

# Identification of the co-expression module-trait relationships and hub genes associated with the diffuse-type gastric cancer

**Zaisheng Ye**

Fujian Provincial Cancer Hospital

**Miao Zheng**

Fujian Provincial Cancer Hospital

**Yi Zeng**

Fujian Provincial Cancer Hospital

**Yunqing Xie**

Fujian Provincial Cancer Hospital

**Shenghong Wei**

Fujian Provincial Cancer Hospital

**Yi Wang**

Fujian Provincial Cancer Hospital

**Zhitao Lin**

Fujian Provincial Cancer Hospital

**Shu Chen**

Fujian Provincial Cancer Hospital

**Zhiwei Wang**

Fujian Provincial Cancer Hospital

**Qihong Zheng**

Fujian Provincial Cancer Hospital

**Luchuan Chen** (✉ [luchuanchen@sina.cn](mailto:luchuanchen@sina.cn))

Fujian Provincial Cancer Hospital

---

## Research Article

**Keywords:** DGC, Co-expression module, WGCNA, Biomarker

**Posted Date:** March 3rd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-242786/v1>

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

[Read Full License](#)

---

**Identification of the co-expression module-trait relationships and hub genes associated with the diffuse-type gastric cancer.**

Zaisheng Ye<sup>1\*</sup> Miao Zheng<sup>2\*</sup> Yi Zeng<sup>1\*</sup> Yunqing Xie<sup>3\*</sup> Shenghong Wei<sup>1\*</sup> Yi Wang<sup>1</sup>  
Zhitao Lin<sup>1</sup> Shu Chen<sup>1</sup> Zhiwei Wang<sup>1</sup> Qihong Zheng<sup>3</sup> Luchuan Chen<sup>1</sup>

1. Department of Gastrointestinal Surgical Oncology, Fujian Cancer Hospital & Fujian Medical University Cancer Hospital, Fuzhou 350014, Fujian Province, China

2. Department of Clinical Laboratory, Fujian Provincial Maternity and Children Hospital, Affiliated Hospital of Fujian Medical University, Fuzhou 350001, Fujian Province, China

3. Department of Fujian Provincial Key Laboratory of Tumor Biotherapy, Fujian Cancer Hospital & Fujian Medical University Cancer Hospital, Fuzhou 350014, Fujian Province, China

\*: These Authors contributed equally to the work.

Corresponding author: Luchuan Chen, Department of Gastrointestinal Surgical Oncology, Fujian Cancer Hospital, No. 420 Fu-ma Road, Fuzhou 350014, China.

E-mail address: luchuanchen@sina.cn.

## **ABSTRACT**

Diffuse-type gastric cancer (DGC) is a gastric cancer subtype based on Laurén histological type, with poor prognosis when compared intestinal-type gastric cancer. It is necessary that comprehensive analysis identifies pathways and genes involved in DGC. RNA expression data of DGC (including 52 samples, top 5000 genes) and fourteen patient clinic traits were downloaded from The Cancer Genome Atlas (TCGA) database. Co-expression modules and module-trait relationships were constructed by weighted gene co-expression network analysis (WGCNA). The identified ten co-expression modules were obtained from DGC samples. The co-expression turquoise module positively correlated with longest dimension. The co-expression magenta module positively correlates with gender, and person neoplasm cancer status. The co-expression green module also positively correlates with pathologic N stage and pathologic stage. Besides, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment obtained some cancer related signal pathways in the co-expression turquoise and green module, such as proliferation, drug resistance, cell metabolism. Additionally, we identified the hub genes in the co-expression turquoise, green and magenta modules. Protein-protein interaction (PPI) and biological processes (BP) analysis were performed on the hub genes in those modules, suggesting the key role of those modules in the development of the cancer. The hub genes were significantly related to Overall survival analysis showed some important hub genes were significant in

DGC patients, including ANK2, GNAO1, MAP6, RERG (turquoise module), MEP1A, GUCY2C, ALDOB (green module), XIST, TTTY15, NCRNA00185, CYorf15B (magenta module). The study provided new biomarkers and potential mechanisms in DGC.

**Keywords:** DGC, Co-expression module, WGCNA, Biomarker

## **INTRODUCTION**

Globally, gastric cancer is the fifth leading cause of cancer and the third leading cause of death from cancer, making up 7% of cases and 9% of deaths [1]. In 1965, Lauren classified gastric cancer into intestinal-type and diffuse-type according to its histological structure and biological behavior [2]. Diffuse-type gastric cancer (DGC) originates from gastric inherent mucosa with poor prognosis, which is poorly differentiated, diffuse growth, lack of cell junction, and generally does not form glandular ducts. Many poorly differentiated adenocarcinomas and signet ring cell carcinomas belong to DGC. Women incidence of DGC is higher than men [3]. Treatments may include some combination of surgery, chemotherapy, radiation therapy and targeted therapy. The poor prognosis could be because most people were diagnosed with advanced disease. Identification of new biomarkers and potential mechanisms are urgent to improve prognosis of DGC [4].

The Cancer Genome Atlas (TCGA), as a large public database, has provided abundant genomic and clinical data of various cancers. So far, a few biomarkers and molecular mechanisms in many kinds of cancers have been

discovered based on TCGA [5]. To identify novel prognostic markers and explore the underlying mechanisms of gastric cancer, weighted gene co-expression network analysis (WGCNA) was used for clustering highly correlated gene group (the co-expression modules) [6]. WGCNA was extensively used in various biological contexts, such as yeast genetics, cancers, genetics data, proteomic data, metabolomics data, and analysis of imaging data [7, 8]. The co-expression module-trait relationships perfectly reflected the correlation between gene-based networks and the clinical phenotypes. Eigengene networks study module relationship and intramodular connectivity help to hub genes in interesting modules. One found a preserved module consisting of 506 genes was associated with clinical traits including pathologic T stage and histologic grade by WGCNA [9]. The study previously also used WGCNA to identified two novel lncRNAs, PCAT18 and LINC01133, associated with GC development and metabolic pathways of gastrointestinal disease and function [10]. However, identification of the co-expression module-trait relationships and hub genes associated with the diffuse-type gastric cancer has not been reported. Diffuse-type gastric cancer is one subtype of gastric cancer with poor prognosis.

In this study, to investigate the DGC-related gene modules and key biomarkers, the gene expression data and clinical traits information were obtained from TCGA. WGCNA identified that co-expression modules (turquoise, magenta,

green modules) were significantly correlated with clinic traits and extracting a series of hub genes. Some of important hub genes were significantly related to overall survival in DGC patients. Our findings might discovery new biomarker and mechanisms for DGC.

## **MATERIALS AND METHODS**

### **Genetic and clinical data of DGC patients**

To download DGC-related data from TCGA database, including 72 DGC sample patients, 20530 genes, and 14 clinical traits. For all these samples, the clinical information including age at initial pathologic diagnosis, gender, longest dimension, lymph node examined count, neoplasm histologic grade, new tumor event after initial treatment, number of lymph nodes positive, pathologic M stage, pathologic N stage, pathologic T stage, pathologic stage, person neoplasm cancer status, radiation therapy, targeted molecular therapy. The gene expression missing value (expression = 0) that more than 20% was excluded, and extracted top 5000 good genes for further study.

### **Construction of weighted gene co-expression network**

To find clusters (modules) of highly correlated genes of DGC sample, WGCNA was performed using the WGCNA R package (<http://systemsbio.ucsd.edu/WGCNAdemo/>). The first step is to remove the outlier DGC samples and perform the hierarchical clustering analysis with the hclust R function. In this step, WGCNA uses a series of 'softthresholding value' to select an accurate scale free  $R^2$  value to construct co-expression modules.

The next step is module detection and the network construction. The modules whose size less than 30 were merged into one module by the `blockwiseConsensusModules` function of WGCNA R package. The third step is to calculate the associations between clinical variable and the co-expression modules, and the module eigengene (ME) is utilized. Therefore, each module's ME represents a summary measure for the overall co-expression network. In addition, the gene significance (GS) was defined as mediated p-value of each gene ( $GS = |g^P|$ ) in the linear regression between gene expression and the clinical traits.

### **The Kyoto Encyclopedia of Genes and Genomes (KEGG) base on the co-expression modules.**

KEGG is a reference knowledge database involving chemical information systems information, and genomic information. KEEG pathway (<https://david.ncifcrf.gov/home.jsp>) enrichment analysis was performed the co-expression modules with a statistical significance level of  $p < 0.05$ .

### **Identify hub genes with molecular complex detection**

The criteria of hub genes searching were molecular complex detection (MCODE) score  $>6$  by Cytoscape software (version 3.2.0; National Resource for Network Biology), and statistical significance of  $p < 0.05$  [11].

### **Construction of the protein-protein interaction (PPI) network and biological processes (BP) analysis of hub genes**

The PPI network of hub genes was constructed by STRING database (<http://string-db.org/cgi/input.pl>). Then we identified some key molecules for further overall survival analysis by KM plotter (<http://kmplot.com/private/index.php?p=home>). BP analysis of hub genes were analyzed with Cytoscape ClueGO (two-sided hypergeometric test, adjusted P value <0.05 corrected with Benjamini-Hochberg).

### **Statistical analysis**

All data were analyzed by R software 3.4.1 (<https://www.r-project.org/>). In all cases,  $p < 0.05$  was considered as statistical significance and corrected by Benjamini-Hochberg for multiple testing.

## **RESULTS**

### **Construction of the co-expression modules of DGC**

The gene expression values in 72 DGC samples (Supplementary table 1) were filtered by WGCNA package good gene function. The top 5000 genes was normalized by limma R package with voom function, then removed the auxiliary data and transposed the expression data for further WGCNA analysis to develop the co-expression modules. The clinical characteristics information of these DGC patients was listed in a table, including age at initial pathologic diagnosis, gender, longest dimension, lymph node examined count, neoplasm histologic grade, new tumor event after initial treatment, number of lymph nodes positive, pathologic M stage, pathologic N stage, pathologic T stage, pathologic stage, person neoplasm cancer status, radiation therapy,

targeted molecular therapy (Supplementary Table 2). Sample clustering was detected by a cutheight to remove outlier, and TCGA-R5-A7ZI-01 sample was removed (Figure 1 A). Sample dendrogram and trait heatmap were plotted based on gene expression data and clinical traits. The corresponding clinical data of DGC were showed the distribution (Figure 1 B). A series of 'softthresholding values' (the power values ( $\beta$ )) were plotted by scale free topology model fit (Figure 2 A and Supplementary Table 3). When the power value was equal to 3, the scale  $R^2$  was up to 0.88 (Figure 2 B). A co-expression network including top 5000 genes was constructed with WGCNA algorithm, and these co-expression modules were represented in different colors (Figure 3A). Heatmap plot of topological overlap model was plotted based on this gene network, which showed correlation between different co-expression modules (Figure 3B).

### **Gene Co-expression modules correspond to clinic traits**

The ten identified modules were constructed module-trait relationships heatmap (Figure 4 and Supplementary Table 4). The number of genes in the ten co-expression modules was shown in Table 1.

Module-trait relationships plot demonstrated that the co-expression turquoise module contained 2156 genes, co-expression red module contained 78 genes, co-expression yellow contained module 220 genes, co-expression black module contained 77 genes, co-expression brown module contained 427 genes, co-expression magenta module contained 30 genes, co-expression

green module contained 125 genes, co-expression blue module contained 577 genes, and co-expression pink module contained 58 genes. Gene co-expression modules correspond to clinic traits, including age at initial pathologic diagnosis, gender, longest dimension, lymph node examined count, neoplasm histologic grade, new tumor event after initial treatment, number of lymph nodes positive, pathologic M stage, pathologic N stage, pathologic T stage, pathologic stage, person neoplasm cancer status, radiation therapy, targeted molecular therapy. Based on the correlation coefficients and p value of module-trait relationships heat map, we further plotted the eigengene dendrogram and clinical trait heatmap (Figure 5). It indicates that three co-expression modules, including turquoise module, magenta module, green module and tan module, that significantly associated with DGC clinic traits (longest dimension, gender, person neoplasm cancer status, pathologic N stage, pathologic stage). What's more, the scatter plots of Gene Significance vs. Module Membership (MM) in this three modules were obtained respectively. The results revealed that MM in co-expression turquoise module was significantly correlated with longest dimension ( $cor=0.55$ ,  $p<0.0001$ ) (Figure 6A). MM in co-expression magenta module was significantly correlated with gender ( $cor=1$ ,  $p<0.0001$ ) and person neoplasm cancer status ( $cor=-0.65$ ,  $p<0.0001$ ) (Figure 6B). MM in co-expression green module was significantly correlated with pathologic N stage ( $cor=0.48$ ,  $p<0.0001$ ) and pathologic stage ( $cor=-0.3$ ,  $p=0.00064$ ) (Figure 6C).

## **KEGG pathway analysis of genes in interested co-expression modules**

KEGG pathway analysis revealed 28 statistically significant signaling pathways to involve genes identified in the co-expression turquoise module and 12 pathways in the co-expression green module (Supplementary table 5). None statistically significant signaling pathways were identified in the co-expression magenta module. It provided clues that DGC might potentially related to multiple signaling pathways, such as such as proliferation, drug resistance, cell metabolism, adhesion molecules. Some of those identified pathway were closely associated with cancer, including calcium signaling pathway, MAPK signaling pathway, ras signaling pathway, transcriptional misregulation in cancer, cell adhesion molecules (CAMs) in turquoise module. Calcium signaling pathway is important for cellular signalling, as once they enter the cytoplasm they exert allosteric regulatory effects on many enzymes and proteins [12]. The MAPK pathway could communicate a signal from a receptor on the surface of the cell to the DNA in the nucleus of the cell [13]. When ras pathway was 'switched on' by the other signals, it subsequently switches on other proteins, which turn on genes involved in cancer cell growth, differentiation and survival [14]. In essence, cell adhesion molecules help cells stick to each other and to their surroundings. Cell adhesion is important in affecting cellular mechanisms of growth, contact inhibition, and apoptosis [15]. Those pathways have investigated as cancer treatments. We also obtained chemical carcinogenesis, drug

metabolism, cytochrome P450, metabolism of xenobiotics by cytochrome P450, metabolic pathway, drug metabolism-other enzymes, glycolysis/ gluconeogenesis pathway in the co-expression green module. Current studies have found that a variety of cytochrome P450 enzymes are closely related to the occurrence of multidrug resistance in tumors [16]. Studies have showed that tumor cells energy metabolism characteristic is aerobic glycolysis, namely Warburg effect [17].

### **Hub genes in the co-expression modules**

Genes with high intramodular connectivity (MCODE score >6 and  $p < 0.05$ ) are considered as intramodular hub genes. A total of 19 hub genes were identified in the co-expression turquoise module, including PYGM, ANGPTL1, PDZD4, CAND2, SYPL2, GNAO1, RNF150, NECAB1, SALL2, HSPB6, PEG3, RERG, HSPB8, MAP6, GPRASP1, LRCH2, ZFP28, ANK2, TRPC1. A total of 17 hub genes were identified in the co-expression magenta module, including NCRNA00185, TMSB4Y, BCORL2, XIST, TTTY14, CYorf15B, PRKY, USP9Y, EIF1AY, TTTY15, NLGN4Y, UTY, CYorf15A, KDM5D, ZFY, DDX3Y, RPS4Y1. A total of 6 hub genes were identified in the co-expression green module, including MEP1A, GUCY2C, CHST5, ALDOB, NR1I2, C14orf176 (Supplementary table 6).

### **BP analysis and PPI of hub genes in the interest co-expression modules**

BP analysis of those hub genes in interested co-expression modules

(turquoise, magenta, and green) revealed that hub genes were mainly distributed in drug export, SCF complex assembly, insecticide metabolic process, postsynaptic membrane assembly, protein sulfation, G-protein coupled receptor catabolic process, protein localization to T-tubule, sequestering of actin monomers, protein demethylation, NADH oxidation (Figure 7A and Supplementary table 7). Those important biological processes suggested the function of hub genes we identified.

The 19 hub genes in the co-expression turquoise module, 17 hub genes in the co-expression magenta module, and 6 hub genes in the co-expression green module were uploaded to STRING for protein–protein interaction (PPI) analysis (Figure 7B and Supplementary table 8). PPI network indicated some hub genes were especially important, including ANK2, MAP6, GNAO1, RERG, PDZD4, ALDOB, MEP1A, and GUCY2C.

Additionally, overall survival analysis showed that some especially important hub genes significant related to survival rate in DGC patients, such as ANK2 (HR=1.48, P= 0.00013), MAP6 (HR=1.82, P= 4e-8), GNAO1 (HR=2.48, P= 1.5e-4), and RERG (HR=2.02, P= 1.2e-10) in the co-expression turquoise module (Figure 8A), and ALDOB (HR=0.63, P= 2.9e-06), MEP1A (HR=0.81, P=0.017), GUCY2C (HR=0.72, P= 0.0015) in the co-expression green module (Figure 8B). For the co-expression magenta module, we found that some long non coding RNAs (lncRNAs). Those genes did not code

proteins, so they would not be enriched in PPI network. Those lncRNAs were also related to DGC survival, including XIST (HR=1.45, P= 0.00013), TTTY15 (HR=1.41, P= 0.00046), NCRNA00185 (HR=1.34, P= 8E-04), CYorf15B (HR=1.34, P= 8E-04) (Figure 8C).

## **DISCUSSION**

Early symptoms of GC may include upper abdominal pain, heartburn, nausea and loss of appetite, but there are no specific clinical symptoms [18]. Most of GC patients were diagnosed as advanced stage and untreatable by surgery. Lauren histologic type is one of the most important factor associated with pattern of recurrence following resection of gastric adenocarcinoma. DGC showed poor prognosis and this is accompanied by lymph nodes metastasis. Women incidence of DGC is higher than men [19]. The study of risk factors, such as pathologic N stage or gender, is necessary [20]. Furthermore, some new targeting therapy of DGC may provide new insights into the treatment for DGC patients.

In this study, a total of ten co-expression gene modules were constructed by the top 5000 genes from the 72 human DGC samples by WGCNA analysis method, which was used to identify the relationship between DGC transcriptome (RNA sequencing data) and clinic traits. The co-expression turquoise module, magenta module, and green module were significantly related to some clinical traits. KEGG pathway analysis of genes in those interested co-expression modules revealed 28 statistically significant signaling

pathways to involve genes identified in the co-expression turquoise module and 12 pathways in the co-expression green module. Here, we would discuss some important cancer related pathways. For example, the co-expression turquoise module was found to be enriched in MAPK signaling pathway, and contained some key genes (FGF18, FGF5, FGF9, FGF14, CACNB2, FGF10, FGF13, CACNB4, MAPT, RASGRP2, FGF1, FGF2, NTF3, PTPN5, CACNG7, PTPRR, CACNG4, FLNC, CACNA2D3, CACNA2D2, RPS6KA6, NTRK1, NTRK2, CACNA1G, CACNA1E, CACNA1F, CACNA1C, DUSP9, PLA2G4E, NGF, PLA2G4D, CACNA1B) in this cell proliferation related pathway. MAPK pathway was an important tumor proliferation-related pathway. Some of the identified genes in MAPK pathway have been reported. Expression of basic fibroblast growth factor (FGF18) was a prognostic indicator in gastric cancer. FGF18 may participate in angiogenesis in gastric cancer and may be involved in tumor invasion and metastasis [21]. Additionally, neurotrophic tyrosine kinase receptor type 2 (NTRK2) promotes proliferation and invasion, relating to cell dedifferentiation, tumor budding, and poor prognosis of GC [22]. The combination of microtubule associated protein-tau (MAPT) and  $\beta$ -tubulin III (TUBB3) was found to predict chemosensitivity to paclitaxel in gastric cancer in vitro and in vivo. This merits further study and may help guide individual therapy [23]. The co-expression turquoise module was found to be enriched in ras signaling pathway, and contained some key genes (FGF18, FGF5, FGF14, FGF9, EFNA3, GNG13, FGF10, FGF13, PAK3, RASGRP2, GNG3, SHC3,

GNG4, FGF1, FGF2, SHC4, GRIN2A, IGF1, HGF, HTR7, NGFR, GNB3, PLA2G3, PLA2G5, PLA2G4E, PLA2G4D, NGF). Ras pathway was closely related with oncogenesis, development of cancer. Some of the identified genes in Ras pathway have been reported. Hepatocyte growth factor (HGF)-induced mesenchymal-epithelial transition factor activation leads to insulin-like growth factor 1 receptor inhibitor unresponsiveness in gastric cancer cells [24]. IGF-1 treatment significantly promoted the proliferation and invasion of SGC-7901 cells, which was inhibited following NT157, piceatannol or si-STAT3 treatment. The present study demonstrated that IGF-1-induced MMP-11 may have facilitated the proliferation and invasion of SGC-7901 cells via the JAK1/STAT3 pathway [25]. The co-expression turquoise module was found to be enriched in cell adhesion molecules (CAMs), and contained some key genes (CADM3, NRXN2, NRXN3, CLDN6, NFASC, NLGN1, NTNG1, CLDN10, L1CAM, NLGN3, CDH2, LRRC4C, NRXN1, CLDN11, NCAM1, NCAM2, ITGA8, NLGN4X, CNTN2, CNTN1, NEGR1). The research on cell adhesion molecules is fast becoming a hot topic in the field of cancer. Some of the identified genes in CAMs pathway have been reported. Neo-expression of N-cadherin (CDH2) in cancer cells is regarded as a significant event in tumor progression via epithelial-mesenchymal transition (EMT). N-cadherin expression significantly affected patient's survival in gastric cancer [26]. NCAM1 gene codes protein CD56. The number of CD3 (-) CD56 (+) and CD3 (+) CD56 (+) cells in GC patients was decreased in metastatic livers compared to those unaffected by

metastases. Therefore the prevalence of tumor-infiltrating lymphocytes seems to be related to the progression of metastatic liver disease and the escape of metastatic cells from the mechanisms of liver immune control [27]. Some of our results were consistent with previous published data, which indicated the new identified biomarkers were reliable. But meanwhile, we also found something new that were not reported before.

The co-expression green module was found to be enriched in Drug metabolism pathway, including drug metabolism-cytochrome P450 (CYP3A4, UGT1A7, GSTA2, FMO5, UGT1A8, ADH4, ADH6, UGT2A3, UGT1A1), metabolism of xenobiotics by cytochrome P450 (CYP3A4, UGT1A7, GSTA2, UGT1A8, SULT2A1, ADH4, ADH6, UGT2A3, UGT1A1), drug metabolism-other enzymes (CYP3A4, UGT1A7, UGT1A8, NAT2, UGT2A3, UGT1A1). Those results provided some drug resistance genes in DGC patients. Genetic polymorphisms of drug-metabolizing enzymes have recently been shown to affect susceptibility to chemical carcinogenesis. However, the molecular mechanisms of individual susceptibility to gastric cancer have not been fully understood [28].

A total of 44 hub genes were identified in the interest co-expression modules, including PYGM, ANGPTL1, PDZD4, CAND2, SYPL2, GNAO1, RNF150, NECAB1, SALL2, HSPB6, PEG3, RERG, HSPB8, MAP6, GPRASP1, LRCH2, ZFP28, ANK2, NCRNA00185, TMSB4Y, BCORL2, XIST, TTTY14, CYorf15B, PRKY, USP9Y, EIF1AY, TTTY15, NLGN4Y, UTY, CYorf15A, KDM5D, ZFY,

DDX3Y, RPS4Y1, MEP1A, GUCY2C, CHST5, ALDOB, NR1I2, and C14orf176. Some of the identified hub genes have been reported as biomarkers in GC. For example, ectopic expression of guanylyl cyclase C (guanylyl cyclase C) was a potential biomarker and therapeutic target in gastric cancer. The over-expression of GCC has been detected in intestinal type GC. GCC siRNA can effectively inhibit the proliferation and invasion of SGC-7901 cells and induce cell apoptosis [29]. Additionally, overall survival analysis showed that some especially important hub genes significant related to survival rate in DGC patients, such as ANK2, MAP6, GNAO1, and RERG in the co-expression turquoise module, and ALDOB, MEP1A, GUCY2C in the co-expression green module. Those new identified overall survival related hub genes might be developed into new biomarkers for DGC.

Furthermore, long noncoding RNAs (lncRNAs) are a cluster of noncoding transcripts, which play key roles in various biological processes in cancer [30]. We found that some lncRNAs in the co-expression magenta module were also related to DGC survival, including XIST, TTTY15, NCRNA00185, and CYorf15B. The result might provide new mechanism for DGC.

In conclusion, this study we used WGCNA to identify the co-expression modules were significantly correlated with five clinic traits. These obtained genes from the co-expression modules enriched in various pathways and cell functions were closely related with the risk factors and

development progress of DGC.

### **Authors' contributions**

Zaisheng Ye, Miao Zheng and Shenghong Wei: Acquisition of data and drafting of the article; Zaisheng Ye, Yunqing Xie and Qihong Zheng: participated in data analysis; Yi Zeng, Yi Wang, Shu Chen, Zhiwei Wang and Zhitao Lin: Analysis and interpretation of the data and statistical expertise; Zaisheng Ye and Luchuan Chen: Conception and design of the study. All of authors reviewed and approved the manuscript.

### **Funding**

This research was sponsored by the National Clinical Key Specialty Construction Program of China and grants from the National Science Foundation Project of Fujian Science and Technology Department (No. 2017J01264 and No. 2018Y0015), Foundation for Fujian Provincial Health Technology Project (No. 2019-ZQN-16, 2019-CXB-9, 2019006) and Startup Fund for scientific research, Fujian Medical University (No. 2017Q1219, 2017Q1220).

### **Consent for Publication**

Not applicable.

### **Conflict-of-interest statement:**

The authors declare no conflicts of interest regarding this manuscript.

### **Ethical Approval**

Not applicable.

### **Data sharing statement:**

No additional data are available.

### **References**

1. Schiemer, M., et al., *Gastric cancer and paraneoplastic dermatomyositis as complications of an unrecognized juvenile polyposis syndrome*. *Z Gastroenterol*, 2019.
2. Korivi, B.R., et al., *Intestinal and diffuse gastric cancer: a retrospective study comparing*

- primary sites*. Clin Imaging, 2019. **56**: p. 33-40.
3. Choi, J.H., et al., *Identification of genomic aberrations associated with lymph node metastasis in diffuse-type gastric cancer*. 2018. **50**(4): p. 6.
  4. Wu, D., et al., *Serum biomarker panels for the diagnosis of gastric cancer*. 2019.
  5. Deng, M., et al., *Web-TCGA: an online platform for integrated analysis of molecular cancer data sets*. BMC Bioinformatics, 2016. **17**: p. 72.
  6. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis*. BMC Bioinformatics, 2008. **9**: p. 559.
  7. Sahu, J., et al., *Revealing shared differential co-expression profiles in rice infected by virus from reoviridae and sequiviridae group*. Gene, 2019. **698**: p. 82-91.
  8. Wang, C.H., et al., *Identification of Key lncRNAs Associated With Atherosclerosis Progression Based on Public Datasets*. Front Genet, 2019. **10**: p. 123.
  9. Chen, J., et al., *Candidate genes in gastric cancer identified by constructing a weighted gene co-expression network*. PeerJ, 2018. **6**: p. e4692.
  10. Foroughi, K. and M. Amini, *Tissue-Specific Down-Regulation of the Long Non-Coding RNAs PCAT18 and LINC01133 in Gastric Cancer Development*. 2018. **19**(12).
  11. Bian, Y., et al., *Identification of Key Genes and Pathways in Post-traumatic Stress Disorder Using Microarray Analysis*. Front Psychol, 2019. **10**: p. 302.
  12. Pchitskaya, E., E. Popugaeva, and I. Bezprozvanny, *Calcium signaling and molecular mechanisms underlying neurodegenerative diseases*. Cell Calcium, 2018. **70**: p. 87-94.
  13. Liu, W. and J. Li, *Theaflavin-3, 3'-Digallate Attenuates Rheumatoid Inflammation in Mice Through the Nuclear Factor-kappaB and MAPK Pathways*. Arch Immunol Ther Exp (Warsz), 2019.
  14. Xie, G., et al., *RDM1 promotes neuroblastoma growth through the RAS-Raf-MEK-ERK pathway*. FEBS Open Bio, 2019. **9**(3): p. 490-497.
  15. Tai, Y., et al., *Axo-axonic Innervation of Neocortical Pyramidal Neurons by GABAergic Chandelier Cells Requires AnkyrinG-Associated L1CAM*. Neuron, 2019.
  16. Kimoto, E., et al., *Mechanistic Evaluation of the Complex Drug-Drug Interactions of Maraviroc: Contribution of Cytochrome P450 3A, P-Glycoprotein and Organic Anion Transporting Polypeptide 1B1*. 2019.
  17. Deng, F., et al., *Tumor-secreted dickkopf2 accelerates aerobic glycolysis and promotes angiogenesis in colorectal cancer*. Theranostics, 2019. **9**(4): p. 1001-1014.
  18. Suh, Y.S. and H.K. Yang, *Screening and Early Detection of Gastric Cancer: East Versus West*. Surg Clin North Am, 2015. **95**(5): p. 1053-66.
  19. Lee, J.H., et al., *Lauren Histologic Type Is the Most Important Factor Associated With Pattern of Recurrence Following Resection of Gastric Adenocarcinoma*. Ann Surg, 2018. **267**(1): p. 105-113.
  20. Liang, J., et al., *[Clinical study on lymph node metastasis regularity in 1456 patients with gastric cancer]*. Zhonghua Wei Chang Wai Ke Za Zhi, 2018. **21**(10): p. 1154-1160.
  21. Zhang, W., et al., *Expression of hepatocyte growth factor and basic fibroblast growth factor as prognostic indicators in gastric cancer*. Anat Rec (Hoboken), 2009. **292**(8): p. 1114-21.
  22. Tanaka, K., et al., *Tropomyosin-related receptor kinase B at the invasive front and tumour cell dedifferentiation in gastric cancer*. Br J Cancer, 2014. **110**(12): p. 2923-34.
  23. Yu, J., et al., *Combination of microtubule associated protein-tau and beta-tubulin III predicts*

- chemosensitivity of paclitaxel in patients with advanced gastric cancer. Eur J Cancer, 2014. 50(13): p. 2328-35.*
24. Liu, R., et al., *Hepatocyte growth factor-induced mesenchymal-epithelial transition factor activation leads to insulin-like growth factor 1 receptor inhibitor unresponsiveness in gastric cancer cells. Oncol Lett, 2018. 16(5): p. 5983-5991.*
  25. Su, C., W. Wang, and C. Wang, *IGF-1-induced MMP-11 expression promotes the proliferation and invasion of gastric cancer cells through the JAK1/STAT3 signaling pathway. Oncol Lett, 2018. 15(5): p. 7000-7006.*
  26. Kamikihara, T., et al., *Clinical implications of N-cadherin expression in gastric cancer. Pathol Int, 2012. 62(3): p. 161-6.*
  27. Gulubova, M., et al., *Decrease in intrahepatic CD56+ lymphocytes in gastric and colorectal cancer patients with liver metastases. Apmis, 2009. 117(12): p. 870-9.*
  28. Suzuki, S., et al., *Relationship between genetic polymorphisms of drug-metabolizing enzymes (CYP1A1, CYP2E1, GSTM1, and NAT2), drinking habits, histological subtypes, and p53 gene point mutations in Japanese patients with gastric cancer. J Gastroenterol, 2004. 39(3): p. 220-30.*
  29. Mao, Z.B., et al., *Ectopic expression of guanylyl cyclase C in gastric cancer as a potential biomarker and therapeutic target. J Dig Dis, 2009. 10(4): p. 272-85.*
  30. Bhan, A., M. Soleimani, and S.S. Mandal, *Long Noncoding RNA and Cancer: A New Paradigm. Cancer Res, 2017. 77(15): p. 3965-3981.*

## Figure legends:

**Figure 1. Sample cluster analysis based on RNA data from TCGA database. A.** Sample clustering to detect outliers based on RNA data. **B.** Sample dendrogram and trait heatmap based on gene expression data and clinical data. a.age at initial pathologic diagnosis, b.gender, c.longest dimension, d.lymph node examined count, e.neoplasm histologic grade, f.new tumor event after initial treatment, g.number of lymph nodes positive, h.pathologic M stage, i.pathologic N stage, j.pathologic T stage, k.pathologic stage, l.person neoplasm cancer status, m.radiation therapy, n.targeted molecular therapy.

**Figure 2. Construction of soft-threshold powers and  $R^2$  of DGC. A.**

Analysis of network topology for various soft-threshold powers. **B.** Check scale-free topology, and here the adjacency matrix was defined using soft-thresholds with  $\beta=3$ .

**Figure 3. Construction of clustering dendrogram**

**A.** Clustering dendrograms of genes, with dissimilarity based on topological overlap, together with assigned module colors. **B.** The heatmap depicts the topological overlap matrix (TOM) among genes based on co-expression modules.

**Figure 4. Analysis of module-trait relationships of DGC based on TCGA data.**

Each row corresponds to a module eigengene, and column to a trait. a.age at initial pathologic diagnosis, b.gender, c.longest dimension, d.lymph node examined count, e.neoplasm histologic grade, f.new tumor event after initial treatment, g.number of lymph nodes positive, h.pathologic M stage, i.pathologic N stage, j.pathologic T stage, k.pathologic stage, l.person neoplasm cancer status, m.radiation therapy, n.targeted molecular therapy.

**Figure 5. The Construction of eigengene dendrogram and heatmap.**

**A.** the dendrogram indicates that turquoise module is highly related to longest dimension. **B.** the dendrogram indicates that magenta module is highly related to gender and person neoplasm cancer status. **C.** the dendrogram indicates that green module is highly related to pathologic N stage and pathologic stage.

**Figure 6. The scatterplot of Gene Significance (GS) for clinical traits vs. Module Membership (MM).**

**A.** A scatterplot of Gene Significance (GS) for

longest dimension vs. Module Membership (MM) in the turquoise module. **B.** A scatterplot of Gene Significance (GS) for gender vs. Module Membership (MM) in the magenta module. A scatterplot of Gene Significance (GS) for person neoplasm cancer status vs. Module Membership (MM) in the magenta module. **C.** A scatterplot of Gene Significance (GS) for pathologic N stage vs. Module Membership (MM) in the green module. A scatterplot of Gene Significance (GS) for pathologic stage vs. Module Membership (MM) in the green module.

**Figure 7. Biological processes analysis and PPI of hub genes. A.**

Biological processes of hub genes involved in the interest co-expression modules. **B.** PPI network of hub genes involved in the co-expression interest module.

**Figure 8. Analysis of overall survival-related hub genes. A.**

overall survival-related hub genes were identified in the co-expression turquoise module, including ANK2, MAP6, GNAO1, and RERG. **B.** overall survival-related hub genes were identified in the co-expression green module, including ALDOB, MEP1A, GUCY2C. **C.** overall survival-related hub genes were identified in the co-expression magenta module, including XIST, TTTY15, NCRNA00185, and CYorf15B.

# Figures

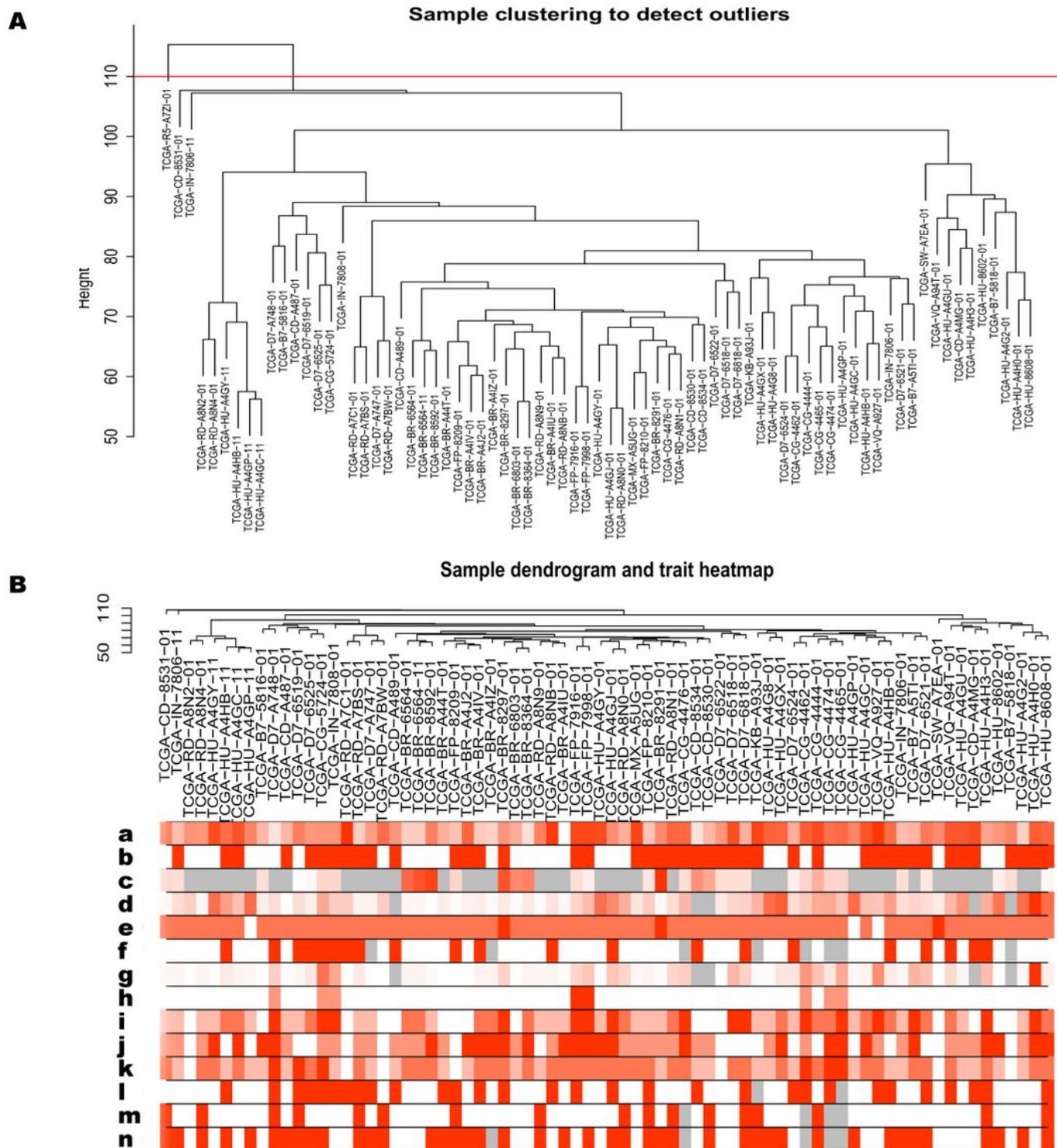
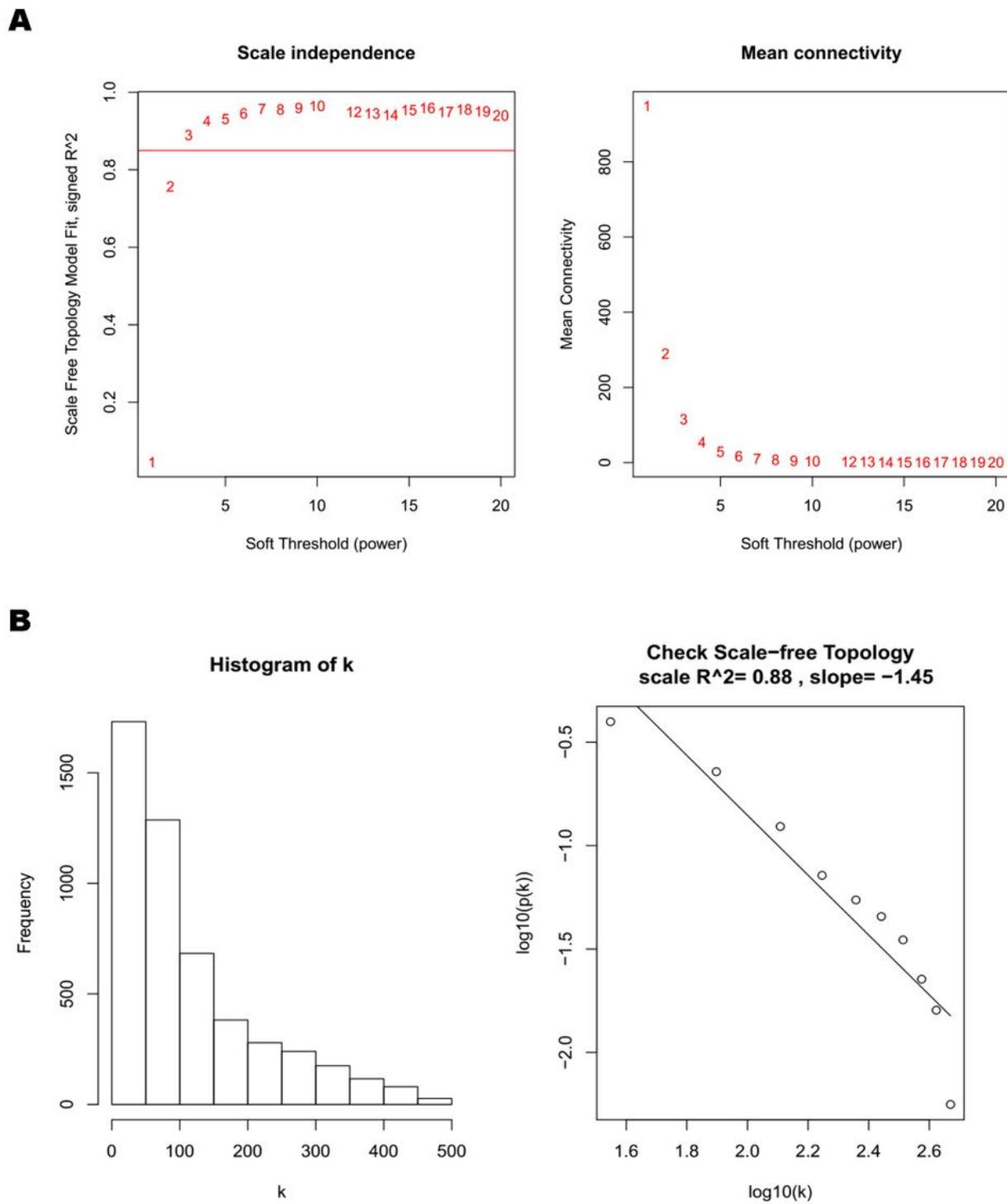


Figure 1

Sample cluster analysis based on RNA data from TCGA database. A. Sample clustering to detect outliers based on RNA data. B. Sample dendrogram and trait heatmap based on gene expression data and clinical data. a.age at initial pathologic diagnosis, b.gender, c.longest dimension, d.lymph node examined

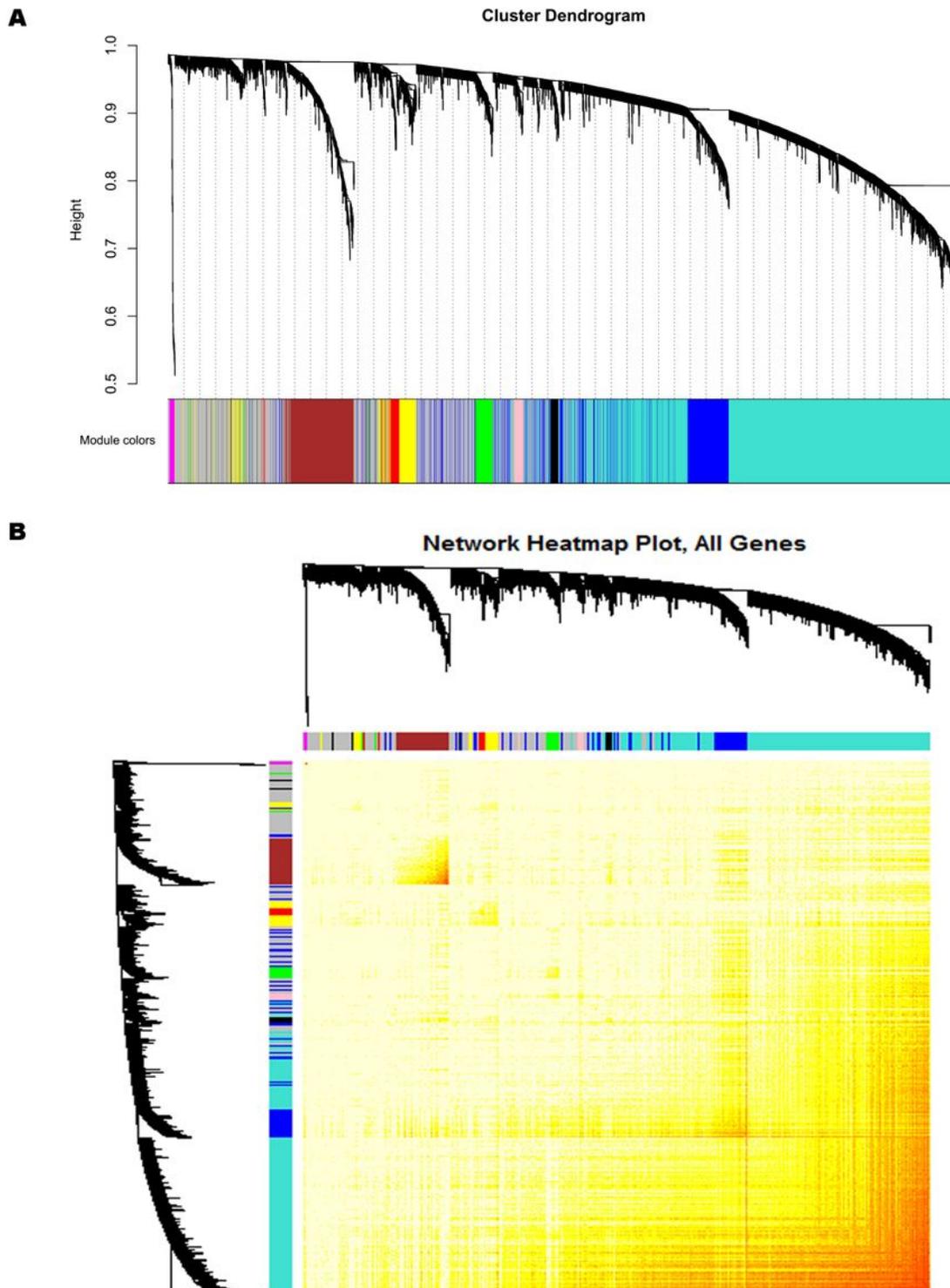
count, e.neoplasm histologic grade, f.new tumor event after initial treatment, g.number of lymph nodes positive, h.pathologic M stage, i.pathologic N stage, j.pathologic T stage, k.pathologic stage, l.person neoplasm cancer status, m.radiation therapy, n.targeted molecular therapy.



**Figure 2**

Construction of soft-threshold powers and  $R^2$  of DGC. A. Analysis of network topology for various soft-threshold powers. B. Check scale-free topology, and here the adjacency matrix was defined using soft-

thresholds with  $\beta=3$ .



**Figure 3**

Construction of clustering dendrogram A. Clustering dendrograms of genes, with dissimilarity based on topological overlap, together with assigned module colors. B. The heatmap depicts the topological overlap matrix (TOM) among genes based on co-expression modules.

## Module-trait relationships

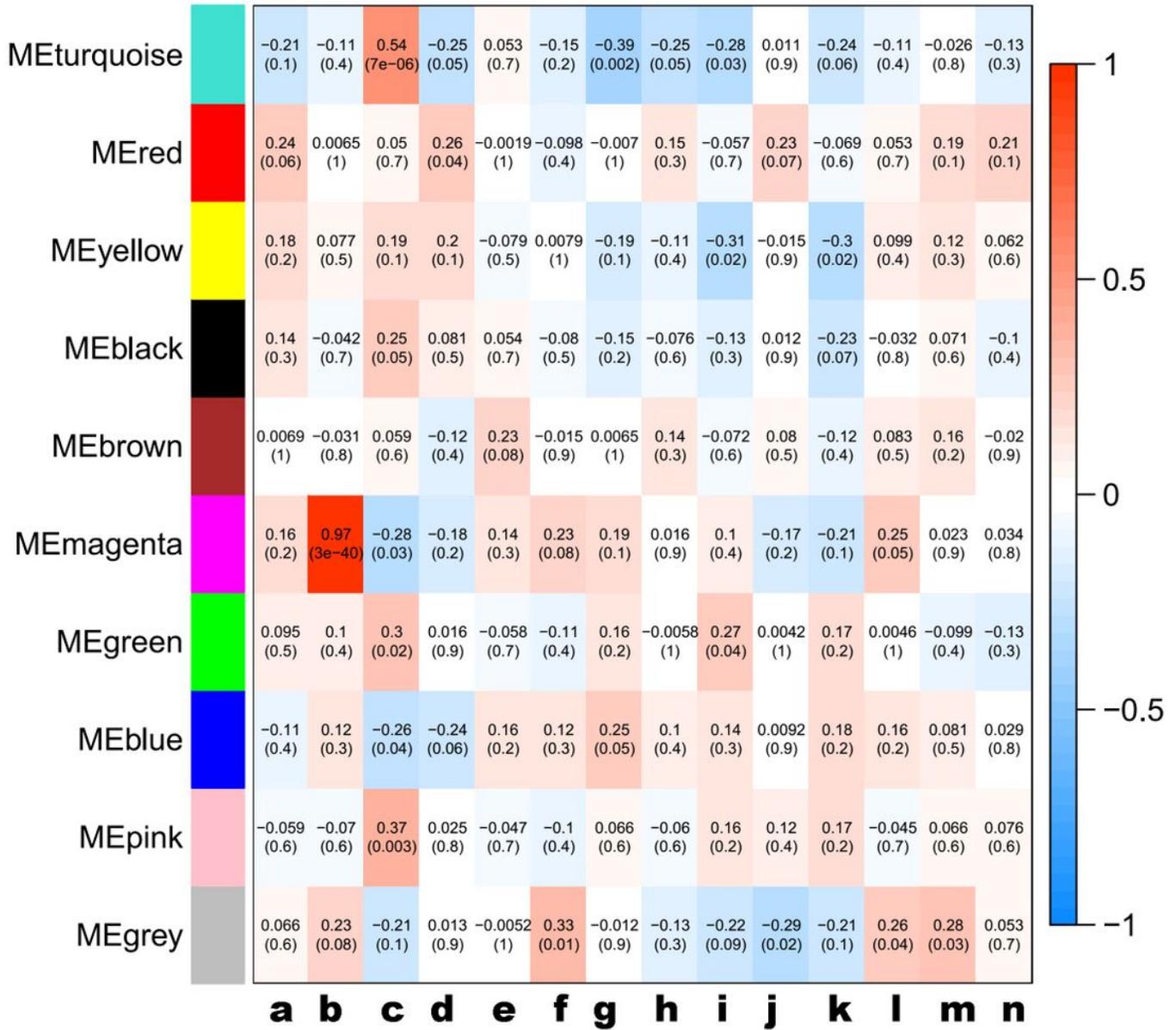
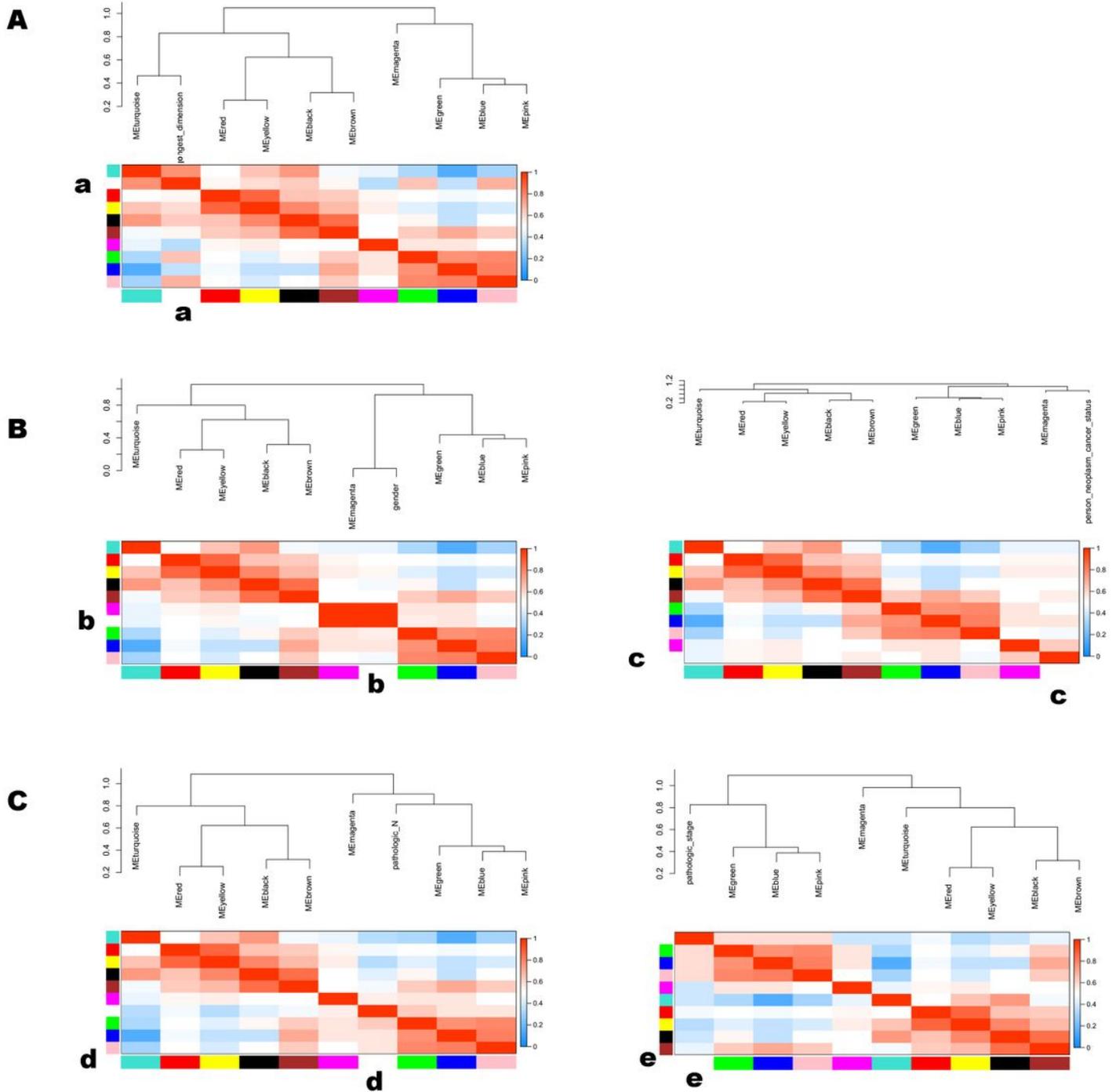


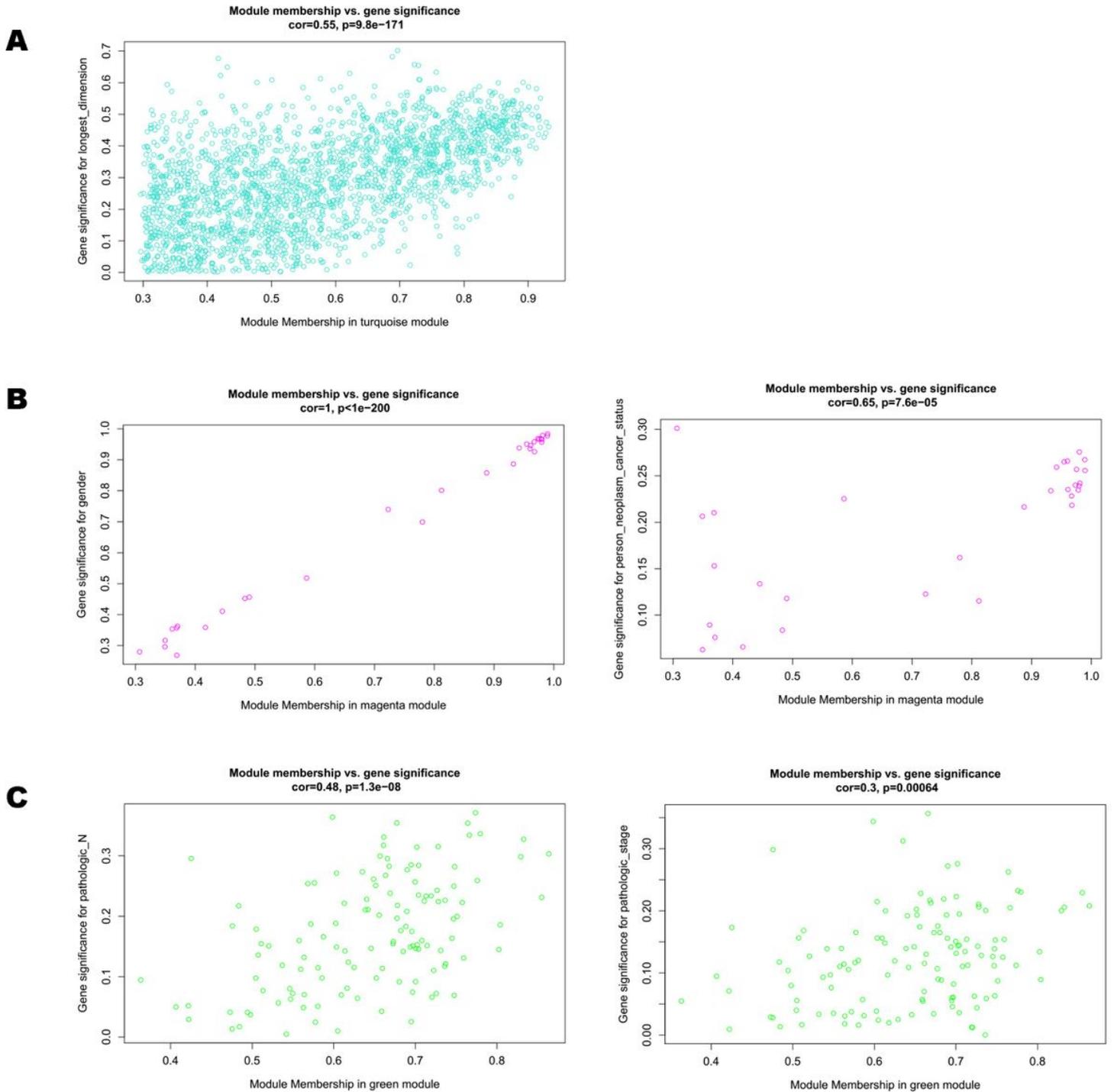
Figure 4

Analysis of module-trait relationships of DGC based on TCGA data. Each row corresponds to a module eigengene, and column to a trait. a.age at initial pathologic diagnosis, b.gender, c.longest dimension, d.lymph node examined count, e.neoplasm histologic grade, f.new tumor event after initial treatment, g.number of lymph nodes positive, h.pathologic M stage, i.pathologic N stage, j.pathologic T stage, k.pathologic stage, l.person neoplasm cancer status, m.radiation therapy, n.targeted molecular therapy.



**Figure 5**

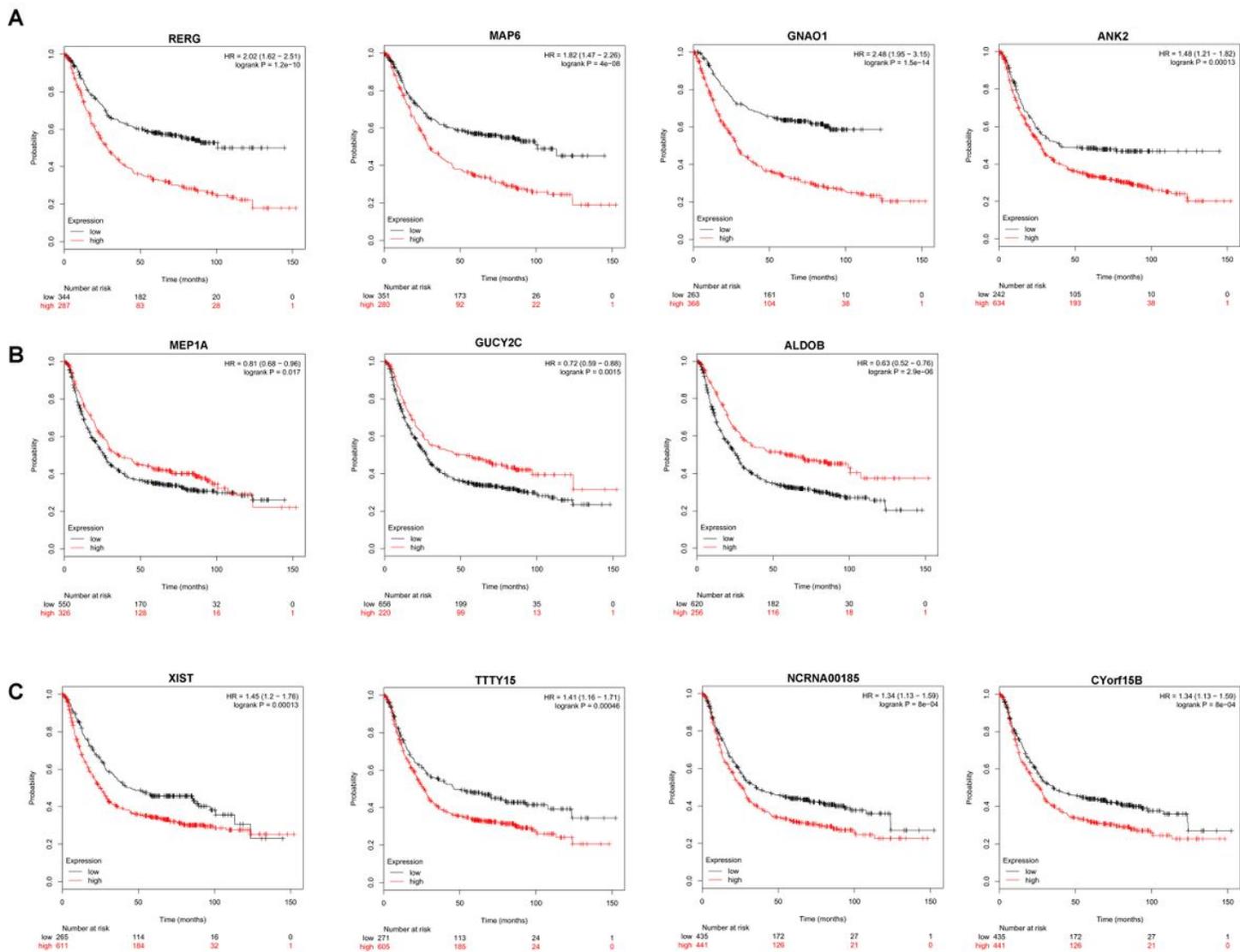
The Construction of eigengene dendrogram and heatmap. A. the dendrogram indicates that turquoise module is highly related to longest dimension. B. the dendrogram indicates that magenta module is highly related to gender and person neoplasm cancer status. C. the dendrogram indicates that green module is highly related to pathologic N stage and pathologic stage.



**Figure 6**

The scatterplot of Gene Significance (GS) for clinical traits vs. Module Membership (MM). A. A scatterplot of Gene Significance (GS) for longest dimension vs. Module Membership (MM) in the turquoise module. B. A scatterplot of Gene Significance (GS) for gender vs. Module Membership (MM) in the magenta module. A scatterplot of Gene Significance (GS) for person neoplasm cancer status vs. Module Membership (MM) in the magenta module. C. A scatterplot of Gene Significance (GS) for pathologic N





**Figure 8**

Analysis of overall survival-related hub genes. A. overall survival-related hub genes were identified in the co-expression turquoise module, including ANK2, MAP6, GNAO1, and RERG. B. overall survival-related hub genes were identified in the co-expression green module, including ALDOB, MEP1A, GUCY2C. C. overall survival-related hub genes were identified in the co-expression magenta module, including XIST, TTTY15, NCRNA00185, and CYorf15B.

## Supplementary Files

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

- [Table1.xlsx](#)
- [Supplementarytable1.xlsx](#)
- [Supplementarytable2.xlsx](#)

- [Supplementarytable3.xlsx](#)
- [Supplementarytable4.xlsx](#)
- [Supplementarytable5.xlsx](#)
- [Supplementarytable6.xlsx](#)
- [Supplementarytable7.xls](#)
- [Supplementarytable8.xlsx](#)