Progress in Nutrition 2020; Vol. 22, Supplement 2: e2020003 DOI: 10.23751/pn.v22i2-S.10179

Original article

Analysis of Gastric Cancer Transcriptomic Data by Bioinformatics Tools and Detection of Candidate Diagnostic Biomarker Genes

Semih Dalkılıç

Fırat University, Faculty of Science, Department of Biology, Molecular Biology and Genetics Program, Elazıg, Turkey.

Abstract *Background:* Gastric cancer is one of the leading cause of deaths in the world and each year many new cases diagnosed worldwide. Although there has been a decrease in its incidence over the past century, gastric cancer is the second leading cause of cancer-related deaths. *Objective*: The main objective of this study is identification of candidate biomarker genes to be used in early diagnosis of gastric cancer. *Methods*: In this study, GSE54129 data set in the Gene Expression Omnibus (GEO) database was used. This data set contains gene expression data of 111 stomach cancer tumor tissues and 21 normal stomach tissues. Bioinformatics analyses performed on raw microarray data (CEL files). All the analyses were performed with Transcriptome Analysis Console 4.0 (TAC) algorithm. *Results:* According to the results, expression level of many genes during neoplastic transformation in gastric cancer significantly changes when compared to healthy control subjects. The upregulated genes which show high fold changes are *SFRP2, EGR1, CHI3L1, COL8A1, NEAT1, INHBA, CXCL8* and *MYL9*. Some of downregulated genes with higher fold changes are *GAST, GIF, GKN2, GKN1, SCGB2A1* and *HRASLS2*. *Conclusion*: These genes have a potential for candidate biomarkers that can be used in the diagnosis or detection of molecular subtypes of gastric cancer.

Keywords: Gastric cancer, Gene expression, Bioinformatics, Biomarker

Introduction

Gastric cancer (GC) is a malignancy with high mortality and morbidity rates. Although there has been a decrease in its incidence over the past century, gastric cancer ranks second in cancer-related deaths and has a 5-year survival rate of 20% (1). It is the most common cancer, especially in East Asia (2). Current molecular genetic data show that, as with other tumors, stomach cancer is composed of various molecular subtypes (3).

Gastric cancer is more common in men than in women, but the incidence is 60-70 years and the average age at diagnosis is 57 years (2, 4, 5). Gastric cancer accounts for 10% of deaths caused by cancer (4, 5, 6). The etiology of gastric cancer is not fully known (5, 7). The main reason for the poor prognosis is the difficulty in diagnosis in early stages and therefore delayed treatment (4, 8). Endoscopic methods used in the diagnosis of gastric cancer are not practical and economical, other methods like as non-invasive molecular diagnostic tests are needed for early diagnosis (7).

GC is a highly heterogeneous disorder, where even similar pathological and medical features lead to multiple outcomes, indicating that previous phases may have expanded the benefits limit to predict the outcome and therapy of patients. Therefore, it is necessary to classify patients with GC with a new approach in prevention and therapy based on Genome Analysis and clinical evidence (9). Because most GC cases are diagnosed in advanced stages when the prognosis is poor and therapeutic options are limited, GC continues being a significant cause of death globally with a high mortality rate (10).

Poor clinical symptoms, poor prognosis, lack of appropriate diagnostic techniques for early diagnosis, and limited biomarkers adversely affect survival rates (11). In the last decade, gene expression microarrays have become a common tool for studying gene expression levels in cancer research. Microarray data are used for a wide variety of analyzes, such as hierarchical clustering, classification, differential expression analysis, and expression mapping of quantitative property loci (12). The identification of biomarkers that can be used in early detection and prognosis of gastric cancer will provide clinically important contributions (13). In addition to diagnostic advantages, molecular analysis of gastric cancer can provide new improvements in the classification of gastric cancer cases individually. This specific and personal molecular classification of cases will make it possible to develop new therapeutic approaches.

In this study, it is aimed to determine the candidate biomarker genes that can be used to diagnose gastric cancer or to evaluate prognosis by performing bioinformatics analysis of genome wide transcriptomic data.

Material and Method

Affymetrix microarray data: The microarray data selected for bioinformatics analysis is stored with the code GSE54129 in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database. (14). This data set includes transcriptomic data of 111 human stomach cancer tumor tissues and 21 normal stomach tissue samples.

CEL files (raw data) of this study were loaded from the NCBI GEO and bioinformatics analyzes were performed on these files. All analyzes were performed with Transcriptome Analysis Console 4.0 (Applied Biosystems). With this dataset, the following analyzes were performed.

1- Preprocessing
2- Variance Filtering
3- Group comparison (Fold change <-4 +4> and p< 0.001)
4- Principle component analysis

As a result of the comparison, a list of genes which differentially expressed between cancer and control group was obtained. This gene list was loaded into DAVID Bioinformatics Tools algorithm and functional clustering analysis, gene set enrichment analysis and pathway analysis were performed.

Functional clustering and gene-set enrichment analysis: The gene list is loaded into the DAVID Bioinformatics Tools for functional cluster analysis, enrichment analysis and pathway analysis (15, 16). With this program, cluster analysis was performed and the genes in the list were clustered according to their functions. The program calculates an enrichment score for each cluster during the analysis. Clusters with an enrichment score greater than 1.3 are considered significant. Then, using this gene list, hierarchical cluster analysis was performed and heat map was obtained. Heatmap shows whether our samples were separated from each other using this list of genes. Then, pathway analyses were performed. Molecular pathways of the genes in the significant clusters were determined and the relation of these molecular pathways with the disease was evaluated.

Results

In the first analysis, the quality control of the data is performed. The scatter and volcano plots obtained from this analysis are shown in Figure 1A and Figure 1B, respectively. As a result of these analyzes, the average log2 fold change of each gene in the list of genes obtained by comparison is given in the scatter plot. X axis shows the control group used in the comparison, Y axis shows the cancer group (Figure 1A). The volcano plot given in Figure 1B shows the p value which calculated for each gene according to the fold change value.

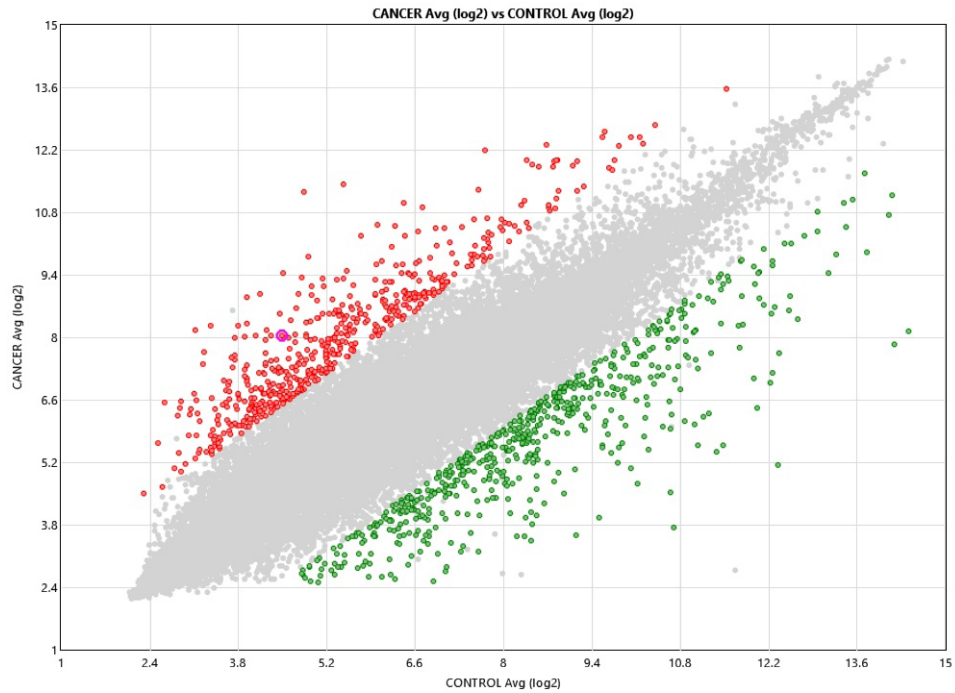

Figure 1A. Scatter plot from control and cancer comparison.

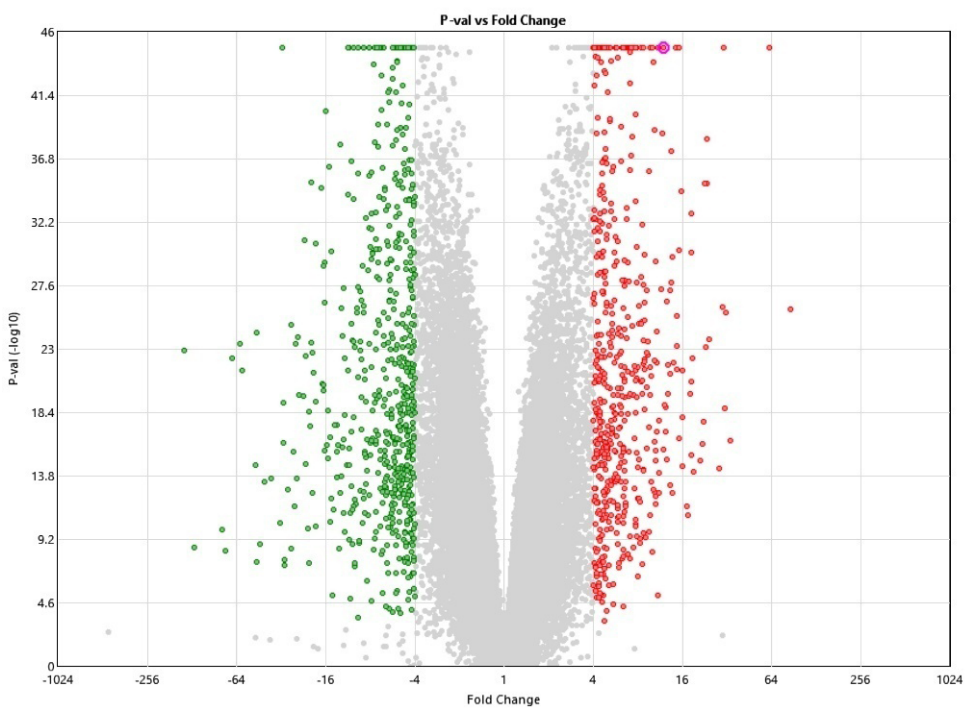

Figure 1B. Volcano plot obtained from control and cancer comparison.

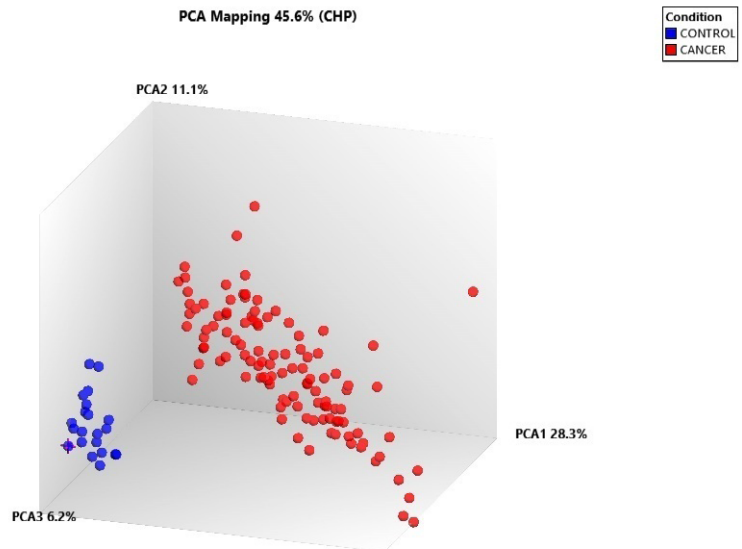

Figure 2A. PCA analysis with Cancer and Control samples.

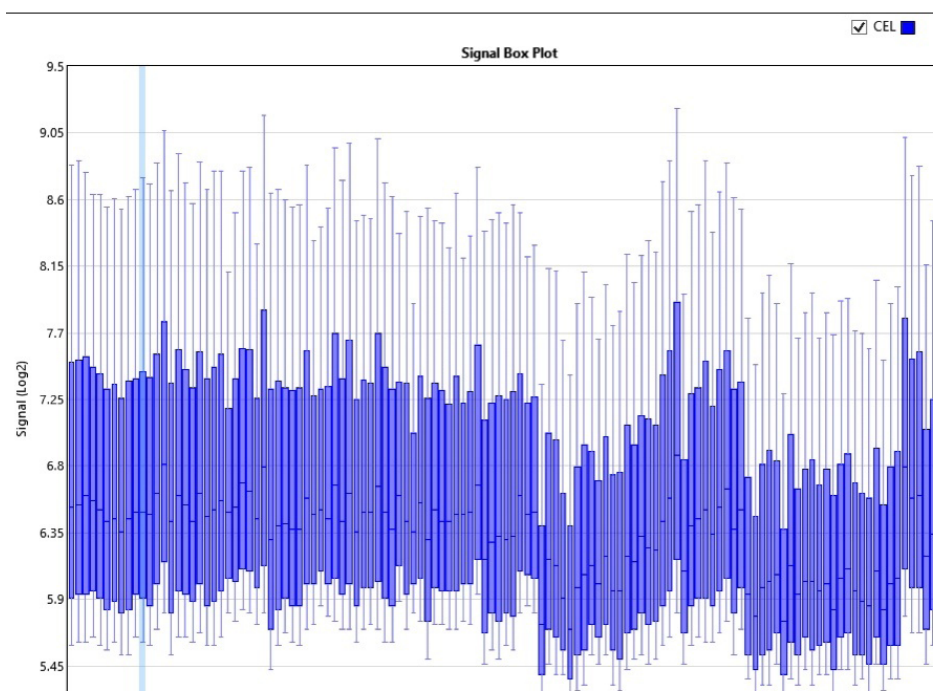

Figure 2B. Signal box plot of the samples.

Principal component analysis (PCA) and signal box plot are shown in Figures 2A and 2B. As a result of the PCA, the distribution of the samples we used is seen in 3-dimensional space and the analysis shows that the samples in the cancer and control groups are grouped separately. The signal box plot shows the signals comes from each sample.

Another analysis performed during the bioinformatics analysis was to compare the groups and identify the genes whose expression levels differ between the two groups. During comparison, the fold change threshold value was determined as ± 4 and p <0.001 value accepted as statistically significant. For each gene, p-value and false discovery rate (FDR) value are also calculated when

calculating the fold change between groups. Benjamini-Hochberg correction was applied on these values.

Table 1 shows the first 50 upregulated genes in the list of genes obtained from the comparison of cancer and control groups**.** Table 2 shows the first 50 downregulated genes obtained from the same comparison. As a result of the comparison of the cancer and control groups, 1113 differentially expressed genes were determined. It was determined that 526 of these genes upregulated and 587 of them downregulated.

Table 1. Top 50 upregulated genes in gastric cancer vs. control comparison

Gene Symbol	Description	Fold Change	P-val	FDR P val
SFRP2	secreted frizzled-related protein 2	85.96	1.33E-26	6.33E-25
EGR1	early growth response 1	62.27	1.40E-45	1.40E-45
CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)	33.47	4.47E-17	5.93E-16
COL8A1	collagen, type VIII, alpha 1	31.24	2.30E-26	1.07E-24
SFRP2	secreted frizzled-related protein 2	30.85	1.75E-19	3.21E-18
NEAT1	nuclear paraspeckle assembly transcript 1	30.36	1.40E-45	1.40E-45
INHBA	inhibin beta A	30.1	8.65E-27	4.24E-25
CXCL8	chemokine (C-X-C motif) ligand 8	28.4	4.00E-15	4.04E-14
MYL9	myosin light chain 9	24.18	1.84E-24	6.65E-23
C1S	complement component 1, s subcomponent	23.62	5.22E-39	1.34E-36
CYR61	cysteine-rich, angiogenic inducer, 61	23.42	1.01E-35	1.55E-33
GREM1	gremlin 1, DAN family BMP antagonist	23.19	6.49E-24	2.17E-22
COL1A1	collagen, type I, alpha 1	22.63	9.58E-36	1.47E-33
SFRP4	secreted frizzled-related protein 4	22.1	1.71E-18	2.75E-17
CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)	21.71	7.32E-17	9.44E-16
OGN	Osteoglycin	19.07	8.01E-15	7.81E-14
THBS2	thrombospondin 2	18.72	4.54E-23	1.36E-21
COL6A2	collagen, type VI, alpha 2	18.52	1.32E-33	1.52E-31
GREM1	gremlin 1, DAN family BMP antagonist	18.49	2.26E-21	5.30E-20
CYR61	cysteine-rich, angiogenic inducer, 61	18.23	1.02E-30	8.11E-29
PLA2G2A	phospholipase A2, group IIA (platelets, synovial fluid)	18.23	4.87E-16	5.58E-15
SLIT2	slit guidance ligand 2	17.98	1.82E-20	3.81E-19
CXCL13	chemokine (C-X-C motif) ligand 13	17.32	1.05E-11	6.80E-11
OGN	Osteoglycin	17.02	2.38E-12	1.67E-11
SPP1	secreted phosphoprotein 1	15.91	3.14E-15	3.22E-14
APOD	apolipoprotein D	15.89	8.53E-19	1.43E-17
SULF1	sulfatase 1	15.88	3.42E-22	8.98E-21
CD74	CD74 molecule, major histocompatibility complex, class II invariant chain	15.2	1.40E-45	2.10E-44
MGP	matrix Gla protein	15.19	3.02E-17	4.11E-16
RGS1	regulator of G-protein signaling 1	15.12	6.04E-31	5.00E-29
GXYLT2	glucoside xylosyltransferase 2	14.79	5.46E-24	1.85E-22
S100A8	S100 calcium binding protein A8	14.57	2.71E-22	7.21E-21

Table 1. Top 50 upregulated genes in gastric cancer vs. control comparison (*Continued*)

Gene Symbol	Description	Fold Change	P-val	FDR P val
FCGR3A; FCGR3B	Fc fragment of IgG, low affinity IIIa, receptor (CD16a)	14.56	3.96E-26	1.76E-24
PKM	pyruvate kinase, muscle	14.32	1.40E-45	1.40E-45
CTHRC1	collagen triple helix repeat containing 1	13.99	9.15E-23	2.62E-21
BGN	Biglycan	13.74	2.00E-30	1.52E-28
SCRG1	stimulator of chondrogenesis 1	13.63	2.54E-13	2.03E-12
DIO2	deiodinase, iodothyronine, type II	13.51	1.41E-28	8.58E-27
RGS1	regulator of G-protein signaling 1	13.51	4.00E-38	8.87E-36
C3	complement component 3	13.2	6.37E-22	1.61E-20
CCDC80; LINC01279	coiled-coil domain containing 80	13.14	6.93E-14	5.99E-13
PDLIM7	PDZ and LIM domain 7 (enigma)	13.13	5.67E-28	3.22E-26
WISP1	WNT1 inducible signaling pathway protein 1	12.75	1.80E-20	3.77E-19
SULF1	sulfatase 1	12.67	1.36E-22	3.79E-21
AQP1	aquaporin 1 (Colton blood group)	12.6	3.54E-27	1.82E-25
HOXC6	homeobox C6	12.55	8.94E-16	9.88E-15
TAGLN	Transgelin	12.43	1.45E-19	2.68E-18
ITGBL1	integrin beta like 1	12.36	9.51E-18	1.39E-16
TAGLN	Transgelin	12.17	1.09E-19	2.05E-18
MALAT1	metastasis associated lung adenocarcinoma transcript 1 (non-protein coding)	11.84	1.40E-45	1.40E-45

Table 2. Top 50 downregulated genes in gastric cancer vs. control comparison

Gene Symbol	Description	Fold Change	P-val	FDR P val
GAST	gastrin	-145.01	1.28E-23	4.08E-22
GIF	gastric intrinsic factor (vitamin B synthesis)	-123.15	2.45E-09	1.15E-08
GKN2	gastrokine 2	-80.63	1.24E-10	6.94E-10
GKN1	gastrokine 1	-76.08	3.85E-09	1.76E-08
SCGB2A1	secretoglobin, family 2A, member 1	-68.84	4.37E-23	1.31E-21
HRASLS2	HRAS-like suppressor 2	-60.76	3.72E-24	1.29E-22
UPK1B	uroplakin 1B	-58.78	3.40E-22	8.94E-21
SST	somatostatin	-47.97	2.62E-15	2.71E-14
ATP4B	ATPase, H+/K+ exchanging, beta polypeptide	-47.02	2.56E-08	1.04E-07
CHGB	chromogranin B	-46.55	6.52E-25	2.50E-23
C6orf58	chromosome 6 open reading frame 58	-44.8	1.23E-09	6.00E-09
FUT9	fucosyltransferase 9 (alpha (1,3) fucosyltransferase)	-41.85	3.68E-14	3.29E-13
DPCR1	diffuse panbronchiolitis critical region 1	-37.42	2.43E-14	2.23E-13
KRT20	keratin 20, type I	-32.73	3.59E-11	2.17E-10
SLC28A2	solute carrier family 28 (concentrative nucleoside transporter), member 2	-30.97	6.48E-17	8.41E-16
UPK1B	uroplakin 1B	-30.76	6.83E-20	1.32E-18

Table 1. Top 50 downregulated genes in gastric cancer vs. control comparison (*Continued*)

Gene Symbol	Description	Fold Change	P-val	FDR P val
MSMB	microseminoprotein, beta-	-30.59	1.75E-08	7.27E-08
MSMB	microseminoprotein, beta-	-30.23	4.25E-08	1.67E-07
DPCR1	diffuse panbronchiolitis critical region 1	-29.06	1.59E-13	1.30E-12
CAPN9	calpain 9	-27.52	1.54E-25	6.32E-24
ORM1; ORM2	orosomucoid 1; orosomucoid 2	-27.47	2.98E-09	1.39E-08
SOSTDC1	sclerostin domain containing 1	-26.62	2.72E-16	3.24E-15
VSIG1	V-set and immunoglobulin domain containing 1	-26.13	2.36E-12	1.66E-11
C6orf58	chromosome 6 open reading frame 58	-25.26	2.16E-14	1.99E-13
CAPN13	calpain 13	-24.69	1.19E-24	4.39E-23
CAPN9	calpain 9	-22.72	2.61E-20	5.32E-19
SCIN	scinderin	-22.42	1.24E-31	1.12E-29
ADGRG2	adhesion G protein-coupled receptor G2	-22.01	3.22E-23	9.76E-22
VSIG1	V-set and immunoglobulin domain containing 1	-21.75	1.89E-15	2.00E-14
GC	group-specific component (vitamin D binding protein)	-21.1	1.04E-10	5.88E-10
SYTL5	synaptotagmin-like 5	-20.78	3.33E-19	5.89E-18
ATP4A	ATPase, H+/K+ exchanging, alpha polypeptide	-20.6	3.05E-08	1.22E-07
MAP7D2	MAP7 domain containing 2	-20.44	3.81E-18	5.87E-17
CWH43	cell wall biogenesis 43 C-terminal homolog	-20.16	8.05E-36	1.26E-33
AKR1C1; AKR1C2	aldo-keto reductase family 1, member C1; aldo-keto reductase family 1, member C2	-20.13	3.11E-24	1.09E-22
KCNJ16	potassium channel, inwardly rectifying subfamily J, member 16	-19.81	2.80E-13	2.22E-12
SCIN	scinderin	-19.6	1.62E-23	5.11E-22
HPGD	hydroxyprostaglandin dehydrogenase 15-(NAD)	-19.24	5.33E-22	1.36E-20
TMEM27	transmembrane protein 27	-18.76	2.05E-31	1.79E-29
MUC5AC	mucin 5AC, oligomeric mucus/gel-forming	-18.68	6.53E-11	3.80E-10
SSTR1	somatostatin receptor 1	-18.57	6.33E-20	1.23E-18
CWH43	cell wall biogenesis 43 C-terminal homolog	-17.24	2.06E-35	2.99E-33
APOBEC1	apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1	-16.75	2.92E-21	6.76E-20
LRRC66	leucine rich repeat containing 66	-16.48	4.00E-21	9.04E-20
SMIM6	small integral membrane protein 6	-16.48	9.17E-30	6.39E-28
ERICH5	glutamate rich 5	-16.47	1.02E-20	2.20E-19
STYK1	serine/threonine/tyrosine kinase 1	-16.36	5.52E-30	3.93E-28
UGT2B15	UDP glucuronosyltransferase 2 family, polypeptide B15	-16.25	2.21E-16	2.67E-15
PKIB	protein kinase (cAMP-dependent, catalytic) inhibitor beta	-16.2	4.04E-27	2.06E-25
SULT1C2	sulfotransferase family 1C member 2	-16.07	5.81E-41	1.86E-38

Hierarchical clustering analysis was performed with the obtained gene lists. Heat Map obtained from this analysis is shown in Figure 3. This analysis was performed using 1113 genes determined as a result of cancer and control group comparison. When we look at the results of this analysis, a complete clustering was determined. It is clear that the samples in the cancer group and the samples in the control group have a unique gene expression profile.

Afterwards, pathway analyzes were performed using gene lists obtained from group comparisons. Table 3 shows the pathway analysis results performed with the list of genes obtained from cancer and control group comparison**.** In the given table, the pathways determined as a result of the analysis and the number of genes sitting in these pathways are shown. The pathway contained the maximum number of genes during the analysis is given first. This results shows that different molecular pathways play an important role in the disease of interest.

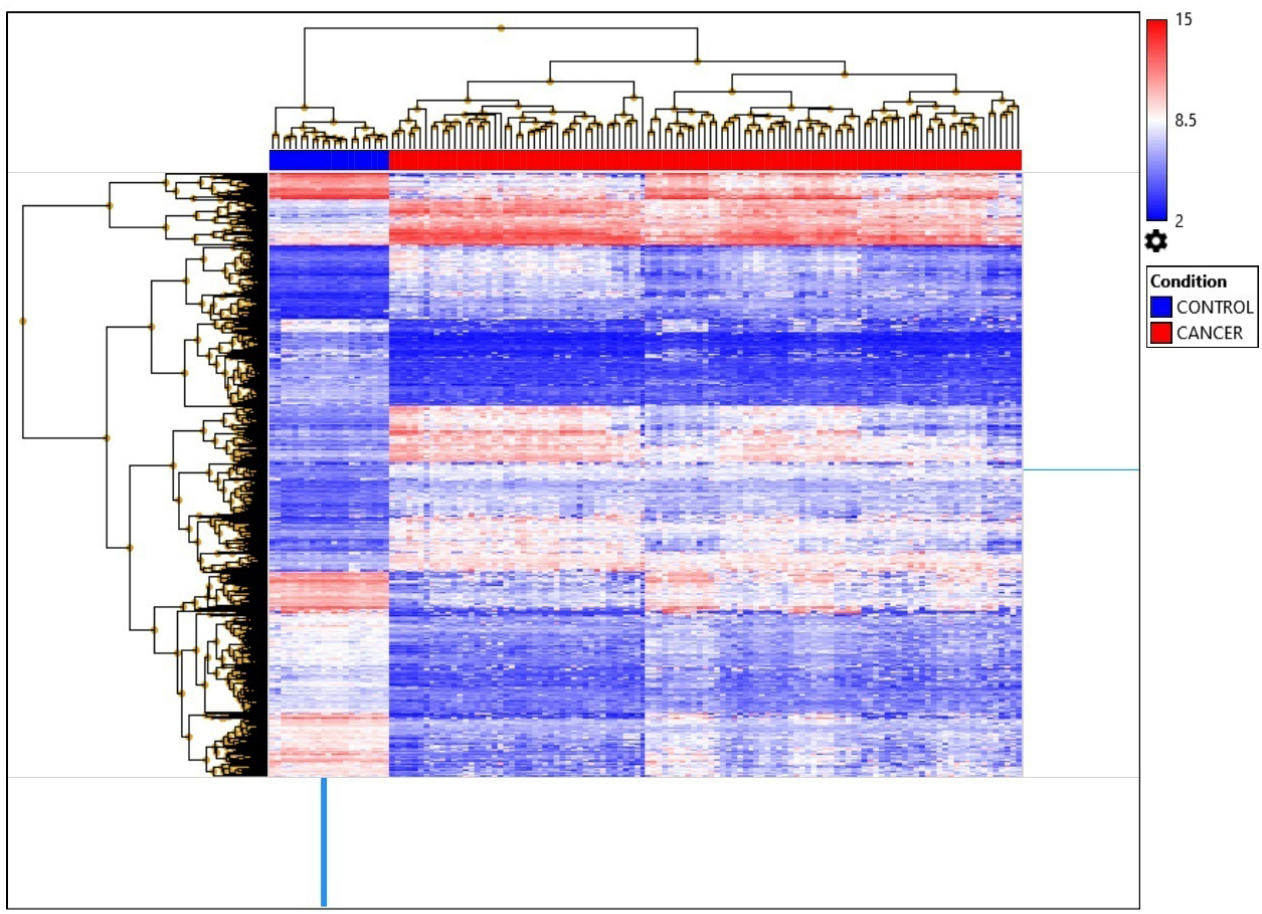

Figure 3. Heat map obtained from hierarchical cluster analysis using the gene list obtained from the comparison of cancer and control group.

Table 3. Pathway analysis results performed with gene list obtained from cancer and control group comparison

Category	Term	Gene Count	P-Value	Benjamini
KEGG_PATHWAY	Metabolism of xenobiotic by cytochrome P450	20	7,7E-10	1,9E-7
KEGG_PATHWAY	Chemical carcinogenesis	20	3,2E-9	4,0E-7
KEGG_PATHWAY	Drug metabolism - cytochrome P450	17	6,7E-8	5,6E-6
KEGG_PATHWAY	Protein digestion and absorption	18	5,5E-7	3,4E-5
KEGG_PATHWAY	Retinol metabolism	14	6,8E-6	3,4E-4
KEGG_PATHWAY	ECM-receptor interaction	16	1,1E-5	4,7E-4
KEGG_PATHWAY	Focal adhesion	26	1,2E-5	4,1E-4
KEGG_PATHWAY	Steroid hormone biosynthesis	13	1,3E-5	3,9E-4
KEGG_PATHWAY	Gastric acid secretion	12	5,7E-4	1,6E-2
KEGG_PATHWAY	Staphylococcus aureus infection	10	8,5E-4	2,1E-2
KEGG_PATHWAY	Leukocyte transendothelial migration	15	9,8E-4	2,2E-2
KEGG_PATHWAY	Cytokine-cytokine receptor interaction	23	2,4E-3	5,0E-2
KEGG_PATHWAY	PI3K-Akt signaling pathway	28	6,4E-3	1,2E-1
KEGG_PATHWAY	Maturity onset diabetes of the young	6	6,6E-3	1,1E-1
KEGG_PATHWAY	Arachidonic acid metabolism	9	7,5E-3	1,2E-1
KEGG_PATHWAY	Pertussis	10	8,4E-3	1,2E-1
KEGG_PATHWAY	Vascular smooth muscle contraction	13	8,9E-3	1,2E-1
KEGG_PATHWAY	Amoebiasis	12	1,1E-2	1,4E-1
KEGG_PATHWAY	Complement and coagulation cascades	9	1,5E-2	1,9E-1
KEGG_PATHWAY	Ether lipid metabolism	7	1,8E-2	2,1E-1
KEGG_PATHWAY	Tight junction	10	2,1E-2	2,2E-1
KEGG_PATHWAY	Tyrosine metabolism	6	2,3E-2	2,3E-1
KEGG_PATHWAY	Ascorbate and aldarate metabolism	5	3,6E-2	3,3E-1
KEGG_PATHWAY	Leishmaniasis	8	4,9E-2	4,1E-1
KEGG_PATHWAY	Drug metabolism - other enzymes	6	6,4E-2	4,9E-1
KEGG_PATHWAY	Pentose and glucuronate interconversions	5	6,8E-2	4,9E-1
KEGG_PATHWAY	cGMP-PKG signaling pathway	13	7,0E-2	4,9E-1
KEGG_PATHWAY	Malaria	6	8,0E-2	5,2E-1
KEGG_PATHWAY	Renin secretion	7	8,0E-2	5,1E-1
KEGG_PATHWAY	Osteoclast differentiation	11	9,0E-2	5,5E-1
KEGG_PATHWAY	Proximal tubule bicarbonate reclamation	4	9,2E-2	5,4E-1
KEGG_PATHWAY	Metabolic pathways	68	9,4E-2	5,4E-1
KEGG_PATHWAY	Glycolysis / Gluconeogenesis	7	9,5E-2	5,3E-1
KEGG_PATHWAY	Phagosome	12	9,7E-2	5,3E-1

Finally, Gene-Set Enrichment Analysis (GSEA) interpretable was performed on gene lists. DAVID Bioinformatics Tools was used during this analysis. DAVID allows converting a very high number of genes into a an interpretable list of fewer genes. During this analysis, the genes in the list are clustered according to their functions. The most important genes and clusters stand out. In this way, the genes that are interested and

important were determined. Table 4 shows the results of functional cluster analysis performed with the list of cancer and control groups. As a result of the analysis, 103 functional clusters were determined. Table 4 shows the first 5 clusters with the biological terms and gene numbers in these clusters. The enrichment score of the most important cluster obtained was calculated as 7.6. All genes in each cluster were investigated one by one with literature search and interested genes were detected.

Table 4. Functional Cluster Analysis that shows first five annotation clusters

Annotation Cluster 1	***Enrichment Score: 7.6***	***Gene Count***	***P_Value***	***Benjamini***
Biological Terms Database	***Biological Terms***			
UP_KEYWORDS	Collagen	20	2.30E-09	1.40E-07
INTERPRO	Collagen triple helix repeat	18	1.10E-08	1.40E-05
GOTERM_CC_DIRECT	collagen trimer	17	6.30E-07	3.90E-05
Annotation Cluster 2	***Enrichment Score: 4.17***	***Gene Count***	***P_Value***	***Benjamini***
Biological Terms Database	***Biological Terms***			
UP_SEQ_FEATURE	domain:CTCK	8	2.40E-05	7.40E-03
INTERPRO	Cystine knot, C-terminal	8	4.20E-05	1.00E-02
SMART	CT	7	3.10E-04	2.90E-02
Annotation Cluster 3	***Enrichment Score: 3.94***	***Gene Count***	***P_Value***	***Benjamini***
Biological Terms Database	***Biological Terms***			
INTERPRO	Netrin domain	8	1.70E-05	5.30E-03
INTERPRO	Tissue inhibitor of metalloproteinases-like, OB-fold	8	4.20E-05	1.00E-02
UP_SEQ_FEATURE	domain:NTR	7	1.10E-04	2.40E-02
INTERPRO	Netrin module, non-TIMP type	6	4.70E-04	6.90E-02
SMART	C345C	6	5.50E-04	2.50E-02
Annotation Cluster 4	***Enrichment Score: 3.09***	***Gene Count***	***P_Value***	***Benjamini***
Biological Terms Database	***Biological Terms***			
INTERPRO	Chemokine interleukin-8-like domain	10	5.40E-05	1.10E-02
GOTERM_MF_DIRECT	chemokine activity	10	1.20E-04	1.80E-02
SMART	SCY	9	3.90E-04	2.70E-02
KEGG_PATHWAY	Chemokine signaling pathway	13	1.70E-01	6.90E-01
Annotation Cluster 5	***Enrichment Score: 2.84***	***Gene Count***	***P_Value***	***Benjamini***
Biological Terms Database	***Biological Terms***			
UP_SEQ_FEATURE	domain:TSP type-1 2	8	1.20E-03	1.40E-01
UP_SEQ_FEATURE	domain:TSP type-1 1	8	1.20E-03	1.40E-01
UP_SEQ_FEATURE	domain:TSP type-1 3	7	2.10E-03	2.00E-01

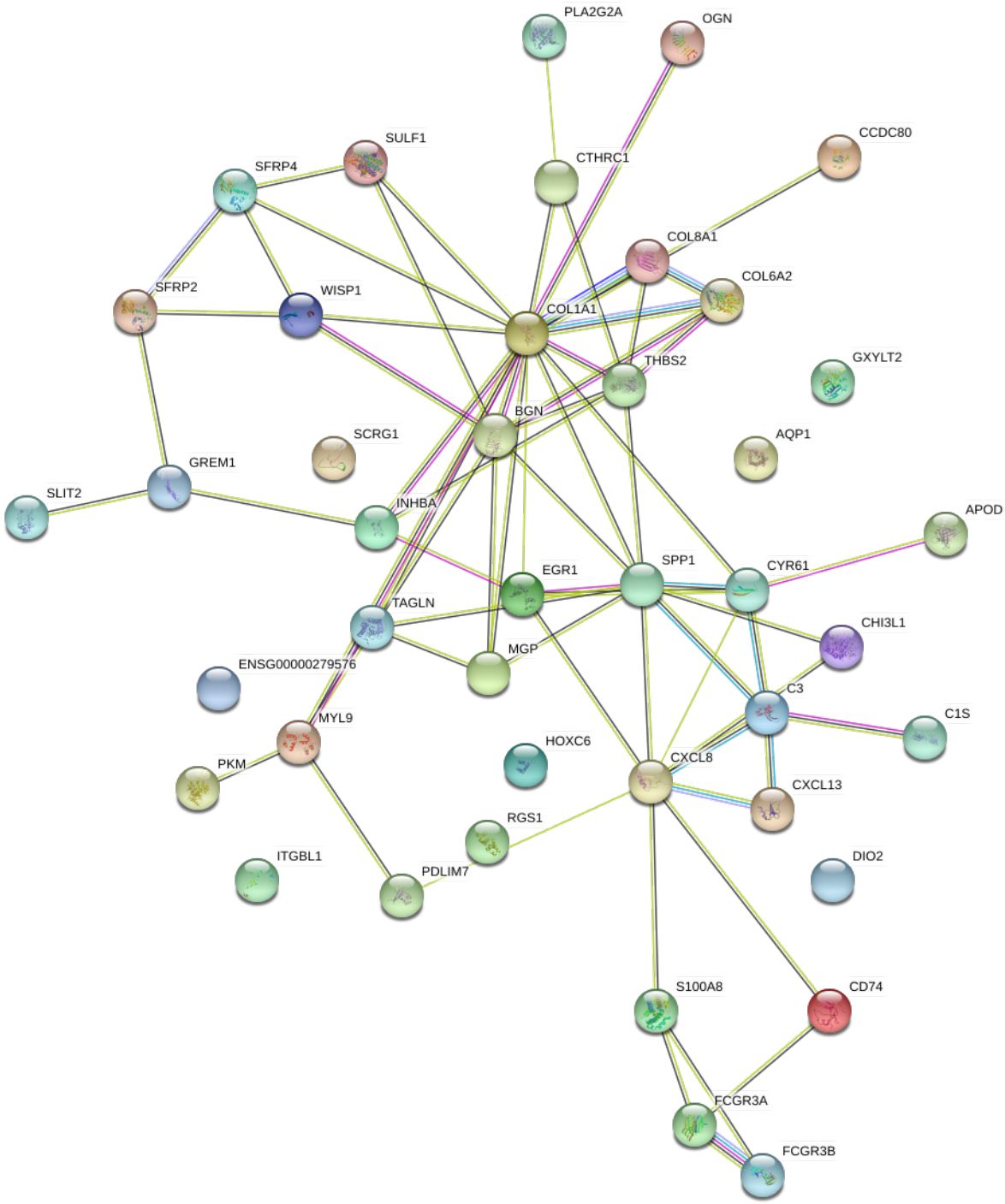

Figure 4. Protein-protein interaction network of the upregulated genes between cancer and control tissue samples.

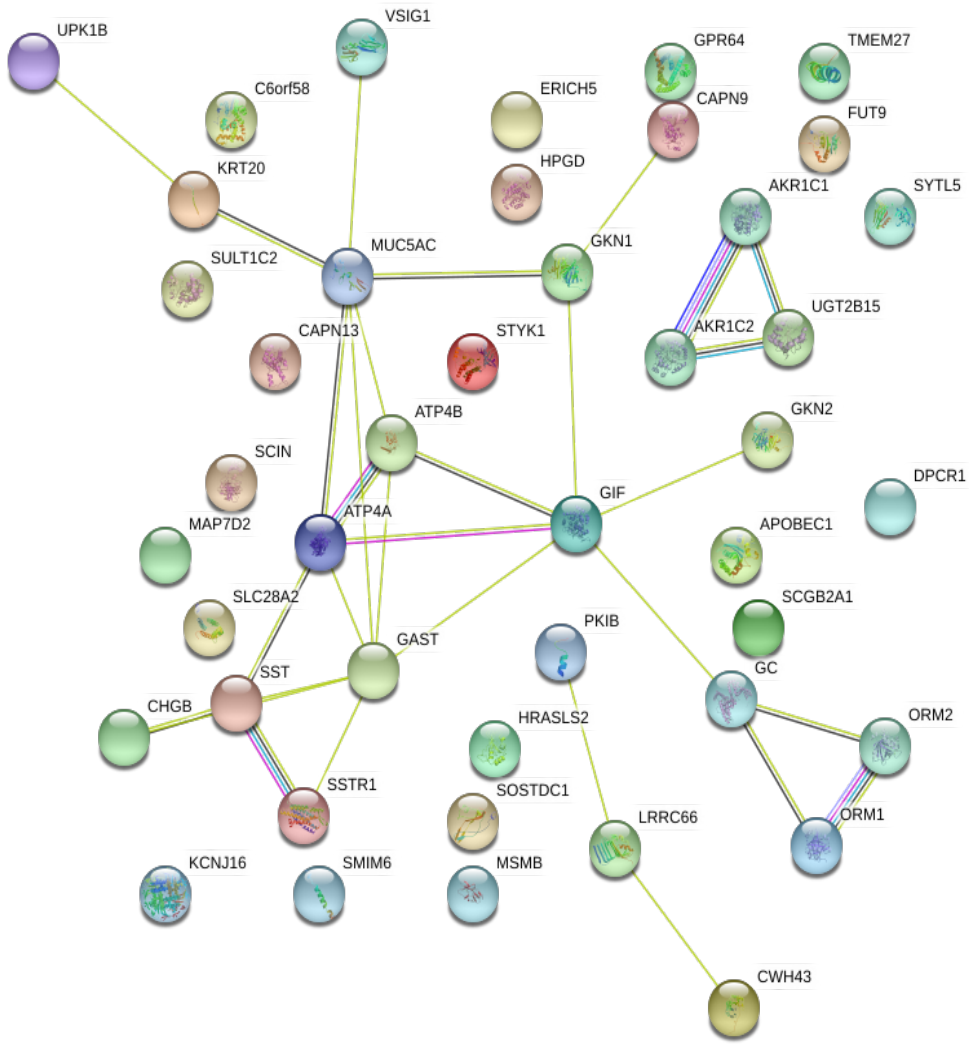

Figure 5. Protein-protein interaction network of the downregulated genes between cancer and control tissue samples.

Discussion

We have already known that the expression level of many genes varies during neoplastic transformation compared to healthy individuals. As a result of bioinformatics analysis, genes with varying levels of expression compared to normal tissue were detected in gastric cancer tissue. Enrichment and functional cluster analyzes was performed with these genes and many prominent genes have been identified. We have showed that the expression level of some of these genes has been shown to significant increase in gastric cancer tissue compared to normal tissue like as *EGR1, CHI3L1, COL8A1, SFRP2, NEAT1, INHBA, CXCL8, MYL9, C1S, CYR61*. Some of the genes we have identified are those that have not yet been specifically studied for gastric cancer.

Genes prominent in the study; *SFRP2* (secreted frizzled-related protein 2), *SFRP4* (secreted frizzled-related protein 4), *SFRP1* (secreted frizzled-related protein 1), *EGR1* (early growth response 1), *COL8A1* (collagen, type VIII, alpha 1), *CXCL8* (chemokine (CXC motif) ligand 8), *C1S* (complement component 1,subcomponent), *COL1A1* (collagen, type I, alpha 1), *THBS2* (thrombospondin 2), *THBS1* (thrombospondin 1), *THBS4* (thrombospondin 4), *ITGBL1* (integrin

beta like 1), *HSPB6* (heat shock protein, alpha-crystallin-related, B6), *FN1* (fibronectin 1), *CEMIP* (cell migration inducing protein, hyaluronan binding) and *RHOB* (ras homologous family member B).

Secreted frizzled-related protein 1 (*SFRP1*) is a gene that belongs to the secreted glycoprotein *SFRP* family (17). A total of 5 *SRFP* genes have been identified in humans. These are *SRFP1, SRFP2, SRFP3, SRFP4* and *SRFP5. SRFP1, SRFP2* and *SRFP4* draw attention in this gene family. These genes show upregulation especially in cancer (17, 25). One of these genes is *SFRP1* and this gene often expressed as a tumor suppressor gene. Because some studies showed that decreased expression levels of this gene in prostate cancer and renal cancers were related to poor prognosis and recurrence (18, 19). Some studies have shown an increase in expression levels, especially in gastric cancers (27). Therefore, there are questionable results in the literature, especially regarding the *SFRP1* gene. The expression level of this gene may show difference depend on the cancer type. According to our results, the expression level of the *SFRP1* gene increased in the cancer group, corroborating the previoulsy published studies. In addition to this gene, the expression levels of *SFRP2* and *SFRP4* genes, which are from the same gene family, showed increased expression levels in our cancer group.

One of these genes, the *MALAT1*, translates a long noncoding RNA (lncRNA) and known as a gene that has been specifically identified as being associated with cancer. Jin et. al. showed that *MALAT1* improves cell proliferation and metastatic features in epithelial ovarian cancer via PI3K-Akt pathway (20). In our study, this gene is among the first 50 upregulated genes. Our pathway analysis results show that PI3K-Akt pathway is statistically significant and 28 genes were showed expression level change within this molecular pathway. Xia et.al. also reported that metastatic lung adenocarcinoma transcript 1 (*MALAT1*) is frequently overexpressed and serves as a prognostic marker in human cancers. Tissue and plasma *MALAT1* levels were significantly higher in gastric cancer patients with distant metastases than in non-distant metastases and healthy controls. They also found that high levels of plasma *MALAT1* were independently associated with poor prognosis for gastric cancer patients. Reported that lowering *MALAT1* expression levels may inhibit cell proliferation, cell cycle progression, migration and invasion, and may promote apoptosis in gastric cancer cells. miR-122-IGF-1R signals correlate with irregular *MALAT1* expression in gastric cancer. These data indicate that *MALAT1* can function as an oncogene in gastric cancer, and that a high level expression of *MALAT1* can serve as a potential biomarker for distant metastasis of gastric cancer (21).

Another important set of genes which shows upregulation in gastric cancer tissues are thrombospondins. Only thrombospondin 2 (*THBS2*) was seen in our top 50 upregulated list but *THBS1* and *THBS4* also showed upregulation in cancer tissues. *THSB1* is an extracellular glycoprotein receptor and regulate cell-matrix and cell-cell interaction. Huang et.al. was demonstrated that *THSB1* is upregulated in gastric cancer tissues and promotes invasion and metastasis (22) In addition to *THSB1*, the genes *THSB2* and *THSB4* showed high expression levels in our analysis.

Pathway analysis were shown that the most affected pathways in gastric cancer pathogenesis are PI3K-Akt pathway, Focal adhesion, Chemical carcinogenesis, and Cytokine-cytokine interaction pathways. In these pathways almost more than 20 genes expression levels were changed during neoplastic transformation. PI3K-Akt pathway has a fundamental role in gastric carcinogenesis as shown previously (28). In this pathway, upon the activation of PI3K, phosphatidylinositol-3,4,5-triphosphate (PIP3) is produced and recruit a set of signaling proteins including Akt to cell membrane which regulate many cellular processes, for example, cell survival, proliferation and growth, and cellular metabolism (23-24).

Conclusion

As a result of the bioinformatics analysis, the target gene list has been determined and it is planned to validate the candidate biomarker genes which were determined as a result of functional clustering analyzes and pathway analyzes, by quantitative real-time PCR (qRT-PCR) in a patient group. We will determine how these genes play a role in the emergence of the disease with molecular analysis to be performed on

the validated genes. In this way, both new information about the molecular mechanism of the disease will be obtained and new drug target molecules will be identified.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

References

1. Petrovchich I, and Ford JM. Genetic predisposition to gastric cancer. Semin Oncol. 2016: 43(5):554–559.
2. Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F,. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. Cancer Epidemiol Biomarkers Prev. 2014: 23(5):700–13.
3. Baretton GB, and Aust DE. Current biomarkers for gastric cancer. Pathologe. 2017: 38(2):93–97.
4. Çaycı HM, Erdoğlu UE, Çantay H, Orman S, Mustafa AKAR, & Demirci H,. Mide kanseri deneyimlerimiz: Tanı ve tedavide geç mi kalıyoruz?. Akademik gastroenteroloji dergisi, 2017: *16*(1), 6–11.
5. Alacalı M. Mide kanseri, mide kanseri taramaları ve mide kanserinden korunma. Ankara Medical Journal, 2012: *12*(4).
6. Jemal A, Bray F, Center MM, Ferlay J, Ward E, & Forman D. Global cancer statistics. CA: a cancer journal for clinicians, 2011: *61*(2), 69–90.
7. Göral V, Yeğilbağlan H, Kaplan A, it D, & Çelik M. Mide kanserinde yeni bir tümör markeri olan CA 72-4'ün yeri. Turkiye Klinikleri Journal of Medical Sciences, 2006: *26*(1), 3–8.
8. Yüksel BC, Uçar NS, Yıldız Y, Berkem H, Özel H, & Hengirmen S. Mide kanserinde standart D2 diseksiyona karşı D1 diseksiyonun mortalite ve morbidite çalışması. Turkish Journal of Surgery/Ulusal Cerrahi Dergisi, 2009: *25*(3).
9. Matsuoka T. and Yashiro M,. "Biomarkers of gastric cancer: current topics and future perspective." World J Gastroenterol . 2018: 24(26): 2818.
10. Necula L, Matei L, et al. "Recent advances in gastric cancer early diagnosis." World J Gastroenterol 2019: 25(17): 2029.
11. Tunca B, Aksoy SA, Mutlu M, & Tekin Ç. Mide Kanserinde Genetik ve Epigenetik Mekanizmaların Rolü. Turkiye Klinikleri Radiation Oncology-Special Topics, 2019: *5*(1), 7–14.
12. Li H, Yu B, Li J, Su L, Yan M, Zhang J, Li Cı, Zhu Z, Liu B,. Characterization of differentially expressed genes involved in pathways associated with gastric cancer. PLoS One. 2015: 30;10(4).
13. Shekari N, Baradaran B, Shanehbandi D, Kazemi T,. Circulating MicroRNAs: Valuable Biomarkers for the Diagnosis and Prognosis of Gastric Cancer. Curr Med Chem. 2018: 25(6):698-714.
14. Yi, Zhu., Xiangwei, Sun., Ji, Lin., Teming, Zhang., Xin, Liu., Xian, Shen., Investigating Potential Molecular Mechanisms of Carcinogenesis and Genes as Biomarkers for Prognosis of Gastric Cancer Based on Integrated Bioinformatics Analysis. Pathology & Oncology Research 2019: 25:1125–1133.
15. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009: 4(1):44–57.
16. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009: 37(1):1–13.
17. Baharudin, R., Tieng, F. Y. F., Lee, L. H., & Ab Mutalib, N. S. Epigenetics of SFRP1: The Dual Roles in Human Cancers. Cancers. 2020: 12(2)–445.
18. Bovolenta P, Esteve P, Ruiz JM, Cisneros E, Lopez-Rios J. Beyond Wnt inhibition: New functions of secreted Frizzled-related proteins in development and disease. J. Cell Sci. 2008: 121, 737–746.
19. Atschekzei F, Hennenlotter J, Jänisch S, Großhennig A, Tränkenschuh W, Waalkes S, Peters I, Dörk T, Merseburger AS, Stenzl A, et al. SFRP1 CpG island methylation locus is associated with renal cell cancer susceptibility and disease recurrence. Epigenetics 2012: 7, 447–457.
20. Jin Y, Feng S.J, Qiu S, Shao N, Zheng J.H. LncRNA MALAT1 promotes proliferation and metastasis in epithelial ovarian cancer via the PI3K-AKT pathway. Eur Rev Med Pharmacol Sci. 2017: 21(14), 3176–3184.
21. Xia H, Chen Q, Chen Y, Ge X, Leng W, Tang Q, Ren M, Chen L, Yuan D, Zhang Y, Liu M, Gong Q, Bi F. The lncRNA MALAT1 is a novel biomarker for gastric cancer metastasis. Oncotarget. 2016: 7(35):56209–56218.
22. Huang T, Wang L, Liu D, Li P, Xiong H, Zhuang L, Qiu H. FGF7/FGFR2 signal promotes invasion and migration in human gastric cancer through upregulation of thrombospondin-1. International journal of oncology. 2017: 50(5), 1501–1512.
23. Dahia, Patricia L. PTEN, a unique tumor suppressor gene. *Endocrine-related cancer*. 2000; 7.2: 115–129.
24. Manning, Brendan D., and Lewis C. Cantley. "AKT/PKB signaling: navigating downstream." Cell. 2007: 129.7 1261-127.
25. Cruz-Hernández CD, Cruz-Burgos M, Cortés-Ramírez SA, et al. SFRP1 increases TMPRSS2-ERG expression promoting neoplastic features in prostate cancer in vitro and in vivo. Cancer Cell Int. 2020;20:312. Published 2020 Jul 16. doi:10.1186/s12935-020-01333-5.
26. Liang CJ, Wang ZW, Chang YW, Lee KC, Lin WH, Lee JL. SFRPs Are Biphasic Modulators of Wnt-Signaling-Elicited Cancer Stem Cell Properties beyond Ex-

tracellular Control. Cell Rep. 2019;28(6):1511-1525.e5. doi:10.1016/j.celrep.2019.07.023.

27. Lin H, Yang G, Ding B, et al. Secreted frizzled-related protein 1 overexpression in gastric cancer: Relationship with radiological findings of dual-energy spectral CT and PET-CT. Sci Rep. 2017;7:42020. Published 2017 Feb 7. doi:10.1038/srep42020.
28. Huang Y, Zhang J, Hou L, et al. Correction to: LncRNA AK023391 promotes tumorigenesis and invasion of gastric cancer through activation of the PI3K/Akt signaling pathway. J Exp Clin Cancer Res. 2020;39(1):155. Published 2020 Aug 12. doi:10.1186/s13046-020-01656-1.