Scientific Research Publishing

Yangtze Medicine, 2021, 5, 141-156
https://www.scirp.org/journal/ym
ISSN Online: 2475-7349
ISSN Print: 2475-7330

Bioinformatics Analysis Revealed Potential Tumor Suppressors (KLF4/CGN), Oncogenes (SHH/LIF) and Biomarkers of Asian Stomach Adenocarcinoma

Yang Zhou[1,2*], Yingying Wang[3*], Junting Cheng[1,2], Ying Zhang[1,2], Wenqi Cai[1,2], Ziwen Han[1,2], Moyu Wang[1,2], Qi Huang[1,2], Xiaochun Peng[1,4#], Hongwu Xin[1,2#]

[1]Laboratory of Oncology, Center for Molecular Medicine, School of Basic Medicine, Health Science Center, Yangtze University, Jingzhou, China
[2]Department of Biochemistry and Molecular Biology, School of Basic Medicine, Health Science Center, Yangtze University, Jingzhou, China
[3]Department of Gynaecology, Comprehensive Cancer Center, Hannover Medical School, Hannover, Germany
[4]Department of Pathophysiology, School of Basic Medicine, Health Science Center, Yangtze University, Jingzhou, China
Email: [#]hongwu_xin@126.com, [#]pxcwd789@sina.com

How to cite this paper: Zhou, Y., Wang, Y.Y., Cheng, J.T., Zhang, Y., Cai, W.Q., Han, Z.W., Wang, M.Y., Huang, Q., Peng, X.C. and Xin, H.W. (2021) Bioinformatics Analysis Revealed Potential Tumor Suppressors (KLF4/CGN), Oncogenes (SHH/LIF) and Biomarkers of Asian Stomach Adenocarcinoma. *Yangtze Medicine*, **5**, 141-156.
https://doi.org/10.4236/ym.2021.52015

Received: December 3, 2020
Accepted: April 12, 2021
Published: April 15, 2021

Abstract

Stomach adenocarcinoma (STAD) is the fifth most prevalent cancer and the third leading cause of cancer-related death in the world and is more common in Asia than in most Western countries. There is an urgent need to identify potential novel oncogenes and tumor suppressor genes, and biomarkers for STAD. 6652 differentially expressed genes were identified between STAD and normal samples based on the transcriptome data analysis of the TCGA and GEO databases. 13 key modules were identified in STAD by WGCNA analysis. 293 potential STAD associated genes were identified from intersection by Venn Diagram. The 293 intersected genes were enriched in cell cortex and infection by GO and KEGG analysis. 10 hub genes were identified from PPI and Cytoscape analyses of the intersected genes. KLF4/CGN low and SHH/LIF high expression were associated with short overall survival of Asian STAD patients. Bioinformatics analysis revealed potential novel tumor suppressors (KLF4/CGN), oncogenes (SHH/LIF) and biomarkers for diagnosis, therapy and prognosis of STAD, specifically for Asian patients.

Keywords

WGCNA (Weighted Correlation Network Analysis), Tumor Suppressors, Oncogenes, Stomach Adenocarcinoma (STAD), Hub Gene

*Authors contributed equally.
#Corresponding authors.

1. Introduction

More than one million people worldwide are diagnosed with gastric adenocarcinoma or stomach adenocarcinoma (STAD) each year. STAD is the fifth most prevalent cancer and the third leading cause of cancer-related death in the world. Gastric cancer is more common in Asia than in most Western countries. The incidence in East Asia is about 50 cases per 100,000 people, about 10 times more than that in North America [1] [2]. There is an urgent need to identify potential novel oncogenes and tumor suppressor genes, and biomarkers of diagnosis, therapy and prognosis specifically for Asian STAD.

Weighted correlation network analysis (WGCNA) [3] is a data reduction and unsupervised classification method. It simplifies the interpretation of many gene responses to multiple synthetic genomes (or modules). The net establishes a link between genes whose expression is related. There will be connections between genes, depending on the value of the correlation (weight). Connectivity between genes is then interpreted as distance, with which genes are grouped into modules. This is the way to reduce many genes to several clusters, the expression of which is quantified by Eigengenes (the first principal component in the module). It assumes that highly related genes in the module are involved in a common biological process.

At present, there are only 4 articles in PubMed in which WGCNA was used to study the gastric cancer transcriptome and none of the 4 articles specifically studied the Asian human STAD transcriptome [4] [5] [6] [7]. Here we used WGCNA together with other bioinformatic tools to identify potential novel oncogenes, tumor suppressor genes, and biomarkers of diagnosis, therapy and prognosis specifically for Asian STAD.

2. Materials and Methods

2.1. Installation of R and Perl

R x64 4.03 language installation package were downloaded from https://www.r-project.org/, Perl installation package downloaded from https://www.perl.org/, and they were installed as instructed.

2.2. Datasets from TCGA

With filters of Cases (Stomach, TCGA, TCGA-STAD, Adenomas and Adeno Carcinomas and Asian) and Files (Transcriptome Profiling, Gene Expression Quantification, HTSeq-Counts and Txt), 74 samples (7 normal tissues and 67 tumors) were filtered out from the TCGA database (https://portal.gdc.cancer.gov/) and the required files were exported via Cart. The exported files were decompressed, and the 74 samples were processed through Perl to merge them together and calculated to get a Matrix file. The matrix file ENSG ID then was converted into Symbol ID by R language.

With filters of Cases (Stomach, TCGA, TCGA-STAD, Adenomas and Adeno Carcinomas and Asian) and Files (Clinical and Bcr xml), then clinical datasets of

443 samples were obtained from TCGA. Clinical information of gene ID, futime, fustat, age, gender, grade, stage, T, M and N was obtained through Perl.

2.3. Datasets from GEO

The Expression Profiling by Array (Series GSE 54129) was obtained from GEO (https://www.ncbi.nlm.nih.gov/). We contained information on 111 human gastric cancer tissues and 21 non-cancerous gastric tissues, which were collected in Ruijin Hospital, SJTU, China. Through Perl processing of the GEO file, we obtained one probe, one matrix file, and one annotation file.

2.4. Differential Expression Analysis

The differential expression analysis environment was constructed with four packages including limma, edgeR, pheatmap and ggplot2 through R software to output the analysis results. Take logFC filter conditions as logfcfilter = 1; After correction, the P value filtering condition was set as fdrFilter = 0.05. TCGA datasets and GEO datasets were analyzed, respectively.

2.5. Construction of WGCNA and Identification of Important Modules

Data was processed using R 4.0.3 software. To ensure the reliability of the network construction, the abnormal samples were deleted. Pearson correlation coefficient was calculated to assess the similarity of gene expression profiles, and then the correlation coefficient between genes was weighted by a power function to obtain a scale-free network. In terms of co-expression, gene modules are densely interconnected gene clusters. WGCNA uses hierarchical clustering to identify gene modules and colors to indicate modules. Dynamic tree cutting was used to identify different modules. In the module selection process, the adjacency matrix (a measure of topological similarity) was converted into a topological covering matrix (TOM, no graph was output due to the limited function of the computer), and the modules were detected by clustering analysis.

The WGCNA and limma packages of R were used. The minimum gene module size was set to 50 to obtain modules of appropriate size and the threshold for merging similar modules was set to 0.25. TCGA datasets and GEO datasets were analyzed, respectively.

2.6. Venn Diagram

TCGA differential expression analysis results, GEO differential expression analysis results, TCGA turquoise module and GEO black module were made into Venn Diagram by Venn Diagram package in R, and the intersection gene data text of the four kinds of data were output at the same time.

2.7. GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) Enrichment Analysis

Intersection gene data text showed gene by symbol ID. We translated it into en-

trezID with R's org.Hs.eg.db package. Then the enrichment analysis was done by R to library clusterProfiler, org.Hs.eg.db, enrichplot and ggplot2 packages. p value filter condition was set as p value filter = 0.05; the adjusted p-value filter condition was set as q value Filter = 1.

2.8. PPI (Protein Interaction Network) Analysis

On the website https://string-db.org/, we input the intersection genes of Venn, selected species, and then got the network diagram of protein interaction. We highlighted the connection of these genes by setting the minimum required interaction score. It can also set medium confidence (0.400) or hide disconnected nodes in the network. We output proteins interactions file and protein interaction diagram.

2.9. Cytoscape

We installed java, then installed Cytoscape. We imported the files obtained from PPI processing into Cytoscape to obtain hub genes.

2.10. HPA

The obtained ten hub genes were introduced into the http://www.proteinatlas.org/, to obtain immunohistochemical images related to the target genes and diseases.

2.11. Survival and Clinical Analysis

Overall survival of STAD patients was analyzed using the Kaplan Meier plot (http://kmplot.com).

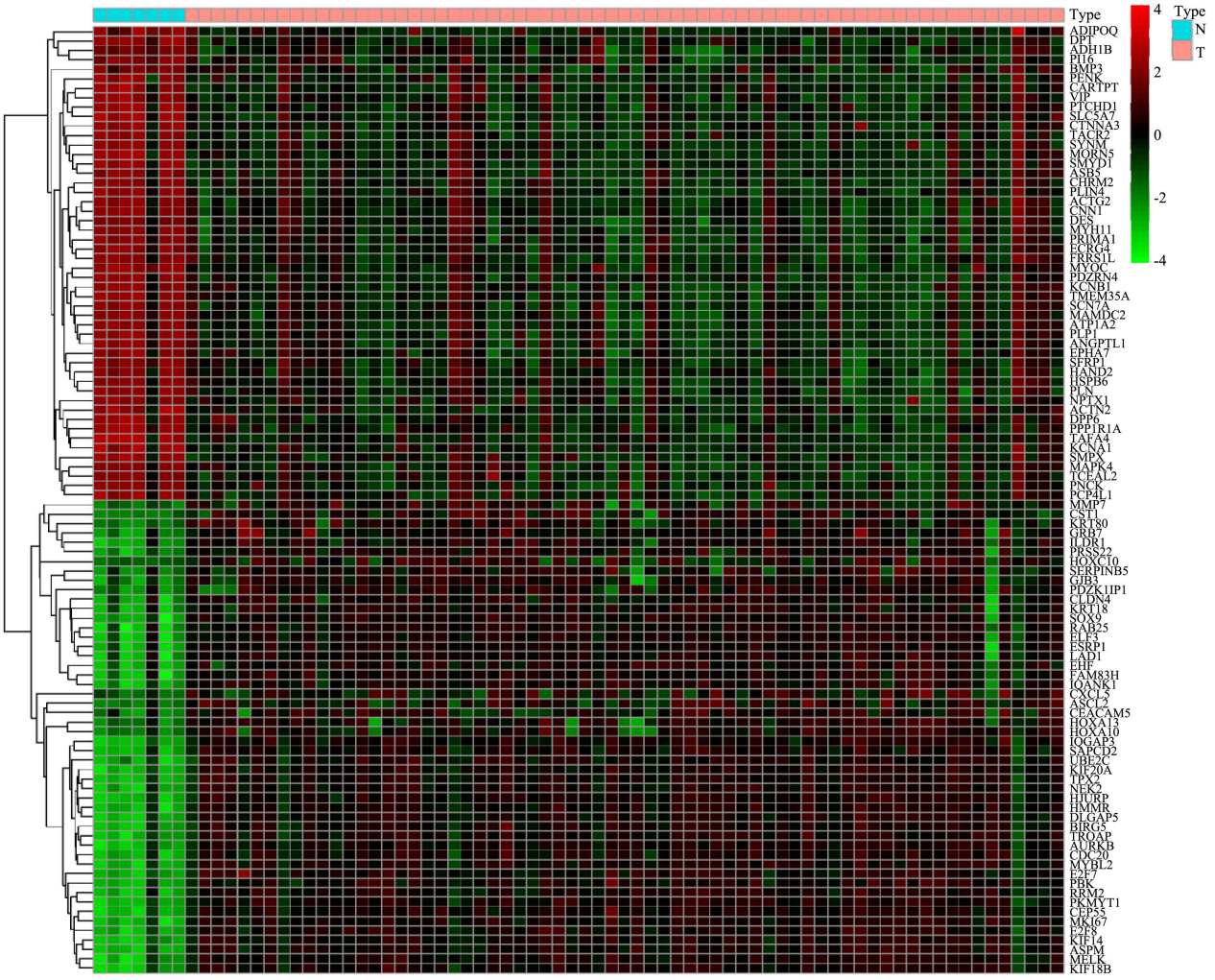
3. Result

3.1. 6652 Differentially Expressed Genes Were Identified between STAD and Normal Samples

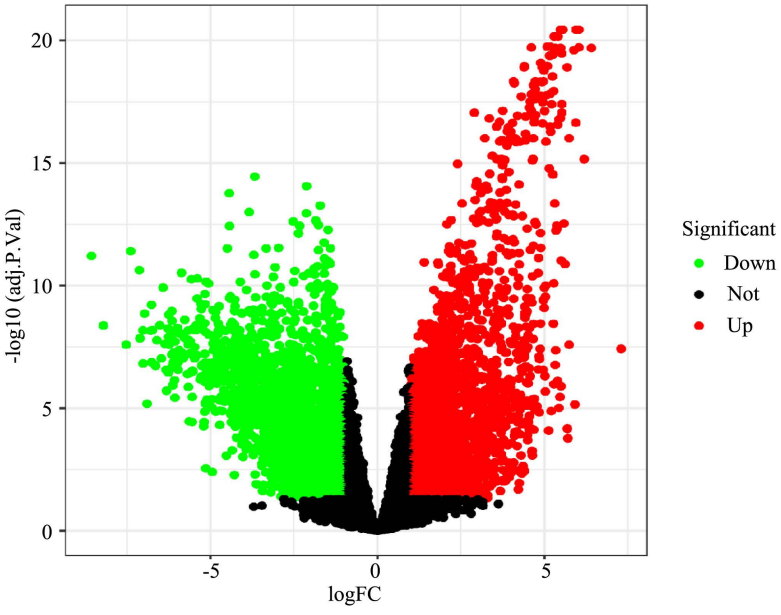
A total of 4859 differentially expressed genes from TCGA and 1793 differentially expressed genes from GEO were identified between STAD and normal samples as shown in **Figure 1(a)** and **Figure 1(b)** (TCGA) and **Figure 1(c)** and **Figure 1(d)** (GEO).

3.2. 13 key Modules Were Identified in STAD by WGCNA Analysis

The first 25% differentially expressed genes in TCGA and GEO data were separately used for cluster analysis through WGCNA package. Hierarchical clustering trees were constructed using the gene expression data with the height threshold limit and screen out outliers. The rest of the data was used to construct and weight the co-expression network. To determine the optimal value of soft threshold (power), the analysis needs were made in a certain range and the scale-free condition. When the power value was set to 1 and 8 (**Figure 2(a)**), the connectivity between genes in the network satisfied the scale-free network distribution. By combining the modules with higher similarity of characteristic


Type
ADIPOQ
DPT
ADH1B
PI16
BMP3
PENK
CARTPT
VIP
PTCHD1
SLC5A7
CTNNA3
TACR2
SYNM
MORN5
SMYD1
ASB5
CHRM2
PLIN4
ACTG2
CNN1
DES
MYH11
PRIMA1
ECRG4
FRRS1L
MYOC
PDZRN4
KCNB1
TMEM35A
SCN7A
MAMDC2
ATP1A2
PLP1
ANGPTL1
EPHA7
SFRP1
HAND2
HSPB6
DPYS
NPTX1
ACTN2
DPP6
PPP1R1A
TAFA4
KCNA1
SMPX
MAPK4
TCEAL2
PNCK
PCP4L1
MMP7
CST1
KRT80
GRB7
IL1R1
PRSS22
HOXC10
SERPINB5
GJB3
PDZK1IP1
CLDN4
KRT18
SOX9
RAB25
ELF3
ESRP1
LAD1
EHF
FAM83H
IQANK1
CXCL5
ASCL2
CEACAM5
HOXA13
HOXA10
IQGAP3
SAPCD2
UBE2C
KIF20A
TPX2
NEK2
HJURP
HMMR
DLGAP5
BIRC5
TROAP
AURKB
CDC20
MYBL2
E2F7
PBK
RRM2
PKMYT1
CEP55
MKI67
E2F8
KIF14
ASPM
MELK
KIF18B
4
2
0
-2
-4
Type
N
T

(a)


20
15
10
5
0
-log10 (adj.P.Val)
-5
0
5
logFC
Significant
Down
Not
Up

(b)

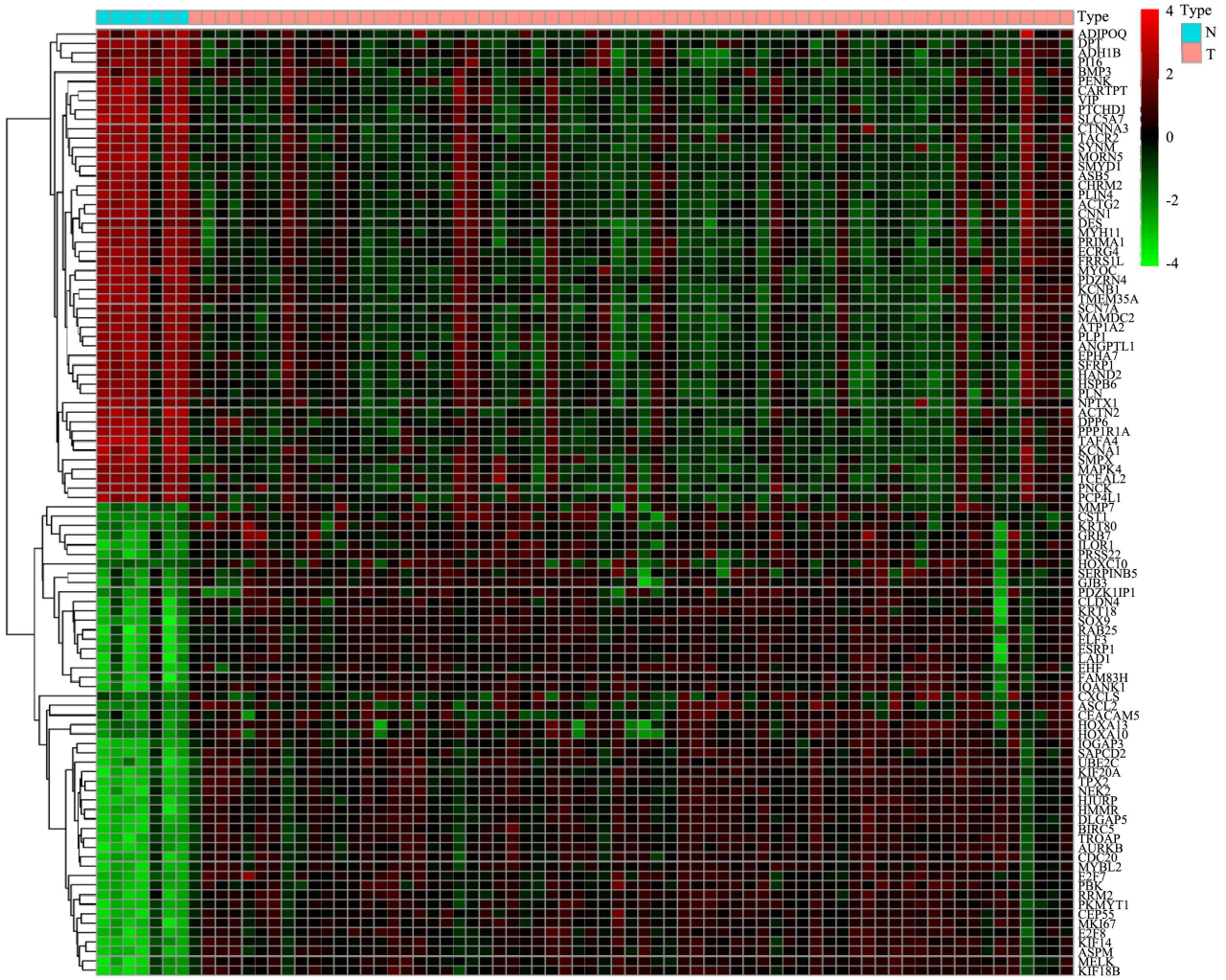
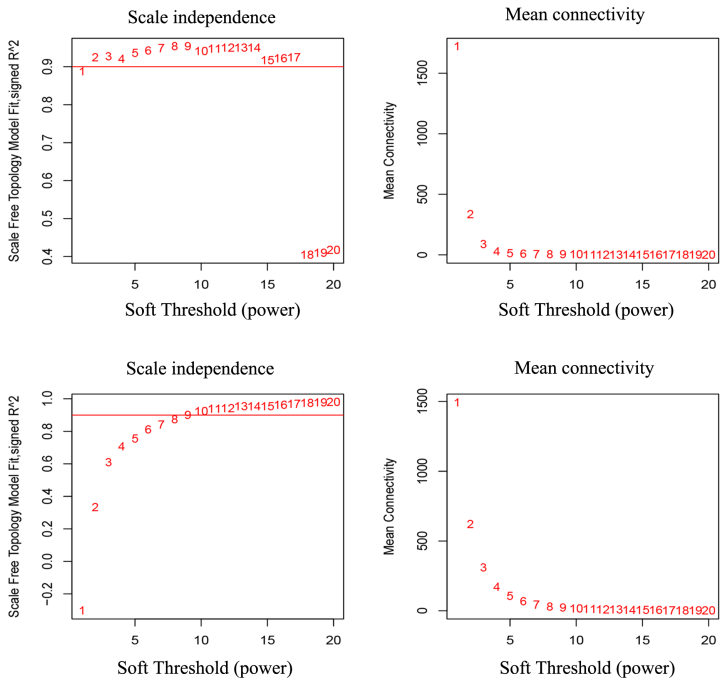
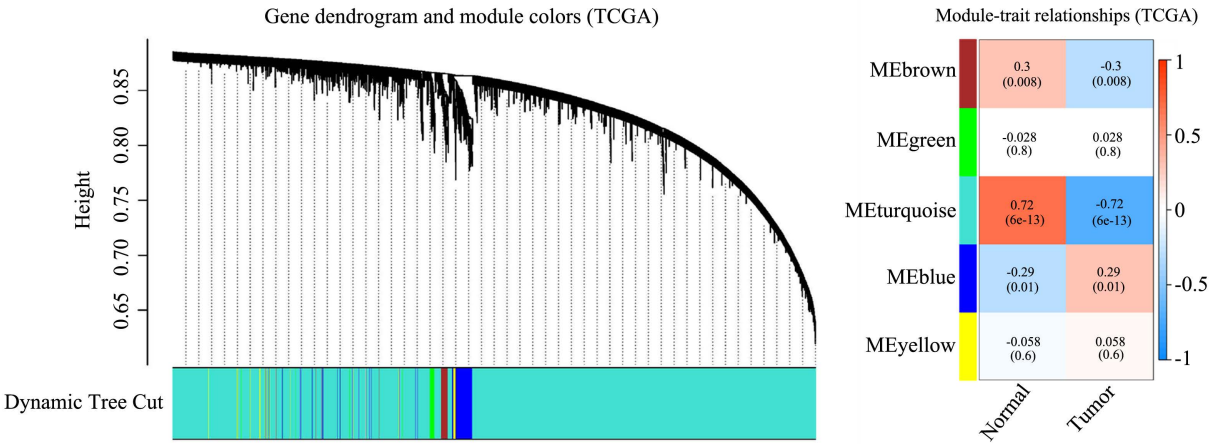


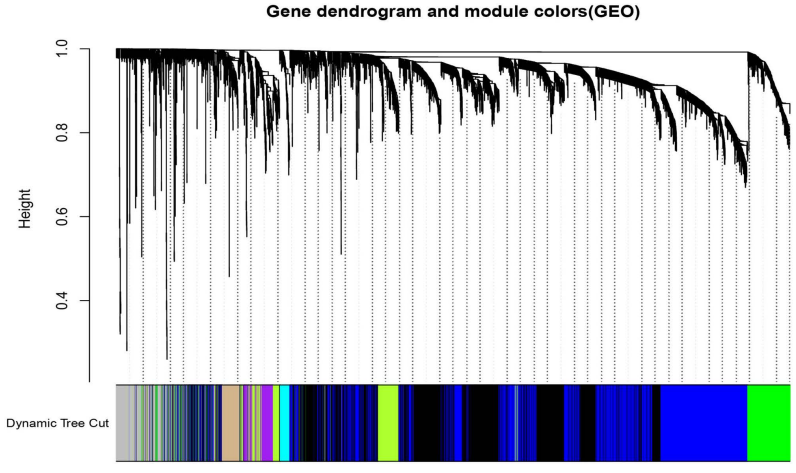
Figure 1. 6652 differentially expressed genes were identified between STAD and normal samples. (a) TCGA-heatmap; (b) TCGA-Volcanic map; (c) GEO-heatmap; (d) GEO-Volcanic map. Red indicates high expression/up-regulation, green indicates low expression/down-regulation, and black indicates no significant difference.


Scale independence
Scale Free Topology Model Fit,signed R^2
Soft Threshold (power)
Mean connectivity
Mean Connectivity
Soft Threshold (power)
Scale independence
Scale Free Topology Model Fit,signed R^2
Soft Threshold (power)
Mean connectivity
Mean Connectivity
Soft Threshold (power)

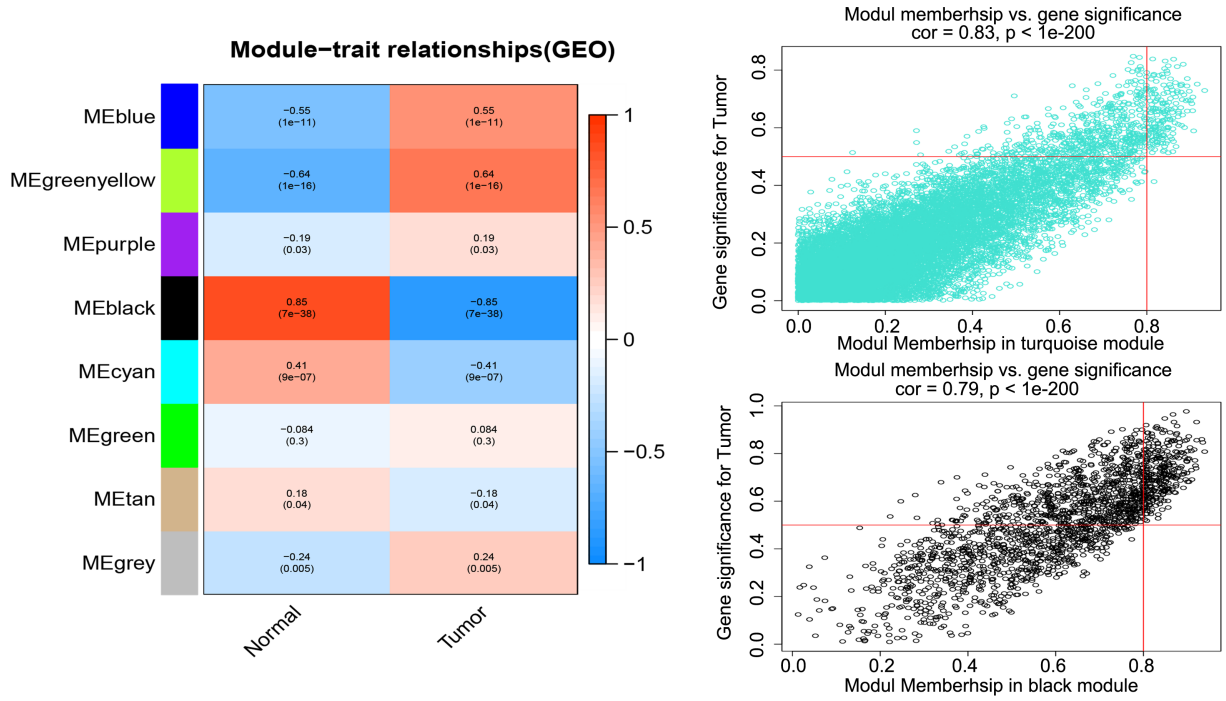
(a)


Gene dendrogram and module colors (TCGA)
Height
Dynamic Tree Cut
Module-trait relationships (TCGA)
MEbrown
MEgreen
MEturquoise
MEblue
MEyellow
0.3 (0.008)
-0.3 (0.008)
-0.028 (0.8)
0.028 (0.8)
0.72 (6e-13)
-0.72 (6e-13)
-0.29 (0.01)
0.29 (0.01)
-0.058 (0.6)
0.058 (0.6)
Normal
Tumor

(b)


Gene dendrogram and module colors(GEO)
Height
Dynamic Tree Cut

(c)



(d)

Figure 2. 13 key modules were identified by WGCNA analysis. (a) Soft Threshold (TCGA). The upper two pictures show the determination of soft threshold based on TCGA differential gene expression data and the power value of 1. The lower two pictures show the determination of soft threshold based on GEO differential gene expression data and the power value of 8; (b) WGCNA analysis of TCGA data. The tree graph of matrix aggregation (left) was based on the gene cluster graph obtained by hierarchical clustering of adjacency-based disambiguation. The color block below the tree graph represents the modules identified by dynamic tree cutting method. In the heatmap of module features (right), the upper number in the color block represents the correlation with STAD, and the following is the P value. Red is positively correlated, and turquoise is negatively correlated; (c) WGCNA analysis of GEO data. The tree graph of matrix aggregation (left) was based on the gene cluster graph obtained by hierarchical clustering of adjacency-based disambiguation. The color block below the tree graph represents the module identified by dynamic tree cutting method. In the heatmap of module features (right), the upper number in the color block represents the correlation with STAD, and the following is the P value. Red is positively correlated, and turquoise is negatively correlated; (d) Module membership vs gene significant. The expression of the genes in turquoise module and black module is more related to tumors.

memes (MEME, a WGCNA term for a module with the same characteristics) using the dynamic mixing shearing method, 5 and 8 MEME modules were finally obtained for TCGA (Figure 2(b)) and GEO (Figure 2(c)) data sets, respectively. We found that the expression of the genes in turquoise modules of TCGA samples and in black modules of GEO samples have the greatest correlation with the tumor tissues (Figure 2(d)).

3.3. 293 Potential STAD Associated Genes Were Identified from Intersection by Venn Diagram

To reduce the false positive rate of the results, the Venn diagram of the above four data sets (TCGA turquoise module, GEO black module, TCGA differentially expressed genes and GEO differentially expressed genes) was used for intersection. There are 293 genes that are intersected in all 4 groups, and pictures and text files were output (Figure 3).

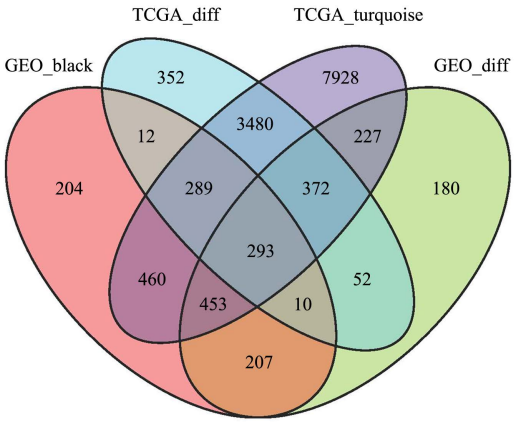


Figure 3. 293 STAD associated genes were identified from intersection by Venn Diagram.

3.4. The 293 Intersected Genes Were Enriched in Cell Cortex and Infection by GO and KEGG Analysis

GO and KEGG analyses were carried out to explore their biological function of the 293 intersected genes identified by Venn Diagram. GO analysis showed that these genes are mainly involved in the processes of cell cortex, secretory granule lumen, carbohydrate binding and actin binding. KEGG analysis showed that these genes were highly correlated with Hepatitis C, pathogenic Escherichia coli infection, amino sugar and nucleotide sugar metabolism, complement and coastal cascades, and Histidine metabolism (**Table 1** and **Figure 4**).

3.5. 10 Hub Genes Were Identified from PPI and Cytoscape Analyses of the Intersected Genes

PPI analysis of the 293 intersected genes identified by Venn Diagram sorted out 354 pairwise links and deleted the genes that were not connected with the subject network (**Figure 5**). Then we imported the interactive file obtained from the PPI analysis into Cytoscape to construct a gene network and screened out 10 hub genes (KLF4, CGN, SHH, LIF, GATA6, FOXA2, OCLN, FOXA1, CLDN1 and NQO1) according to their degree of associations (**Table 2** and **Table 3**, **Figure 5**). Among the 10 hub genes, KLF4 and CGN expression was decreased while other 8 gene expression was increased in STAD tumors compared to normal tissues (**Table 3**). When we imported into the HPA official website, immunohistochemical images of malignant gastric cancer tissues verified the expression of these hub genes.

3.6. KLF4/CGN Low and SHH/LIF High Expression Were Associated with Short Overall Survival of Asian STAD Patients

Interestingly, using the Kaplan Meier plot, we found that SHH/LIF high expression among tumors was associated with short overall survival of Asian STAD patients, while low expression of KLF4/CGN and the other 6 hub genes among tumors was associated with short overall survival of Asian STAD patients (**Table 4** and **Figure 6**). The association of the gene expression among tumors and

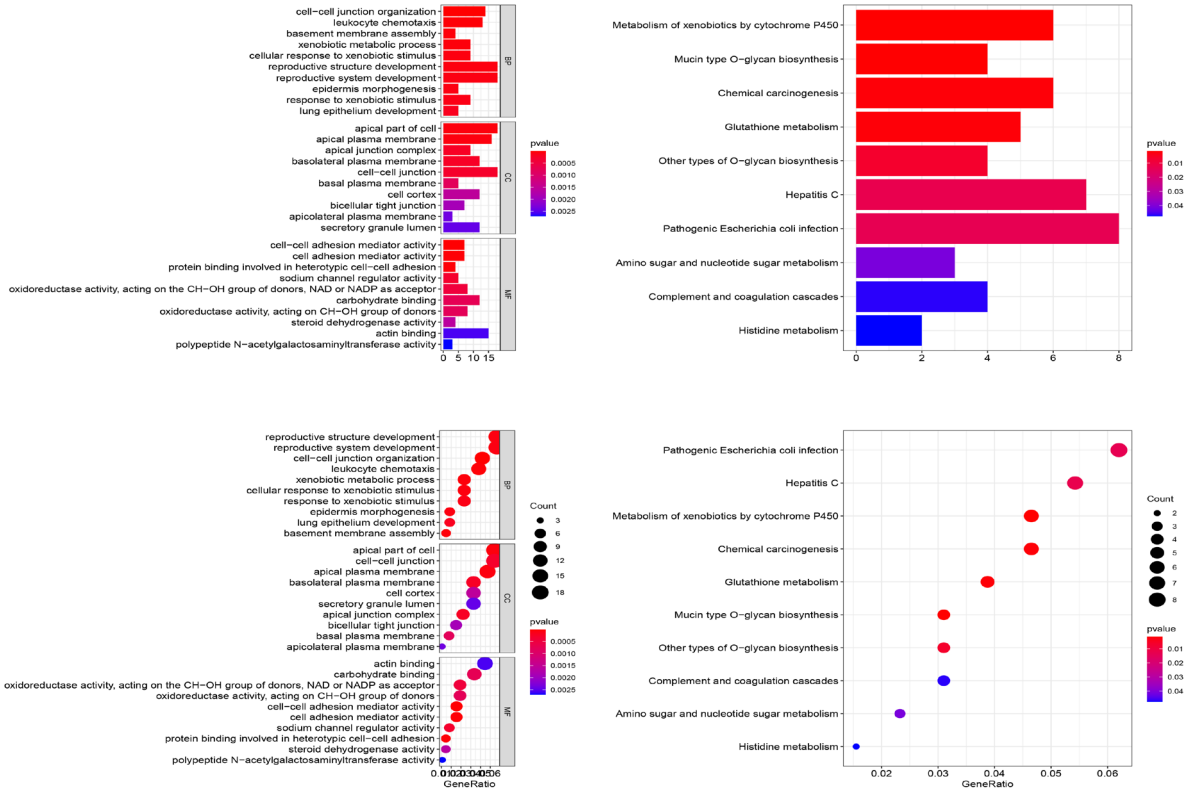


Figure 4. The 293 intersected genes were enriched in cell cortex and infection by GO (left) and KEGG (right) analysis.

Table 1. The most significant biological function of the 293 STAD associated Venn intersection gene from KEGG.

ID	Description	Gene Ratio	Bg Ratio	P value	P. adjust	q value	Gene ID	Count
hsa00980	Metabolism of xenobiotics by cytochrome P450	6/129	78/8075	0.001484	0.103158	0.10313	AKR1C1/AKR7A3/CBR1/CYP3A5/EPHX1/GSTM2	6
hsa00512	Mucin type O-glycan biosynthesis	4/129	32/8075	0.001581	0.103158	0.10313	GALNT5/GALNT6/GALNT7/GCNT1	4
hsa05204	Chemical carcinogenesis	6/129	83/8075	0.002044	0.103158	0.10313	CBR1/CYP2C18/CYP3A5/EPHX1/GSTM2/NAT1	6
hsa00480	Glutathione metabolism	5/129	57/8075	0.002074	0.103158	0.10313	CHAC2/G6PD/GPX2/GSTM2/IDH1	5
hsa00514	Other types of O-glycan biosynthesis	4/129	47/8075	0.006522	0.259581	0.259512	GALNT5/GALNT6/GALNT7/PLOD3	4
hsa05160	Hepatitis C	7/129	157/8075	0.012718	0.381891	0.38179	CD81/CLDN1/CLDN23/IRF3/OAS1/OCLN/TLR3	7
hsa05130	Pathogenic Escherichia coli infection	8/129	197/8075	0.013433	0.381891	0.38179	BAIAP2L1/CLDN1/CLDN23/IL18/MYO5B/MYO5C/OCLN/TUBB2A	8
hsa00520	Amino sugar and nucleotide sugar metabolism	3/129	48/8075	0.040911	0.926281	0.926036	CYB5R3/GFPT1/GMDS	3
hsa04610	Complement and coagulation cascades	4/129	85/8075	0.046745	0.926281	0.926036	BDKRB1/F5/PLAU/SERPINA1	4
hsa00340	Histidine metabolism	2/129	22/8075	0.047501	0.926281	0.926036	ALDH2/MAOA	2

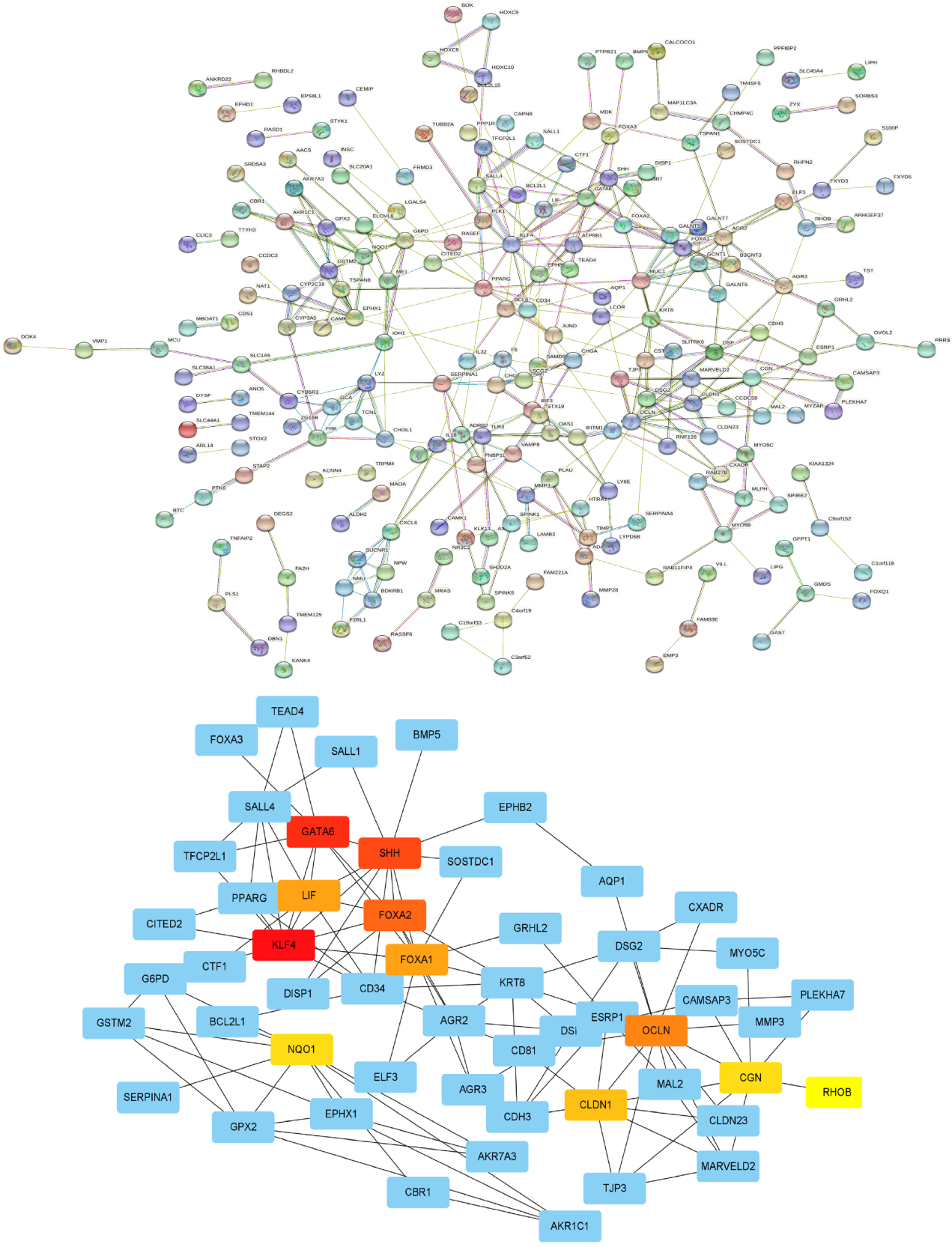


Figure 5. 10 hub genes were identified from PPI and Cytoscape analyses of the intersected genes.

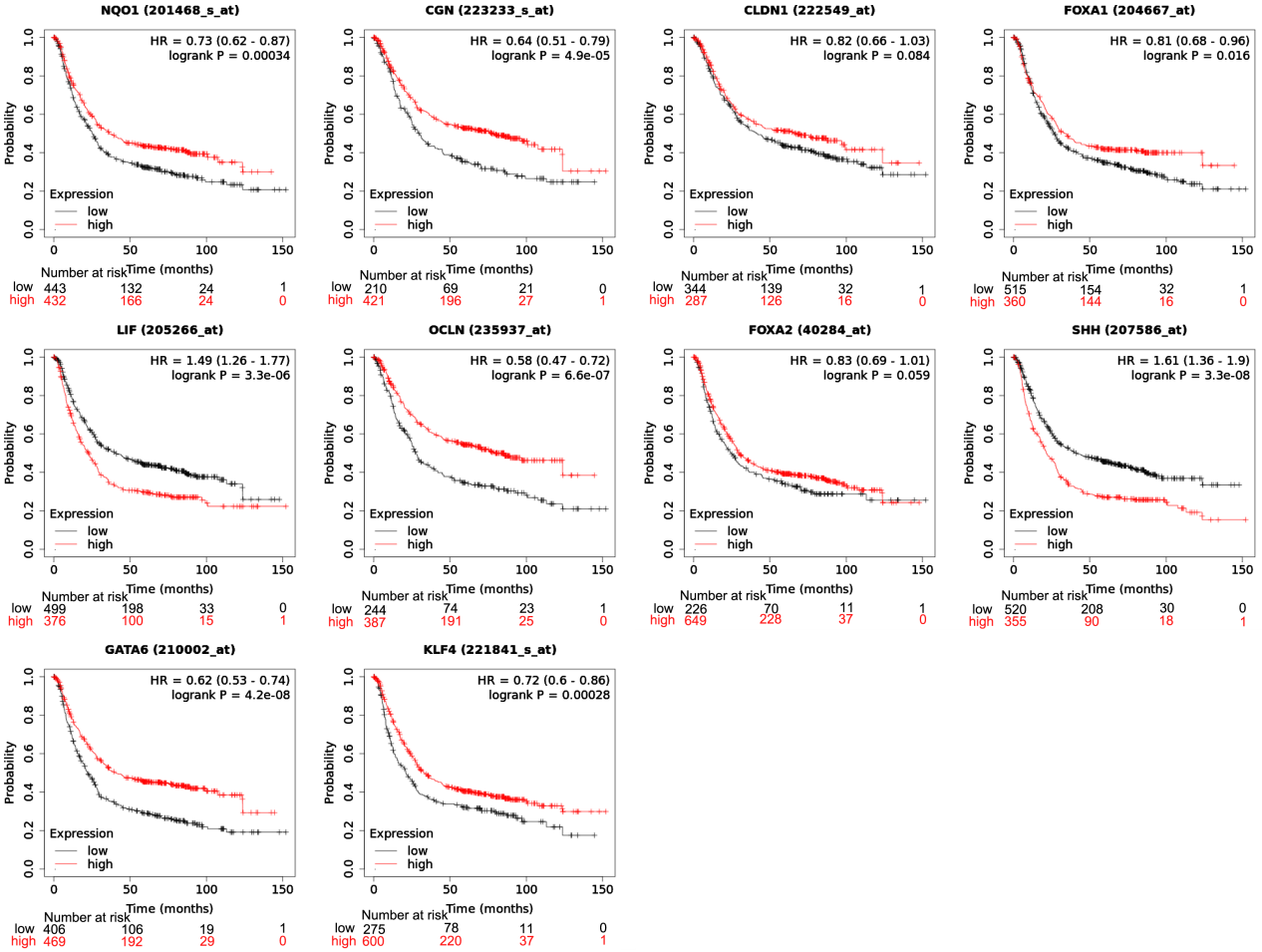


Figure 6. KLF4/CGN low and SHH/LIF high expression were associated with short overall survival of Asian STAD patients. The 10 genes shown in the figure are negatively (LIF/SHH) or positively (the other 8 genes) correlated with the survival time of STAD patients.

Table 2. The 10 hub genes filtered based on MCC information.

Node name	MCC	DMNC	MNC	Degree	EPC	Bottle Neck	Ec Centricity	Closeness	Radiality	Betweenness	Stress	Clustering Coefficient
KLF4	73	0.35915	10	11	58.277	13	0.11257	63.44286	6.21607	2416.518	8984	0.32727
GATA6	63	0.4082	8	9	54.523	3	0.0985	53.76071	5.72936	597.1098	2556	0.38889
SHH	61	0.37904	8	11	55.244	6	0.11257	59.64286	6.03529	1517.783	4498	0.23636
FOXA2	56	0.37904	8	8	56.063	1	0.0985	56.77738	5.91477	539.4949	2704	0.46429
OCLN	40	0.32929	7	13	55.116	14	0.11257	61.89286	6.08628	3128.249	9986	0.16667
LIF	37	0.47549	6	7	52.922	2	0.11257	56.45952	5.9704	639.8622	2770	0.47619
FOXA1	37	0.2864	9	10	55.625	10	0.0985	58.49405	5.95186	1700.16	6308	0.26667
CLDN1	30	0.38039	6	8	49.53	2	0.11257	54.16905	5.79889	726.2893	2198	0.32143
CGN	28	0.56839	4	8	46.468	4	0.0985	51.07262	5.53931	1034.018	2754	0.25
NQO1	28	0.36588	7	9	43.906	5	0.13134	55.8	5.83134	2021.798	5364	0.27778

Note: maximal clique centrality (MCC); density of maximum neighborhood component (DMNC); maximum neighborhood component (MNC); edge percolated component (EPC).

Table 3. Differential expression of the 10 hub genes in STAD tumors verse normal tissues (GEOdiff).

ID	logFC	Ave Expr	t	P. Value	adj. P. Val	B
KLF4	−1.47302	6.06431	−3.1081	0.00265	0.007416	−2.49024
GATA6	2.097037	5.957883	4.016353	0.000138	0.000587	0.289208
SHH	3.041936	2.593291	3.930826	0.000185	0.00076	0.00668
FOXA2	4.817927	4.158278	6.301899	1.77E−08	2.65E−07	8.97395
OCLN	3.418704	4.562556	6.173057	3.04E−08	4.22E−07	8.441581
LIF	2.095345	3.932531	5.407506	7.15E−07	6.35E−06	5.361607
FOXA1	4.845426	4.857299	6.639157	4.19E−09	7.78E−08	10.38263
CLDN1	3.484039	5.44208	4.770593	8.71E−06	5.46E−05	2.938582
CGNL1	−2.66988	2.885747	−4.12225	9.5E−05	0.000426	0.644452
NQO1	1.778712	7.197352	2.88561	0.005084	0.012958	−3.08934

LogFC means log (expression in cancer/expression in normal tissue). The higher the value, the higher the expression levels in cancer tissues are. Minus of LogFC means the expression is lower in cancer tissues compared to that in normal tissues from GEO datasets series GSE 54129.

Table 4. The survival times of the 10 hub genes.

Gene	Low expression cohort (months)	High expression cohort (months)	P value
KLF4	21.2	33.27	0.00028
GATA6	22.2	40.7	4.2e−08
SHH	41.2	21.6	3.3e−08
FOXA2	24.9	29.8	0.059
OCLN	28.03	85.6	6.6e−07
LIF	40.2	22.87	3.3e−06
FOXA1	26.7	33.27	0.016
CLDN1	41.2	67.1	0.084
CGN	30	77.2	4.9e−05
NQO1	25.2	36.4	0.00034

overall survival are consistent with the KLF4/CGN low expression or SHH/LIF high expression in the STAD tumors compared to normal tissues (Table 3). Our findings support that KLF4/CGN is potential tumor suppressor and SHH/LIF is potential oncogene for STAD of Asian patients.

4. Discussion and Conclusion

STAD is a kind of high incidence and mortality tumor in adults, especially in Asian, and there is an urgent need to identify potential novel oncogenes, tumor suppressor genes, and biomarkers of diagnosis, therapy and prognosis, specifically for Asian STAD. In this study, the gene expression data of the STAD transcriptome were analyzed, and a total of 4859 differentially expressed genes from

TCGA and 1793 genes from GEO were identified in STAD, followed by identification of the five modules of TCGA and eight modules of GEO. These differentially expressed genes and the tumor's associated module genes were subjected to Venn intersection, and 293 intersected genes were obtained. These intersected genes are enriched in a few biological processes and biological functions, such as chemotaxis, inflammatory reactions, angiogenesis, cell cycle, etc., which may be related to the occurrence and development of the cancer. Ten hub genes were screened out from these 293 intersected genes for association analysis with survival in patients with STAD. Our results showed that KLF4/CGN low and SHH/LIF high expression were associated with short overall survival of Asian STAD patients. Our bioinformatics analysis revealed potential novel tumor suppressors (KLF4/CGN) and oncogenes (SHH/LIF) and biomarkers for diagnosis, therapy and prognosis of STAD, specifically for Asian patients [8] [9] [10].

In our study, the system biology-based methods, including WGCNA, were used to identify the 10 network hub genes related to STAD, namely KLF4, CGN, LIF, SHH, GATA6, FOXA2, OCLN, FOXA1, CLDN1 and NQO1. While KLF4/CGN low and SHH/LIF high expression were associated with short overall survival of Asian STAD patients, the other 6 hub genes (GATA6, FOXA2, OCLN, FOXA1, CLDN1 and NQO1) were all expressed higher in STAD tumors compared to the normal tissues, but higher expression in tumors showed longer overall survival of STAD patients. The inconsistence may suggest that the 6 hub genes may be passenger genes or expressed as an active compensation to suppress tumor progression. It has been shown that KLF4, CGN, GATA6, FOXA2, OCLN, FOXA1, CLDN1 and NQO1 can inhibit the occurrence and development of tumors [11]-[17], while SHH and LIF promote tumor occurrence and development [18] [19]. LIF promotes tumorigenesis and metastasis of breast cancer, Shh Bladder promotes cancer stemness and tumorigenesis. At present, experimental studies on CGN are not found. KLF4 can inhibit proliferation and invasion of breast cancer [20], reduce the impact of chemotherapy on colon cancer cells [21]. As an important signal transduction pathway for the occurrence and development of cancer, SHH has become an important target gene for the treatment of medulloblastoma and pancreatic cancer [22] [23]. LIF can be induced by HIF-2α and promotes tumor progression, metastasis, and chemical resistance in a variety of solid tumors [24] [25]. The potential novel tumor suppressors and oncogenes and biomarkers identified here need to be further validated.

Some limitations of our study should be mentioned. First, this was a retrospective design study, not a prospective cohort study. In addition, a large sample size was required to verify our findings. Thirdly, these results may be validated experimentally in future.

Funding

This work was supported by National Natural Science Foundation of China (81872412).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Schumacher, S.E., *et al.* (2017) Somatic Copy Number Alterations in Gastric Adenocarcinomas among Asian and Western Patients. *PLoS ONE*, **12**, e0176045. https://doi.org/10.1371/journal.pone.0176045

[2] Zhou, L., *et al.* (2020) Exploring TCGA Database for Identification of Potential Prognostic Genes in Stomach Adenocarcinoma. *Cancer Cell International*, **20**, Article No. 264. https://doi.org/10.1186/s12935-020-01351-3

[3] Niemira, M., *et al.* (2019) Molecular Signature of Subtypes of Non-Small-Cell Lung Cancer by Large-Scale Transcriptional Profiling: Identification of Key Modules and Genes by Weighted Gene Co-Expression Network Analysis (WGCNA). *Cancers* (*Basel*), **12**, 37. https://doi.org/10.3390/cancers12010037

[4] Li, N., *et al.* (2020) Integrative Analysis of Differential lncRNA/mRNA Expression Profiling in *Helicobacter pylori* Infection-Associated Gastric Carcinogenesis. *Frontiers in Microbiology*, **11**, 880.

[5] Xiang, Z., *et al.* (2019) A Positive Feedback between IDO1 Metabolite and COL12A1 via MAPK Pathway to Promote Gastric Cancer Metastasis. *Journal of Experimental & Clinical Cancer Research*, **38**, Article No. 314. https://doi.org/10.1186/s13046-019-1318-5

[6] Zhang, Y., *et al.* (2018) LASSO-Based Cox-PH Model Identifies an 11-lncRNA Signature for Prognosis Prediction in Gastric Cancer. *Molecular Medicine Reports*, **18**, 5579-5593. https://doi.org/10.3892/mmr.2018.9567

[7] Zhao, X., *et al.* (2016) Discovery of Signature Genes in Gastric Cancer Associated with Prognosis. *Neoplasma*, **63**, 239-245. https://doi.org/10.4149/209_150531N303

[8] Krstic, M., *et al.* (2013) KLF4 Expression and Apoptosis-Related Markers in Gastric Cancer. *Journal BUON*, **18**, 695-702.

[9] Samadani, A.A., *et al.* (2019) CDX1/2 and KLF5 Expression and Epigenetic Modulation of Sonic Hedgehog Signaling in Gastric Adenocarcinoma. *Pathology & Oncology Research*, **25**, 1215-1222. https://doi.org/10.1007/s12253-019-00594-4

[10] Sulahian, R., *et al.* (2014) An Integrative Analysis Reveals Functional Targets of GATA6 Transcriptional Regulation in Gastric Cancer. *Oncogene*, **33**, 5637-5648. https://doi.org/10.1038/onc.2013.517

[11] Adams, E.J., *et al.* (2019) FOXA1 Mutations Alter Pioneering Activity, Differentiation and Prostate Cancer Phenotypes. *Nature*, **571**, 408-412. https://doi.org/10.1038/s41586-019-1318-9

[12] Conceição, A.L., *et al.* (2017) Downregulation of OCLN and GAS1 in Clear Cell Renal Cell Carcinoma. *Oncology Reports*, **37**, 1487-1496. https://doi.org/10.3892/or.2017.5414

[13] Liu, J., *et al.* (2018) Coordination of FOXA2 and SIRT6 Suppresses the Hepatocellular Carcinoma Progression through ZEB2 Inhibition. *Cancer Management and Research*, **10**, 391-402. https://doi.org/10.2147/CMAR.S150552

[14] Mahati, S., *et al.* (2017) miR-29a Suppresses Growth and Migration of Hepatocellular Carcinoma by Regulating CLDN1. *Biochemical and Biophysical Research Communications*, **486**, 732-737. https://doi.org/10.1016/j.bbrc.2017.03.110

[15] Martinelli, P., *et al.* (2017) GATA6 Regulates EMT and Tumour Dissemination, and Is a Marker of Response to Adjuvant Chemotherapy in Pancreatic Cancer. *Gut*, **66**, 1665-1676. https://doi.org/10.1136/gutjnl-2015-311256

[16] Motea, E.A., *et al.* (2019) NQO1-Dependent, Tumor-Selective Radiosensitization of Non-Small Cell Lung Cancers. *Clinical Cancer Research*, **25**, 2601-2609. https://doi.org/10.1158/1078-0432.CCR-18-2560

[17] Wang, X., *et al.* (2020) The Deubiquitinase USP10 Regulates KLF4 Stability and Suppresses Lung Tumorigenesis. *Cell Death & Differentiation*, **27**, 1747-1764. https://doi.org/10.1038/s41418-019-0458-7

[18] Li, X., *et al.* (2014) LIF Promotes Tumorigenesis and Metastasis of Breast Cancer through the AKT-mTOR Pathway. *Oncotarget*, **5**, 788-801. https://doi.org/10.18632/oncotarget.1772

[19] Syed, I.S., Pedram, A. and Farhat, W.A. (2016) Role of Sonic Hedgehog (Shh) Signaling in Bladder Cancer Stemness and Tumorigenesis. *Current Urology Reports*, **17**, 11. https://doi.org/10.1007/s11934-015-0568-9

[20] Nagata, T., *et al.* (2017) KLF4 and NANOG Are Prognostic Biomarkers for Triple-Negative Breast Cancer. *Breast Cancer*, **24**, 326-335. https://doi.org/10.1007/s12282-016-0708-1

[21] Yadav, S.S., *et al.* (2019) KLF4 Sensitizes the Colon Cancer Cell HCT-15 to Cisplatin by Altering the Expression of HMGB1 and hTERT. *Life Sciences*, **220**, 169-176. https://doi.org/10.1016/j.lfs.2019.02.005

[22] Catenacci, D.V., *et al.* (2015) Randomized Phase Ib/II Study of Gemcitabine plus Placebo or Vismodegib, a Hedgehog Pathway Inhibitor, in Patients with Metastatic Pancreatic Cancer. *Journal of Clinical Oncology*, **33**, 4284-4292. https://doi.org/10.1200/JCO.2015.62.8719

[23] Samkari, A., White, J. and Packer, R. (2015) SHH Inhibitors for the Treatment of Medulloblastoma. *Expert Review of Neurotherapeutics*, **15**, 763-770. https://doi.org/10.1586/14737175.2015.1052796

[24] Wu, L., *et al.* (2015) HIF-2α Mediates Hypoxia-Induced LIF Expression in Human Colorectal Cancer Cells. *Oncotarget*, **6**, 4406-4417. https://doi.org/10.18632/oncotarget.3017

[25] Zhang, L., *et al.* (2016) Expression of HIF-2α and VEGF in Cervical Squamous Cell Carcinoma and Its Clinical Significance. *BioMed Research International*, **2016**, Article ID: 5631935. https://doi.org/10.1155/2016/5631935