

Weighted gene co-expression network analysis identifies potential candidate biomarkers for adenocarcinoma of the esophagogastric junction

Zhiyong Lai

Shanxi Bethune hospital

Wenhui Yang

Shanxi Bethune hospital

Weibin Li

Shanxi Bethune hospital

Tiantian Zhang

Shanxi Bethune hospital

Kai Jia

Shanxi Bethune hospital

Kewei Jiang

Peking University People's Hospital

Jun Xu (✉ junxutyg@163.com)

Shanxi Dayi Hospital, Shanxi Academy of Medical Sciences <https://orcid.org/0000-0002-9721-7569>

Primary research

Keywords: Adenocarcinoma of the esophagogastric junction, Weighted gene co-expression network analysis, Modules, Biomarker

Posted Date: February 24th, 2020

DOI: <https://doi.org/10.21203/rs.2.24360/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Gastric cancer (GC) is the fifth most common kind of malignant tumor, and commonly leads to death. As a subtype of gastric cancer, adenocarcinoma of the esophagogastric junction (AEG), accounts for about 50% of all gastric cancer cases. So far, the systematic co-expression analysis of this tumor has not fully explained its pathogenesis. The purpose of this study was to construct RNAs co-expression networks to predict candidate hub genes associated with the tumorigenesis of AEG.

Methods: The RNA-seq data of 22 AEG patients was processed with weighted gene co-expression network analysis strategy. Differentiate the modules with clinical tumor markers and preservation, and carry out gene ontology and pathway enrichment analysis. We identified the co-expression modules and used GO and KEGG terms to investigate the functional enrichment of co-expression genes, suggesting that blue and brown modules are related to the biological processes of tumorigenesis.

Results: Twenty-five distinct co-expression gene modules were identified, and as the top hub genes of tumorigenic gene modules, CD93, TRIM28, SLC3A2, CBX4, PATL1, and ZNF473 with high intramodular connectivity were assumed as intramodular hub genes in AEG.

Conclusion: The weighted gene co-expression network analysis conducted in this study screened out CD93, TRIM28, SLC3A2, CBX4, PATL1, and ZNF473 may act as candidate biomarker in GC and AEG.

1. Background

Gastric cancer (GC) is a common malignant tumor of the digestive tract with the fifth-highest incidence and the third leading cause of death worldwide[1]. According to the data from the National Central Cancer Registry, there were an estimated 697,100 new GC cases and 498,000 GC deaths in China in 2015 alone[2]. Although the overall incidence of GC worldwide has decreased in recent decades, as a subtype of GC, the incidence of adenocarcinoma of the esophagogastric junction (AEG) has increased significantly in both Western countries and East Asian countries[3-8]. In China, especially in areas where there is a high incidence of traditional esophageal cancer, the number of new AEG cases accounts for about 50% of all GC cases, thus AEG is a serious threat to the health of Chinese people[6].

AEG can be classified according to Siewert's classification into Type I, Type II, and Type III. Regardless of the anatomical location of the tumor, the three types AEG exhibit similarities in epidemiology, morphology and other aspects, but there are significant differences in the pathogenesis [4, 9, 10]. Overall, it is not clear whether this classification reflects the process of pathogenesis. Next-generation sequencing (NGS) was adopted in the Cancer Genome Atlas and other projects to analyze genomic and RNA profile changes of different cancer types, including GC[11-15]. The ultimate goal of the cancer genome mapping and tumor RNA profiling are to make precise medicine possible by adjusting treatments for each patient[16].

Although there have been sequencing studies of AEG, there has been no comprehensive study of AEG from Chinese populations. Genomic sequencing and transcriptome RNA sequencing have the potential to

achieve precision medicine for cancer patients. So far, the large-scale of genomic and RNA sequencing of patients has been limited to Western populations. In order to recognize and understand the potential race and geographic differences and to investigate the widespread application of RNA-seq to Chinese population, RNA sequencing technology was used to characterize clinically actionable driver events in 22 patients with AEG in Shanxi Province of China.

Past studies aimed at identifying diagnostic biomarkers have done so by comparing the expression of particular genes between cancerous and non-cancerous samples. Newly developed systems biology strategies offer new means of analyzing datasets to identify putative biomarkers. For examples, the weighted gene co-expression network analysis (WGCNA) approach is a strategy which aims to account for the fact that differential expression is an oversimplification of the complexities of changes in gene expression patterns. WGCNA instead assessed correlations between specific genes across a range of samples, generating similarity-based gene networks that allow for a global assessment of patterns of gene expression. Those genes which are very highly correlated in terms of their expression across datasets within the resultant network can then be arranged into functional module, allowing for a systems-level analysis of changes in transcriptomic dynamics in a complex dataset. Using these generated modules and gene correlation networks, it is possible to determine which genes are related to one another and how these genes related to specific traits of interest in a given sample[17, 18]. WGCNA also allows researchers to identify certain central genes, known as hub genes, which are the most heavily connected genes within a given correlational network. This strategy can be used to compare expression networks between, for example, health and disease samples, thereby identifying common modules within each set of samples and allowing for a comparison of differences in modules between the two resultant networks. Those modules and the associated hub genes that are uniquely associated with specific disease states may be linked to underlying disease pathology, thus identifying them as potential clinical targets or diagnostic and prognostic biomarkers of disease[17-20].

The WGCNA approach has not been applied to AEG RNA-seq-based gene expression data. Therefore, in this study, our goal is to build co-expression modules from AEG's RNA-seq data. Using this approach, we were able to identify co-expressed gene modules that are specific to clinical tumor markers. The gene ontology, KEGG pathway enrichment and co-expression network analysis were carried out for the modules of interest, and the hub genes were recognized, which helped to identify the potential biological functions of candidate hub genes. These findings provide a set of genes that may serve as indicators of malignant potential and prognosis of AEG.

2. Materials And Methods

2.1 RNA-seq data of AEG samples

Twenty-two AEG tumor tissue samples and paired normal gastric mucosa samples were obtained from the Department of General Surgery, Shanxi Cancer Hospital (Table S1). All of the tissue donors included in this study provided their informed consent. The study was approved by the Human Ethics Committee

of the Shanxi Medical University (China). Methods were carried out in “accordance” with the approved guidelines. The samples were immediately quick-frozen in liquid nitrogen after surgery resection and were stored in refrigerator at -80°C until RNA extraction. The integrity of RNA was assessed by the analysis of rRNA band integrity using 1.0% formaldehyde denaturing gel electrophoresis. Only RNA samples with $1.8 < A_{260}/A_{280} < 2.0$ and $28S/18S \text{ rRNA} > 2$ were used for subsequent experiment. The diagnosis of AEG was confirmed pathologically. First, we characterized expression levels of RNA transcripts using RNA-seq analysis from 22 paired normal and cancerous AEG tissues. Each sample was sequenced (200×) with Illumina HiSeq platform (Illumina, San Diego, CA). RNA-seq was conducted in the laboratory of BGI gene Technology Corporation in Shenzhen, China.

2.2 Weighted gene co-expression network analysis of AEG RNA-seq data

The RNAs co-expression networks were constructed by using WGCNA program kit in Computer R language package [17]. First, the square Euclidean distance of each sample was calculated by the function adjacency method, and the connectivity of the total sample network calculated by the distance was standardized by the function scale [21]. Cluster analysis was carried out with FLASH CLUST software package. We chose $\beta=11$ as the soft-thresholding power to build a scale-free network, resulting in a scale-free topology index of 0.8. The NovelBio Cloud Analysis Platform was used, with support from Novel Bioinformatics Ltd., Co.

2.3 Identifying gene modules associated with tumor markers

The relationship between gene module characteristics and serum tumor marker measurements was used to estimate the relationship between modules and tumor markers, so as to identify different modules that were highly correlated with specific phenotypes. The gene significance (GS) was determined by the absolute value of the correlation between the RNA expression profile and each trait, and the module membership (MM) was valued by the correlation between the RNA expression profile and the eigengene of each module [22]. The correlations between module eigengenes (MEs) and clinical tumor markers including CEA, CA199, CA242, AFP, CA724, CA50, SCC, TPA, TPS, and VEGF were appraised by Pearson's correlation tests.

2.4 KEGG pathway and Gene ontology analysis

In order to probe into the biological functions and pathways of genes in modules, the Gene Ontology and the pathways of Kyoto Encyclopedia of Genes and Genomes (KEGG) were annotated and visualized by cluster analyzer in Computer R language package [23]. The differentially expressed RNAs with $P < 0.05$ and foldchange > 2 were used to annotate the GO terms and KEGG pathway analyses using DAVID. The GO items can be divided into three categories: biological process (BP), cell component (CC), and molecular function (MF). BP is regarded as the most important among the diseases. The differentially expressed genes were subjected to pathway analysis, based on the KEGG database. The significant GO terms and pathways were identified by Fisher's exact test, and the results were filtered for enriched GO

terms and pathways with a false discovery rate (FDR)-corrected $P < 0.01$. The KEGG pathway and Gene ontology analysis was performed using the NovelBio Cloud Analysis Platform.

2.5 The hub genes in modules

In WGCNA analysis, the expression value of the whole gene in the module is generally standardized, and the values are all assigned as $[-1,1]$, which facilitates the comparison between modules. A ME (module engine) represents the value of a module's characteristic genes[17]. For each module, there is only one ME. ME is the first principal component (PC1) calculated and analyzed based on all gene expression data in the module, and is assigned between $[-1,1]$. The gene expression of modules is also standardized and assigned as a value between $[-1,1]$. KME, also known as module membership, or MM, represents the relevance between gene expression and module ME, and is a value between $[-1,1]$. The closer the absolute value of KME is to 1, the stronger the correlation is between the gene and the module, and the closer the gene is to the core of the module. Thus, the KME and its statistical P Value can be used as screening indexes to identify a hub gene[20]. In this analysis, the calculated saliency of hub gene with high KME is consistent with the saliency of the module. In blue and brown modules, we selected these hub genes with MM values greater than 0.8. From this set, these hub genes with fold change (Tumor vs Normal, up and down regulated) of more than 120% were identified and used to construct a gene co-expression network of the module that could be visualized with Cytoscape software.

2.6 The co-expression networks of RNA profiling

For different module genes, The Pearson Correlation Coefficient were calculated and significantly related genes were selected to construct the co-expression network [24]. In co-expression network, degree is a major index to measure the centrality in the network. It reflects the significance of a gene. In short, degree is regarded as the number of links from one node to other nodes. Furthermore, to learn some network properties, k-cores as the other parameter was introduced to simplify graph topology analysis. The k-core is defined as all genes connected to at least k other genes in the subnetwork[25]. The purpose of this strategy is to identify the core hub genes in modules.

3. Results

3.1 The RNA expression profiles for AEG

Cancer and normal samples were subjected to massive transcriptome RNA sequencing, and we obtained excellent sequencing depth, with average coverage that was approximately 200-fold coverage of the human transcriptome. Next, we identified differentially expressed RNAs between AEG tissues and normal gastric mucosa. Standardization of RNAs expression profile was gauged by fragments per kilobase of exon per million fragments mapped (FPKM). A total of 63,677 mRNA or non-coding RNA transcripts were detected. To study these RNA-seq data and to potentially identify RNA transcript levels that differed between the two groups of samples, more stringent filtering criteria were used (Fold change ≥ 2 or Fold change ≤ 0.5 , False Discovery Rate < 0.05). With this analysis, a total of 3,311 mRNAs and 2,013 non-

coding RNAs were identified differentially expressed in AEG tumor tissue samples versus paired normal gastric mucosa samples (Figure.1 a, b, c). In addition, The GO and KEGG enrichment analysis were performed to explore the biological functions of those differentially expressed genes using Cluster Profiler software. A cut-off of 0.05 of the adjusted p value was used. To annotate the genes, we used a gene ontology system that considers three domains: Molecular Functions, Cellular Components, and Biological Processes (Figure.1 d, e).

3.2 Gene co-expression modules of WGCNA

WGCNA was performed to carry out the architectural analysis of the AEG RNAs transcriptome and to construct its co-expression modules. We analyzed 40,523 gene expression profiles derived from 22 AEG tumor tissues from our RNA expression data. The FLASH CLUST software package was carried out to perform cluster analysis on these samples (Figure 2a). We selected $\beta=11$ as the soft-thresholding power to built a scale-free network resulting in a scale-free topology index of 0.8 (Figure 2b and c). Twenty-five distinct co-expression gene modules were identified that contained 178-3039 genes per module (Figure 2d, e and f).

3.3 Modules with clinical significance and preservation

Modules with RNAs expression profile patterns and interactions with modules associated with clinical traits were identified based on the correlation between module eigengenes (MEs) and clinical traits. In order to evaluate the clinical significance of each modules, correlations between MEs index and commonly-used clinical tumor marker traits (CEA, CA199, CA242, AFP, CA724, CA50, SCC, TPA, TPS and VEGF) were calculated and analyzed (Table S2). To analyze gene weight co-expression, tumor markers were used as priority correlation analysis indicators. This is a reasonable strategy, because when logical grouping is used for association analysis, tumor markers are highly reliable as measured numerical indicators and continuous variables. As the result, there were six separate modules mostly positively correlated with clinical tumor marker, and these modules are colored purple, cyan, blue, grey60, brown, and turquoise (Figure 3a, b and c). The purple module was positively correlated to CEA ($p < 0.05$). The cyan module was positively correlated to CA199 and CA242 ($p < 0.05$). The blue module was positively correlated to CA724, SCC, and TPS ($p < 0.05$). The grey60 module was positively correlated to CA199, CA242, CA724, CA50, and VEGF ($p < 0.05$). The brown module was positively correlated to TPA ($p < 0.05$). The turquoise module was positively correlated to AFP ($p < 0.05$).

3.4 The functional enrichment analysis in gene modules of interest

In the study, the blue and brown modules were of particular interest. There were 325 differentially expressed RNAs (Tumor vs Normal, False Discovery Rate < 0.05) in blue module and mapped to the GO database to determine their potential biological functions. The bioinformatics results suggested that functions of extracellular matrix organization, inflammatory response, collagen catabolic process, extracellular matrix disassembly, and cell adhesion were the top five enriched Biological Process groups. In addition, differentially expressed RNAs in blue module were also involved in immune response,

angiogenesis, positive regulation of cell proliferation, and vascular endothelial growth factor production, which were all known to be related to tumorigenesis (Figure 4a). KEGG pathway analysis indicated that these differentially expressed RNAs in blue module participating in ECM-receptor interaction, TNF signaling pathway, NF-kappa B signaling pathway, focal adhesion, Jak-STAT signaling pathway, and the PI3K-Akt signaling pathway. All are tumorigenesis-related signal pathways (Figure 4b).

The brown module includes 1,207 differentially expressed RNAs (Tumor vs Normal, False Discovery Rate < 0.05). Mapping of these RNAs to the GO database showed enrichments in functions of cell cycle, cell division and DNA replication. In addition, RNAs in this group were also involved in chromosome segregation, microtubule-based movement, and ATP catabolic process, all processes related to tumor cell division and replication (Figure 4c). KEGG pathway analysis suggested that these genes participated in the p53 signaling pathway, viral carcinogenesis, pathways in cancer, and ubiquitin mediated proteolysis. All are tumorigenesis-related signal pathways (Figure 4d).

3.5 Module visualization and hub gene identification

In the blue and brown modules, 200 RNA molecules (Figure 5a) and 157 RNA molecules (Figure 5b) were respectively differentially expressed (Tumor vs Normal; Fold change ≥ 1.2 Fold change ≤ 0.83 ; FDR<0.05) and analyzed for inclusion in a co-expression network. Within the co-expression networks, two potentially related RNAs are connected by a string. This connection indicates a correlation between different RNAs and a potential regulatory relationship (Figure 5a and b). The co-expression network analysis showed that one RNA may correlate with one to dozens of others RNAs, suggesting an interaction between critical RNAs in AEG. After analysis, we found that the expression of CD93 was directly related to the remainder of the 199 genes in the blue module, indicating a high degree of relativity. Therefore, the CD93 could be a candidate clinical biomarker in AEG patients, and as a candidate gene of AEG, which merits further research (Figure 5a).

The co-expression network structure of brown module RNAs was completely different from that of the blue module. For each pair of RNAs, the solid lines represent the positive correlation interactions between RNA molecules, while the dashed lines represent the negative correlation interactions. High structural cohesion levels are presented using red color (Figure 5b).

3.6 Identification of candidate hub genes

The intra-module connectivity of each RNAs was determined by summing the connection strengths with other module RNAs molecules and the RNAs with high intramodular connectivity are initially assumed as intramodular hub genes[26]. As the hub gene with the highest connectivity in the blue module, CD93 was highly expressed in AEG tumors as well as in the GC cases of TCGA dataset and the validation of GC dataset of TCGA were associated with tumor grade (Table 1, Figure 6a and b)[27]. The RNA-seq data of TCGA and survival information were subjected to survival analysis. The results showed that GC patients with lower expression of CD93 had longer overall survival (Tumor vs Normal, FDR<0.05) (Figure 6c)[27]. These bioinformatics and clinical data suggest that CD93 may play an important biological role in

tumorigenesis, both in AEG and other types of GC. Therefore, CD93 has the potential to be considered a biomarker and further study is warranted.

TRIM28, SLC3A2, CBX4, PATL1, and ZNF473 were the top hub genes with the high kME values in the brown module (Table S3), and based on the statistical analysis, the functions of these genes are potentially consistent with brown module. These five genes are also at the center of the network, indicating that they are likely to play a key role in the gene co-expression network. All of them were expressed at a high level both in the GC cases of TCGA dataset and in our AEG dataset (Tumor vs Normal, FDR<0.05) (Figure S1-S5)[27]. In addition, the RNA-seq data of TCGA and survival information were subjected to survival analysis, similarly, the results showed that patients with lower expression of TRIM28, SLC3A2, CBX4, PATL1, and ZNF473 had longer overall survival (Tumor vs Normal, FDR<0.05) (Figure S1-S5). It is interesting to note that, TRIM28, SLC3A2, CBX4, PATL1, and ZNF473 were positively highly expressed in most malignant tumor tissues with TCGA pan-cancer analysis[27]. According to bioinformatics analysis results, the data set supported that these genes might be associated with tumor cell proliferation, which was consistent with enrichment analysis of GO and KEGG pathways of the brown module. Therefore, TRIM28, SLC3A2, CBX4, PATL1, and ZNF473 have the potential to be considered candidate biomarkers and further study is warranted.

4. Discussion

The present study aimed to explore the transcriptomic landscape of RNAs in AEG and sought the most significant mRNA biomarker in the pathogenesis of AEG. Thus, the RNA sequencing technology was used to characterize clinically actionable driver events in 22 patients with AEG. The high-throughput platform provides a tool for medical oncology with broad clinical application prospects. However, it is difficult to utilize such a large number of genes in clinical applications. WGCNA, one of regulatory network methods, is based on power law distribution and constructing weighted networks by a soft thresholding. It can detect modules, identify hub genes, and recognize candidate genes or modules relating to external information. In the present study, a global RNAs co-expression network was constructed to identify a number of candidate hub genes in the carcinogenesis of AEG. A total of 25 co-expression gene modules were built for 40,523 RNA transcripts from 22 human AEG samples by the WGCNA method. Compared with other methods, WGCNA has many advantages since this strategy focuses on the relations between co-expression gene modules and clinical traits, providing convincing results of bioinformatics analysis[28]. RNA molecules in the same module are supposed to be functionally related. Therefore, the purpose of this study is to identify AEG tumorigenesis related modules and central hub genes, which may serve as biomarkers for AEG detection or treatment [26]. Through WGCNA analysis, we identified six gene modules that mostly relate to commonly-used tumor markers. We assume that the gene module, which is expressed closely with tumor markers, enriches shared functional annotations, may provide the best predicted candidate gene set, and may be the basis for a specific biological process.

According to correlation study by module-trait associations (Figure 3), modules purple, cyan, blue, grey60, brown, and turquoise were mostly positively correlated with clinical tumor marker. We next performed GO

and KEGG pathways analysis to obtain insight into the potential functions of the genes in modules. It is noteworthy that genes in the blue and brown modules were found to be involved in immune response, angiogenesis, positive regulation of cell proliferation, chromosome segregation, microtubule-based movement, ATP catabolic process, and vascular endothelial growth factor production, activities that are related to cancer (Figure 4). KEGG pathway enrichment analysis suggested these genes participate in ECM-receptor interaction, Jak-STAT signaling pathway, focal adhesion, TNF signaling pathway, p53 signaling pathway, viral carcinogenesis, ubiquitin mediated proteolysis, NF-kappa B signaling pathway, and the PI3K-Akt signaling pathway. These are all cancer-related signal pathways (Figure 4). WGCNA analysis and correlation study by module-trait associations allowed identification of the blue and brown modules as candidate hub gene sets.

The protein CD 93 is a cell surface glycoprotein, which was initially identified as a myeloid cell specific marker. More recent work of CD93 has identified a role in intercellular adhesion and the clearance of apoptotic cells[29, 30]. Furthermore, previous studies have shown that CD93 protein was highly expressed in tumor tissue and tumor-related vascular endothelial cells, but low in non-proliferating endothelial cells[31], and overexpression of CD93 promoting the malignancy of nasopharyngeal carcinoma and glioma [32, 33]. According to the meta-analysis data of breast, kidney and head and neck cancer cases, CD93 was positively correlated with tumor angiogenesis, so CD93 was identified as one of the top 20 core genes of primary tumor angiogenesis[34]. At present, the research on the mechanism of CD93 regulating tumorigenesis is scarce. The research of its molecular mechanism is mainly focused on vascular endothelial cell adhesion, vascular regeneration, and promoting tumor growth. As a member of the membrane protein family, we speculated that during the occurrence of GC and AEG, there should be ligand proteins binding to CD93, which promoted the biological process of tumor occurrence and development. To date, the relationship between CD93 and GC has not been reported, but our bioinformatics analysis suggested that CD93 may play an important biological role in tumorigenesis, both in AEG and GC and it has the potential to be considered a biomarker and further study is warranted.

In the test and validation TCGA datasets, TRIM28, SLC3A2, CBX4, PATL1, and ZNF473 were inversely associated with prognosis. On the basis of Kaplan–Meier plotter, the results showed that GC patients with lower expression of TRIM28, SLC3A2, CBX4, PATL1, and ZNF473 had longer overall survival, and remarkably, these hub genes were positively highly expressed in most malignant tumor tissues with TGCA pan-cancer analysis (Tumor vs Normal, FDR<0.05) (Figure S1-S5). Although TRIM28, SLC3A2, CBX4, PATL1, and ZNF473 have not been reported in AEG and GC studies, the results of our study suggested that these hub genes may be associated with cancer cells proliferation. The enrichment analysis confirmed that the hub genes in the brown module are related to tumor cell division, replication, the p53 signaling pathway, viral carcinogenesis, pathways in cancer, and ubiquitin-mediated proteolysis. Therefore, as top hub gene in brown module, TRIM28, SLC3A2, CBX4, PATL1, and ZNF473 may be considered biomarkers and future work should investigate the molecular mechanisms of these genes in AEG, GC even in pan-cancers.

5. Conclusions

In summary, through WGCNA analysis, our study identified CD93, TRIM28, SLC3A2, CBX4, PATL1, and ZNF473 as candidate biomarker genes in AEG. For most of these genes, this was the first association with GC and AEG. These genes were expressed at a high level both in TCGA datasets and our AEG dataset, indicating they may contribute to the tumorigenesis of GC and AEG. Nevertheless, the present study has some limitations, because of the data were only analyzed by bioinformatics, lacking the validation of molecular biology experiments. Overall, the identified candidate hub genes merit further study to provide greater insight into understanding GC and AEG, with the ultimate goal of personalized therapy.

Declarations

Ethics approval and consent to participate

The present study was approved by the Human Ethics Committee of the Shanxi Medical University (Taiyuan, Shanxi province of China) and written informed consent was obtained from all subjects. Methods were carried out in “accordance” with the approved guidelines.

Consent for publication

All authors have read, edited, and approved the manuscript for publication.

Availability of data and materials

The data used and analyzed during this study are available from the corresponding author on request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the Leading Clinical Department Sponsored By ‘136’ Healthcare Improvement Project of Shanxi Province.

Author contributions

Zhiyong Lai, Wenhui Yang, Kewei Jiang and Jun Xu acquired the data and created a draft of the manuscript; Kai Jia, Tiantian Zhang, and Weibin Li collected the data and performed the statistical analysis and analyzed the results; Jun Xu and Kewei Jiang revised and approved the final version of the manuscript. All authors read and approved the manuscript and agree to be accountable for all aspects of the research in ensuring that the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

We thank members of our laboratory for helpful discussion and Novel Bioinformatics Ltd., Co for the support of bioinformatics analysis with their NovelBio Cloud Analysis Platform.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA Cancer J Clin* 2018, **68**(6):394-424.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J: **Cancer statistics in China, 2015.** *CA Cancer J Clin* 2016, **66**(2):115-132.
3. Devesa SS, Blot WJ, Fraumeni JJ: **Changing patterns in the incidence of esophageal and gastric carcinoma in the United States.** *CANCER-AM CANCER SOC* 1998, **83**(10):2049-2053.
4. Bai JG, Lv Y, Dang CX: **Adenocarcinoma of the Esophagogastric Junction in China according to Siewert's classification.** *JPN J CLIN ONCOL* 2006, **36**(6):364-367.
5. Blot WJ, Devesa SS, Kneller RW, Fraumeni JJ: **Rising incidence of adenocarcinoma of the esophagus and gastric cardia.** *JAMA* 1991, **265**(10):1287-1289.
6. Zhou Y, Zhang Z, Zhang Z, Wu J, Ren D, Yan X, Wang Q, Wang Y, Wang H, Zhang J *et al*: **A rising trend of gastric cardia cancer in Gansu Province of China.** *CANCER LETT* 2008, **269**(1):18-25.
7. Chung JW, Lee GH, Choi KS, Kim DH, Jung KW, Song HJ, Choi KD, Jung HY, Kim JH, Yook JH *et al*: **Unchanging trend of esophagogastric junction adenocarcinoma in Korea: experience at a single institution based on Siewert's classification.** *DIS ESOPHAGUS* 2009, **22**(8):676-681.
8. Kusano C, Gotoda T, Khor CJ, Katai H, Kato H, Taniguchi H, Shimoda T: **Changing trends in the proportion of adenocarcinoma of the esophagogastric junction in a large tertiary referral center in Japan.** *J Gastroenterol Hepatol* 2008, **23**(11):1662-1665.
9. Marsman WA, Birjmohun RS, van Rees BP, Caspers E, Johan G, Offerhaus A, Bosma PJ, Jan J, van Lanschot B: **Loss of heterozygosity and immunohistochemistry of adenocarcinomas of the esophagus and gastric cardia.** *CLIN CANCER RES* 2004, **10**(24):8479-8485.
10. Ott K, Weber WA, Fink U, Helmberger H, Becker K, Stein HJ, Muller J, Schwaiger M, Siewert JR: **Fluorodeoxyglucose-positron emission tomography in adenocarcinomas of the distal esophagus and cardia.** *WORLD J SURG* 2003, **27**(9):1035-1039.
11. **Comprehensive molecular characterization of gastric adenocarcinoma.** *NATURE* 2014, **513**(7517):202-209.
12. **Integrated genomic characterization of oesophageal carcinoma.** *NATURE* 2017, **541**(7636):169-175.
13. **Comprehensive molecular characterization of human colon and rectal cancer.** *NATURE* 2012, **487**(7407):330-337.

14. **Comprehensive genomic characterization of squamous cell lung cancers.** *NATURE* 2012, **489**(7417):519-525.
15. **Comprehensive molecular portraits of human breast tumours.** *NATURE* 2012, **490**(7418):61-70.
16. Nagahashi M, Wakai T, Shimada Y, Ichikawa H, Kameyama H, Kobayashi T, Sakata J, Yagi R, Sato N, Kitagawa Y *et al.* **Genomic landscape of colorectal cancer in Japan: clinical implications of comprehensive genomic sequencing for precision medicine.** *GENOME MED* 2016, **8**(1):136.
17. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC BIOINFORMATICS* 2008, **9**:559.
18. Yepes S, Lopez R, Andrade RE, Rodriguez-Urrego PA, Lopez-Kleine L, Torres MM: **Co-expressed miRNAs in gastric adenocarcinoma.** *GENOMICS* 2016, **108**(2):93-101.
19. Giulietti M, Occhipinti G, Principato G, Piva F: **Weighted gene co-expression network analysis reveals key genes involved in pancreatic ductal adenocarcinoma development.** *Cell Oncol (Dordr)* 2016, **39**(4):379-388.
20. Chen J, Wang X, Hu B, He Y, Qian X, Wang W: **Candidate genes in gastric cancer identified by constructing a weighted gene co-expression network.** *PEERJ* 2018, **6**:e4692.
21. Liu X, Hu AX, Zhao JL, Chen FL: **Identification of Key Gene Modules in Human Osteosarcoma by Co-Expression Analysis Weighted Gene Co-Expression Network Analysis (WGCNA).** *J CELL BIOCHEM* 2017, **118**(11):3953-3959.
22. Shi K, Bing ZT, Cao GQ, Guo L, Cao YN, Jiang HO, Zhang MX: **Identify the signature genes for diagnose of uveal melanoma by weight gene co-expression network analysis.** *Int J Ophthalmol* 2015, **8**(2):269-274.
23. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters.** *OMICS* 2012, **16**(5):284-287.
24. Prieto C, Risueno A, Fontanillo C, De Las RJ: **Human gene coexpression landscape: confident network derived from tissue transcriptomic profiles.** *PLOS ONE* 2008, **3**(12):e3911.
25. Vazquez A, Dobrin R, Sergi D, Eckmann JP, Oltvai ZN, Barabasi AL: **The topological relationship between the large-scale attributes and local interaction patterns of complex networks.** *Proc Natl Acad Sci U S A* 2004, **101**(52):17940-17945.
26. Wan Q, Tang J, Han Y, Wang D: **Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma.** *EXP EYE RES* 2018, **166**:13-20.
27. Chandrashekar DS, Bashel B, Balasubramanya S, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, Varambally S: **UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses.** *NEOPLASIA* 2017, **19**(8):649-658.
28. Chou WC, Cheng AL, Brotto M, Chuang CY: **Visual gene-network analysis reveals the cancer gene co-expression in human endometrial cancer.** *BMC GENOMICS* 2014, **15**:300.
29. Greenlee-Wacker MC, Galvan MD, Bohlson SS: **CD93: recent advances and implications in disease.** *CURR DRUG TARGETS* 2012, **13**(3):411-420.

30. Nepomuceno RR, Henschen-Edman AH, Burgess WH, Tenner AJ: **cDNA cloning and primary structure analysis of C1qR(P), the human C1q/MBL/SPA receptor that mediates enhanced phagocytosis in vitro.** *IMMUNITY* 1997, **6**(2):119-129.
31. Ronca R, Benkheil M, Mitola S, Struyf S, Liekens S: **Tumor angiogenesis revisited: Regulators and clinical implications.** *MED RES REV* 2017, **37**(6):1231-1274.
32. Langenkamp E, Zhang L, Lugano R, Huang H, Elhassan TE, Georganaki M, Bazzar W, Loof J, Trendelenburg G, Essand M *et al.*: **Elevated expression of the C-type lectin CD93 in the glioblastoma vasculature regulates cytoskeletal rearrangements that enhance vessel function and reduce host survival.** *CANCER RES* 2015, **75**(21):4504-4516.
33. Bao L, Tang M, Zhang Q, You B, Shan Y, Shi S, Li L, Hu S, You Y: **Elevated expression of CD93 promotes angiogenesis and tumor growth in nasopharyngeal carcinoma.** *Biochem Biophys Res Commun* 2016, **476**(4):467-474.
34. Masiero M, Simoes FC, Han HD, Snell C, Peterkin T, Bridges E, Mangala LS, Wu SY, Pradeep S, Li D *et al.*: **A core human primary tumor angiogenesis signature identifies the endothelial orphan receptor ELTD1 as a key regulator of angiogenesis.** *CANCER CELL* 2013, **24**(2):229-241.

Table

Table 1. Expression of CD93 in GC of TCGA dataset based on tumor grade

Comparison (VS)	Grade 1	Grade 2	Grade 3
Statistical significance	Well differentiated (low grade)	Moderately differentiated (intermediate grade)	Poorly differentiated (high grade)
Normal	1.106850E-01	1.93101999999001E-05 (FDR<0.05)	3.47899486996539E-11 (FDR<0.05)
Grade 1	--	3.350600E-01	0.000000E+00 (FDR<0.05)
Grade 2	--	--	4.944000E-03 (FDR<0.05)

Figures

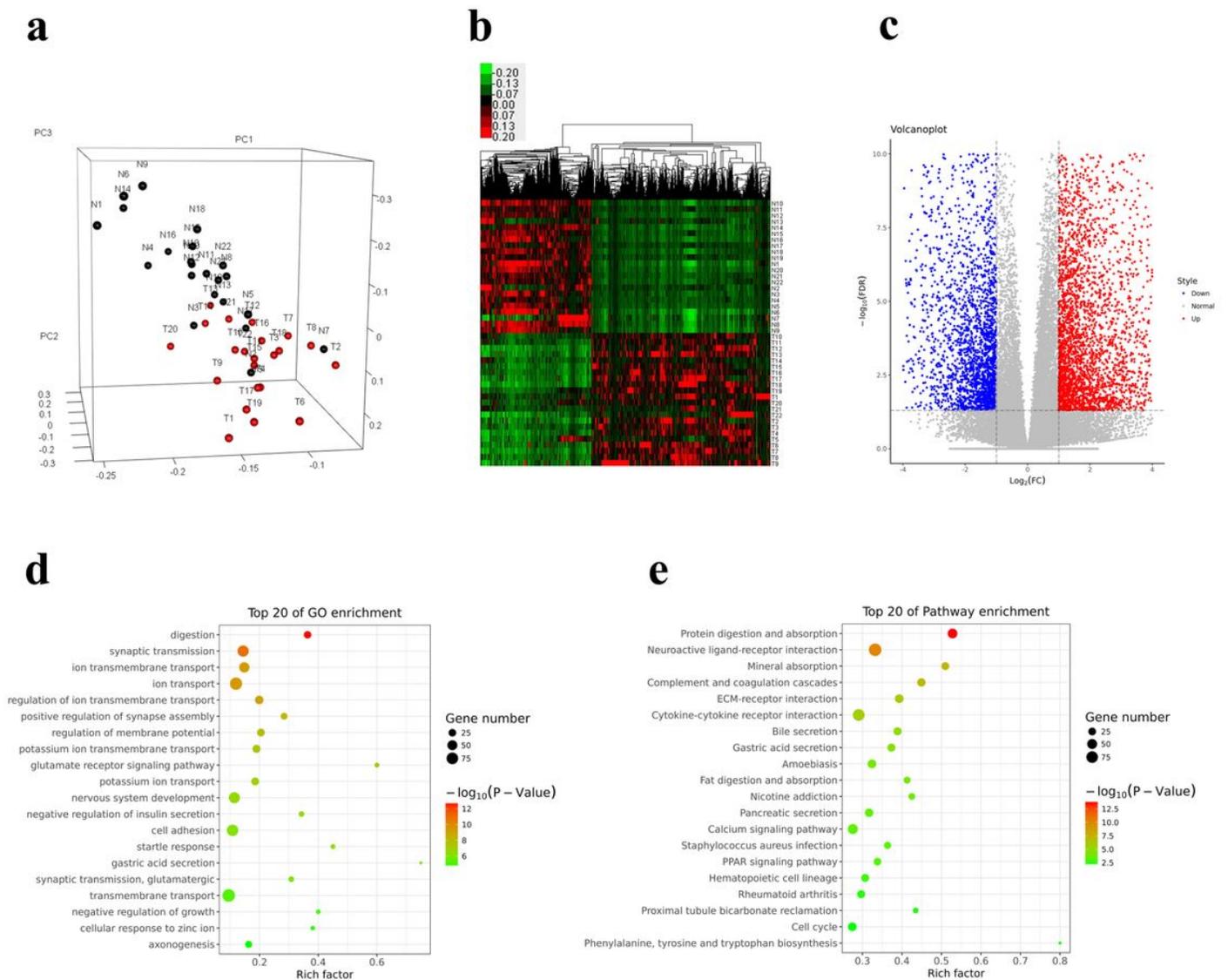


Figure 1

Analysis of AEG RNA-Seq data. a: Principal component analysis of RNA-seq data of AEG, Red dots and black dots represent AEG tissues and paired normal gastric mucosa tissues, respectively. b: Heat map of the mRNA and noncoding RNA profiles in AEG tissues versus paired normal gastric mucosa tissues. Red indicates high relative expression and green indicates low relative expression. RNA expression levels hierarchically clustered on the y-axis, and the tissue samples are hierarchically clustered on the x-axis (Fold change ≥ 2 , $p < 0.05$). Expression levels are presented in red and green, indicating upregulated and downregulated RNAs, respectively. Numbers marked with T and N are from twenty-two paired AEG tissues and paired adjacent normal tissues, respectively. c: Volcano plot of differentially-expressed mRNAs and noncoding RNAs between AEG tissue versus paired normal gastric mucosa tissue samples. Volcano plots were constructed using fold-change values and p-values. The vertical lines correspond to 2.0-fold up- and downregulation between AEG tissue versus paired normal gastric mucosa tissue samples (T vs N), and the horizontal lines represent p-values. Red plot points represent differentially expressed mRNAs and

noncoding RNAs with statistical significance. d: Analysis of deregulated genes in AEG by GO enrichment. e: Analysis of deregulated genes in AEG by Pathway enrichment.

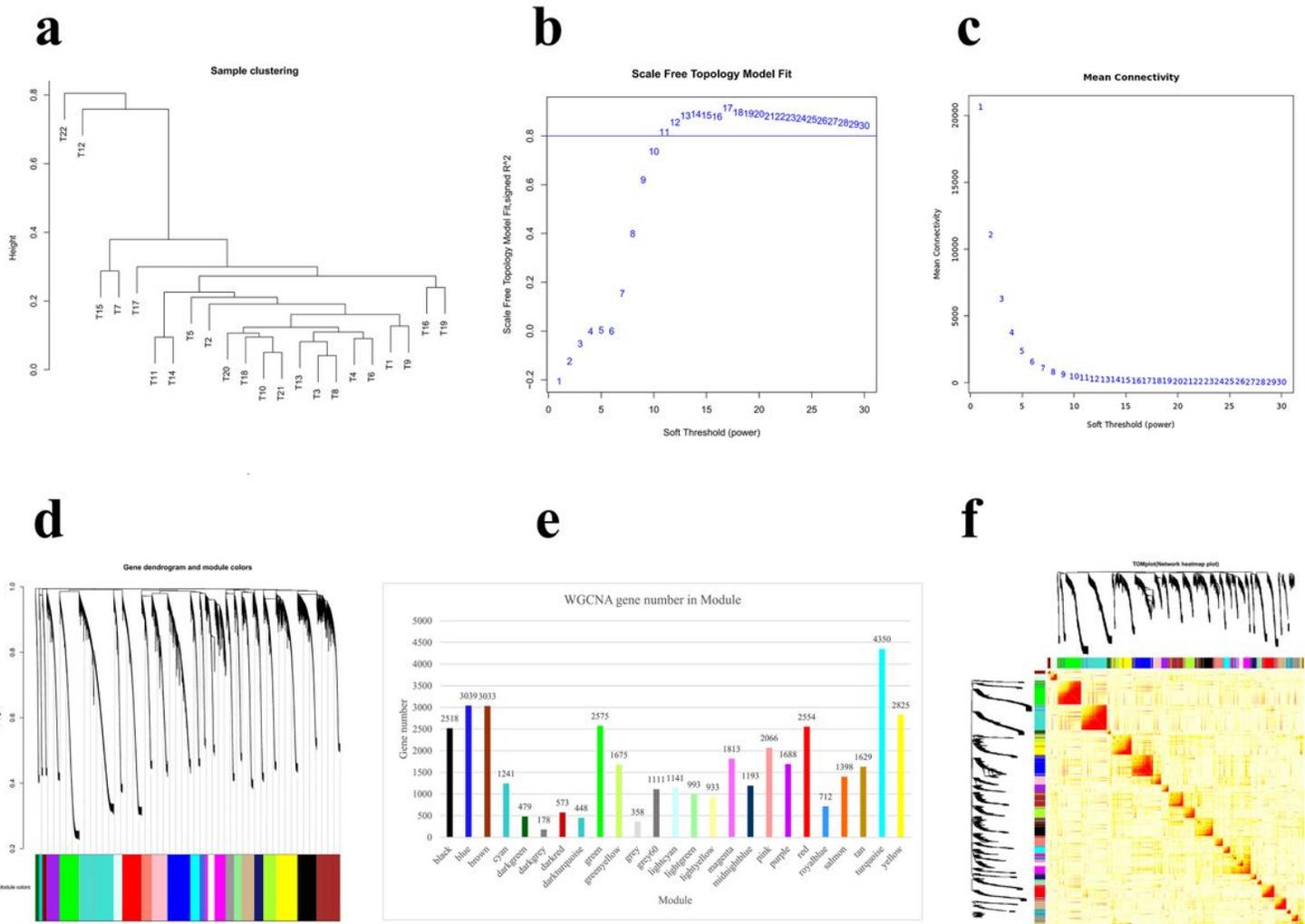


Figure 2

Weighted gene co-expression network analysis of AEG RNA-Seq data. a: Sample clustering to detect outliers. All AEG samples are included in the clusters. b: The scale-free fitting exponent (γ) as a function of soft threshold power (β) (x axis). c: Analysis of the mean connectivity (degree, y-axis) for various soft-thresholding powers (x-axis). d: Hierarchical clustering and module colors. As a result, 25 co-expression modules were constructed and are shown in different colors. These modules included different numbers of genes. e: The number of genes in the 25 modules. f: The heatmap depicts the Topological Overlap Matrix (TOM) among all genes in the analysis. Hierarchical clustering dendrogram branch correspond to each module. In a thermal diagram, a gradual saturation of yellow and red indicates a high correlation.

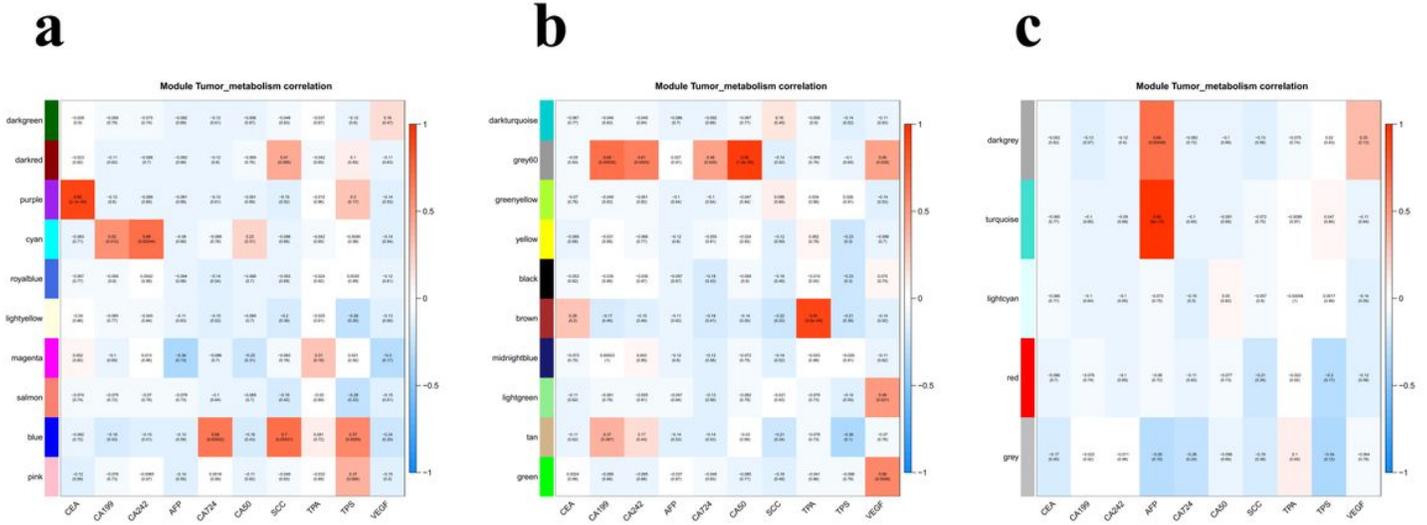


Figure 3

The correlation of module traits was evaluated by the correlation between MEs and clinical traits. Each row corresponds to a module eigen gene, and each column corresponds to a clinical tumor marker trait (CEA, CA199, CA242, AFP, CA724, CA50, SCC, TPA, TPS and VEGF). Each cell contains a corresponding correlation and p-value. The table is color-coded by correlation based on the color legend.

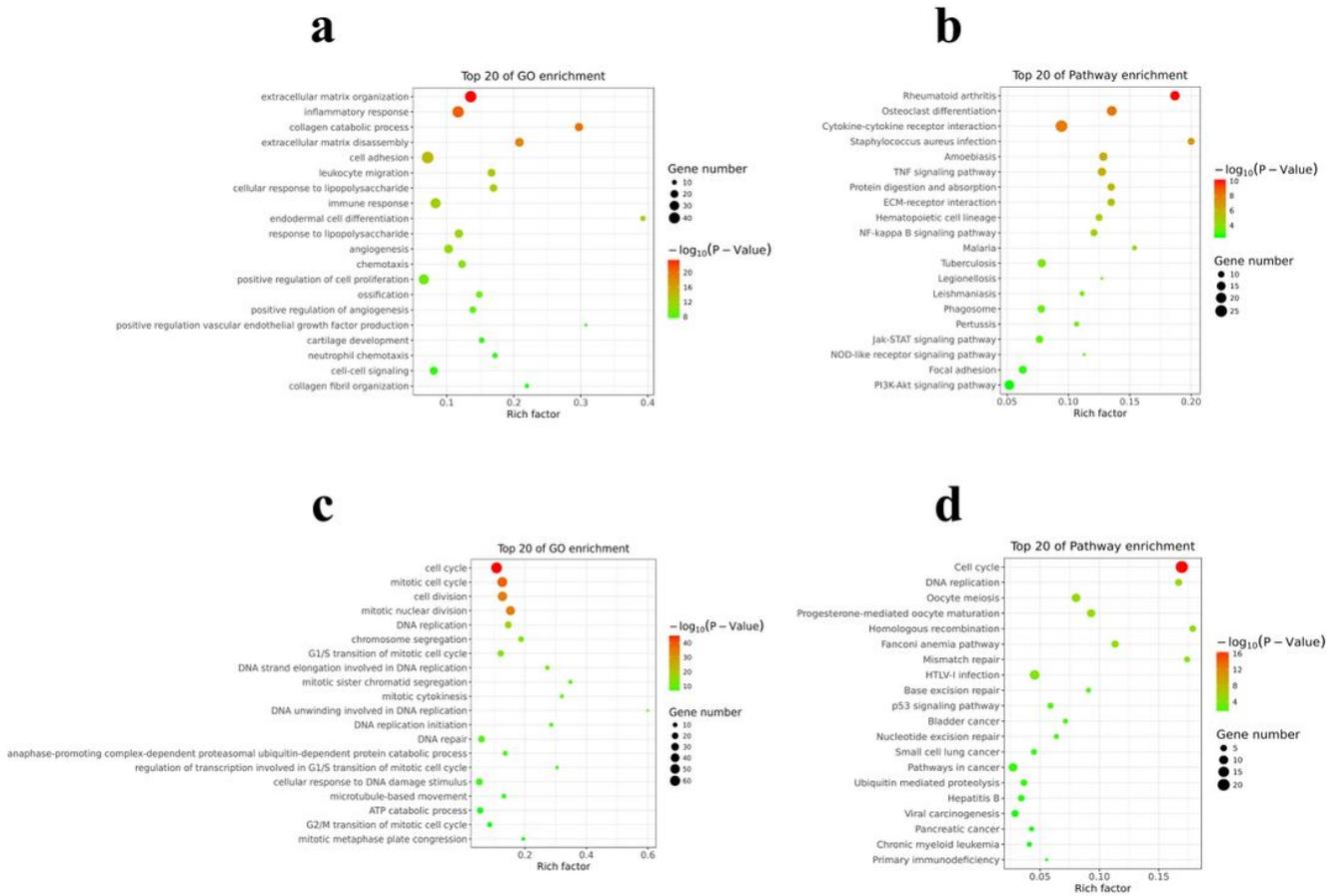


Figure 4

The most significantly enriched GO annotations and pathways of genes in the blue and brown modules. a: Top 20 significantly enriched biological process GO annotations in the blue module; b: Top 20 significantly enriched KEGG pathways in the blue module. c: Top 20 significantly enriched biological process GO annotations in the brown module; d: Top 20 significantly enriched KEGG pathways in the brown module.

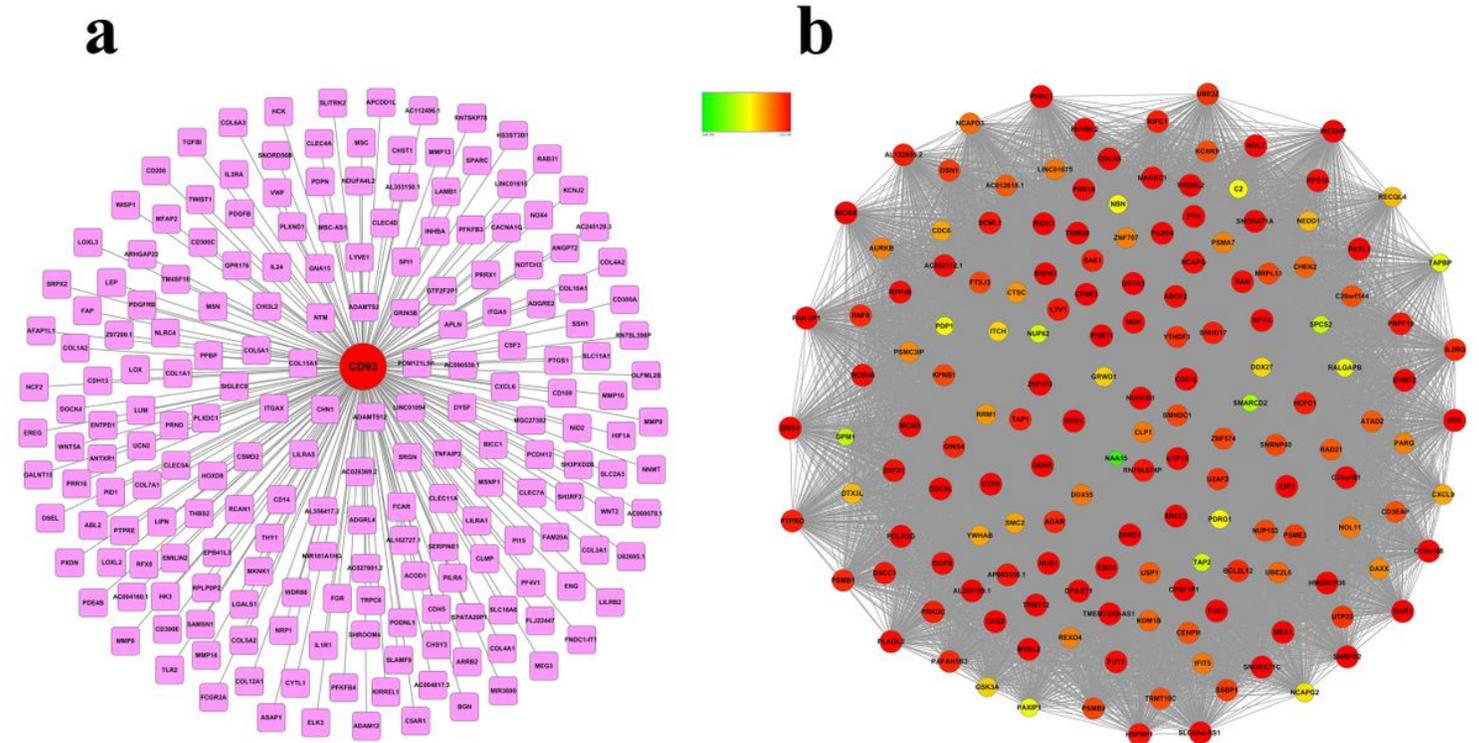


Figure 5

Co-expression network of hub genes in the blue and brown modules. Nodes and lines represent genes and correlation between genes. Red indicates a high cohesion level. a: Co-expression network of hub genes in the blue module. B: Co-expression network of hub genes in the brown module.

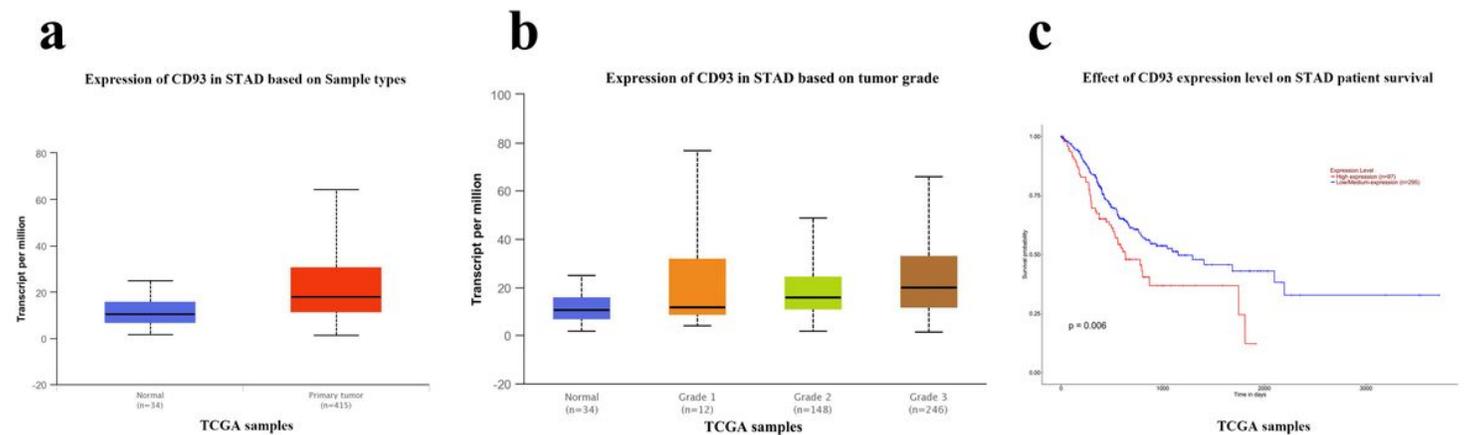


Figure 6

Expression and Kaplan-Meier curves of CD93 in TCGA GC dataset. a: Expression of CD93 in TCGA GC dataset. b: Expression of CD93 in GC based on tumor grade. c: Kaplan-Meier curves of CD93 in GC dataset.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalFigures.zip](#)