Construct ion of co - expression modules related to survival by WGCNA and identif ication of potential prognostic biomarkers in glioblastoma

Jing Zhou (zhoujingsx07@163.com)
Shanxi University of Traditional Chinese Medicine
Hao Guo (guohao2795@163.com)
Shanxi Provincial Peoples Hospital
Likun Liu (llkun133@126.com)
Shanxi University of Traditional Chinese Medicine
Shulan Hao
Shanxi University of Traditional Chinese Medicine
Zhi Guo
Shanxi University of Traditional Chinese Medicine
Fupeng Zhang
Shanxi University of Traditional Chinese Medicine
Zhi Wang
Shanxi Provincial Peoples Hospital
Weiwei Zhang
Shanxi Provincial Peoples Hospital

Research

Keywords: glioblastoma; weighted gene co-expression network analysis; hub gene; cell cycle; The Cancer Genome Atlas

Posted Date: May 12th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-27332/v1

Abstract

Background: Glioblastoma (GBM) is a malignant brain tumor with poor prognosis. However, the potential pathogenesis and therapeutic target s still needs to be explored.

Methods: Herein, The Cancer Genome Atlas (TCGA) expression profile data and clinical information were downloaded, and the Weighted Gene Co-expression Network Analysis (WGCNA) was conducted. Subsequently, hub genes which closely related to the poor prognosis of GBM were obtained. Further, the relationship between the genes of interest and prognosis of GBM patients, and immune microenvironment were analysed. Patients from TCGA were divided into high- and low-risk groups.

Results: Top 10 hub genes (CDC20, NCAPH, CDCA5, BUB1, CDCA8, PBK, KIF2C, TPX2, TTK and TOP2A) were obtained. Then, we performed single-gene analysis on CDCA5 and CDCA8 as genes of interest. We found that their expression levels were closely related to overall survival (OS). They had good correlations with the genes that regulate cell cycle in p53 signaling pathway. Moreover, it revealed that high amplication of CDCA5 was correlated with CD8 + T cells while CDCA8 with CD4 + T cells, and the expression of these two genes showed a significant difference in OS.

Conclusions: These results might provide new molecular targets and intervention strategy for GBM.

Introduction

Glioblastoma (GBM) is the most common malignant tumor of the central nervous system (CNS) in adults [1, 2]. It comprises 45.2% of CNS tumors and 54% of all gliomas [3]. Corresponding therapeutics, for instance, maximum surgical resection, comprehensive radiotherapy and chemotherapy has been applied to clinical practice. Yet the advances in treatment means have not concomitant with prominent amelioration in outcomes until recently. The five-year survival rate is still less than 10% [4-6]. Plenty of studies have been conducted to identify the underlying pathogenesis mechanisms, however, it’s still not been illuminated [7-9]. In particular, there are few related studies on the expression modules of GBM, which has brought certain difficulties to the identification of key genes in the occurrence and recurrence of disease. The potential heterogeneities and complexities of GBM make it difficult to identify reliable factors for determining effective clinical treatment. Hence, it is urgently needed to uncover efficient molecular targets which can clinically significance contribute to the personalized treatment and improve prognosis for GBM patients.

As a newly invented systematic biology approach, Weighted Gene Co-expression Network Analysis (WGCNA), has been used to describe the connectivity of gene clusters inside the comprehensive network and assess the correlations of gene modules with different clinical features [10, 11]. Distinguished from other analysis method, WGCNA hierarchical clustering methods focused on the whole genome information instead of previous selected genes to overview of the signature of gene networks in phenotypes which can avoid bias and subject judgement. [12]. WGCNA has been widely used in the study of multiple diseases [13-15]. By constructing a co-expression network of genes and an identification

module, WGCNA can investigate hub genes closely related to clinical phenotypes, which will provide us a beacon of hope for discovering new molecular biomarkers and therapeutic targets in GBM.

In the present study, we acquired the clinical information of GBM patients from The Cancer Genome Atlas (TCGA) database. These patients were defined as high- risk and low-risk groups respectively in accordance with the follow-up time and survival status. The gene co-expression networks of these two groups were constructed by WGCNA, then the modules related to prognosis were identified and the core genes in the modules were obtained. Through the screening and functional enrichment analysis of the hub genes in the prognosis-related specific modules, two genes, cell division cycle associated 5 (CDCA5) and cell division cycle associated 8 (CDCA8), which are vital to the prognosis of GBM patients have been anchored. Additionally, we performed a single-gene analysis of CDCA5 and CDCA8 to further validate our prediction. These findings may greatly help us develop new therapeutic targets and improve GBM patient's clinical outcomes.

Materials And Methods

Data acquisition and samples grouping

TCGA database (http://www.cancer.gov/about-nci/organization/ccg/research/structural- genomics/tcga) is a landmark cancer genomics database, which mainly contains clinical data of various human cancers, such as genome variation, mRNA expression, miRNA expression, methylation and other data. Our study included a total of 142 GBM patients with complete clinical information from TCGA database. Patients were divided into high- and low-risk group according to the follow-up time and survival status. A total of 93 patients with a follow-up of less than 60 months and the survival status of death were defined as high-risk patients, while 49 patients of the rest were defined as low-risk patients. We downloaded and used TCGA level 3 FPKM RNA-seq and clinical data for subsequent WGCNA network construction.

WGCNA co-expression network construction and significant module

identification

In our study, we constructed the gene co-expression network of high- and low-risk GBM patients respectively via the standard procedure of WGCNA[10]. The WGCNA package in R (http://www.r-project.org/) was used to WGCNA installation, data reading and import. The data were obtained by removing genes with zero variance between groups and including the first 75% of gene sets with Median Absolute Deviation (MAD) for further analysis. The filtering principle of soft threshold was to make the constructed network more consistent with the characteristics of scale-free network. The weighted adjacency matrix was transformed into a topological overlap matrix (TOM) to estimate its connectivity in the network. The hierarchical clustering method was used to construct the clustering tree structure of the TOM. Different branches of the cluster tree represented different gene modules, and different modules were represented by different colors.

Based on their weighted correlation coefficients, genes were classified on the grounds of their expression patterns. Finally, tens of thousands of genes were divided into multiple modules according to gene expression patterns. Comparing the two non-preserved modules of high- and low-risk GBM patients with the two co-expression networks, the module with the minimum value of Z-summary score was the specific module of the high-risk group. We identified the hub gene of this non-preserved module by the degree of genes linkage and performed functional enrichment analysis on them.

Functional enrichment analysis of hub genes

The clusterProfiler package in R[16] was used to annotate hub genes to fully discover and explore their functional correlations. The tools of Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were used to assess the relevant functional categories. The *P*-value of less than 0.05 and q-value of less than 0.1 were set as the threshold.

Further single-gene analysis of the interest genes in hub genes

We intended to further tap the relationship between interest genes in hub genes and GBM progression through single-gene analysis. Firstly, the relationship between the genes of interest and the prognosis of GBM patients was obtained through the PrognoScan database[17] (http://dna00.bio.kyutech.ac.jp/PrognoScan/index.html). Then we analyzed all RNA-seq data on GBM in the TCGA database. The correlation coefficient of greater than 0.4 and the *P*-value of less than 0.001 were set as the filtering condition. Then we obtained the genes co-expressed with the interested genes. After screening the genes related to the p53 signaling pathway, the "pheatmap" and" gplot "packages were used to draw the heatmap for the correlation analysis of the interested genes. At last, the relationship with immune microenvironment and immune cell content was obtained by TIMER[18]Tumor Immune Estimation Resource, https://cistrome.shinyapps.io/timer/). All statistical analyses were conducted using R 3.6.0 (https://www.r-project.org/). All tests were performed bilaterally, $P<0.05$ indicated a difference with statistical significance.

Results

Data preprocessing and soft threshold screening

Considering the sensitivity of WGCNA to the effect of batch processing. We first preprocessed the data sets of all 93 GBM patients in the high-risk group and 49 GBM patients in the low-risk group from the TCGA database. By removing genes with zero variance between groups and including the first 75% of gene sets with MAD, we obtained the gene sets for following analysis. Subsequently, we used the hclust function to confirm the effect of batch removal from the dataset and to see if there were any outliers. The results showed that these data sets had not been corrected due to the batch removal effect, and the tree graph and samples were clustered in the correct random order (Figure 1). Due to the premise of WGCNA algorithm needs to assume that gene network is subject to scale-free distribution. Thus, we next needed to screen out appropriate soft threshold (power) to make the constructed network more consistent with

the characteristics of scale-free network. We set the soft threshold as 5 (high-risk group) and 10 (low-risk group) respectively to meet the selected criteria of power value (Figure 2). By calculating the scale-free topology fitting index, the value of R-square reached 0.9 (Figure 3). All the above results further verified and illustrated the feasibility of WGCNA.

Construction of co-expression networks and identification of modules

We constructed two co-expression networks of the high- and low-risk GBM patients. Hierarchical clustering analysis was conducted based on weighted correlation, and the clustering results were segmented according to the set criteria to obtain different gene modules (Figure 4). By using WGCNA for the low-risk group, we identified ten modules of different sizes, and used branches of the cluster tree and different colors to represent them. Then the high-risk group network was mapped to the low-risk group network modules. This approach helped us identify non-preserved modules. Non-preserved modules could explain the change of network properties between low- and high-risk group networks. In addition, these non-preserved modules may be related to survival status of GBM patients and tumor progression. The results showed that the module tags were still clustered together in the low-risk group network, indicating that the preservation of this module was well. To validate the stability of WGCNA, we used the module Preservation function to calculate the module preservation. The saved median and Z-summary score were showed for different color modules (Figure 5). The turquoise module had the highest Z-summary score, which indicated that it retained the network characteristics of the high-risk group network. However, the black module with the lowest Z-summary score menat a low degree of preservation, indicating that the prognosis level could be distinguished between high- and low-risk patients. The black module was the specific module of the high-risk group.

Identification and functional enrichment analysis of hub genes

To identify key nodes associated with prognosis, we performed a more detailed analysis of the black module. Because it was minimally preserved between networks and could be used to distinguish between samples of high- and low-risk GBM patients. As a result, a heat map of 50 core genes was obtained (Figure 6), which were identified in the high-risk patient network and may play important roles in shortening the survival time of GBM patients. Then, the Cytoscape software was applied to calculate the strength of the intra-module connectivity of each gene for the non-conservative modules. We sorted these genes take scores as the standard and finally got the top 10 hub genes (CDC20、NCAPH、CDCA5、BUB1、CDCA8、PBK、KIF2C、TPX2、TTK and TOP2A).

GO enrichment analysis was performed on the genes in the black module. The results demonstrated that the biological process of these 50 core genes mainly enriched in three aspects: Cell cycle, Progesterone–mediated oocyte maturation and Oocyte meiosis (Figure 7A). In addition, we showed in detail that each gene corresponded to a specific link in the cell cycle (Figure 7B). By further enrichment analysis of genes and the interactions in cell cycle, we found that these genes mainly played important roles in the processes of chromosome segregation, organelle fission, nuclear division and mitotic nuclear division (Figure 7C).

Single-gene analysis of CDCA5 and CDCA8

Through further searched for literatures of the top 10 identified hub genes, we found that there scarcely no reports about the mechanism of CDCA5 and CDCA8 with GBM since which were important regulatory proteins in the cell cycle in cancer. Thus, CDCA5 and CDCA8 were tremendously expected to be as new therapeutic targets for GBM. Hence, we conducted single-gene analysis on CDCA5 and CDCA8 to explore their roles in GBM.

Gene expression profiles of CDCA5 and CDCA8 were obtained from the NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo/): GSE4412. GBM Patients were divided into high and low expression groups according to the CDCA5 and CDCA8 gene expression levels. The PrognoScan database was used to analyze their relationship with the prognosis of GBM patients. The results showed a significant difference in overall survival (OS) between the two groups, the OS in both CDCA5 and CDCA8 high expression groups were dramatically shorten when compared with the low expression groups (Figure 8A). Subsequently, the co-expression analysis of CDCA5 and CDCA8 with the whole genome of TCGA database were conducted to screen the gene correlations related to cell cycle in the p53 signaling pathway. The correlation analysis heatmaps were made by the "pheatmap" and" gplot "packages (Figure 8B). We found that both of these two genes had good correlations with the genes which regulated the cell cycle in p53 signaling pathway: CCND1, CCNB1, CCNB2, CCNE1, CDK1, and CDK2 (Figure 8C:CDCA5 and Figure 8D:CDCA8). These results suggested that CDCA5 and CDCA8 may be involved in the signal regulation of p53 pathway by affecting relevant genes in the cell cycle. Then, the relationship between these two genes and immune microenvironment of GBM was obtained by TIMER. We studied the differential expression of the CDCA5 and CDCA8 in tumors and normal tissues of multiple cancer species, the relationship between expression levels and copy number variations of genes and the levels of infiltration of six immune cells (B cell, $CD8^+$T cell, $CD4^+$T cell, macrophage, neutrophil, dendritic cell). The results revealed that CDCA5 and CDCA8 were differentially expressed in tumor and normal tissues of multiple cancer species, showing a tendency of up-regulation (Figure 9A). Figure 9B showed that there was no significant correlation between the expression levels of CDCA5 and CDCA8 and the six types of immune cells. However, the high amplication of CDCA5 was obviously correlated with $CD8^+$T cells in GBM (Figure 9C). Similarly, the high amplication of CDCA8 had a significant correlation with $CD4^+$T cells in GBM (Figure 9C). It was indicated that CDCA5 and CDCA8 may affect the immune microenvironment of GBM through this mechanism, leading to the malignant progression of GBM.

Discussion

GBM is the most common primary malignant brain tumor in adults, with a poor prognosis and high mortality due to its highly aggressive characteristics [19]. As an incurable malignant brain tumor, surgical resection, radiotherapy and medical therapy may never be achieved [20, 21]. The advances in GBM therapy have not concomitant with prominent amelioration in outcomes until recently [22]. Therefore, exploring molecular targets and therapeutic means are urged needed [23]. Years of molecular studies have identified many key links that affect the development and progression of GBM [24]. Especially with

the progress of high-throughput genome technology make it possible to find more potential molecular markers by using bioinformatics methods.

In the present study, data of GBM patients with complete clinical information were obtained from TCGA database. The patients were divided into high-risk and low-risk groups according to the follow-up time and survival status of the patients. Our study was the first to construct co-expression modules related to survival by WGCNA in two groups of patients. Compared with other methods, WGCNA has many obvious advantages. Benefit from its analysis focuses on the association between co-expression modules and clinical features of interest, the analytical results have better reliability and biological significance[25]. We analyzed the conservation of all modules in the high-risk group and low-risk group. Due to the low degree of preservation between the high- and low-risk group, the non-conservative module of the two co-expression networks (that was, the module with the minimum preservation Z-summary score) was considered as the specific module of the high-risk group after cluster analysis. The preservation Z-summary score results from figure 5 showed that the black module was identified to be the lowest conservative module due to its lowest Z-summary value. Therefore, we focused on the black module to explore the influencing factors related to survival of patients in the high-risk group.

We identified 50 key genes from the black module and plotted a heat map. These genes were the key genes that affect the survival time and survival status of patients with GBM. Our further analysis of these genes by GO showed that these prognostic genes were mainly related to cell cycle. In particular, it was most closely related to the key links: chromosome segregation, organelle fission and nuclear division. Therefore, we speculated that these genes may shorten the cell cycle and accelerate the replication of cancer cells by regulating the cell cycle of cancer cells, resulting in the rapid spread of cancer cells. In order to explore the specific mechanism of the effect of these genes on survival, we screened the top 10 genes (CDC20, NCAPH, CDCA5, BUB1, CDCA8, PBK, KIF2C, TPX2, TTK and TOP2A). By retrieving related literature, we found that CDCA5 and CDCA8, as important regulatory proteins in the cell cycle in cancer, were recognized as oncogenes[26-29]. However, compared with other genes, there scarcely no reports about the mechanism of CDCA5 and CDCA8 with GBM. Thus, conduct study on the specific mechanism of CDCA5 and CDCA8 to GBM malignant progression might have vital clinical significance.

We carried out single gene analysis of CDCA5 and CDCA8 respectively to explore their potential mechanism. Firstly, we obtained and verified the effects of the two on the overall survival (OS) time from the chip data of GSE4412. The results showed that the OS probability of patients with high expression of CDCA5 and CDCA8 in tumor tissues was significantly decreased. This was consistent with the effect of the two on the survival time of patients in other types of tumors[30-32]. Secondly, we analyzed the correlation between the two and all genes involved in the p53 signaling pathway. p53, as a star tumor suppressor gene, has been demonstrated can regulate cell cycle and prevent cell cancerization. It is referred to as the "guardian of the genome" by the scientific community. Generally, more than 50% of cancer patients have mutations and inactivation of p53 gene[33, 34]. The p53 signaling pathway, as the most influential signaling pathway in the tumor field, also has a significant impact on the incidence of different cancers[35]. In particular, it plays an important role in regulating cell cycle. Fortunately, we find

the cell cycle regulation-related proteins CDCA5 and CDCA8 were significantly correlated with CCND1, CCNB1, CCNB2, CCNE1, CDK1 and CDK2 (correlation coefficient > 0.4, $P < 0.05$). These genes were key genes that regulating the cell cycle in the p53 signaling pathway. CCND1 is a protein encoded by the human CCND1 gene. It forms a complex with CDK4 or CDK6 and acts as a regulatory subunit, which is essential for the transition from G1 to S phase of cells. Mutations, amplification or overexpression of the gene could change the cell cycle process. These phenomena often occurre in many tumors and may cause tumorigenesis[36, 37]. CCNB1/2, as a vital member of the cyclin family, is an important cell cycle regulator related to G2/M detection points in cells. It regulates cyclin-dependent kinase 1 (CDK1) and forms a complex with it to phosphorylate the substrate, initiate cells from G1/S phase to G2/M phase, and promote mitosis. Plenty of evidence indicates that CCNB1/2 dysfunction is an early event in tumorigenesis, and its unregulate expression can be observed in many human tumors, including breast cancer, lung cancer, and brain cancer[38, 39]. CCNE1 plays crucial role in regulating the cell from G1 to S phase. It forms a complex by binding and activating CDK2, which plays a very important role in inducing the synchronization of DNA replication, centrosome replication and regulation, chromosome reconstruction, and histone synthesis. It has been reported in the literature that the high expression of CCNE1 was closely related to the poor clinical prognosis of patients with various malignancies such as ovarian, bladder, and colon cancer[40, 41]. Therefore, we speculated that the effect of CDCA5 and CDCA8 on the prognosis of GBM patients may be achieved by participating in the regulation of cell cycle in the p53 pathway. In particular, the co-expression of these key genes leaded to dysfunction at G1/S checkpoint and/or G2/M checkpoint through overexpression of these genes, leading to active replication of cancer cells and malignant tumor progression.

In recent years, tumor immunotherapy has become a novel focus in cancers. More and more studies have focused on the infiltration of immune cells in tumor tissues to explore the relationship between tumor microenvironment and clinical outcomes[42, 43]. By applying online tool-TIMER, we attempted to explore the relationship between CDCA5 and CDCA8 and immune cells in GBM. We found that the expression of these two genes was significantly up-regulated in GBM as well as various other types of cancer tissues compared to normal tissues. This present study revealed that CDCA5 and CDCA8 with high amplication had significant effects on $CD8^+$ T cell and $CD4^+$ T cell respectively in GBM patients under different gene copy states. Thus, both of these two genes may also influence the prognosis to some extent by regulating the immune microenvironment of GBM patients. However, the specific relationship between these two genes and immunity of GBM need to be further explored.

In conclusion, this study used WGCNA to construct co-expression modules related to the survival of GBM patients. We identified the least conserved module and identified hub genes associated with poor prognosis in patients of GBM. We focused on the potential pathways and molecular mechanisms of CDCA5 and CDCA8 regulation of cell cycle in the p53 pathway and their effects on the immune microenvironment. It provided new molecular targets and intervention strategy for improving the prognosis of GBM patients.

Abbreviations

GBM: glioblastoma; WGCNA: weighted gene co-expression network analysis; TIMER: Tumor Immune Estimation Resource; TCGA: The Cancer Genome Atlas; CDCA5: cell division cycle associated 5; CDCA8: cell division cycle associated 8; MAD: Median Absolute Deviation; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; TOM: topological overlap matrix; OS: overall survival; CDK1:cyclin-dependent kinase 1.

Declarations

Acknowledgments

This work benefited from the database of TCGA and GEO. We are grateful for the access to the resources and the efforts of the staff to expand and improve the databases.

Authors' contributions

JZ, HG and LKL designed and conceived the study. FPZ, SLH and ZG searched and collected the majority of the data. ZW and WWZ performed the bioinformatics analyses. JZ and HG were major contributors in writing the manuscript. All authors have read and approved the final version of this manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (No.81904183), the Project of Health Commission of Shanxi Province (No. 2018078), the Natural Science Foundation of Shanxi Province (No. 201801D121300)the cultivation project of Science and Technology Innovation Ability of Shanxi University of Chinese Medicine (2019PY-017) and the Project of Shanxi Province Academy of Traditional Chinese Medicine (201901).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that there are no conflicts of interest.

Author details

[1]Department of oncology, Shanxi Province Academy of Traditional Chinese Medicine, Shanxi Province Hospital of Traditional Chinese Medicine, Taiyuan, Shanxi 030012, China

[2]Department of National Traditional Chinese Medicine Clinical Research Base, Shanxi Province Hospital of Traditional Chinese Medicine, Taiyuan, Shanxi 030012, China

[3]Department of Anesthesiology, Shanxi provincial people's Hospital, Taiyuan 030000, China

References

1.Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P and Ellison DW. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol. 2016; 131(6):803-820. https://doi.org/10.1007/s00401-016-1545-1

2.Morgan LL. The epidemiology of glioma in adults: a "state of the science" review. Neuro Oncol. 2015; 17(4):623-624. https://doi.org/10.1093/neuonc/nou358

3.Ostrom QT, Gittleman H, Farah P, Ondracek A, Chen Y, Wolinsky Y, Stroup NE, Kruchko C and Barnholtz-Sloan JS. CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 2006-2010. Neuro Oncol. 2013; 15 Suppl 2:ii1-56. https://doi.org/10.1093/neuonc/not151

4.Ryken TC, Kalkanis SN, Buatti JM, Olson JJ and Committee ACJG. The role of cytoreductive surgery in the management of progressive glioblastoma : a systematic review and evidence-based clinical practice guideline. J Neurooncol. 2014; 118(3):479-488. https://doi.org/10.1007/s11060-013-1336-7

5.Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJ, Janzer RC, Ludwin SK, Allgeier A, Fisher B, Belanger K, Hau P, Brandes AA, Gijtenbeek J, et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. Lancet Oncol. 2009; 10(5):459-466. https://doi.org/10.1016/S1470-2045(09)70025-7

6.Weller M, van den Bent M, Hopkins K, Tonn JC, Stupp R, Falini A, Cohen-Jonathan-Moyal E, Frappaz D, Henriksson R, Balana C, Chinot O, Ram Z, Reifenberger G, et al. EANO guideline for the diagnosis and treatment of anaplastic gliomas and glioblastoma. Lancet Oncol. 2014; 15(9):e395-403.

https://doi.org/10.1016/S1470-2045(14)70011-7

7.Dong P, Yu B, Pan L, Tian X and Liu F. Identification of Key Genes and Pathways in Triple-Negative Breast Cancer by Integrated Bioinformatics Analysis. Biomed Res Int. 2018; 2018:2760918.

https://doi.org/10.1155/2018/2760918

8.Zhou Y, Yang L, Zhang X, Chen R, Chen X, Tang W and Zhang M. Identification of Potential Biomarkers in Glioblastoma through Bioinformatic Analysis and Evaluating Their Prognostic Value. Biomed Res Int. 2019; 2019:6581576. https://doi.org/10.1155/2019/6581576

9.Zhuang L, Yang Z and Meng Z. Upregulation of BUB1B, CCNB1, CDC7, CDC20, and MCM3 in Tumor Tissues Predicted Worse Overall Survival and Disease-Free Survival in Hepatocellular Carcinoma Patients. Biomed Res Int. 2018; 2018:7897346. https://doi.org/10.1155/2018/7897346

10.Langfelder P and Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559. https://doi.org/10.1186/1471-2105-9-559

11.Zhang B and Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005; 4:Article17. https://doi.org/10.2202/1544-6115.1128

12.Wan Q, Tang J, Han Y and Wang D. Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma. Exp Eye Res. 2018; 166:13-20. https://doi.org/10.1016/j.exer.2017.10.007

13.Org E, Blum Y, Kasela S, Mehrabian M, Kuusisto J, Kangas AJ, Soininen P, Wang Z, Ala-Korpela M, Hazen SL, Laakso M and Lusis AJ. Relationships between gut microbiota, plasma metabolites, and metabolic syndrome traits in the METSIM cohort. Genome Biol. 2017; 18(1):70. https://doi.org/10.1186/s13059-017-1194-2

14.Shi K, Bing ZT, Cao GQ, Guo L, Cao YN, Jiang HO and 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. https://doi.org/10.3980/j.issn.2222-3959.2015.02.10

15.van Dam S, Vosa U, van der Graaf A, Franke L and de Magalhaes JP. Gene co-expression analysis for functional classification and gene-disease predictions. Brief Bioinform. 2018; 19(4):575-592. https://doi.org/10.1093/bib/bbw139

16.Yu G, Wang LG, Han Y and He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16(5):284-287. https://doi.org/10.1089/omi.2011.0118

17.Mizuno H, Kitada K, Nakai K and Sarai A. PrognoScan: a new database for meta-analysis of the prognostic value of genes. BMC Med Genomics. 2009; 2:18. https://doi.org/10.1186/1755-8794-2-18

18.Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B and Liu XS. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer Res. 2017; 77(21):e108-e110. https://doi.org/10.1158/0008-5472.CAN-17-0307

19.Furnari FB, Fenton T, Bachoo RM, Mukasa A, Stommel JM, Stegh A, Hahn WC, Ligon KL, Louis DN, Brennan C, Chin L, DePinho RA and Cavenee WK. Malignant astrocytic glioma: genetics, biology, and paths to treatment. Genes Dev. 2007; 21(21):2683-2710. https://doi.org/10.1101/gad.1596707

20.Sette P, Amankulor N, Li A, Marzulli M, Leronni D, Zhang M, Goins WF, Kaur B, Bolyard C, Cripe TP, Yu J, Chiocca EA, Glorioso JC, et al. GBM-Targeted oHSV Armed with Matrix Metalloproteinase 9 Enhances Anti-tumor Activity and Animal Survival. Mol Ther Oncolytics. 2019; 15:214-222. https://doi.org/10.1016/j.omto.2019.10.005

21.Grossman SA, Ye X, Piantadosi S, Desideri S, Nabors LB, Rosenfeld M, Fisher J and Consortium NC. Survival of patients with newly diagnosed glioblastoma treated with radiation and temozolomide in research studies in the United States. Clin Cancer Res. 2010; 16(8):2443-2449. https://doi.org/10.1158/1078-0432.CCR-09-3106

22.Network TC. Corrigendum: Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2013; 494(7438):506. https://doi.org/10.1038/nature11903

23.Montemurro N. Glioblastoma Multiforme and Genetic Mutations: The Issue Is Not Over Yet. An Overview of the Current Literature. J Neurol Surg A Cent Eur Neurosurg. 2020; 81(1):64-70. https://doi.org/10.1055/s-0039-1688911

24.Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008; 455(7216):1061-1068. https://doi.org/10.1038/nature07385

25.Chou WC, Cheng AL, Brotto M and Chuang CY. Visual gene-network analysis reveals the cancer gene co-expression in human endometrial cancer. BMC Genomics. 2014; 15:300. https://doi.org/10.1186/1471-2164-15-300

26.Xu J, Zhu C, Yu Y, Wu W, Cao J, Li Z, Dai J, Wang C, Tang Y, Zhu Q, Wang J, Wen W, Xue L, et al. Systematic cancer-testis gene expression analysis identified CDCA5 as a potential therapeutic target in esophageal squamous cell carcinoma. EBioMedicine. 2019; 46:54-65. https://doi.org/10.1016/j.ebiom.2019.07.030

27.Ci C, Tang B, Lyu D, Liu W, Qiang D, Ji X, Qiu X, Chen L and Ding W. Overexpression of CDCA8 promotes the malignant progression of cutaneous melanoma and leads to poor prognosis. Int J Mol Med. 2019; 43(1):404-412. https://doi.org/ 10.3892/ijmm.2018.3985

28.Tian Y, Wu J, Chagas C, Du Y, Lyu H, He Y, Qi S, Peng Y and Hu J. CDCA5 overexpression is an Indicator of poor prognosis in patients with hepatocellular carcinoma (HCC). BMC Cancer. 2018; 18(1):1187. https://doi.org/ 10.1186/s12885-018-5072-4

29.Bi Y, Chen S, Jiang J, Yao J, Wang G, Zhou Q and Li S. CDCA8 expression and its clinical relevance in patients with bladder cancer. Medicine (Baltimore). 2018; 97(34):e11899. https://doi.org/10.1097/MD.0000000000011899

30.Bu Y, Shi L, Yu D, Liang Z and Li W. CDCA8 is a key mediator of estrogen-stimulated cell proliferation in breast cancer cells. Gene. 2019; 703:1-6. https://doi.org/10.1016/j.gene.2019.04.006

31.Shen Z, Yu X, Zheng Y, Lai X, Li J, Hong Y, Zhang H, Chen C, Su Z and Guo R. CDCA5 regulates proliferation in hepatocellular carcinoma and has potential as a negative prognostic marker. Onco Targets Ther. 2018; 11:891-901. https://doi.org/10.2147/OTT.S154754

32.Phan NN, Wang CY, Li KL, Chen CF, Chiao CC, Yu HG, Huang PL and Lin YC. Distinct expression of CDCA3, CDCA5, and CDCA8 leads to shorter relapse free survival in breast cancer patient. Oncotarget. 2018; 9(6):6977-6992. https://doi.org/10.18632/oncotarget.24059

33.Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MDM, Miller CA, Welch JS, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013; 502(7471):333-339. https://doi.org/10.1038/nature12634

34.Soussi T, Ishioka C, Claustres M and Beroud C. Locus-specific mutation databases: pitfalls and good practice based on the p53 experience. Nat Rev Cancer. 2006; 6(1):83-90. https://doi.org/10.1038/nrc1783

35.Stracquadanio G, Wang X, Wallace MD, Grawenda AM, Zhang P, Hewitt J, Zeron-Medina J, Castro-Giner F, Tomlinson IP, Goding CR, Cygan KJ, Fairbrother WG, Thomas LF, et al. The importance of p53 pathway genetics in inherited and somatic cancer genomes. Nat Rev Cancer. 2016; 16(4):251-265. https://doi.org/10.1038/nrc.2016.15

36.Buschges R, Weber RG, Actor B, Lichter P, Collins VP and Reifenberger G. Amplification and expression of cyclin D genes (CCND1, CCND2 and CCND3) in human malignant gliomas. Brain Pathol. 1999; 9(3):435-442; discussion 432-433. https://doi.org/10.1111/j.1750-3639.1999.tb00532.x

37.Betticher DC, Heighway J, Hasleton PS, Altermatt HJ, Ryder WD, Cerny T and Thatcher N. Prognostic significance of CCND1 (cyclin D1) overexpression in primary resected non-small-cell lung cancer. Br J Cancer. 1996; 73(3):294-300. https://doi.org/10.1038/bjc.1996.52

38.Mao P, Bao G, Wang YC, Du CW, Yu X, Guo XY, Li RC and Wang MD. PDZ-Binding Kinase-Dependent Transcriptional Regulation of CCNB2 Promotes Tumorigenesis and Radio-Resistance in Glioblastoma. Transl Oncol. 2019; 13(2):287-294. https://doi.org/10.1016/j.tranon.2019.09.011

39.Qian X, Song X, He Y, Yang Z, Sun T, Wang J, Zhu G, Xing W and You C. CCNB2 overexpression is a poor prognostic biomarker in Chinese NSCLC patients. Biomed Pharmacother. 2015; 74:222-227. https://doi.org/10.1016/j.biopha.2015.08.004

40.Raab M, Kobayashi NF, Becker S, Kurunci-Csacsko E, Kramer A, Strebhardt K and Sanhaji M. Boosting the apoptotic response of high-grade serous ovarian cancers with CCNE1 amplification to paclitaxel in vitro by targeting APC/C and the pro-survival protein MCL-1. Int J Cancer. 2020; 146(4):1086-1098. https://doi.org/10.1002/ijc.32559

41.Jiang B, Wu D, Huang L and Fang H. miR-424-5p inhibited malignant behavior of colorectal cancer cells by targeting CCNE1. Panminerva Med. 2019. https://doi.org/ 10.23736/S0031-0808.19.03708-X

42.Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, Coussens LM, Gabrilovich DI, Ostrand-Rosenberg S, Hedrick CC, Vonderheide RH, Pittet MJ, Jain RK, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med. 2018; 24(5):541-550. https://doi.org/10.1038/s41591-018-0014-x

43.Senbabaoglu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, Miao D, Ostrovnaya I, Drill E, Luna A, Weinhold N, Lee W, Manley BJ, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. Genome Biol. 2016; 17(1):231. https://doi.org/10.1186/s13059-016-1092-z

Figures

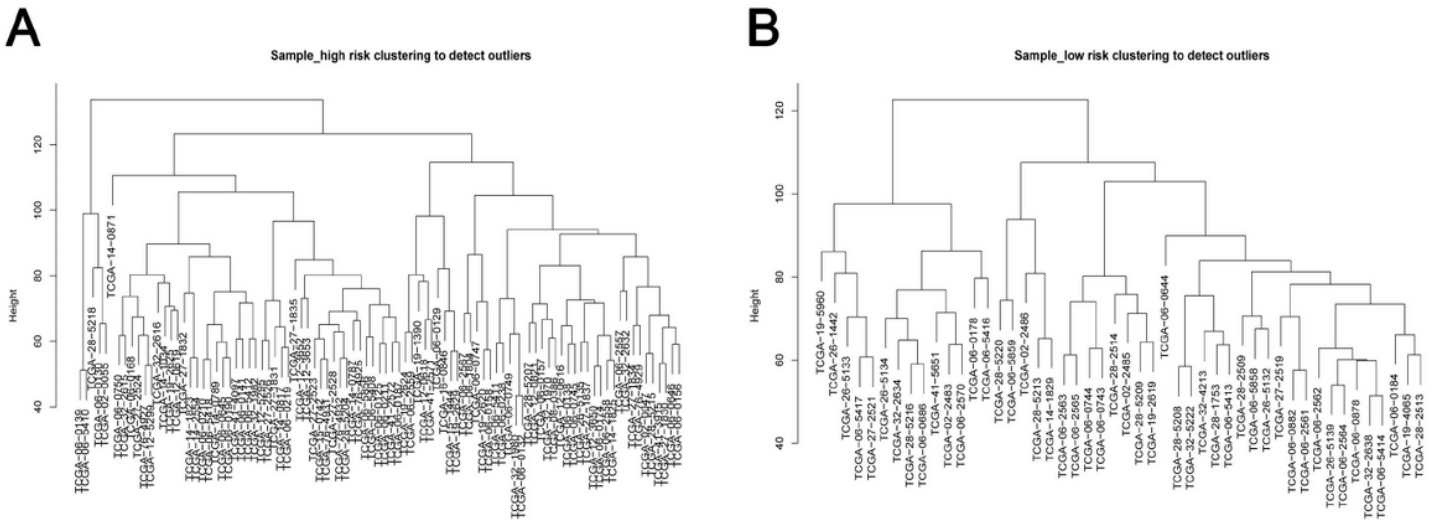

Figure 1

Clustering dendrogram of samples based on their Euclidean distance. The hclust function was used to detect outliers, resulting in clustering dendrograms with samples that were correctly sorted in random order. (A) The clustering dendrogram of high-risk samples to detect outliers. (B) The clustering dendrogram of low-risk samples to detect outliers.

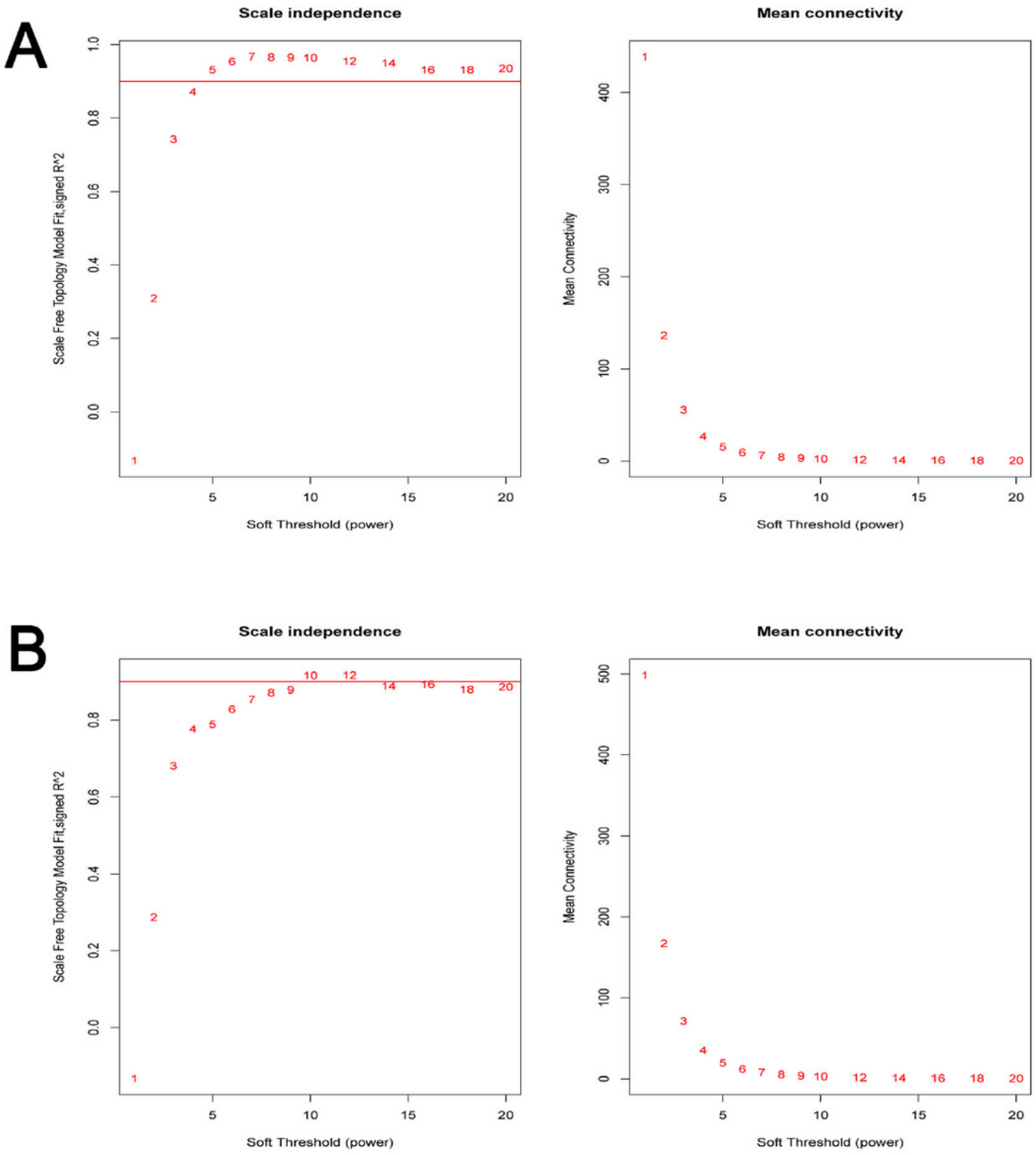

Figure 2

The optimal soft threshold power of the WGCNA was selected to make our genes distribution conforms to the scale-free network by calculating the scale-free topological fit index and the average connectivity. (A) The scale independence and the mean connectivity of the WGCNA analysis of the high-risk samples. (B) The scale independence and the mean connectivity of the WGCNA analysis of the low-risk samples.

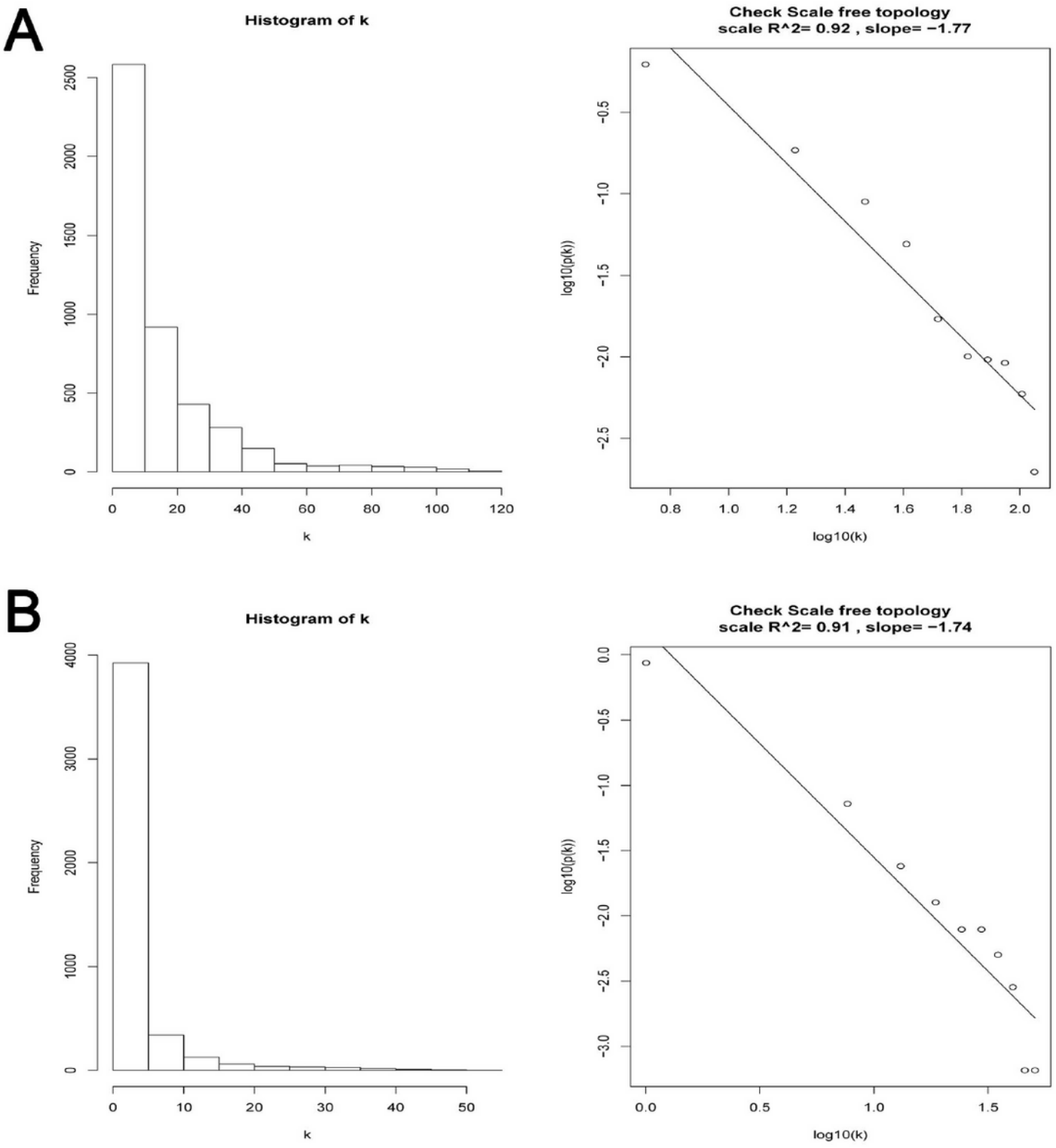

Figure 3

The optimal soft threshold power of the WGCNA was verified to build the scale-free network by drawing the histogram of k and calculating the correlation coefficient between k and p (k). (A) The histogram of k and the correlation coefficient between k and p (k) of the high-risk samples. (B) The histogram of k and the correlation coefficient between k and p (k) of the low-risk samples.

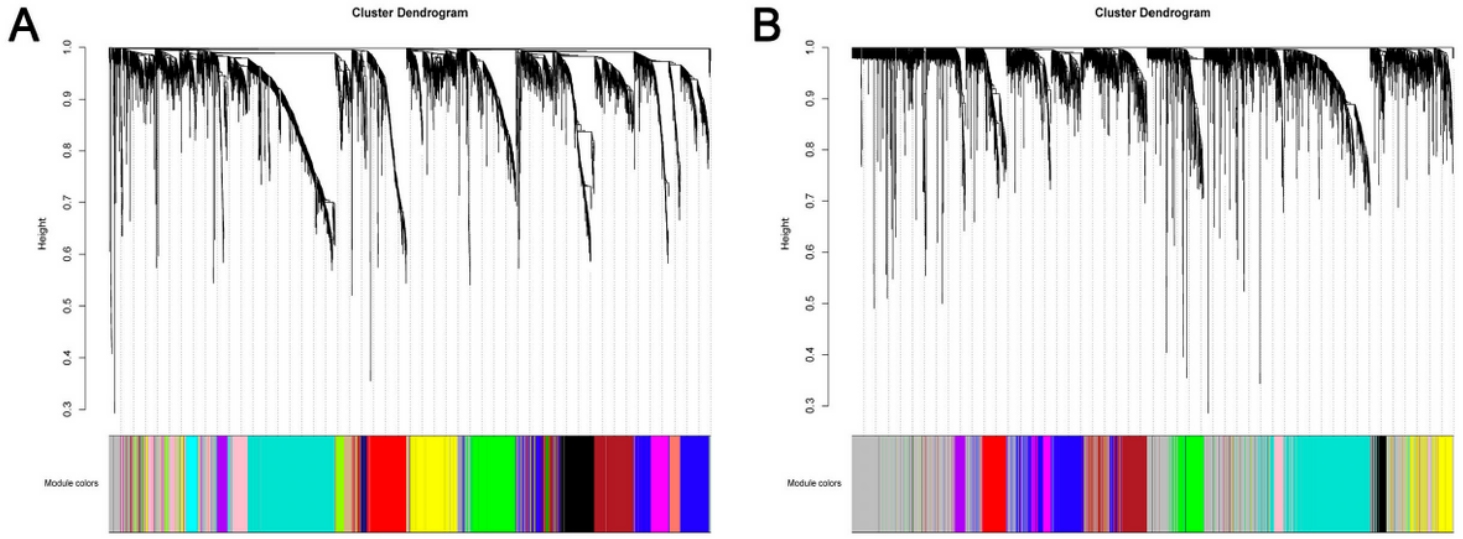

Figure 4

Clustering dendrograms and modules identified by WGCNA. The co-expression modules were constructed and were shown in different colors. (A) High-risk samples. (B) Low-risk samples.

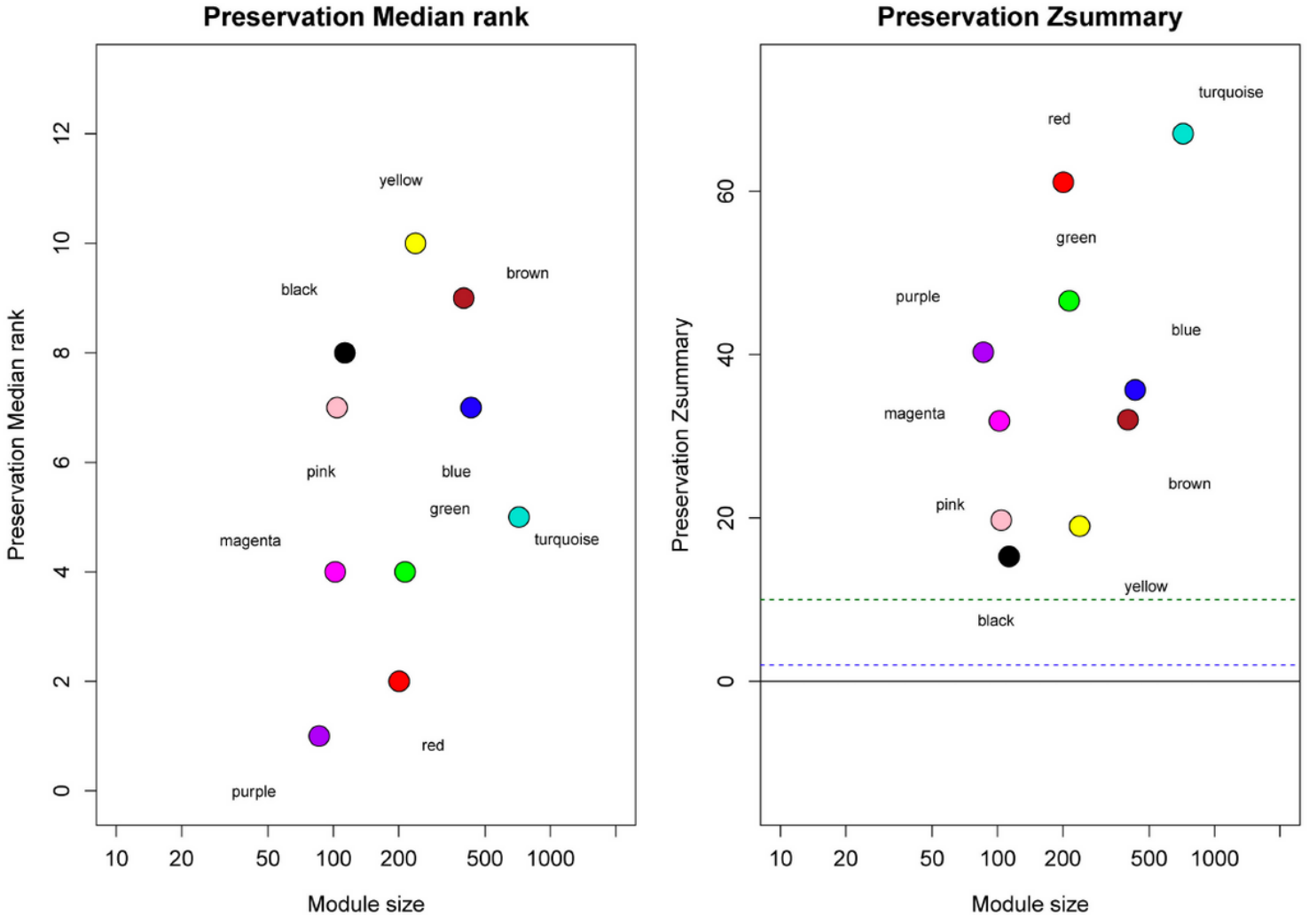

Figure 5

Characterization of the co-expression modules identified by WGCNA. (A) The preservation median rank of ten co-expression modules. (B) The preservation Z-summary score of ten co-expression modules.

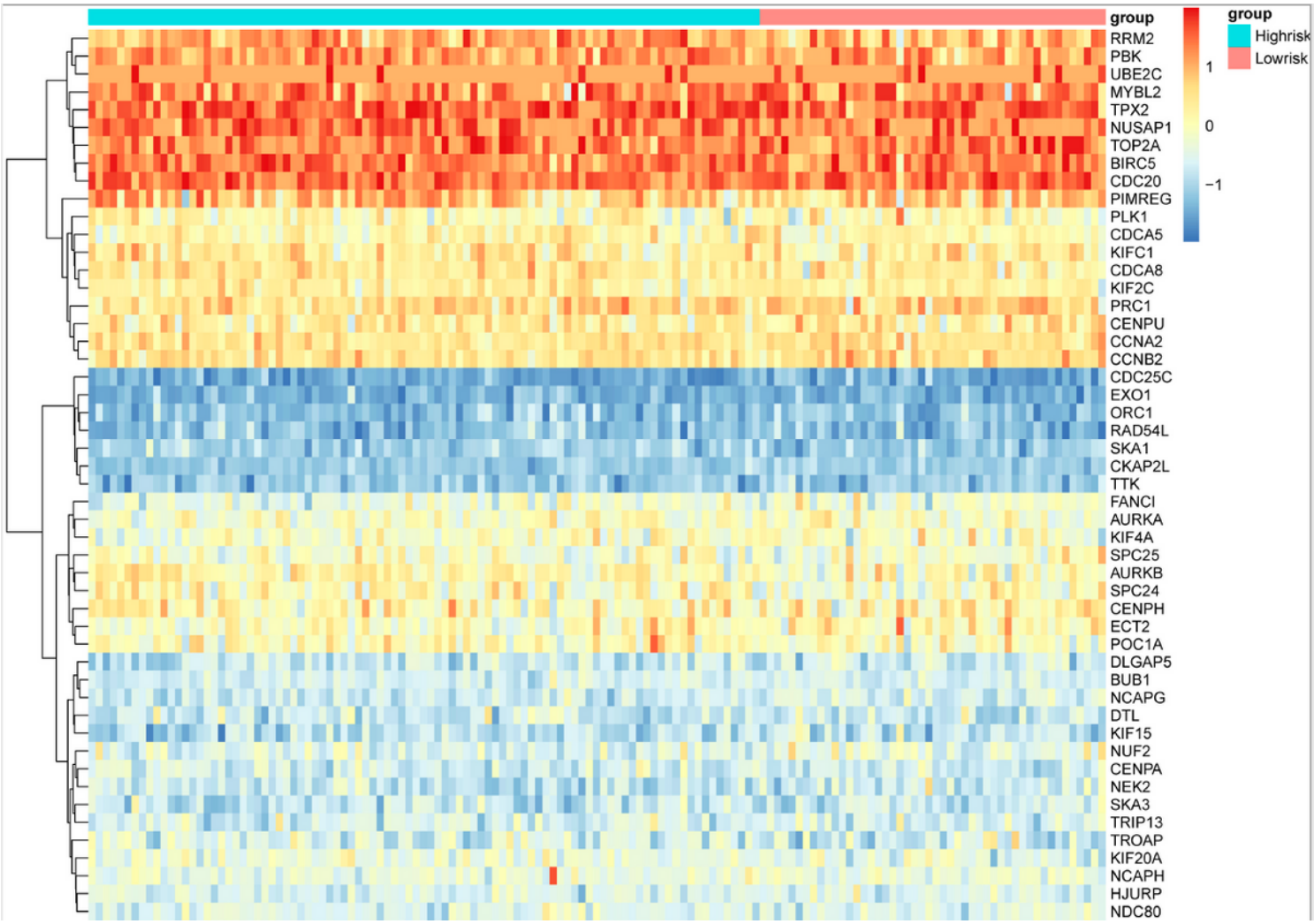

Figure 6

The heat map of the hub genes in the black module that were identified in the high-risk GBM patients.

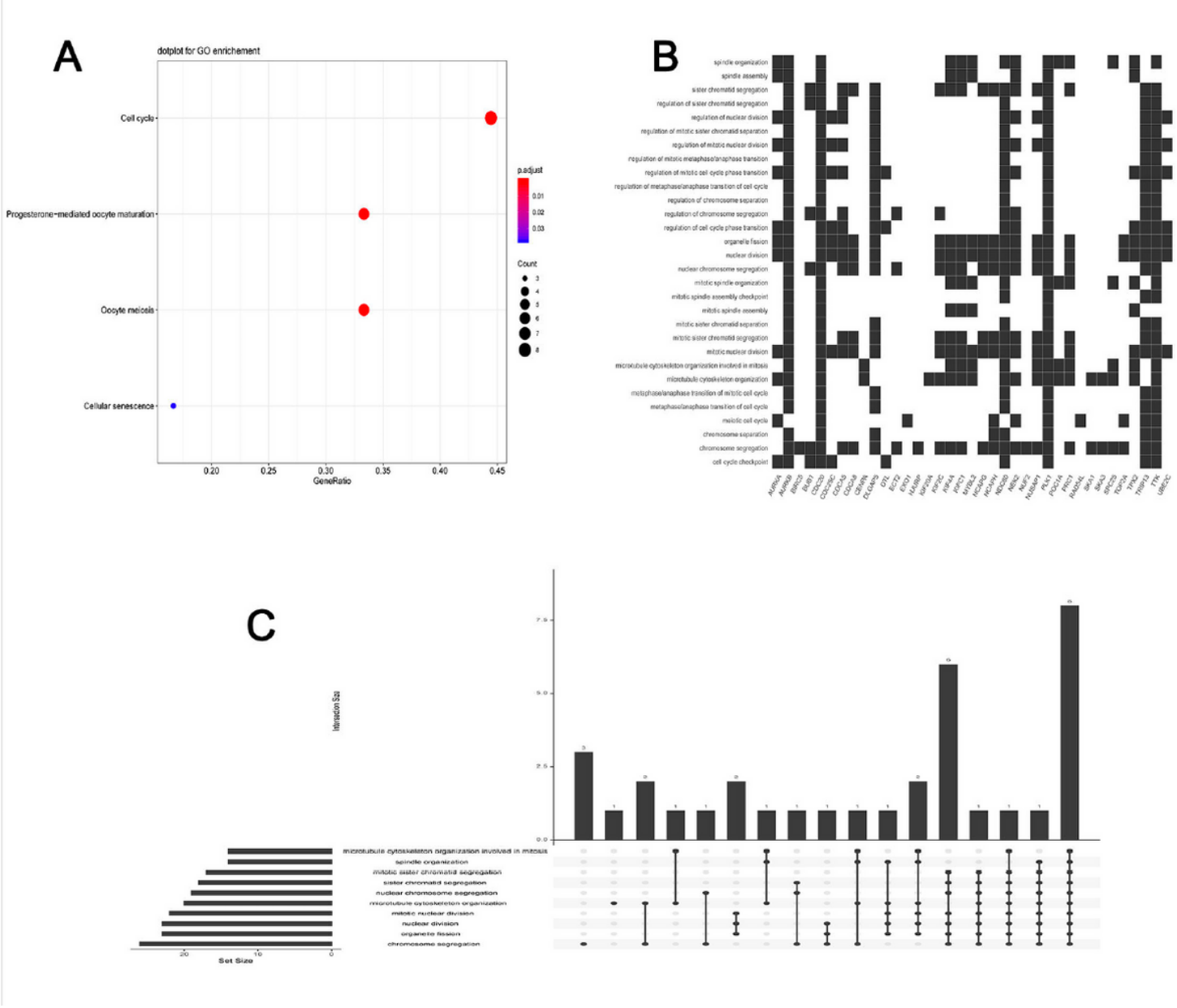

Figure 7

Go enrichment analysis of hub genes in the black module. (A) The bubble plot showed the top enriched biological process of these genes. (B) The scatter plot showed the distribution of hub genes involved in specific links in the cell cycle. (C) The UpSet plot showed the interactions among the ten links of cell cycle.

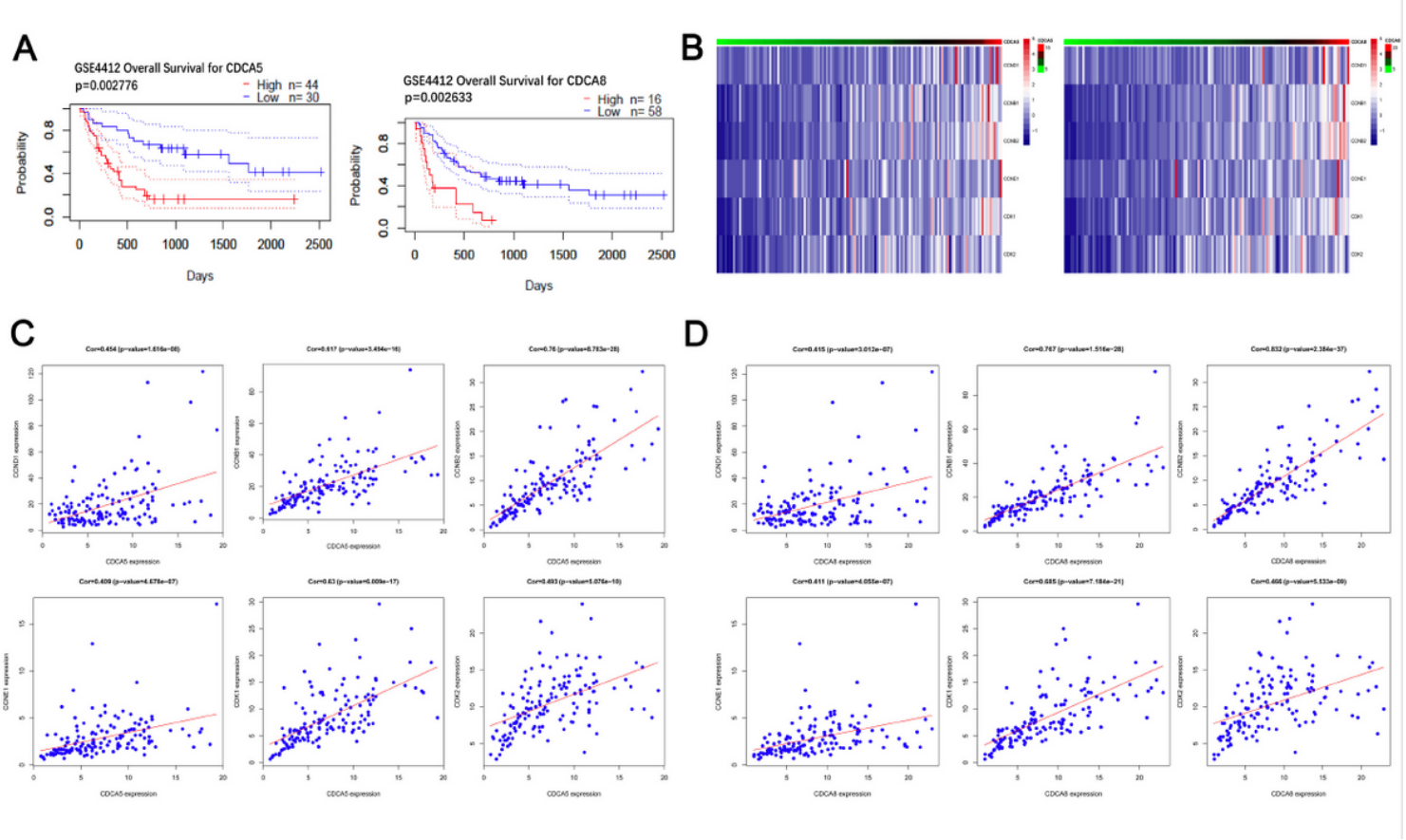

Figure 8

Single gene analysis to study the relationship between CDCA5 and CDCA8 and cell cycle related genes in the p53 pathway. (A) The overall survival rate of high- and low-risk patients based on the expression levels of CDCA5 and CDCA8 in the GSE4412 dataset. (B) The heatmaps of CDCA5 and CDCA8 and cell cycle related genes expression profiles in the p53 pathway. (C) The scatter plot showed the correlation between CDCA5 and cell cycle related genes in the p53 pathway. (D) The scatter plot showed the correlation between CDCA8 and cell cycle related genes in the p53 pathway.

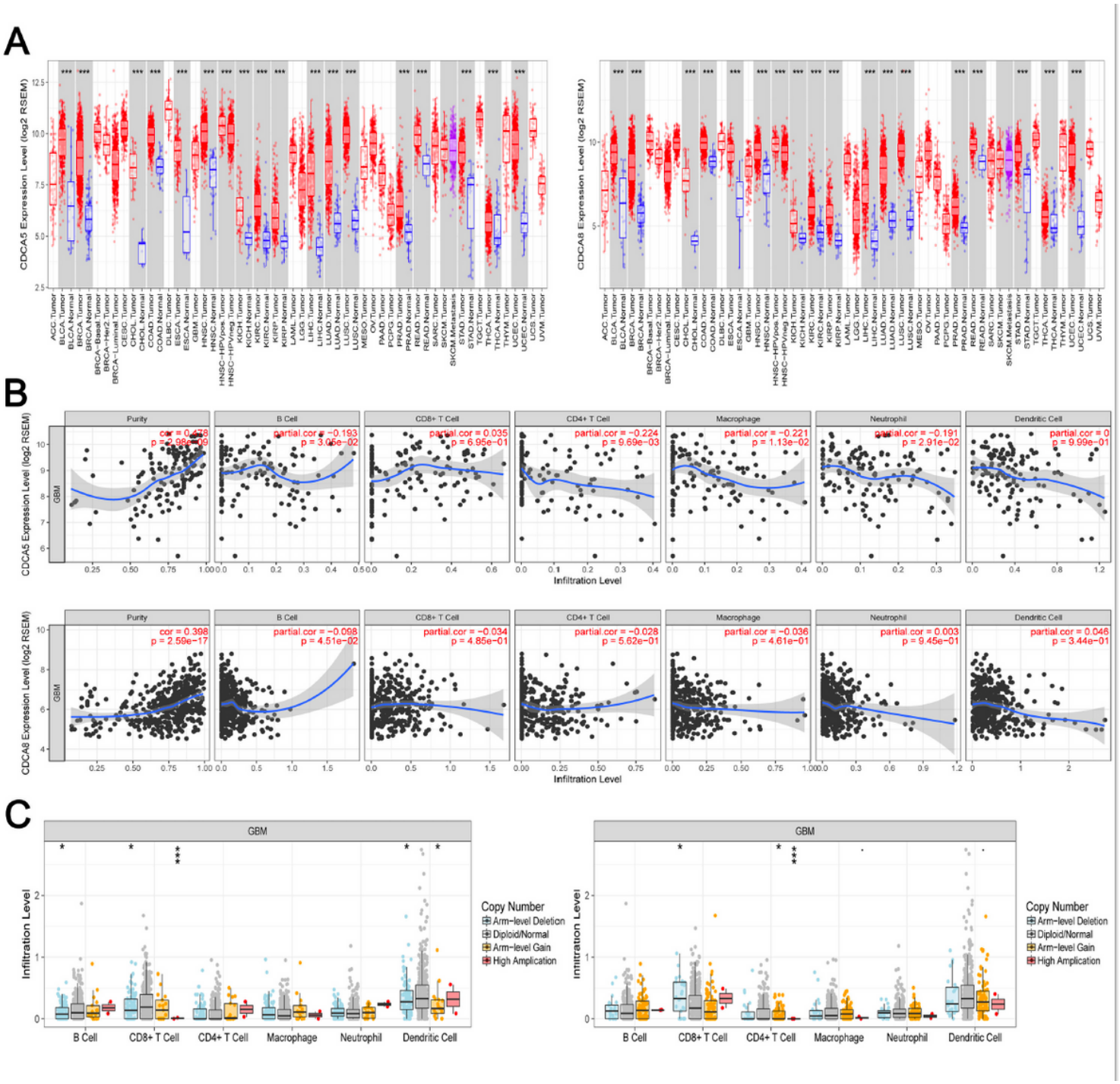

Figure 9

Single gene analysis to study the relationship between CDCA5 and CDCA8 and tumor microenvironment. (A) The differential expression of CDCA5 and CDCA8 in tumors and normal tissues of multiple cancer species. (B) The relationship between the expression levels of CDCA5 and CDCA8 and the six types of immune cells. (C) The relationship between expression levels and copy number variations of CDCA5 and CDCA8 and the levels of infiltration of six immune cells. *p<0.05, ***p<0.001.