Investigation of Potential Genetic Biomarkers and Molecular Mechanism of Ulcerative Colitis Utilizing Bioinformatics Analysis

CURRENT STATUS: POSTED

Jiaqi Zhang  zjq405@163.com
China Academy of Chinese Medical Sciences Xiyuan Hospital
Corresponding Author
ORCiD: 0000-0002-3143-5651

Xue Wang
Experimental research center of China Academy of Chinese Medicine Sciences

Lin Xu
China Academy of Chinese Medical Sciences Xiyuan Hospital

Zedan Zhang
China Academy of Chinese Medical Sciences Xiyuan Hospital

Fengyun Wang
China Academy of Chinese Medical Sciences Xiyuan Hospital

Xudong Tang
China Academy of Chinese Medical Sciences

DOI:
10.21203/rs.2.13329/v1

SUBJECT AREAS
Epigenetics & Genomics

KEYWORDS
Ulcerative Colitis, Bioinformatics Analysis, Biomarkers, WGCNA
Abstract

Objectives: To reveal the molecular mechanisms of ulcerative colitis (UC) and provide potential biomarkers for UC gene therapy. Methods: We downloaded the GSE87473 microarray dataset from the Gene Expression Omnibus (GEO) and identified the differentially expressed genes (DEGs) between UC samples and normal samples. Then, a module partition analysis was performed based on a weighted gene co-expression network analysis (WGCNA), followed by pathway and functional enrichment analyses. Furthermore, we investigated the hub genes. At last, data validation was performed to ensure the reliability of the hub genes. Results: Between UC group and normal group, 988 DEGs were investigated. The DEGs were clustered into 5 modules using WGCNA. These DEGs were mainly enriched in functions such as the immune response, the inflammatory response and chemotaxis, and they were mainly enriched in KEGG pathways such as the cytokine-cytokine receptor interaction, chemokine signaling pathway, and complement and coagulation cascades. The hub genes, including dual oxidase maturation factor 2 (DUOXA2), serum amyloid A (SAA) 1 and SAA2, TNFAIP3-interacting protein 3 (TNIP3), C-X-C motif chemokine (CXCL1), solute carrier family 6 member 14 (SLC6A14) and complement decay-accelerating factor (CD antigen CD55), were revealed as potential tissue biomarkers for UC diagnosis or treatment. Conclusions: This study provides supportive evidence that DUOX2, A-SAA, TNIP3, CXCL1, SLC6A14 and CD55 might be used as potential biomarkers for tissue biopsy of UC, especially SLC6A14 and CD55, which may be new targets for UC gene therapy. Moreover, the DUOX2/DUOX2, NF-κB /TNIP3 and CXCL1/CXCR2 pathways might play an important role in the progression of UC through the chemokine signaling pathway and inflammatory response.

Introduction
Ulcerative colitis (UC) is a chronic non-specific inflammation of the rectum and colon whose etiology and pathogenesis are not yet well defined[1]. UC has a high incidence in Western countries and with increasing incidence in the developing countries[2]. The etiology of UC is considered to be multifactorial, including genetic and environmental factors such as urban lifestyles, dietary factors, high levels of hygiene, and gut microbiota, all of which are associated with disease progression; however, the pathogenesis of UC remains unclear[3]. Bioinformatics can be effectively used to analyze UC microarray data, providing theoretical reference for further exploration of the mechanisms of inflammatory bowel disease, and help to find potential target genes. As the latest bioinformatics research method, WGCNA is commonly used to reveal differences between genes in different samples[4].

In this study, UC gene expression data uploaded by K Li et al. were downloaded. We identified the DEGs between UC samples and normal samples. Then, a module partition analysis was performed based on a WGCNA, followed by pathway and functional enrichment analyses. Then, data validation was performed to ensure the reliability of the hub genes. This study forecasts the molecular mechanism of UC and the potential biomarkers for UC therapy.

Materials And Methods

2.1. Microarray data. The gene expression profile of GSE87473 was obtained from the GEO database[5] (http://www.ncbi.nlm.nih.gov/geo/). A total of 127 mucosal biopsy samples were obtained from 106 UC patients and 21 control subjects for subsequent analysis.

2.2. Data Preprocessing and DEG Analysis. There were a total of 20741 probes in the present dataset. GEO2R(http://www.ncbi.nlm.nih.gov/geo/geo2r/) is based on R that comes with the GEO databases, which was used to identify DEGs between UC and control samples. |log-fold change (LFC)| >1 and P-values<0.05 were selected as the thresholds for
2.3. **WGCNA Analysis.** The co expression network analysis was performed using WGCNA (version: 1.63)[6]. WGCNA is a systematic biological method for constructing scale-free networks using gene expression data. First, we selected the soft threshold for network construction. The soft threshold was used to transform the similarity matrix of gene expression into adjacency matrix, which enhances strong correlation and weakens correlation at the exponential level. Second, the adjacency matrix was transformed into a topological matrix. Based on TOM, we used average-linkage hierarchical clustering method to cluster genes. According to the standard of hybrid dynamic cut tree, we set the minimum number of base 30 for each gene network module. After determining the gene module by dynamic shearing method, we calculated the eigenvectors of each module in turn, then clustered the modules, merged the nearer modules into new modules, and set height = 0.25[7]. Third, we calculated the module eigengene (ME) of each module, which represents the expression level for each module. We also calculated the correlation between the clinical traits and ME in each module. At last, we calculated the gene significance (GS) of each gene in the module, which represented the correlation between the genes and sample.

2.4. **Function and Pathway Enrichment Analysis.** We used the DAVID 6.8 (https://david.ncifcrf.gov) software for the GO-biological function (GO-BP) and KEGG pathway analyses of the genes in main modules. We selected the P-false discovery rate (FDR) of <0.05 as the threshold for the identification of significant GO-BP terms and KEGG pathways.

2.5. **Hub Genes Investigation.** According to the feature vector of each module, the correlation of the gene expression in the module was analyzed by WGCNA. Genes with correlations greater than 0.9 in each module were considered hub genes.
2.6. *Data Validation*. To verify the robustness of hub genes, the microarray data of GSE75214[8] ([HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene)
version]), which included 108 tissue samples (97UC samples and 11control samples), was
downloaded from the GEO database. GraphPad Prism 7.00 software was used to calculate the
area under the curve (AUC).

**Results**

3.1. *DEGs between UC Samples and Normal Samples*. We identified 988 DEGs,
including 466 up-regulated DEGs and 522 down-regulated DEGs with \( p_{\text{FDR}} < 0.05 \) and |LFC| > 1. The heatmap and volcano plot are shown in Figures 1(A) and 1(B). Obviously, the
heatmap showed that these DEGs could be used to distinguish UC from control samples.

3.2. *WGCNA Analysis*. We performed WGCNA analysis using the 988 DEGs. The co-
expression network is a scale-free network, which means the logarithm \( \log(k) \) of a node
with a connection degree of \( k \) is negatively correlated with the logarithm \( \log(P(k)) \) of the
probability of occurrence of the node, and the correlation coefficient is greater than 0.8. R
software package WGCNA was used to build a weighted co-expression network. To ensure
that the network was a scale-free network, we chose a soft threshold of \( \beta = 6 \) (Figure 1C).
The DEGs was clustered into 5 modules, described here as including turquoise (510 DEGs),
blue (393 DEGs), brown (48DEGs), yellow (30DEGs) and gray (7 DEGs) (Figure 1(D)). The
turquoise and blue modules were down-regulated, while the brown and yellow modules
were up-regulated. Moreover, the turquoise module (correlation index: -0.68, \( P = 3.0E-18 \))
was negatively correlated with the disease presence, the yellow (correlation index: 0.51, \( P
\) = 6.0E-10), blue (correlation index: 0.62, \( P = 1.0E-14 \)) and brown modules (correlation
index: 0.73, \( P = 3E-22 \)) were positively correlated with the disease presence, while the
turquoise module (correlation index: -0.65, \( P = 1.0E-16 \)) was negatively correlated with
the disease extent, the yellow (correlation index: 0.42, \(P = 1.0\times10^{-16}\)) , blue (correlation index: 0.43, \(P = 4\times10^{-7}\)) and brown modules (correlation index: 0.52, \(P = 6\times10^{-10}\)) were positively correlated with the disease extent (Figure 3A). The average gene significance (GS) for each module indicated that the brown module was the most related to the disease presence, and the turquoise module was most related to the disease extent (limited or extensive) (Figure 3B).

3.3. Functional and Pathway Enrichment for DEGs. The top 3 GO-BP and KEGG terms enriched by DEGs are shown in Table 1 and Figure 4. The DEGs in the brown module were mainly involved in functions such as inflammatory response (\(P = 4.88\times10^{-7}\)) and pathways such as the chemokine signaling pathway (\(P = 0.004195\)). The DEGs in the turquoise module were mainly involved in functions such as the oxidation-reduction process (\(P = 9.70\times10^{-3}\)) and pathways such as metabolic pathways (\(P = 2.8\times10^{-9}\)).

3.4. Hub Genes. The brown module was most relevant to the disease; therefore, we analyzed the correlation of gene expression in the brown module in the following study. Figure 5 shows that dual oxidase maturation factor 2 (DUOXA2), serum amyloid A (SAA) 1 and SAA2, TNFAIP3-interacting protein 3 (TNIP3), C-X-C motif chemokine (CXCL1), Solute carrier family 6 member 14 (SLC6A14) and Complement decay-accelerating factor (CD antigen CD55) were selected as hub genes.

3.5. Data Validation. To verify the robustness of the hub genes, the validation data GSE75214 was obtained from the GEO database. We performed ROC curve analysis using GraphPad Prism 7.00. The results of the analysis showed that the hub genes related to UC, including DUOXA2, SAA1, SAA2, TNIP3, CXCL1, SLC6A14 and CD55, were identified as potential tissue biopsy molecules for UC diagnosis (Table 2 and Figure 6).

Discussion

UC is a common chronic colorectal inflammation that is caused by the dysregulation of the
immune response of the genetically susceptible host to an intraluminal antigen[9,10] with recurrent mucus pus and bloody stools, abdominal pain and diarrhea as the main clinical manifestations. With the emergence of UC research, many potential biomarkers for early diagnosis or treatment of UC have been identified after the development of biology technology. However, the mechanism of UC is still unknown.

UC gene expression data were analyzed by WGCNA in this study. We screened a total of 988 DEGs between UC samples and control samples, and identified 5 modules. The brown module was related to the occurrence of UC, and the turquoise module was related to the development of UC. Furthermore, the DEGs in the brown module were mainly involved in functions such as inflammatory responses and pathways such as chemokine signaling pathways. The DEGs in the turquoise module were mainly involved in functions such as oxidation-reduction processes and pathways such as various metabolic pathways. The hub genes DUOXA2, SAA1, SAA2, TNIP3, CXCL1, SLC6A14 and CD55 were identified as potential tissue biomarkers for UC diagnosis.

DUOXA2 is the maturation partner of DUOX2, the results demonstrated that both DUOX2 and DUOXA2 were up-regulated in association with inflammation in UC, as well as in colorectal cancer, UC-associated adenomas and colorectal dysplasia[11]. DUOX2/DUOXA2 provided the NADPH oxidase family that directly generates \( \text{H}_2\text{O}_2 \), which is involved in inflammatory reactions and can cause DNA single-strand and double-strand breaks[12,13]. These results suggest that DUOX2/DUOXA2 may be involved in the development and carcinogenesis of UC. SAA is a sensitive acute phase protein and a precursor of amyloid A (AA). Because it has higher sensitivity and specificity than C-reactive protein (CRP), SAA can better reflect inflammation, making SAA a protein of current great interest in medical and health fields. Furthermore, as a pro-inflammatory cytokine, SAA plays an important role in the occurrence and development of autoimmune diseases, such as ankylosing
spondylitis (AS), rheumatoid arthritis (RA), and inflammatory bowel disease (IBD) [14–16]. Acute SAA (A-SAA), including SAA1 and SAA2, are produced by hepatocytes and then enter the blood. The SAA1 gene is expressed in colon carcinomas, and it is a chemoattractant that promotes the migration, adhesion and tissue infiltration of monocytes and polymorphonuclear leukocytes [17]. Microarray assessments showed significant over expression of SAA1 in UC specimens. SAA1 plays a role in local inflammation in the microenvironment of malignant tissue and is expressed in colon carcinomas [18]. Therefore, SAA may be a new early predictive clinical marker and a target for designing novel selective inhibitors for therapeutic intervention of UC and UC-associated CRC.

Nuclear factor (NF)-κB plays an important role in the immune system, and inappropriate NF-κB activity is associated with many inflammatory diseases and autoimmune diseases. TNFAIP3 is a key participant in the negative feedback regulation of NF-κB signal transduction in response to multiple stimuli [19]. Recent genetic studies have demonstrated that several mutations in human TNFAIP3 locus are associated with UC, Crohn’s disease, RA, type 1 diabetes mellitus and systemic lupus erythematosus [20–24]. Lee SH showed that TNIP3 over expression down regulated the activation of signal transducer and activator of transcription 3 (STAT3) and NF-κB. At the same time, TNIP3 also decreased the expression of IL-17 in CD4+ T cells in spleen slices. On the contral, TNIP3 over expression increased the expression of FOXP3 in CD4+ T cells [25]. These results further demonstrate the importance of TNIP3 in reducing inflammation and preventing UC.

CXCL1 was originally found in melanoma and was also expressed in macrophages, neutrophils and epithelial cells. CXCL1 acts by specifically binding to its receptor, C-X-C chemokine receptor type 2 (CXCR2) [26]. Recent studies have shown that the CXCL1/CXCR2 signaling pathway and chemotactic central granulocytes regulate the
inflammatory response; moreover, the pathway causes tumor cell proliferation, angiogenesis, and lymph angiogenesis and promotes tumor invasion and vascular metastasis, which plays an important role in the occurrence and development of tumors, etc.[27]. Previous studies have shown that CXCL1 was concomitantly expressed in ulcerative colitis and decreased during treatment with topical corticosteroids[28]. Recent research has shown that CXCL1 is critical for metastasis in colorectal cancer and pre-metastatic niche formation. The specific mechanism is that primary malignant cell secreted Vascular endothelial growth factor A(VEGF-A) stimulates tumor-associated macrophages to produce CXCL1, which recruits CXCR2-positive myeloid-derived suppressor cells to form a pre-metastatic niche to promote liver metastases[29].

SLC6A14, expressed on the apical membrane of epithelial cells, is a Na\(^+\)/Cl\(^-\)-dependent neutral and cationic amino acid transporter[30]. Due to its broad specificity and concentrative transport mechanisms, SLC6A14 is hypothesized to be involved in nutrient uptake[31]. Moreover, changes of SLC6A1 expression are related to multiple diseases, such as obesity, cystic fibrosis, colitis, male infertility and various epithelial cancers[32–35]. Multiple sequencing or microarray studies have shown that SLC6A14 was upregulated in UC patients[5,8,36,58]. This indicates that the up-regulation of SLC6A14 is a common phenomenon in intestinal inflammation. SLC6A14 may be involved in colonic inflammation by regulating glutamine(a substrate for SLC6A14) and nitric oxide synthase 2 (coordinated upregulation with SLC6A14 in inflamed cells)[37,–38]. A recent study showed that abnormal permeability of the inner mucus was observed in some UC patients, with a significant reduction in SLC26A3, which provides the bicarbonate necessary for colonic mucosa formation, indicted that decreased mucus barrier in colon was an early event in ulcerative colitis pathogenesis. Whether SLC6A14 and SLC26A3 have synergistic effects in the pathogenesis of UC? [39] However, to date, the relationship between increased
expression of SLC6A14 and inflammation remains unknown.

CD55, known as decay-accelerating factor, is a member of the regulators of complement activation (RCA) family [40]. It is a major regulator of complement activation alternative and classical pathways and is expressed in all serum-exposed cells [41]. CD55 physiologically acts as an inhibitor of the complement system but it is also broadly expressed in malignant tumors [42,43]. Previous studies have shown increased expression of CD55 in stools and colonic mucosa of disease activity in patients with UC [44,45]. The results of functional and pathway DEG enrichment in this study show that the biological functions involved in the pathogenesis of UC include the inflammatory response, innate immune response and chemotaxis, indicating that the pathogenesis of UC was multifactorial, involving epithelial barrier defects, genetic predisposition, environmental factors, and dysregulated immune responses. The six hub genes screened in this study are not only related to mucosal inflammation but they also accelerate the progression of colon cancer, so they should be given proper attention in the treatment of UC. Emerging studies have demonstrated that aberrant gut microbiota are involved in the development and progression of ulcerative colitis [46,47]; therefore, aberrant gut microbiota may also participate in the pathogenesis of UC through the above pathways. Do abnormal gut microbiota affect the expression of host genes, or vice versa? The specific mechanism still needs to be explored.

Although we found six hub genes closely related to UC and confirmed the robustness of their diagnostic value, which may be useful for us to improve our understanding of the molecular mechanism of UC and as a potential prognostic and diagnostic biomarker, we still need further molecular biological experiments to confirm our hypothesis.

Conclusions

In conclusion, DUOXA2, A-SAA, TNIP3, CXCL1, SLC6A14 and CD55 might be used as potential
biomarkers for UC tissue biopsy, especially SLC6A14 and CD55, which may be new targets for UC gene therapy. Furthermore, the DUOX2/DUOXA2, NF-κB/TNIP3 and CXCL1/CXCR2 pathways may play important roles in UC progression via the inflammatory response.

Abbreviations

AS: Ankylosing Spondylitis; AUC: Area Under the Curve; CD: Complement Decay-accelerating Factor; CXCL1: C-X-C motif chemokine; CXCR2: C-X-C chemokine receptor type 2; DEG: Differentially Expressed Gene; DUOXA2: Dual Oxidase Maturation Factor 2; FDR: False Discovery Rate; GEO: Gene Expression Omnibus; GO-BP: GO-biological function; GS: Gene Significance; IBD: Inflammatory Bowel Disease; LFC: Log-Fold Change; ME: Module Eigengene; NF: Nuclear factor; RA: Rheumatoid Arthritis; SAA: Serum Amyloid A; SLC6A14: Solute Carrier family 6 member 14; STAT3: Signal Transducer and Activator of Transcription 3; TNIP3: TNFAIP3-interacting protein 3; UC: Ulcerative Colitis; WGCNA: Weighted Gene Co-expression Network Analysis;

Declarations

Acknowledgements

Not applicable.

Funding

This study is supported by the National Natural Science Foundation of China (grants 81830118 and 81804078).

Availability of data and materials

The datasets analyzed during the current study are available in the GeneExpression Omnibus with the accession GSE87473 and GSE75214.

Authors’ contributions

(i) Guarantor of the article: Xudong-Tang, (ii) Study concept and design: Fengyun-Wang;

Acquisition of data: Jiaqi-Zhang and Xue-Wang; Analysis and interpretation of data: Jiaqi-
Zhang and Xue-Wang and Li-Xu; Drafting of the manuscript: Jiaqi-Zhang, Xue Wang and Zedan-Zhang; (iii). 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.

1Department of Gastroenterology, Xiyuan Hospital of China Academy of Chinese Medical Sciences, Beijing 100091, China; 2Experimental Research Center of China Academy of Chinese Medical Sciences, Beijing 100700, China; 3China Academy of Chinese Medical Sciences, Beijing 100700, China.

*both authors contributed equally to this work.

Supported by: The work was supported by the National Nature Science Foundation of China (Grant No. 81830118) and the National Natural Science Youth Fund Project (Grant No. 81804078).

Correspondence to: Xu-Dong Tang, Doctor, China Academy of Chinese Medical Sciences, No. 16, Inner South Street, Dongzhimen, Dongcheng District, Beijing 100700, China. E-mail: txdly@sina.com.

References

1. Haberman Y, Karns R, Dexheimer PJ, et al. Ulcerative colitis mucosal transcriptomes reveal mitochondriopathy and personalized mechanisms underlying disease severity and treatment response. Nat Commun. 2019;10:38.
2. Ramos GP, Papadakis KA. Mechanisms of Disease: Inflammatory Bowel Diseases. Mayo Clin Proc. 2019;94:155-165.

3. Armstrong H, Alipour M, Valcheva R, et al. Host immunoglobulin G selectively identifies pathobionts in pediatric inflammatory bowel diseases. Microbiome. 2019;7:1.

4. P. Langfelder and S. Horvath, “WGCNA: an R package for weighted correlation network analysis,” BMC Bioinformatics, 2008,9:599.

5. Barrett T, Troup DB, Wilhite SE, et al. NCBI GEO: mining tens of millions of expression profiles—database and tools update. Nucleic Acids Res. 2007;35(Database issue):D760–765.

6. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008.

7. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics. 2008;24:719–720.

8. Vancamelbeke M, Vanuytsel T, Farré R, et al. Genetic and Transcriptomic Bases of Intestinal Epithelial Barrier Dysfunction in Inflammatory Bowel Disease. Inflamm Bowel Dis. 2017;23:1718–1729.

9. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. Lancet. 2017;389:1756–1770.

10. Rubin D T, Ananthakrishnan A N, Siegel C A, et al. ACG Clinical Guideline: Ulcerative Colitis in Adults[]. American Journal of Gastroenterology, 2019;114:384-413.

11. MacFie TS, Poulson R, Parker A, et al. DUOX2 and DUOXA2 form the predominant enzyme system capable of producing the reactive oxygen species H2O2 in active ulcerative colitis and are modulated by 5-aminosalicylic acid. Inflamm Bowel Dis. 2014;20:514-524.

12. Muzza M, Fugazzola L. Disorders of H2O2 generation[]. Best Practice & Research Clinical Endocrinology & Metabolism, 2017;31:225.

13. Sarr D, Tóth E, Gingerich A, et al. Antimicrobial actions of dual oxidases and lactoperoxidase[].
Malle E, De Beer F C. Human serum amyloid A (SAA) protein: a prominent acute-phase reactant for clinical practice[J]. European Journal of Clinical Investigation, 2015; 26:427–435.

O’Reilly S, Cant R, Ciechomska M, et al. Serum amyloid A induces interleukin-6 in dermal fibroblasts via Toll-like receptor 2, interleukin-1 receptor-associated kinase 4 and nuclear factor-κB[J]. Immunology, 2015; 143:331–340.

Ye R D, Sun L. Emerging functions of serum amyloid A in inflammation[J]. Journal of Leukocyte Biology, 2015; 98:923.

Gutfeld O, Prus D, Ackerman Z, et al. Expression of serum amyloid A, in normal, dysplastic, and neoplastic human colonic mucosa: implication for a role in colonic tumorigenesis.[J]. Journal of Histochemistry & Cytochemistry, 2006;54:63–73.

Schaffer M W, Smith J C, Roy T G, et al. Abstract LB–428: Upregulation of serum amyloid A1 (SAA1) gene is associated with ulcerative colitis-related colon cancer[J]. Cancer Research, 2011, 71(8 Supplement):LB–428-LB–428.

Coornaert B, Carpentier I, Beyaert R. A20: Central Gatekeeper in Inflammation and Immunity[J]. Journal of Biological Chemistry, 2009;284:8217.

Zheng C, Huang Y, Ye Z, et al. Infantile Onset Intractable Inflammatory Bowel Disease Due to Novel Heterozygous Mutations in TNFAIP3 (A20)[J]. Inflammatory Bowel Diseases, 2018;24:2613.

Billmann-Born S, Till A, Arlt A, et al. Genome-Wide Expression Profiling Identifies an Impairment of Negative Feedback Signals in the Crohn’s Disease-Associated NOD2 Variant L1007fsinsC[J]. The Journal of Immunology, 2011; 186:4027–4038.

Matmati M, Jacques P, Maelfait J, et al. A20 (TNFAIP3) deficiency in myeloid cells triggers erosive polyarthritis resembling rheumatoid arthritis.[J]. Nature Genetics, 2011; 43:908–912.

Xiao L, Qin H, Wu J, et al. Association of TNFAIP3 and TNIP1 polymorphisms with systemic lupus erythematosus risk: A meta-analysis[J]. Gene, 2018:S0378111918305523.
24. Fukaya M, Brorsson C A, Meyerovich K, et al. A20 Inhibits β-Cell Apoptosis by Multiple Mechanisms and Predicts Residual β-Cell Function in Type 1 Diabetes[J]. Molecular Endocrinology, 2016; 30:48.

25. Lee S H, Lee H R, Kwon J Y, et al. A20 ameliorates inflammatory bowel disease in mice via inhibiting NF-κB and STAT3 activation.[J]. Immunology Letters, 2018;28:198.

26. Mantovani A, Savino B, Locati M, et al. The chemokine system in cancer biology and therapy[J]. Cytokine Growth Factor Rev, 2010; 21:27–39.

27. Acharyya S, Oskarsson T, Vanharanta S, et al. A CXCL1 paracrine network links cancer chemoresistance and metastasis[J]. Cell, 2012;150:165–178.

28. Egesten A, Eliasson M, Olin Al, et al. The proinflammatory CXC-chemokines GRO-alpha/CXCL1 and MIG/CXCL9 are concomitantly expressed in ulcerative colitis and decrease during treatment with topical corticosteroids. Int J Colorectal Dis. 2007;22:1421–1427.

29. Wang D, Sun H, Wei J, et al. CXCL1 Is Critical for Premetastatic Niche Formation and Metastasis in Colorectal Cancer[J]. Cancer Research, 2017;77: 3655–3665.

30. Karunakaran S, Ramachandran S, Coothankandaswamy V, et al. SLC6A14 (ATB0,+) protein, a highly concentrative and broad specific amino acid transporter, is a novel and effective drug target for treatment of estrogen receptor-positive breast cancer.[J]. Journal of Biological Chemistry, 2011;286:31830–31838.

31. Rudnick G, Krämer R, Blakely R D, et al. The SLC6 Transporters: perspective on structure, functions, regulation and models for transporter dysfunction (2014)[J]. Pflügers Archiv - European Journal of Physiology, 2014; 466:25–42.

32. Sikder MOF, Yang S, Ganapathy V, Bhutia YD. The Na+/Cl—Coupled, Broad-Specific, Amino Acid Transporter SLC6A14 (ATB0,+) : Emerging Roles in Multiple Diseases and Therapeutic Potential for Treatment and Diagnosis. AAPS J. 2017;20:12.

33. Babu E, Bhutia Y D, Ramachandran S, et al. Deletion of the amino acid transporter Slc6a14
suppresses tumor growth in spontaneous mouse models of breast cancer.[J]. Biochemical Journal, 2015;469:17-23.

34. Coothankandaswamy V, Cao S, Xu Y, et al. Amino acid transporter SLC6A14 is a novel and effective drug target for pancreatic cancer[J]. Br J Pharmacol, 2016;173:3292-3306.

35. Sewda K, Enkemann S, Gillies R J, et al. Abstract 3178: Identification and validation of markers for colon adenoma and adenocarcinoma[J]. Cancer Research, 2011; 71(8 Supplement):3178-3178.

36. Flach C F, Eriksson A, Jennische E, et al. Detection of elafin as a candidate marker for ulcerative colitis by whole-genome microarray screening[J]. Inflammatory Bowel Diseases, 2006;12:837-842.

37. Novak E A, Mollen K P. Mitochondrial dysfunction in inflammatory bowel disease[J]. Frontiers in Cell & Developmental Biology, 2015;3:62.

38. Kolios G, Valatas V, Ward S G. Nitric oxide in inflammatory bowel disease: a universal messenger in an unsolved puzzle[J]. Insect Science, 2010;113:427-437.

39. van der Post S, Jabbar KS, Birchenough G, et al. Structural weakening of the colonic mucus barrier is an early event in ulcerative colitis pathogenesis. Gut. 2019;0,1-10

40. Hamann J, Vogel B, van Schijndel G M, et al. The seven-span transmembrane receptor CD97 has a cellular ligand (CD55, DAF)[J]. Journal of Experimental Medicine, 1996;184:1185-1189.

41. He Z, Wu H, Jiao Y, et al. Expression and prognostic value of CD97 and its ligand CD55 in pancreatic cancer.[J]. Oncology Letters, 2015;9:793-797.

42. Saygin C, Wiechert A, Rao V S, et al. CD55 regulates self-renewal and cisplatin resistance in endometrioid tumors[J]. Journal of Experimental Medicine, 2017;214:2715.

43. Cimmino F, Avitabile M, Pezone L, et al. CD55 is a HIF-2α marker with anti-adhesive and pro-invasing properties in neuroblastoma:[J]. Oncogenesis, 2016;5:e212.

44. Inaba T, Mizuno M, Ohya S, et al. Decay-accelerating factor (DAF) in stool specimens as a marker of disease activity in patients with ulcerative colitis (UC).[J]. Clinical & Experimental Immunology, 2010;112:237-241.
45. Andoh A, Fujiyama Y, Kimura T, et al. Increased expression of decay-accelerating factor (CD55) in the inflamed mucosa of patients with ulcerative colitis[J]. Pathophysiology, 1998;5:105-110.

46. Axelrad JE, Joelson A, Green PHR, et al. Enteric Infections Are Common in Patients with Flares of Inflammatory Bowel Disease. Am J Gastroenterol. 2018;113:1530–1539.

47. Forbes JD, Chen CY, Knox NC, et al. A comparative study of the gut microbiota in immune-mediated inflammatory diseases-does a common dysbiose exist?. Microbiome. 2018;6:221

Tables

Table 1: The results for GO-BP function and KEGG pathway enrichment analysis (top 3 in Brown and Turquoise module are listed).

| Module  | GO-BP terms            | P_value  | KEGG terms                              | P_val  |
|---------|------------------------|----------|-----------------------------------------|--------|
| Brown   | Inflammatory response  | 4.88E-07 | Chemokine signaling pathway             | 0.004  |
|         | Innate immune response |          | Cytokine-cytokine receptor interaction  | 0.010  |
|         | Chemotaxis             | 1.39E-06 | Complement and coagulation cascades     | 0.001  |
|         | Oxidation-reduction process | 9.7E-3 | Metabolic pathways                      | 2.8E-1  |
|         | Transport              | 1.4E-5   | Drug metabolism - cytochrome P450        | 1.3E-1  |
|         | Metabolic process      | 1.2E-10  | Chemical carcinogenesis                  | 2.1E-1  |

Table 2: Results of AUCs for hub genes

| Hub genes | UC vs. Normal | AUC | P-value | 95%CI       |
|-----------|---------------|-----|---------|-------------|
| Disease-related |               |     |         |             |
| DUOXA2     |               | 0.8894 | 0.0001 | 0.826 to 0.9528 |
| SAA1 /// SAA2 |            | 0.8097 | 0.0008 | 0.6975 to 0.9220 |
| TNIP3      |               | 0.8969 | 0.0002 | 0.8366 to 0.9572 |
| CXCL1      |               | 0.8857 | 0.0001 | 0.8151 to 0.9562 |
| SLC6A14    |               | 0.9822 | 0.0001 | 0.9606 to 1.004 |
| CD55       |               | 0.9297 | 0.0001 | 0.8814 to 0.978 |

Figures
The heat map, volcano plot, and weighted gene coexpression network analysis (WGCNA) of differentially expressed genes (DEGs) between the UC group and the control group. (A) The heat map for DEGs. (B) The volcano plot for DEGs. Gray dots represent the genes that are not differentially expressed, red dots represent the upregulated genes, and the blue dots represent the downregulated genes. (C) Determination of the soft threshold in the WGCNA algorithm. The approximate scale-free fit index can be attained at the soft-thresholding power of 6. (D) Clustering dendrograms showing 4 modules that contain a group of highly connected genes. Each designated color represents a certain gene module.
Figure 2

The module expression pattern. The heatmap represents the expression of genes where each row represents a gene and each column represents a sample. The red color in the heatmap represents upregulated genes, while the green color represents downregulated genes. The bar charts represent the eigengene profiles of four WGCNA modules; the color of the bar chart represents the color of the related module.
Relationships of module eigengenes and the samples. (A) The module-trait relationships. The number in the first row of the square on the right is the correlation coefficient to the UC group shown at the top of each row with the p values printed below the correlations in parentheses, and the number on the left is the correlation coefficient to the disease extent (limited or extensive) of the UC group. The rows are colored based on the correlation of the module to the UC group: red for a positive correlation and blue for a negative correlation. (B) The average gene significance (GS) of all genes (i.e., module significance, MS) of each module. Modules with greater MS values were considered to have more connection with the disease.
Figure 4

The results for GO-BP function and KEGG pathway enrichment analysis (the top20 in the Brown and Turquoise colored modules are listed). (A) The GO-BP function enrichment of DEGs in the brown module. (B) KEGG pathway enrichment of the DEGs in the brown module. (C) The GO-BP function enrichment of the DEGs in the turquoise module. (D) KEGG pathway enrichment of the DEGs in the turquoise module.
Figure 5

Hub genes in the brown module. A total of six genes were selected as hub genes, and the correlation coefficients (ranging from 0.90-0.93) and p-values are shown in the lower right corner of each image.
Predicted ROC curves of the UC hub genes. The prediction of UC vs. control was robust, and the area under the curve (AUC) of 6 hub genes for UC vs. control ranged from 0.8097 to 0.9822.