2014). Pancreatic adenocarcinoma, version 2.2014. *Journal of the National Comprehensive Cancer Network*, *12*(8), 1083–1093.

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

Tyagi, S., Gupta, P., Saini, A. S., Kaushal, C., & Sharma, S. (2011). The peroxisome proliferator-activated receptor: A family of nuclear receptors role in various diseases. *Journal of Advanced Pharmaceutical Technology & Research*, *2*(4), 236–240. https://doi.org/10.4103/2231-4040.90879

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

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.

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

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.

Zhou, W., Sokoll, L. J., Bruzek, D. J., Zhang, L., Velculescu, V. E., Goldin, S. B., Hruban, R. H., Kern, S. E., Hamilton, S. R., & Chan, D. W. (1998). Identifying markers for pancreatic cancer by gene expression analysis. *Cancer Epidemiology and Prevention Biomarkers*, *7*(2), 109–112.

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