

# The Oxidative Stress Pathway May be an Important Pathway Leading to The Recurrence of Ewing Sarcoma — Based On Weighted Gene Co-Expression Network Analysis

**Jianfeng Li**

Zhuhai City People's Hospital

**Shaoyu Hu**

Zhuhai City People's Hospital

**Song Hao**

Zhuhai City People's Hospital

**Shengjia Huang**

Zhuhai City People's Hospital

**Yi Qin**

Zhuhai City People's Hospital

**Jianjun Wang** (✉ [spinewang7@163.com](mailto:spinewang7@163.com))

Zhuhai City People's Hospital

---

## Research article

**Keywords:** The Oxidative Stress Pathway, Ewing sarcoma, Differentially Expressed Genes, Recurrence

**Posted Date:** April 9th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-21707/v1>

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

[Read Full License](#)

---

# Abstract

**Background** The role of gene and pathway in recurrence of Ewing sarcoma (ES) was not clear. Thus, we investigated the biological role and underlying mechanism of gene and pathway in recurrence of ES.

**Methods** Data sets of patients with ES were collected from the GEO database. We used dataset GSE63155 and GSE63156 to construct co-expression networks by weighted gene co-expression network analysis (WGCNA). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed by Database for Annotation, Visualization and Integrated Discovery (DAVID).

**Results** We can find that genes with significant interactions in the genes of the recurrence group include SRSF11, TRIM39, SOCS3, NUPL2, COPS5. They work primarily through the oxidative stress pathway.

**Conclusion** Through our research, for the first time found that ES by SRSF11 TRIM39, SOCS3, NUPL2, COPS5 interaction, activation of phosphorylation of bone and oxidative stress is affecting tumor recurrence.

## Introduction

Ewing sarcoma (ES) is a rare malignant bone and soft tissue sarcoma. It is the second most common primary malignant bone tumor. It has a very high mortality rate and is more common in adolescents and young people[1]. In fact, ES exhibits a large degree of undifferentiated and "dry" phenotype, contributing to its clinical invasiveness[2]. Among ES patients, 60-75% of patients benefit from multimodal treatment. Therefore more than 30% of patients have limited response to treatment, which is usually first based on evaluating their histological response after neoadjuvant chemotherapy[3]. Therefore, it is particularly important to find the target sensitive to ES drugs for the relapsing type ES. In combination with the latest advances in molecular biology, it is found that the occurrence and development of ES is related to the activation of oncogenes such as SOX2, EGR1 and ARID1A and the inactivation of tumor suppressor genes such as p53[4]. However, the occurrence of tumor is a complex and gradual process with multiple factors, multiple stages and multiple steps. The molecular pathogenesis of relapsed ES is still unclear, preventing the recurrence of ES and improving the survival rate of patients is the key. In order to judge relapsed ES early and improve its therapeutic effect, it is necessary to better understand the pathogenesis of relapsed ES. Weighted gene co-expression network analysis (WGCNA) is a new tool for analyzing potential gene modules in gene expression data[5]. The analysis of the core genes of recurrent and non-recurrent ES by WGCNA can bring great convenience to our clinical drug development. To this end, we searched for the core genes and pathways of recurrent type and non-recurrent type ES through the inclusion of two queues and the method of WGCNA.

## Patients And Methods

### Data information

ES from NCBI gene expression data sets Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>), queue GSE63155 and GSE63156 were included in the two groups. GSE63155 consists of 28 primary and 18 recurrent tumors on the platform of [huex-1\_0-st] Affymetrix Human Exon 1.0st Array [transcript (gene) version]. GSE63156 consists of 23 primary and 16 recurrent tumors on the platform of [huex-1\_0-st] Affymetrix Human Exon 1.0st Array [transcript (gene) version]. The affy and annotated R packets were used to process the raw data, generate expression matrices, and match the probes with their genetic symbols.

## **Construction of WGCNA**

The raw data for GSE63155 and GSE63156 was downloaded from the GEO database. We preprocessed and normalized the raw data using the affy package (R environment, version 3.4.3). The parameters were set to RMA (background correction) and impute(supplementary missing value). We selected the first 5,000 genes of WGCNA by sequencing the SD from large to small (including primary and recurrent tumors). The R package was used for WGCNA and the pick Soft Threshold function was used to predict the power parameters. The scale-free topological fitting index of multiple efficiencies was calculated to provide an appropriate soft threshold efficiency for protein network construction. The adjacency is defined as the sum of the adjacency of a gene and all other genes, and the network generation is carried out to measure the network connectivity of the gene. Hierarchical clustering function was used to divide genes with similar expression profiles into modules according to TOM differences with minimum size of 50. By calculating the differences between the characteristic genes of each module, a dividing line was selected to merge each module. 1000 genes were randomly selected and the gene network was visualized. Finally, the gene network of feature genes is visualized.

## **Identification of clinical significant modules**

We identified the relevant modules by calculating the correlation between MEs and recurrence. Subsequently, gene significance (GS) was defined as  $\log_{10}$  conversion of p value ( $GS = \lg P$ ) in linear regression of gene expression and clinical information. Module significance (MS) is defined as the average GS of all genes in a module. In general, of all the selected modules, the module with the highest absolute value of MS was considered to be related to recurrence.

## **Identification of hub genes in key modules**

The core gene expression in ES was verified on two independent ES tissue data sets (GSE63155 and GSE63156). By looking for the common core genes of primary ES and relapsing ES affecting different tissue datasets, we included the WGCNA core genes of the two datasets to find the common genes, which helped us to find the most important genes.

## **Differentially expressed genes functional enrichment and pathway analysis**

The DAVID database was used for GO analysis of the identified differentially expressed genes. DAVID provides comprehensive and systematic annotation tools for elucidating the biological significance of

genes. GO analysis includes molecular function analysis, biological process analysis and cell component analysis [6]. The difference genes detected by DESeq2 software were used as the enrichment foreground, and the reference genes in human database were used as the enrichment analysis background.  $P < 0.05$  was selected as the significantly enriched GO pathway, and the number of different genes in each significantly enriched GO and pathway was counted. We plot through the Gplot package of R language and achieve visual results [7].

### **Survival analysis of DEGs**

To assess the prognostic value of a particular gene, patient samples were divided into two groups based on the median expression of the gene (high expression vs low expression). Kaplan-meier survival map was used to analyze the overall survival (OS) of patients with ES, and the core genes were respectively passed through R language to obtain the kaplan-meier survival map.

### **Hub gene analysis and identification**

The genes in the co-expression module were uploaded to the search tool to retrieve the interacting genes/proteins (STRING) online database (version 11.0) to evaluate PPI information and construct the functional protein association network. Interactions with a composite score above 0.9 were considered significant. STRING was used to detect the PPI central gene, and the connectivity was set to the top 5%. The weight network of each common expression module is weighted by STRING. The PPI central gene in the top weighted network is considered to be the true central gene [8].

### **Statistical Analyses**

Pearsons 2 test was used to investigate the differences in the clinical and pathological parameters of each group. Overall survival rate (OS) was calculated by survival analysis. OS was defined as the time from the initial diagnosis of ES to death from any cause. The survival rate was statistically analyzed by kaplan-meier, and the difference of survival curve was statistically analyzed by log-rank test. Spearman's rank correlation (rs) was used to investigate the relationship between relapsed and non-relapsed groups. All statistical analyses were performed using IBM SPSS 20.0 (IBM Corp., Armonk, NY, USA) and R version 3.3.0 (The R Foundation). P value was bilateral, and  $P < 0.05$  was considered statistically significant.

## **Results**

### **Construction of weighted co-expression network and identification of key modules**

To construct a gene co-expression network, the original data of ES (GSE63155 and GSE63156) were downloaded from the GEO database. R is used for background correction and normalization, and the original data is preprocessed in the same way. The probe was matched with the gene symbol by r-packet annotation, and the probe matching multiple genes was removed. For the gene matched by multiple probes, the median was taken as the final expression value. In the end, we had a total of 24,991 genes. We calculated the SD values of each gene, sequenced them from large to small, and selected the first

5,000 genes as WGCNA. The function of fashClust in the WGCNA package was used for cluster analysis of 5000 genes (Figure 1A and Figure 2A). The selection of soft threshold power is an important step in the construction of WGCNA. The network topology of 1 ~ 20 threshold weights was analyzed, and the scale independence and average connectivity of WGCNA were determined. As shown in Figure 1B, C and Figure 2B, C, the power value 9 was selected as the lowest power (0.9) of the scaling free topological fit index to generate a hierarchical clustering tree (dendrogram) of 5,000 genes. We set the MEDiss Thres to 0.25 to merge similar modules (Figure 3A and Figure 4A) and generated 65 modules (Figure 5A and Figure 6A). The gene statistics in each module are shown in Table1 and 2. Genes that cannot be included in any module are put into the grey module and removed in subsequent analysis.

### **Correlation between modules and identification of key modules**

We analyzed the interaction between 65 modules and drew a network heat map (FIG. 5A and 6A). The results show that the modules are independent from each other, and the modules have high independence from each other, and the gene expression of each module is relatively independent. In addition, we calculated the feature genes and clustered them according to their correlations to explore the co-expression similarity of all the modules (FIG. 5B and 6B). We found that the 65 modules are mainly divided into two clusters. A heat map based on the adjacency relationship shows similar results. In GSE63155, Blue module, black module and recurrence group are positively correlated, while brown module and Blue module are negatively correlated with non-recurrence group (figure 7). In GSE63156, Brown module, turquoise module and recurrence group are positively correlated, while blue module and non-recurrence group are negatively correlated (figure 8).

### **Look for common core genes**

The core differentially expressed genes were verified and the core genes were found in two independent ES data sets (GSE63155, GSE63156). By analyzing the genes that play a key role in recurrence in GSE63155, we can find that Blue module and black module are positively correlated in the recurrence group through WGCNA, while GSE63156 is Brown module and turquoise module. Euler diagram shows that they share 12 genes (see figure 9). Second, through WGCNA, we can find that brown module and Blue module are negatively correlated in the non-recurrence group, while GSE63156 is Blue module. By Euler diagram, we found that they Shared 26 genes (see figure 10). So we have 38 genes that are critical to the recurrence group.

### **Function enrichment analysis in two key modules**

In order to study the role of core genes, we conducted enrichment analysis and explored the pathways involved in BP and these two key modules. GO enrichment of BP by DAVID shows that, Recurrence of core set of genes is mainly enriched in regulation of phosphorus metabolic process, regulation of phosphate metabolic process, regulation of phosphorylation, the response to hypoxia, the response to oxygen levels, Regulation of protein amino acid phosphorylation (FIG. 11,12,13,).

## Survival analysis of Core gene

Genes with significant interactions in the recurrence group included PSEN2, GLT1D1, MAP1A, RBM43, FDF1, SOCS3, RPL5, SRSF11, COPS5, SMG7, TSPAN2, CYP2C19, AAMP, GPR137B, TGFBR1, HEBP2, C2orf57, GRM1, CDK2AP2, WDR5B, NUPL2, BPIL1, CLDN11, TNFRSF1A, ADIPOQ, TRIB2, FOXN1, CRYM, ELMO2, PENK, TPK1, FN3KRP, CINP, KCNK3, LDHA, TFPI, POU2F2, FGF17. In order to further study the influence of core genes on the prognosis of patients, we studied the influence of core genes on the prognosis by grouping the expression amount. Relapse group at the core of the correlated with prognosis of genes and gene TGFBR1, TFPI, SRSF11, SOCS3, RPL5, RBM43, POU2F2, NUPL2, MAP1A, GRM1, GLT1D1, FOXN1, FN3KRP, FGF17, CYP2C19, CRYM, COPS5, ELMO2, CLDN11, C2ORF57, CINP, ADIPOQ, AAMP, 16, 17 (Fig 14, 15).

## Identification of hub genes in the Salmon module and Darkorange module

We submitted the core gene set related to prognosis to the protein interaction of STRING, with the binding confidence interval of truncation set at 0.4. In Plugin Molecular Complex Detection (MCODE), significant models with strong protein-protein connections are calculated and selected, with default parameters (degree cut  $\geq 2$ , node score cut  $\geq 2$ , k-core  $\geq 2$ , maximum depth = 100). In order to  $P < 0.05$  was considered statistically significant. The candidate genes of node degree were sequenced and the core genes were selected for further analysis. Figure 18 shows the hub genes for SRSF11 TRIM39, SOCS3, NUPL2, COPS5.

## Discussion

ES (ES) is a bone and soft tissue tumor that occurs primarily in children and young adults. At the molecular level, it is characterized by a chromosomal translocation between the EWSR1 gene (ES break point region 1) and a gene (mostly FLI1) in the E26 mutation-specific or e-26 (ETS) family that produces the drive fusion gene, with few other genomic changes. The clinical manifestations of the disease still vary from person to person. Although efforts have been made to improve treatment options and outcomes, patients in the diagnosis of metastatic or recurrent disease remain a dire prognosis [9]. The most commonly used initial chemotherapy strategies are three, depending in part on institutional preferences and patient age. Despite the widespread use of chemotherapy, at least a quarter of patients will initially relapse, with an even higher recurrence rate of early metastatic disease, with treatment failure occurring in 50-80% of patients, depending on the site of metastasis. There are few studies on the causes of recurrence of ES and most of them are based on clinically relevant studies. Therefore, we urgently need to master the molecular mechanism of recurrence of ES, so as to facilitate our treatment and diagnosis. Therefore, we first through the analysis of GEO data sets, by incorporating two independent ES (GSE63155 GSE63156) data sets on core differentially expressed genes are verified and looking for the core found SRSF11, TRIM39, SOCS3, NUPL2, COPS5 by influencing the phosphorylation of bone and oxidative stress path is the main factor affecting the recurrence of ES.

Oxidative stress is an important factor affecting tumors. Reactive oxygen species (ROS) and reactive nitrogen (RNS) are highly reactive by-products of energy produced during physiological processes (such

as metabolism). ROS are prone to interact with lipids, proteins and DNA, leading to oxidative damage, which, if unchecked, may lead to the occurrence and development of a variety of pathologies. Levels of ROS increase in cancer cells compared to normal cells. Although ROS production and removal responses are essentially the same in cancer cells and normal cells, increased ROS production and decreased detoxification can lead to oxidative stress and loss of REDOX control [10]. ROS play an important role in oncogenic signaling, and elevated ROS levels are an important feature of many highly invasive cancers, possibly including ewing tumors [11]. There is strong evidence that many malignancies have a permanently active "oxidative stress phenotype" that leads to increased invasion, which is considered another marker of cancer recurrence [12]. Among other effects, mitochondrial respiratory defects inactivate PTEN in a redox-dependent manner, leading to activation of the PI3K/Akt pathway for survival [13]. ROS dependent inactivation of PTEN (through the oxidation of bound cysteine) and activation of the PI3K/Akt pathway highlights the role of ROS as a second receptor that regulates intracellular signal transduction to determine cell fate. ROS dependent MAP kinase cascade regulation also plays an important role in cell fate regulation through REDOX dependent regulation of ASK1 [14]. The ROS/Jab1/Trx signaling pathway may be significantly associated with the recurrence and metastasis of tumor patients [15]. In this study, oxidative stress pathway was significantly upregulated in the recurrence group, which may be closely related to tumor recurrence.

SRSF11 is a member of the SR protein family. Members of the SR protein family have an RS domain rich in serine/arginine (S/R) repeat sequence, which plays an important role in the assembly of RNA splicing and the regulation of alternative splicing. They are characterized by 1 ~ 2 RS domains and RNA binding domains, wherein the RS domain consists of serine/arginine (S/R) rich repeats. At present, it has been found that there are a large number of RS domain proteins in organisms, among which about 50 species of higher organisms participate in RNA splicing[16]. The splicing factor activity of this family is regulated by reversible phosphorylation mediated by protein kinases belonging to SRPK and CLK families and by kinases activated by different signaling pathways such as MAPK, PI3K and Akt [17]. They are involved in tumor apoptosis by activating ROS and other oxidative stress pathways through interaction, and affect recurrence and metastasis [18]. Tripartite motif 39 (TRIM39), also known as the Ring finger protein 23 (RNF23), belongs to the class with annular structure domain, B - box structure and curly structure domains for characteristics of protein family. In recent years, TRIM39 stabilized apoptosis regulator 1 (moap-1) by inhibiting the APC/C cdh1-mediated polyubiquitination process[19]. Some research results confirmed the causal role of TRIM39 in regulating the process of cell cycle, and the balance between cell stagnation and apoptosis after DNA destruction by stabilizing p21, which may affect tumor progression through the regulation of cell cycle, and play a certain role in DNA damage caused by oxidative stress [20]. As a negative regulator of cytokine signal transduction, SOCS3 is involved in the regulation of various signal transduction pathways. As an important molecule of the SOCS family, a large number of studies have shown that it can be induced by a variety of inflammatory factors and anti-inflammatory factors, and inhibit the signal transduction of a variety of immune molecules. It plays an important role in the occurrence and development of inflammatory diseases, viral infections, obesity and tumors [21]. Studies have shown that SOCS3 can affect tumor proliferation and development through Src/STAT3 and

c-cbl mediated oxidative stress [22]. NUPL2 is a member of the nuclear pore gene family, which may play an important role in DNA damage caused by oxidative stress and affect tumor progression [23]. The photogenic factor complex constitutively photomorphogenic 9 signalosome subunit 5 (COPS5) was first referred to as c-jun activation domain-binding protein-I (Jabl). COPS5 is an evolutionarily conserved multifunctional protein that regulates many cellular processes, including signal transduction, cell proliferation, apoptosis, cell cycle control, DNA damage repair, and tumorigenesis[24]. SRSF11 TRIM39, SOCS3 NUPL2, COPS5 interaction, through activation of phosphorylation of bone and oxidative stress is ES recurred.

## **Conclusion**

Through our research, for the first time found that ES by SRSF11 TRIM39, SOCS3, NUPL2, COPS5 interaction, activation of phosphorylation of bone and oxidative stress is affecting tumor recurrence. This could make it easier for us to find precise treatments.

## **Declarations**

### **Acknowledgements**

None.

### **Conflicts of interest**

All authors declare that they have no conflicts of interest.

### **Author contributions**

YQ, JL, JH and SYH conceived and designed the study. SJH, JL, and SYH performed the data analysis. JL and SYH wrote the paper. All authors read and approved the manuscript.

### **Consent for publication**

Not applicable.

### **Detail of funding**

None.

### **Ethics committee**

The information of patients with ES obtained from the GEO database does not contain any identifiers and is public. Due to the retrospective nature of the study, the patient's informed consent was not required. All data base information were deidentified by United States institutional review board approval.This

analysis does not involve interaction with human subjects or the use of personally identifiable information.

## References

1. Dirksen U, Brennan B, Le Deley MC, Cozic N, van den Berg H, Bhadri V, Brichard B, Claude L, Craft A, Amler S *et al*: **High-Dose Chemotherapy Compared With Standard Chemotherapy and Lung Radiation in ESWith Pulmonary Metastases: Results of the European Ewing Tumour Working Initiative of National Groups, 99 Trial and EWING 2008**. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2019;Jco1900915.
2. Grunewald TGP, Cidre-Aranaz F, Surdez D, Tomazou EM, de Alava E, Kovar H, Sorensen PH, Delattre O, Dirksen U: **Ewing sarcoma**. *Nature reviews Disease primers* 2018, **4**(1):5.
3. Whelan J, Le Deley MC, Dirksen U, Le Teuff G, Brennan B, Gaspar N, Hawkins DS, Amler S, Bauer S, Bielack S *et al*: **High-Dose Chemotherapy and Blood Autologous Stem-Cell Rescue Compared With Standard Chemotherapy in Localized High-Risk Ewing Sarcoma: Results of Euro-E.W.I.N.G.99 and Ewing-2008**. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2018;Jco2018782516.
4. Selvanathan SP, Graham GT, Grego AR, Baker TM, Hogg JR, Simpson M, Batish M, Crompton B, Stegmaier K, Tomazou EM *et al*: **EWS-FLI1 modulated alternative splicing of ARID1A reveals novel oncogenic function through the BAF complex**. *Nucleic acids research* 2019, **47**(18):9619-9636.
5. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis**. *Statistical applications in genetics and molecular biology* 2005, **4**:Article17.
6. **The Gene Ontology (GO) project in 2006**. *Nucleic acids research* 2006, **34**(Database issue):D322-326.
7. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the unification of biology**. **The Gene Ontology Consortium**. *Nature genetics* 2000, **25**(1):25-29.
8. Guo D, Wang H, Sun L, Liu S, Du S, Qiao W, Wang W, Hou G, Zhang K, Li C *et al*: **Identification of key gene modules and hub genes of human mantle cell lymphoma by coexpression network analysis**. *PeerJ* 2020, **8**:e8843.
9. Italiano A, Mir O, Mathoulin-Pelissier S, Penel N, Piperno-Neumann S, Bompas E, Chevreau C, Duffaud F, Entz-Werle N, Saada E *et al*: **Cabozantinib in patients with advanced ESor osteosarcoma (CABONE): a multicentre, single-arm, phase 2 trial**. *The Lancet Oncology* 2020, **21**(3):446-455.
10. Smith DG, Magwere T, Burchill SA: **Oxidative stress and therapeutic opportunities: focus on the ES family of tumors**. *Expert review of anticancer therapy* 2011, **11**(2):229-249.
11. Grunewald TG, Diebold I, Esposito I, Plehm S, Hauer K, Thiel U, da Silva-Buttkus P, Neff F, Unland R, Muller-Tidow C *et al*: **STEAP1 is associated with the invasive and oxidative stress phenotype of Ewing tumors**. *Molecular cancer research : MCR* 2012, **10**(1):52-65.

12. Luo J, Solimini NL, Elledge SJ: **Principles of cancer therapy: oncogene and non-oncogene addiction.** *Cell* 2009, **136**(5):823-837.
13. Pelicano H, Xu RH, Du M, Feng L, Sasaki R, Carew JS, Hu Y, Ramdas L, Hu L, Keating MJ *et al*: **Mitochondrial respiration defects in cancer cells cause activation of Akt survival pathway through a redox-mediated mechanism.** *The Journal of cell biology* 2006, **175**(6):913-923.
14. Matsuzawa A, Ichijo H: **Redox control of cell fate by MAP kinase: physiological roles of ASK1-MAP kinase pathway in stress signaling.** *Biochimica et biophysica acta* 2008, **1780**(11):1325-1336.
15. Zhou F, Pan Y, Wei Y, Zhang R, Bai G, Shen Q, Meng S, Le XF, Andreeff M, Claret FX: **Jab1/Csn5-Thioredoxin Signaling in Relapsed Acute Monocytic Leukemia under Oxidative Stress.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2017, **23**(15):4450-4461.
16. Fu XD, Maniatis T: **Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus.** *Nature* 1990, **343**(6257):437-441.
17. da Silva MR, Moreira GA, Goncalves da Silva RA, de Almeida Alves Barbosa E, Pais Siqueira R, Teixeira RR, Almeida MR, Silva Junior A, Fietto JL, Bressan GC: **Splicing Regulators and Their Roles in Cancer Biology and Therapy.** *BioMed research international* 2015, **2015**:150514.
18. Kedzierska H, Piekielko-Witkowska A: **Splicing factors of SR and hnRNP families as regulators of apoptosis in cancer.** *Cancer letters* 2017, **396**:53-65.
19. Huang NJ, Zhang L, Tang W, Chen C, Yang CS, Kornbluth S: **The Trim39 ubiquitin ligase inhibits APC/CCdh1-mediated degradation of the Bax activator MOAP-1.** *The Journal of cell biology* 2012, **197**(3):361-367.
20. Zhang L, Mei Y, Fu NY, Guan L, Xie W, Liu HH, Yu CD, Yin Z, Yu VC, You H: **TRIM39 regulates cell cycle progression and DNA damage responses via stabilizing p21.** *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(51):20937-20942.
21. Luckey MA, Kim TH, Prakhar P, Keller HR, Crossman A, Choi S, Love PE, Walsh STR, Park JH: **SOCS3 is a suppressor of gammac cytokine signaling and constrains generation of murine Foxp3(+) regulatory T cells.** *European journal of immunology* 2020.
22. Lin L, Cao L, Liu Y, Wang K, Zhang X, Qin X, Zhao D, Hao J, Chang Y, Huang X *et al*: **B7-H3 promotes multiple myeloma cell survival and proliferation by ROS-dependent activation of Src/STAT3 and c-Cbl-mediated degradation of SOCS3.** *Leukemia* 2019, **33**(6):1475-1486.
23. Lim KS, Wong RW: **Targeting Nucleoporin POM121-Importin beta Axis in Prostate Cancer.** *Cell chemical biology* 2018, **25**(9):1056-1058.
24. Li P, Gao L, Cui T, Zhang W, Zhao Z, Chen L: **Cops5 safeguards genomic stability of embryonic stem cells through regulating cellular metabolism and DNA repair.** *Proceedings of the National Academy of Sciences of the United States of America* 2020, **117**(5):2519-2525.

## Tables

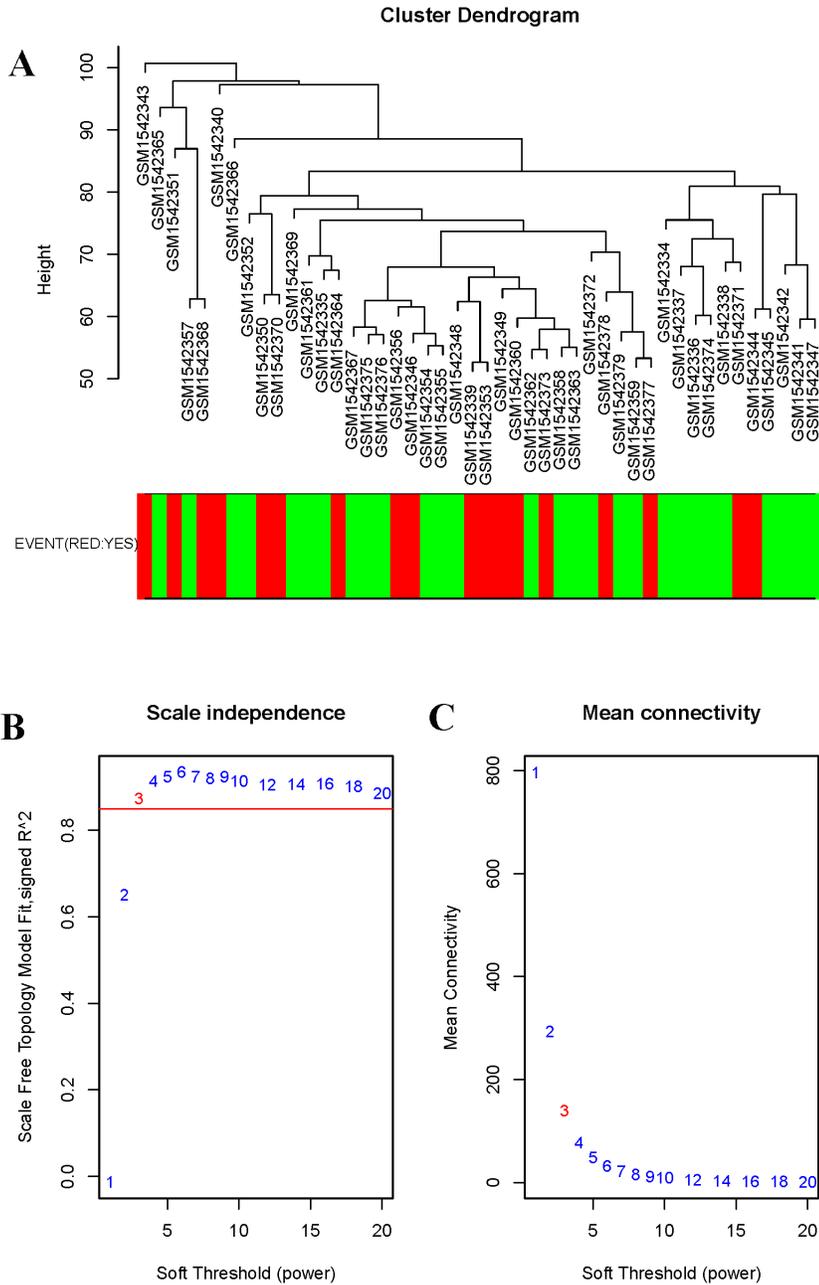
Table 1 Gene statistics in each module of GSE63155.

| Module      | Genes |
|-------------|-------|
| black       | 57    |
| blue        | 647   |
| brown       | 362   |
| cyan        | 11    |
| green       | 146   |
| greenyellow | 34    |
| grey        | 418   |
| magenta     | 50    |
| pink        | 54    |
| purple      | 38    |
| red         | 112   |
| salmon      | 18    |
| tan         | 23    |
| turquoise   | 1357  |
| yellow      | 222   |

Table 2 Gene statistics in each module of GSE63156.

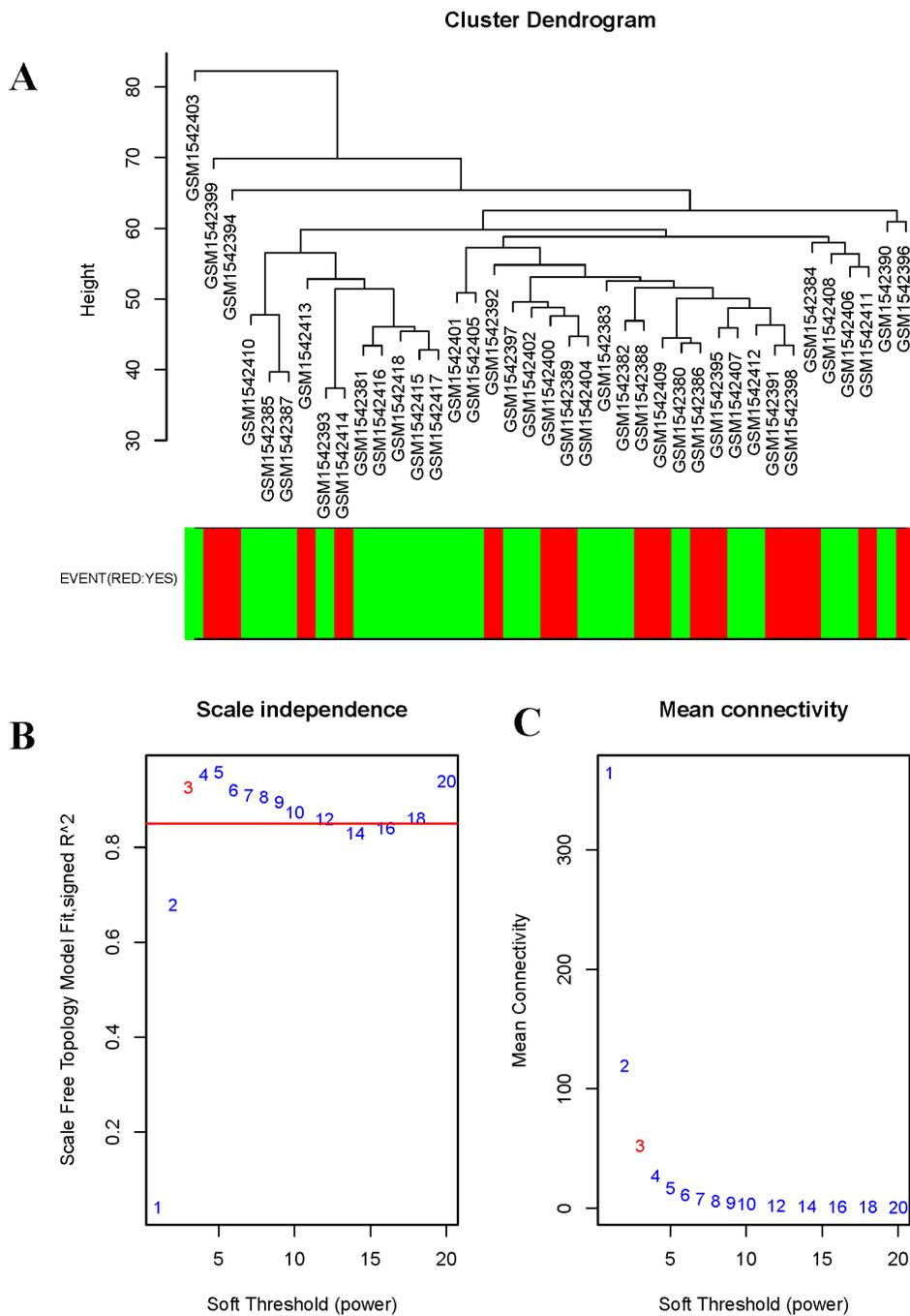
| Module    | Genes |
|-----------|-------|
| black     | 27    |
| blue      | 391   |
| brown     | 320   |
| green     | 73    |
| grey      | 309   |
| magenta   | 24    |
| pink      | 25    |
| red       | 28    |
| turquoise | 507   |
| yellow    | 146   |

## Figures



**Figure 1**

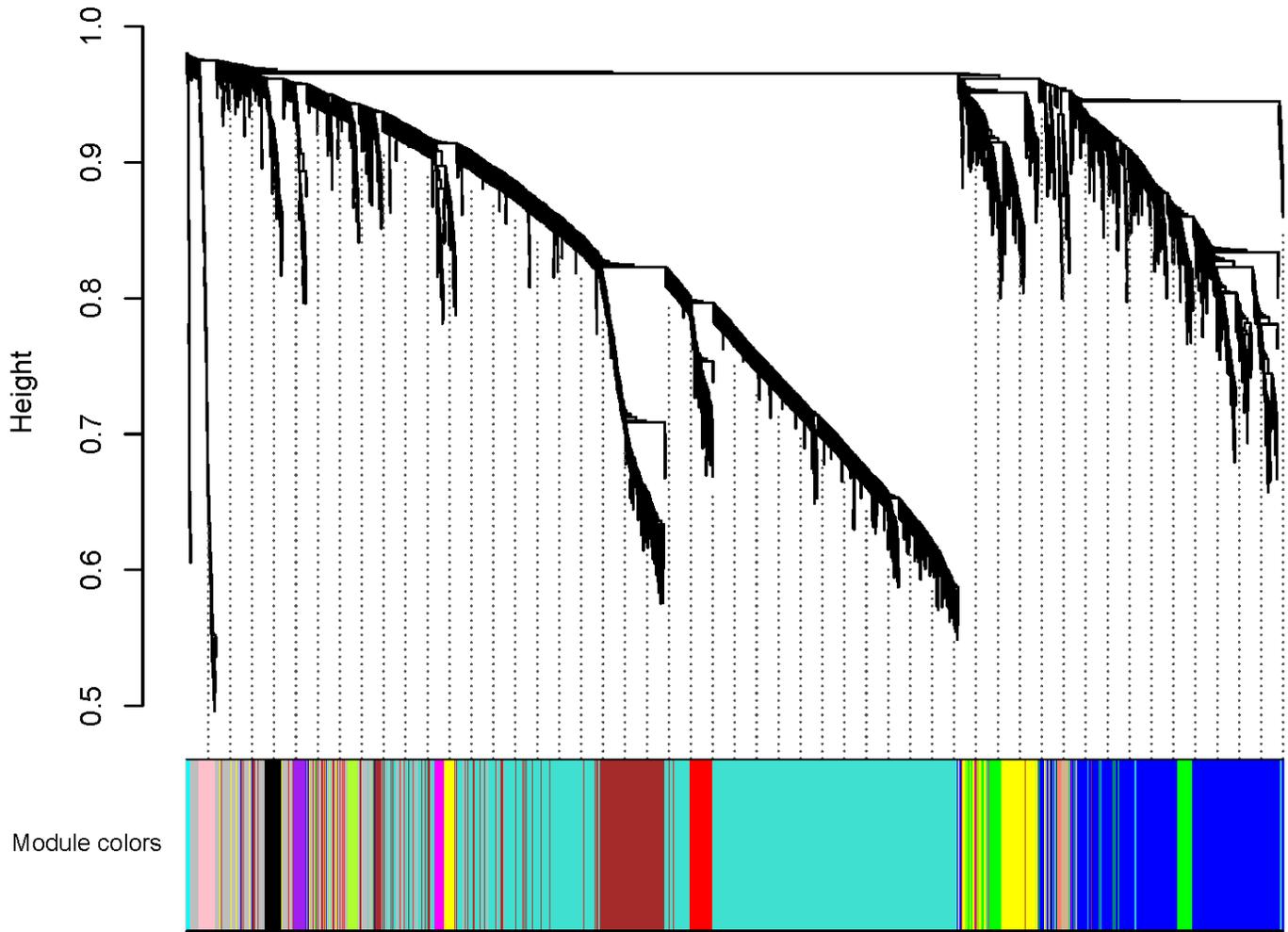
Clustering of samples and determination of soft-thresholding power. A The clustering was based on the expression data of GSE63155. The top 5,000 genes with the highest sD values were used for the analysis by Wgcna. The color intensity was proportional to expression status (recurrence, and Non-recurrence). B Analysis of the scale-free fit index for various soft-thresholding powers ( $\beta$ ). Analysis of the mean connectivity for various soft-thresholding powers. In all, 4 was the most fit power value.



**Figure 2**

Clustering of samples and determination of soft-thresholding power. A The clustering was based on the expression data of GSE63156. The top 5,000 genes with the highest sD values were used for the analysis by Wgcna. The color intensity was proportional to expression status (recurrence, and Non-recurrence). B Analysis of the scale-free fit index for various soft-thresholding powers ( $\beta$ ). Analysis of the mean connectivity for various soft-thresholding powers. In all, 4 was the most fit power value.

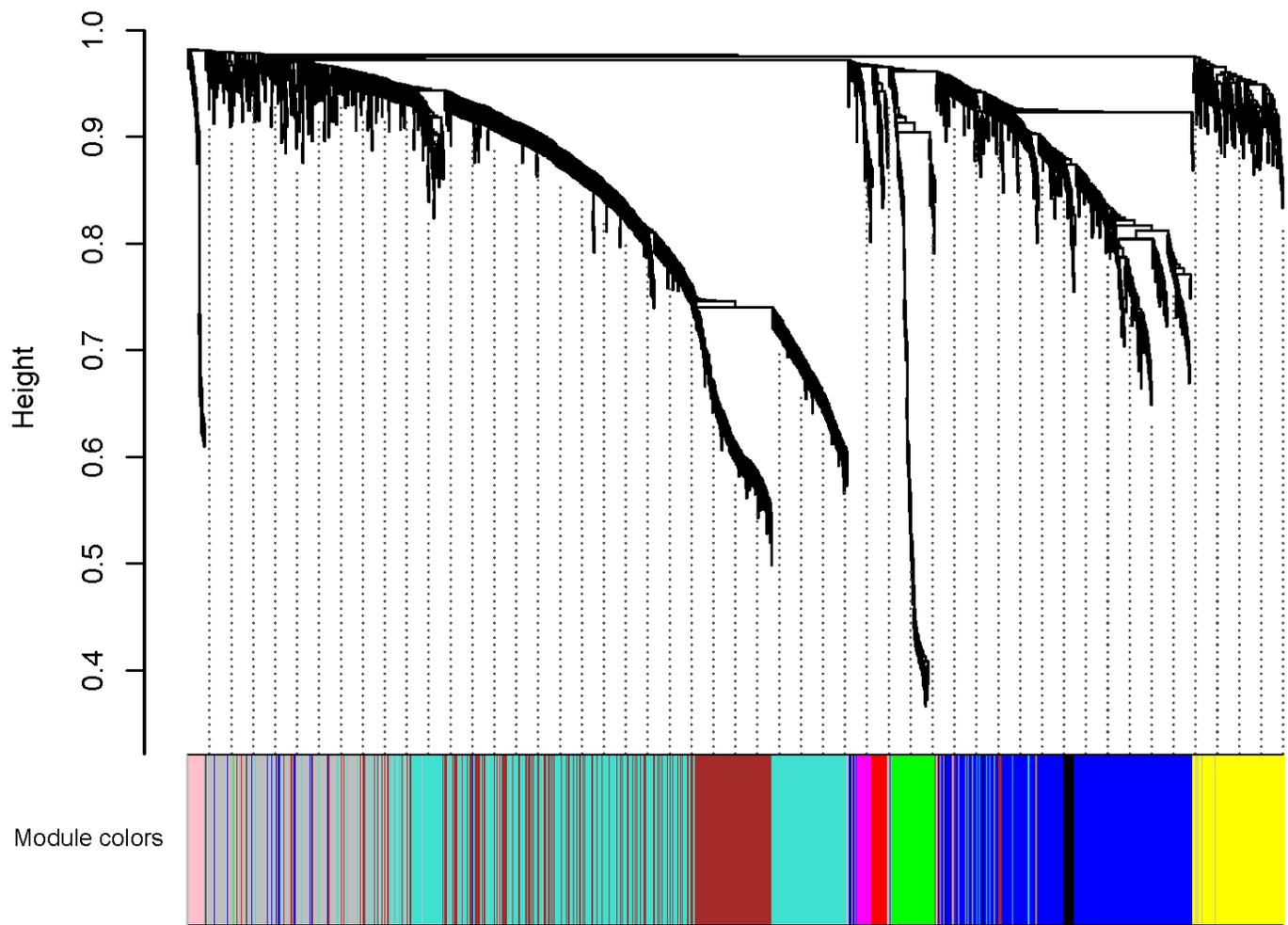
### Cluster Dendrogram



**Figure 3**

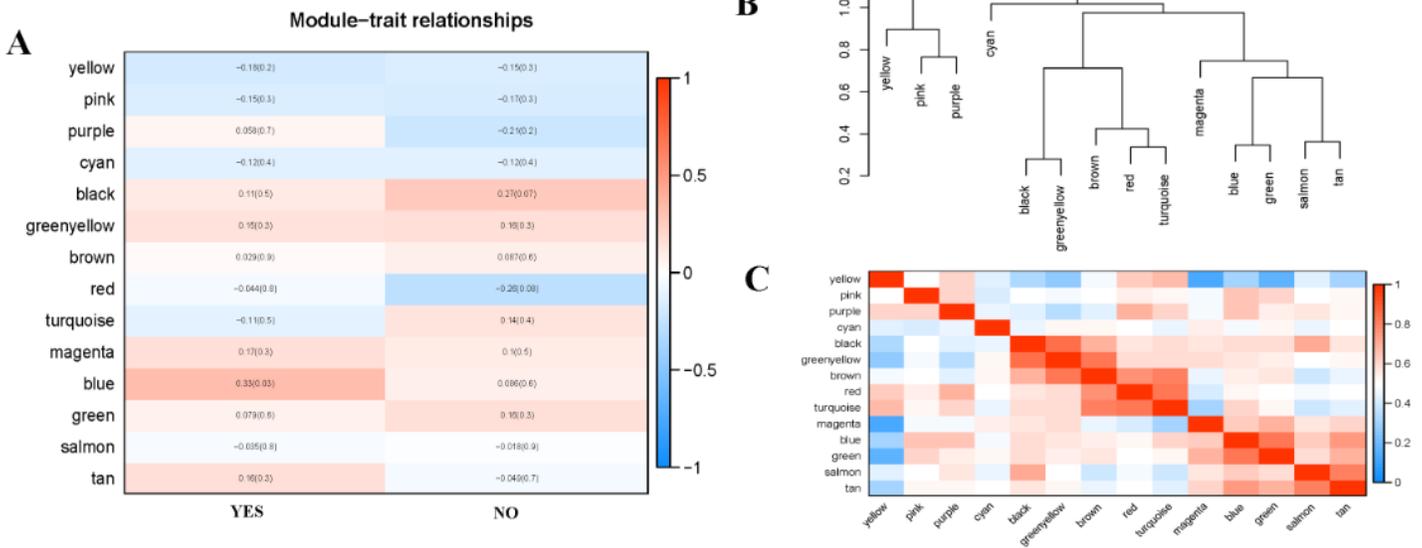
Construction of co-expression modules by Wgcna package in r. The cluster dendrogram of genes in GSE63155. Each branch in the Fig represents one gene, and every color below represents one co-expression module.

## Cluster Dendrogram



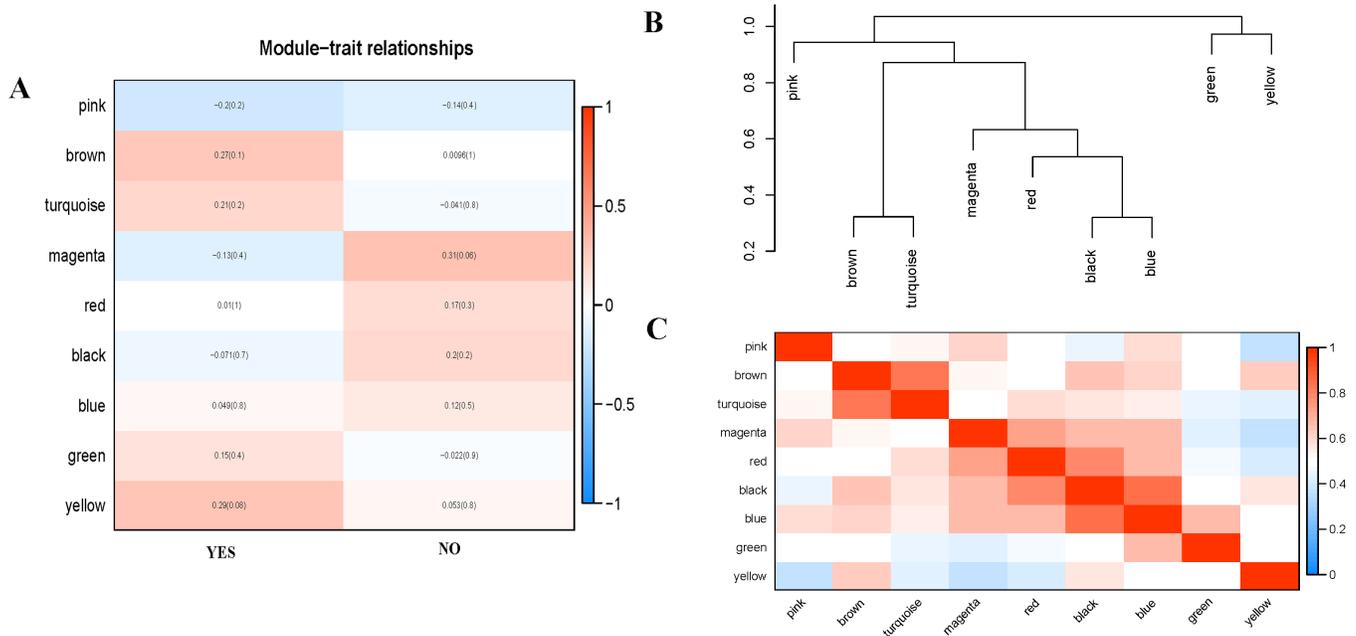
**Figure 4**

Construction of co-expression modules by Wgcna package in r. The cluster dendrogram of genes in GSE63156. Each branch in the Fig represents one gene, and every color below represents one co-expression module.



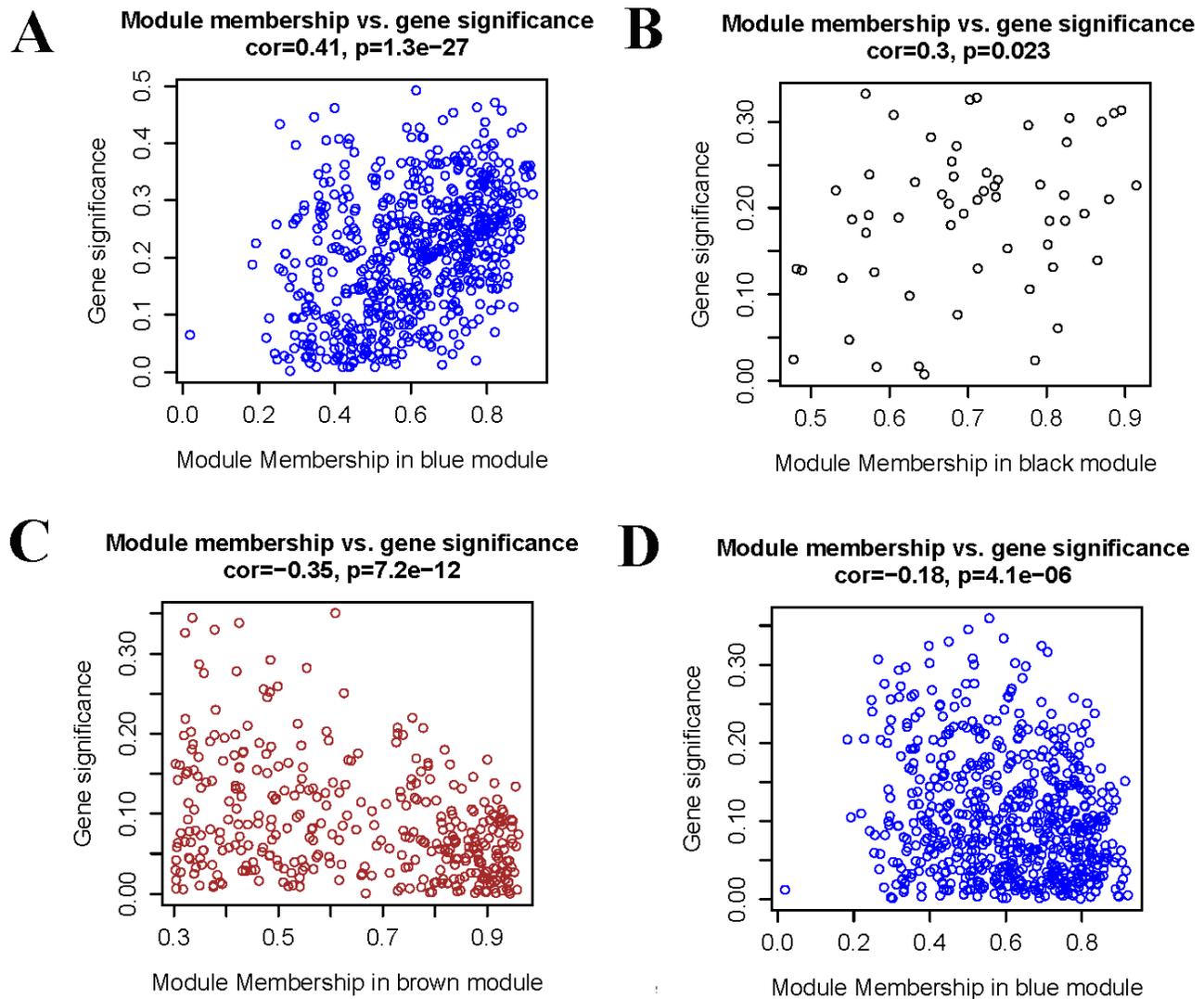
**Figure 5**

A Heatmap of the correlation between module eigengenes and the expression status of GSE63155 in recurrence. The Blue module, black module module was the most positively correlated with recurrence, and the brown module and Blue module was the most positively correlated with Non- recurrence. B Hierarchical clustering of module hub genes that summarize the modules yielded in the clustering analysis. C Heatmap plot of the adjacencies in the hub gene network.



**Figure 6**

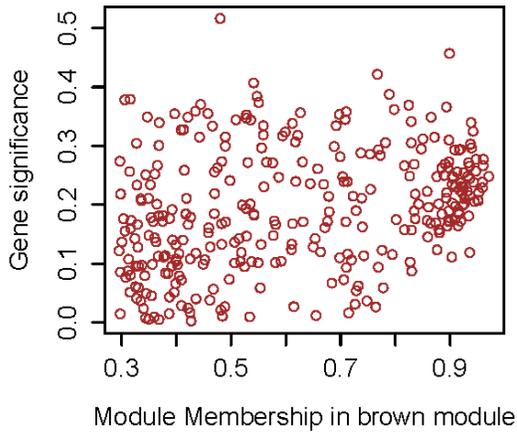
A Heatmap of the correlation between module eigengenes and the expression status of GSE63156 in recurrence. The Blue module, black module module was the most positively correlated with recurrence, and the brown module and Blue module was the most positively correlated with Non- recurrence. B Hierarchical clustering of module hub genes that summarize the modules yielded in the clustering analysis. C Heatmap plot of the adjacencies in the hub gene network.



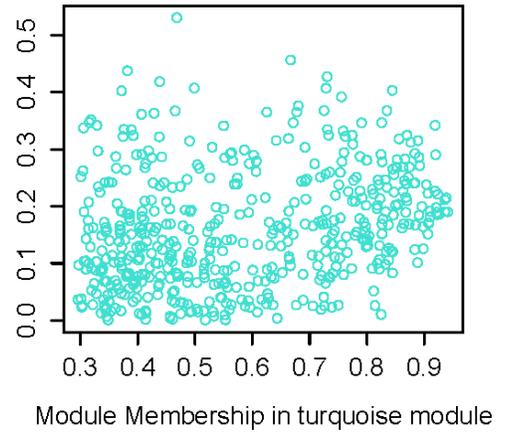
**Figure 7**

A Scatter plot of module eigengenes in the Bule module. B Scatter plot of module eigengenes in the black module module. C scatter plot of module eigengenes in the brown module module. D Scatter plot of module eigengenes in the Blue module module.

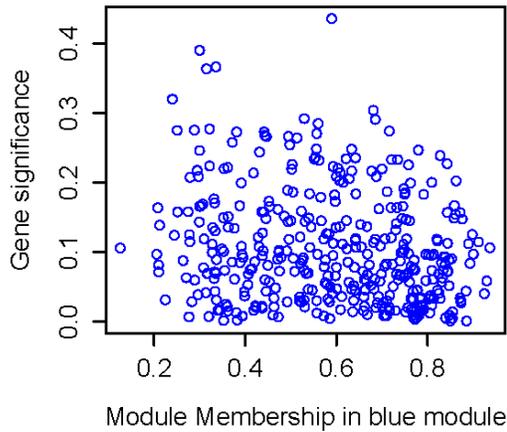
**A** Module membership vs. gene significance  
cor=0.3, p=4.4e-08



**B** Module membership vs. gene significance  
cor=0.27, p=6.4e-10

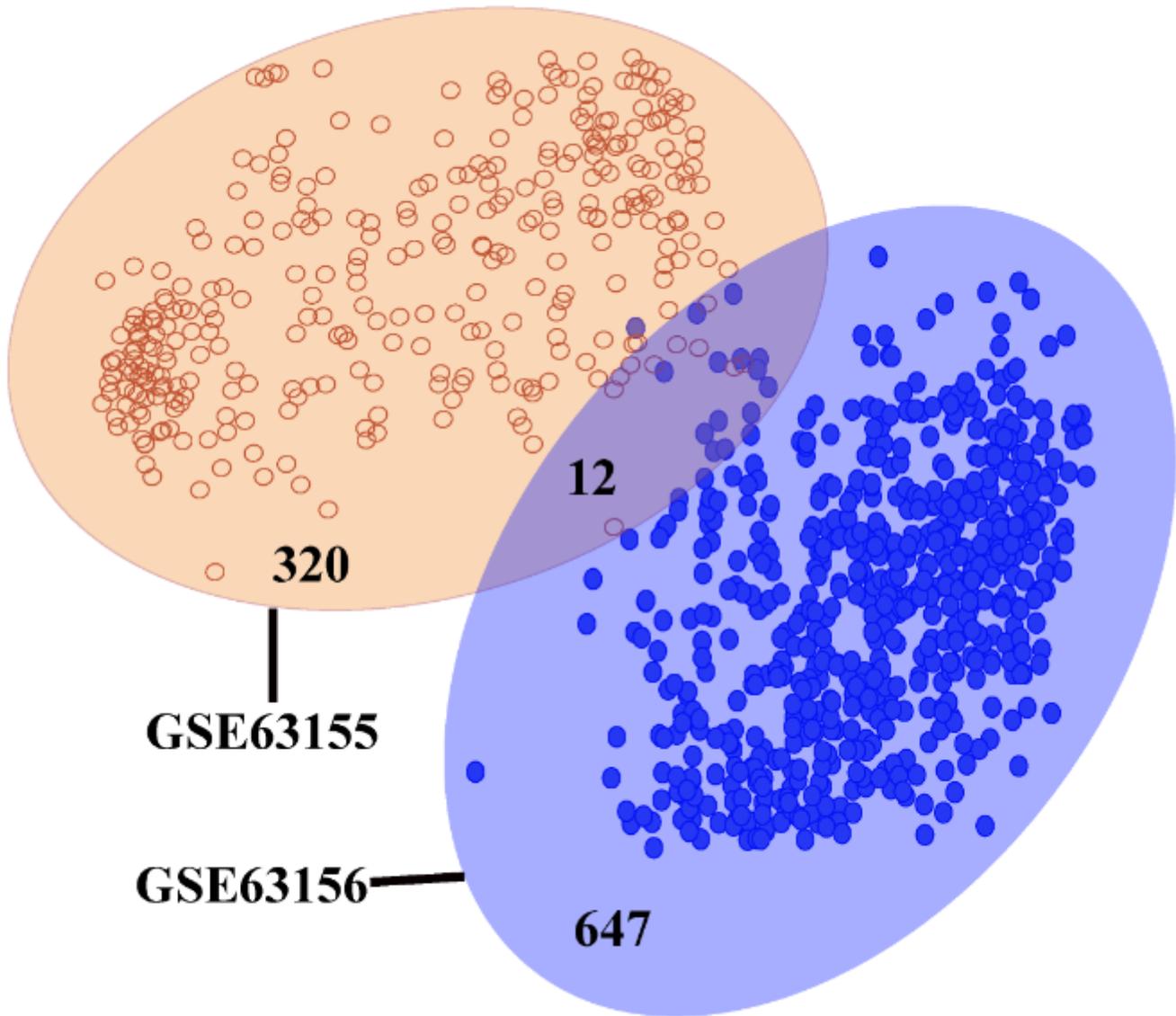


**C** Module membership vs. gene significance  
cor=-0.21, p=2.8e-05



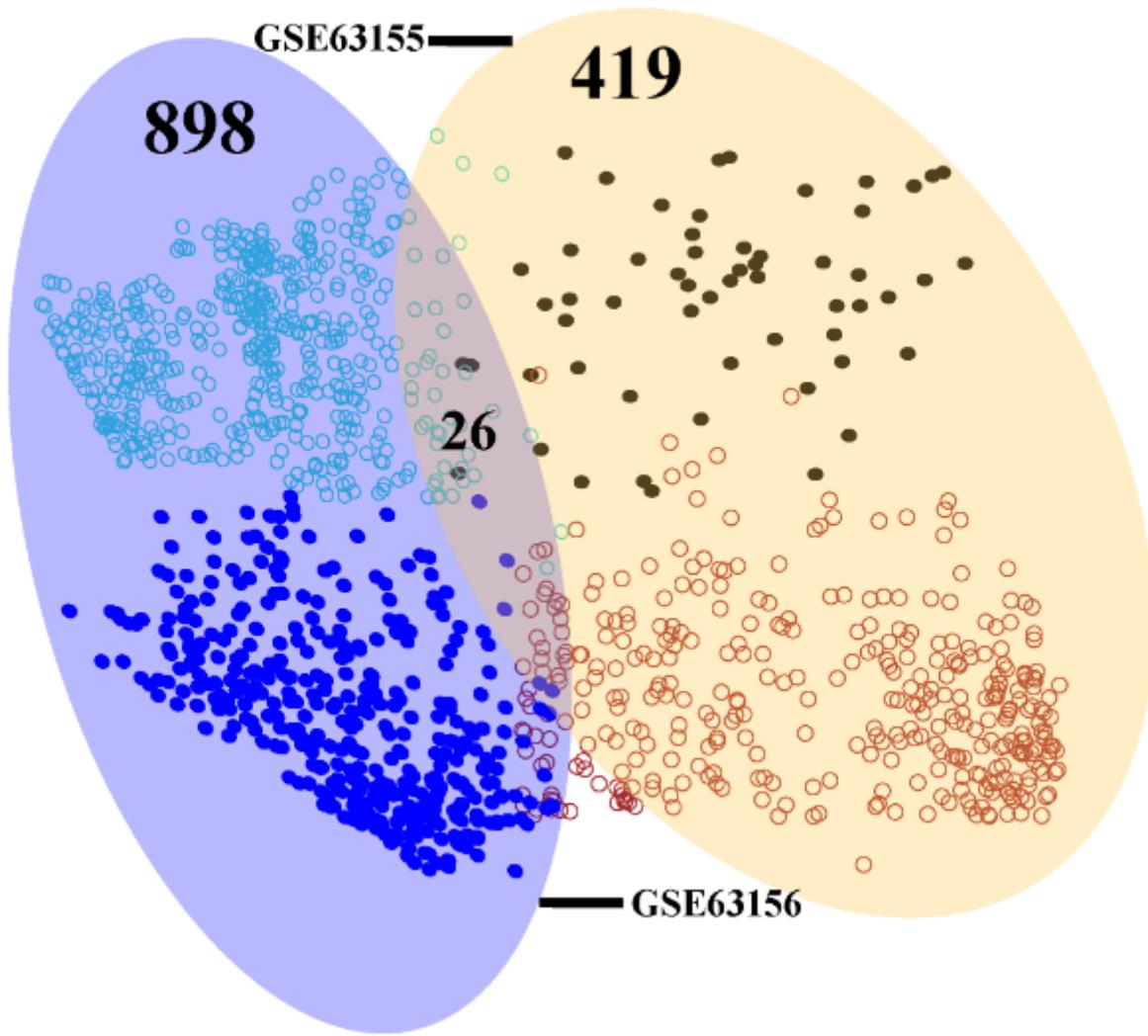
**Figure 8**

A Scatter plot of module eigengenes in the Brown module. B Scatter plot of module eigengenes in the turquoise module module. C Scatter plot of module eigengenes in the blue module module.



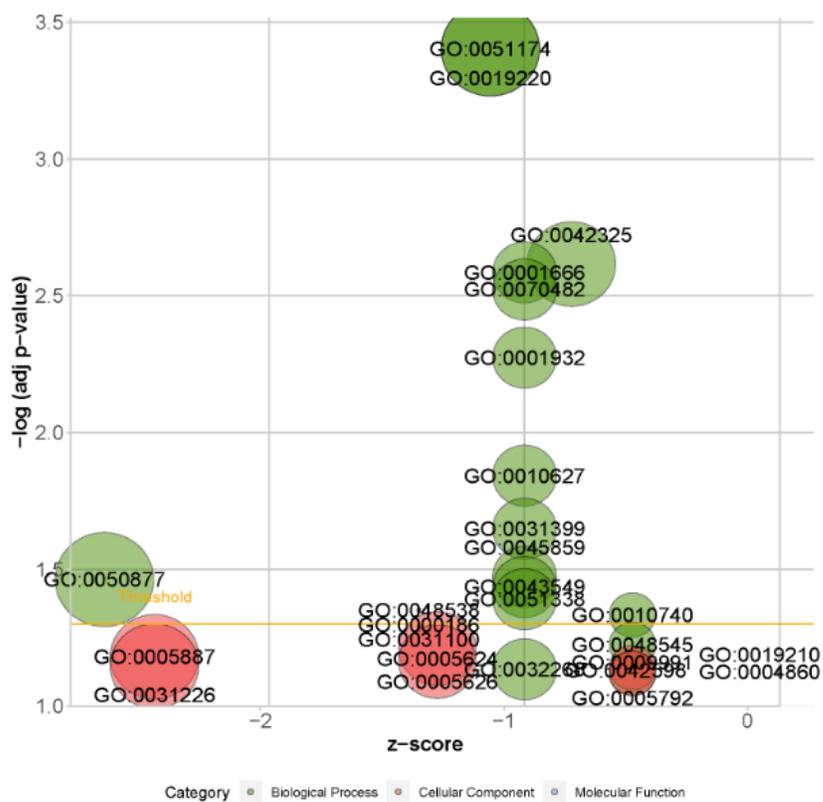
**Figure 9**

Euler diagram of the number of differentially expressed genes identified as significant (FDR adjusted  $p < 0.05$ ) for the Positive correlation of recurrence. Of 967 genes, 12 were identified as significant.



**Figure 10**

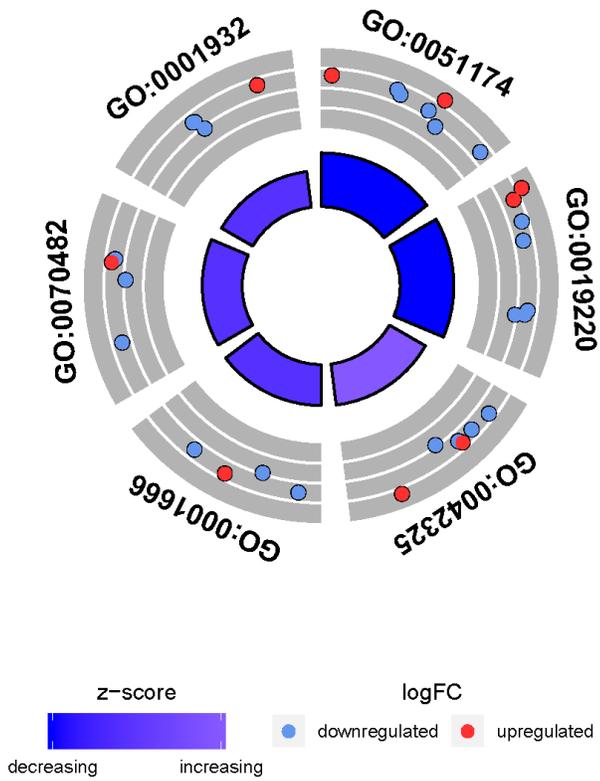
Euler diagram of the number of differentially expressed genes identified as significant (FDR adjusted  $p < 0.05$ ) for the Negative correlation of Non-recurrence. Of 1317 genes, 26 were identified as significant.



| ID         | Description                                      |
|------------|--|
| GO:0051174 | regulation of phosphorus metabolic process       |
| GO:0019220 | regulation of phosphate metabolic process        |
| GO:0042325 | regulation of phosphorylation                    |
| GO:0001666 | response to hypoxia                              |
| GO:0070482 | response to oxygen levels                        |
| GO:0001932 | regulation of protein amino acid phosphorylation |
| GO:0010627 | regulation of protein kinase cascade             |
| GO:0031399 | regulation of protein modification process       |
| GO:0045859 | regulation of protein kinase activity            |
| GO:0050877 | neurological system process                      |
| GO:0043549 | regulation of kinase activity                    |
| GO:0051338 | regulation of transferase activity               |
| GO:0048538 | thymus development                               |
| GO:0010740 | positive regulation of protein kinase cascade    |
| GO:0000186 | activation of MAPKK activity                     |
| GO:0031100 | organ regeneration                               |
| GO:0048545 | response to steroid hormone stimulus             |
| GO:0032268 | regulation of cellular protein metabolic process |
| GO:0009991 | response to extracellular stimulus               |
| GO:0005624 | membrane fraction                                |
| GO:0005887 | integral to plasma membrane                      |
| GO:0005626 | insoluble fraction                               |
| GO:0031226 | intrinsic to plasma membrane                     |
| GO:0005792 | microsome  |
| GO:0042598 | vesicular fraction                               |
| GO:0004860 | protein kinase inhibitor activity                |
| GO:0019210 | kinase inhibitor activity                        |

Figure 11

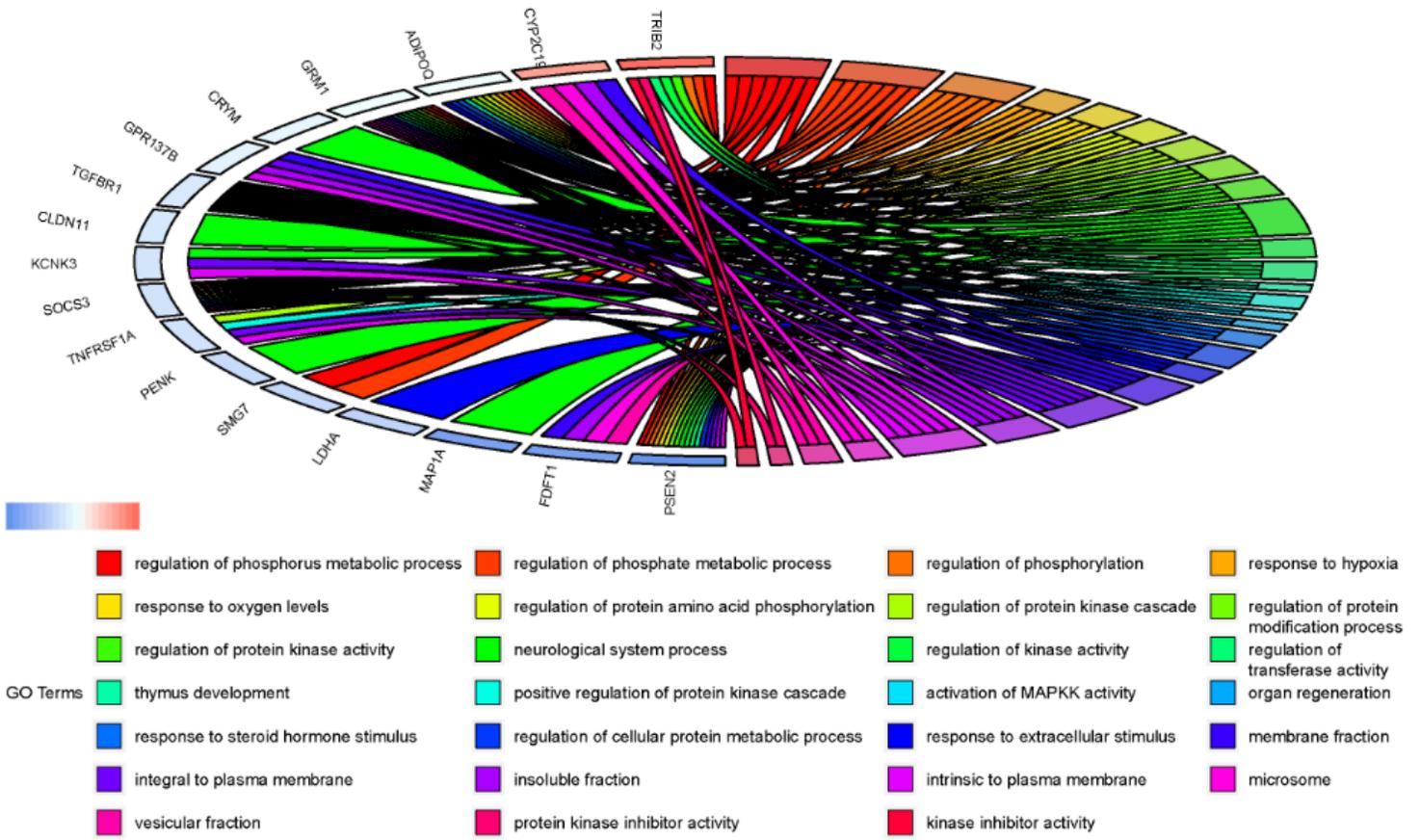
Functional enrichment and pathway analysis of the hub gene.



| ID         | Description                                      |
|------------|--|
| GO:0051174 | regulation of phosphorus metabolic process       |
| GO:0019220 | regulation of phosphate metabolic process        |
| GO:0042325 | regulation of phosphorylation                    |
| GO:0001666 | response to hypoxia                              |
| GO:0070482 | response to oxygen levels                        |
| GO:0001932 | regulation of protein amino acid phosphorylation |

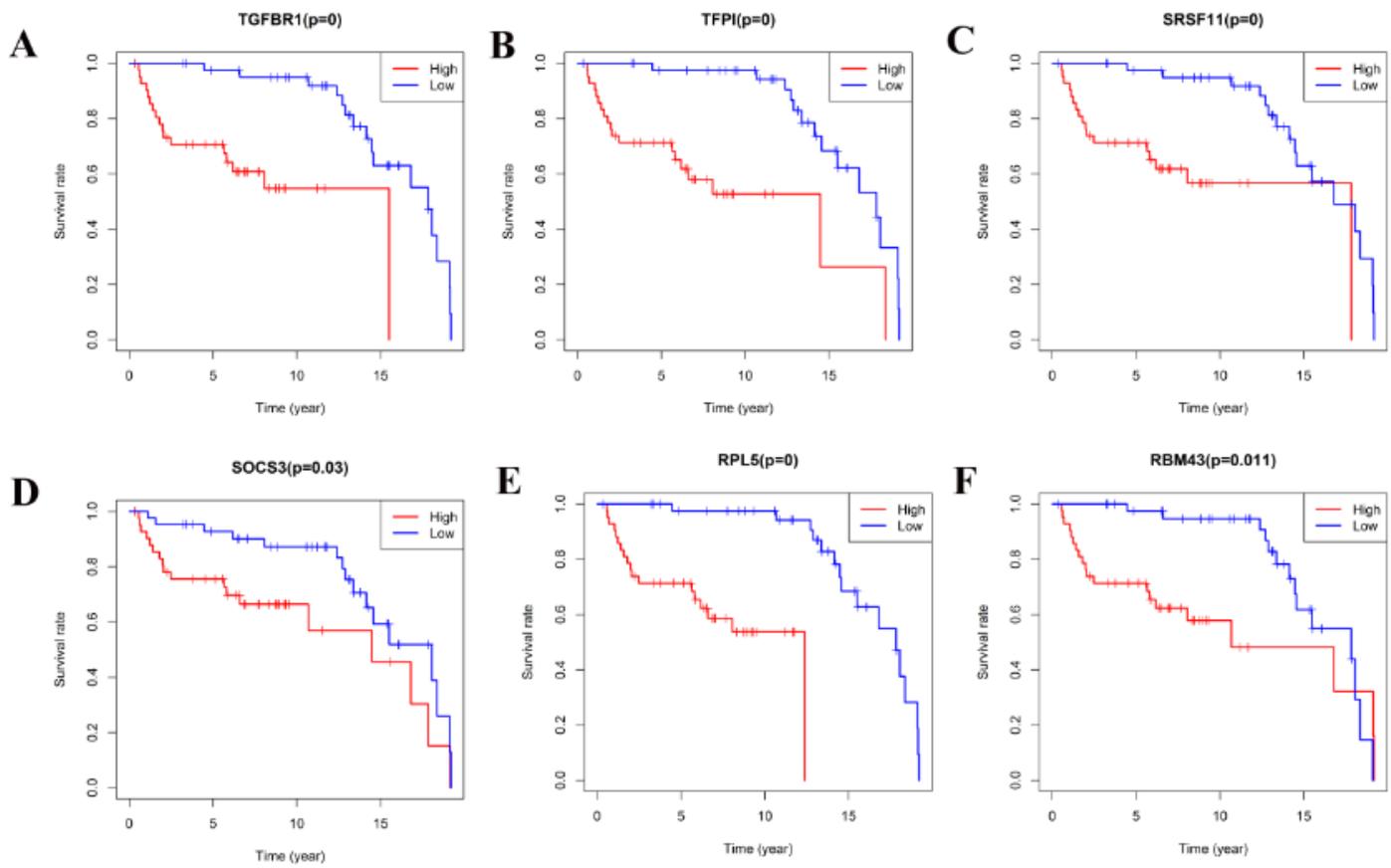
Figure 12

Hub Functional enrichment and pathway analysis of the hub gene.



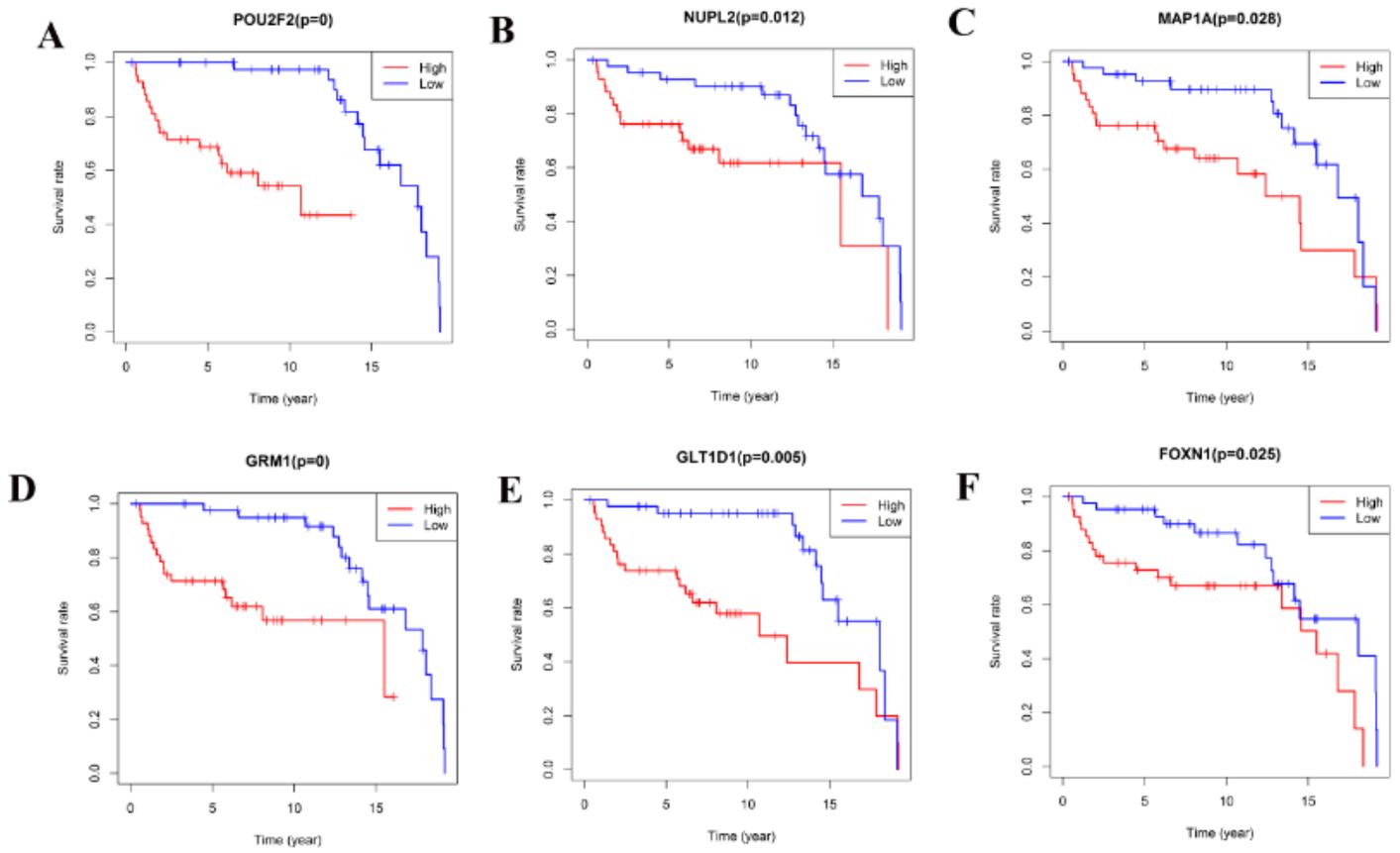
**Figure 13**

Functional enrichment and pathway correlate with gene of the hub gene.



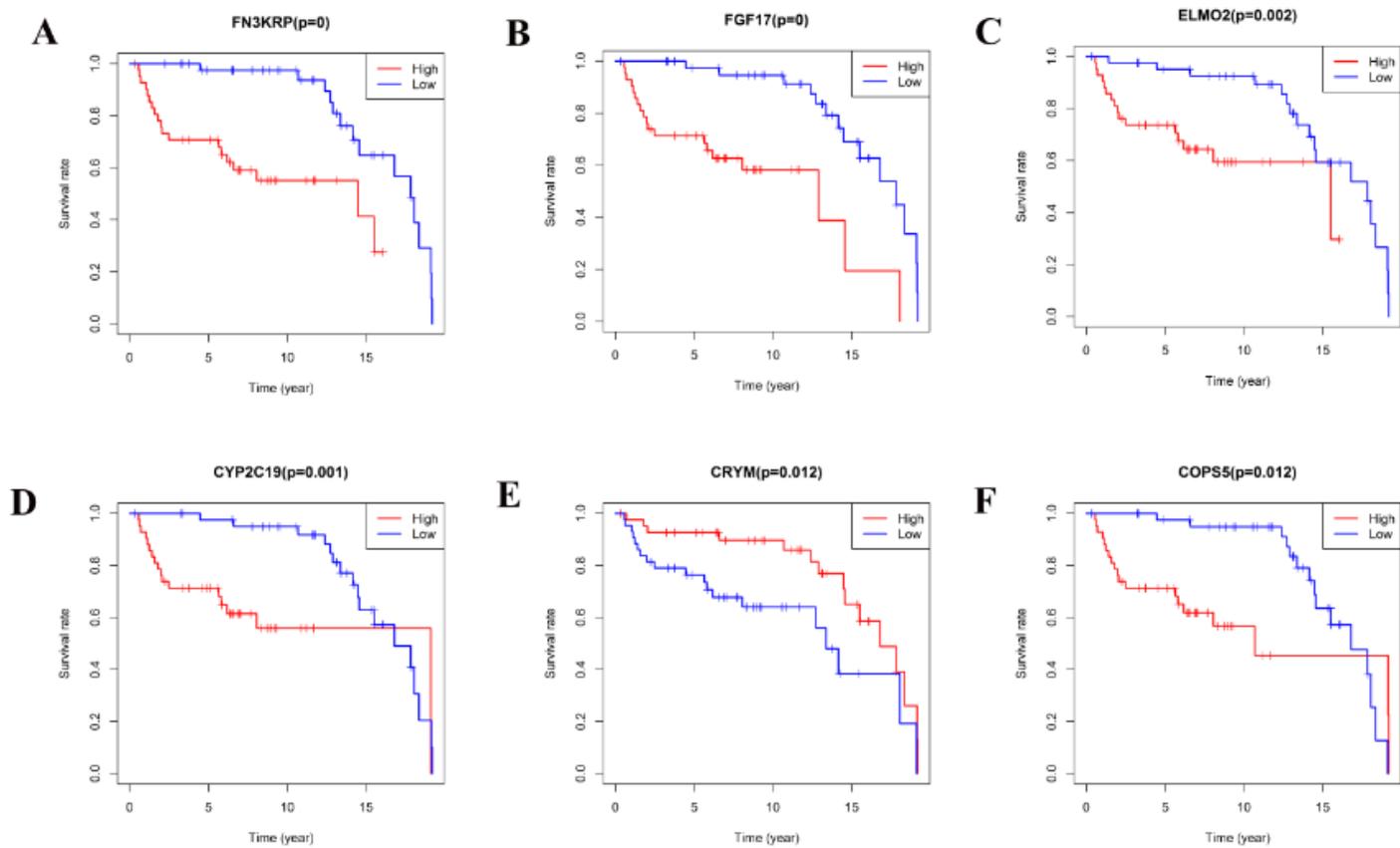
**Figure 14**

Survival analysis of hub genes. A–F Hub genes were TGFBR1, TFP1, SRSF11, SOCS3, RPL5 respectively.



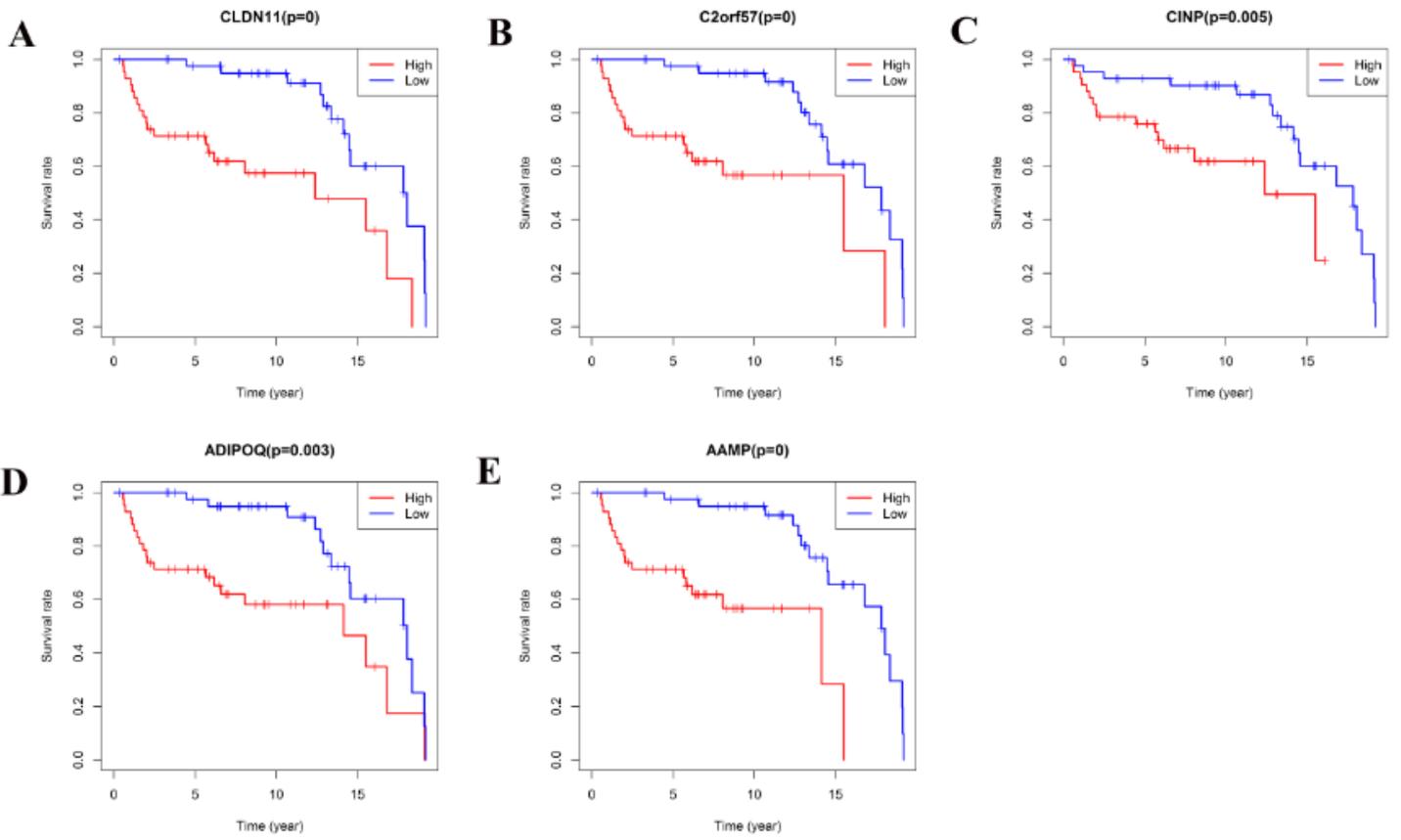
**Figure 15**

Survival analysis of hub genes. A–F Hub genes were POU2F2, NUPL2, MAP1A, GRM1, GLT1D1, FOXN1 respectively.



**Figure 16**

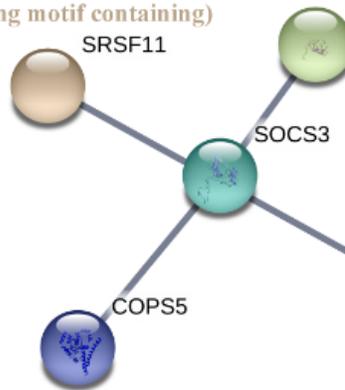
Survival analysis of hub genes. A–F Hub genes were FN3KRP, FGF17, ELMO2, CYP2C19, CRYM, COPS5 respectively.



**Figure 17**

Survival analysis of hub genes. A–E Hub genes were CLDN11, C2ORF57, CINP, ADIPOQ, AAMP respectively.

(Serine/arginine-rich splicing factor 11; May function in pre-mRNA splicing; RNA binding motif containing)



(E3 ubiquitin-protein ligase TRIM39; E3 ubiquitin-protein ligase. May facilitate apoptosis by inhibiting APC/C-Cdh1-mediated poly-ubiquitination and subsequent proteasome-mediated degradation of the pro-apoptotic protein MOAPI )

(Suppressor of cytokine signaling 3; SOCS family proteins form part of a classical negative feedback system that regulates cytokine signal transduction. SOCS3 is involved in negative regulation of cytokines that signal through the JAK/STAT pathway).

NUPL2  
(Nucleoporin-like protein 2; Required for the export of mRNAs containing poly(A) tails from the nucleus into the cytoplasm. In case of infection by HIV-1, it may participate in the docking of viral Vpr at the nuclear envelope)

(COP9 signalosome complex subunit 5; Probable protease subunit of the COP9 signalosome complex (CSN), a complex involved in various cellular and developmental processes.)

## Figure 18

The top hub genes in the core gene.