Identification of new biomarkers for pancreatic cancer management: A bioinformatics analysis

Somayeh Jafari
Birjand university of Medical Sciences
Milad Ravan
Shiraz University of Medical Sciences
Hamid Aria
Fasa University of Medical Sciences
Amir Atapour (amir.atapoor58@yahoo.com)
Shiraz University of Medical Sciences
Gholamreza Anani Sarab
Birjand University of Medical Sciences

Research Article

Keywords: Pancreatic adenocarcinoma, Biomarkers, Bioinformatics analysis, Histone, Non-coding RNA

Posted Date: June 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1757565/v1

Abstract

Background

Pancreatic adenocarcinoma is one of the highly invasive and the seventh most common cause of death among cancers worldwide. To identify key genes and the involved mechanisms in pancreatic adenocarcinoma, we used bioinformatics analyzes in our study to introduce potential biomarkers in pancreatic cancer management.

Methods

In this study, gene expression profiles of pancreatic adenocarcinoma patients and normal adjacent tissues were screened and downloaded from The Cancer Genom Atlas (TCGA) bioinformatics database. Differentially expressed genes (DEGs) were identified between normal and pancreatic cancer gene expression signatures using R software. Then, Enrichment analysis of DEGs [including Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis] was performed by an enrichr (interactive and collaborative HTML5 gene list enrichment analysis) web-based tool. The protein-protein interaction (PPI) network was also constructed using STRING (Search Tool for the Retrieval of Interacting Genes) and Cytoscape software to identify the hub genes according to the top 100 DEGs in pancreatic adenocarcinoma.

Results

In our study, more than 2000 DEGs with variable log2 fold (LFC) were identified among 34,706 genes. Principal component analysis showed that the top 20 DEGs, including H1-4, H1-5, H4C3, H4C2, RN7SL2, RN7SL3, RN7SL4P, RN7SKP80, SCARNA12, SCARNA10, SCARNA5, SCARNA7, SCARNA6, SCARNA21, SCARNA9, SCARNA13, SNORA73B, SNORA53, SNORA54 with 99.91% probability might distinguish pancreatic adenocarcinoma from normal tissue. GO analysis of these 20 top DEGs showed that they have more enriched in negative regulation of gene silencing, negative regulation of chromatin organization, negative regulation of chromatin silencing, nucleosome positioning, regulation of chromatin silencing and nucleosomal DNA binding. KEGG analysis identified an association between pancreatic adenocarcinoma and systemic lupus erythematosus, alcoholism, neutrophil extracellular trap formation, and viral carcinogenesis. In protein-protein interaction (PPI) network analysis, we found that different types of histone-encoding genes are involved as hub genes in the carcinogenesis of pancreatic adenocarcinoma.

Conclusions

Our bioinformatics analysis showed that the DEGs and hub genes as key genes identified in this study may serve as new biomarkers in the near future for better management of pancreatic cancer. Although, H1.3 is currently one of the prognostic biomarkers in pancreatic cancer.

Background

Pancreatic cancer is a highly invasive and immunosuppressive cancer of the digestive system and the seventh most common cause of death among cancers worldwide (1, 2). The incidence rate of this cancer is increasing annually in the world. So, it is estimated to become the second leading cause of cancer death in 2030 (1, 3). For various reasons such as the placement of the pancreas deep in the abdomen, late-onset of clinical manifestations, high invasiveness, and early metastasis, pancreatic cancer is diagnosed in the advanced stages in more than 75% of patients (4, 5). The prognosis of this cancer is very poor and only 7% of patients have a 5-year overall survival. Hence, most patients die six months after being diagnosed with pancreatic cancer (4). Pancreatic adenocarcinoma is one of the most common subtypes of pancreatic cancer, accounting for approximately 90% of cases. Therefore, in many studies, pancreatic adenocarcinoma is used as an equal to pancreatic cancer (6, 7). Many risk factors including smoking, obesity, pancreatitis, family history, specific genetic polymorphisms, and diabetes are known in pancreatic cancer (8). Surgery, immunotherapy, radiotherapy, and chemotherapy are standard therapy in the treatment of pancreatic cancer, which unfortunately has unsatisfactory outcomes and their side effects reduce the patient's quality of life (4, 9, 10). Also, current diagnostic methods such as biomarkers, imaging examination, etc. have many limitations in diagnosing pancreatic cancer, especially early diagnosis (11).

Gene expression signature includes the expression data of a set of genes in a disease, characterized by high-throughput sequencing methods. Examining the signature in certain diseases such as cancers and comparing it with the normal cell pattern can indicate differentially expressed genes (DEGs). Analysis of recognized DEGs by bioinformatics methods can lead to the identification of pathogenic mechanisms in the disease, helping diagnosis, management and prognosis. Bioinformatics analysis, a powerful and preferred method, investigates gene expression raw data obtained from high-throughput methods and identifies differences between normal and disease. High-throughput sequencing technologies are modern, cost-effective and high-speed approaches that allow the measurement of a very large number of gene transcripts in parallel at the same time (12–14).

In this study, we selected pancreatic adenocarcinoma gene expression profiles from The Cancer Genome Atlas (TCGA) database and used bioinformatics tools to screen the DEGs in this cancer. Also, the enrichment analysis including Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis was performed based on enrichr. Then, we used the STRING database and Cytoscape software to construct a protein-protein interaction (PPI) network to identify the hub genes in pancreatic adenocarcinoma. Finally, we found a set of key genes involved in the biological processes and underlying diseases of pancreatic adenocarcinoma that we hope will serve as novel biomarkers shortly to better manage of this cancer.

Methods

Screening database

We extracted gene expression profiles of 30 pancreatic adenocarcinoma and 52 normal adjacent tissue obtained by RNA-sequencing methods from The TCGA bioinformatics database in the htseq.counts file format. Each gene expression profile belonged to a patient. The cases were selected based on the following criteria: First, patients were clustered into two groups with pancreatic adenocarcinoma (as a cancer group) and normal group. The search was screened for the following criteria in the cancer group: i) primary site was the pancreas. ii) disease type was adenomas and adenocarcinomas. iii) sample type was a primary tumor. iv) gender was female (13 cases) and male (17 cases) in two separate searches. and v) race was white. Also, the search criteria included the following items for the normal group: i) primary site was the pancreas. ii) sample type was solid normal tissue. iii) gender was female (20 cases) and male (32 cases) in two separate searches. and iv) race was white.

Because of the low number of non-white patients recorded in the TCGA database, the cases were selected from the white race only. Also, to prevent the effect of other malignant diseases on gene expression, patients with a history of malignancy other than pancreatic adenocarcinoma were removed using data filtering in the TCGA database.

Data collection and preprocessing

After applying the above filtering, there were about 797 pancreatic adenocarcinoma cases in the TCGA database. But only the expression data for 30 cases were available. Each patient's gene expression profile was downloaded separately and converted into a single data set using R programming language version 3.6.3. Preliminary data included more than 60,000 genes in 82 tissue samples. In the downloaded gene expression data file, the expression index of each gene was the ensemble gene ID. In this step, the gene index was converted to a gene symbol using the R language and the BioMart package. Of course, some gene IDs were removed from the genes groups because the equivalent gene symbol has not been defined yet. Then, a data set including the expression data of 34706 different genes in 82 tissue samples was obtained. This data set was analyzed in R software in the following order.

Normalization of gene expression profile data

The expression levels of the same mRNAs in the same cells should be in the same range. But different laboratory methods, as well as the variable length of genes, falsely affect the measured level of mRNA. Therefore, raw data need to be normalized before comparison. For this purpose, the DESeq2 package, one of the most reliable software packages for finding genes with distinct expressions was used. This package normalizes the data according to the depth of sequencing and RNA structure (15, 16).

In the next step, the mean expression of housekeeping genes was compared between normal and adenocarcinoma tissue samples. This comparison is made to ensure the quality of gene expression profile data. Housekeeping genes are used as a standard in gene expression studies. The expression of housekeeping genes in each living cell is expected to remain unchanged. To perform this step, the mean expression of these genes in both normal and adenocarcinoma groups was obtained and plotted on a point graph. The bisector was plotted and considered as a criterion for comparing gene expression in the two groups. The closer the points are to the bisector, the more similar the expression in the two groups (17).

Univariate analysis using an independent t-test to find DEGs

Gene expression differentiation can be interpreted in the form of over-expression or down-expression from normal to cancer. In order to specify the DEGs, gene expression differentiation in normal and adenocarcinoma groups has been evaluated and compared by an independent t-test. The independent t-test answers whether there is a statistically significant difference between the means of the two independent groups (18).

P-value adjustment with false discovery rate

P-value adjustment was performed using a false discovery rate according to Benjamini- Hochberg method. When the number of variables is much larger than the number of cases, we deal with large data and the number of tested statistical assumptions is high. When the number of assumptions rises, the chance of observing rare phenomena also increases and enhances the probability of rejecting the null hypothesis or the first type (α) error, while the null hypothesis is true (19, 20).

Data Visualization based on DEGs

The resulting data were shown as heatmap diagrams and clustered with dendrogram diagrams. A heatmap is a type of data visualization in which different colors and intensities in diverse cells represent the value of each cell in the map or matrix. The dendrogram is a tree-like diagram drawn for clustering samples at the edge of the heatmap. The main package used to draw diagrams and data Visualization was ggplot2 (21, 22).

Principal component analysis

In principal component analysis, linear combinations of variables (here DEGs) are constructed to explain the greatest dispersion between normal and cancerous data. It also reduces the dimensionality of big data. In this way, according to the coefficients of variables in these linear combinations, the role of them in the data dispersion can also be understood (23, 24).

In the present study, this analysis was performed to answer the question of the extent to which all the DEGs than all the gene profiles can justify the differentiation between cancer and normal. To ensure no interference in dimensional reduction, principal component analysis was performed between a set of 20 random genes and the top 20 DEGs.

Enrichment analysis of DEGs

In this step, while a list of DEGs has been obtained, it is necessary to determine which cell processes and functions are significantly affected. For this purpose, enrichr web-based tool was used for the evaluation of the top 20 DEGs. enrichr is an enrichment analysis tool that provides interpretable output from input DEGs function by interacting with other databases related to gene function, such as the KEGG and GO. DEGs clustering in enrichr was performed based on cellular components (CC), biological processes (BP), molecular function (MF) and diseases. We selected enriched functional clusters with a cutoff of $p < 0.05$ in this study (25, 26).

Establishment of the PPI network

In this part of study, we utilized the online tool STRING (Search Tool for the Retrieval of Interacting Genes) version 11.5 to map the PPI network of the first found DEGs. Then, Cytoscape software version 3.8.2 was used to visualize and analyze the PPIs network. The maximum number of interactors = 0 and confidence score ≥ 0.4 were selected. PPI network analysis methods are an important tool or for obtaining a comprehensive view of biological processes and describing physical interactions between proteins (27–29).

Results

In this study, a data set including the expression data of 34,706 different genes in 82 tissue samples (30 samples of pancreatic adenocarcinoma and 52 normal adjacent tissue samples) was analyzed.

Normalization of gene expression profile data

For this purpose, EIF1B, ATF1, PABPN1, ATF2, ATF4, ATF6, EIF1 and EIF6 genes were selected from 2176 existing housekeeping genes in the dataset and their mean expression between normal and adenocarcinoma groups was calculated. As expected and shown in the upper part of Fig. 1, the expression of these genes, did not differ significantly between both groups and all of them are located close to the bisector.

Comparison of gene expression pattern between normal and adenocarcinoma tissue samples

Considering a significance level of less than 0.05 for the p-value adjusted by the Benjamini- Hochberg method among 34706 genes studied, more than 2000 DEGs were found in the adenocarcinoma group compared to normal. All DEGs are downregulated in adenocarcinoma than in normal tissue. As well, no association was found between DEGs and gender. The top 100 DEGs are shown in Table 1, in order of the highest to the lowest p-value. It should be noted that to visualize the genes better, their mean expression at the log base 2 scale has been used. Figure 2 demonstrates the clustering of adenocarcinoma and normal samples based on DEGs in the heatmap and dendrogram diagrams. As seen in this figure, the adenocarcinoma group has formed clusters on the right side of the image.

Table 1
the top 100 DEGs most significantly downregulated in this study with highest to lowest p-value.

No.	DEGs	log2 fold change (LFC)	No.	DEGs	log2 fold change (LFC)	No.	DEGs	LFC log2 fold change (LFC)	No.	DEGs	log2 fold change (LFC)
1	SCARNA7	-9.872191923	21	STARD13-AS	–6.586138982	41	CHORDC1P4	-6.138641061	61	SNORA49	-9.875984788
2	SCARNA6	-9.136192135	22	H4C5	-7.003969725	42	PRKCA-AS1	-7.192091768	62	MAPK6P3	-6.312623969
3	RN7SL2	-8.483314601	23	ZBTB20	-5.575275315	43	CYP2U1-AS1	-5.936144925	63	KLF7-IT1	-4.772020331
4	RN7SL3	-9.588784426	24	SNORA2C	-7.362649511	44	H2AC14	-8.335678769	64	DIAPH1-AS1	-7.242754254
5	SCARNA21	-9.034023563	25	SNORD94	-5.905646794	45	SNHG22	-4.906674691	65	H4C8	-4.283608225
6	RN7SKP80	-8.067611544	26	H4C4	-7.400316925	46	H2BC13	-6.700252189	66	RPL3P6	-5.407536123
7	SNORA73B	-8.539411443	27	KCNJ13	-6.707619781	47	H2AC17	-6.293731146	67	ATP1A3	5.128644134
8	H4C3	-9.376120494	28	SNORA71A	-5.845018752	48	MMADHCP2	-6.548399417	68	TAS2R30	-7.975290336
9	SCARNA5	-10.24679653	29	H2AC20	-5.667385694	49	FOXP1-IT1	-6.595447618	69	TATDN1P1	-6.019952411
10	RN7SL4P	-8.546085682	30	RN7SL5P	-8.826507055	50	BMP2KL	-8.580881172	70	SNORA12	-7.258979178
11	SNORA53	-7.979723615	31	KRT18P31	-6.625644209	51	RNY1	-9.808546291	71	RNU5B-1	-10.34663995
12	H1-4	-7.276132193	32	MRTFA-AS1	-5.704040111	52	H2BC3	-9.708488287	72	KRT18P57	-6.261010722
13	SCARNA9	-7.107154892	33	POU5F2	-6.643217509	53	H1-3	-6.851746017	73	SNORA23	-7.520624732
14	SCARNA13	-7.073519738	34	MIR3609	-9.828421275	54	H2AC12	-8.239600629	74	H2BC6	-5.604485706
15	SNORA54	-9.297686808	35	MALAT1	-5.106068934	55	SYT5	5.789893437	75	PHBP2	-5.492474884
16	H4C2	-8.971268737	36	ZNF460	-5.153904002	56	H2AC21	-7.111029988	76	MARK2P8	-6.28853337
17	SNORD17	-8.439895562	37	RN7SL648P	-8.058409434	57	H2BC17	-6.1974086	77	RBMS3-AS2	-6.51653129
18	SCARNA12	-6.924204908	38	RNU5A-1	-9.872167322	58	UHRF2P1	-8.035063773	78	H3C7	-8.457592556
19	SCARNA10	-9.544811817	39	LCMT1-AS2	-5.755003979	59	H2AC4	-8.486362848	79	H3C11	-7.976305302
20	H1-5	-8.570333537	40	RLIMP1	-7.241890326	60	MIR181A1HG	-6.495394522	80	ADAM20	-5.086396233

Principal component analysis results

In this study, when all the DEGs were used in the principal component analysis, they were able to distinguish alone 38.29% of the adenocarcinoma sample from the normal sample than all gene profiles (Fig. 3). Thus, the use of all DEGs cannot differentiate between cancerous and normal samples. But in contrast, the principal component analysis of the top 20 DEGs was able to separate both groups by 99.91% probability (Fig. 4).

Enrichment analysis of DEGs

As mentioned in the previous section, the top 20 DEGs are more likely to distinguish between the two groups of normal and adenocarcinoma. Hence, we used the enrichr web-based software to analyze the 20 top DEGs to obtain the GO and KEGG pathways. As shown earlier, all DEGs were downregulated in this study. By analyzing GO enrichment, we found that the DEGs in BP were more enriched in negative regulation of gene silencing, negative regulation of chromatin organization, negative regulation of chromatin silencing, nucleosome positioning and regulation of chromatin silencing (Fig. 5, Table 2). As for MF, the DEGs were just enriched in nucleosomal DNA binding (Fig. 5, Table 2). Enrichment analysis of the top 20 genes in the CC category had no significant results (p value > 0.05). The KEGG pathway analysis results showed that the DEGs were significantly enriched in systemic lupus erythematosus, alcoholism, neutrophil extracellular trap formation and viral carcinogenesis (Fig. 6, Table 2).

Table 2
Enrichment analysis of DEGs

Expression	analysis	Category	Term	P value	Involved genes
Down regulated	GO analysis	Biological process (BP)	negative regulation of gene silencing (GO:0060969)	0.0000736129304247206	H1-5;H1-4
			negative regulation of chromatin organization (GO:1905268)	0.0000736129304247206	H1-5;H1-4
			negative regulation of chromatin silencing (GO:0031936)	0.0000736129304247206	H1-5;H1-4
			nucleosome positioning (GO:0016584)	0.0000989757503698103	H1-5;H1-4
			regulation of chromatin silencing (GO:0031935)	0.000678284581571347	H1-5;H1-4
			DNA packaging (GO:0006323)	0.001293831	H1-5;H1-4
			negative regulation of DNA metabolic process (GO:0051053)	0.002513274	H1-5;H1-4
			regulation of DNA recombination (GO:0000018)	0.002994673	H1-5;H1-4
			chromosome condensation (GO:0030261)	0.003055262	H1-5;H1-4
			negative regulation of DNA recombination (GO:0045910)	0.004001058	H1-5;H1-4
			positive regulation of gene expression, epigenetic (GO:0045815)	0.004001058	H1-5;H1-4
			chromatin assembly (GO:0031497)	0.005981624	H1-5;H1-4
			nucleosome organization (GO:0034728)	0.009068935	H1-5;H1-4
			positive regulation of histone H3-K9 methylation (GO:0051574)	0.014957134	H1-5
			establishment of protein localization to chromatin (GO:0071169)	0.015946713	H1-5
			regulation of histone H3-K9 methylation (GO:0051570)	0.02238257	H1-5
			establishment of protein localization to chromosome (GO:0070199)	0.022810625	H1-5
			protein localization to chromatin (GO:0071168)	0.034669113	H1-5
			positive regulation of histone methylation (GO:0031062)	0.040568327	H1-5
		Molecular function (MF)	nucleosomal DNA binding (GO:0031492)	0.000492368756376955	H1-5;H1-4
		Cellular Compartment (CC)	no significant results		
	KEGG analysis		Systemic lupus erythematosus	0.00793534776758775	H4C2;H4C3
			Alcoholism	0.014642007150812	H4C2;H4C3
			Neutrophil extracellular trap formation	0.0150924344910985	H4C2;H4C3
			Viral carcinogenesis	0.0172729137884516	H4C2;H4C3

PPI network analysis of the DEGs

Analysis of the top 20 DEGs in the string database showed a network of 3 nodes containing HIST1H1B, HIST1H1E and HIST1H4F genes (Fig. 7). This network could not be analyzed in Cytoscape due to its small size. Therefore, the top 100 DEGs were selected and placed in the string database for PPI network analysis. Investigation of the resulting network in the Cytoscape showed 14 nodes and 178 edges (Fig. 8A). The degree of connectivity for each genes is specified in Fig. 7B. The genes with the highest degree of connectivity in this study were considered hub genes. Hub genes in the PPI network have the most connection with other nodes and can play a vital role in gene expression. Hub genes in this study included HIST1H1B, HIST1H1D, HIST1H4F, HIST1H2AE, HIST1H1E, HIST1H2AJ, HIST1H2BL, HIST2H2AB, HIST1H3J, HIST1H2AI and HIST2H2AC (Fig. 8B).

Discussion

Pancreatic adenocarcinoma is one of the most invasive human cancers that has become increasingly prevalent in recent years. As mentioned, it is estimated that by 2030, this cancer will be the second leading cause of death among cancers. Therefore, the identification of sensitive and specific biomarkers for the early diagnosis and treatment of pancreatic adenocarcinoma, as well as predicting its survival and prognosis is crucial. High-throughput studies can find genes expression differences and important molecular pathways in both normal and cancerous cases, leading to the development of biomarkers for better

management of pancreatic adenocarcinoma. In this study, a data set including the expression data of 34,706 different genes in 82 samples of pancreatic adenocarcinoma and normal was analyzed in the TCGA database. Meanwhile, considering the significance level less than 0.05, 2000 DEGs were found in cancer samples compared to normal. All of these DEGs are down-expressed in pancreatic adenocarcinoma than in normal tissue. To have a deep understanding of the function of these DEGs, we performed enrichment analysis and PPI network analysis to screen the genes and pathways associated with pancreatic adenocarcinoma that are more important in the development and progression of pancreatic cancer.

In this study, the results of the principal component analysis showed that the top 20 DEGs including 4 genes encoding histone protein called H1-4, H1-5, H4C3 and H4C2 and 16 genes encoding non-coding RNA named RN7SL2, RN7SL3, RN7SL4P, RN7SKP80, SCARNA12, SCARNA10, SCARNA5, SCARNA7, SCARNA6, SCARNA21, SCARNA9, SCARNA13, SNORA73B, SNORA53, SNORA54 and SNORD17 with 99.91% probability diagnose pancreatic adenocarcinoma cases from normal (Table 1). Other studies have revealed that histones, as chromatin-regenerating proteins, are important in cancer biology. They undergo severe changes during cancer and may involve in causing the disease. Among the four genes encoding histone, histone H1 is both a cancer promoter and a potential diagnostic biomarker in different malignancies (30–32). About the majority of non-coding RNAs found in the present study, their exact molecular function was unclear and there were few articles about them. Several studies have shown that long-noncoding RNAs such as RN7SL2 and RN7SL4P are overexpressed in patients with multiple myeloma(33). RN7SL2 is also abundant in the cancer patient's plasma (34). In contrast, another report presented that the RN7SL3 is downregulated in hepatocellular carcinoma (35). SNORA73 is known as a chromatin-associated snoRNA and is effective in genome stability (36, 37). SNORA54 has been studied in many human cancers such as breast, melanoma, lymphoma and myeloma. This snoRNA has upregulated in most cancer patients, but down-expressed in patients with melanoma (38, 39). According to the literature, SNORD17 is also overexpressed in cases of hepatocellular carcinoma and its upregulation is usually associated with poor clinical outcomes (40, 41).

Unfortunately, no pancreatic cancer study was performed on the found non-coding RNAs expression and function in the paper. However, one study demonstrated that SCARNA6 is overexpressed in patients with autism spectrum disorders (42). This finding can be important due to the close relationship between gene expression in the pancreas and nerve tissues. SCARNA7 has also been shown to be correlated with many cancers such as breast, prostate and non-small cell lung cancer. This SCARNA is usually upregulated in breast cancer and associated with poor prognosis (43–45). One other study revealed that SCARNA9 was significantly overexpressed in colon cancer. In contrast, another article suggested that downregulation of SCARNA9 are negatively associated with endometrial cancer (46, 47). Numerous studies have investigated the expression of SCARNA10 in liver fibrosis and hepatocellular carcinoma. They showed that the expression of SCARNA10 increased in these two disorders and is usually associated with the physio-pathological features of the diseases. Hence, this SCARNA has been introduced as a diagnostic biomarker and therapeutic target in liver fibrosis and hepatocellular carcinoma. Silencing of SCARNA10 coding gene in hepatocytes has been displayed down-expression of TGFβ, TGFβRI, SMAD2, SMAD3 and KLF6 (48–51). SCARNA13 is also highly expressed in hepatocellular carcinoma and is involved in tumorigenesis and metastasis (52–54).

GO analysis of the top 20 DEGs in this study presented that they are mainly enriched in the pathways associated with negative regulation of gene silencing, negative regulation of chromatin organization, negative regulation of chromatin silencing, nucleosome positioning, regulation of chromatin silencing, DNA packaging, negative regulation of DNA metabolic process, regulation of DNA recombination, chromosome condensation, negative regulation of DNA recombination, positive regulation of gene expression, epigenetic, chromatin assembly, nucleosome organization, positive regulation of histone H3-K9 methylation, the establishment of protein localization to chromatin, Regulation of histone H3-K9 methylation, the establishment of protein localization to chromosome, protein localization to chromatin, positive regulation of histone methylation and nucleosomal DNA binding (Table 2). Of course, among the first 20 DEGs only two genes, H1-4 and H1-5, were identified as influential genes in GO analysis.

In this study, the top 20 DEGs KEGG analysis showed a relationship between pancreatic cancer and diseases such as systemic lupus erythematosus, alcoholism, neutrophil extracellular trap formation and viral carcinogenesis due to the function of H4C2 and H4C3 genes. Many other studies have proved that alcoholism (consumption of high amounts of alcohol) is one of the critical risk factors in the progression and development of pancreatic adenocarcinoma, especially in patients with KRAS mutations (55–58). There have been many reports on the effect of neutrophil extracellular trap formation on pancreatic cancer, but its exact role in the development of this cancer is still unknown. one article has pointed to the anti-cancer effects of neutrophil extracellular trap, but most studies have emphasized the function of neutrophil extracellular trap formation in tumor symptoms exacerbation, resistance to immunotherapy and induction of migration and invasion in pancreatic cancer cells. Neutrophil extracellular trap formation has even been suggested to be involved in predicting the survival of pancreatic adenocarcinoma patients after surgery (59–62). There is ample evidence linking systemic lupus erythematosus to the risk of developing various cancer types. In a meta-analysis study, systemic lupus erythematosus was associated with an increased risk of pancreatic cancer (63). But in another study, no significant relationship was found between the two diseases (64). As in our research, other studies show the role of viral infections in pancreatic carcinogenesis. Some of these viruses include the SARS-COVID-19 family and the hepatitis family (B and C). Certainly, careful monitoring of patients with any of the diseases may help in the early diagnosis of pancreatic cancer and improve the prognosis of these patients (65–67).

In the last part of this study, PPI network analysis was performed for the top 100 DEGs found in this study. As announced in the results section (Fig. 7), 14 nodes were identified with these DEGs. Eleven of them with a degree of connectivity equal to 13 were selected as hub genes. Interestingly, all of the hub genes were histones encoding genes. Of course, our study also found 23 histone encoding genes among the top 100 DEGs. These genes include H4C3, H1-4, H4C2, H1-5, H4C5, H4C4, H2AC20, H2AC14, H2BC13, H2AC17, H2BC3, H1-3, H2AC12, H2AC21, H2BC17, H2AC4, H4C8, H2BC6, H3C7, H3C11, H4C1, H4C6 and H4C13, indicating the role of histones in the development of pancreatic cancer. Numerous reports suggest that histone gene expression profiles in many types of cancer such as breast, lung, prostate, kidney and pancreas may involve in patients' prognosis. For example, the expression of histone H1.3 in pancreatic adenocarcinoma patients can predict the clinical outcome after pancreatic surgery. Therefore, H1.3 is identified as one of the prognostic biomarkers in pancreatic cancer (30, 32, 68, 69). Finally, much studies on histones and non-coding RNAs found in this study should be performed to determine their role and function in pancreatic cancer.

Conclusion

Briefly, we present a gene expression profile analysis of patients with pancreatic adenocarcinoma. We identified DEGs and their associated biological pathways and showed that they could be the link between pancreatic adenocarcinoma and several diseases. During this study, we identified many key genes that we believe could serve as potential candidates for the diagnosis and prognosis of pancreatic adenocarcinoma in the near future. As, the role of some of them, such as H1.3, is now identified as a prognostic biomarker. However, more extensive studies are needed to determine the role of each of these genes in the diagnosis, treatment, and prognosis of pancreatic adenocarcinoma.

Abbreviations

TCGA: The Cancer Genome Atlas; DEGs: Differentially expressed genes; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; PPI: protein-protein interaction; STRING: Search Tool for the Retrieval of Interacting Genes; LFC: log2 fold changes; H1-4: Histone 1-4; RN7SL2: RNA Component Of Signal Recognition Particle 7SL2; RN7SL4P: RNA, 7SL, Cytoplasmic 4, Pseudogene; RN7SKP80: RN7SK Pseudogene 80; SCARNA12: Small Cajal Body-Specific RNA 12; SNORA73B: Small Nucleolar RNA, H/ACA Box 73B.

Declarations

- Ethical Approval and Consent to participate

Not applicable

- Consent for publication

Not applicable

- Availability of supporting data

Not applicable

- Competing interests

The authors declare that they have no competing interests.

- **Funding**

Not applicable

- Authors' contributions

SJ, MR and HA. conceptualized and wrote the manuscript. AA, GAS. helped with modification of the manuscript. AA, GAS. conceived the study and provided advice. All authors participated in manuscript editing and read and approved the final version.

- Acknowledgements

Not applicable

- Authors' information

[1]Department of Molecular Medicine, School of Medicine, Birjand university of Medical Sciences, Birjand, Iran. [2]Student Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran.[3]Noncommunicable Diseases Research Center, Fasa University of Medical Sciences, Fasa, Iran. [4]Department of Immunology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran.[5]Department of Medical Biotechnology, School of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences, Shiraz, Iran. [6]Cellular & Molecular Research Center, Birjand University of Medical Sciences, Birjand, Iran

References

1. Ma C, Cui Z, Wang Y, Zhang L, Wen J, Guo H, et al. Bioinformatics analysis reveals TSPAN1 as a candidate biomarker of progression and prognosis in pancreatic cancer. Bosnian Journal of Basic Medical Sciences. 2021;21(1):47.
2. https://gco.iarc.fr/today
3. Conway JR, Herrmann D, Evans TJ, Morton JP, Timpson P. Combating pancreatic cancer with PI3K pathway inhibitors in the era of personalised medicine. Gut. 2019;68(4):742–58.
4. McGuigan A, Kelly P, Turkington RC, Jones C, Coleman HG, McCain RS. Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes. World journal of gastroenterology. 2018;24(43):4846.
5. Moutinho-Ribeiro P, Macedo G, Melo SA. Pancreatic cancer diagnosis and management: has the time come to prick the bubble? Frontiers in endocrinology. 2019;9:779.
6. Pancreas ESGoCTot. European evidence-based guidelines on pancreatic cystic neoplasms. Gut. 2018;67(5):789–804.
7. Haeberle L, Esposito I. Pathology of pancreatic cancer. Translational gastroenterology and hepatology. 2019;4.

8. Rawla P, Sunkara T, Gaduputi V. Epidemiology of pancreatic cancer: global trends, etiology and risk factors. World journal of oncology. 2019;10(1):10.
9. Luo D, Carter KA, Miranda D, Lovell JF. Chemophototherapy: an emerging treatment option for solid tumors. Advanced Science. 2017;4(1):1600106.
10. Schizas D, Charalampakis N, Kole C, Economopoulou P, Koustas E, Gkotsis E, et al. Immunotherapy for pancreatic cancer: A 2020 update. Cancer Treatment Reviews. 2020:102016.
11. Li J, Zhu C, Yue P, Zheng T, Li Y, Wang B, et al. Identification of glycolysis related pathways in pancreatic adenocarcinoma and liver hepatocellular carcinoma based on TCGA and GEO datasets. Cancer cell international. 2021;21(1):1–13.
12. Rao C, Huisman DH, Vieira HM, Frodyma DE, Neilsen BK, Chakraborty B, et al. A gene expression high-throughput screen (GE-HTS) for coordinated detection of functionally similar effectors in cancer. Cancers. 2020;12(11):3143.
13. De Wolf H, Cougnaud L, Van Hoorde K, De Bondt A, Wegner JK, Ceulemans H, et al. High-throughput gene expression profiles to define drug similarity and predict compound activity. Assay and drug development technologies. 2018;16(3):162–76.
14. Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. Nature biotechnology. 2019;37(12):1452–7.
15. Fundel K, Haag J, Gebhard P, Zimmer R, Aigner T. Normalization strategies for mRNA expression data in cartilage research. Osteoarthritis and cartilage. 2008;16(8):947–55.
16. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):1–21.
17. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. TRENDS in Genetics. 2013;29(10):569–74.
18. Gerald B. A brief review of independent, dependent and one sample t-test. International journal of applied mathematics and theoretical physics. 2018;4(2):50–4.
19. Alberton BA, Nichols TE, Gamba HR, Winkler AM. Multiple testing correction over contrasts for brain imaging. NeuroImage. 2020;216:116760.
20. Noble WS. How does multiple testing correction work? Nature biotechnology. 2009;27(12):1135–7.
21. Wickham H. ggplot2: elegant graphics for data analysis: springer; 2016.
22. Ahlmann-Eltze C, Patil I. ggsignif: R Package for Displaying Significance Brackets for'ggplot2'. 2021.
23. Karamizadeh S, Abdullah SM, Manaf AA, Zamani M, Hooman A. An overview of principal component analysis. Journal of Signal and Information Processing. 2020;4.
24. Abdi H, Williams LJ. Principal component analysis. Wiley interdisciplinary reviews: computational statistics. 2010;2(4):433–59.
25. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic acids research. 2016;44(W1):W90-W7.
26. Xie Z, Bailey A, Kuleshov MV, Clarke DJ, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with enrichr. Current protocols. 2021;1(3):e90.
27. Faro S, Lecroq T, Borzi S, Di Mauro S, Maggio A, editors. The String Matching Algorithms Research Tool. Stringology; 2016.
28. Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. Data mining in proteomics: Springer; 2011. p. 291–303.
29. Otasek D, Morris JH, Bouças J, Pico AR, Demchak B. Cytoscape automation: empowering workflow-based network analysis. Genome biology. 2019;20(1):1–15.
30. Izquierdo-Bouldstridge A, Bustillos A, Bonet-Costa C, Aribau-Miralbés P, García-Gomis D, Dabad M, et al. Histone H1 depletion triggers an interferon response in cancer cells via activation of heterochromatic repeats. Nucleic Acids Research. 2017;45(20):11622–42.
31. Scaffidi P. Histone H1 alterations in cancer. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms. 2016;1859(3):533–9.
32. Harshman SW, Hoover ME, Huang C, Branson OE, Chaney SB, Cheney CM, et al. Histone H1 phosphorylation in breast cancer. Journal of proteome research. 2014;13(5):2453–67.
33. Chen M, Mithraprabhu S, Ramachandran M, Choi K, Khong T, Spencer A. Utility of circulating cell-free RNA analysis for the characterization of global transcriptome profiles of multiple myeloma patients. Cancers. 2019;11(6):887.
34. Chen S, Jin Y, Wang S, Xing S, Wu Y, Tao Y, et al. Cancer Type Classification Using Plasma Cell Free RNAs Derived From Human and Microbes. 2021.
35. Giovannini C, Fornari F, Indio V, Trerè D, Renzulli M, Vasuri F, et al. Direct antiviral treatments for hepatitis C virus have off-target effects of oncologic relevance in hepatocellular carcinoma. Cancers. 2020;12(9):2674.
36. Han C, Sun L-Y, Luo X-Q, Pan Q, Sun Y-M, Zeng Z-C, et al. Chromatin-associated orphan snoRNA regulates DNA damage-mediated differentiation via a non-canonical complex. Cell Reports. 2022;38(13):110421.
37. Toms D, Pan B, Bai Y, Li J. Small RNA sequencing reveals distinct nuclear microRNAs in pig granulosa cells during ovarian follicle growth. Journal of ovarian research. 2021;14(1):1–12.
38. Rahman MM, Lai YC, Husna AA, Chen HW, Tanaka Y, Kawaguchi H, et al. Aberrantly expressed snoRNA, snRNA, piRNA and tRFs in canine melanoma. Veterinary and Comparative Oncology. 2020;18(3):353–61.
39. Bratkovič T, Božič J, Rogelj B. Functional diversity of small nucleolar RNAs. Nucleic acids research. 2020;48(4):1627–51.
40. Han S, Xie Y, Yang X, Dai S, Dai X. Small Nucleolar RNA and Small Nucleolar RNA Host Gene Signatures as Biomarkers for Pancreatic Cancer. 2020.
41. Liang J, Li G, Liao J, Huang Z, Wen J, Wang Y, et al. Non-coding small nucleolar RNA SNORD17 promotes the progression of hepatocellular carcinoma through a positive feedback loop upon p53 inactivation. Cell Death & Differentiation. 2022:1–16.

42. Ander BP, Barger N, Stamova B, Sharp FR, Schumann CM. Atypical miRNA expression in temporal cortex associated with dysregulation of immune, cell cycle, and other pathways in autism spectrum disorders. Molecular autism. 2015;6(1):1–13.
43. Goyal B, Yadav SRM, Awasthee N, Gupta S, Kunnumakkara AB, Gupta SC. Diagnostic, prognostic, and therapeutic significance of long non-coding RNA MALAT1 in cancer. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer. 2021;1875(2):188502.
44. Narayanaswamy PB, Baral TK, Haller H, Dumler I, Acharya K, Kiyan Y. Transcriptomic pathway analysis of urokinase receptor silenced breast cancer cells: a microarray study. Oncotarget. 2017;8(60):101572.
45. Iacobas DA. Powerful quantifiers for cancer transcriptomics. World journal of clinical oncology. 2020;11(9):679.
46. Wang X, Chen K, Wang Z, Xu Y, Dai L, Bai T, et al. Using Immune-Related Long Non-coding Ribonucleic Acids to Develop a Novel Prognosis Signature and Predict the Immune Landscape of Colon Cancer. Frontiers in cell and developmental biology. 2021;9.
47. Shi R, Wang Z, Zhang J, Yu Z, An L, Wei S, et al. N6-Methyladenosine-Related Long Noncoding RNAs as Potential Prognosis Biomarkers for Endometrial Cancer. International Journal of General Medicine. 2021;14:8249.
48. Han Y, Jiang W, Wang Y, Zhao M, Li Y, Ren L. Serum long non-coding RNA SCARNA10 serves as a potential diagnostic biomarker for hepatocellular carcinoma. BMC cancer. 2022;22(1):1–8.
49. Wu Z, Huang S, Zheng X, Gu S, Xu Q, Gong Y, et al. Regulatory long non–coding RNAs of hepatic stellate cells in liver fibrosis. Experimental and Therapeutic Medicine. 2021;21(4):1-.
50. He Z, Yang D, Fan X, Zhang M, Li Y, Gu X, et al. The roles and mechanisms of lncRNAs in liver fibrosis. International journal of molecular sciences. 2020;21(4):1482.
51. Zhang K, Han Y, Hu Z, Zhang Z, Shao S, Yao Q, et al. SCARNA10, a nuclear-retained long non-coding RNA, promotes liver fibrosis and serves as a potential biomarker. Theranostics. 2019;9(12):3622.
52. Yang X, Sun L, Wang L, Yao B, Mo H, Yang W. LncRNA SNHG7 accelerates the proliferation, migration and invasion of hepatocellular carcinoma cells via regulating miR-122-5p and RPL4. Biomedicine & Pharmacotherapy. 2019;118:109386.
53. Lan T, Yuan K, Yan X, Xu L, Liao H, Hao X, et al. LncRNA SNHG10 facilitates hepatocarcinogenesis and metastasis by modulating its homolog SCARNA13 via a positive feedback loop. Cancer research. 2019;79(13):3220–34.
54. Xin H, Yan Z, Cao J. Long non-coding RNA ABHD11-AS1 boosts gastric cancer development by regulating miR-361-3p/PDPK1 signalling. The Journal of Biochemistry. 2020;168(5):465–76.
55. Clement EJ, Law HC-H, Qiao F, Noe D, Trevino JG, Woods NT. Combined Alcohol Exposure and KRAS Mutation in Human Pancreatic Ductal Epithelial Cells Induces Proliferation and Alters Subtype Signatures Determined by Multi-Omics Analysis. Cancers. 2022;14(8):1968.
56. Rahman F, Cotterchio M, Cleary SP, Gallinger S. Association between alcohol consumption and pancreatic cancer risk: a case-control study. PLoS One. 2015;10(4):e0124489.
57. Wang Y-T, Gou Y-W, Jin W-W, Xiao M, Fang H-Y. Association between alcohol intake and the risk of pancreatic cancer: a dose–response meta-analysis of cohort studies. BMC cancer. 2016;16(1):1–11.
58. Koyanagi YN, Oze I, Kasugai Y, Kawakatsu Y, Taniyama Y, Hara K, et al. New insights into the genetic contribution of ALDH2 rs671 in pancreatic carcinogenesis: Evaluation by mediation analysis. Cancer Science. 2022;113(4):1441.
59. Shahzad MH, Feng L, Su X, Brassard A, Dhoparee-Doomah I, Ferri LE, et al. Neutrophil Extracellular Traps in Cancer Therapy Resistance. Cancers. 2022;14(5):1359.
60. Shao B-Z, Yao Y, Li J-P, Chai N-L, Linghu E-Q. The Role of Neutrophil Extracellular Traps in Cancer. Frontiers in Oncology. 2021:3098.
61. Chen Y, Han L, Qiu X, Wang G, Zheng J. Neutrophil Extracellular Traps in Digestive Cancers: Warrior or Accomplice. Frontiers in Oncology. 2021:4914.
62. Jin W, Yin H, Li H, Yu XJ, Xu HX, Liu L. Neutrophil extracellular DNA traps promote pancreatic cancer cells migration and invasion by activating EGFR/ERK pathway. Journal of cellular and molecular medicine. 2021;25(12):5443–56.
63. Seo M-S, Yeo J, Hwang IC, Shim J-Y. Risk of pancreatic cancer in patients with systemic lupus erythematosus: a meta-analysis. Clinical rheumatology. 2019;38(11):3109–16.
64. Song L, Wang Y, Zhang J, Song N, Xu X, Lu Y. The risks of cancer development in systemic lupus erythematosus (SLE) patients: a systematic review and meta-analysis. Arthritis research & therapy. 2018;20(1):1–13.
65. Abramczyk U, Nowaczyński M, Słomczyński A, Wojnicz P, Zatyka P, Kuzan A. Consequences of COVID-19 for the Pancreas. International Journal of Molecular Sciences. 2022;23(2):864.
66. Gheorghe G, Diaconu CC, Ionescu V, Constantinescu G, Bacalbasa N, Bungau S, et al. Risk Factors for Pancreatic Cancer: Emerging Role of Viral Hepatitis. Journal of Personalized Medicine. 2022;12(1):83.
67. Ebrahimi Sadrabadi A, Bereimipour A, Jalili A, Gholipurmalekabadi M, Farhadihosseinabadi B, Seifalian AM. The risk of pancreatic adenocarcinoma following SARS-CoV family infection. Scientific reports. 2021;11(1):1–13.
68. Bauden M, Kristl T, Sasor A, Andersson B, Marko-Varga G, Andersson R, et al. Histone profiling reveals the H1. 3 histone variant as a prognostic biomarker for pancreatic ductal adenocarcinoma. BMC cancer. 2017;17(1):1–9.
69. Rattray AM, Müller B. The control of histone gene expression. Biochemical Society Transactions. 2012;40(4):880–5.

Figures

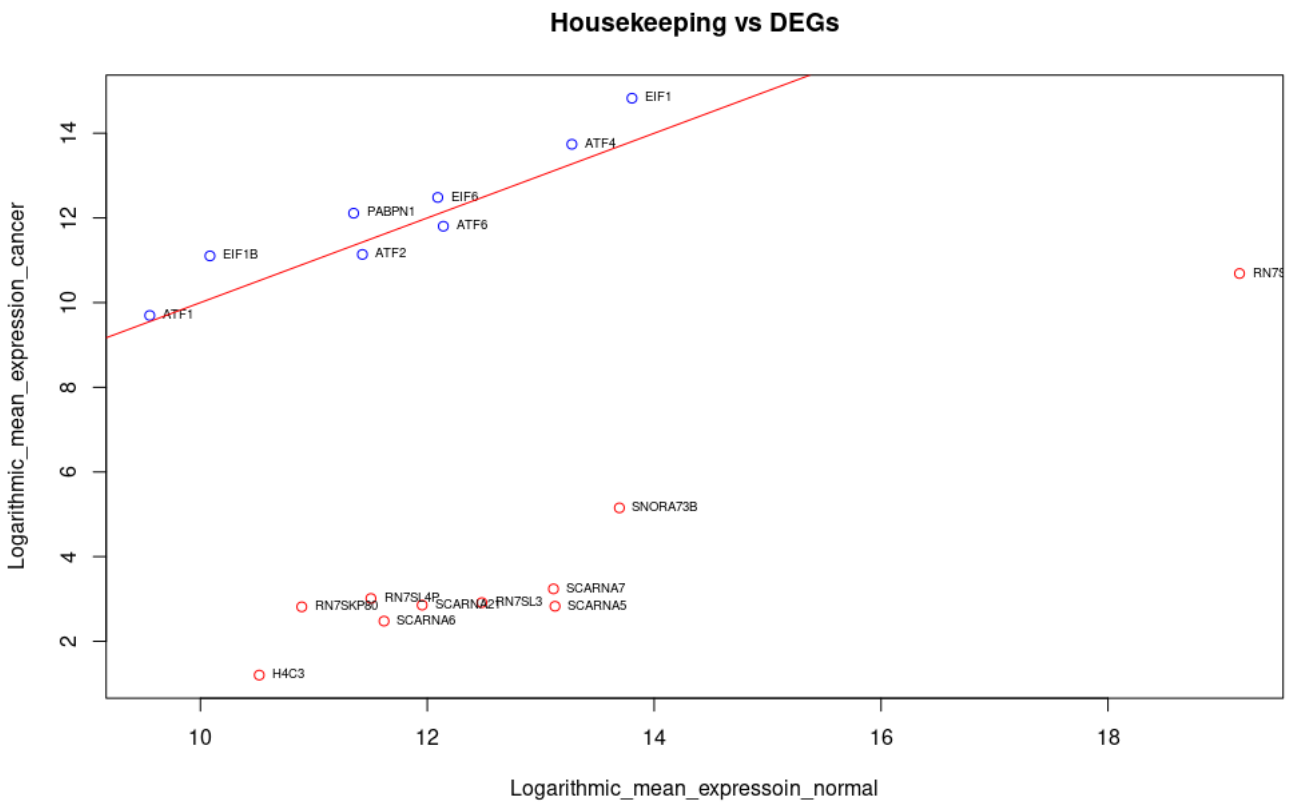

Figure 1

Normalization of gene expression profile data. The mean expression of housekeeping genes among normal and cancerous samples has been calculated and shown in the upper part of the figure. Also, the expression value of the top 10 DEGs in both normal and adenocarcinoma groups has been depicted

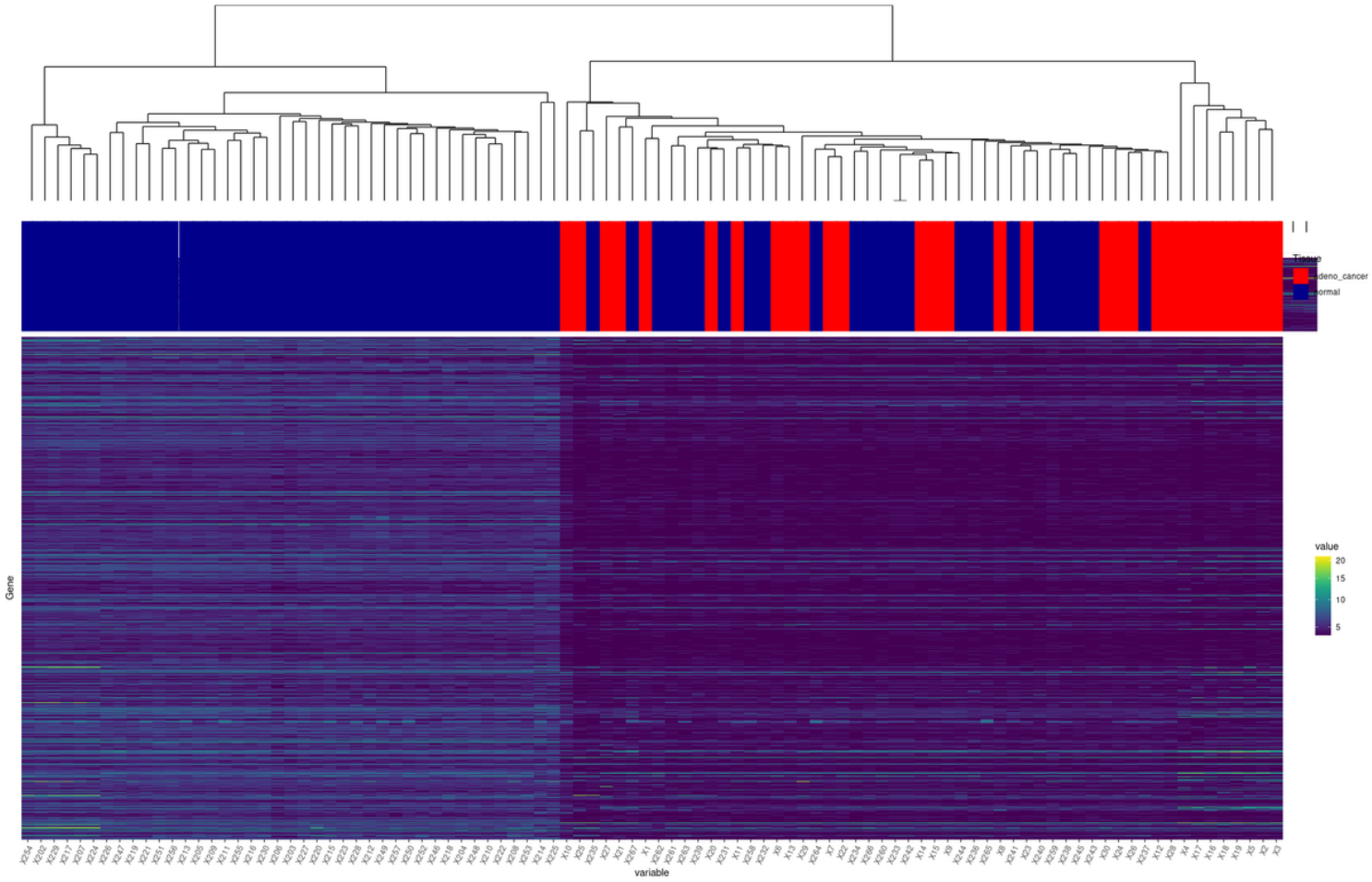

Figure 2

Clustering of adenocarcinoma and normal samples based on DEGs

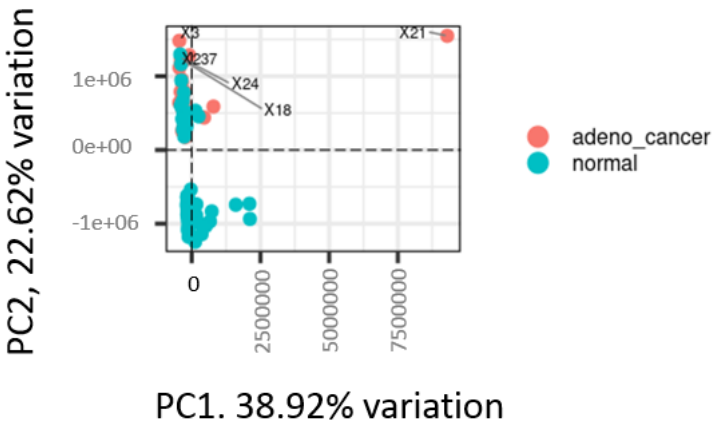

Figure 3

The ability of principal components to explain the variation among samples using all DEGs.

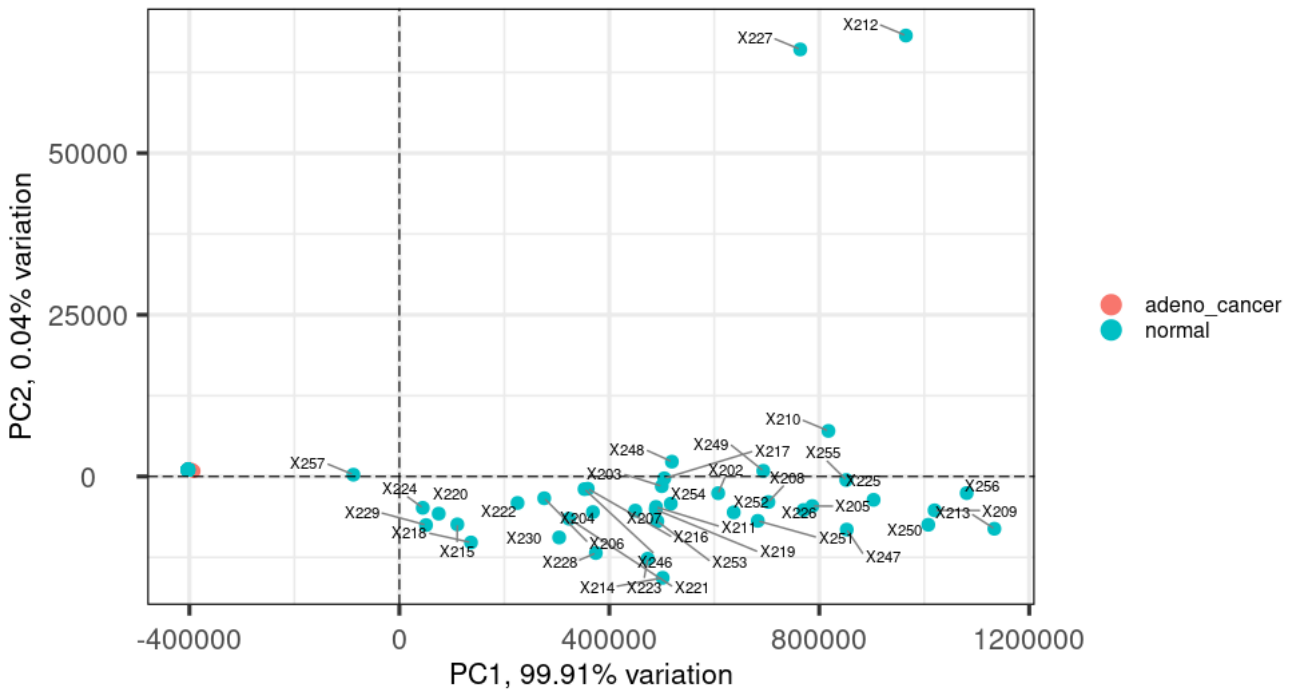

Figure 4

The ability of principal components to explain the variation among samples using the top 20 DEGs.

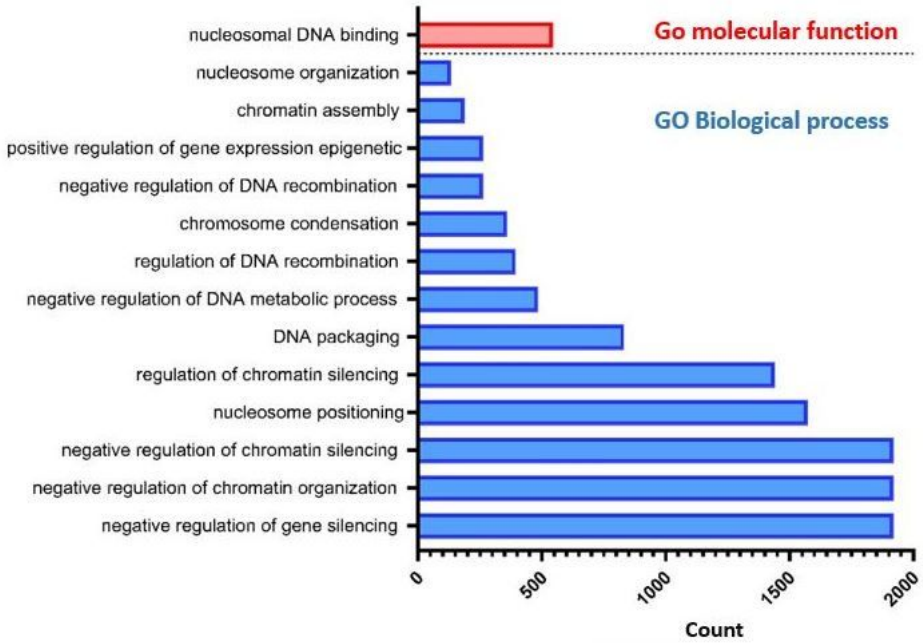

Figure 5

The most significant Go enrichment analysis of top 20 DEGs

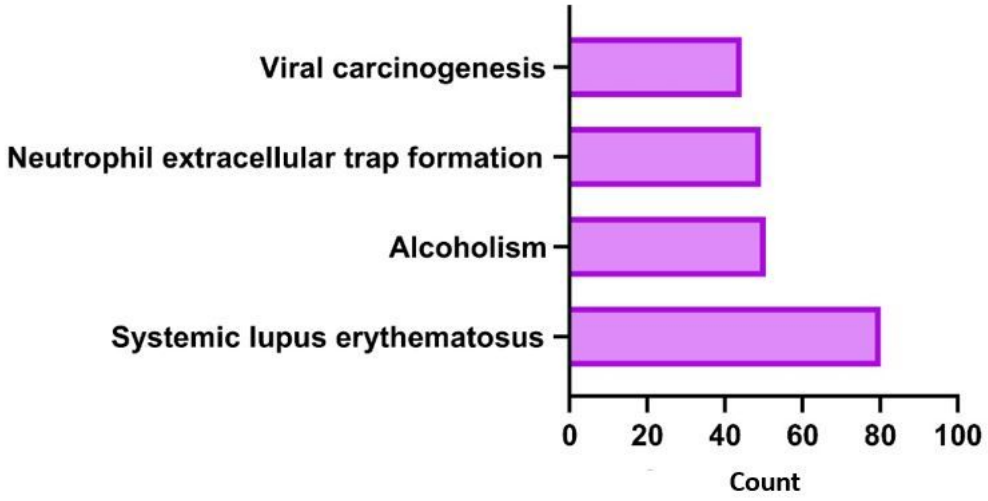

Figure 6

KEGG Enrichment analysis of top 20 DEGs

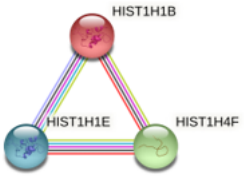

Figure 7

The PPI network of 20 top DEGs.

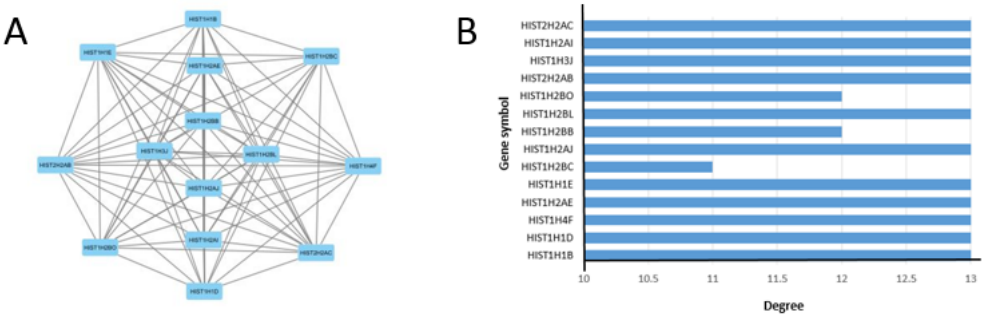

Figure 8

Protein-protein interaction (PPI) network. A. The PPI network of top 100 DEGs. B. The connectivity degree of hub genes. Genes with a degree of connectivity equal to 13 were considered hub genes.