



# Gene Expression Characteristics of Tumor and Adjacent Non-Tumor Tissues of Pancreatic Ductal Adenocarcinoma (PDAC) In-Silico

Emine Güven<sup>1\*</sup>

<sup>1</sup>Department of Biomedical Engineering, Engineering Faculty, Düzce University, Yörük/Düzce Merkez/Düzce, 81620, Turkey

\*Corresponding author: Emine Güven. Department of Biomedical Engineering, Faculty of Engineering, Duzce University, Duzce, Turkey. Tel/Fax: +90-3185421037, E-mail: [emine.guven@duzce.edu.tr](mailto:emine.guven@duzce.edu.tr)

**Background:** One of the deadliest and most prevalent cancer is pancreatic ductal adenocarcinoma (PDAC). Microarray has become an important tool in the research of PDAC genes and target therapeutic drugs.

**Objectives:** This study intends to clarify the promising prognostic and biomarker targets in PDAC using GSE78229 and GSE62452 datasets, publicly accessible at the Gene Expression Omnibus database.

**Materials and Methods:** Utilizing GEOquery, Bio base, gplots, and ggplot2 packages in the R program, this study detects 428 differentially expressed genes that are further applied to build a co-expression network by the weighted correlation network analysis (WGCNA). The turquoise module presented a higher correlation with PDAC progression. 79 candidate genes were selected based on the co-expression and protein-protein interaction (PPI) networks. In addition, the functional enrichment analysis was studied.

**Results:** Five significant KEGG pathways linked to PDAC were detected, in which the endoplasmic reticulum protein processing pathway was remarked to be vital. The resulting 19 hub genes as HSPA4, PABPC1, HSP90B1, PPP1CC, USP9X, EIF2S3, MSN, RAB10, BMP2, P4HB, UBC, B2M, SLC25A5, MMP7, SPTBN1, RALB, DNAJB1, CENPE, and PDIA6 were identified by the Network Analyst web tool founded on PPI network by the STRING. These were identified as the most connected hub proteins. The quantification of the expression of levels and survival probabilities were analyzed overall survival (OS) of the real hub genes and were investigated by Kaplan–Meier (KM) plotter through The Cancer Genome Atlas Program (TCGA) database.

**Conclusions:** The protein-protein interactions and KEGG pathway enrichment by DAVID indicated that some pathways were involved in PDAC, such as “pathways in cancer (hsa05200)”, “protein processing in the endoplasmic reticulum (hsa04141)”, “antigen processing and presentation (hsa04612)”, “dopaminergic synapse (hsa04728)”, and “measles (hsa05162)”; in which these pathways, the “protein processing in endoplasmic reticulum (hsa04141)”, was further studied because of its closely relationship with PDAC. The rest of the hub genes reviewed throughout the study might be promising targets for diagnosing and treating PDAC and relevant diseases.

**Keywords:** Biomarker; Co-expression; Differentially expressed genes; Endoplasmic reticulum; Gene ontology pathway enrichment; Pancreatic ductal adenocarcinoma

## 1. Background

Pancreatic ductal adenocarcinoma (PDAC), one of the deadliest cancer types, generally has a five-year viability fraction of 3% because of diagnosis at a late grade (1). Even though there are advanced treatment procedures such as pancreatectomy, radiotherapy, adjuvant and

neoadjuvant chemotherapies, and palliative care pancreatectomy remains the most effective treatment, particularly for the initial stage of pancreatic cancer (2, 3, 4). Therefore, an updated understanding of pancreatic cancer is fundamental, and a straightforward mechanism is essential for personalized and curable

therapies to improve patient survival.

The generation of countless gene expression profiles of pathological samples has been directed by the advancement of high throughput sequencing that is publicly reachable via the GEO database (5). Whereas only a small part of these datasets has been studied, the different facets of the machinery of pancreatic tumor fast growth and resilience to therapies should be on focus. The deposited datasets are re-analyzed and used to offer beneficial outcomes for further examination in-silico.

The protein expression changes in the progression and expansion of PDAC and associated diseases demand extensive research (6). Moreover, the relations among DEGs detected co-expression network construction, particularly protein-protein interaction (PPI) networks, should elucidate underlying signaling pathways. (7, 8). By investigating hub nodes globally and between tumor and healthy samples of the most significant co-expression module of DEGs and forming PPI networks, this study aims to uncover the biological pathway and genetic mechanisms of PDAC and associated diseases growth.

The experimental studies (9, 10) have concentrated only on detecting the significant genes so far. Thereby, two GEO datasets are studied which contained paired samples of tumor and adjacent non-tumor samples. The analysis provided the detection of the DEGs by the co-expression network analysis. GO and KEGG pathway enrichment investigation were subsequently studied to examine the biological process, cellular component, and molecular function of the pathways, genes, and proteins. Moreover, a PPI network was constructed, and the associated pathway was investigated to detect the hub genes of PDAC.

## 2. Objectives

In this study, two GEO datasets were chosen, which included tumor and adjacent non-tumor tissues of the microarray expression datasets. The weighted

correlation network analysis was utilized as an exploratory tool, while gene filtering approach was utilized to spot the clusters (modules) of highly correlated genes based on DEGs. GO and KEGG pathway enrichments was studied, and biological process, cellular component, and molecular function of the pathways were highlighted. The PPI network revealed the real hub genes could assist as promising targets for the therapies of PDAC. This study aims to contribute further understanding of the machinery of PDAC development and its subsequent core genes.

## 3. Materials and Methods

### 3.1. The Gene Expression Datasets

The expression datasets of mRNA of human pancreatic cancer were downloaded from NCBI GEO database with the accession GSE78229 and GSE62452 (11, 12). **Table 1** illustrates a summary of the datasets. The datasets contain an entirety of expression of 33,297 probes of 111 samples, i.e., 50 tumors and 61 adjacent normal tissues.

### 3.2. Gene Expression Data Analysis Codes

Analysis was performed in the R programming language that can be reached at GitHub repository. Before conducting the analyses, the datasets with the low quality and low number of reads were filtered out, while the remainder of expression set was converted to a base-2 logarithmic measure. The expression values were then normalized by taking the averages of the samples.

### 3.3. Differentially Expressed Genes

Prior to transformation of gene expression values to base-2 logarithmic scale, datasets were pulled out by utilizing GEOquery package in Bioconductor (13). The statistical significance threshold is fixed at p-value  $< 0.05$  and  $|\log_2(\text{fold cut-off})| > 5$  to detect DEGs between each data set via t-test for further analysis.

**Table 1.** Summary of the PDAC data sets.

| GEO data set | Samples                      | Expressed Probes                                    |
|--------------|------------------------------|---|
| GSE78229     | 50 PDAC tumor tissue         | 33,297 probes on Affymetrix Human Gene 1.0 ST Array |
| GSE62452     | 61 adjacent non-tumor tissue | 33,297 probes on Affymetrix Human Gene 1.0 ST Array |

The study utilized `gplots` package using `heatmap.2` function to construct heatmaps of DEGs (14)

### 3.4. Weighted Co-Expression Network Construction

The differentially expressed genes with presented expression values were used to identify the scale-free gene modules of co-expression and highly correlated genes created by WGCNA. The soft threshold power  $\beta$  was set to 5, the lowest power based on the scale-free topology to build a weighted gene network (15). In addition, the parameter `blockSize` assigned to 30 and `TOMType` were not given. Lastly, the topological overlap matrix (TOM) was computed by adjacency transformation and chose the value (1-TOM) to the distance for detecting hierarchical clustering genes and modules.

### 3.5. GO and KEGG Enrichments of the Pathways

Prior to gene expression measurements of annotations for the DEGs in hub module, probe IDs were matched to the official gene symbols utilizing Biomart package (16) in R program. Subsequently, GO annotations of biological processes, molecular functions, and cellular components via Database for Annotation, Visualization and Integrated Discovery 6.8 (DAVID 6.8) were studied (16). Each annotation type was retrieved utilizing DAVID and KEGG (17, 18). All annotated pathways were carefully reviewed and further partitioned according to the characteristics of their biological and molecular meanings. Bar plot of GO and KEGG pathways were created using `ggplot` function of `ggplot2` package (19).

### 3.6. The PPI Network

Network Analyst offers the study of the PPI networks for the DEGs in hub module and expression values, utilizing STRING Interactome (20). To verify the outcomes, DAVID is matched with Network Analyst enrichments performed with KEGG (21).

### 3.7. Survival Analysis

Survival analysis via Kaplan–Meier (KM) plotter patients from The Cancer Genome Atlas Program (TCGA) dataset (22) was conducted. The examination was performed using common hub genes counting on expression profiles values in PDAC datasets.

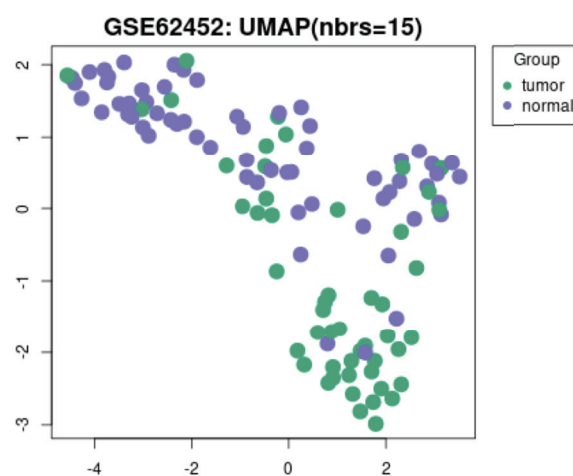
## 4. Results

### 4.1. Exploratory Data Analysis

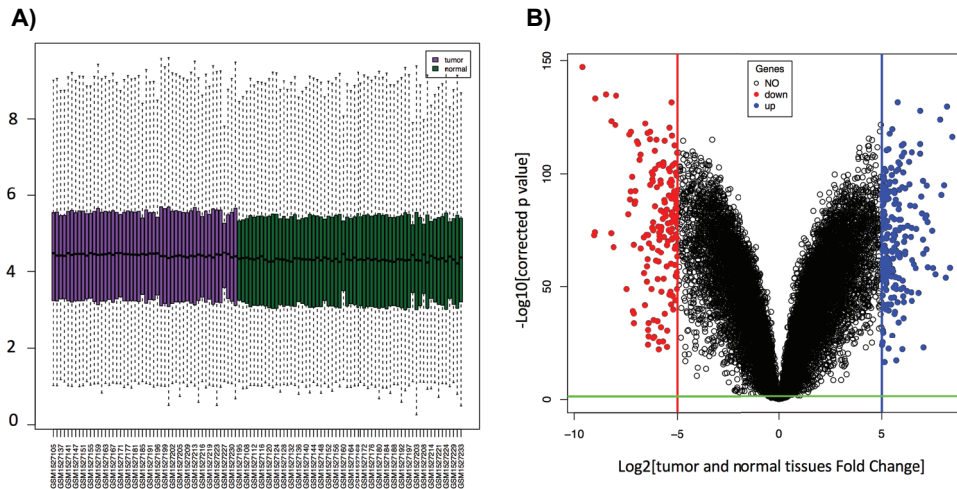
**Figure 1** shows combined microarray datasets to examine quality control via Uniform Manifold Approximation and Projection (UMAP), which is a dimension reduction method useful for visualizing clusters or groups of samples and relative proximities. The number of nearest neighbors used in the calculation is indicated in the plot (23). **Figure 2A** illustrates the boxplot of the non-normalized base-2 logarithmic transformation of the raw gene expression datasets with tumor tissue and adjacent normal tissue samples. The examination detected 207 up-regulated DEGs and 221 down-regulated DEGs, which were shown with a volcano plot (**Fig. 2B**). The expression values were used to confer the tumor and normal tissues separation, and a heatmap was designed. (**Fig. S1**)

### 4.2. Weighted Co-Expression Network Construction and Detection of Significant Modules

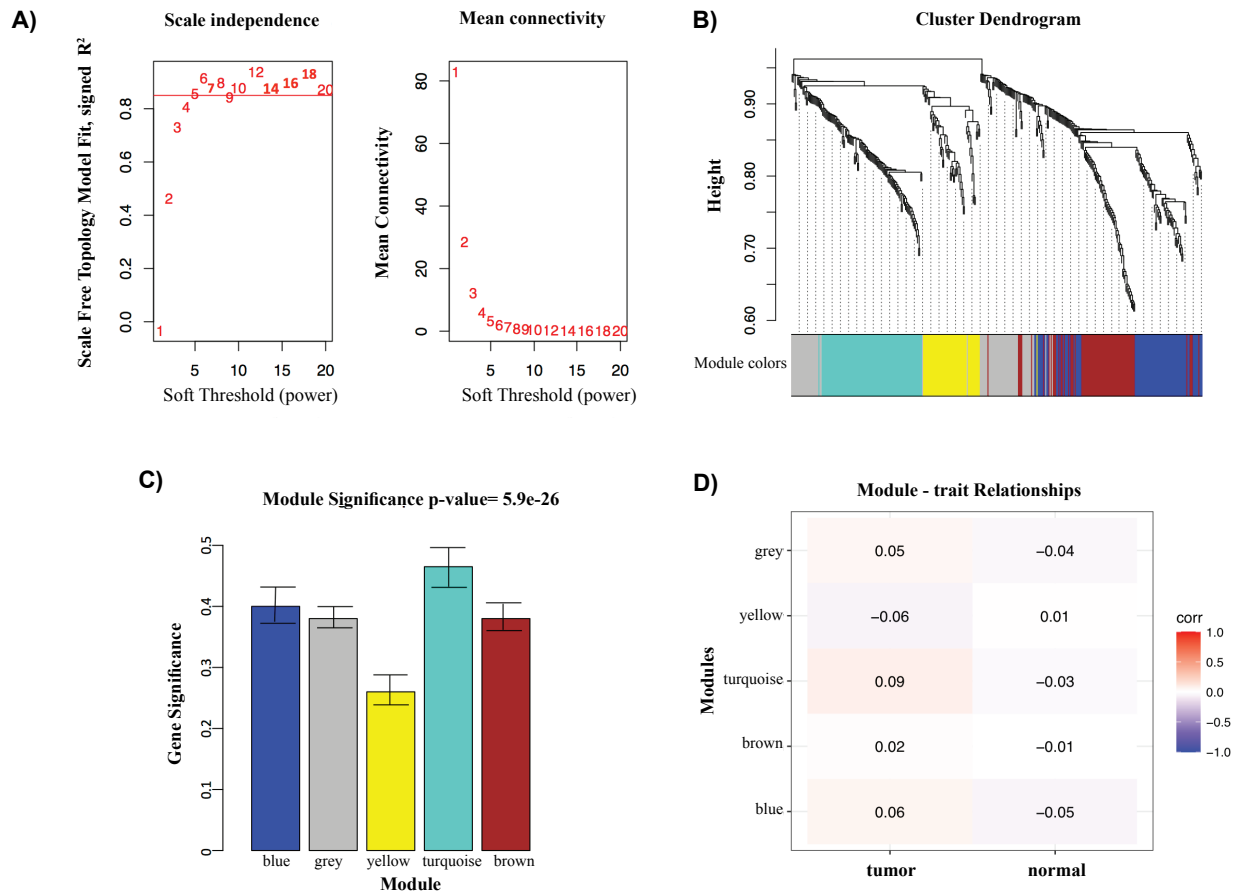
To provide a scale-independent network, a power of  $\beta = 5$  (unscaled  $R^2 = 0.85$ ) was chosen (**Fig. 3A**). Based on tumor and normal tissues of PDAC, by setting DEGs with similar expression types into modules utilizing mean linkage-clustering method and five modules were found (**Fig. 3B**). The study applied two methodologies to examine the connection among each module and DEGs in the PDAC. Module extracts of the turquoise module revealed a higher correlation with disease progression than other modules (**Fig. 3C**).



**Figure 1.** The raw GSE62452 and GSE78229 microarray datasets subject to quality control to examine high-dimensional structure of datasets using UMAP.



**Figure 2.** A) The boxplot explains the gene expression values of samples of the original datasets without normalization. B) Plots displaying the expression difference in PDAC tumor and normal tissues. Black illustrates no change (NO), red illustrates low-expressed (Down), and blue represents over-expressed (Up) DEGs, FC, fold change.



**Figure 3.** A) Ascertainment of soft-thresholding power and the examination of the scale-free fit index for several soft-thresholding powers. B) Dendrogram of the differentially expressed genes. The DEGs are clustered based on a dissimilarity measure (1- TOM) highly clustered in the turquoise module. C) Distribution of mean gene significance and errors in the modules linked with the PDAC traits. D) Heatmap of the correlation between module eigengenes and the tumor and normal samples (traits) of the PDAC.

In addition, the turquoise module was positively associated with tumor tissues but was negatively associated with adjacent normal tissues (blue shading) (**Fig. 3D**). Thus, this study distinguished the turquoise module of PDAC as the most connected to each trait (tumor vs. normal) based on these two methods.

#### 4.3. GO Enrichment and KEGG Pathways

In **Table 2**, the top DEGs in turquoise module were ranked according to the gene counts and  $p$ -value < 0.05 criteria. The common hub DEGs mainly enriched were associated with GO:0000904~cell morphogenesis involved in differentiation GO:0007409~axonogenesis, GO:0016192~vesicle-mediated transport, GO:0061564~axon development, and GO:0000902~cell morphogenesis in the 'biological process' group (**Fig. 4A**). As for the 'molecular function' group GO:0050839~cell adhesion molecule binding, GO:0032403~protein complex binding, GO:0097367~carbohydrate derivative binding, GO:0005543~phospholipid binding, and GO:0008289~lipid binding (**Fig. 4B**) were identified. Furthermore, the 'cellular component' gene ontology enrichment analysis revealed pathways such as GO:0070062~extracellular exosome, GO:1903561~extracellular vesicle, GO:0043230~extracellular organelle, GO:0044421~extracellular region part, and GO:0005912~adherens junction (**Fig. 4C**).

KEGG pathway study results revealed which common hub genes in turquoise module were considerably enriched in has05200: pathways in cancer, hsa04141: protein processing in endoplasmic reticulum, hsa04612: antigen processing and presentation, hsa04728: dopaminergic synapse, and hsa05162: measles as shown in **Figure 4D**. Among these pathways, ER protein processing pathway might have a vital influence on multiple protein process.

#### 4.4. The PPI Network and KEGG Enrichment Analysis

**Figure 4E** demonstrates the PPI network of the hub genes of the turquoise module in the PDAC tumor and adjacent non-tumor tissues. According to the STRING database, all the genes in the turquoise module of the 449 nodes and 559 edges were represented by red to yellow color, relative to the degree of connectivity in the network of the weighted gene co-expression. Subsequently, the significant proteins in the PPI network were selected as the mutual genes to be subject to KEGG enrichment analysis. The KEGG pathway analysis reveals involvement protein processing in endoplasmic reticulum,

mRNA surveillance pathway, estrogen signaling pathway, measles, and cellular senescence. This confirms the role of ER stress in protein processing pathway of the common hub genes. HSPA4, PABPC1, HSP90B1, PPP1CC, USP9X, EIF2S3, MSN, RAB10, BMPR2, P4HB, UBC, B2M, SLC25A5, MMP7, SPTBN1, RALB, DNAJB1, CENPE, and PDIA6 were identified as the most connected hub proteins (**Tables 2 and 3**).

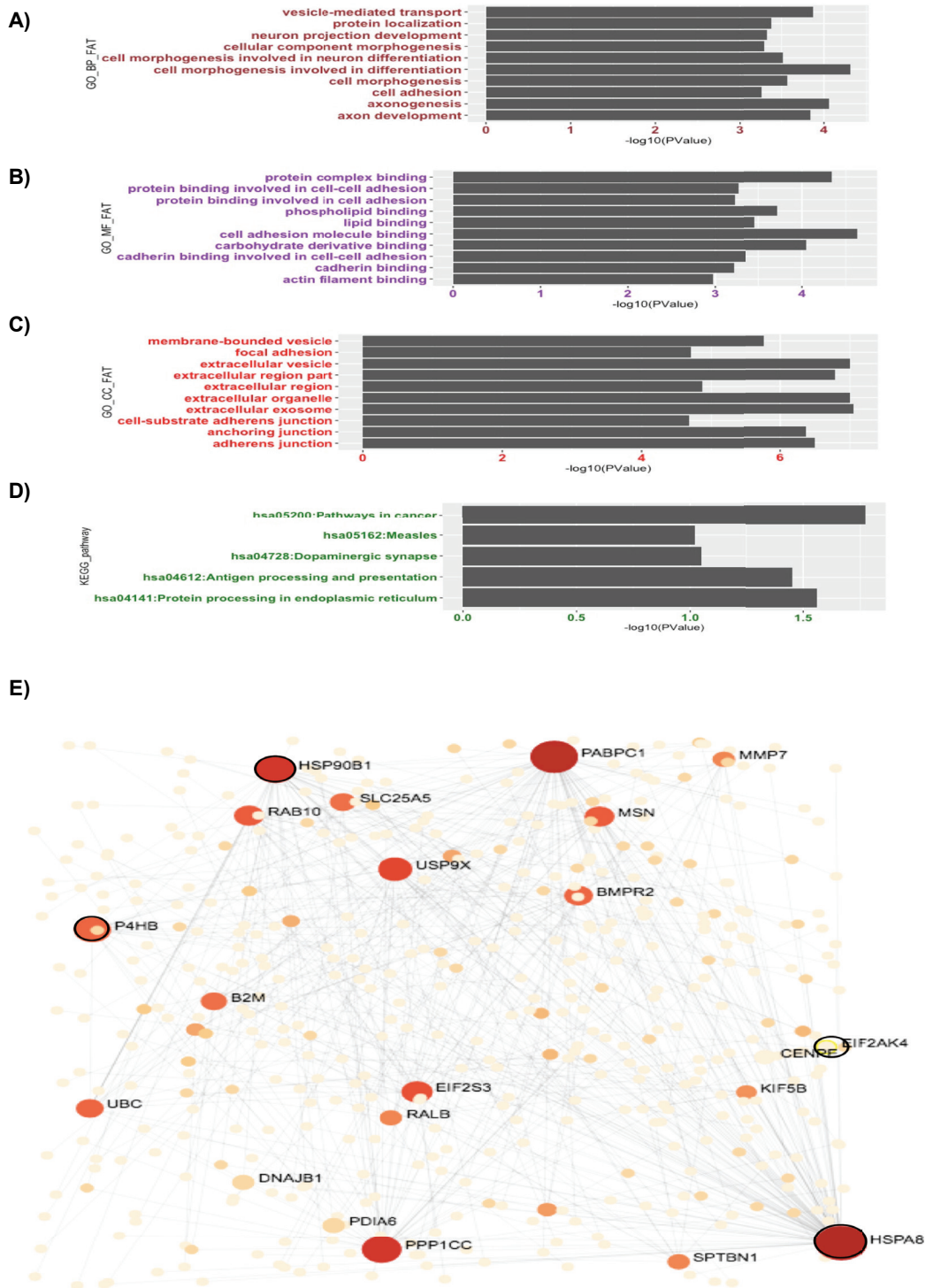
#### 4.5. The Role of the Endoplasmic Reticulum Protein Processing Pathway

The turquoise module, in which 19 common network genes reveal the "ER protein processing" pathway through KEGG enrichment analysis, is strongly linked to pancreatic cancer and other related diseases. ER stress is a disparity within the protein-folding capacity of ER and its protein pack that results from the collection of cranking proteins (24, 25). It has been considered to be engaged in the Parkinson's, Alzheimer's, and other deformational diseases. These diseases are caused by a few particular morbidic unfolding proteins (26, 27).

Primary hub genes investigation associated with ER protein processing pathway was performed, as presented in **Figure 4E**. The network is constructed based on specifically engaged hub genes with black circled nodes in ER protein processing pathway. The study hypothesized that by virtue of over-expression of ER-associated proteins, the entire ER protein processing pathway might be unsettled in PDAC. The key genes of ER protein processing pathway can be listed as HSPA4, HSP90B1, P4HB, and EIF2AK4. The most expressed HSP gene family performs a key position of the ER protein processing pathway. HSP90B1 would be a gene that is related with this pathway, folding and transforming molecular chaperones with key roles in organizing other proteins (28). Other common hub genes are listed as HSPA4 and P4HB in PDAC tumor and normal tissues. These findings verify the vital duty of the ER protein processing pathway engaged in PDAC and associated diseases treatment, proposing updated molecular biomarkers to the essential drug agents.

#### 4.6. Survival Analysis

The datasets containing 50 PDAC tumor tissues and 61 adjacent normal tissues were separately subjected to survival analysis in the TCGA database. To anticipate the prognostic values of the eight real hub genes, KM plotter was employed (**Fig. 5A-G**).



**Figure 4.** The bar plot shows 30 top GO annotations regarding **A)** biological processes (BP), **B)** molecular function (MF), and **C)** cellular component (CC). They are represented by axes text color brown, purple, and red respectively. **D)** The top five KEGG pathway enrichments of the turquoise module is shown. **E)** The gradual color change from red to the yellow node was comparable to the degree of connectivity in the weighted gene co-expression network; positive correlation in the red, and negative correlation in the yellow nodes. Black-circled nodes present the ER protein processing pathway associated genes.

**Table 2.** Gene expression dataset retrieved with top significant pathways GO annotations analysis of the turquoise module in PDAC

| Category  | Term   | Count | %     | Pvalue   | Genes   |
|-----------|--|-------|-------|----------|---|
| GO_BP_FAT | GO:0000904~cell morphogenesis involved in differentiation        | 12    | 13.33 | 4.89E-05 | RAB10, BMPR2, USP9X, KIF5B, SLITRK5, DNAJB1, CENPE, CTNNA2, MT3, ANTXR1, SPTBN1, CORO1C   |
| GO_BP_FAT | GO:0007409~axonogenesis  | 9     | 10    | 8.67E-05 | RAB10, BMPR2, USP9X, KIF5B, SLITRK5, CENPE, CTNNA2, MT3, SPTBN1   |
| GO_BP_FAT | GO:0016192~vesicle-mediated transport                            | 16    | 17.78 | 1.34E-04 | CDC42SE1, HSPA4, RALB, COG6, EIF2AK1, MSN, SNX31, HSP90B1, CORO1C, JCHAIN, RAB10, MYO1B, GOLPH3, KIF5B, B2M, SPTBN1                         |
| GO_BP_FAT | GO:0061564~axon development                                      | 9     | 10    | 1.45E-04 | RAB10, BMPR2, USP9X, KIF5B, SLITRK5, CENPE, CTNNA2, MT3, SPTBN1   |
| GO_BP_FAT | GO:0000902~cell morphogenesis                                    | 14    | 15.56 | 2.72E-04 | CDC42SE1, BMPR2, USP9X, MSN, DNAJB1, CENPE, ANTXR1, MT3, CORO1C, RAB10, KIF5B, SLITRK5, CTNNA2, SPTBN1                                      |
| GO_BP_FAT | GO:0048667~cell morphogenesis involved in neuron differentiation | 9     | 10    | 3.03E-04 | RAB10, BMPR2, USP9X, KIF5B, SLITRK5, CENPE, CTNNA2, MT3, SPTBN1   |
| GO_BP_FAT | GO:0008104~protein localization                                  | 20    | 22.22 | 4.20E-04 | PDIA6, HSPA4, RALB, COG6, NCOA4, MSN, SNX31, MT3, UBAC2, HSP90B1, CORO1C, RAB10, GOLPH3, KIF5B, RAB18, SLC25A5, GLUL, PAM, SPTBN1, FBN1     |
| GO_BP_FAT | GO:0031175~neuron projection development                         | 11    | 12.22 | 4.69E-04 | RAB10, BMPR2, USP9X, KIF5B, SLITRK5, DNAJB1, CENPE, CTNNA2, B2M, MT3, SPTBN1  |
| GO_BP_FAT | GO:0032989~cellular component morphogenesis                      | 14    | 15.56 | 5.02E-04 | CDC42SE1, BMPR2, USP9X, MSN, DNAJB1, CENPE, ANTXR1, MT3, CORO1C, RAB10, KIF5B, SLITRK5, CTNNA2, SPTBN1                                      |
| GO_BP_FAT | GO:0007155~cell adhesion   | 16    | 17.78 | 5.45E-04 | PDIA6, HSPA4, MSN, DNAJB1, ANTXR1, CORO1C, RAB10, MYO1B, GOLPH3, VCAN, EIF2S3, KIF5B, CTNNA2, B2M, SPTBN1, FBN1                             |
| GO_BP_FAT | GO:0000904~cell morphogenesis involved in differentiation        | 12    | 13.33 | 4.89E-05 | RAB10, BMPR2, USP9X, KIF5B, SLITRK5, DNAJB1, CENPE, CTNNA2, MT3, ANTXR1, SPTBN1, CORO1C   |
| GO_MF_FAT | GO:0050839~cell adhesion molecule binding                        | 10    | 11.11 | 2.36E-05 | RAB10, HSPA4, MYO1B, EIF2S3, KIF5B, MSN, CTNNA2, P4HB, SPTBN1, FBN1   |
| GO_MF_FAT | GO:0032403~protein complex binding                               | 12    | 13.33 | 4.59E-05 | HSPA4, PDIA6, PPP1CC, MYO1B, PLS3, CTNNA2, P4HB, ANTXR1, SPTBN1, CORO1C, FBN1, JCHAIN   |
| GO_MF_FAT | GO:0097367~carbohydrate derivative binding                       | 20    | 22.22 | 8.89E-05 | HSPA4, RALB, BMPR2, MMP7, EIF2AK1, HSP90B1, JCHAIN, RAB10, MYO1B, SMCHD1, VCAN, EIF2S3, KIF5B, PDE3A, RAB18, YME1L1, GLUL, TRPM6, B2M, FBN1 |

|           |  |    |       |          |  |
|-----------|--|----|-------|----------|--|
| GO_MF_FAT | GO:0005543~phospholipid binding                            | 8  | 8.89  | 1.92E-04 | HSPA4, GOLPH3, MYO1B, ANXA5, PON1, ANXA2P2, SNX31, SPTBN1  |
| GO_MF_FAT | GO:0008289~lipid binding                                   | 10 | 11.11 | 3.48E-04 | HSPA4, GOLPH3, MYO1B, FABP7, ANXA5, PON1, ANXA2P2, SNX31, CD1B, SPTBN1   |
| GO_MF_FAT | GO:0098641~cadherin binding involved in cell-cell adhesion | 7  | 7.78  | 4.50E-04 | RAB10, HSPA4, MYO1B, EIF2S3, KIF5B, CTNNA2, SPTBN1   |
| GO_MF_FAT | GO:0098632~protein binding involved in cell-cell adhesion  | 7  | 7.78  | 5.48E-04 | RAB10, HSPA4, MYO1B, EIF2S3, KIF5B, CTNNA2, SPTBN1   |
| GO_MF_FAT | GO:0098631~protein binding involved in cell adhesion       | 7  | 7.78  | 5.98E-04 | RAB10, HSPA4, MYO1B, EIF2S3, KIF5B, CTNNA2, SPTBN1   |
| GO_MF_FAT | GO:0045296~cadherin binding                                | 7  | 7.78  | 6.08E-04 | RAB10, HSPA4, MYO1B, EIF2S3, KIF5B, CTNNA2, SPTBN1   |
| GO_MF_FAT | GO:0051015~actin filament binding                          | 5  | 5.56  | 0.001071 | MYO1B, PLS3, CTNNA2, ANTXR1, CORO1C  |
| GO_CC_FAT | GO:0070062~extracellular exosome                           | 30 | 33.33 | 8.98E-08 | C5ORF46, RALB, PON1, ANTXR1, N4BP2L2, HSP90B1, JCHAIN, GLUL, B2M, SPTBN1, PDIA6, HSPA4, PARP4, MMP7, ANXA5, MSN, DNAJB1, GNG12, RAB10, MYO1B, EIF2S3, SUB1, RAB18, SERINC1, P4HB, PABPC1, ANXA2P2, SLC25A5, PAM, FBN1                                |
| GO_CC_FAT | GO:1903561~extracellular vesicle                           | 30 | 33.33 | 1.01E-07 | C5ORF46, RALB, PON1, ANTXR1, N4BP2L2, HSP90B1, JCHAIN, GLUL, B2M, SPTBN1, PDIA6, HSPA4, PARP4, MMP7, ANXA5, MSN, DNAJB1, GNG12, RAB10, MYO1B, EIF2S3, SUB1, RAB18, SERINC1, P4HB, PABPC1, ANXA2P2, SLC25A5, PAM, FBN1                                |
| GO_CC_FAT | GO:0043230~extracellular organelle                         | 30 | 33.33 | 1.01E-07 | C5ORF46, RALB, PON1, ANTXR1, N4BP2L2, HSP90B1, JCHAIN, GLUL, B2M, SPTBN1, PDIA6, HSPA4, PARP4, MMP7, ANXA5, MSN, DNAJB1, GNG12, RAB10, MYO1B, EIF2S3, SUB1, RAB18, SERINC1, P4HB, PABPC1, ANXA2P2, SLC25A5, PAM, FBN1                                |
| GO_CC_FAT | GO:0044421~extracellular region part                       | 35 | 38.89 | 1.63E-07 | C5ORF46, RALB, Bmpr2, PON1, ANTXR1, N4BP2L2, HSP90B1, JCHAIN, GLUL, B2M, SPTBN1, PDIA6, HSPA4, CPA1, PARP4, MMP7, ANXA5, MSN, DNAJB1, CENPE, GNG12, MT3, RAB10, MYO1B, VCAN, EIF2S3, SUB1, RAB18, SERINC1, P4HB, PABPC1, ANXA2P2, SLC25A5, PAM, FBN1 |
| GO_CC_FAT | GO:0005912~adherens junction                               | 15 | 16.67 | 3.22E-07 | HSPA4, ANXA5, MSN, HSP90B1, CORO1C, RAB10, PPP1CC, MYO1B, EIF2S3, KIF5B, P4HB, PABPC1, CTNNA2, B2M, SPTBN1   |

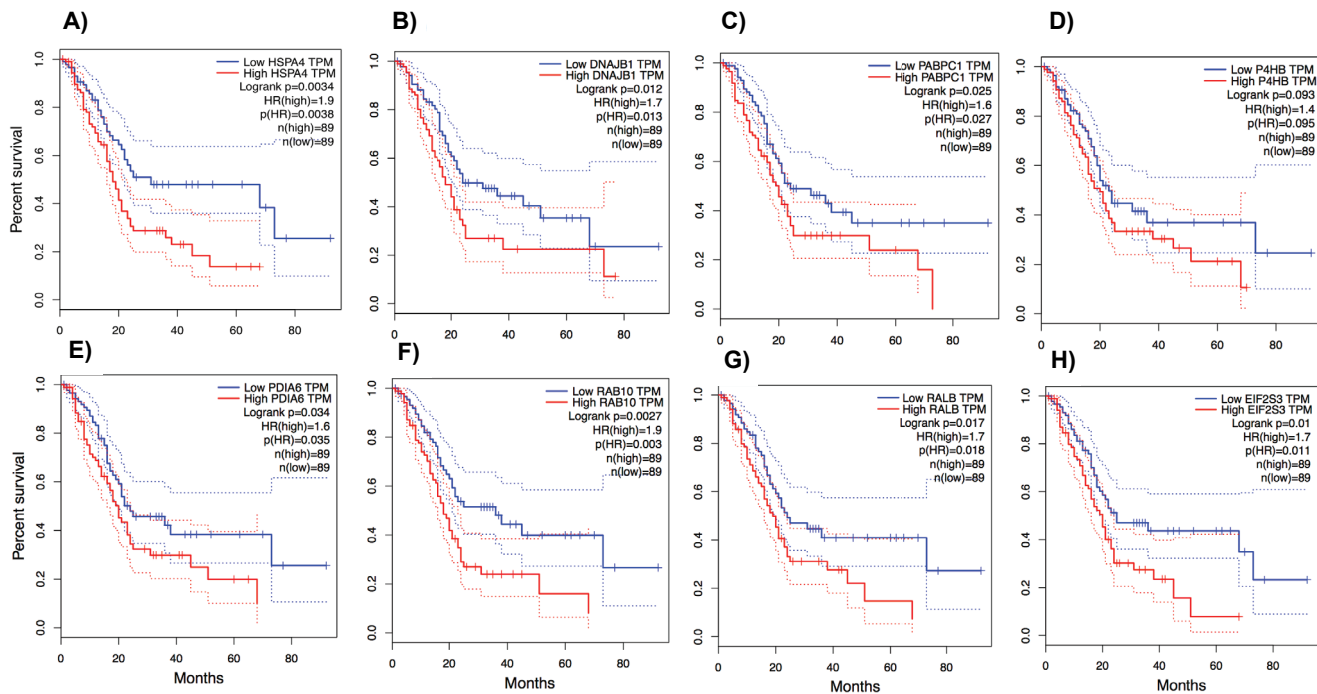


|              |  |    |       |          |  |
|--------------|--|----|-------|----------|--|
| GO_CC_FAT    | GO:0070161~anchoring junction                        | 15 | 16.67 | 4.32E-07 | HSPA4, ANXA5, MSN, HSP90B1, CORO1C, RAB10, PPP1CC, MYO1B, EIF2S3, KIF5B, P4HB, PABPC1, CTNNA2, B2M, SPTBN1   |
| GO_CC_FAT    | GO:0031988~membrane-bounded vesicle                  | 32 | 35.56 | 1.70E-06 | C5ORF46, RALB, PON1, ANTXR1, N4BP2L2, HSP90B1, JCHAIN, KIF5B, GLUL, B2M, SPTBN1, PDIA6, HSPA4, PARP4, MMP7, ANXA5, MSN, DNAJB1, GNG12, MT3, RAB10, MYO1B, EIF2S3, SUB1, RAB18, SERINC1, P4HB, PABPC1, ANXA2P2, SLC25A5, PAM, FBN1                    |
| GO_CC_FAT    | GO:0005576~extracellular region                      | 35 | 38.89 | 1.35E-05 | C5ORF46, RALB, BMPR2, PON1, ANTXR1, N4BP2L2, HSP90B1, JCHAIN, GLUL, B2M, SPTBN1, PDIA6, HSPA4, CPA1, PARP4, MMP7, ANXA5, MSN, DNAJB1, CENPE, GNG12, MT3, RAB10, MYO1B, VCAN, EIF2S3, SUB1, RAB18, SERINC1, P4HB, PABPC1, ANXA2P2, SLC25A5, PAM, FBN1 |
| GO_CC_FAT    | GO:0005925~focal adhesion                            | 10 | 11.11 | 1.94E-05 | RAB10, HSPA4, PPP1CC, ANXA5, MSN, P4HB, PABPC1, B2M, HSP90B1, CORO1C   |
| KEGG_PATHWAY | hsa05200:Pathways in cancer                          | 6  | 6.67  | 0.0168   | RALB, NCOA4, DNAJB1, CTNNA2, GNG12, HSP90B1  |
| KEGG_PATHWAY | hsa04141:Protein processing in endoplasmic reticulum | 4  | 4.44  | 0.0276   | HSPA4, EIF2AK1, P4HB, HSP90B1  |
| KEGG_PATHWAY | hsa04612:Antigen processing and presentation         | 3  | 3.33  | 0.0353   | HSPA4, PDIA6, B2M  |
| KEGG_PATHWAY | hsa04728:Dopaminergic synapse                        | 3  | 3.33  | 0.0891   | PPP1CC, KIF5B, GNG12   |
| KEGG_PATHWAY | hsa05162:Measles                                     | 3  | 3.33  | 0.0951   | HSPA4, EIF2AK1, MSN  |

## 5. Discussion

The predominance of PDAC and the associated survival rates have decreased in the former decade (29). Therefore, accurate and immediate diagnoses of the PDAC and the improvement of paramount remedies are vital. Previous research identified actual hub genes in PDAC that were declared to be of diagnostic importance. (8). This study comprehensively analyzed the combination of GSE78229 and GSE62452 datasets from patients with PDAC of microarray gene expression profiles holding the expressions of 50 tumor and 61 non-tumor tissues (**Fig. 2** and **Table 1**). In this study, the DEGs in the datasets were subject to create a co-expression network by WGCNA (15). After which, GO functional enrichments and KEGG pathway analyses were performed on the turquoise module, which was

revealed as the most significant module (**Tables 2, 3,** and **Fig. 3**). KEGG pathway study resulted mostly in the turquoise module was entailed in protein processing in endoplasmic reticulum, mRNA surveillance pathway, estrogen signaling pathway, measles, and cellular senescence. (**Fig. 4D** and **Table 2**). The DEGs in turquoise module are related to other pathways such as cell morphogenesis involved in differentiation (30), vesicle-mediated transport (31), axon guidance development (32), and cell morphogenesis (30). These group pathways might be of importance in the 'biological process' (**Fig. 4A**). To obtain an in-depth understanding of these common hub genes, this study analyzed the constructed PPI network (**Fig. 4E**), which revealed eight real hub genes such as *HSPA4*, *DNAJB1*, *PABPC1*, *P4HB*, *PDIA6*, *RAB10*, *RALB*, *EIF2S3*, and *CELPE*.



**Figure 5.** Survival study of the real hub genes in PDAC datasets. Effect of expression levels on PDAC patients of survival. **A)** *HSPA4* (Logrank p = 0.0034), **B)** *DNAJB1* (Logrank p = 0.012), **C)** *PABBC1* (Logrank p = 0.025) **D)** *P4HB* (Logrank p = 0.093), **E)** *PDIA6* (Logrank p = 0.034), **F)** *RAB10* (Logrank p = 0.0027), **G)** *RALB* (Logrank p = 0.017), and **H)** *EIF2S3* (Logrank p = 0.01). Red lines and blue lines represent high expression and low expression of the real hub genes, respectively.

These hub genes are strongly linked with progression of PDAC and they may serve as therapeutic targets. Moreover, mutations employed in ER protein processing pathway genes and several associated pathway genes contribute to the growth of the PDAC (**Fig. 4E**). Metabolic and functional changes in ribosomal and ER protein processing pathways are evident in pancreatic cancer (33, 34). In addition to their standard roles, ER protein processing pathway further rules metabolism characteristics of aggregation of misfolded proteins in the endoplasmic reticulum, which drives ER stress and then triggers the unfolded protein response (UPR) signaling pathway (35). Heat Shock Protein Family A (Hsp70) Member 4 (*HSPA4*) was detected as one of the core genes with the biggest degree of connectivity (**Table 3**). One study revealed that *HSPA4* expression is associated with increased patient survival in PDAC (36) DnaJ Heat Shock Protein Family (Hsp40) Member B1 (*DNAJB1*) is a protein-coding gene that could serve as an adverse prognostic factor for overall survival

and relapse-free survival (37). Overexpression of *DNAJB1* is linked to progression and recurrence of cholangiocarcinoma (38). Furthermore, it is reported by Cui X, *et al.* that *DNAJB1* was able to promote cancer cell proliferation in the lung cancer cell line A549 and suppress apoptosis through ubiquitin degradation of *PDCD5* (39). Taken together, the weighted gene co-expression analysis results of two GEO microarray datasets of PDAC indicated that dopaminergic synapse (*hsa94728*) (40), antigen processing and presentation (*hsa04612*) (41), and protein processing in ER (*hsa04141*) (42) pathways participate in the onset and development of PDAC. This project applied survival study to filter real hub genes with a significant Logrank of a p-value. A total of eight genes (*HSPA4*, *DNAJB1*, *PABBC1*, *P4HB*, *PDIA6*, *RAB10*, *RALB*, and *EIF2S3*) were especially promising, and they may be potential biomarkers for prognosis (**Fig. 5A-G**).

**Table 3.** The top 15 common hub genes of PPI network of the turquoise module in PDAC tumor and non-tumor tissues gene expression data.

| Gene ID | Genes   | Nodes | Betweenness centrality | Expression | Fold Cut-off |
|---------|---------|-------|------------------------|------------|--------------|
| 3312    | HSPA4   | 103   | 38655.36               | 7.26105469 | 7.26         |
| 26986   | PABPC1  | 76    | 30322.54               | 5.54081207 | 5.54         |
| 7184    | HSP90B1 | 47    | 18422.09               | 6.67241036 | 6.67         |
| 5501    | PPP1CC  | 45    | 26263.34               | 5.83112829 | 5.83         |
| 8239    | USP9X   | 31    | 11968.88               | 4.62837378 | 4.62         |
| 1968    | EIF2S3  | 24    | 7240.32                | 5.25950649 | 5.25         |
| 4478    | MSN     | 21    | 18382.39               | 5.39104874 | 5.39         |
| 10890   | RAB10   | 21    | 16158.94               | 3.86893387 | 3.86         |
| 659     | BMPR2   | 19    | 8073                   | 4.48815865 | 4.48         |
| 5034    | P4HB    | 18    | 7982.73                | 6.27703351 | 6.27         |
| 7316    | UBC     | 17    | 61998.78               | 0          | 6.89         |
| 567     | B2M     | 15    | 6307                   | 6.00173613 | 6.00         |
| 292     | SLC25A5 | 15    | 5591.53                | 4.94497369 | 4.94         |
| 4316    | MMP7    | 10    | 9309                   | 5.35016667 | 5.35         |
| 6711    | SPTBN1  | 10    | 3685.11                | 5.38178162 | 5.38         |
| 5899    | RALB    | 10    | 3223.23                | 3.82931216 | 3.82         |
| 3337    | DNAJB1  | 10    | 3750.13                | 3.24341789 | 3.78         |
| 1062    | CENPE   | 9     | 3730                   | 3.35601213 | 5.10         |
| 10130   | PDIA6   | 9     | 4500                   | 4.27824156 | 5.21         |

## 6. Conclusion

An accurate and faster identification of PDAC and the advancement of strong particular remedies are of importance. This study identified 428 DEGs. Afterwards, the DEGs are subject to build a co-expression network by WGCNA. KEGG pathway study of common hub genes in turquoise module was considerably enriched in hsa05200: pathways in cancer, hsa04141: protein processing in endoplasmic reticulum, hsa04612: antigen processing and presentation, hsa04728: dopaminergic synapse, and hsa05162: measles, in which the ER protein processing pathway was remarked to be important. Nineteen hub genes were identified via PPI network. A total of eight genes (HSPA4, DNAJB1, PABPC1, P4HB, PDIA6, RAB10, RALB, and EIF2S3) were confirmed through survival analysis, and they may be potential biomarkers for

prognosis of the PDAC.

## Abbreviations

DEGs: Differentially Expressed Genes  
 PDAC: Pancreatic Adenocarcinoma  
 WGCNA: The Weighted Correlation Network Analysis  
 DAVID: The Database for Annotation, Visualization and Integrated Discovery  
 KEGG: Kyoto Encyclopedia of Genes and Genomes  
 GEO: Gene Expression Omnibus  
 GO: Gene Ontology  
 ER: Endoplasmic Reticulum  
 PPI: Protein-protein Interaction  
 BP: Biological Process  
 MF: Molecular Function  
 CC: Cellular Component

## References

- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics. *CA Cancer J Clin*. 2021;**71**(1):7-33. doi:10.3322/caac.21654
- Gillen S, Schuster T, Büschenfelde CM zum, Friess H, Kleeff J. Preoperative/Neoadjuvant Therapy in Pancreatic Cancer: A Systematic Review and Meta-analysis of Response and Resection Percentages. *PLOS med*. 2010;**7**(4):e1000267. doi:10.1371/journal.pmed.1000267
- Tempero MA, Malafa MP, Behrman SW, Benson AB, Casper ES, Chiorean EG. Pancreatic adenocarcinoma. *J Natl Compr Canc Netw*. 2014;**12**(8):1083-1093.
- Lambert A, Schwarz L, Borbath I, Henry A, VanLaethem J-L, Malka D, et al. An update on treatment options for pancreatic adenocarcinoma. *Ther Adv Med Oncol*. 2019;**2**(11):175883591987556. doi:10.1177/1758835919875568
- Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, et al. NCBI GEO: mining tens of millions of expression profiles-database and tools update. *Nucleic Acids Res*. 2007;**35**(suppl\_1):D760-D765. doi:10.1093/nar/gkl887
- Grützmann R, Boriss H, Ammerpohl O, Lüttges J, Kalthoff H, Schackert HK, et al. Meta-analysis of microarray data on pancreatic cancer defines a set of commonly dysregulated genes. *Oncogene*. 2005;**24**(32):5079-5088. doi:10.1038/sj.onc.1208696
- Clarke C, Madden S F, Doolan P, Aherne S T, Joyce H, O'driscoll L, Clynes M, et al. Correlating transcriptial networks to breast cancer survival: a large-scale coexpression analysis. *Carcinogenesis*. 2013;**34**(10):2300-2308. doi:10.1093/carcin/bgt208
- Zou Z, Cheng Y, Jiang Y, Liu S, Zhang M, Liu J, Zhao Q, et al. Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis. *Int J Biol Sci*. 2018;**14**(2):124. doi:10.7150/ijbs.22619
- Lv K, Yang J, Sun J, Guan J. Identification of key candidate genes for pancreatic cancer by bioinformatics analysis. *Exp Ther Med*. 2019;**18**(1):451-458. doi:10.3892/etm.2019.7619
- Kwon M-S, Kim Y, Lee S, Namkung J, Yun T, Yi SG, et al. Integrative analysis of multi-omics data for identifying multi-markers for diagnosing pancreatic cancer. *BMC Genomics*. 2015;**16**(9):S4. doi:10.1186/1471-2164-16-s9-s4
- Wang J, Yang S, He P, Schetter AJ, Gaedcke J, Ghadimi BM, et al. Endothelial Nitric Oxide Synthase Traffic Inducer (NOSTRIN) is a Negative Regulator of Disease Aggressiveness in Pancreatic Cancer. *Clin Cancer Res*. 2016;**22**(24):5992-6001. doi:10.1158/1078-0432.ccr-16-0511
- Yang S, He P, Wang J, Schetter A, Tang W, Funamizu N. A Novel MIF Signaling Pathway Drives the Malignant Character of Pancreatic Cancer by Targeting NR3C2. *Cancer Res*. 2016;**76**(13):3838-3850. doi:10.1158/0008-5472.can-15-2841
- Konopka T, Konopka MT. R-package: umap. Uniform Manifold Approximation and Projection. 2018.
- Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods*. 2015;**12**(2):115-121. doi:10.1038/nmeth.3252
- Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: Various R programming tools for plotting data. *R Package Version*. 2009;**2**(4):1.
- Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;**4**(1). doi:10.2202/1544-6115.1128
- Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*. 2005;**21**(16):3439-3440. doi:10.1093/bioinformatics/bti525
- Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;**4**(1):44. doi:10.1038/nprot.2008.211
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;**44**(D1):D457-D462. doi:10.1093/nar/gkv1070
- Gómez-Rubio V. ggplot2-elegant graphics for data analysis. *J Stat Softw*. 2017;**77**(1):1-3. doi:10.18637/jss.v077.b02
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;**43**(D1):D447-D52. doi:10.1093/nar/gku1003
- Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J, et al. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res*. 2019;**47**(W1):W234W241. doi:10.1093/nar/gkz240
- Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;**45**(W1):W98-W102. doi:10.1093/nar/gkx247
- Kelly E, Greene CM, Carroll TP, McElvaney NG, O'Neill SJ. Selenoprotein S/SEPS1 modifies endoplasmic reticulum stress in Z variant alpha1-antitrypsin deficiency. *J Biol Chem*. 2009;**284**(25):16891-16897. doi:10.1074/jbc.M109.006288
- Hidvegi T, Schmidt BZ, Hale P, Perlmutter DH. Accumulation of mutant alpha1-antitrypsin Z in the endoplasmic reticulum activates caspases-4 and -12, NFkappaB, and BAP31 but not the unfolded protein response. *J Biol Chem*. 2005;**280**(47):39002-39015. doi:10.1074/jbc.M508652200
- Forman MS, Lee VM-Y, Trojanowski JQ. "Unfolding" pathways in neurodegenerative disease. *Trends Neurosci*. 2003;**26**(8):407-410. doi:10.1016/s0166-2236(03)00197-8
- Gow A, Sharma R. The unfolded protein response in protein aggregating diseases. *Neuromolecular Med*. 2003;**4**(1-2):73-94. doi:10.1385/nmm.4:1-2:73
- Wang, X, Chen, M, Zhou, J, Zhang, X. HSP27, 70 and 90, anti-apoptotic proteins, in clinical cancer therapy. *Int J Oncol*. 2014;**45**(1):18-30. doi:10.3892/ijo.2014.2399.
- Siegel RL, Miller KD, Jemal A. *Cancer statistics*. 2018. *CA Cancer J Clin*. 2018;**68**(1):7-30. doi:10.3322/caac.21442
- Koorstra J B M, Feldmann G, Habbe N, Maitra, A. Morphogenesis of pancreatic cancer: role of pancreatic intraepithelial neoplasia (PanINs). *Langenbecks Arch Surg*. 2008;**393**(4):561-570. doi:10.1007/s00423-008-0282-x
- Moeng S, Son S W, Lee J S, Lee H Y, Kim T H, Choi S Y, Park, J K, et al. Extracellular vesicles (evs) and pancreatic

- cancer: From the role of evs to the interference with ev-mediated reciprocal communication. *Biomedicines*. 2020;**8**(8):267. doi:10.3390/biomedicines8080267
32. Biankin A V, Waddell N, Kassahn K S, Gingras M C, Muthuswamy L B, Johns A L, Grimmond S M, *et al*. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature*.2012;**491**(7424):399-405. doi:10.1038/nature11547
  33. Grant TJ, Hua K, Singh A. Molecular Pathogenesis of Pancreatic Cancer. *Prog Mol Biol Transl Sci*. 2016;**144**:241-75.
  34. Wang W, Nag S, Zhang X, Wang M-H, Wang H, Zhou J, *et al*. Ribosomal Proteins and Human Diseases: Pathogenesis, Molecular Mechanisms, and Therapeutic Implications. *Med Res Rev*. 2015;**35**(2):225-285. doi:10.1002/med.21327
  35. Lin S, Zhang J, Chen H, Chen K, Lai F, Luo J, Tong H, *et al*. Involvement of endoplasmic reticulum stress in capsaicin-induced apoptosis of human pancreatic cancer cells. *Evid based Complement Altern Med* 2013;**2013**. doi:10.1155/2013/629750
  36. Gehrman M, Radons J, Molls M, Multhoff G. The therapeutic implications of clinically applied modifiers of heat shock protein 70 (Hsp70) expression by tumor cells. *Cell Stress Chaperones*. 2008;**13**(1):1-10. doi:10.1379/csc-295
  37. Kong L, Liu P, Fei X, Wu T, Wang Z, Zhang B, Tan X, *et al*. A Prognostic Prediction Model Developed Based on Four CpG Sites and Weighted Correlation Network Analysis Identified DNAJB1 as a Novel Biomarker for Pancreatic Cancer. *Front Oncol*. 2020;**10**:1716. doi:10.3389/fonc.2020.01716
  38. Ren H, Luo M, Chen J, Zhou Y, Li X, Zhan Y, *et al*. Identification of TPD52 and DNAJB1 as two novel bile biomarkers for cholangiocarcinoma by iTRAQbased quantitative proteomics analysis. *Oncol Rep*. 2019;**42**:2622–2634. doi.:10.3892/or.2019.7387
  39. Cui X, Choi HK, Choi YS, Park SY, Sung GJ, Lee YH, *et al*. DNAJB1 destabilizes PDCD5 to suppress p53-mediated apoptosis. *Cancer Lett*.2015;**357**:307-315. doi:10.1016/j.canlet.2014.11.041
  40. Pandha H, Rigg A, John J, Lemoine N. Loss of expression of antigen-presenting molecules in human pancreatic cancer and pancreatic cancer cell lines. *Clin Exp Immunol*. 2007;**148**(1):127-135. doi:10.1111/j.1365-2249.2006.03289.x
  41. Ustione A, Piston DW, Harris PE. Minireview: Dopaminergic regulation of insulin secretion from the pancreatic islet. *Mol Endocrinol*. 2013;**27**(8):1198-1207. doi:10.1210/me.2013-1083
  42. Lin, S, Zhang, J, Chen, H, Chen, K, Lai, F, Luo, J, Tong, H, *et al*. Involvement of endoplasmic reticulum stress in capsaicin-induced apoptosis of human pancreatic cancer cells. *Evid Based Complementary Altern Med*. 2013;**2013**. doi:10.1155/2013/629750