Identification of molecular subgroups of ulcerative colitis by gene expression profile

Tenggang Ma (maxdoc@163.com)
Shandong University of Traditional Chinese Medicine https://orcid.org/0000-0003-3336-7928
Hongyu Zhang
Shandong University of Traditional Chinese Medicine
Ziwen Feng
Shandong University of Traditional Chinese Medicine
Renzhong Wang
Shandong University of Traditional Chinese Medicine Affiliated Hospital https://orcid.org/0000-0002-2850-285X

Research Article

Keywords: ulcerative colitis (UC), bioinformatics, molecular subtype, WGCNA module, Transcriptome classification

Posted Date: June 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-543947/v1

Abstract

Ulcerative colitis (UC) is a chronic inflammatory disease that is prone to recurrent attacks. It has complex pathogenesis, which is closely related to genetics, constitution, dietary habits, and environmental factors. From the comprehensive Gene Expression Omnibus database (GEO), we retrieved gene expression profiles and classified 197 cases of UC into three subgroups for the purpose of predicting the basic molecular characteristics for different types of ulcerative colitis. As expected, each group showed its own clinical peculiarity and way of presentation. In this article, consensus clustering was used to divide the sample into three. The WGCNA analysis was applied to evaluate specific modules and reveal transcriptional differences among the subgroups. Subsequently, pathway and function enrichment analysis was conducted based on WGCNA. In subgroup Ⅰ, fructose and mannose metabolism pathway and cell cycle 11/36 pathway are up-regulated, which could be an indicator of exacerbation. Furthermore, the hematopoietic cell lineage pathway, which was significantly up-regulated in subgroup Ⅱ, might be indicating a disease correlation. In subgroup Ⅲ, the gene expression pattern of the peroxisome pathway is similar to the normal group, which may indicate an early stage of UC. Although no significant prognostic difference existing among the groups, there were significant differences in their underlying biological characteristics. This suggests that transcriptome classifications also represent risk factors for different disease states and ages. In summary, the bioinformatics techniques used in this study contribute to identifying molecular subtypes for diagnosing human ulcerative colitis. The transcriptome classification of UC cases suggests that each subgroup may have its own gene expression pattern and pathway, providing further personalized treatment guidance for patients with ulcerative colitis.

1. Introduction

Ulcerative colitis is closely related to immune, genetic and environmental factors, which is increasing in incidence and prevalence year by year (Yadav V et al.,2016). Ulcerative colitis is a concern worldwide (Zhang YZ et al.,2014). In the United States, approximately 910,000 residents are troubled by UC (Ramos GP et al.,2018). The peak incidence is mainly concentrated in people aged 15 to 30 and 50 to 70 years old (Abraham C et al.,2009), and a considerable number of people have severely affected their work and life. In terms of the degree of colon involvement, ulcerative colitis is distributed in layers. It is mainly characterized by inflammation of the mucosal surface (Ordás I et al.,2012). Besides, the disruption of intestinal homeostasis can also lead to the occurrence of chronic colitis (Maloy, K. J. & Powrie, F,2011). It manifests as bloody diarrhea, abdominal pain, fatigue, and fecal incontinence (Kobayashi T et al.,2020). Ulcerative colitis recurs, and patients need to take medication for life, which leads to drug dependence and affects the life quality seriously (Rol Á et al.,2021& Torres J et al.,2015). Bioinformatics is an emerging field. It provides new ideas and basis for treating diseases from genomics, pharmacogenomics, pharmacological metabolomics transcriptomics, etc. Accordingly, it promotes the development of medicine (Oulas A et al.,2019). The recent genome-wide association study (GWAS) used scRNA-seq evaluated the gene expression patterns of specific subgroups of IBD intestinal cells. The results indicated that T cell subgroups might aggravate tissue damage (Smillie CS et al.,2019). A genome-wide expression

sequence study revealed that the differentially up-regulated genes in ulcerative colitis biopsy tissues include SAA1, DEFA5□DEFA6, MMP3, and MMP7 (Noble CL et al.,2018). A recent DNA hypo-methylation study with high-throughput sequencing technologies indicated that severe hypomethylation of UC is mediated by interactions between epithelium and lamina propria. Its anti-inflammatory genes are IL10, SIGLEC5, CD86, and CLMP and members of inflammasomes NLRP3 and NLRC4 (Taman H et al.,2021). Biotherapy with tumor necrosis factor (TNF) inhibitors has different efficacy in different subgroups of ulcerative colitis (Olesen CM et al.,2016). This may be because patients in different subgroups have their own gene expression modes.

2. Materials And Methods

2.1 Data resources

We use Geo Public Functional Genomics DataRepository (http://www.ncbi.nlm.nih.gov/geo/) to acquire the three gene expression profile data set GSE48958, GSE59071, GSE87466 and perl package for data processing.

2.2 Removal of batch effects and consensus clustering

We use the limma and sva packages in the R/conductor package (v4.0.3) (https://www.r-project.org/) to the performance of batch correction on the data to remove the influence of batch effects. Limma is an R/Bioconductor software package. It offers a systematic solution for analyzing provided gene expression data. The COMBAT function in the SVA package was used to normalize the gene expression values. To evaluate the elimination of effects, we apply the principal component analysis to the data. The batch effect between different platforms and batches would be removed.

Consensus clustering was performed by the "ConsensusClusterPlus" package in the R language on the samples of the experimental group. We set the maximum cluster figure to 10. And then, we apply the K-means algorithm for consensus clustering. The calculation of the K-means algorithm is carried out by Spearman distance. According to the consensus matrix result obtained, set the consensus score (> 0.8). The sample distance was calculated by Spearman distance, and the seed value was set to be 1 to 6. Then consistency score was conducted. The higher the consistency score is, the better the stability of the typing will be.

2.3 The clinical features of groups

Compare the proportions of active components of the disease activity among three subgroups. Comparisons between groups are implemented by a paired comparison test. Furthermore, a paired Wilcoxon rank-sum test was selected to check discrepancies in the age and disease status between subgroups.

2.4 Peculiar up-regulated genes of subgroups

Comparing the cases in the designated subgroup with the cases in each subgroup can clarify the up-regulated genes in the subgroup. Set $p < 0.05$, the threshold of mean difference > 0.2. For a certain gene, the mean difference was calculated by subtracting the average expression in the control from the cases in the subgroup.

2.5 WGCNA analysis

Weighted gene co-expression network analysis (WGCNA) is a typical system biology method (Langfelder P et al.,2018). We can use WGCNA to find modules of genes that are highly related. The module's core gene is considered the gene most closely related to disease, which has crucial biological significance in guiding diseases (Goh, K. I et al.,2007&Wang, T et al.,2017). WGCNA uses subgroup-specific signals to express associated patterns of the genes in the sample. At the same time, it can also clarify underlying effective modules of biofunctionality of each subgroup. The marginal value of the non-scale network was concluded according to the maximum R2. The average clustering and dynamic clustering methods were used to construct clustering tree and divide genes into modules, respectively. In the WGCNA software package, the Spearman's method was used to calculate the correlation coefficient. Then we can reach the relevant p-value on the clinical peculiarity and effective modules through cor function.

2.6 GSEA analysis

Gene Set Enrichment Analysis (GSEA) (Subramanian A et al.,2007), also known as gene probe enrichment analysis, is a calculation method of revealing genome expression data, analyzing and explaining the changes at the level of transcriptomics coordinating pathways between two biological states. Relevant biological pathways published in authoritative journals and experimental co-expression data are used to classify gene probes. A series of calculations based on the correlation can determine whether the probe set can reveal the distribution of related genomes on the phenotype. We use the GSEA v4.1.0 version for enrichment analysis.

2.7 Overrepresentation enrichment analysis

According to Fisher's exact test, overrepresentation enrichment analysis (ORA) was performed on the specific biomarkers of the subgroups. We perform this test in R software, which contained hypergeometric test. For each module, we chose KEGG pathways of the top 10 dominating ones. These gene sets came from the Molecular Signatures Database (MsigDB), one of the most accurate and all-around gene set databases for gene set enrichment analysis (Liberzon A et al.,2015).

3 Results

3.1 Characteristics of UC samples

To obtain reliable sample data, we selected three groups of ulcerative colitis mucosal biopsy samples in the GEO database, a total of 237 samples, including 197 UC samples and 40 healthy samples. The genetic data derives from the GEO database, and the three data sets are GSE87466(Telesco S et

al,2019，n = 108，UC = 87), GSE48958，Van der Goten J et al，2018，n = 21，UC = 13，, GSE59071，Arijs I et al，2018，n = 97，UC = 13，respectively. Besides, the GSE87466 dataset provides information on age, and the GSE48958 provides data on whether the state of the disease is active or not. Information on gender and race was not available.

3.2 Eliminate batch effects through cross-platform standardization

We applied the combat method to batch calibration on different platforms of data sets. The WGCNA package constructed the modules. At last, we chose 16454 genes on 237 samples. First, we select the first two primary components to cluster the samples (Fig. 1A), which were determined according to the expression value unnormalized and carried out in batches. Then batch effect was eliminated. Then through the normalization process, a scatter diagram was obtained (Fig. 1B), which indicated batch effect was eliminated.

FIGURE 1 Primary component analysis on samples from a different platform. Different colors represent samples from different data sets. The scatter plot points are due to the first two primary components of the gene expression profile (PC1 and PC2) to visualize the sample. (A) Before batch-effect removal, while (B) removes batch effects. Figure 1A is the result before batch calibration, while Fig. 1B is the result after batch calibration. The batch effect has been eliminated successfully.

3.3 Consensus clustering of UC cases

We obtained expression (GSE48958, GSE59071, GSE87466) data from the Gene Expression Omnibus (GEO) database of NCBI. We applied the R package to recognize the abnormally methylated differentially expressed genes (MDEG). The clusterProfiler software package was executed to perform enrichment analysis on functions and pathways. Gene Expression Synthesis (GEO) objects were extracted by using R/Bioconductor package GEOquery. The GEO object consists of probe sets, gene array, and clinical features. The followed-up analyses were performed according to the GEO sample data. After removing the batch effect of gene expression profiles, consensus clustering (unsupervised clustering method) divided 237 UC cases into subgroups. According to the cluster consensus matrix, when the cluster consensus score (> 0.8), it had the highest stability. Therefore, the samples were divided into three groups. Then we performed consensus clustering and obtained three subtypes, among which there were 99, 58, and 48 cases in these groups, respectively. The analysis was based on gene expression profiles. Their expression patterns were significantly different. In contrast, based on the consensus matrix, a high degree of similarity in gene expression patterns was observed in each subgroup (Fig. 2A). According to the cluster consensus matrix result, when the cluster consensus score is higher than 0.8, it had the highest stability.

FIGURE 2 Consensus cluster on gene expression profiles in patients with ulcerative colitis (UC). (A) A consensus matrix that the cluster number is 3. The heat map is up to the minimum consensus score of

the subtypes (> 0.8). (B)The bar graph represents the consensus scores of the subgroups with the number of clusters ranging from 2 to 10.

3.4 Comparison of clinical characteristics of subgroups

We conduct research on the clinical features of the three subtypes. The disease state and age data of UC cases were selected from the GSE48958 data set. Precisely, the active proportion in subtype III was markedly lower than subtype I and II,but it is evident that no remarkable divergence between subtype I and II on the proportion of active (Fig. 3A). However,there are no apparent differences between the three subgroups on age (Fig. 3B). Then the proportion of active diseases was visualized. Compare whether there was a difference in the proportion of active diseases between different types (Fig. 3A). The calculation of the pairwise function shows the proportion of diseases in the active state had a significant difference when the difference of p-value < 0.001.

FIGURE 3 The paired comparison between three subtypes on clinical features (PCA). (A) The bar-plot represents the proportion of active state. (B) displays the age of each subgroup. The abscissa is the classification of the samples into three groups, Ⅰ, Ⅱ, and Ⅲ. The ordinate represents the proportion of active diseases in figure A and the age of figure B. ***p-value < 0.001.

3.5 Transcriptome classification on GSEA

Gene Set Enrichment Analysis (GSEA) is an effective calculation means. It can determine the difference between gene sets. We mainly focus on the statistical discrepancy between biological features. There is no doubt that the discrepancy should be markedly and consistently. We used the predefined gene sets from functional explanation and ranked the genes of double sorts of samples. The ranking was mainly based on the differential expression. Subsequently, we checked the previous gene set enriched on the sorting table and identified if they were at the top or bottom. Through GSEA enrichment analysis, the specific differential genes in each type can be obtained, and whether the type is still different from the normal sample. The differential gene in the typing was the differential gene compared with the normal sample, indicating that the differential gene in the typing was also different when the typing compared with the normal sample (Fig. 4). The original chip data contained 16454 probes. No probe set = > gene symbol collapsing was requested, so all 16454 features were used. The genome size filter ($15 \leq n \leq 5000$) leads to screening out one gene set. One gene set was remarkably enriched on the condition of p-value < 1%.

The heat map was drawn with the limma and pheatmap packages. Analysis of the difference of classification was performed. Set the absolute value of the mean filter > 0.2 and the corrected p-value filter < 0.05 as the differential gene. The average value of multiple rows of genes was taken, deleted the genes whose expression was 0 in all samples. Limma was applied to analyze the difference and extract the up-regulated genes in each type.

Take the intersection of expression data and typing data to get the module to which each gene belongs. The samples were typed, and genes are annotated. The subgroup-I samples were up-regulated in the

green-yellow gene module and salmon gene module. The subgroup-II samples were up-regulated in the brown gene module and cyan gene module and down-regulated in the black gene module. The subgroup-III samples showed up-regulation in the black gene module and down-regulation in the brown and salmon gene modules.

In order to reveal the differences between the transcriptome UC subgroups, each subgroup made up the WGCNA level of gene expression. Differential expression analysis pairwise between every two subtypes. The number of genes up-regulated in subgroups I, II, and III was determined to be 497, 2839, and 2835 (mean deviation > 0.2, $p < 0.05$). In addition, differential expression analysis was performed by contrasting the gene expression profile of every subtype with the control profile. The subpopulation, specifically up-regulated genes, was significantly up-regulated in the casecontrol (Fig. 4A,4B,4C).

FIGURE 4 Overview of ES scores in operation and the position of gene set members on the ranking list. Three subgroups specifically up-regulated gene expression patterns. In the enrichment map(A), (B), and (C), compared with the control group, the up-regulated genes of the specific subgroup have higher expression. (D) Gene expression value was shown in the heat map for six WGCNA modules. Different colors represent the level of gene expression. Low expression was shown in blue, and high expression was shown in red. The vertical line represents the differential gene of each type, and the abscissa represents the difference between the type n and the normal sample. The closer to the left lies, the more significant the genetic differentiation.

3.6 WGCNA analysis and modules

Before establishing an appropriate gene co-expression module, a suitable power value was chosen for the comparative value. The target gene expression level was extracted. Clustering was performed in the absence of missing values and power index range 1:20 to find the best power value. Obtained a scatter plot of the fitting index and power value and a scatter plot of connectivity and power value (Fig. 5A). Therefore, according to the R2 value is not less than 0.8, the slope is close to -1, we chose the best power value to be 7 in subsequent analysis.

Meanwhile, it can be found that when the scale independence was > 0.8, there was higher average connectivity. The gene distance was calculated, and the genes were clustered. Then dynamic cutting module identification was performed. Each module identifies at least 60 genes and obtains the module to which each gene belongs. Find similar modules, clustering was performed, and the module was gotten that each gene ultimately belongs to. The module data and the clinical shape data were intersected, the clinical data and the module data were tested, the p-value of the correlation test was obtained, and the visualization was performed to obtain the heat map of the module and the trait data. Find the module where the gene was located, and classify the gene. The limma and WGCNA packages were used to perform WGCNA analysis. Subsequently, the amount of the genes was calculated in every module. At last, according to genes amount, we depict these modules. We used different colors to represent and arranged from high to low.

FIGURE 5 WGCNA analysis and modules. (A) Scale independence and mean connectivity, based on which the optimal power value can be obtained. (B) Functional characterization on WGCNA modules and clinical peculiarity. The correlation coefficients of UC's age and disease state are shown in orange and blue, respectively. The ordinate represents the module to which the gene belongs, and the abscissa represents the clinical traits. The correlation analysis between the module and clinical traits shows that if p-value < 0.05, it means that the module is correlated with clinical traits. Blue represents a negative correlation, and orange represents a positive correlation.

3.7 Association of WGCNA modules and Clinical features.

We computed the pertinence coefficient or p-value between each module's activity level or age and the characteristic gene so that the correlation between clinical features and WGCNA modules could be studied (Fig. 5B). It is worth noting that the feature vector represents the feature gene on the gene presentation array of every module. It was not difficult to see that module 2 was positively correlated with age, while module 6 was negatively correlated. Modules 1, 5, and 6 were related to disease states actively, while modules 2, 3, and 4 were passively correlated with disease states. In addition, 1,3,4 modules were displayed regardless of age, which may be indicated that there was no correlation between disease progression and age. It can be seen from our research that WGCNA was correlated with certain clinical characteristics, such as disease states.

3.8 GO and KEGG enrichment analysis

The enrichment analysis was performed by using clusterProfiler and enrichplot in the R package. Set p-value□0.05 and perform GO enrichment analysis on BP, CC, and MF, respectively. We compared the functions between modules through GO enrichment analysis. KEGG was a comparison between module pathways, and the enrichment conditions were the same as those of GO. BH or Benjamini & Hochberg(1995)was chosen as the calibration method. Then obtain the GO (Fig. 6) and KEGG (Fig. 7) enrichment pathway bubble chart.

FIGURE 6 GO enrichment bubble graph. Figure 7 KEGG enrichment bubble graph. The ordinate of bubble chart represents the name of the GO, and the abscissa represents the name of a module. The color is displayed according to the p-value. The redder the color, the more significant the enrichment of GO in the module. ▲: p-value□0.05,Significantly enriched;●□p-value□0.05□Enrichment is not significantly.

4. Discussion

This research initially used omics data such as disease gene expression profiling and high-throughput sequencing to classify the transcriptome and differentiate subgroups based on the clinical characteristics of the disease. For instance, a study using serum miRNAs to identify murine colitis subtypes showed that circulating miRNA expression profiles identified in mice can predict the disease state of UC patients, and Th2 may be a cytokine that mediates UC. Its accuracy rate can reach 83.3% in mice and 100% in humans (Viennois E et al.,2017). While in another immunophenotype and risk gene analysis study, it was found that in human intraepithelial T cells CD8Trm and γδ T cells, the fundamental expression of CD39 at the T

cell level was higher. In comparison, the corresponding expression was lower in mice. This difference might be due to differences in living conditions (Huang B et al.,2019). Unlike these studies, we further divided the disease into subgroups based on the study of ulcerative colitis and studied the different clinical characteristics of its transcriptomics.

This study analyzed UC cases, which consisted of three independent GEO data sets and normal control gene molecular subtypes. The data included clinically active and inactive patients. The data of different platforms was normalized, and different platforms or continuous additional effects were successfully removed. In addition, according to gene presentation profiles, we divided the 237 UC cases into three subtypes successfully. In subsequent analysis, functional modules of specific subgroups were displayed by transcriptional classification. Apparent correlations have been noted between transcriptional classification and clinical features. Comparing the three groups, subgroup I and II had a higher proportion than that of III, suggesting that subtype I and II displayed a more serious UC condition. It was meaningful that we had extensive sample data. Meanwhile, consensus score we clustered > 0.8 revealed that the transcriptional classification was reliable. Transcriptional categorization. In general, the transcriptional classification of UC instance was bound up with clinical features, particular modules' function, and pathways, which has profound guiding significance in UC disease treatment.

Our research use subgroup and WGCNA to better comprehend the gene expression profiles of ulcerative colitis. KEGG enrichment revealed that the peroxisome pathway was up-regulated in subgroup III, which can restore cells and cause inactivation of E. coli (Cevallos SA et al.,2021). PPARγ activates the function of regulating inflammation and immune response (Pedersen G et al.,2010). In UC, the sfTSLP expressed downregulating, and in the colonic mucosa, the PPARγ directly affects the transcriptional regulation of sfTSLP gene (Martin Mena A et al.,2017& Pedersen G et al.,2010). In addition, since the pathway expression mode is the same as normal group's, subgroup III is considered the bud stage of UC disease. For patients with ulcerative colitis, the expression of peroxisome proliferation-activated receptor (PPAR) γ located in the colonic epithelium decreases, which may be a crucial factor leading to inflammation and intestinal dysfunction (Bouguen G et al.,2015). Peroxisome inhibits the activity of NF-κB through negative regulation of transcription factors to achieve anti-inflammatory effects (Cao H et al.,2018). The ulcerative colitis (UC) model shows that berberine can change the levels of various fecal metabolites related to various pathways such as the citrate cycle (Liao Z et al.,2019). Valine can target colonic epithelial cells, high-affinity peptide transporter 1 (PepT1), which exhibits anti-inflammatory properties (Wu Y et al.,2019).

As we have seen, the hematopoietic cell lineage pathway of subgroup II is up-regulated, leading to a better response of granulocyte/monocyte apheresis (GMA) (Yokoyama Y et al.,2015). In UC patients, granulocyte/monocyte adsorption (GMA) ,associated with the arthritic monocyte phenotype, is used as the Adacolumn to adsorb leukocytes of the adherent lineage selectively (Takeda S et al.,2010). The lin – c-kit + cell subset expresses CD33, a microscopic lineage cell marker that can maintain the immune replacement of intestinal antigens (Chinen H et al.,2007). The hematopoietic cell lineage is the foremost approach leading to UC inflammation. Lineage initiation and lineage transformation of transcription factors can induce lineage transformation of committed cells (Defendenti C et al.,2012). Clearing these

cells from the circulatory system can enhance the efficacy of drugs and promote disease treatment (Sáez-González E et al.,2017). Experiments with ulcerative colitis in rats confirmed that adhesion molecules could achieve anti-ulcer effects and reduce inflammation by reducing the infiltration of neutrophils to obtain inhibition and antioxidant effects (Abdallah DM et al.,2011&Marafini I et al.,2014). After S100a9 antibody treatment, essential pathways related to CAC such as the cytokine-cytokine receptor interaction pathway are inhibited (Zhang X et al.,2017). Fructose and mannose metabolism can improve the intestinal flora and regulate intestinal function (Liu C et al.,2021). The destruction of the critical enzyme function of TGS is found in the "galactose metabolism" and "fructose and mannose metabolism" pathways, which are fructose kinase α- and β-galactosidase, which can cause the decomposition of TGS (Metzler-Zebeli BU et al.,2019& Jin M et al.,2018). Excessive cell death may lead to chronic inflammation, and irregular intestinal epithelial cells may lead to ulcerative colitis (Günther C et al.,2012). Different variations of cell cycle determine the difference effector, leading to ulcerative colitis or Crohn's disease (Sturm A et al.,2004). The exon sequencing results of ulcerative colitis showed that the expression of DUOX2 and DUOXA2 increased. At the same time, the cell cycle expression was up-regulated (Vinayaga-Pavan M et al.,2019). The crucial of sphingolipid metabolism is ceramide, which promotes the circulatory function of cells (Gomez-Larrauri A et al.,2020). In the synthesis of thyroid hormone, it has been confirmed that carbonic anhydrase II exists in the serum of patients with ulcerative colitis (Alver A et al.,2007). Colon disease may be induced by oocyte division and cell cycle (Wu D et al.,2016).

5. Conclusions

The case data of the three subgroups did not differ significantly in age. Although there is little difference in whether a group of patients is clinically active, there is a big difference between the two in terms of biological characteristics. Different from subgroup II, subgroup I is active in fructose and mannose metabolism, cell cycle, thyroid hormone synthesis, and oocyte meiosis. In contrast, subgroup II is active in hematopoietic cell lineage, osteoclast differentiation, cell adhesion molecules, B cell receptor signaling pathway, cytokine – cytokine receptor interaction, and other aspects. Subgroup III is diverse from subgroup I and subgroup II, which is active in Peroxisome, Thermogenesis, Citrate cycle, Valine, leucine, and isoleucine degradation, etc. Since the expression pattern is similar to normal cases, it suggests that subgroup III may be in the early stage of the disease. In summary, our data provide a comparison of the clinical characteristics of transcriptomics data of UC patients between different subgroups.

Our research of the subgroup classification shows that three subgroups have their own expression patterns and indicate different stages of disease development. This suggests that different subgroups should implement proper treatment plans that are suitable for each more accurately. Our research results require more follow-up data and experimental methods for more precise subgroup classification and will be strongly confirmed in the future.

Declarations

All authors have contributed to the creation of this manuscript for important intellectual content and read and approved the final manuscript.

Fundings

The funding is provided by the authors ourselves and there are no other funding supporters.

Ethics approval and consent to participate

Not applicable.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the GEO repository, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87466; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48958; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59071

Authors' contributions

Ma Tenggang prepared for the data collection in the early stage of the article, and performed follow-up analysis to complete the main part of the manuscript. Zhang Hongyu analyzes the data, makes graphs and tables and complete some part of the manuscript. Feng Ziwen carried out data proofreading, article layout and content editing. Wang Renzhong guided and revised this article.

Acknowledgements

Thank you Professor Wang for his teaching guidance and friends for their support and help.

Competing Interests

The authors declare that they have no conflicts of interest.

References

1. Abdallah DM, Ismael NR. Resveratrol abrogates adhesion molecules and protects against TNBS-induced ulcerative colitis in rats. Can J Physiol Pharmacol. 2011;89(11):811–8. doi:10.1139/y11-080.
2. Abraham C, Cho JH. Inflammatory bowel disease. N Engl J Med. 2009 Nov 19;361(21):2066-78. doi: 10.1056/NEJMra0804647. PMID: 19923578; PMCID: PMC3491806.
3. Alver A, Menteşe A, Karahan SC, Erem C, Keha EE, Arikan MK, Eminağaoğlu MS, Deger O. Increased serum anti-carbonic anhydrase II antibodies in patients with Graves' disease. Exp Clin Endocrinol Diabetes. 2007 May;115(5):287 – 91. doi: 10.1055/s-2007-960498. PMID: 17516290.

4. Bouguen G, Langlois A, Djouina M, Branche J, Koriche D, Dewaeles E, Mongy A, Auwerx J, Colombel JF, Desreumaux P, Dubuquoy L, Bertin B. Intestinal steroidogenesis controls PPARγ expression in the colon and is impaired during ulcerative colitis. Gut. 2015 Jun;64(6):901–10. doi:10.1136/gutjnl-2014-307618. Epub 2014 Jul 22. PMID: 25053717.
5. Cao H, Liu J, Shen P, Cai J, Han Y, Zhu K, Fu Y, Zhang N, Zhang Z, Cao Y. Protective Effect of Naringin on DSS-Induced Ulcerative Colitis in Mice. J Agric Food Chem. 2018 Dec 19;66(50):13133–13140. doi: 10.1021/acs.jafc.8b03942. Epub 2018 Dec 4. PMID: 30472831.
6. Cevallos SA, Lee JY, Velazquez EM, Foegeding NJ, Shelton CD, Tiffany CR, Parry BH, Stull-Lane AR, Olsan EE, Savage HP, Nguyen H, Ghanaat SS, Byndloss AJ, Agu IO, Tsolis RM, Byndloss MX, Bäumler AJ. 5-Aminosalicylic Acid Ameliorates Colitis and Checks Dysbiotic Escherichia coli Expansion by Activating PPAR-γ Signaling in the Intestinal Epithelium. mBio. 2021 Jan 19;12(1):e03227-20. doi: 10.1128/mBio.03227-20. PMID: 33468700; PMCID: PMC7845635.
7. Chinen H, Matsuoka K, Sato T, Kamada N, Okamoto S, Hisamatsu T, Kobayashi T, Hasegawa H, Sugita A, Kinjo F, Fujita J, Hibi T. Lamina propria c-kit + immune precursors reside in human adult intestine and differentiate into natural killer cells. Gastroenterology. 2007 Aug;133(2):559–73. doi: 10.1053/j.gastro.2007.05.017. Epub 2007 May 21. PMID: 17681176.
8. Defendenti C, Grosso S, Atzeni F, Croce A, Senesi O, Saibeni S, Bollani S, Almasio PL, Bruno S, Sarzi-Puttini P. Unusual B cell morphology in inflammatory bowel disease. Pathol Res Pract. 2012 Jul 15;208(7):387 – 91. doi: 10.1016/j.prp.2012.05.003. Epub 2012 Jun 2. PMID: 22658383.
9. Goh KI, Cusick ME, alle V, Childs D, Vidal B, M., and Barabasi AL. (2007).The human disease network. Proc. Natl. Acad. Sci. U.S.A. 104, 8685–8690.doi: 10.1073/pnas.0701361104.
10. Gomez-Larrauri A, Presa N, Dominguez-Herrera A, Ouro A, Trueba M, Gomez-Muñoz A. Role of bioactive sphingolipids in physiology and pathology. Essays Biochem. 2020 Sep 23;64(3):579–589. doi: 10.1042/EBC20190091. PMID: 32579188.
11. Günther C, Neumann H, Neurath MF, Becker C. Apoptosis, necrosis and necroptosis: cell death regulation in the intestinal epithelium. Gut. 2013 Jul;62(7):1062-71. doi: 10.1136/gutjnl-2011-301364. Epub 2012 Jun 11. PMID: 22689519.
12. Huang B, Chen Z, Geng L, Wang J, Liang H, Cao Y, Chen H, Huang W, Su M, Wang H, Xu Y, Liu Y, Lu B, Xian H, Li H, Li H, Ren L, Xie J, Ye L, Wang H, Zhao J, Chen P, Zhang L, Zhao S, Zhang T, Xu B, Che D, Si W, Gu X, Zeng L, Wang Y, Li D, Zhan Y, Delfouneso D, Lew AM, Cui J, Tang WH, Zhang Y, Gong S, Bai F, Yang M, Zhang Y. Mucosal Profiling of Pediatric-Onset Colitis and IBD Reveals Common Pathogenics and Therapeutic Pathways. Cell. 2019 Nov 14;179(5):1160–1176.e24. doi: 10.1016/j.cell.2019.10.027. PMID: 31730855.
13. Jin M, Zhang H, Wang J, Shao D, Yang H, Huang Q, Shi J, Xu C, Zhao K. Response of intestinal metabolome to polysaccharides from mycelia of Ganoderma lucidum. Int J Biol Macromol. 2019 Feb 1;122:723–731. doi: 10.1016/j.ijbiomac.2018.10.224. Epub 2018 Nov 2. PMID: 30395901.
14. Kobayashi T, Siegmund B, Le Berre C, et al. Ulcerative colitis. Nat Rev Dis Prim. 2020;6:74.

15. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008 Dec;29:9:559. doi:10.1186/1471-2105-9-559. PMID: 19114008; PMCID: PMC2631488.
16. Liao Z, Zhang S, Liu W, Zou B, Lin L, Chen M, Liu D, Wang M, Li L, Cai Y, Liao Q, Xie Z. LC-MS-based metabolomics analysis of Berberine treatment in ulcerative colitis rats. J Chromatogr B Analyt Technol Biomed Life Sci. 2019 Dec 1;1133:121848. doi: 10.1016/j.jchromb.2019.121848. Epub 2019 Nov 1. PMID: 31756623.
17. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015 Dec 23;1(6):417–425. doi: 10.1016/j.cels.2015.12.004. PMID: 26771021; PMCID: PMC4707969.
18. Liu C, Du P, Guo Y, Xie Y, Yu H, Yao W, Cheng Y, Qian H. Extraction, characterization of aloe polysaccharides and the in-depth analysis of its prebiotic effects on mice gut microbiota. Carbohydr Polym. 2021 Jun 1;261:117874. doi: 10.1016/j.carbpol.2021.117874. Epub 2021 Mar 2. PMID: 33766361.
19. Maloy KJ, Powrie F. Intestinal homeostasis and its breakdown in inflammatory bowel disease. Nature. 2011 Jun 15;474(7351):298–306. doi: 10.1038/nature10208. PMID: 21677746.
20. Marafini I, Sedda S, Pallone F, Monteleone G. Targeting integrins and adhesion molecules to combat inflammatory bowel disease. Inflamm Bowel Dis. 2014 Oct;20(10):1885-9. doi: 10.1097/MIB.0000000000000091. PMID: 25215614.
21. Martin Mena A, Langlois A, Speca S, Schneider L, Desreumaux P, Dubuquoy L, Bertin B. The Expression of the Short Isoform of Thymic Stromal Lymphopoietin in the Colon Is Regulated by the Nuclear Receptor Peroxisome Proliferator Activated Receptor-Gamma and Is Impaired during Ulcerative Colitis. Front Immunol. 2017 Sep 4;8:1052. doi: 10.3389/fimmu.2017.01052. PMID: 28928735; PMCID: PMC5591373.
22. Metzler-Zebeli BU, Newman MA, Grüll D, Zebeli Q. Functional adaptations in the cecal and colonic metagenomes associated with the consumption of transglycosylated starch in a pig model. BMC Microbiol. 2019 May 2;19(1):87. doi: 10.1186/s12866-019-1462-2. PMID: 31046662; PMCID: PMC6498482.
23. Noble CL, Abbas AR, Cornelius J, Lees CW, Ho GT, Toy K, Modrusan Z, Pal N, Zhong F, Chalasani S, Clark H, Arnott ID, Penman ID, Satsangi J, Diehl L. Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis. Gut. 2008 Oct;57(10):1398 – 405. doi: 10.1136/gut.2008.148395. Epub 2008 Jun 3. PMID: 18523026.
24. Olesen CM, Coskun M, Peyrin-Biroulet L, Nielsen OH. Mechanisms behind efficacy of tumor necrosis factor inhibitors in inflammatory bowel diseases. Pharmacol Ther. 2016 Mar;159:110–9. doi:10.1016/j.pharmthera.2016.01.001. Epub 2016 Jan 22. PMID: 26808166.
25. Ordás I, Eckmann L, Talamini M, Baumgart DC, Sandborn WJ. Ulcerative colitis. Lancet. 2012;380(9853):1606–19. https://doi.org/10.1016/S0140-6736(12)60150-0.

26. Oulas A, Minadakis G, Zachariou M, Sokratous K, Bourdakou MM, Spyrou GM. Systems Bioinformatics: increasing precision of computational diagnostics and therapeutics through network-based approaches. Brief Bioinform. 2019 May 21;20(3):806–824. doi: 10.1093/bib/bbx151. PMID: 29186305; PMCID: PMC6585387.
27. Pedersen G, Brynskov J. Topical rosiglitazone treatment improves ulcerative colitis by restoring peroxisome proliferator-activated receptor-gamma activity. Am J Gastroenterol. 2010 Jul;105(7):1595–603. doi:10.1038/ajg.2009.749. Epub 2010 Jan 19. PMID: 20087330.
28. Pedersen G, Brynskov J. Topical rosiglitazone treatment improves ulcerative colitis by restoring peroxisome proliferator-activated receptor-gamma activity. Am J Gastroenterol. 2010 Jul;105(7):1595–603. doi:10.1038/ajg.2009.749. Epub 2010 Jan 19. PMID: 20087330.
29. Ramos GP, Papadakis KA. Mechanisms of Disease: Inflammatory Bowel Diseases. Mayo Clin Proc. 2019 Jan;94(1):155–165. doi: 10.1016/j.mayocp.2018.09.013. PMID: 30611442; PMCID: PMC6386158.
30. Rol Á, Todorovski T, Martin-Malpartida P, Escolà A, Gonzalez-Rey E, Aragón E, Verdaguer X, Vallès-Miret M, Farrera-Sinfreu J, Puig E, Fernández-Carneado J, Ponsati B, Delgado M, Riera A, Macias MJ. Structure-based design of a Cortistatin analogue with immunomodulatory activity in models of inflammatory bowel disease. Nat Commun. 2021 Mar 25;12(1):1869. doi: 10.1038/s41467-021-22076-5. PMID: 33767180.
31. Sáez-González E, Moret I, Alvarez-Sotomayor D, Díaz-Jaime FC, Cerrillo E, Iborra M, Nos P, Beltrán B. Immunological Mechanisms of Adsorptive Cytapheresis in Inflammatory Bowel Disease. Dig Dis Sci. 2017 Jun;62(6):1417–1425. doi: 10.1007/s10620-017-4577-z. Epub 2017 Apr 21. PMID: 28432476.
32. Smillie CS, Biton M, Ordovas-Montanes J, Sullivan KM, Burgin G, Graham DB, Herbst RH, Rogel N, Slyper M, Waldman J, Sud M, Andrews E, Velonias G, Haber AL, Jagadeesh K, Vickovic S, Yao J, Stevens C, Dionne D, Nguyen LT, Villani AC, Hofree M, Creasey EA, Huang H, Rozenblatt-Rosen O, Garber JJ, Khalili H, Desch AN, Daly MJ, Ananthakrishnan AN, Shalek AK, Xavier RJ, Regev A. Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell. 2019 Jul 25;178(3):714–730.e22. doi: 10.1016/j.cell.2019.06.029. PMID: 31348891; PMCID: PMC6662628.
33. Sturm A, Leite AZ, Danese S, Krivacic KA, West GA, Mohr S, Jacobberger JW, Fiocchi C. Divergent cell cycle kinetics underlie the distinct functional capacity of mucosal T cells in Crohn's disease and ulcerative colitis. Gut. 2004 Nov;53(11):1624–31. doi:10.1136/gut.2003.033613. PMID: 15479683; PMCID: PMC1774268.
34. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. Bioinformatics. 2007 Dec 1;23(23):3251-3. doi: 10.1093/bioinformatics/btm369. Epub 2007 Jul 20. PMID: 17644558.
35. Takeda S, Sato T, Katsuno T, Nakagawa T, Noguchi Y, Yokosuka O, Saito Y. Adsorptive depletion of alpha4 integrin(hi)- and CX3CR1hi-expressing proinflammatory monocytes in patients with ulcerative colitis. Dig Dis Sci. 2010 Jul;55(7):1886–95. doi:10.1007/s10620-009-0974-2. Epub 2009 Nov 12. PMID: 19908144.

36. Taman H, Fenton CG, Anderssen E, Florholmen J, Paulssen RH. DNA hypo-methylation facilitates anti-inflammatory responses in severe ulcerative colitis. PLoS One. 2021 Apr 1;16(4):e0248905. doi: 10.1371/journal.pone.0248905. PMID: 33793617; PMCID: PMC8016308.
37. Torres J, Boyapati RK, Kennedy NA, Louis E, Colombel JF, Satsangi J. Systematic Review of Effects of Withdrawal of Immunomodulators or Biologic Agents From Patients With Inflammatory Bowel Disease. Gastroenterology. 2015 Dec;149(7):1716–30. doi:10.1053/j.gastro.2015.08.055. Epub 2015 Sep 14. PMID: 26381892.
38. Viennois E, Zhao Y, Han MK, Xiao B, Zhang M, Prasad M, Wang L, Merlin D. Serum miRNA signature diagnoses and discriminates murine colitis subtypes and predicts ulcerative colitis in humans. Sci Rep. 2017 May 31;7(1):2520. doi: 10.1038/s41598-017-02782-1. PMID: 28566745; PMCID: PMC5451415.
39. Vinayaga-Pavan M, Frampton M, Pontikos N, Levine AP, Smith PJ, Jonasson JG, Björnsson ES, Segal AW, Smith AM. Elevation in Cell Cycle and Protein Metabolism Gene Transcription in Inactive Colonic Tissue From Icelandic Patients With Ulcerative Colitis. Inflamm Bowel Dis. 2019 Jan 10;25(2):317–327. doi: 10.1093/ibd/izy350. PMID: 30452647; PMCID: PMC6327231.
40. Wang T, He X, Liu X, Liu Y, Zhang W, Huang Q, et al. Weighted gene co-expression network analysis identifies FKBP11 as a key regulator in acute aortic dissection through a NF-kB dependent pathway. Front Physiol. 2017;8:1010. doi:10.3389/fphys.2017.01010.
41. Wu D, Li Q, Song G, Lu J. Identification of disrupted pathways in ulcerative colitis-related colorectal carcinoma by systematic tracking the dysregulated modules. J BUON. 2016;21(2):366–74.
42. Wu Y, Sun M, Wang D, Li G, Huang J, Tan S, Bao L, Li Q, Li G, Si L. A PepT1 mediated medicinal nano-system for targeted delivery of cyclosporine A to alleviate acute severe ulcerative colitis. Biomater Sci. 2019 Oct 1;7(10):4299–4309. doi: 10.1039/c9bm00925f. Epub 2019 Aug 13. PMID: 31408067.
43. Yadav V, Varum F, Bravo R, Furrer E, Bojic D, Basit AW. Inflammatory bowel disease: exploring gut pathophysiology for novel therapeutic targets. Transl Res. 2016 Oct;176:38–68. doi: 10.1016/j.trsl.2016.04.009. Epub 2016 May 6. PMID: 27220087.
44. Yokoyama Y, Watanabe K, Ito H, Nishishita M, Sawada K, Okuyama Y, Okazaki K, Fujii H, Nakase H, Masuda T, Fukunaga K, Andoh A, Nakamura S. Factors associated with treatment outcome, and long-term prognosis of patients with ulcerative colitis undergoing selective depletion of myeloid lineage leucocytes: a prospective multicenter study. Cytotherapy. 2015 May;17(5):680–8. doi: 10.1016/j.jcyt.2015.02.007. Epub 2015 Mar 21. PMID: 25804800.
45. Zhang X, Wei L, Wang J, Qin Z, Wang J, Lu Y, Zheng X, Peng Q, Ye Q, Ai F, Liu P, Wang S, Li G, Shen S, Ma J. Suppression Colitis and Colitis-Associated Colon Cancer by Anti-S100a9 Antibody in Mice. Front Immunol. 2017 Dec 13;8:1774. doi: 10.3389/fimmu.2017.01774. PMID: 29326691; PMCID: PMC5733461.
46. Zhang YZ, Li YY. Inflammatory bowel disease: pathogenesis. World J Gastroenterol. 2014 Jan 7;20(1):91 – 9. doi: 10.3748/wjg.v20.i1.91. PMID: 24415861; PMCID: PMC3886036.

Tables

Due to technical limitations, table 1, 2 is only available as a download in the Supplemental Files section.

Figures

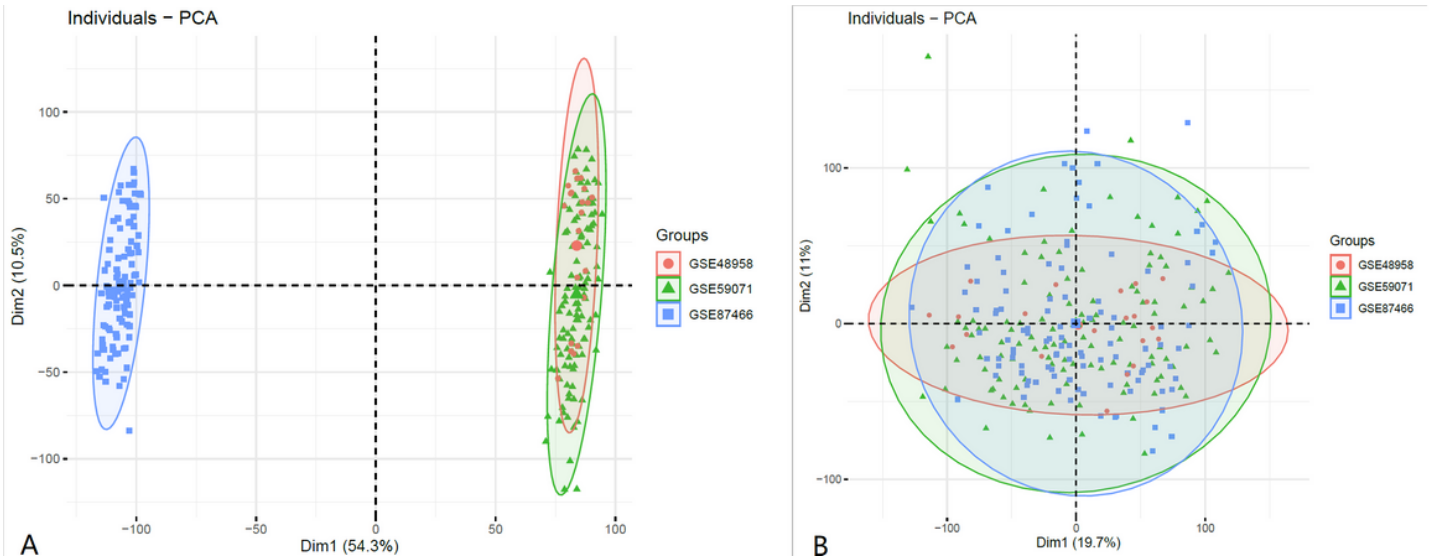

Figure 1

Primary component analysis on samples from a different platform. Different colors represent samples from different data sets. The scatter plot points are due to the first two primary components of the gene expression profile (PC1 and PC2) to visualize the sample. (A) Before batch-effect removal, while (B) removes batch effects. Figure 1A is the result before batch calibration, while Figure 1B is the result after batch calibration. The batch effect has been eliminated successfully.

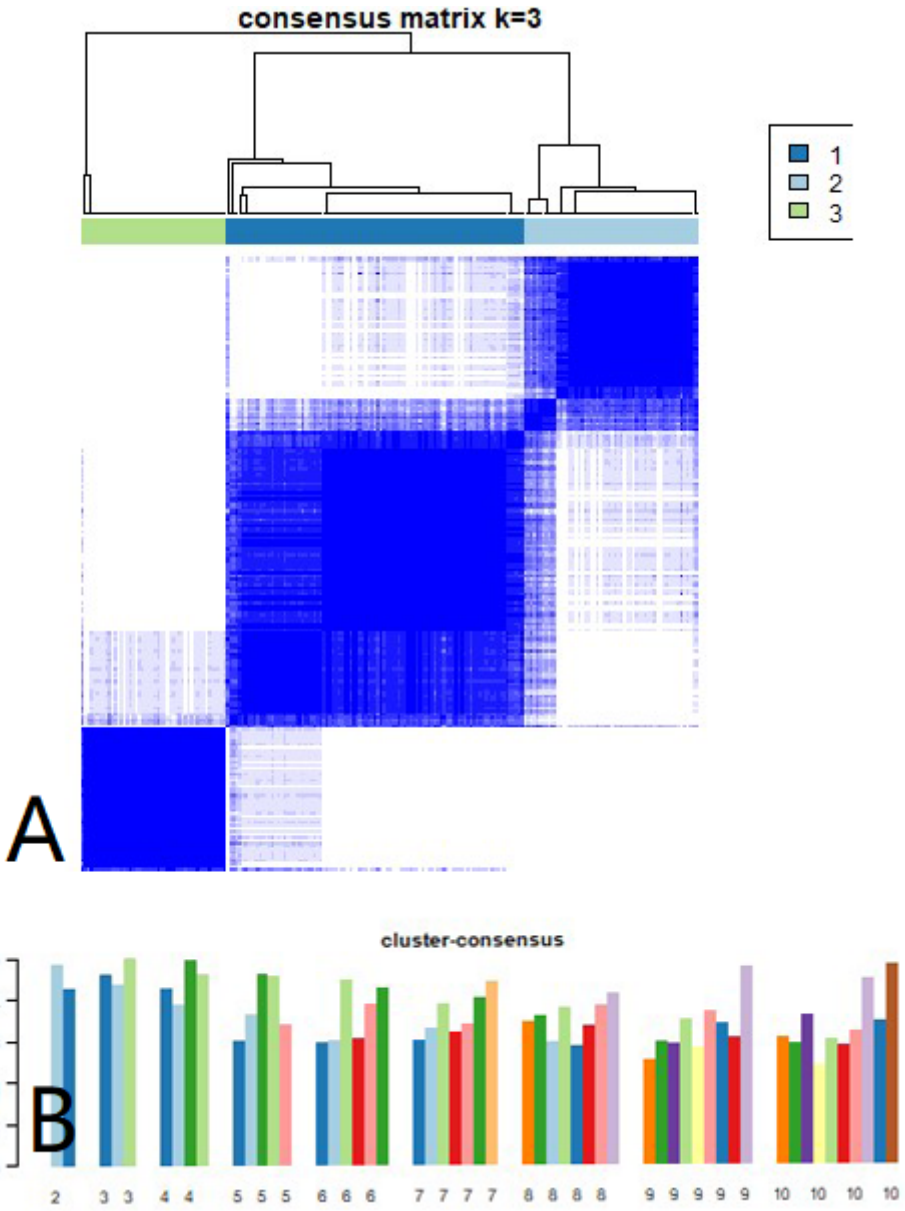

Figure 2

Consensus cluster on gene expression profiles in patients with ulcerative colitis (UC). (A) A consensus matrix that the cluster number is 3. The heat map is up to the minimum consensus score of the subtypes (> 0.8). (B)The bar graph represents the consensus scores of the subgroups with the number of clusters ranging from 2 to 10.

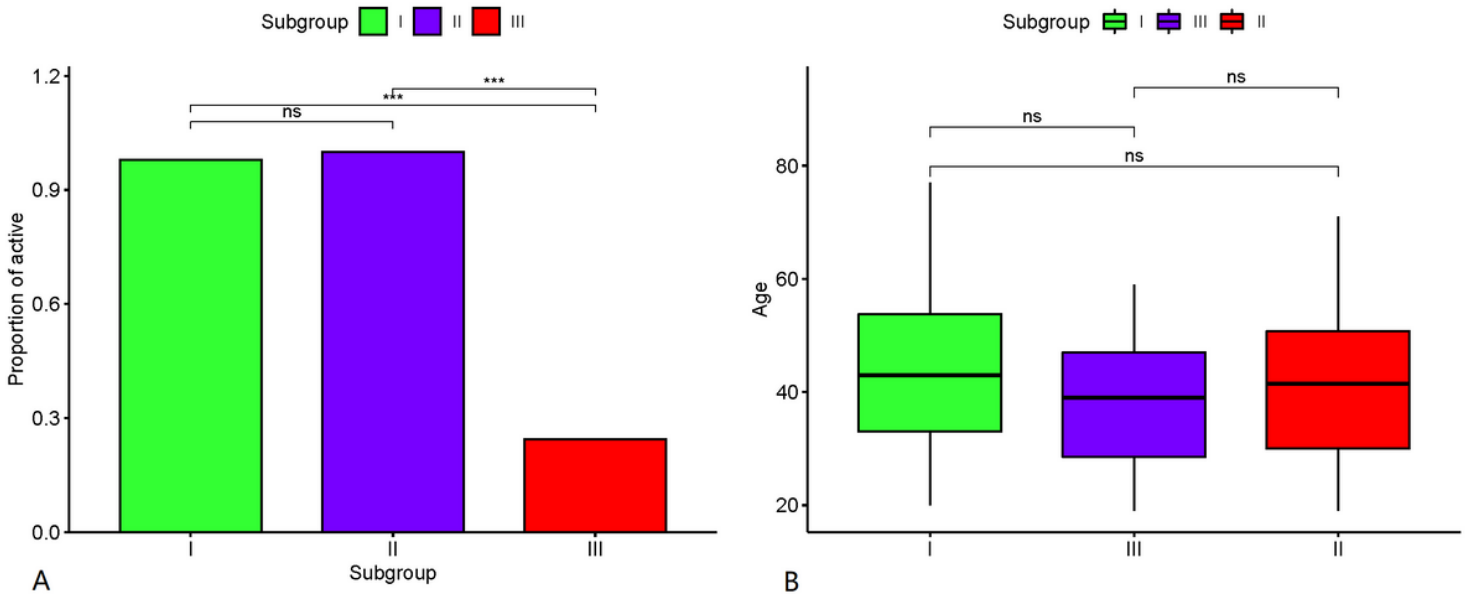

Figure 3

The paired comparison between three subtypes on clinical features (PCA). (A) The bar-plot represents the proportion of active state. (B) displays the age of each subgroup. The abscissa is the classification of the samples into three groups, Ⅰ, Ⅱ, and Ⅲ. The ordinate represents the proportion of active diseases in figure A and the age of figure B. ***p-value<0.001.

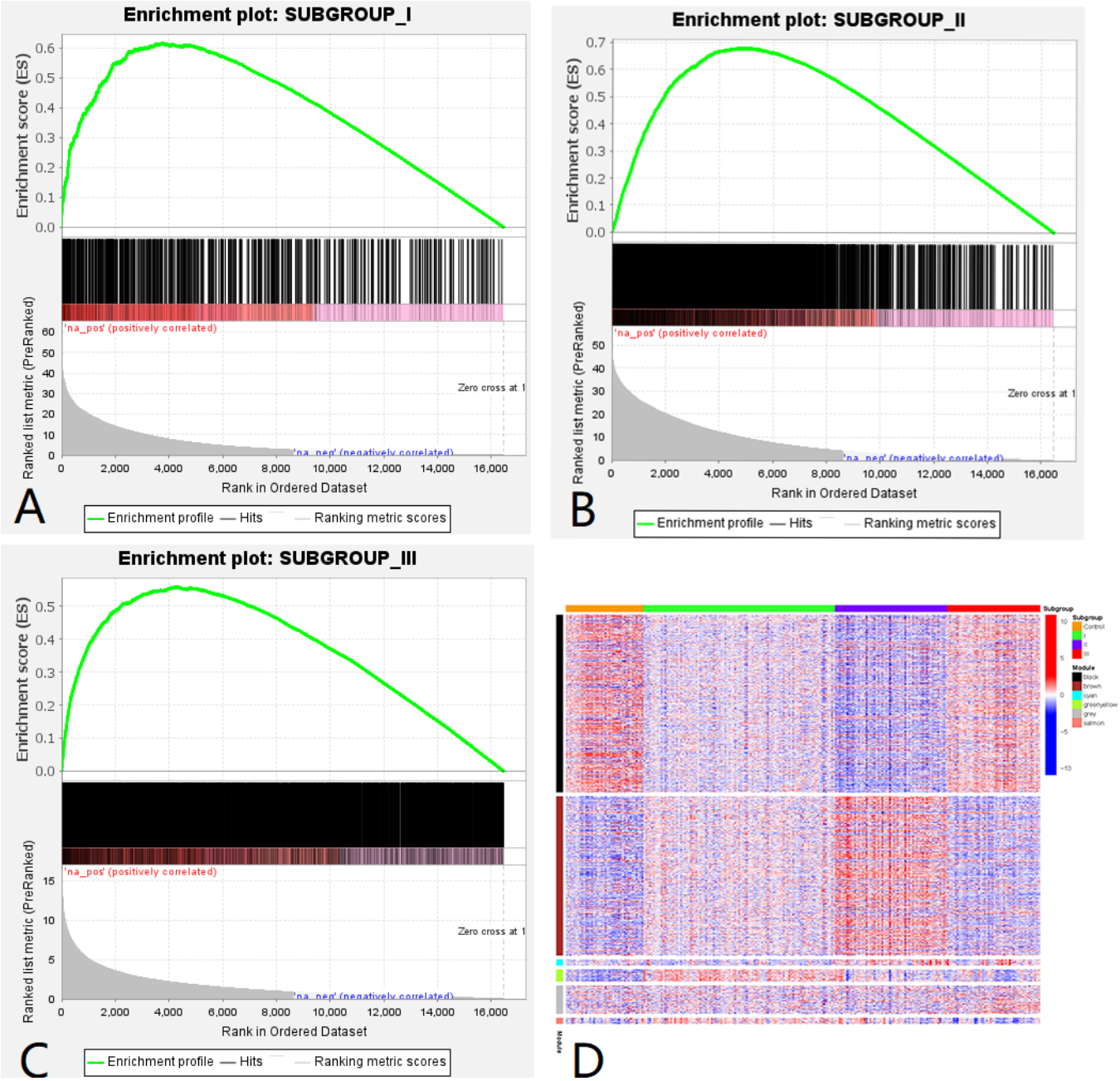

Figure 4

Overview of ES scores in operation and the position of gene set members on the ranking list. Three subgroups specifically up-regulated gene expression patterns. In the enrichment map(A), (B), and (C), compared with the control group, the up-regulated genes of the specific subgroup have higher expression. (D) Gene expression value was shown in the heat map for six WGCNA modules. Different colors represent the level of gene expression. Low expression was shown in blue, and high expression was shown in red. The vertical line represents the differential gene of each type, and the abscissa represents the difference between the type n and the normal sample. The closer to the left lies, the more significant the genetic differentiation.

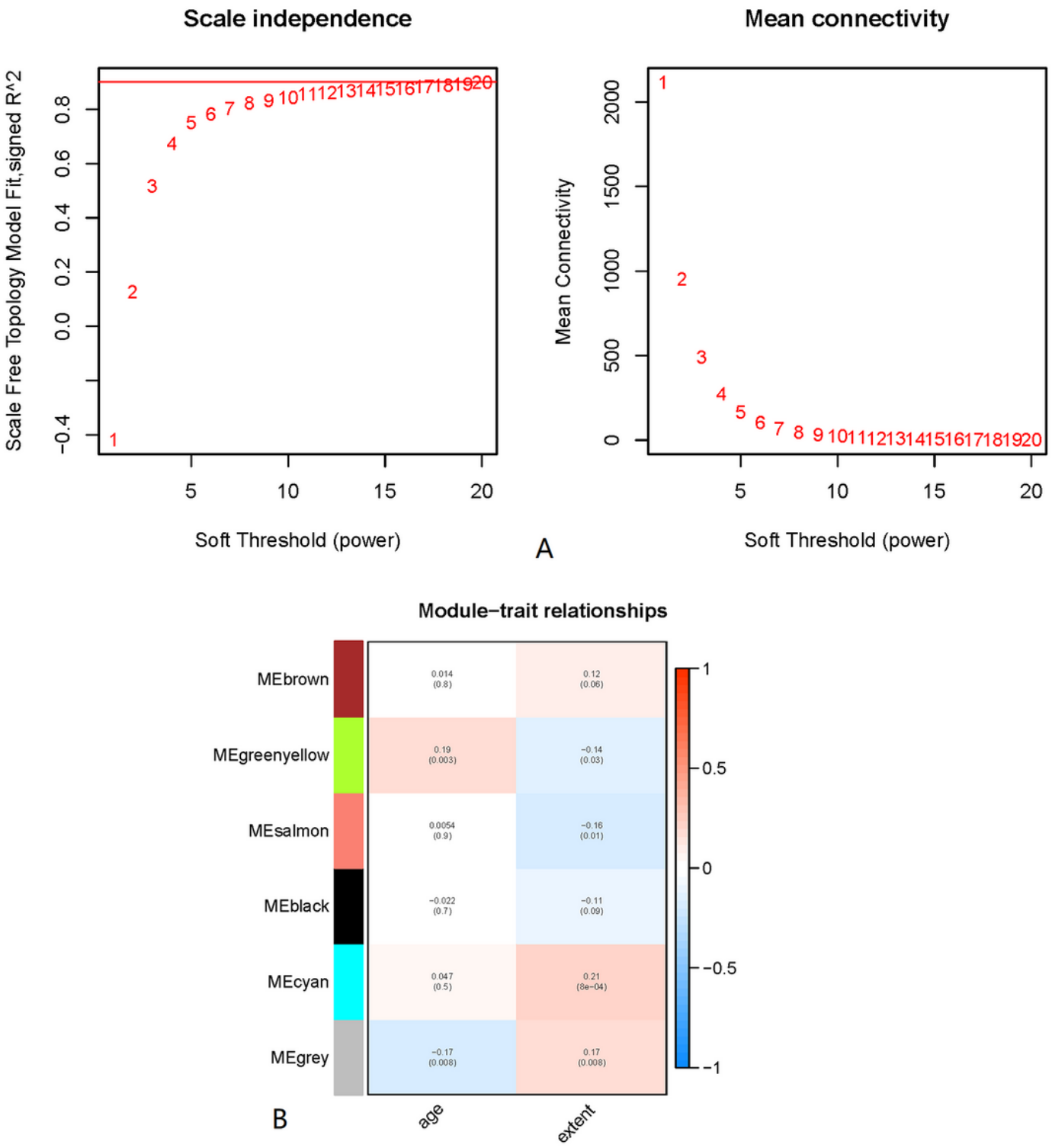

Figure 5

WGCNA analysis and modules. (A) Scale independence and mean connectivity, based on which the optimal power value can be obtained. (B) Functional characterization on WGCNA modules and clinical peculiarity. The correlation coefficients of UC's age and disease state are shown in orange and blue, respectively. The ordinate represents the module to which the gene belongs, and the abscissa represents the clinical traits. The correlation analysis between the module and clinical traits shows that if p-

value<0.05, it means that the module is correlated with clinical traits. Blue represents a negative correlation, and orange represents a positive correlation.

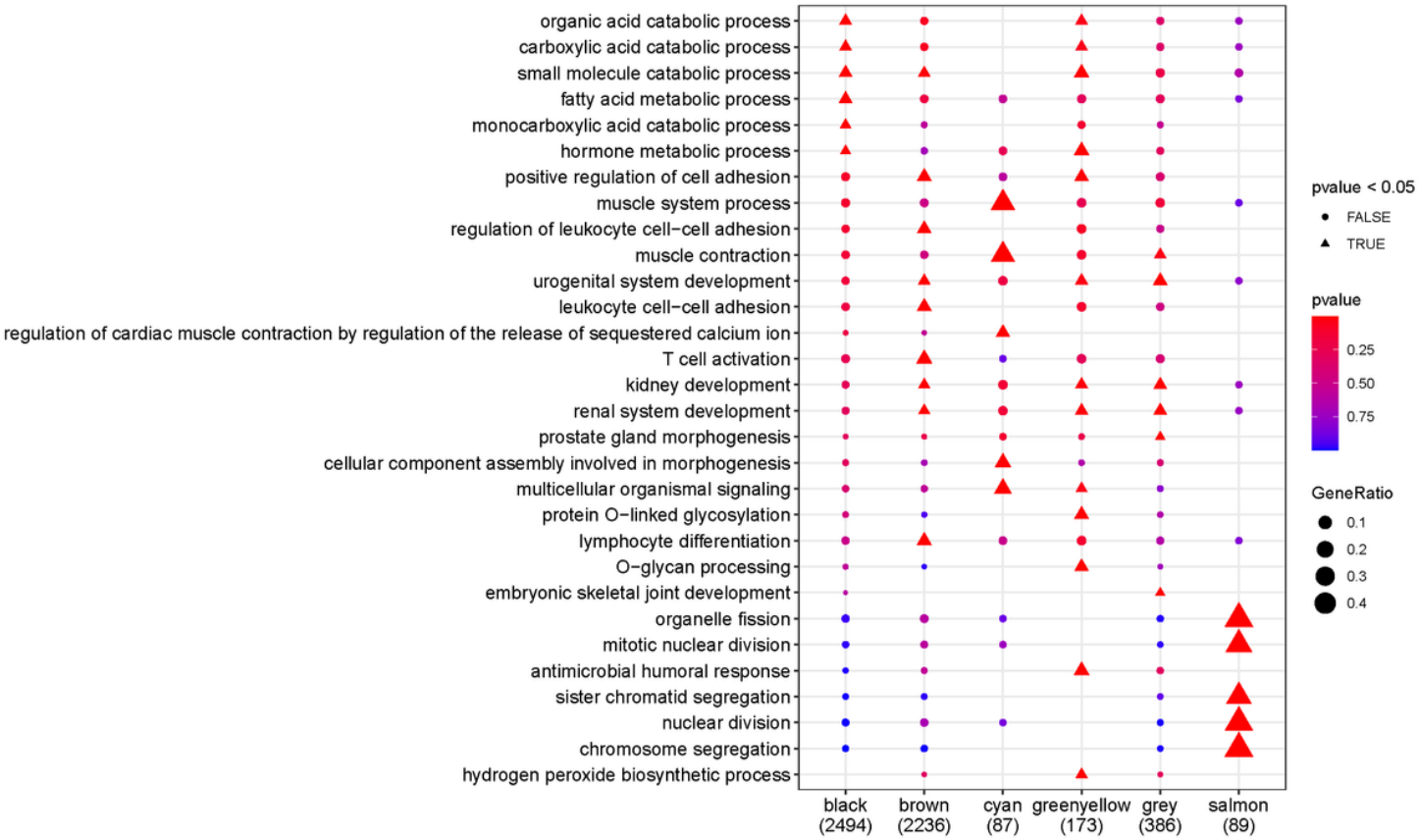

Figure 6

GO enrichment bubble graph.

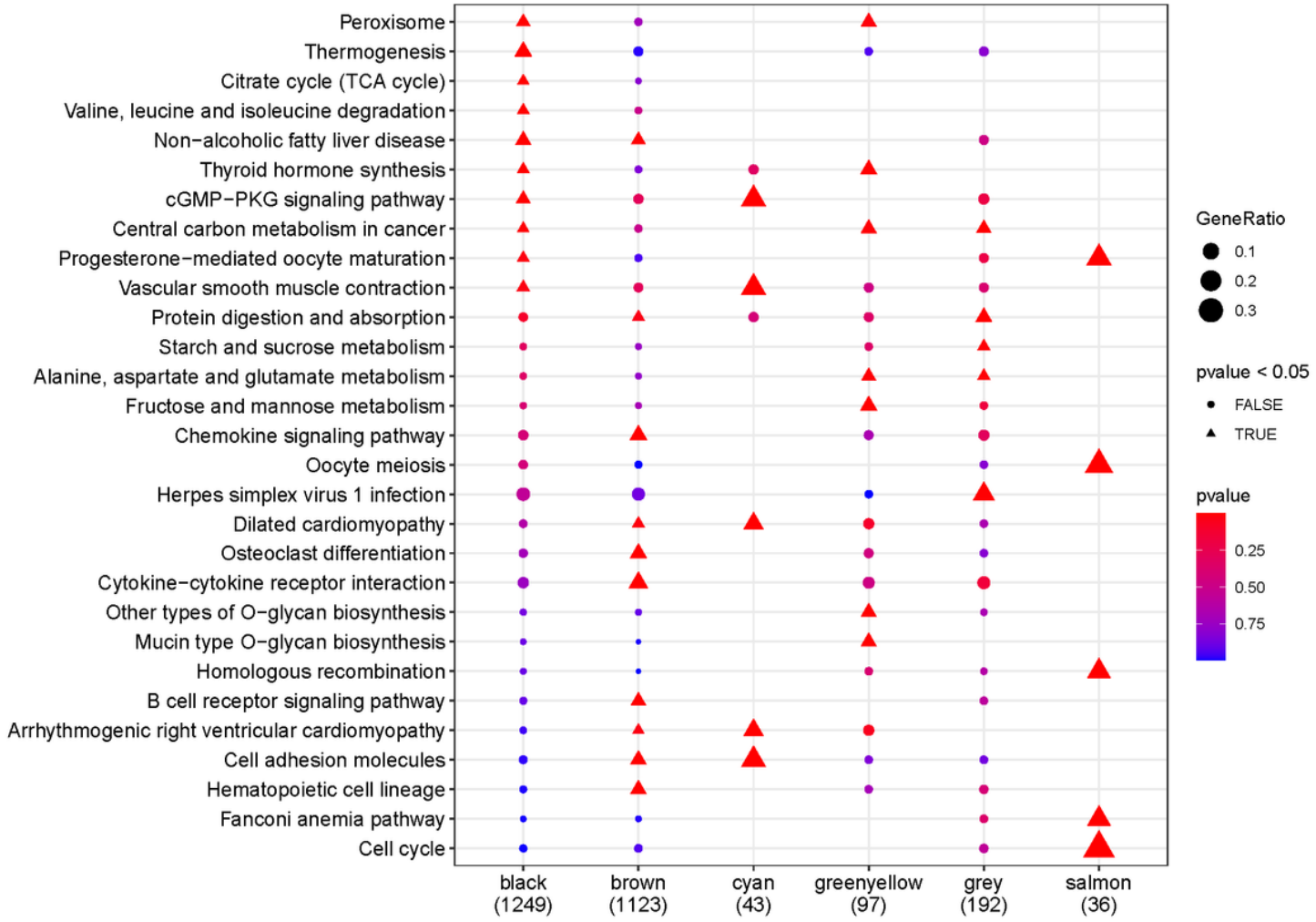

Figure 7

KEGG enrichment bubble graph. The ordinate of bubble chart represents the name of the GO, and the abscissa represents the name of a module. The color is displayed according to the p-value. The redder the color, the more significant the enrichment of GO in the module. ▲：p-value＜0.05：Significantly enriched；●：p-value＞0.05：Enrichment is not significantly.

Supplementary Files

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

- Table.csv