

Identification of B3GNT6 Gene Associated with Development of Colon Carcinoma by Weighted Gene Co-expression Network Analysis

Lebin Yuan

The Second Affiliated Hospital of Nanchang University

Fei Cheng

The Second Affiliated Hospital of Nanchang University

Zhao Wu

The Second Affiliated Hospital of Nanchang University

Wei Shen (✉ shenwei18189@163.com)

The Second Affiliated Hospital of Nanchang University

Research Article

Keywords: colon cancer, differential gene expression analysis, weighted gene co-expression network analysis, differentially co-expressed genes, biomarkers

Posted Date: March 25th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1442716/v1>

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

[Read Full License](#)

Abstract

Colon cancer (COAD) stem cells are resistant to cancer treatment, so they are easy to lead to tumor progression after routine treatment failure. However, little is known about the molecular mechanism of cancer cells. The purpose of our study is to identify differential gene modules and representative candidate biomarkers for the clinical prognosis of patients, help to predict the prognosis and reveal the mechanism of cancer progression. In our study, based on the tumor Genome Atlas (TCGA) COAD database and gene expression profiles of GSE44076 from the Gene Expression Omnibus (GEO), the differentially co-expressed genes in COAD and normal tissues were explored. Combined with weighted gene co-expression network analysis (WGCNA) and differential gene expression analysis, 451 differential co-expression genes were screened out. As shown by functional annotation analysis using R clusterProfiler software package, these genes are mainly enriched in lipid catabolic process (biological process), cell apical part (cellular component) and phosphate hydrolase activity (molecular function). In addition, in the protein-protein interaction (PPI) network, 20 hub genes (ACAA2, FABP1, ACOX1, EHHADH, CPT2, ACADS, CPT1A, MT1G, MT1E, MT1X, MT1H, PPARGC1A, ACSS2, MT2A, MT1F, CRAT, UGT2B17, B3GNT6, and MUC4) were identified using Cytoscape's CytoHubba plug-in. Based on Kaplan–Meier univariate survival analysis and UALCAN database analysis, lower expression of B3GNT6 was associated with better overall survival (OS), which could greatly influence cancer development. Also, online TIMER and CIBERSORT database showed that there was a great correlation between high B3GNT6 gene expression and immune infiltrating cells in poor tumor progression. Finally, we verified the protein levels of B3GNT6 through HPA database. In conclusion, our study shows that B3GNT6 can play an important role in the progression of colon cancer and serve as potential biomarkers for future diagnosis and treatment.

Introduction

Colon adenocarcinoma (COAD) is a common malignant tumor, which endangers human health[1]. In 2020, there were more than 1.9 million new cases of colon cancer in the world with epidemiological analysis[2]. It is estimated that by 2030, there will be more than 2.2 million new cases and 1.1 million deaths[3]. The tumorigenesis and progression of COAD is considered to be a multi-step, multi-stage and multi-gene cellular genetic disease. Although the progress of treatment including surgical resection, neoadjuvant radiotherapy and chemotherapy, postoperative radiotherapy and chemotherapy, targeted therapy and immunotherapy have significantly improved the prognosis of patients with COAD, its prognosis is far from ideal especially in patients with advanced diseases[4]. Therefore, more research is needed on useful new biomarkers that may have therapeutic or prognostic value for patients with colon cancer.

Nowadays, microarray technology is widely used to study the molecular mechanism of disease and finding disease-specific biomarkers, which has given scholars more perspectives to study the characteristics and treatment of cancer, especially colon cancer[5]. Weighted Gene Co-expression Network Analysis (WGCNA), which is a systems biology method to describe the correlation patterns among genes

via microarray samples, help physicians to understand the gene function and gene association from genome-wide expression[6]. Co-expression modules of highly correlated genes and interested modules associated with clinical traits can be ascertained by WGCNA[7, 8], which providing great insight into predicting the functions of co-expression genes and finding hub genes that play key roles in human diseases[9, 10]. Moreover, another powerful analysis within transcriptomics is differential gene expression analysis, which is an amicable online data processing tool for studying molecular mechanisms underlying genome regulation and discovering quantitative changes in expression levels between experimental groups and control groups[7, 11]. Potential biomarkers of some specific diseases can be found through gene expression differences. Thus, combining the results of WGCNA and differential gene expression analysis to improve the recognition ability of highly related genes as candidate biomarkers to assist us better research cancer.

In this study, we downloaded the original mRNA expression data of COAD from the TCGA and GEO databases, differential co-expression gene was obtained by sophisticated WGCNA and differential gene expression analysis. In order to better understand COAD development, functional enrichment and protein-protein interaction (PPI) analysis combined with survival analysis were explored by integrated online database. Furthermore, taking advantage of Cytoscape software to construct the module gene network and identify the hub genes. By analyzing differentially co-expressed genes, this study provides a latent basis for clinical diagnosis or treatment to comprehend the etiology and potential molecular events of COAD.

Materials And Methods

Data Collection from TCGA and GEO Database

The gene expression profiles of COAD were downloaded from TCGA (<https://portal.gdc.cancer.gov/>) and GEO (<https://www.ncbi.nlm.nih.gov/gds>). In the TCGA database, all data on COAD and corresponding clinical information were freely downloaded by TCGA GDC Data portal[12]. There were 437 COAD samples, including 398 tumor and 39 normal tissues. After filtering using function `rpkm` in `edgeR` package [13], which is calculated by dividing gene counts by gene length, a total of 14,280 genes with RPKM values were subject to our next analysis. On the other hand, Microarray data include a gene expression profile data set (GSE44076) and GPL13667 platform [HG-U219] Affymetrix human genome U219 array. GSE44076 consisted of 98 tumor samples and 148 paired normal tissues from patients with COAD, which was downloaded from GEO website. We used PERL (version 5.30.1) operation arithmetic software to sort out the platform file and probe matrix input file[14]. PERL was used to process data by matching it with a specific probe according to the relationship between gene name and probe matrix, and finally convert the probe matrix into gene matrix.

Construction of Gene Co-expression Module and Detection of Differentially Expressed Genes (DEGs)

Co-expression networks boost network-based gene screening methods which can be used to identify candidate biomarkers and therapeutic targets. In our study, the gene expression data profiles of TCGA-

COAD and GSE44076 were constructed to gene co-expression networks using the WGCNA package in R programming language(4.1.2) [6]. The Go.db (version 3.12.1), preprocessCore (version 1.52.1), impute (version 1.64.0), and limma package were used to hold and handle the obtained datasets in Bioconductor. Using the correlation coefficient and the corresponding p value, we obtained several modules, which reflected the relationship between normal tissue and COAD. After all the modules obtained from WGCNA analysis, based on the most significant correlation, we selected the best colon cancer-related gene module and obtained the information module about genes. Differential expression analyses on RNA-Sequencing and microarray data were integrated by the R limma[15]. The p-value was adjusted by t value and to control for the false discovery Rate (FDR). Genes with the cutoff criteria of $|\log_{2}FC| \geq 1.0$ and $adj. P < 0.05$ were regarded as DEGs. The DEGs of the TCGA-COAD and GSE44076 dataset were visualized as a volcano plot and hot plot by using the R package ggplot2[16, 17]. Subsequently, the overlapping genes between DEGs and co-expression genes that were extracted from the co-expression network were used to identify potential prognostic genes, which were presented as a Venn diagram using the R package Venn Diagram[18].

Functional Enrichment Analysis and Hub Genes Identification

The functional enrichment analysis included Gene Ontology (GO) enrichment which comprised biological process (BP), molecular function (MF) and cellular component (CC), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis by DEGs. Dose [19](3.16.0), clusterProfiler [20] (3.18.1) and enrichplot (1.10.2) package in Bioconductor and colorspace, stringi, ggplot2 package which are analyzed in R. The results of GO enrichment and KEGG enrichment analysis were calculated based on $P < 0.05$. The STRING (Search Tool for the Retrieval of Interacting Genes) online database allowed us to build a protein-protein interaction network (PPI) network based on selected genes. Based on a maximum required interaction score of 0.9, We used Cytoscape[21] (version 3.9.0) software to complete the analysis and visualization of the PPI network. In a co-expression network, Maximal Clique Centrality (MCC) algorithm was reported to be the most effective method of finding hub nodes. The MCC of each node is calculated by CytoHubba[22], plug-in in Cytoscape, computed the MCC of each node to perform the hub genes analysis. In this study, the genes with the top 20 MCC values were considered as the hub genes. GEPIA(<http://gepia.cancer-pku.cn/#index>) is an open online database that provides gene expression profiles of all tumor samples and paired normal tissues. In our study, GEPIA database was used to verify the hub gene expression.

Correlation Analysis between Target Genes and Colon Carcinoma

Using the survival package in R software, we performed Kaplan–Meier univariate survival analysis to investigate the relationship between overall survival rate (OS) and hub genes. The survival related hub genes with $\log_{2}rank P < 0.05$ was statistically significant. UALCAN (ualcan.path.uab.edu/home) is a comprehensive, user-friendly, and interactive web resource for analyzing cancer OMICS data, which allow users to identify biomarkers or provide graphs and plots depicting expression profile and patient survival information[23]. We used UALCAN for quantification of B3GNT6 gene relative expression across different

grades of tumors and healthy tissue, also several other types of clinicopathological characteristics. Moreover, to further explore the potential immunomodulatory mechanism of B3GNT6 in the regulation of tumor-infiltrating immune cells, we used online TIMER (<https://cistrome.shinyapps.io/timer/>) database. R's CIBERSORT[24] software package was used to detect the proportion of 22 immune cells with high and low significant gene expression in COAD samples so as to make us better explore the relationship between tumor development and genes.

Validation of Protein Expressions by the HPA Database

The Human Protein Atlas (HPA) (<https://www.proteinatlas.org/>) is a rich and reliable database that provides a large number of transcriptomic and proteomic data of specific human tissues and cells[25]. It can automatically generate immunohistochemical images for scholars to use. IHC based protein expression pattern is the most common application of immunostaining to detect the relative position and abundance of proteins[26]. HPA was used to directly view immunohistochemical pictures of protein expression of B3GNT6 between normal and tumor samples in this research.

Results

Construction of Weighted Gene Co-expression Network Analysis

Based on the standard $P < 0.05$, $|\log_{2}FC| > 1$, we used R package to analyze the TCGA-COAD and GSE44076 data respectively to obtain meaningful differentially expressed genes. The top 50 with the most significant difference were selected to draw the heatmap and volcano map, in which the red part represents the up-regulated genes and the green part represents the down-regulated genes as shown in Fig. 1,2. Subsequently, A gene co-expression network was constructed from TCGA-COAD and GSE44076 data sets using WGCNA software package to find functional clusters in COAD patients. In the case of assigning color to each module, 10 modules of TCGA-COA (Fig. 3A) gain color assignment (remove a gray module that is not assigned to any cluster), and 7 modules in GSE44076 (Fig. 4A) get different colors in this article. Subsequently, we mapped the module feature relationship to assess the association between each module and two clinical features (cancer and normal), The outcomes of the module trait relationship are shown in Fig. 3B,4B. It was found that the brown module ($r = 0.86$, $p = 3e-131$) in TCGA-COAD and the turquoise module ($r = 0.89$, $p = 2e-87$) in GSE44076 had the positively highest correlation with normal tissues respectively (Fig. 3C,4C).

Identification between DEG Databases and Co-expression Modules

Based on the cut-off criteria of $|\log_{2}FC| \geq 1.0$ and $\text{adj. } P < 0.05$, It was found that 1913 DEGs in TCGA dataset and 206 DEGs in GSE44076 dataset were dislocate in tumor tissues,480 and 417 co-expressed genes were found in the brown module of TCGA dataset and the turquoise module of GSE44076 Separately. Finally, a total of 451 intersection genes were extracted to verify the genes of the co-expression modules (Fig. 4A).

Functional Enrichment Analyses and Identification of Hub Gene

The clusterProfiler software package was used for gene enrichment analysis to further understand the potential functions of 451 genes overlapping with DEG lists and two co-expression modules. After screening the GO enrichment analysis, we observed that biological processes (BP) mainly focus on lipid catabolic process (Fig. 5 and Table 1). The results of cell composition (CC) showed that these genes were mainly involved in the apical part of cell. In molecular function (MF) analysis, phosphoric ester hydrolase activity was considered to be related to 451 genes. Besides, Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of top ten genes in DEGs showed that the genes are mainly related to mineral absorption and PPAR signaling pathway (Fig. 6, Table 2).

Table 1
The GO function enrichment analysis of DEGs in COAD

| GO | Category | Description | Count | % | Pvalue | Qvalue |
|------------|----------|---|-------|-------|----------|----------|
| GO:0016042 | BP | Lipid catabolic process | 30 | 0.073 | 1.90E-11 | 6.17E-08 |
| GO:0006631 | BP | Fatty acid metabolic process | 29 | 0.071 | 8.83E-06 | 7.83E-06 |
| GO:0006066 | BP | alcohol metabolic process | 27 | 0.066 | 1.03E-05 | 9.13E-06 |
| GO:0044242 | BP | cellular lipid catabolic process | 23 | 0.056 | 5.02E-07 | 5.02E-07 |
| GO:0007586 | BP | digestion | 16 | 0.039 | 4.98E-08 | 1.62E-05 |
| GO:0045177 | CC | apical part of cell | 39 | 0.114 | 1.01E-13 | 3.31E-11 |
| GO:0016324 | CC | apical plasma membrane | 34 | 0.100 | 1.66E-12 | 2.70E-10 |
| GO:0031253 | CC | cell projection membrane | 25 | 0.073 | 2.20E-07 | 1.19E-05 |
| GO:0098862 | CC | cluster of actin-based cell projections | 19 | 0.055 | 2.84E-09 | 1.85E-07 |
| GO:0005903 | CC | brush border | 18 | 0.052 | 1.67E-11 | 1.81E-09 |
| GO:0042578 | MC | phosphoric ester hydrolase activity | 25 | 0.073 | 1.10E-06 | 1.50E-04 |
| GO:0008509 | MC | anion transmembrane transporter activity | 21 | 0.050 | 1.10E-05 | 6.01E-04 |
| GO:0008081 | MC | phosphoric diester hydrolase activity | 15 | 0.035 | 1.41E-09 | 7.69E-07 |
| GO:0016616 | MC | oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor | 14 | 0.033 | 8.74E-07 | 1.50E-04 |
| GO:0016616 | MC | oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor | 14 | 0.033 | 8.74E-07 | 1.50E-04 |

Table 2
The KEGG function enrichment analysis of DEGs in COAD

| GO | Description | Count | % | Pvalue | Qvalue |
|----------|-----------------------------------|-------|-------|----------|-------------|
| hsa04978 | Mineral absorption | 10 | 0.043 | 6.23E-06 | 0.000764207 |
| hsa03320 | PPAR signaling pathway | 10 | 0.043 | 4.71E-05 | 0.003854135 |
| hsa04530 | Tight junction | 10 | 0.043 | 2.24E-02 | 0.189207 |
| hsa00071 | Fatty acid degradation | 9 | 0.038 | 2.55E-06 | 0.000625852 |
| hsa00830 | Retinol metabolism | 9 | 0.038 | 1.20E-04 | 0.004907811 |
| hsa00982 | Drug metabolism - cytochrome P450 | 9 | 0.038 | 1.88E-04 | 0.005812421 |
| hsa04976 | Bile secretion | 9 | 0.038 | 9.23E-04 | 0.022638297 |
| hsa04936 | Alcoholic liver disease | 9 | 0.038 | 1.99E-02 | 0.178525176 |
| hsa01212 | Fatty acid metabolism | 8 | 0.034 | 1.90E-04 | 0.005812421 |
| hsa00983 | Drug metabolism - other enzymes | 8 | 0.034 | 1.90E-03 | 0.038794788 |

The PPI network between overlapping genes was established by using STRING database. The hub genes selected in the PPI network shows the MCC algorithm using the CytoHubba plug-in (Fig. 4B). According to the statistics of MCC, the top 20 genes with the highest scores are selected as the hub genes include Acetyl-CoA acyltransferase2 (ACAA2), Acyl-CoA dehydrogenase short chain (ACADS), Fatty acid binding protein 1 (FABP1), Carnitine palmitoyltransferase 1A (CPT1A), Acyl-CoA oxidase 1(ACOX1), Metallothionein 1G (MT1G), Metallothionein 1E (MT1E), Enoyl-CoA hydratase and 3-hydroxyacyl CoA dehydrogenase (EHHADH), Carnitine palmitoyltransferase 2 (CPT2), Acyl-CoA dehydrogenase medium chain (ACADM), Metallothionein 1X (MT1X), Metallothionein 1H (MT1H), PPARG coactivator 1 alpha (PPARGC1A), Acyl-CoA synthetase short chain family member 2 (ACSS2), Metallothionein 2A (MT2A), Metallothionein 1F (MT1F), Carnitine O-acetyltransferase (CRAT), UDP glucuronosyltransferase family 2 member B17 (UGT2B17), Beta-1,3-N-acetylglucosaminyltransferase 6 (B3GNT6), and Mucin 4 (MUC4) as shown in Fig. 4C. After screening 20 hub genes through the CytoHubba plug-in, we verified the expression level of hub gene in patients by convenient online database GEPIA. Compared with normal tissues, 20 hub genes in COAD were significantly down regulated as shown in Fig. 7C,7D and supplement 1,2.

Correlation of Hub Gene Expression and Colon Carcinoma

Kaplan Meier plotter used R survival software package to analyze OS of 20 hub genes to investigate the prognostic value of hub genes in COAD patients. Among the 20 hub genes, Kaplan Meier analysis showed that the low expression level of CPT1A and B3GNT6 was significantly correlated with better OS in patients with COAD ($P < 0.05$) (Fig. 7A,7B and supplement 3,4). Moreover, according to the UALCAN database, the protein level of B3GNT6 gene in tumor tissues was significantly lower than that in normal tissues (Fig. 8A). There was an obvious correlated among clinical stages, Gender and age with low

expression level of B3GNT6 (Fig. 8B, C, D). Also, we applied TIMER to analysis that low expression level of B3GNT6 is significant in different tissue include colon adenocarcinoma, esophageal carcinoma and prostate adenocarcinoma (Fig. 9A). TIMER database was used to analyze the correlation between B3GNT6 expression and immune infiltrating cells in COAD. The results showed that the negative expression of the B3GNT6 gene was related to the expression levels of B cells ($r = 0.1708$, $P < 0.05$) and CD4 + T cells ($r = 0.0986$, $P < 0.05$) (Fig. 9B). CIBERSORT analysis showed that the expression level of B3GNT6 was related to tumor immune cell infiltration include NK cell activated ($p = 0.030$), Macrophages M1 ($P = 0.006$) and Mast cells activated ($p = 0.037$) (Fig. 9C). All the above observations confirmed that the deregulation of B3GNT6 expression was interrelated with COAD.

Expression of the B3GNT6 Gene in HPA

By searching the HPA database, we entered B3GNT6 as the target gene into the website. The output showed a significant difference, in which B3GNT6 was lightly stained in colon cancer tissues, while it was significantly stained in normal colon tissues (Fig. 10).

Discussion

COAD is the most common intestinal tumor in the world, among the 36 cancers estimated worldwide in 2018, the number of new cases and related deaths of colon cancer ranked fourth, with an estimated 1100000 new cases[27, 28]. Although the treatment of colon cancer has improved, the prognosis of patients is usually poor due to tumor recurrence and metastasis, even if the tumor tissue is removed before tumor metastasis. Precise molecular targeted therapy gives patients more treatment opportunities. For example, Yu et al discovered that ZIC2 gene knockout inhibited the growth of colon cancer cells, prevented the transformation of cell cycle from G0 / G1 phase to S phase, and inhibited tumor formation[29]. Wang et al. had confirmed that CPT1A mediated fat oxidation increased the metastatic ability and drug resistance of colorectal cancer cells which could be a meaningful biomarker[30]. Therefore, we choose B3GNT6 for further research.

In this study, we identified 451 important genes COAD differentially expressed genes (DEG) using integrated bioinformatic analysis, these results suggested that these modular genes play an important role in the occurrence and development of COAD. The functional annotation analysis of cluster profiler software package showed that these genes were mainly enriched in metabolic processes and signaling transport pathways, which is basic process of cell function change and progress. Furthermore, based on the MCC scoring function of CytoHubba plug-in in Cytoscape, the first 20 COAD related genes were screened out (namely ACAA2, FABP1, ACOX1, EHHADH, CPT2, ACADS, CPT1A, MT1G, MT1E, MT1X, MT1H, PPARGC1A, ACSS2, MT2A, MT1F, CRAT, UGT2B17, B3GNT6, and MUC4). It was found that all their expression patterns were down regulated in cells when compared with normal control group and tumor tissue. We also applied GEPIA database to verify the reliability of the data. Among them, the down-regulated expression of CPT1A and B3GNT6 was significantly correlated with the high overall survival

rate of colon cancer. Finally, we systematically analyzed the clinical correlation with B3GNT6 in colon carcinoma.

B3GNT6, the protein encoded by this gene is β -1,3-n-acetylglucosamine transferase, which adds an N-acetylglucosamine moiety to N-acetylgalactosamine modified serine or threonine[31]. The encoded enzyme is responsible for creating the core 3 structure of O-glycan, which is an important component of mucin glycoprotein[32]. At present, there are few studies on the molecular mechanism of B3GNT6. Research reports say that the increase of B3GNT6 mRNA plays an important role in the formation and maintenance of impaired intestinal mucus stability during zinc deficiency[33]. Expression of B3GNT6 in human prostate cancer cells inhibits tumor growth and metastasis[34]. Wang et al show that core B3GNT6 affect EMT-MTE plasticity of colorectal cancer cells through MUC1/p53/mir-200c dependent signal cascade[35]. B3GNT6 is down regulated in colon cancer and deeply inhibits the metastatic potential of cancer cells[36]. That was consistent with our finding of survival analysis.

Tumor immunotherapy is a hot spot in tumor therapy, which is widely used in the research and treatment of COAD[37]. The immune infiltration of tumor cells and lymph node metastasis are closely related to the prognosis of colon cancer. TIMER database analysis showed that the expression level of B3GNT6 in COAD was positively correlated with the expression level of CD4 + T cells and B cells. The expression level of B3GNT6 in COAD was positively correlated with the expression level of B cells, CD4 + T cells, macrophages, neutrophils and dendritic cells. Also, the proportion of 22 tumor immune cells in COAD was determined by CIBERSORT analysis, the result showed that B3GNT6 may affect the immune infiltration of COAD by affecting the expression of NK cell activated, Macrophages M1 and Mast cells activated. B cells and NK cells are important immune cells in the body and have a wide range of anti-tumor effects[38, 39]. Enhancement of NK cells in lung cancer can block the signal pathway of transforming growth factor B and play an anti-tumor role[40]. Germain et al. found that lung cancer patients with high-density B cells had a better prognosis[41]. We deem that the low expression of B3GNT6 genes may trigger antitumor immune response, indicating they may play a significant role in the immune system. All these studies have strengthened the link between B3GNT6 expression and cancer progression. However, there are some unavoidable defects in our research. First, the small number of data sets may lead to the deviation of results. Second, the molecular mechanism of immunity needs more experimental such as western blotting and prospective studies to explore and testify our hypothesis.

In conclusion, by integrating WGCNA and differential gene expression analysis, our study mined an important survival related genes B3GNT6, which may predict the prognosis of COAD.

Declarations

DATA AVAILABILITY STATEMENT

All available data were analyzed in this study. These can be found here: TCGA (<https://portal.gdc.cancer.gov/>), GEO (<https://www.ncbi.nlm.nih.gov/gds>), STRING (<https://cn.string-db.org/>), GEPIA(<http://gepia.cancer-pku.cn/#index>), R(<https://www.r-project.org/>),

TIMER(<https://cistrome.shinyapps.io/timer/>) and HPA (<https://www.proteinatlas.org/>), UALCAN(Ualcan.path.uab.edu/analysis)

ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

FUNDING

This research was supported by the Natural Science Foundation of Jiangxi Provincial [20171BAB205064].

AUTHOR CONTRIBUTIONS

LB, FC wrote the paper. ZW edited the paper. All authors contributed to the article and approved the submitted version. All authors have no related financial or non-financial conflicts of interest.

Acknowledgments

We acknowledge the TCGA, GEO, HPA, GEPIA2, R, UALCAN, STRING and other tools for free use.

References

1. Siegel, R.L., et al., *Colorectal cancer statistics, 2020*. CA Cancer J Clin, 2020. **70**(3): p. 145–164.
2. Sung, H., et al., *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*. CA Cancer J Clin, 2021. **71**(3): p. 209–249.
3. Arnold, M., et al., *Global patterns and trends in colorectal cancer incidence and mortality*. Gut, 2017. **66**(4): p. 683–691.
4. Dekker, E., et al., *Colorectal cancer*. Lancet, 2019. **394**(10207): p. 1467–1480.
5. Can, T., *Introduction to bioinformatics*. Methods Mol Biol, 2014. **1107**: p. 51–71.
6. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis*. BMC Bioinformatics, 2008. **9**: p. 559.
7. Long, J., et al., *Transcriptional landscape of cholangiocarcinoma revealed by weighted gene coexpression network analysis*. Brief Bioinform, 2021. **22**(4).
8. Langfelder, P. and S. Horvath, *Fast R Functions for Robust Correlations and Hierarchical Clustering*. J Stat Softw, 2012. **46**(11).
9. Shi, M., et al., *APC(CDC20)-mediated degradation of PHD3 stabilizes HIF-1a and promotes tumorigenesis in hepatocellular carcinoma*. Cancer Lett, 2021. **496**: p. 144–155.

10. Yang, Y., et al., *Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types*. Nat Commun, 2014. **5**: p. 3231.
11. Guillotin, D., et al., *Transcriptome analysis of IPF fibroblastic foci identifies key pathways involved in fibrogenesis*. Thorax, 2021. **76**(1): p. 73–82.
12. Colaprico, A., et al., *TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data*. Nucleic Acids Res, 2016. **44**(8): p. e71.
13. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data*. Bioinformatics, 2010. **26**(1): p. 139–40.
14. Fourment, M. and M.R. Gillings, *A comparison of common programming languages used in bioinformatics*. BMC Bioinformatics, 2008. **9**: p. 82.
15. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies*. Nucleic Acids Res, 2015. **43**(7): p. e47.
16. Ito, K. and D. Murphy, *Application of ggplot2 to Pharmacometric Graphics*. CPT Pharmacometrics Syst Pharmacol, 2013. **2**(10): p. e79.
17. Postma, M. and J. Goedhart, *PlotsOfData-A web app for visualizing data together with their summaries*. PLoS Biol, 2019. **17**(3): p. e3000202.
18. Gao, C.H., G. Yu, and P. Cai, *ggVennDiagram: An Intuitive, Easy-to-Use, and Highly Customizable R Package to Generate Venn Diagram*. Front Genet, 2021. **12**: p. 706907.
19. Ritz, C., et al., *Dose-Response Analysis Using R*. PLoS One, 2015. **10**(12): p. e0146021.
20. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters*. Omics, 2012. **16**(5): p. 284–7.
21. Otasek, D., et al., *Cytoscape Automation: empowering workflow-based network analysis*. Genome Biol, 2019. **20**(1): p. 185.
22. Chin, C.H., et al., *cytoHubba: identifying hub objects and sub-networks from complex interactome*. BMC Syst Biol, 2014. **8 Suppl 4**(Suppl 4): p. S11.
23. Chandrashekar, D.S., et al., *UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses*. Neoplasia, 2017. **19**(8): p. 649–658.
24. Newman, A.M., et al., *Robust enumeration of cell subsets from tissue expression profiles*. Nat Methods, 2015. **12**(5): p. 453–7.
25. Colwill, K. and S. Gräslund, *A roadmap to generate renewable protein binders to the human proteome*. Nat Methods, 2011. **8**(7): p. 551–8.
26. Cyriac, G. and L. Gandhi, *Emerging biomarkers for immune checkpoint inhibition in lung cancer*. Semin Cancer Biol, 2018. **52**(Pt 2): p. 269–277.
27. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2019*. CA Cancer J Clin, 2019. **69**(1): p. 7–34.
28. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. CA Cancer J Clin, 2018. **68**(6): p. 394–424.

29. Xu, Z., et al., *Multilevel regulation of Wnt signaling by Zic2 in colon cancer due to mutation of β -catenin*. *Cell Death Dis*, 2021. **12**(6): p. 584.
30. Wang, Y.N., et al., *CPT1A-mediated fatty acid oxidation promotes colorectal cancer cell metastasis by inhibiting anoikis*. *Oncogene*, 2018. **37**(46): p. 6025–6040.
31. Iwai, T., et al., *Molecular cloning and characterization of a novel UDP-GlcNAc:GalNAc-peptide beta1,3-N-acetylglucosaminyltransferase (beta 3Gn-T6), an enzyme synthesizing the core 3 structure of O-glycans*. *J Biol Chem*, 2002. **277**(15): p. 12802–9.
32. Ota, T., et al., *Complete sequencing and characterization of 21,243 full-length human cDNAs*. *Nat Genet*, 2004. **36**(1): p. 40–5.
33. Maares, M., et al., *Zinc Deficiency Disturbs Mucin Expression, O-Glycosylation and Secretion by Intestinal Goblet Cells*. *Int J Mol Sci*, 2020. **21**(17).
34. Lee, S.H., et al., *Core3 O-glycan synthase suppresses tumor formation and metastasis of prostate carcinoma PC3 and LNCaP cells through down-regulation of alpha2beta1 integrin complex*. *J Biol Chem*, 2009. **284**(25): p. 17157–17169.
35. Ye, J., et al., *Core 3 mucin-type O-glycan restoration in colorectal cancer cells promotes MUC1/p53/miR-200c-dependent epithelial identity*. *Oncogene*, 2017. **36**(46): p. 6391–6407.
36. Iwai, T., et al., *Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells*. *Proc Natl Acad Sci U S A*, 2005. **102**(12): p. 4572–7.
37. Liu, X.S., et al., *NPM1 Is a Prognostic Biomarker Involved in Immune Infiltration of Lung Adenocarcinoma and Associated With m6A Modification and Glycolysis*. *Front Immunol*, 2021. **12**: p. 724741.
38. Bruno, T.C., *New predictors for immunotherapy responses sharpen our view of the tumour microenvironment*. *Nature*, 2020. **577**(7791): p. 474–476.
39. Shevtsov, M. and G. Multhoff, *Immunological and Translational Aspects of NK Cell-Based Antitumor Immunotherapies*. *Front Immunol*, 2016. **7**: p. 492.
40. Germain, C., et al., *Presence of B cells in tertiary lymphoid structures is associated with a protective immunity in patients with lung cancer*. *Am J Respir Crit Care Med*, 2014. **189**(7): p. 832–44.
41. Yang, B., et al., *Blocking transforming growth factor- β signaling pathway augments antitumor effect of adoptive NK-92 cell therapy*. *Int Immunopharmacol*, 2013. **17**(2): p. 198–204.

Figures

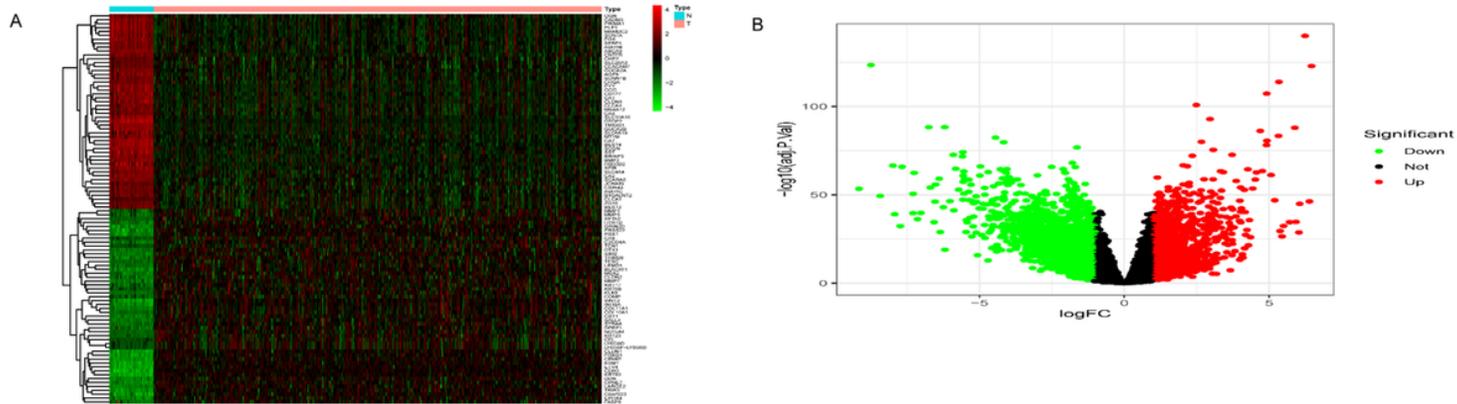


Figure 1

The first 50 significant differentially expressed genes obtained in the TCGA-COAD dataset. (A)Heatmap; (B)Volcano diagram.

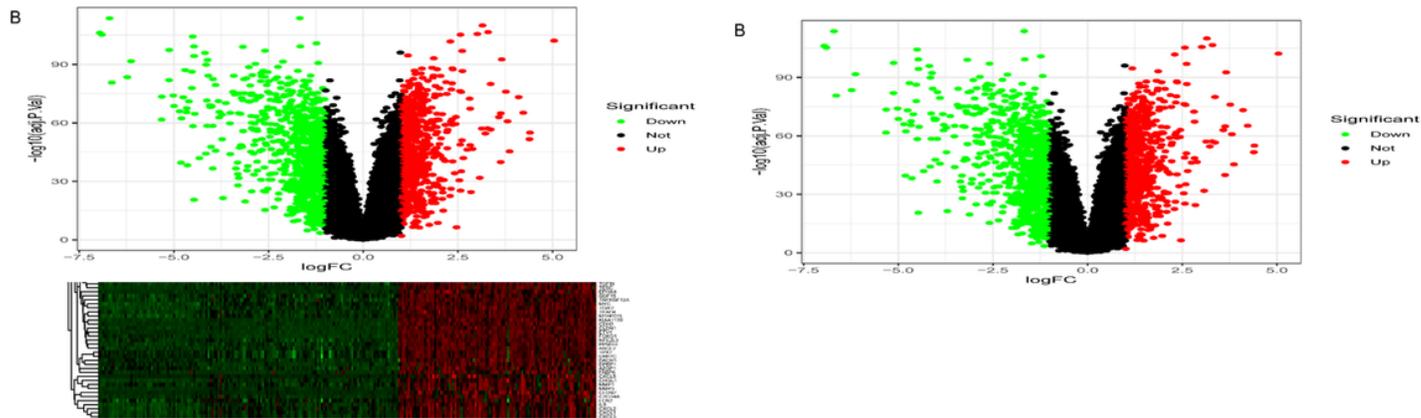


Figure 2

The first 50 significant differentially expressed genes obtained in the GSE44076 dataset. (A)Heatmap; (B)Volcano diagram.

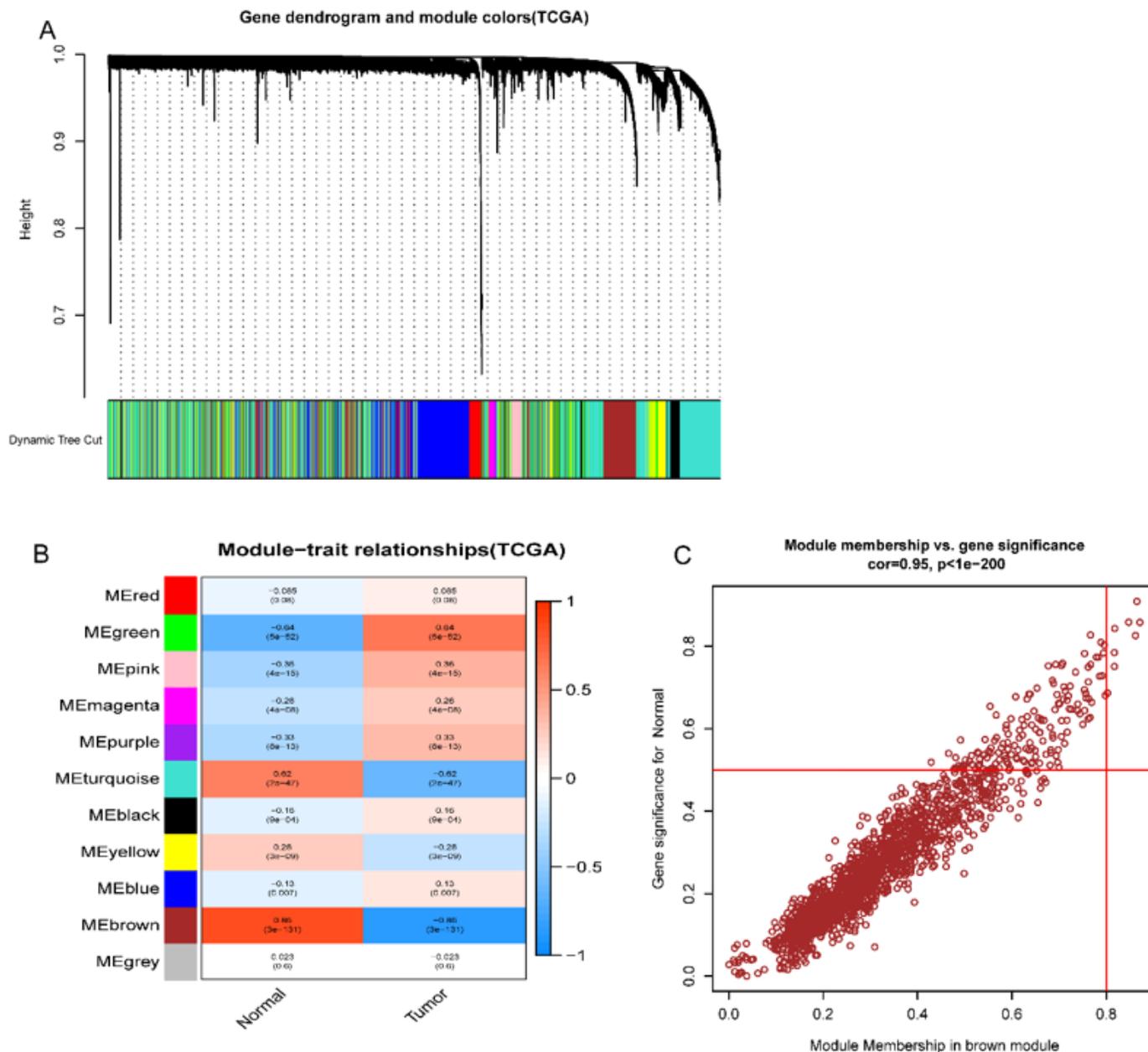


Figure 3

Figure 3 Identify the modules related to clinical information in the TCGA-COAD dataset. (A) The clustering tree module of co-expression network is sorted by gene hierarchical clustering method based on 1-TOM matrix. Each module is assigned a different color. (B) Module-trait relationship, each row corresponds to a color module, and each column corresponds to a clinical feature (cancer and normal). Each cell contains the corresponding correlation and P value. (C) Brown module showed the highest positive correlation in normal tissues.

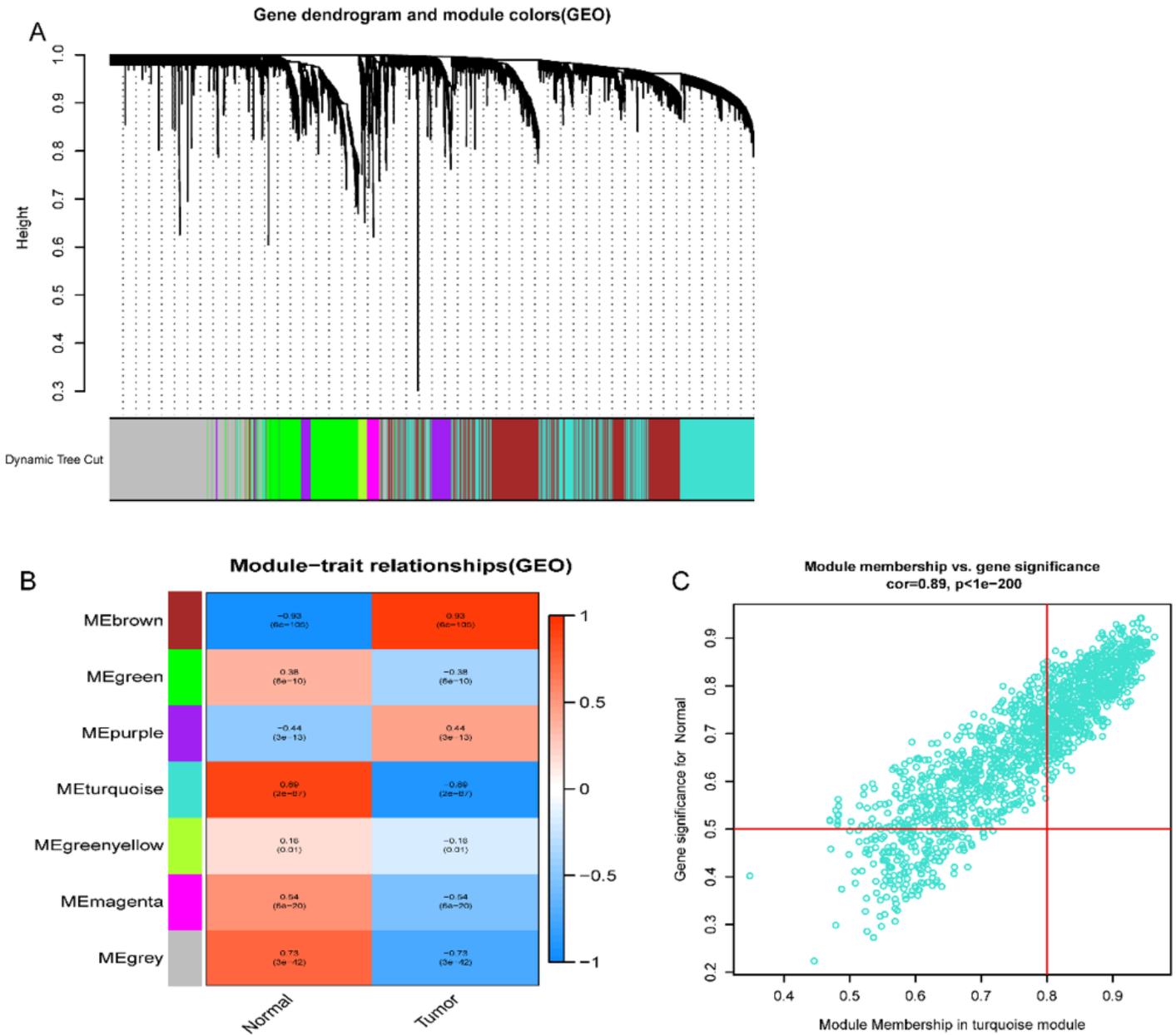


Figure 4

Figure 3 Identify modules related to clinical information in the GSE44076 dataset. (A) The clustering tree of co-expression network modules was sorted by gene hierarchical clustering based on 1-TOM matrix. Each module was assigned a different color. (B) Module trait relationship. Each row corresponds to a color module and each column corresponds to a clinical feature (cancer and normal). Each cell contains the corresponding correlation and P value. (C) turquoise module showed the highest positive correlation in normal tissues.

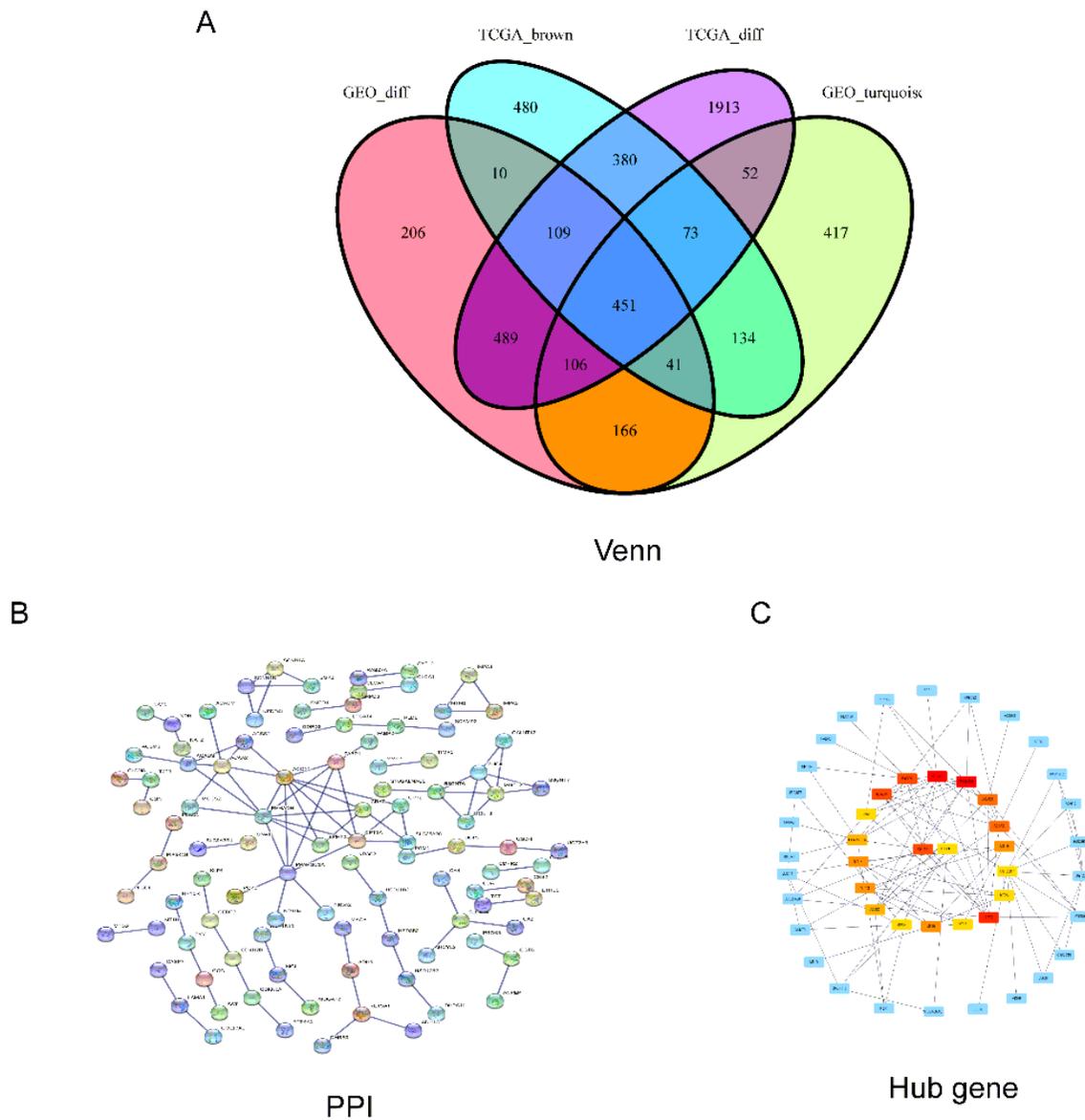
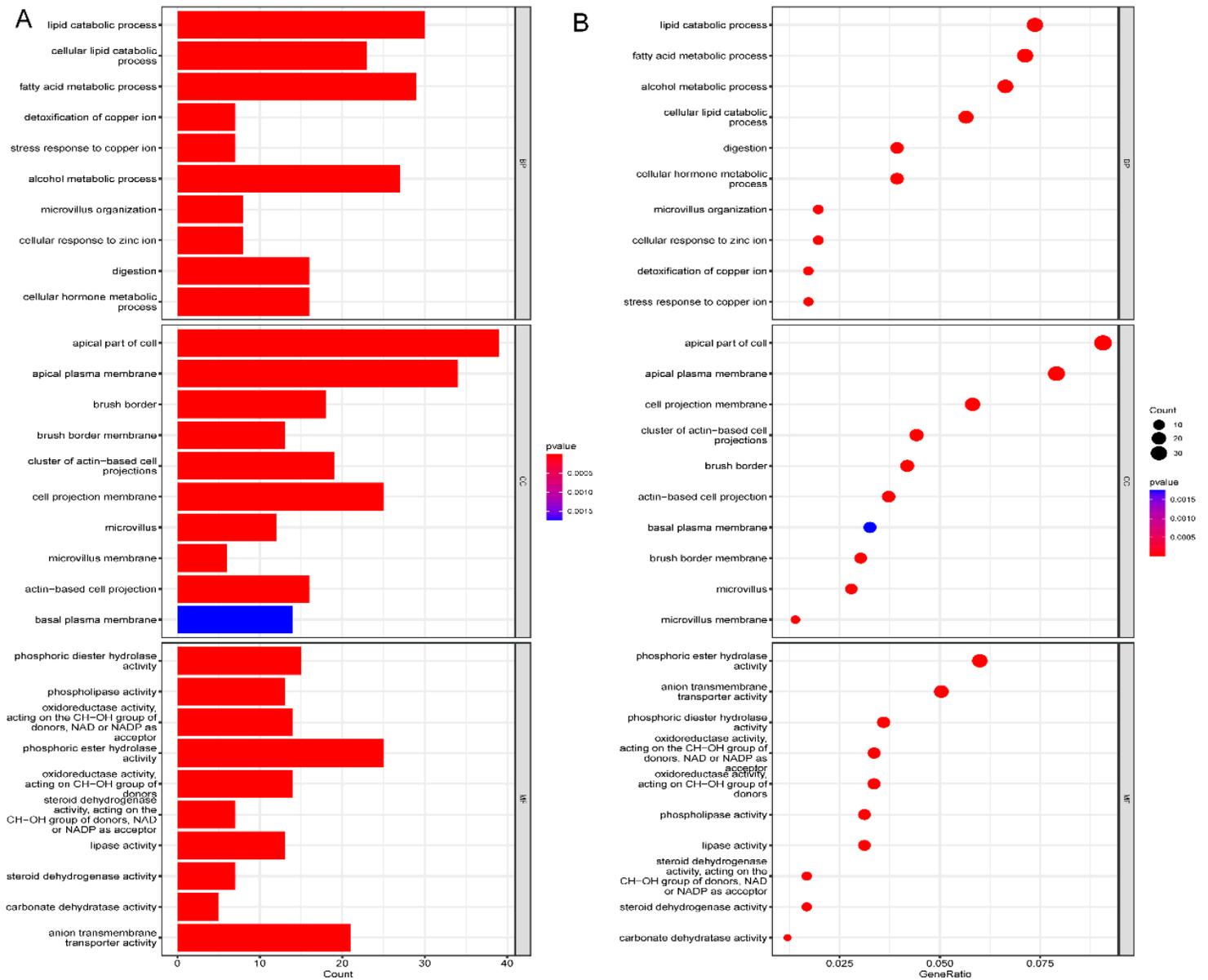


Figure 5

Figure 4 Acquisition and analysis of differentially expressed genes. (A) The genes in DEG tables and were co-expressed by Venn diagram. A total of 451 overlapping genes were located at the intersection of DEG list and two co-expression modules. (B) PPI network was established by using STRING database. (C) 20 hub genes were obtained by CytoHubba in Cytoscape.



GO

Figure 6

Figure 5 Gene Ontology (GO) enrichment analysis of top 10 genes in DEGs include top ten in BP, CC, MF.

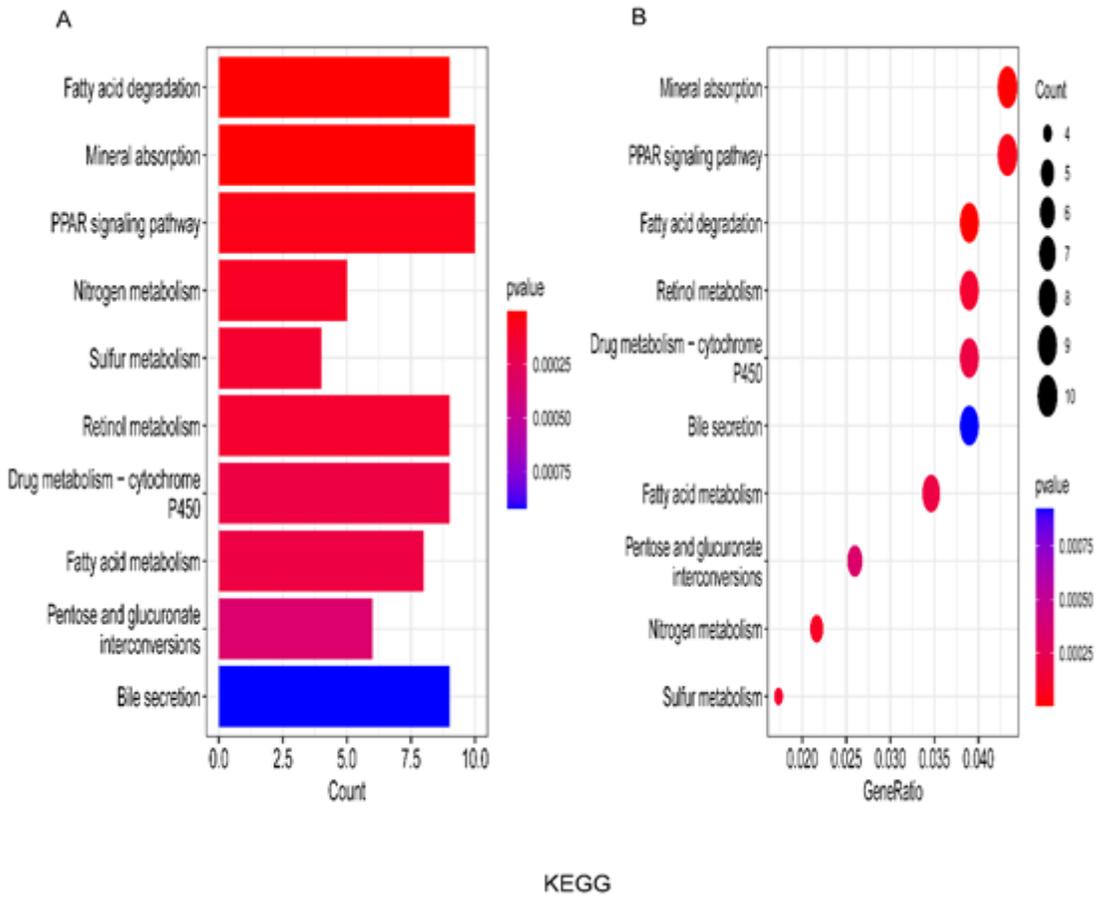


Figure 7

Figure 6 Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of top ten genes in DEGs.

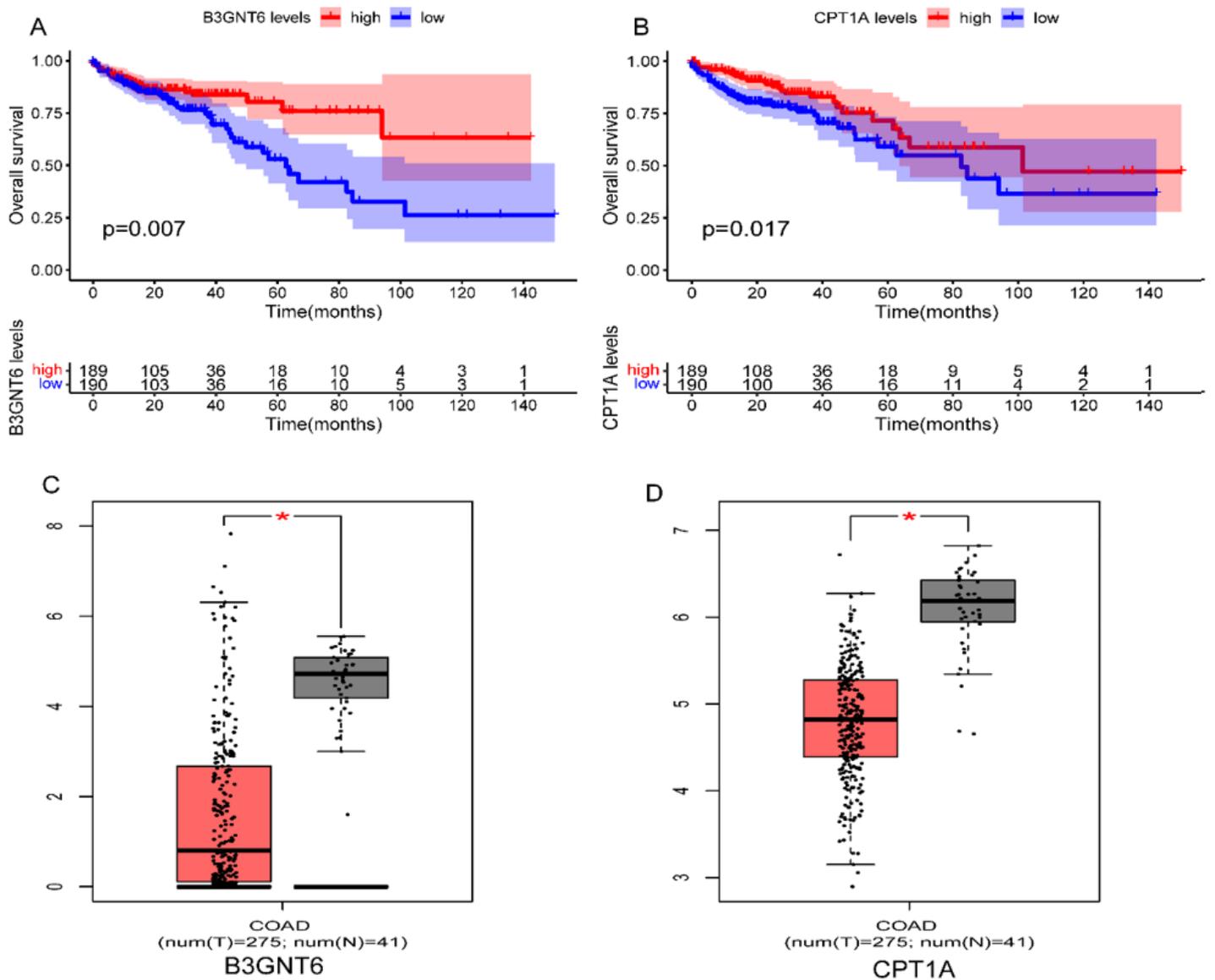


Figure 8

Figure 7 (A) Overall survival (OS) analysis of meaningful CPT1A gene in COAD patients from Kaplan–Meier univariate survival analysis. (B) Overall survival (OS) analysis of meaningful B3GNT6 gene in COAD patients from Kaplan–Meier univariate survival analysis. (C) Disease free survival (DFS) analysis of significant ACADS gene in COAD patients in GEPIA2 database. (D) Disease free survival (DFS) analysis of meaning CPT2 gene in COAD patients in GEPIA2 database.

Figure 9

Figure 8 Relationship between B3GNT6 gene expression and clinical features by UALCAN. (A) Expression of B3GNT6 gene in tumor tissues and normal tissues;(B) Comparison of B3GNT6 gene in different stages

of tumor tissue;(C) Comparison of B3GNT6 gene in different genders of tumor tissues;(D) Comparison of B3GNT6 gene in tumor tissues at different ages. * $P < 0.05$; ** $P < 0.001$; *** $P < 0.0001$.

Figure 10

Figure 9 Explore the relationship between B3GNT6 gene and immune microenvironment. (A) Expression of B3GNT6 gene in Pan-cancer by TIMER. (B) Correlation between B3GNT6 expression and immune infiltrating cells in COAD. (C) Expression level of B3GNT6 was related to tumor immune cell infiltration by CIBERSORT analysis.

Figure 11

Figure 10 Immunohistochemistry of B3GNT6 gene in COAD and normal tissues in human protein atlas (HPA) database. (A) Protein level of B3GNT6 in normal colonic mucosa. (B) Protein level of B3GNT6 in COAD.

Supplementary Files

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

- [supplement.docx](#)