

CLCA1 Has a Prognostic Indicator for the Colorectal Cancer by Weighted Gene Co-Expression Network Analysis

jiajng lin (✉ 627919928@qq.com)

First Affiliated Hospital of Fujian Medical University

lingzhi yang

First Affiliated Hospital of Fujian Medical University

suyong lin

First Affiliated Hospital of Fujian Medical University

zihua chen

First Affiliated Hospital of Fujian Medical University

shaoqin chen

First Affiliated Hospital of Fujian Medical University

Research Article

Keywords: Colorectal cancer, differential gene expression analysis, weighted gene co-expression network analysis, the differential co-expression genes, biomarkers

Posted Date: February 19th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-209748/v1>

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

[Read Full License](#)

CLCA1 has a prognostic indicator for the colorectal cancer by Weighted Gene Co-expression Network Analysis

jiating lin, lingzhi yang, suyong lin, zhihua chen,shaoqin chen

Department of Gastrointestinal Surgery, The First Affiliated Hospital of Fujian Medical University, Fuzhou, Fujian 350004, P.R. China

Correspondence to: Professor Shao Qin Chen, Department of Gastrointestinal Surgery, The First Affiliated Hospital of Fujian Medical University, 20 Chazhong Road, Taijiang, Fuzhou, Fujian 350004, P.R. China

E-mail: chenshaoqin1613@163.com

Abstract

Colorectal cancer (CRC) has become the second most common digestive tract tumor. Even though the means to treat colon cancer have improved, patients prognosis is low due to the lack of accurate molecular targets. Hence, it urgently demanded better biomarkers for prognosis and progression of colon cancer. This study explores the hub gene associated with the prognosis of colorectal cancer and further analyzes the hub gene function. In this study, all genes mRNA expression data were from the cancer genome atlas (TCGA) colon cancer database and the Gene Expression Omnibus (GEO). These databases were used to screen the differentially expressed co-genes between colon cancer tissue and normal tissue. Weighted Gene Co-expression Network Analysis screened out a total of 103 differential co-expression genes (WGCNA). According to the R cluster profile package annotation analysis, these genes biological functions mainly concentrate on energy metabolism. Moreover, in the protein-protein interaction (PPI) network, the CytoHubba plugin of Cytoscape was used to screen out ten genes (CLCA1,

ZG16, GUCA2B, GUCA2A, CLCA4, SLC26A3, MS4A12, GCG, SI, and NR1H4). According to the survival analysis results, high expression of CLCA1 has better overall survival and disease-free survival in patients with CRC. Simultaneously, the mRNA expression of CLCA1 in normal tissues was higher than that in CRC tissues. Besides, there were significant differences in the expression of CLCA1 in pathological stage, T stage, and M stage. By using a gene set enrichment analysis, we found several considerable enrichment pathways in the high-groups. CIBERSORT analysis for the proportion of TICs revealed that B-cell naive, dendritic cells, plasma cells, and CD4+ T cells were positively correlated with CLCA1 expression, suggesting that CLCA1 might be responsible for the preservation of immune-dominant status for TME. Finally, in the Human Protein Atlas (HPA) database, the protein level of CLCA1 in the colorectal cancer samples decreased, consistent with the down-regulation of the mRNA expression level CLCA1. To sum up, by integrating WGCNA with differential gene expression analysis, this research generated a significant survival correlative gene called CLCA1 that can predict prognosis prediction in colon cancer.

Keywords: Colorectal cancer, differential gene expression analysis, weighted gene co-expression network analysis, the differential co-expression genes, biomarkers

Introduce

Colorectal cancer (CRC) is the most fatally and commonly diagnosed cancer globally in gastrointestinal tumors [1]. Currently, the early colorectal adenocarcinoma can be surgically removed with a good prognosis. Although great progress has been made in diagnosis of CRC, the fatality rate of disease is very high. Some studies have shown that patients with stage I–III and stage IV CRC recurrence rate were 30% and 65% after curative treatment, respectively [2,3]. Therefore, understanding prognostic biomarkers may effectively contribute to guidance to selection of treatment and predicting survival rate. With the development of microarray technologies, gene expression profiles have been widely used to tumor research. Microarray analyses have related to the comparison between tumor and normal samples[4]. With the rapidly development of bioinformatics analysis, weighted gene co-expression network analysis (WGCNA) has become an emerging technique for analyzing

expressed genes (DEGs), but also high degree of interconnection between genes.[5]. In WGCNA, the basic concept is construction of co-expression modules, which are clusters of genes maintaining consistent expression patterns and even playing similar biological roles, and these modules are derived from the data of mRNA expression profiles by performing unsupervised hierarchical clustering [6,7]. Moreover, WGCNA analyzes the internal relationships between modules and sample traits, which provides a way to reveal the specific hidden mechanisms.[5]. WGCNA has been widely applied to identify gene modules associated with clinical parameters in many cancers. The usefulness of WGCNA in the detection of the colon cancer recurrence gene module has been demonstrated in many literatures.[7,8].

In the current study, WGCNA and differential gene expression analysis were used for analysis of the mRNA expression data of colon cancer, which from the TCGA and GEO databases. The authors explored development of colon cancer by intersecting functional enrichment, protein-protein interaction (PPI) analysis and survival analysis. By analyzing different co-expressed genes, this study further reveals the potential molecular mechanism of tumor recurrence and makes certain contribution to clinical diagnosis or treatment.

Material and methods

The flow chart of this study is shown in Figure 1. Specific steps will be shown in the following sub-heading.

Data download and preprocessing

In the TCGA database, we select the colon cancer category and download the gene expression profiles related to colon cancer. In addition, we downloaded gene expression profiles from the GEO database with colorectal cancer as the keyword. All clinical information and mRNA data related to CRC were obtained after the R package TCGA-biolinks [9]. There were 473 colorectal cancer and 41 normal tissues in the TCGA database. These data were preprocessed as follows: 1) Samples lacking clinical data were removed;

2) Samples with FPKM of 0 were excluded; 3) The CPM (count per million) > 1 gene was retained; 4) Samples with OS < 1 month were removed. Moreover, the expression profiles of GSE44861 were obtained using the R package GEOquery[10]. The GSE44861 included 49 tumor samples and 48 paired normal tissues. These data were preprocessed as follows: 1) Annotate the relationship between chip probes and the human gene SYMBOL using the Bioconductor package; 2) Samples with OS < 1 month were removed.

Application of WGCNA in identifying key Co-expression modules

Through the WGCNA package, we constructed gene co-expression networks based on the data in TCGA and GSE44861, respectively [11]. Firstly, the Pearson correlation analysis is carried out with all pairwise genes to construct the co-expression matrix. Secondly, we converted the Pearson correlation matrix to an adjacency matrix (scale-free network). To decide the most appropriate β value, we calculated the scale-free fit index and meant connectivity for each supposed β from 3 to 20. To select the proper β value, we set β (from 3 to 20) to calculate the scale-free fit index. Next, transform the adjacency matrix to TOM (topological overlap matrix), reducing the false correlation. Finally, we divided the gene into several co-expression modules by different TOM values. To further identify functional modules, based on the algorithm, the characteristic correlation of clinical modules is constructed[12]. To better understand the logic, we found a more detailed method of the WGCNA [12].

Differential Expression Analysis and Interaction with the Modules of Interest

The R package limma has a powerful function, which can label the biological function of genes and screen different genes [13]. Firstly, we used a limma package to screen out the differentially expressed genes in the TCGA and GSE44861 datasets. The inclusion criteria for screening are as follows: 1) cut-off value = $|\log FC| \geq 1.0$; 2) $P < 0.05$. The

screened genes were then shown using volcanic maps[14]. Finally, the obtained DEGs gene and co-expression genes were analyzed interactively to obtain the potential prognostic genes overlapping part [15].

Functional Annotation for Genes of Interest

In order to reveal the biological functions and signaling pathways of the screened genes, Gene Ontology (GO)[16] enrichment analysis and Kyoto Encyclopedia of Genes and Genomes[17] were fully utilized. With cluster Profiler package in R package, we carried out enrichment analysis on genes of Interest, and the results were shown in bubble diagrams. The enrichment standard was set as $p < 0.05$. Among the many products, we mainly focus on the result of the biological process.

Construction of PPI and Screening of Hub Genes

After enrichment analysis, the genes were screened by a visual network model to determine the hub genes. Firstly, we input the genes into the STRING online tool to build a network model. Secondly, we choose the genes with a score ≥ 0.4 to build a network model visualized by Cytoscape (v3.7.2) [18]. Then, the best central node was found in the co-expression network. Finally, the hub genes were the top ten in MCC value, which was calculated by CytoHubba.

Verification of the Expression Patterns and the Prognostic Values of Hub Genes

Survival analysis of the hub genes was performed to verify the prognostic value of the genes. In the TCGA database, we use the survival package under the R environment to obtain relevant survival analysis data. We divided the patients into the high expression group and the low expression group to explore the relationship between the two groups

overall survival (OS). Only patients with completed follow-up times were included for survival analysis. We used the Kaplan-Meier method to conduct a bilateral logarithmic test to explore the differences between the two groups, and $P < 0.05$ was considered statistically significant. Moreover, the association between disease-free survival (DFS) and hub genes expressed in CRC patients was determined using the online tool GEPIA2 [19].

Validation of Protein Expressions of Survival-Related Hub Genes by the HPA Database

The Human Protein Atlas (<https://www.proteinatlas.org>) is a website that contains immunohistochemistry-based expression data for near 20 widespread kinds of cancers, and each tumor type includes 12 individual tumors[20]. In this study, the differences in CLCA1 in normal and CRC tissues were compared by immunohistochemical images.

Gene Set Enrichment Analysis

To understand the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of CLCA1 signature, a gene set enrichment analysis (GSEA) was used to analyze the enrichment terms in the entire TCGA cohort. The GSEA version was 4.11, and meeting $NOM\ p < 0.05$ and $FDR\ q < 0.25$ were set to be statistically significant.

TICs Profile

CIBERSORT (<https://cibersort.stanford.edu/>) was used to estimate the abundance of TIC in all tumor samples. Quality filtering criteria were $p < 0.05$, and detailed methods were detailed in this study.

Statistical Analysis

Mann-Whitney U test tested the difference in CLCA1 expression between CRC tissues and adjacent nontumor tissues. The differences in CLCA1 among multiple groups were compared by Kruskal–Wallis test. Chi-square (χ^2) test was used to evaluate the interrelation between CLCA1 expression and clinicopathological parameters. Kaplan–Meier analysis and logrank test were used to compare the significant differences in survival rates between the high- and the low- CLCA1 expression groups. All statistical analyses were performed with R software (version 4.03), and $P < 0.05$ was used to determine the significance level.

RESULTS

Weighted Gene Co-expression Modules construction

In order to move forward a single step to discover the functional clusters in CRC patients, the gene co-expression networks were made up of the TCGA-CRC and GSE44861 datasets with the WGCNA package. TCGA-CRC dataset and GSE44861 dataset were divided into 6 modules (Figure 2A) and 15 modules (Figure 3A) respectively. In order to verify the relationship between each module and two clinical traits (cancer and normal), we plotted the adjacency heatmap of all genes. As was shown in Figure 2B, 3B, different colors represent different modules, the results of the module-trait that the green module in the TCGA-CRC and blue module in the GSE44861 were found to have the highest association with normal tissues (green module: $r = 0.8$, $p = 1e-10$; blue module: $r = 0.68$, $p = 1e-14$).

The DEG analysis and identification of CRC relation Modules

Base on the cut-off criteria of $|\log FC| \geq 1.0$ and $\text{adj. } P < 0.05$, 1135 DEGs in the TCGA dataset (Figure 4A) and 278 DEGs in the GSE44861 dataset (Figure 4B) were identified by the Limma-package. Differential expression analysis was performed on the RNA-

sequencing data from TCGA and GEO. In Figure 4C which pointed out that, 1154 co-expression genes which were identified in the green module of TCGA dataset and 531 co-expression genes which were identified in the blue module of GSE44861. In the end, 103 overlapping genes were extracted from co-expression modules (Figure 4C).

Go and KEGG pathway Enrichment Analyses for the 103 Genes

To explore the biological functions and signal pathways, we carried out GO and KEGG pathway enrichment analyses on genes in CRC related. Figure 5 represents the top 10 terms of GO-BP and KEGG enrichment analyses. In the GO-BP enrichment analysis (Figure 5A), 103 genes were mainly concentrated in steroid metabolic process (gene count = 11, $p = 1.18E-06$) and primary alcohol metabolic process (gene count = 4, $p = 8.94E-07$). In the KEGG pathway enrichment analysis (Figure 5B) were mainly enriched in Nitrogen metabolism (gene count = 5, $p = 1.55E-07$) and Ascorbate and aldarate metabolism (gene count = 3, $p = 1.22E-03$).

PPI Network Construction and Hub Genes Identification

The PPI network of overlapping genes was established by Using the STRING database (Figure 6A). The centrality of genes was analyzed in the sub-network with MCC method. The results of the hub genes from the PPI network were shown in Figure 6B. The results are shown in Figure 8 and the top ten highest-scored genes are highlighted. These including CLCA1, ZG16, GUCA2B, GUCA2A, CLCA4, SLC26A3, MS4A12, GCG, SI, NR1H4, were selected as the hub genes (Figure 6B).

Verification of the Prognostic Values and Protein Expression of Hub Genes

We did a survival analysis of the ten hub genes through the use of Kaplan–Meier

plotter which took advantage of the R survival package. The results showed that high gene expression of CLCA1, GCG, and GUCA2A was positively correlated with longer survival time ($P < 0.05$) (Figure 7A, 7B, 7C). In addition, the DFS analysis of ten hub genes was performed by the GEPIA2 database. The results showed that high gene expression of CLCA1 was positively correlated with longer survival time ($P < 0.05$) (Figure 8A). Moreover, the protein levels of CLCA1 in colon tumor tissues are lower than those in normal tissues in the HPA database (Figure 10). Therefore, it is concluded that high-expression of CLCA1 is associated with good prognosis and higher overall survival in CRC patients.

The Difference in CLCA1 Expression in CRC

In TCGA database, there were 514 tissues (including 473 CRC tissues and 41 adjacent nontumor tissues) associated with the mRNA expression level of CLCA1 gene. As shown in this scatter plot (Figure 9A), mRNA expression profiles of CLCA1 in colon tumor tissues was significantly down-regulated compared with normal tissues ($P < 0.001$). Furthermore, the different expression group of CLCA1 was analyzed. The results shown that different expression group is associated with pathological stage ($P < 0.05$, Figure 9B), T stage ($P < 0.05$, Figure 9C), and M stage ($P < 0.05$, Figure 9E).

Identification of CLCA1-Related Signaling Pathways by GSEA

The function of CLCA1 was explored through the mining of TCGA database. In addition, CLCA1-related signaling pathways were explored through GSEA. According to NES, FDR q-value, and nominal p-value, the signal pathways with obvious enrichment were selected. Figure 11 exhibits the top-10 up-regulated in the high-expression group of CLCA1. The results revealed that the CLCA1 was related to biological processes of "Ascorbate and aldarate metabolism", "Butanoate metabolism", "Citrate cycle tca cycle", "Fatty acid metabolism", "Glycolysis gluconeogenesis", and "peroxisome" in colon.

Correlation of CLCA1 With the Proportion of TICs

In order to study the correlation between CLCA1 expression and immune microenvironment, the proportion of tumor filtered subset of the immune was analyzed by making full use of CIBERSORT algorithm. In the colorectal cancer samples, we finally constructed 22 kinds of immune cells profiles (Figure 12). On the basis of the results the difference and correlation analyses revealed that there were 6 kinds of TICs associated with the expression of CLCA1 (Figure 13). Among them (B-cell naive, dendritic cells resting, plasma cells, T cells CD4 memory resting, macrophage M0 and NK cells), there are 4 kinds of TICs were positively associated with the expression of CLCA1, including B-cell naive, dendritic cells resting, plasma cells, T cells CD4 memory resting; and 2 kinds of TICs were negatively correlated with CLCA1 expression, including macrophage M0 and NK cells resting.

DISCUSSION

In recent years, Colon cancer has become the second most common digestive tract tumor[21]. The mechanisms underlying colon cancer development and progression remain unclear. Even though the means to treat colon cancer has improved, the prognosis of patients is poor due to the lack of accurate molecular targets. Hence, it was urgently demand better biomarkers for prognosis and progression of colon cancer.

In this paper, bioinformatics analysis methods were fully utilized to identify the 103 significant gene, which has the same expression trends in the TCGA and GSE44861 databases. According to KEGG and GO results, these genes were mainly concentrated in tumor energy metabolism. Cell energy metabolism is divided into aerobic oxidation and glycolysis. In most cases, tumor cells prefer to generate energy through glycolysis under aerobic conditions, which is the basis for tumor cell proliferation and invasion.[22-24]. In addition, 10 genes were screened out by the MCC scores (namely CLCA1, ZG16, GUCA2B, GUCA2A, CLCA4, SLC26A3, MS4A12, GCG, SI, NR1H4).

Among them, CLCA1 high expression, and that was highly relevant to good overall

survival in colon cancers. Furthermore, CLCA1 expression levels were different in groups classified including the basis of pathological stage, T stage, and M stage. The expression of CLCA1 increased with high differentiation, and which was downregulated with rising tumor stage. Then, we revealed several significantly enriched pathways by the GSEA. In the patient with high-risk scores subgroup, these pathways are primarily enriched in the “Ascorbate and aldarate metabolism”, “Butanoate metabolism”, “Citrate cycle tca cycle”, “Fatty acid metabolism”, “Glycolysis gluconeogenesis”, and “peroxisome” in cancer (Figure 5A). From this perspective, high-risk patients may benefit more from metabolic therapy. However, we need to explore the relationship further.

In addition, the Immune infiltration analysis for CLCA1 was carried out. As was known to all, CACA1 was closely related to the functions of B lymphocytes. Previous studies have shown a positive correlation between naive B cells and CACA1 expression in asthma though the CIBERSORT analysis for the proportion of TICs. Ching have found that epithelial cells could release CLCA1 to activate airway macrophages and eventually induce the pro-inflammatory response in the airway by inducing IL-1 β , IL-6, TNF- α and IL-8 [25]. However, it has been shown that the non-hydrolase mutant CLCA1 retains the ability to activate macrophages.[26]. Although these findings indicate that CLCA1 is closely related to the tumorigenesis mechanism, the relationship is still further study

Calcium-activated chloride channel (CLCA) regulators are proteins with a symmetrical homocysteine motif in the terminal tail of the amino group[27]. The loci of the human CLCA genes are at chromosome 1p31-1p22 [28]. Calcium-activated chloride channel regulator 1 (CLCA1) is a group of secreted self-cleaving protein that activates calcium-dependent chloride channels. Recently, the function of CLCA1 in metalloprotease property and involvement in mucus homeostasis and immune modulation has become a research hotspot. More and more studies have shown that CLCA1 may be involved in the pathophysiological process of colon cancer. There is growing interest in utilizing CLCA1 as a diagnostic, prognostic and predictive biomarker, as well as a potential therapeutic target. Whether CLCA1 can be used as a potential target for the treatment of colon cancer has attracted wide attention[29]. Knockdown of CLCA1 in Caco-2 cell line has been shown to inhibit cell differentiation and promote cell proliferation [30]. In vitro experiments further

showed that CLCA1 may inhibit the proliferation and metastasis of colon cancer by inhibiting the Wnt/beta-catenin signaling pathway and the epithelial-mesenchymal transition process. In addition, in vivo experiments showed that overexpression of CLCA1 could inhibit the proliferation and metastasis of colon cancer[31]. It has been proved that CLCA1 is closely related to TMEM16A, and CLCA1 can regulate TMEM16A to participate in the proliferation, migration and invasion of colorectal cancer cells [32-34]. It is well known that c-myc is a pro-oncogene and its products closely regulate cell proliferation and apoptosis. It has been reported that there is a close relationship between CLCA1 and c-myc transcription[35]. However, the specific mechanism of action remains to be studied

In addition, Chen et al. have pointed out that chemotherapy has a poor prognosis in rectal cancer patients with high expression of CLCA1[36]. Theoretically speaking, the high expression of CLCA1 should have a better prognosis in rectal cancer patients, so why the different results? We believe that this discrepancy may be caused by differences in statistical data, sample size, patient characteristics, and treatment before sampling. Therefore, more studies are needed to explain the effect of CLCA1 expression level on the prognosis of colon cancer.

In our study, compared with normal tissues, the expression levels of CLCA1 in colon cancer tissues were significantly down-regulated with statistical significance. Moreover, previous studies have shown that high expression of CLCA1 is positively correlated with prognosis in patients with colorectal cancer. This conclusion is consistent with our research results.

However, there are some limitations in our research. Such as, this paper lacks experimental verification. Despite the high quality of RNA-seq data in TCGA, in vitro and in vivo experiments are still needed to further verify mRNA expression of CLCA1 in colon cancer. In addition, important information such as tumor size is not available, and the clinical information was incomplete. Moreover, there was a shortage of specific details, such as surgical treatment and surgical details, which are crucial to the prognosis of patients. In the end, like all biogenetic studies, our study was unable to explore the results of CLCA1 in different subtypes of colorectal cancer (Left colon, right colon, sigmoid and rectal cancers). In despite of the fact that we provided a comprehensive bioinformatics

analysis to identify potential diagnostic genes between cancer and normal tissues, it may be difficult to make accurate results for each patient with colon cancer subtypes.

To sum up, by integrating WGCNA with differential gene expression analysis, this research generated the significant survival correlative gene, which was called CLCA1 that has potential for prognosis prediction in colon cancer.

Acknowledgments

Not applicable.

Funding

This study was supported by the Fujian province teaching reform project (grant no. FBJG20190063) and the Fujian medical university Key teaching reform Project (grant no. Y19001).

Availability of data and materials

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

Authors' contributions

SQC designed the study, obtained funding, analyzed the data and revised the manuscript. JJL drafted the manuscript and supervised the study. SQC was a supporter of the funding (grant no. FBJG20190063) and participating in the drafting and subsequent revision of the manuscript. SYL, ZHC and LZY collected and analyzed the data. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

1. Workman L, Brandt HM, Jackson D, Annang L, Hurley TG A gender-based examination of colorectal cancer screening participation among participants in a colon cancer awareness and educational program in two South Carolina counties. In: Apha Meeting & Exposition, 2012.

2. Mortazavi A, Shaukat A, Othman E, Kepner JL, Fakih MG, Kuvshinoff BW, Smith JL, Javle MM (2005) Postoperative computed tomography scan surveillance for patients with stage II and III colorectal cancer: worthy of further study? *American Journal of Clinical Oncology* 28 (1):30-35

3. Villanueva L, Hernan DIF, Ahumada M, Cardemil B, Rosas J, Araya H, Hales C, Madrid E, Roa I, Orlandi L (2013) P-0225MULTICENTRE, OBSERVATIONAL, CHILEAN TRIAL OF PATIENTS WITH COLORECTAL CANCER STAGE III AND IV: ESCORRECTO GOCCHI 2008-01. *Annals of Oncology* 24 (suppl 4):iv99-iv100

4. Johnson CD, Byrom M, Bader AG, Slack FJ, Brown D, Ovcharenko D, Kelnar K (2008) FUNCTIONS AND TARGETS OF LET-7 MICRO RNAS.

5. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* 9 (1):559

6. Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH (2008) Functional organization of the transcriptome in human brain. *Nature neuroscience* 11 (11):1271

7. He Z, Sun M, Ke Y, Lin R, Xiao Y, Zhou S, Zhao H, Wang Y, Zhou F, Zhou Y (2017) Identifying biomarkers of papillary renal cell carcinoma associated with pathological stage by weighted gene co-expression network analysis. *Oncotarget* 8 (17):27904

8. Zhai X, Xue Q, Liu Q, Guo Y, Chen Z (2017) Colon cancer recurrence-associated genes revealed by WGCNA co-expression network analysis. *Molecular Medicine Reports* 16 (5):6499-6505

9. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I (2016) TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic acids research* 44 (8):e71-e71

10. Sean D, Meltzer PS, GEOquery B (2007) v. 23 n. 14. July,

11. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* 4 (1)

12. Wang CC, Li CY, Cai J-H, Sheu PC-Y, Tsai JJ, Wu M-Y, Li C-J, Hou M-F (2019) Identification of prognostic candidate genes in breast cancer by integrated bioinformatic analysis. *Journal of clinical medicine* 8 (8):1160

-
13. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 43 (7):e47-e47
 14. Wickham H (2016) ggplot2: elegant graphics for data analysis. springer,
 15. Chen H, Boutros PC (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC bioinformatics* 12 (1):1-7
 16. Lu Y, Rosenfeld R, Simon I, Nau GJ, Bar-Joseph Z (2008) A probabilistic generative model for GO enrichment analysis. *Nucleic acids research* 36 (17):e109-e109
 17. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 28 (1):27-30
 18. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 13 (11):2498-2504
 19. Tang Z, Kang B, Li C, Chen T, Zhang Z (2019) GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic acids research* 47 (W1):W556-W560
 20. Asplund A, Edqvist PHD, Schwenk JM, Pontén F (2012) Antibodies for profiling the human proteome—The Human Protein Atlas as a resource for cancer research. *Proteomics* 12 (13):2067-2077
 21. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 68 (6):394-424
 22. Wallace DC (2012) Mitochondria and cancer. *Nature Reviews Cancer* 12 (10):685-698
 23. Icard P, Shulman S, Farhat D, Steyaert J-M, Alifano M, Lincet H (2018) How the Warburg effect supports aggressiveness and drug resistance of cancer cells? *Drug Resistance Updates* 38:1-11
 24. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *cell* 144 (5):646-674
 25. Chan A, Weillbach F, Toyka K, Gold R (2005) Mitoxantrone induces cell death in peripheral blood leucocytes of multiple sclerosis patients. *Clinical & Experimental Immunology* 139 (1):152-158
 26. Duddy M, Niino M, Adatia F, Hebert S, Freedman M, Atkins H, Kim HJ, Bar-Or A (2007) Distinct effector cytokine profiles of memory and naive human B cell subsets and implication in multiple sclerosis. *The Journal of Immunology* 178 (10):6092-6099
 27. Li X, Hu W, Zhou J, Huang Y, Peng J, Yuan Y, Yu J, Zheng S (2017) CLCA1 suppresses colorectal cancer aggressiveness via inhibition of the Wnt/beta-catenin signaling pathway. *Cell Communication and Signaling* 15 (1):1-13
 28. Gruber AD, Elble RC, Ji H-L, Schreur KD, Fuller CM, Pauli BU (1998) Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca²⁺-activated Cl⁻ channel proteins. *Genomics* 54 (2):200-214
 29. Duvvuri U, Shiwerski DJ, Xiao D, Bertrand C, Huang X, Edinger RS, Rock JR, Harfe BD, Henson BJ, Kunzelmann K (2012) TMEM16A induces MAPK and contributes directly to tumorigenesis and cancer progression. *Cancer research* 72 (13):3270-3281
 30. Woo S-R, Corrales L, Gajewski TF (2015) Innate immune recognition of cancer. *Annual*

review of immunology 33:445-474

31. Alevy YG, Patel AC, Romero AG, Patel DA, Tucker J, Roswit WT, Miller CA, Heier RF, Byers DE, Brett TJ (2012) IL-13–induced airway mucus production is attenuated by MAPK13 inhibition. *The Journal of clinical investigation* 122 (12):4555-4568

32. Rottgen TS, Nickerson AJ, Rajendran VM (2018) Calcium-Activated Cl⁻ Channel: Insights on the Molecular Identity in Epithelial Tissues. *International journal of molecular sciences* 19 (5):1432

33. Sui Y, Wu F, Lv J, Li H, Li X, Du Z, Sun M, Zheng Y, Yang L, Zhong L (2015) Identification of the novel TMEM16A inhibitor dehydroandrographolide and its anticancer activity on SW620 cells. *PloS one* 10 (12):e0144715

34. Sui Y, Sun M, Wu F, Yang L, Di W, Zhang G, Zhong L, Ma Z, Zheng J, Fang X (2014) Inhibition of TMEM16A expression suppresses growth and invasion in human colorectal cancer cells. *PloS one* 9 (12):e115443

35. Stephen A. Bustin S-RL, Sina Dorudi (2001) Expression of the Ca²⁺-activated chloride channel genes CLCA1 and CLCA2 is downregulated in human colorectal cancer. *Dna & Cell Biology* 20 (6):331-338

36. Tzu-Ju C, Hong-Lin H, Yow-Ling S, Ching-Chieh Y, Li-Ching L, Yu-Feng T, Shang-Hung C (2018) High chloride channel accessory 1 expression predicts poor prognoses in patients with rectal cancer receiving chemoradiotherapy. *International journal of medical sciences* 15 (11):1171-1178

Figure 1 Study design and workflow of this study.

Figure 2 Identification of modules associated with the clinical information in the TCGA colon cancer dataset.

(A) The Cluster dendrogram of co-expression network modules was ordered by a hierarchical clustering of genes based on the 1-TOM matrix. Each module was assigned different colors.

(B) Module-trait relationships. Each row corresponds to a color module and column corresponds to a clinical trait (cancer and normal). Each cell contains the corresponding correlation and P-value.

Figure 3 Identification of modules associated with clinical information in the GSE44861 dataset.

(A) The Cluster dendrogram of co-expression network modules was ordered by a hierarchical clustering of genes based on the 1-TOM matrix. Each module was assigned different colors.

(B) Module-trait relationships. Each row corresponds to a color module and each column correlates to a clinical trait (cancer and normal). Each cell contains the corresponding correlation and P-value.

Figure 4 Identification of differentially expressed genes (DEGs) among the TCGA and GSE44861 datasets of colon cancer with the cut-off criteria of $|\logFC| \geq 1.0$ and $\text{adj. } P < 0.05$.

(A) Volcano plot of DEGs in the TCGA dataset.

(B) Volcano plot of DEGs in the GSE6631 dataset.

(C) The Venn diagram of genes among DEG lists and co-expression module.

In total, 103 overlapping genes in the intersection of DEG lists and two co-expression modules.

Figure 5 Gene Ontology (GO) enrichment analysis for the genes in the brown module. The color represents the adjusted p-values (BH), and the size of the spots represents the gene number.

Figure 6 Visualization of the protein-protein interaction (PPI) network and the candidate hub genes.

(A) PPI network of the genes between DEG lists and two co-expression modules. The blue nodes represent the genes. Edges indicate interaction associations between nodes.

(B) Identification of the hub genes from the PPI network using maximal clique centrality (MCC) algorithm. Edges represent the protein-protein associations.

The red nodes represent genes with a high MCC scores, while the yellow node represent genes with a low MCC score.

Figure 7 Overall survival (OS) analysis of 10 hub genes in colon cancer patients from the TCGA database.

(A) Survival analysis for CLCA1 in colon cancer

(B) Survival analysis for GCG in colon cancer

(C) Survival analysis for GUCA2A in colon cancer

(D) Survival analysis for SLC26A3 in colon cancer

(E) Survival analysis for ZG16 in colon cancer

(F) Survival analysis for CLCA4 in colon cancer

(G) Survival analysis for GUCA2B in colon cancer

(H) Survival analysis for MS4A12 in colon cancer

(I) Survival analysis for NRIH4 in colon cancer

(J) Survival analysis for SI in colon cancer

The patients were stratified into high-level group (red) and low-level group (green) according to median expression of the gene. Log-rank $P < 0.05$ was considered to be a statistically significant difference.

Figure 8 Disease-free survival (DFS) analysis of 10 hub genes in colon cancer patients from the GEPIA2 database.

(A) Survival analysis for CLCA1 in colon cancer

(B) Survival analysis for GCG in colon cancer

(C) Survival analysis for GUCA2A in colon cancer

(D) Survival analysis for SLC26A3 in colon cancer

(E) Survival analysis for ZG16 in colon cancer

(F) Survival analysis for CLCA4 in colon cancer

(G) Survival analysis for GUCA2B in colon cancer

(H) Survival analysis for MS4A12 in colon cancer

(I) Survival analysis for NRIH4 in colon cancer

(J) Survival analysis for SI in colon cancer

The patients were stratified into high-level group (red) and low-level group (green) according to median expression of the gene. Log-rank $P < 0.05$ was considered to be a statistically significant difference.

Figure 9 The expression of CLCA1 and its association with clinicopathological parameters based on TCGA data. (A) Comparison of CLCA1 expression between colon cancer tissues and adjacent nontumor tissues. The expression of CLCA1 is grouped by pathological stage (B), T stage (C), N stage (D), and M stage (E).

Figure 10 Immunohistochemistry of the CLCA1 gene in colon cancer and normal tissues from the human protein atlas (HPA) database.

(A) Protein levels of CLCA1 in colon cancer tissues (antibody CAB047315; staining: not detected; intensity: negative; quantity: none).

(B) Protein levels of CLCA1 in normal colon tissues (antibody CAB047315; staining: high; intensity: strong; quantity: >75%).

Figure 11 A merged enrichment plot from gene set enrichment analysis including enrichment score and gene sets.

Figure 12 TIC profile in tumor samples and correlation analysis. (A) Barplot showing the proportion of 21 kinds of TICs in colon cancer samples. Column names of plot were sample ID. (B) Heatmap showing the correlation between 21 kinds of TICs and numeric in each tiny box indicating the p value of correlation between two kinds of cells. The shade of each tiny color box represented corresponding correlation value between two cells, and Pearson coefficient was used for significance test.

Figure 13 Correlation of TICs proportion with CLCA1 expression.

(A) Violin plot showed the ratio differentiation of 22 kinds of immune cells between colon tumor samples with low or high CLCA1 expression relative to the median of CLCA1 expression level, and Wilcoxon rank sum was used for the significance test.

(B) Scatter plot showed the correlation of 7 kinds of TICs proportion with the CLCA1 expression ($p < 0.05$). The red line in each plot was fitted linear model indicating the proportion tropism of the immune cell along with CLCA1 expression, and Pearson coefficient was used for the correlation test.

(C) Venn plot displayed eight kinds of TICs correlated with CLCA1 expression codetermined by difference and correlation tests displayed in violin and scatter plots, respectively.

Figures

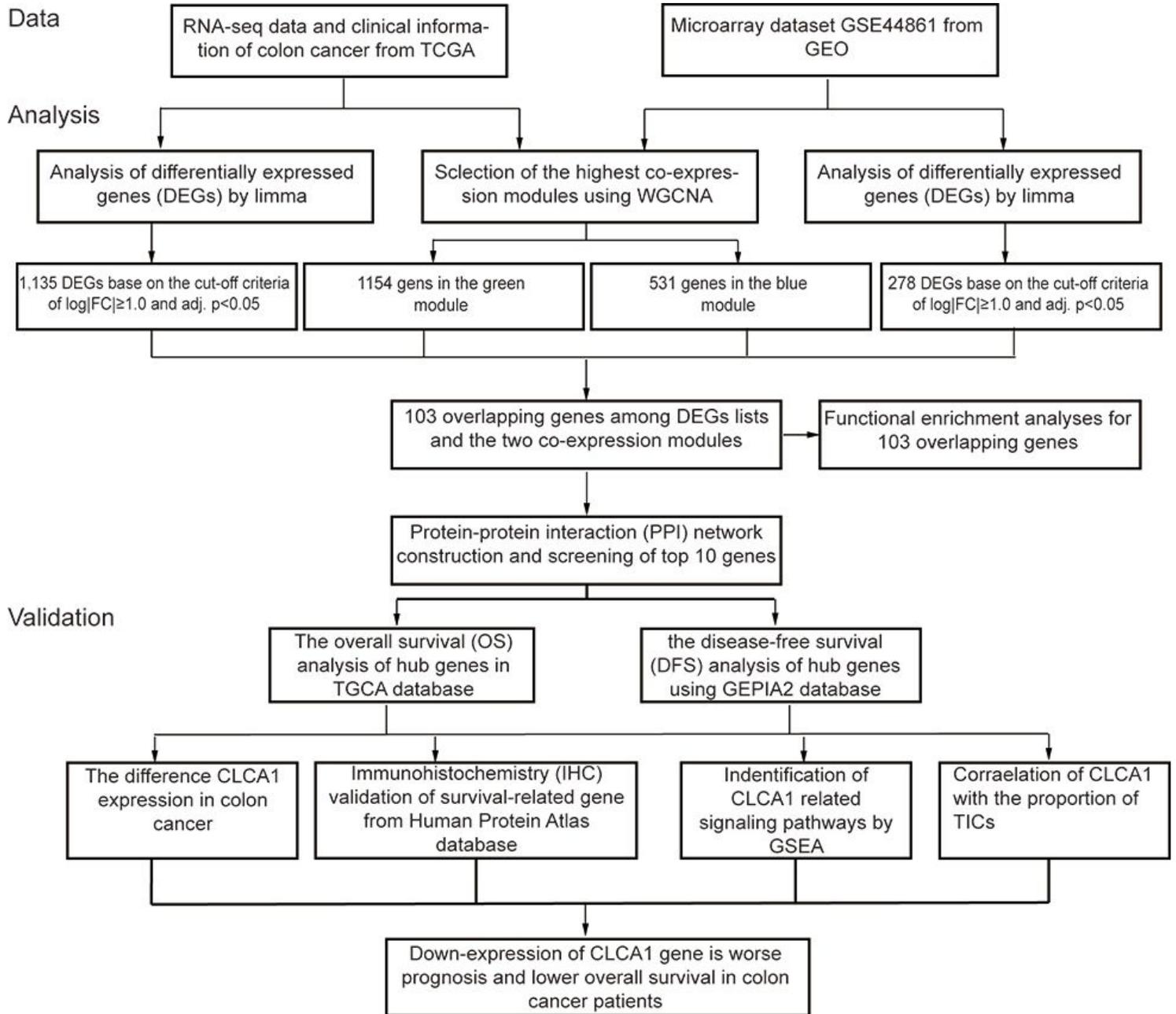
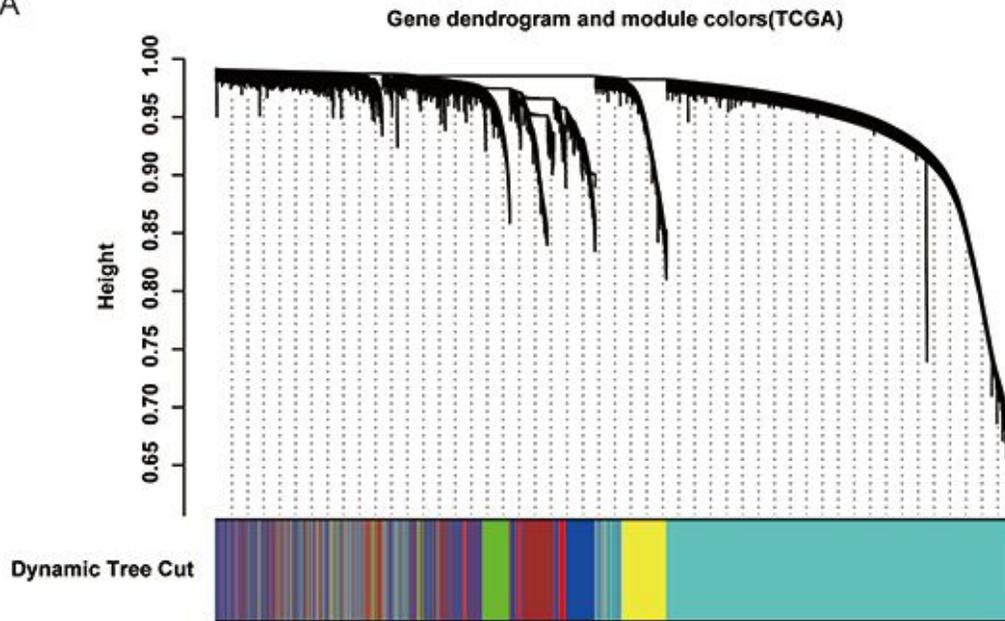


Figure 1

Study design and workflow of this study.

Figure 2

A



B

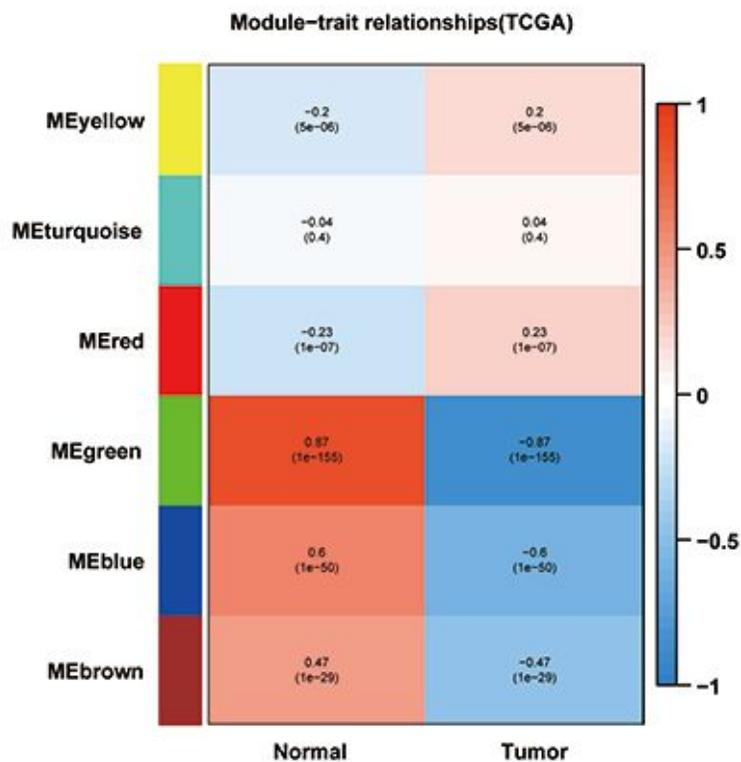


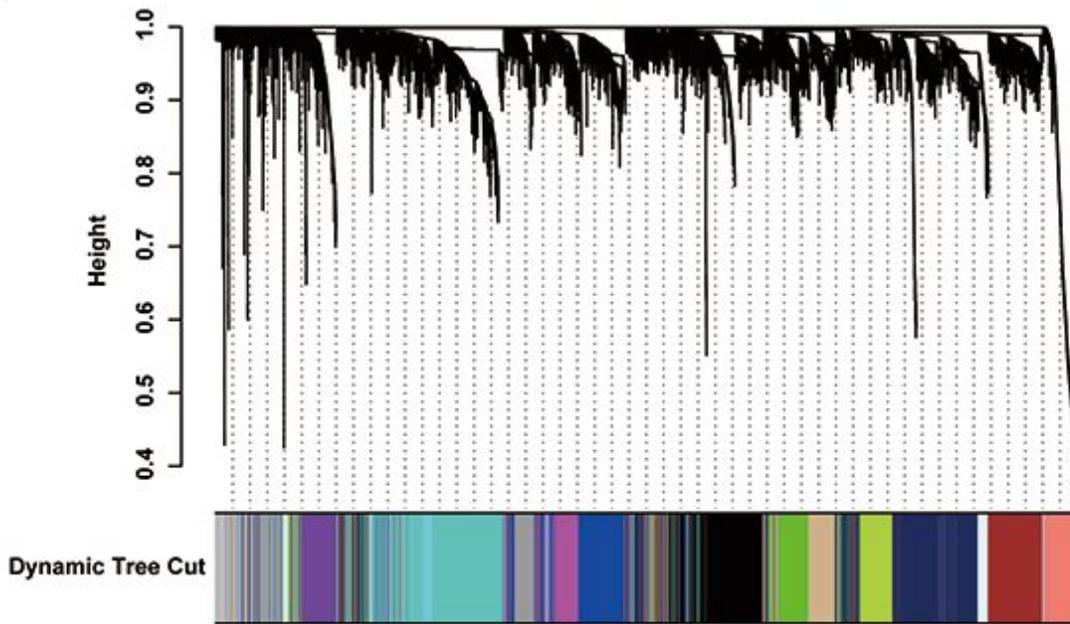
Figure 2

Identification of modules associated with the clinical information in the TCGA colon cancer dataset. (A) The Cluster dendrogram of co-expression network modules was ordered by a hierarchical clustering of genes based on the 1-TOM matrix. Each module was assigned different colors. (B) Module-trait relationships. Each row corresponds to a color module and column corresponds to a clinical trait (cancer and normal). Each cell contains the corresponding correlation and P-value.

figure3

A

Gene dendrogram and module colors(GEO)



B

Module-trait relationships(GEO)

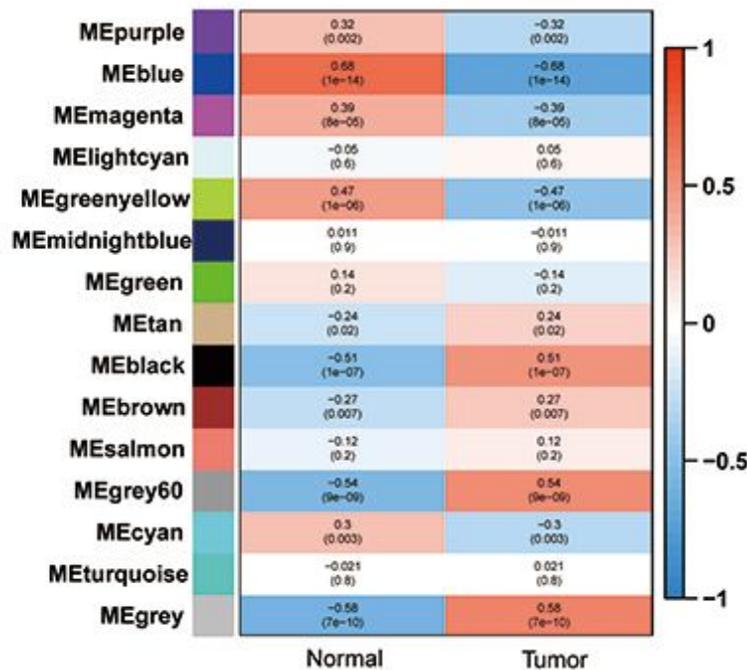


Figure 3

Identification of modules associated with clinical information in the GSE44861 dataset. (A) The Cluster dendrogram of co-expression network modules was ordered by a hierarchical clustering of genes based on the 1-TOM matrix. Each module was assigned different colors. (B) Module-trait relationships. Each row corresponds to a color module and each column correlates to a clinical trait (cancer and normal). Each cell contains the corresponding correlation and P-value.

figure 4

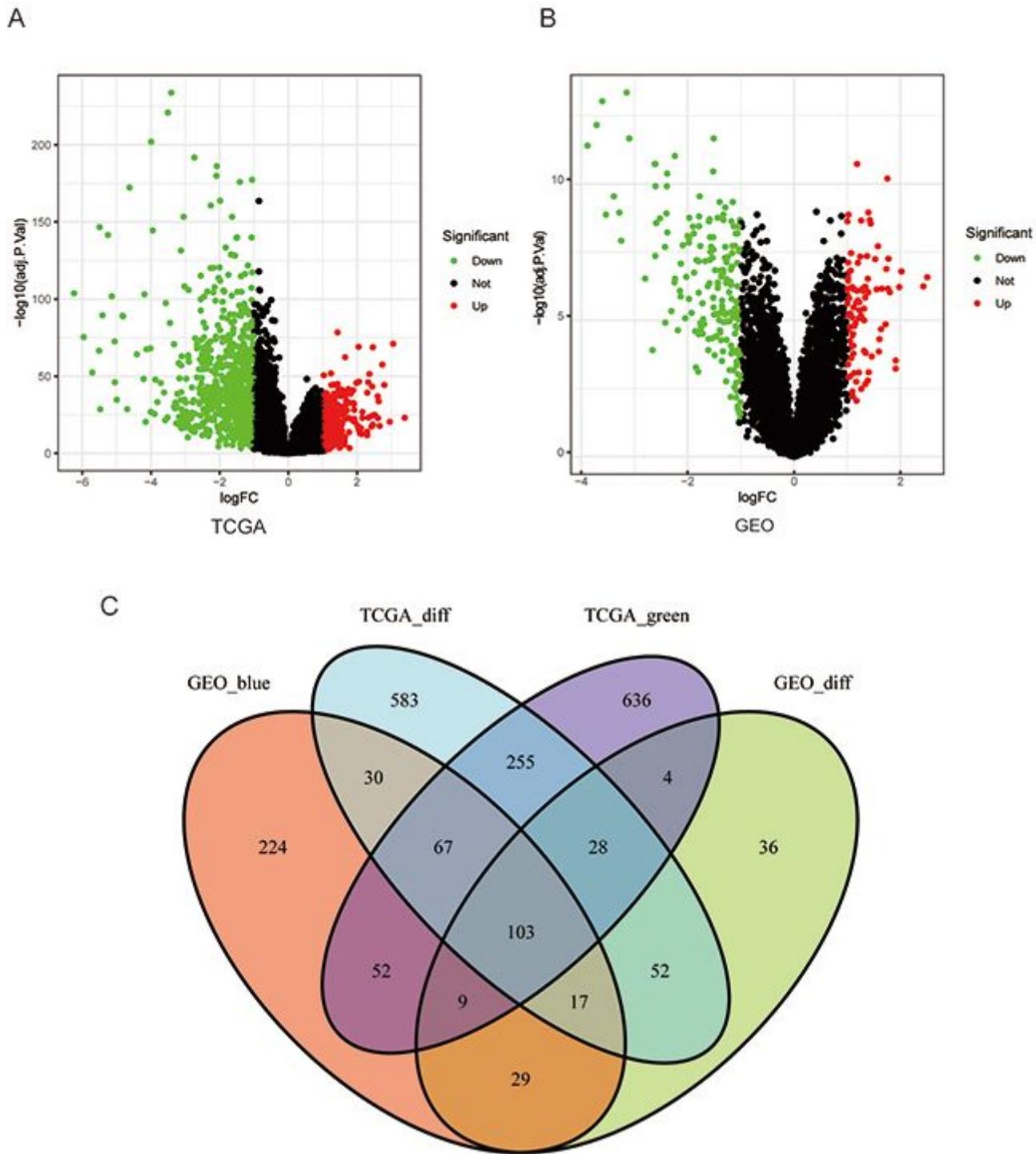
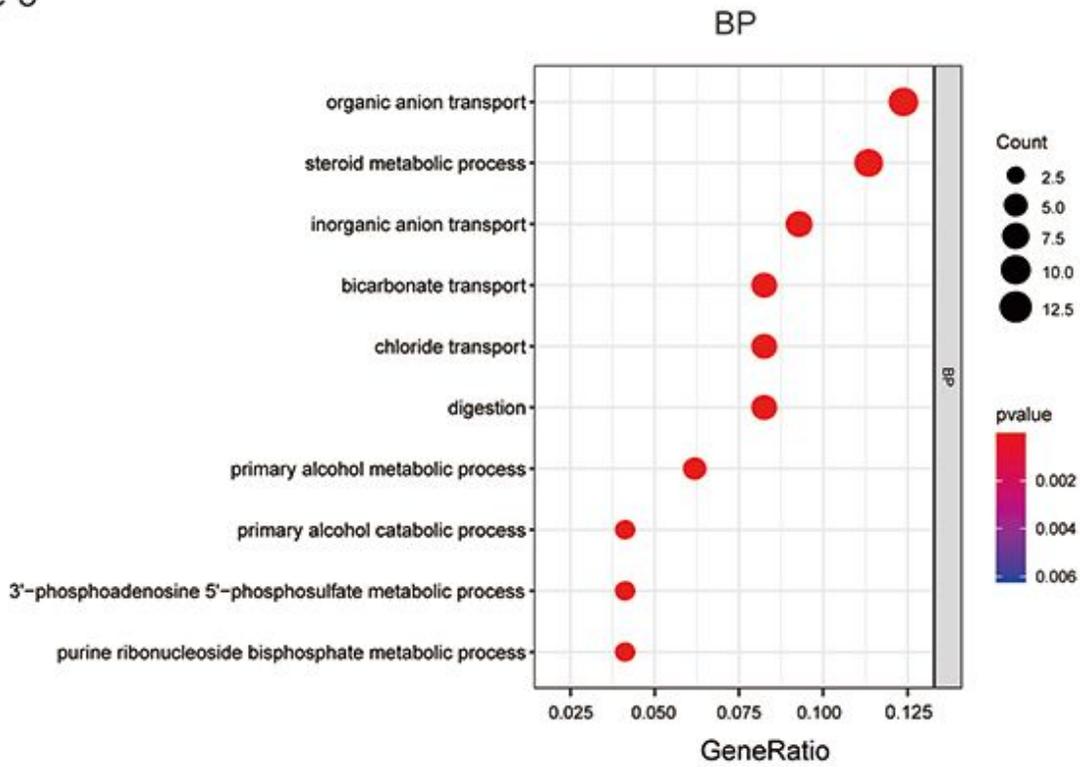


Figure 4

Identification of differentially expressed genes (DEGs) among the TCGA and GSE44861 datasets of colon cancer with the cut-off criteria of $|\logFC| \geq 1.0$ and $\text{adj. } P < 0.05$. (A) Volcano plot of DEGs in the TCGA dataset. (B) Volcano plot of DEGs in the GSE6631 dataset. (C) The Venn diagram of genes among DEG lists and co-expression module. In total, 103 overlapping genes in the intersection of DEG lists and two co-expression modules.

Figure 5

A



B

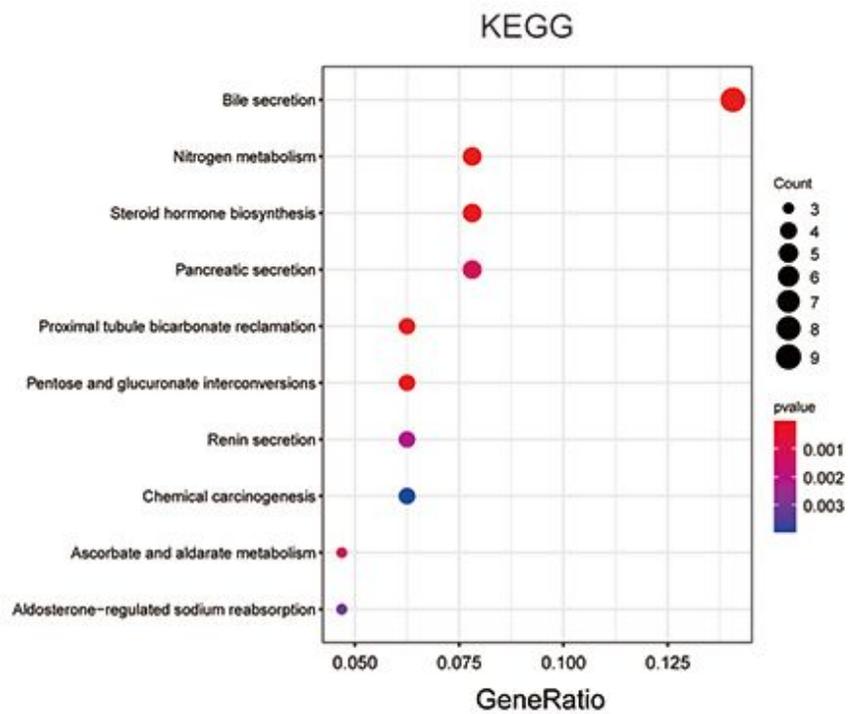
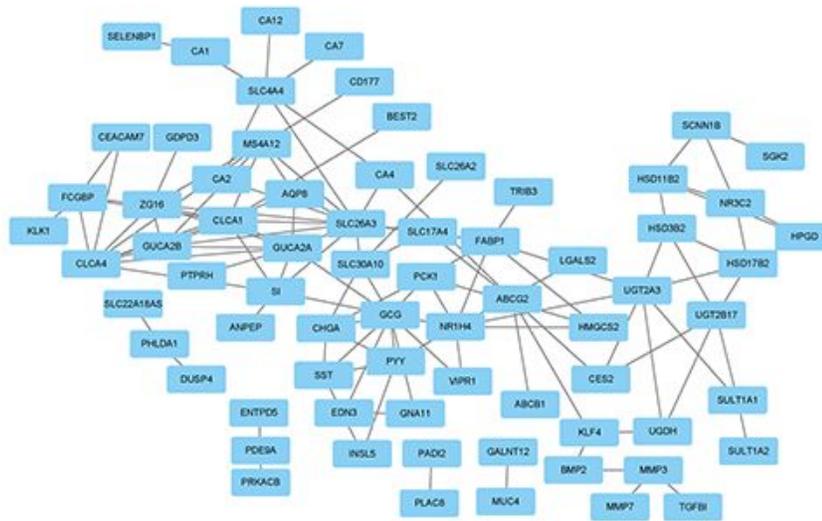


Figure 5

Gene Ontology (GO) enrichment analysis for the genes in the brown module. The color represents the adjusted p-values (BH), and the size of the spots represents the gene number.

Figure 6

A



B

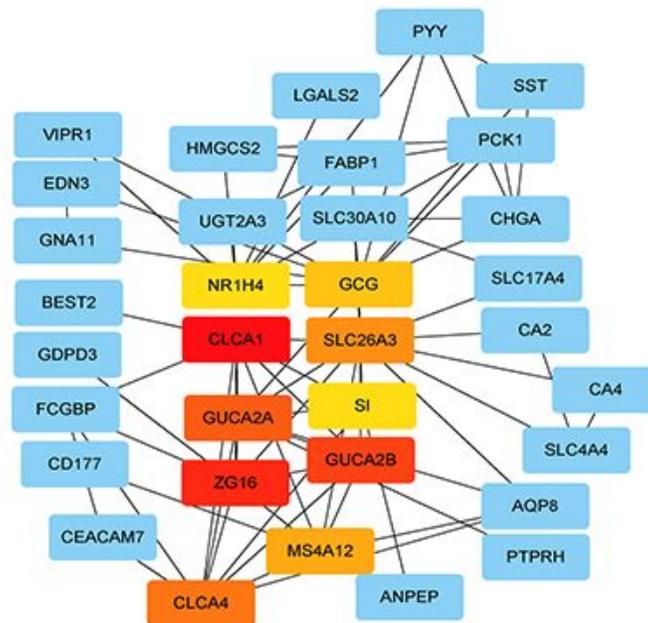


Figure 6

Visualization of the protein-protein interaction (PPI) network and the candidate hub genes. (A) PPI network of the genes between DEG lists and two co-expression modules. The blue nodes represent the genes. Edges indicate interaction associations between nodes. (B) Identification of the hub genes from the PPI network using maximal clique centrality (MCC) algorithm. Edges represent the protein-protein

associations. The red nodes represent genes with a high MCC scores, while the yellow node represent genes with a low MCC score.

Figure 7

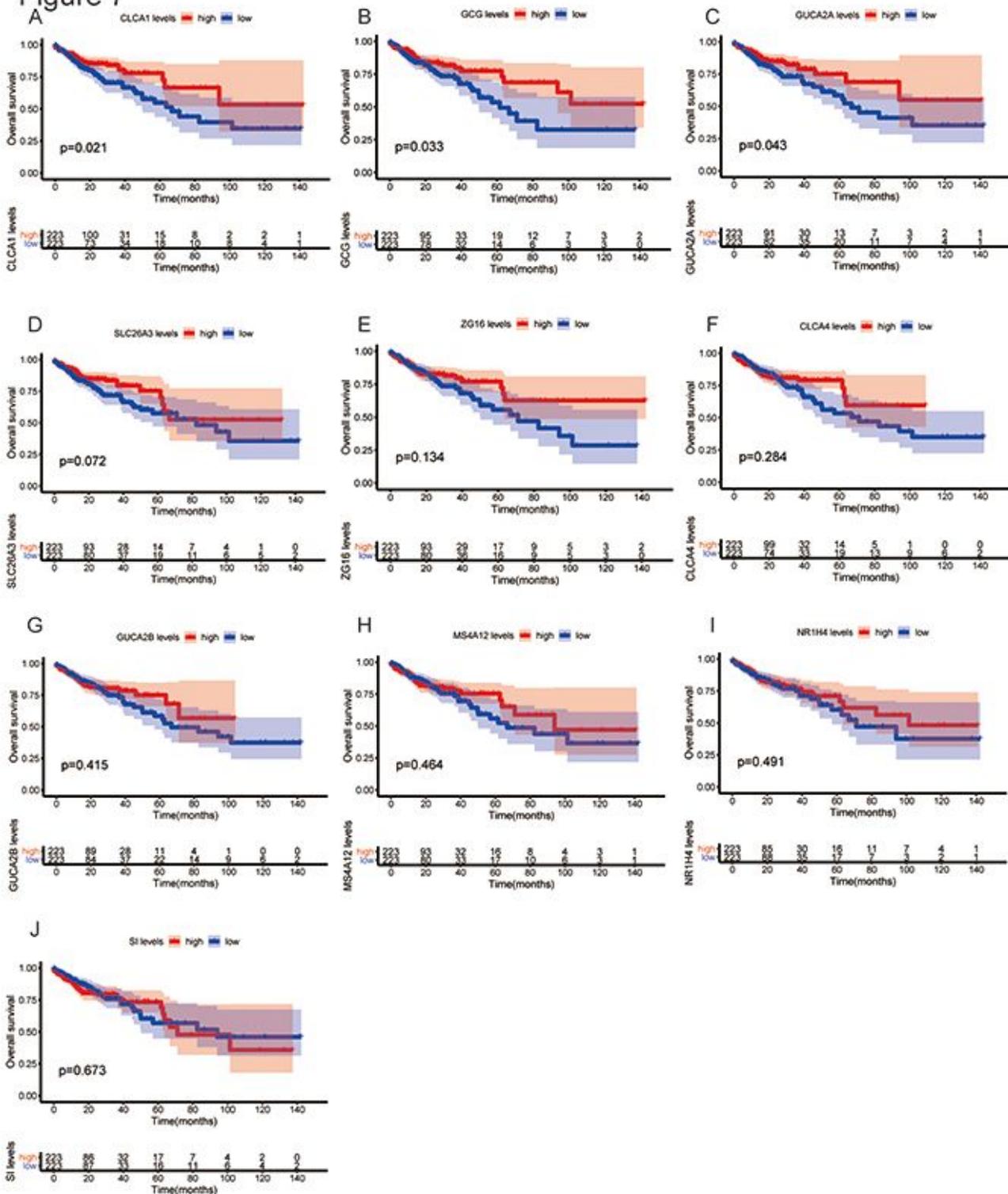


Figure 7

Overall survival (OS) analysis of 10 hub genes in colon cancer patients from the TCGA database. (A) Survival analysis for CLCA1 in colon cancer (B) Survival analysis for GCG in colon cancer (C) Survival analysis for GUCA2A in colon cancer (D) Survival analysis for SLC26A3 in colon cancer (E) Survival

analysis for ZG16 in colon cancer (F) Survival analysis for CLCA4 in colon cancer (G) Survival analysis for GUCA2B in colon cancer (H) Survival analysis for MS4A12 in colon cancer (I) Survival analysis for NRIH4 in colon cancer (J) Survival analysis for SI in colon cancer The patients were stratified into high-level group (red) and low-level group (green) according to median expression of the gene. Log-rank $P < 0.05$ was considered to be a statistically significant difference.

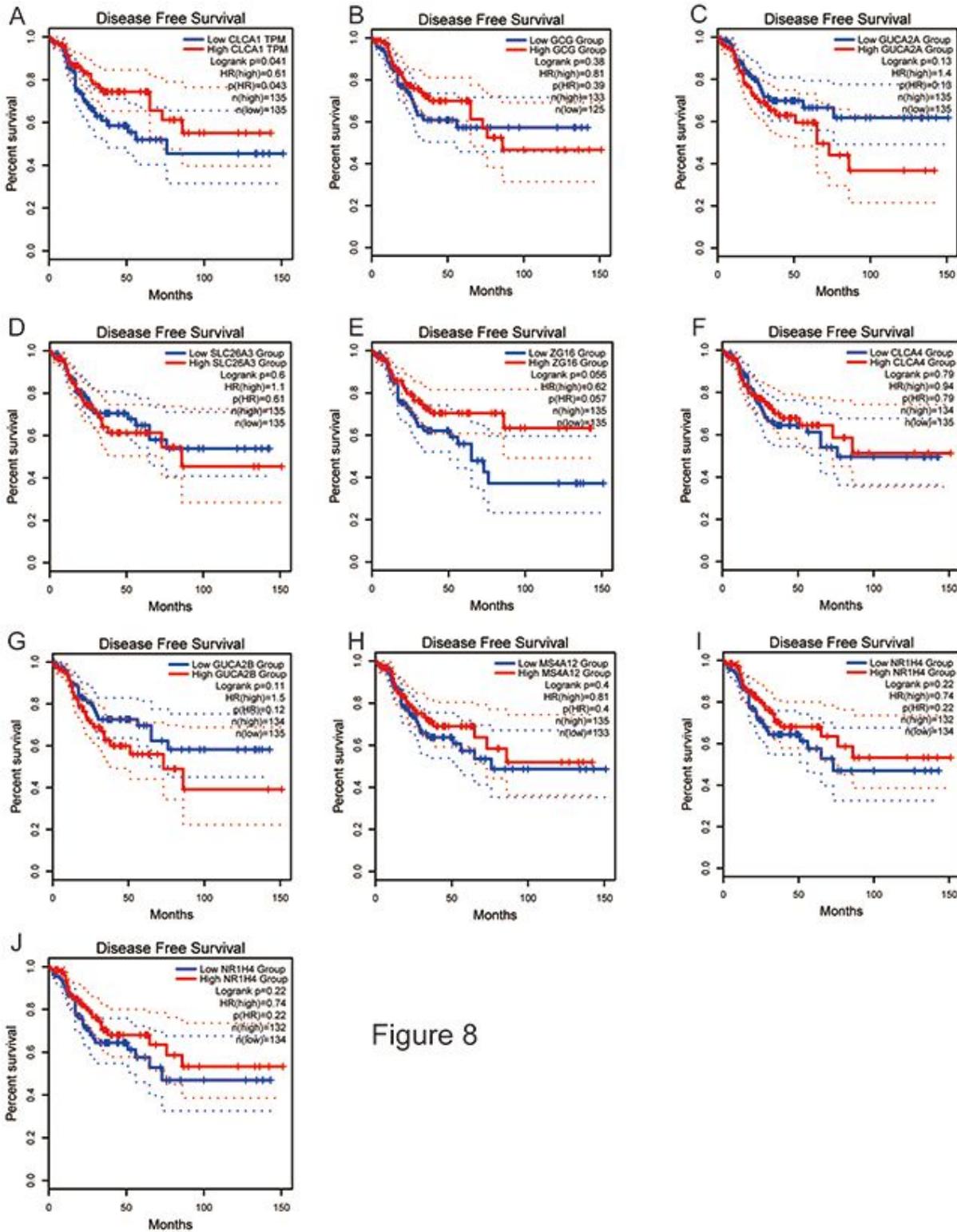


Figure 8

Figure 8

Disease-free survival (DFS) analysis of 10 hub genes in colon cancer patients from the GEPIA2 database. (A) Survival analysis for CLCA1 in colon cancer (B) Survival analysis for GCG in colon cancer (C) Survival analysis for GUCA2A in colon cancer (D) Survival analysis for SLC26A3 in colon cancer (E) Survival analysis for ZG16 in colon cancer (F) Survival analysis for CLCA4 in colon cancer (G) Survival analysis for GUCA2B in colon cancer (H) Survival analysis for MS4A12 in colon cancer (I) Survival analysis for NRIH4 in colon cancer (J) Survival analysis for SI in colon cancer The patients were stratified into high-level group (red) and low-level group (green) according to median expression of the gene. Log-rank $P < 0.05$ was considered to be a statistically significant difference.

Figure 9

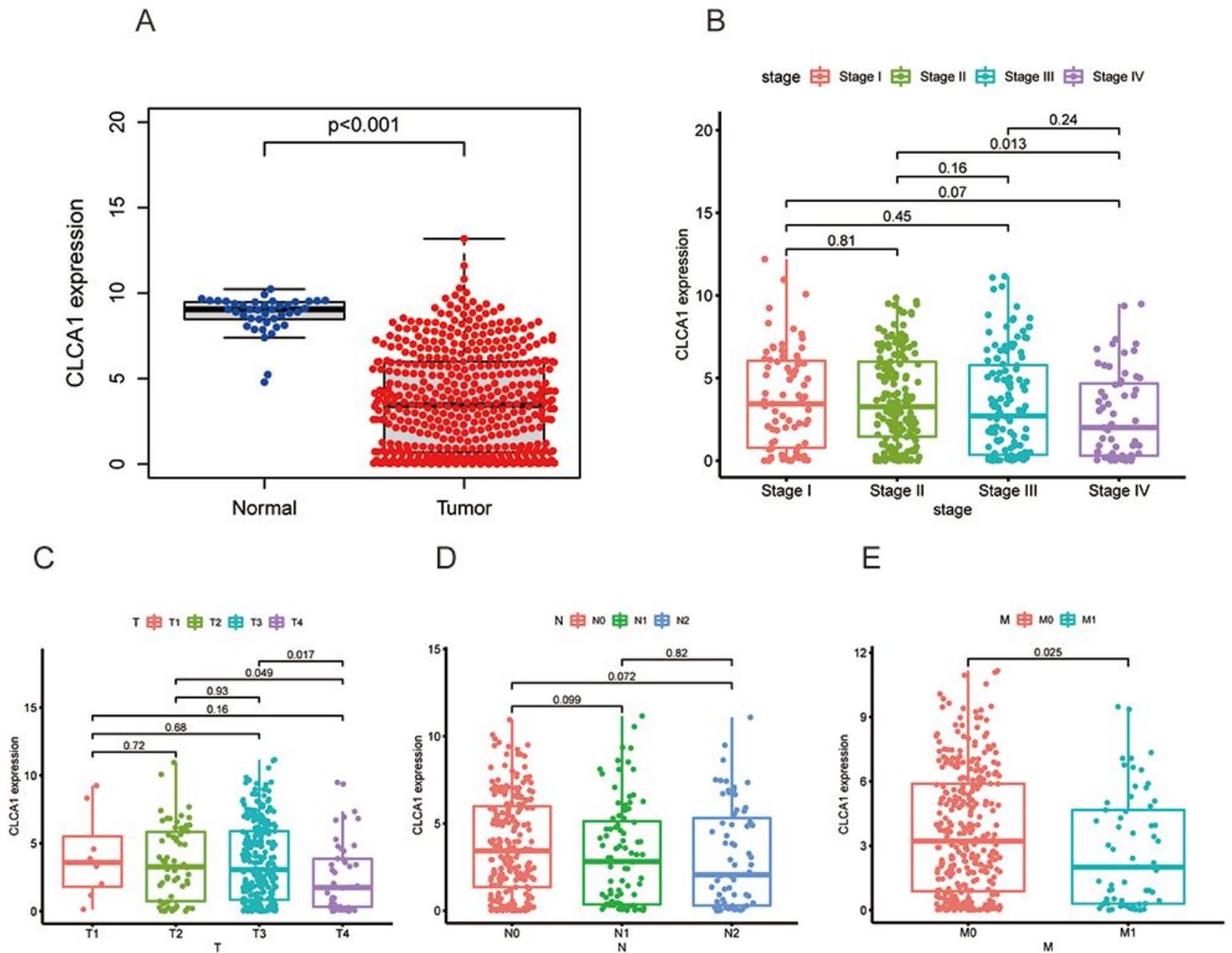
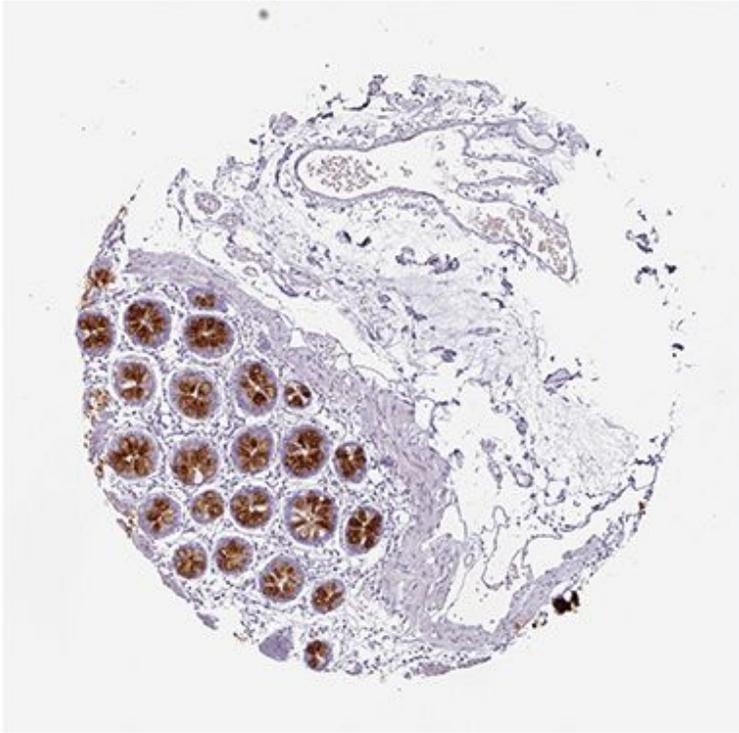


Figure 9

The expression of CLCA1 and its association with clinicopathological parameters based on TCGA data. (A) Comparison of CLCA1 expression between colon cancer tissues and adjacent nontumor tissues. The expression of CLCA1 is grouped by pathological stage (B), T stage (C), N stage (D), and M stage (E).

normal



tumor

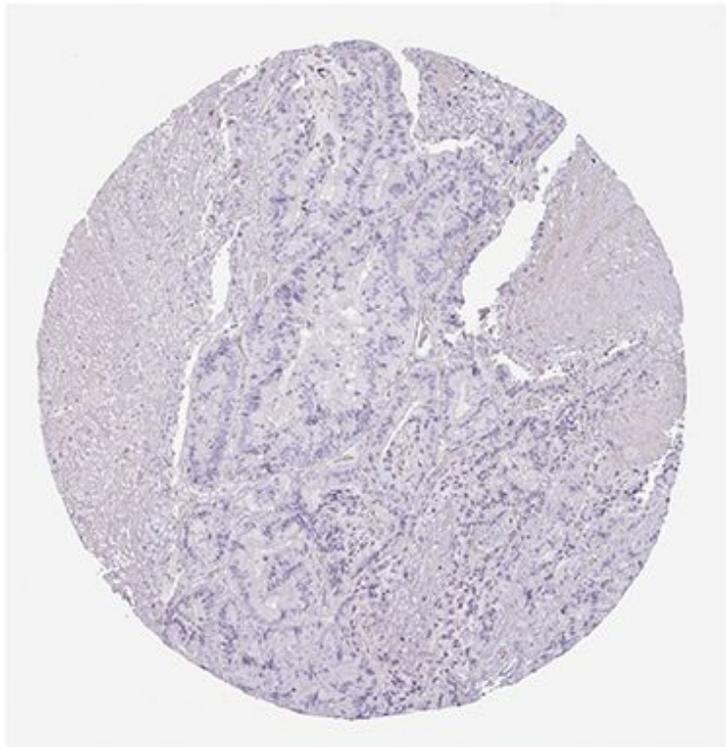


Figure 10

Immunohistochemistry of the CLCA1 gene in colon cancer and normal tissues from the human protein atlas (HPA) database. (A) Protein levels of CLCA1 in colon cancer tissues (antibody CAB047315; staining: not detected; intensity: negative; quantity: none). (B) Protein levels of CLCA1 in normal colon tissues (antibody CAB047315; staining: high; intensity: strong; quantity: >75%).

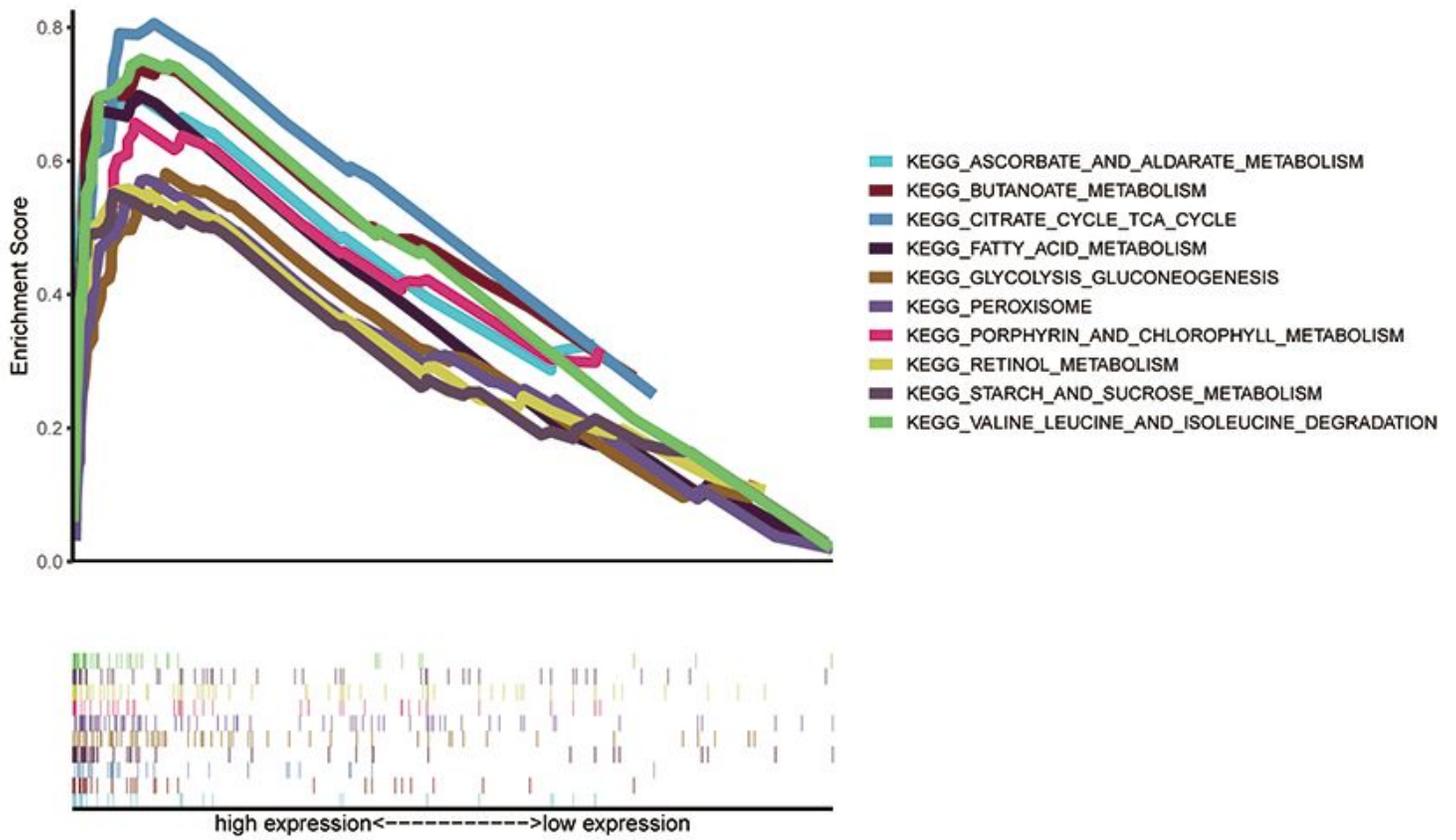


Figure 11

A merged enrichment plot from gene set enrichment analysis including enrichment score and gene sets.

Figure 12

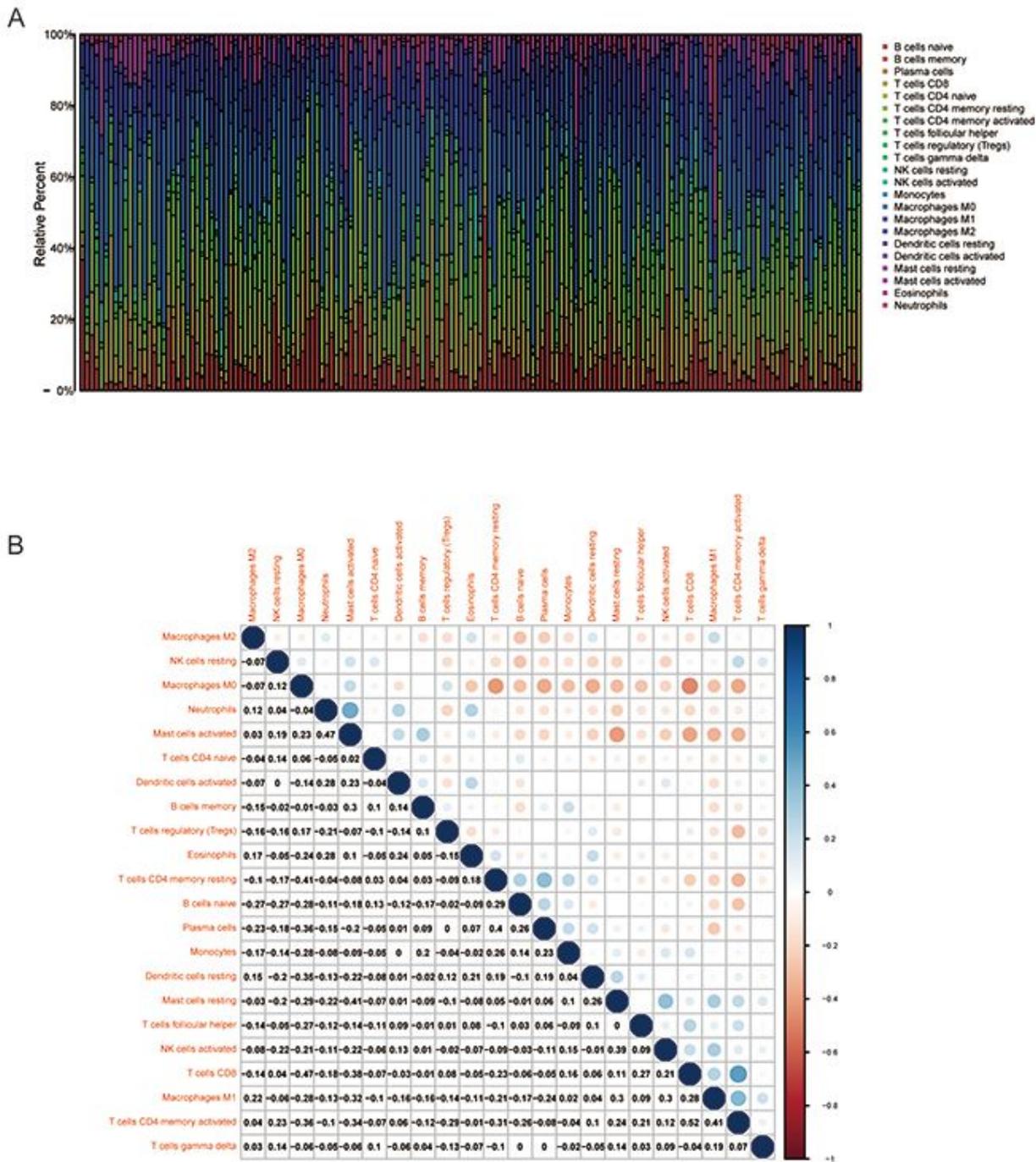


Figure 12

TIC profile in tumor samples and correlation analysis. (A) Barplot showing the proportion of 21 kinds of TICs in colon cancer samples. Column names of plot were sample ID. (B) Heatmap showing the correlation between 21 kinds of TICs and numeric in each tiny box indicating the p value of correlation between two kinds of cells. The shade of each tiny color box represented corresponding correlation value between two cells, and Pearson coefficient was used for significance test.

Figure 13

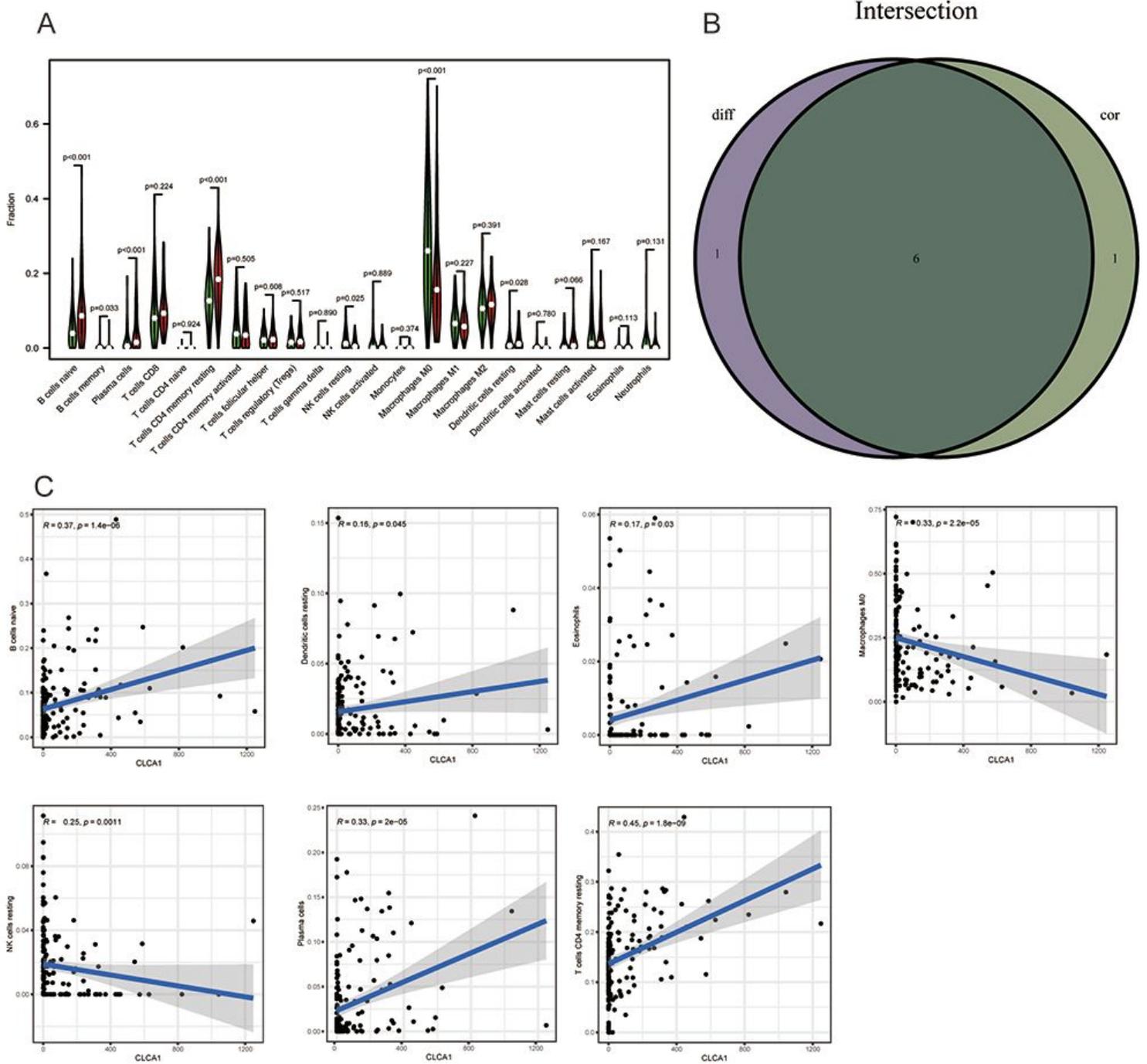


Figure 13

Correlation of TICs proportion with CLCA1 expression. (A) Violin plot showed the ratio differentiation of 22 kinds of immune cells between colon tumor samples with low or high CLCA1 expression relative to the median of CLCA1 expression level, and Wilcoxon rank sum was used for the significance test. (B) Scatter plot showed the correlation of 7 kinds of TICs proportion with the CLCA1 expression ($p < 0.05$). The red line in each plot was fitted linear model indicating the proportion tropism of the immune cell along with CLCA1 expression, and Pearson coefficient was used for the correlation test. (C) Venn plot displayed eight

kinds of TICs correlated with CLCA1 expression codetermined by difference and correlation tests displayed in violin and scatter plots, respectively.