Identification of differentially expressed genes in ulcerative colitis and verification in a colitis mouse model by bioinformatics analyses

Lei Shi, Xiao Han, Jun-Xiang Li, Yu-Ting Liao, Fu-Shun Kou, Zhi-Bin Wang, Rui Shi, Xing-Jie Zhao, Zhong-Mei Sun, Yu Hao

ORCID number: Lei Shi 0000-0002-7925-5166; Xiao Han 0000-0003-1340-9692; Jun-Xiang Li 0000-0001-7590-9444; Yu-Ting Liao 0000-0003-1821-048X; Fu-Shun Kou 0000-0003-1841-3324; Zhi-Bin Wang 0000-0001-9274-635X; Rui Shi 0000-0002-9374-0897; Xing-Jie Zhao 0000-0002-7918-9576; Zhong-Mei Sun 0000-0001-8667-7306; Yu Hao 0000-0002-8544-5580.

Author contributions: Shi L wrote the article and Han X edited it; Liao YT, Kou FS and Sun ZM performed the animal experiments; Wang ZB, Shi R and Zhao XJ analyzed the bioinformatics data; Hao Y and Li JX conducted the study.

Supported by Chinese Medicine Inheritance and Innovation “One Hundred Million” Talent Project Qihuang Scholar (to Li JX); The National Key R&D Program of China during the 13th Five-Year Plan Period, No. 2018YFC1705405; and The 66th China Postdoctoral Science Foundation, No. 2019M660575.

Institutional review board statement: The data of ulcerative colitis we analyzed in this study were all from the National Center for Biotechnology Information-Gene Expression Omnibus (NCBI-GEO) databases.

Abstract

BACKGROUND
Ulcerative colitis (UC) is an inflammatory bowel disease that is difficult to diagnose and treat. To date, the degree of inflammation in patients with UC has mainly been determined by measuring the levels of nonspecific indicators, such as C-reactive protein and the erythrocyte sedimentation rate, but these indicators have an unsatisfactory specificity. In this study, we performed bioinformatics analysis using data from the National Center for Biotechnology Information-Gene Expression Omnibus (NCBI-GEO) databases and verified the selected core genes in a mouse model of dextran sulfate sodium (DSS)-induced colitis.

AIM
To identify UC-related differentially expressed genes (DEGs) using a bioinformatics analysis and verify them in vivo and to identify novel biomarkers and the underlying mechanisms of UC.

METHODS
Two microarray datasets from the NCBI-GEO database were used, and DEGs between patients with UC and healthy controls were analyzed using GEO2R and Venn diagrams. We annotated these genes based on their functions and signaling pathways, and then protein-protein interactions (PPIs) were identified using the Search Tool for the Retrieval of Interacting Genes. The data were further analyzed with Cytoscape software and the Molecular Complex Detection (MCODE) app.
The core genes were selected and a Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis was performed. Finally, colitis model mice were established by administering DSS, and the top three core genes were verified in colitis mice using real-time polymerase chain reaction (PCR).

RESULTS

One hundred and seventy-seven DEGs, 118 upregulated and 59 downregulated, were initially identified from the GEO2R analysis and predominantly participated in inflammation-related pathways. Seven clusters with close interactions in UC formed: Seventeen core genes were upregulated [C-X-C motif chemokine ligand 13 (CXCL13), C-X-C motif chemokine receptor 2 (CXCR2), CXCL9, CXCL5, C-C motif chemokine ligand 18, interleukin 1 beta, matrix metallopeptidase 9, CXCL3, formyl peptide receptor 1, complement component 3, CXCL8, CXCL1, CXCL10, CXCL2, CXCL6, CXCL11 and hydroxycarboxylic acid receptor 3] and one was downregulated [neuropeptide Y receptor Y1 (NYP1R)] in the top cluster according to the PPI and MCODE analyses. These genes were substantially enriched in the cytokine-cytokine receptor interaction and chemokine signaling pathways. The top three core genes (CXCL13, NYP1R, and CXCR2) were selected and verified in a mouse model of colitis using real-time PCR. Increased expression was observed compared with the control mice, but only CXCR2 expression was significantly different.

CONCLUSION

Core DEGs identified in UC are related to inflammation and immunity inflammation, indicating that these reactions are core features of the pathogenesis of UC. CXCR2 may reflect the degree of inflammation in patients with UC.

Key Words: Ulcerative colitis; Bioinformatics analysis; C-X-C motif chemokine ligand 13; Neuropeptide Y receptor Y1; C-X-C motif chemokine receptor 2; Colitis model mice

©The Author(s) 2020. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: Two microarray datasets were used, and differentially expressed genes were analyzed. Seventeen core genes were upregulated, and one was downregulated. These genes were markedly enriched in the cytokine-cytokine receptor interaction and chemokine signaling pathways. The top three core genes [C-X-C motif chemokine ligand 13, neuropeptide Y receptor Y1, C-X-C motif chemokine receptor 2 (CXCR2)] were verified in a dextran sulfate sodium-induced colitis model mice by real-time polymerase chain reaction and showed an increased expression, but only CXCR2 was statistically different. CXCR2 may represent a new biomarker to determine the degree of inflammation or a treatment target in ulcerative colitis.

Citation: Shi L, Han X, Li JX, Liao YT, Kou FS, Wang ZB, Shi R, Zhao XJ, Sun ZM, Hao Y. Identification of differentially expressed genes in ulcerative colitis and verification in a colitis mouse model by bioinformatics analyses. World J Gastroenterol 2020; 26(39): 5983-5996
URL: https://www.wjgnet.com/1007-9327/full/v26/i39/5983.htm
DOI: https://dx.doi.org/10.3748/wjg.v26.i39.5983

INTRODUCTION

Ulcerative colitis (UC), a nonspecific inflammatory disease that occurs in the colon or rectum, is the most common type of inflammatory bowel disease (IBD) and tends to occur in young and middle-aged people. Symptoms of UC, such as diarrhea, bloody stools with mucus-pus and abdominal pain, substantially affect quality of life[1][2]. Because UC is difficult to cure, readily relapses and has a high risk of cancer, UC has been classified as a refractory disease by the World Health Organization.

Epidemiological studies have confirmed that the annual incidence of UC worldwide is 10.5-14 cases per 100000 people, and the prevalence rate is approximately 246.7 cases
per 100000 people\(^1\). Currently, the number of patients in Europe and the United States accounts for 0.5% of the global population, and the incidence and prevalence rates in urban areas have exceeded those in rural areas. Since the 1990s, the number of children who were initially treated for IBD in Western developed countries has gradually increased. In the past 20 years, the incidence and prevalence of IBD in developed countries have begun to stabilize and plateau, but in other countries an obvious increase in the incidence and prevalence still occurs, particularly in South America and East Asia\(^4\). Recently, a systematic review of 140 studies in 2018 showed that the incidence of IBD in children increased in both developed and developing countries\(^5\).

The pathogenesis of UC is complex, and the interaction of multiple factors may lead to its occurrence. With contributions from the environment, mental factors and intestinal flora, the intestinal barrier of genetically susceptible people is destroyed, and the immune system is dysfunctional, resulting in excessively hyperactive immune reactions and inflammation\(^6\). The combination of C-reactive protein and fecal calprotectin levels may be beneficial for dynamically diagnosing and monitoring the progress of UC at present\(^7\). However, the limited number of significant and specific biomarkers for UC has become an increasingly prominent problem in its diagnosis and treatment.

Microarray technology reveals numerous genes that are activated in different tissues as well as their physiological and pathological statuses and has been regarded as a novel approach for clarifying the mechanisms underlying different diseases\(^8\). In recent years with the optimization of gene sequencing platforms, differentially expressed genes (DEGs) have been identified using bioinformatics analyses\(^9\). To date, several studies have reported the bioinformatics analysis of IBD using arrays or chips\(^10\), but the analysis of UC is still lacking. Thus, bioinformatics methods may help us study and more clearly understand the underlying mechanisms of UC\(^11\).

In this study, we applied two databases from Gene Expression Omnibus (GEO), GEO2R and online tools for constructing Venn diagrams to identify the DEGs, including upregulated and downregulated genes. Then, the Database for Annotation, Visualization and Integrated Discovery (DAVID) was used to analyze the DEGs based on the molecular function (MF), cellular component (CC), and biological process (BP) and different Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways\(^12\). Third, we constructed a protein-protein interaction (PPI) network using Cytoscape and the Molecular Complex Detection (MCODE) app for further analysis. Using these methods, core genes were chosen, and the KEGG pathway enrichment analysis was repeated. Finally, we verified the top three core genes in the first cluster selected by MCODE in colon tissues from mice with dextran sulfate sodium (DSS)-induced colitis using real-time polymerase chain reaction (PCR).

**MATERIALS AND METHODS**

**Microarray data**

Microarray data were obtained from the National Center for Biotechnology Information-GEO, a free public database of microarray and gene profiles, and we obtained the gene expression profiles of GSE92415 and GSE87466 in colon mucosal tissues from patients with UC and healthy individuals. Microarray data included in GSE92415 and GSE87466 were obtained from GPL13158 Platforms (HT_HG-U133_Plus_PM) Affymetrix HT HG-U133+ PM Array Plate) and included 162 colonic mucosal tissues from patients with UC (who were treated with the anti-TNF agent golimumab and placebo) and 21 healthy colon mucosal tissues and 87 samples from patients with UC and 21 normal colon mucosal tissues, respectively.

**DEG analysis**

DEGs were identified between specimens from patients with UC and normal colonic mucosa using the GEO2R online tool with a |logFC| > 2 and adjusted P value < 0.05. Then, the raw data in TXT format were imported into the online tool to construct a Venn diagram and identify the common DEGs among the two datasets. The DEGs with a logFC < 0 were considered downregulated genes, while the DEGs with a logFC > 0 were considered upregulated genes.

**Gene Ontology and KEGG pathway enrichment analyses**

Gene Ontology (GO) analysis is a commonly used method for identifying the biological properties of high-throughput transcriptome or genome data for genes and their RNA or protein products\(^13\). KEGG is a collection of databases for genomes,
diseases, biological pathways, drugs and chemical materials. DAVID is an online bioinformatics tool designed to identify a large number of gene or protein functions\[^{10}\]. We used DAVID and performed a GO analysis to visualize the enrichment of DEGs in BP, MF, cellular components and KEGG pathways.

**PPI network and module analysis**

PPI information was evaluated using an online tool, the Search Tool for the Retrieval of Interacting Genes (STRING). A tab-separated file containing the values for the network from STRING was imported into Cytoscape (continuous mapping for columns depended on the frequency of genes; default size of the column = 30.0; upregulated genes were red and downregulated genes were blue), which was applied to detect the potential interactions between these DEGs\[^{16}\]. In addition, the MCODE app in Cytoscape was used to assess the modules of the PPI network (cutoff = 2, max depth = 100, k-core = 2, node score cutoff = 0.2) and to identify the core genes in the top cluster in UC. Meanwhile, the KEGG pathway enrichment analysis of these core DEGs was repeated.

**Establishment of the mouse model of colitis**

We complied with the ethics standard for research activity established by the Animal Ethics Committee of Beijing University of Chinese Medicine in accordance with the guidelines issued by the Regulations of Beijing Laboratory Animal Management. Fourteen male specific-pathogen-free (SPF) C57BL/6 mice (weighing 20 ± 2 g) were procured from SPF Biological Technology Co., Ltd., Beijing, China (Certificate No. SCXK [jing] 2019-0010). They were housed in the SPF animal center of Beijing University of Chinese Medicine at a constant temperature of 20-26 °C and humidity of 50% to 60% in a light-controlled environment with a 12-h light/dark cycle. In addition, fodder and sterilized water were supplied. After adaptive feeding for 1 wk, these mice were divided into the control (n = 4) and model (n = 10) groups using the random number table method. Mice in the model group drank 3.5% (weight/volume) DSS (average molecular weight 36000-50000, MP Biomedicals) for 7 consecutive days\[^{17}\].

**Histological analysis of colon tissues from the colitis model mice**

On the 7th d of the experiment, all the mice were fasted for 24 h. On the 8th d, mice were anesthetized by ether inhalation until the superficial reflex disappeared, and the mice were sacrificed by cervical spondylolisthesis. After disinfecting and preparing the skin, the abdomen was incised to separate the colon tissue. Specimens of colon tissues (three 1 cm pieces) from each mouse were fixed with a 10% neutral formaldehyde solution for at least 24 h. The lesion score was judged by the histological criterion after hematoxylin-eosin staining of 4 μm thick sections: 0, no signs of inflammation; 1, a low level of leukocyte infiltration (10-30 leukocytes per high-power field); 2, a moderate level of leukocyte infiltration (31-70 leukocytes per high-power field); 3, a high level of leukocyte infiltration (> 71 leukocytes per high-power field), high vascular density and thickening of the bowel wall; and 4, transmural infiltrations, loss of goblet cells, high vascular density, strong bowel wall thickening, ulcerations and cryptic abscesses\[^{18}\].

**Verification of top three core genes in the colitis model mice using real-time PCR**

The other tissues were placed in cryotubes, frozen in liquid nitrogen overnight and stored at -80 °C. Total RNA was extracted from the colon tissues using TRIzol reagent (Invitrogen, United States). Specific primers were used to amplify the core genes (Table 1), and the expression of each gene was normalized to β-actin with the standard curve method. Reverse transcription was performed at 42 °C for 60 min, and reverse transcriptase was inactivated at 70 °C for 15 min. The relative quantitative analysis was performed using the 2-ΔΔCT approach.

**Statistical analysis**

The histological lesion score and relative expression levels obtained using real-time PCR were analyzed with SPSS (version 22.0, IBM, Corp., Armonk, NY, United States). Data were distributed normally, and the two groups were compared with an independent-sample t test. A P value < 0.05 indicated a statistically significant difference.
Table 1 Sequences of primers

| Gene     | Sequence, 5' to 3'          |
|----------|----------------------------|
| β-actin  | Forward: GAGATTACTGCTCTGGCTCCTA |
|          | Reverse: GGACTCATCGTACTCTCGCTTG |
| CXCL13   | Forward: CCAAAACCATCATTGTTCTTT |
|          | Reverse: CTGAAGTCCATCTGGCAAACTCT |
| NPY1R    | Forward: CCAGTGAGACCAAGCGAATCAACA |
|          | Reverse: GCGGTGAGGTGACAGAACAGAA |
| CXCR2    | Forward: ATGGTGGAACTTGTGGCTCTAGC |
|          | Reverse: CCATTCCTGGTCCTGGCATTCC |

CXCL13: C-X-C motif chemokine ligand 13; NPY1R: Neuropeptide Y receptor Y1; CXCR2: C-X-C motif chemokine receptor 2.

RESULTS

Identification of DEGs

One hundred and ninety-six DEGs were identified in the GSE92415 database, including 130 upregulated and 66 downregulated genes. One hundred and eighty-four DEGs were identified in the GSE87466 database, including 122 upregulated and 62 downregulated genes. Among the two databases, 177 genes (118 upregulated and 59 downregulated) overlapped (Figure 1A and 1B).

Functional enrichment of DEGs in the GO analysis

Three categories of DEGs, MF, BP and CC, were classified by performing a GO analysis. Among these GO functions, extracellular region (GO: 0005576), defense response (GO: 0006952), extracellular region part (GO: 0044421), response to wounding (GO: 0009611), extracellular space (GO: 0005615) and inflammatory response (GO: 0006954) were the top six terms associated with UC with more than 30 genes identified in each category (Figure 1C). Upregulated genes were mainly enriched in defense response, inflammatory response, response to wounding, extracellular region, extracellular region part, extracellular space, cytokine activity, chemokine activity and chemokine receptor binding and were the top three enriched terms, depending on the P value of the respective categories (Table 2, P < 0.05). Downregulated genes were mainly enriched in transmembrane transport, carboxylic acid transport, organic acid transport, cell fraction, membrane fraction, insoluble fraction, symporter activity, xenobiotic-transporting ATPase activity and multidrug transporter activity and were the top three enriched terms, based on the P value of the respective categories (Table 3, P < 0.05). The KEGG pathway enrichment analysis suggested that DEGs predominantly participated in inflammation-related pathways, including the chemokine signaling pathway, cytokine-cytokine receptor interaction and complement and coagulation cascades (Table 4, P < 0.05).

Analysis of the PPI network using Cytoscape

One hundred and seventy-seven DEGs were entered into the PPI network using the Search Tool for the Retrieval of Interacting Genes, including 130 nodes and 639 edges. The thickness of the edges was adjusted based on the level of the combined score. The larger the area of a protein node, the more protein nodes that interacted with it. Upregulated genes were shown in red, and downregulated genes were shown in blue (Figure 2A). Then, we applied the MCODE app for further analysis and identified seven clusters with close interactions in patients with UC. Seventeen core genes were upregulated [C-X-C motif chemokine ligand 13 (CXCL13), C-X-C motif chemokine receptor 2 (CXCR2), CXCL9, CXCL5, C-C motif chemokine ligand 18 (CCL18), interleukin 1 beta (IL1B), matrix metallopeptidase 9 (MMP9), CXCL3, formyl peptide receptor 1 (FPR1), complement component 3 (C3), CXCL8, CXCL1, CXCL10, CXCL2, CXCL6, CXL11 and hydroxycarboxylic acid receptor 3 (HCAR3)], and one gene was downregulated [neuropeptide Y receptor Y1 (NPY1R)] in the first significant cluster (Figure 2B).
| Category | Term | Count | Genes |
|----------|------|-------|-------|
| BP       | Defense response (GO: 0006952) | 40 | CXCL1, KYNU, S100A8, FGR, C3, CXCL3, CXCL2, S100A9, CXCL9, CXCR2, CXCL6, DEF84A, CXCL11, CXCL10, REG3A, SERPINA3, CSF3R, VNN1, IL1B, LTF, SERPINA1, CFI, SPP1, SELP, CR1, PDGN, CFB, IL1RN, C4BPB, IDO1, CCL18, S100A12, IL1R2, TNFAIP6, PROK2, DEF6, CXCL13, DEF5, PLA2G2D, DMBT1. |
| BP       | Inflammatory response (GO: 0006954) | 31 | CXCL1, S100A8, C3, CXCL3, CXCL2, S100A9, CXCL9, CXCR2, CXCL6, CXCL11, CXCL10, REG3A, SERPINA3, VNN1, IL1B, SERPINA1, CFI, SPP1, SELP, CR1, PDGN, CFB, IL1RN, C4BPB, IDO1, CCL18, S100A12, TNFAIP6, PROK2, CXCL13, PLA2G2D. |
| BP       | Response to wounding (GO: 0009611) | 33 | CXCL1, S100A8, C3, TNC, CXCL3, CXCL2, S100A9, CXCL9, CXCR2, CXCL6, CXCL11, CDH3, CXCL10, REG3A, SERPINA3, VNN1, IL1B, SERPINA1, CFI, SPP1, SELP, CR1, PDPN, CFB, IL1RN, C4BPB, IDO1, CCL18, S100A12, TNFAIP6, PROK2, CXCL13, PLA2G2D. |
| CC       | Extracellular region (GO: 0005576) | 55 | REG4, MMP9, MMP7, CXCL11, MMP3, MMP1, CXCL10, REG3A, SERPINA3, CSF3R, IL1B, LTF, SERPINA1, KLK10, CFI, FCGR3B, CYR61, IL13RA2, EGFL6, IL24, TCN1, MMP12, MMP10, PROK2, GZMK, DEF6, SERPINB5, DEF5, STC1, PLA2G2D, CXCL1, CXCL5, C3, CXCL3, TNC, CXCL2, CXCL9, CXCL6, DEF84A, GREM1, TIMP1, THBS2, OLFM1, SPP1, SELP, CFB, IL1RN, CHI3L1, CHI3L2, C4BPB, CXCL18, CCN2, CXCL13, PI3, DMBT1. |
| CC       | Extracellular region part (GO: 0044421) | 38 | CXCL1, CXCL5, C3, MMP9, TNC, CXCL3, CXCL2, MMP7, CXCL9, CXCL6, CXCL11, MMP3, CREM, C3, CXCL10, TIMP1, REG3A, CXCL10, CXCL13, CXCL5, C3, CXCL3, TNC, CXCL2, CXCL9, CXCL6, DEF84A, GREM1, TIMP1, THBS2, OLFM1, SPP1, SELP, CFB, IL1RN, CHI3L1, CHI3L2, C4BPB, CCL18, CN2, CXCL13, PI3, DMBT1. |
| CC       | Extracellular space (GO: 0005615) | 33 | CXCL1, CXCL5, C3, MMP9, TNC, CXCL3, CXCL2, MMP7, CXCL9, CXCL6, CXCL11, MMP3, CREM, CXCL10, TIMP1, REG3A, CXCL10, CXCL13, CXCL5, C3, CXCL3, TNC, CXCL2, CXCL9, CXCL6, DEF84A, GREM1, TIMP1, THBS2, OLFM1, SPP1, SELP, CFB, IL1RN, CHI3L1, CHI3L2, C4BPB, CCL18, CN2, CXCL13, PI3, DMBT1. |
| MF       | Cytokine activity (GO: 00035125) | 16 | CXCL1, NAMPT, CXCL5, CXCL3, IL1RN, CXCL2, CXCL9, CXCL6, IL24, GREM1, CXCL11, CXCL3, MMP3, REG3A, IL1B, SERPINA1, CFI, OLFM1, SPP1, IL13RA2, SELP, EGF6, IL1RN, CHI3L1, CHI3L2, IL24, CCL18, MMP12, MMP10, SERPINB5, DEF6, CXCL13, DEF5, PI3, STC1. |
| MF       | Chemokine activity (GO: 0008009) | 10 | CXCL1, CXCL5, CXCL3, CXCL2, CXCL9, CXCL6, CXCL11, CXCL18, CXCL10. |
| MF       | Chemokine receptor binding (GO: 0042379) | 10 | CXCL1, CXCL5, CXCL3, CXCL2, CXCL9, CXCL6, CXCL11, CXCL18, CXCL10. |

MF: Molecular function; CC: Cellular component; BP: Biological process.

Repeated KEGG pathway enrichment analysis of 18 core genes

The KEGG pathway enrichment analysis was repeated using the DAVID GO analysis to identify the pathways in which these 18 core DEGs in the top cluster (18 nodes, 142 edges, and a score of 16.706) were involved ($P < 0.05$). These DEGs were markedly enriched in the cytokine-cytokine receptor interaction and chemokine signaling pathways (Table 5).

Verification of the top three genes in the colitis model mice

Finally, we verified the top three genes (CXCL13, NPY1R and CXCR2) among the 18 core DEGs in the colitis model mice. Four mice in the model group were sacrificed during the experimental period because of the severity of disease, and the colon tissues of all the remaining mice (control: $n = 4$; colitis model: $n = 6$) were observed under a microscope. The normal four-layer structure of intestinal tissues, goblet cells and crypts disappeared, and inflammatory cells had infiltrated the submucosa in the
Table 3 Top three enriched terms in the respective categories of downregulated differentially expressed genes identified in ulcerative colitis (P < 0.05)

| Category | Term | Count | Genes |
|----------|------|-------|-------|
| BP       | Transmembrane transport (GO: 0055085) | 9 | TRPM6, SLC23A1, SLC16A1, SLC25A34, SLC17A4, ABCB1, AQP7, SLC30A10, SLC26A2. |
| BP       | Carboxylic acid transport (GO: 0046942) | 5 | SLC38A4, SLC23A1, SLC16A1, AQP8, SLC3A1. |
| BP       | Organic acid transport (GO: 0015849) | 5 | SLC38A4, SLC23A1, SLC16A1, AQP8, SLC3A1. |
| CC       | Cell fraction (GO: 0000267) | 12 | HSD3B2, SLC23A1, SLC16A1, AQP8, SLC17A4, CYP2B6, MEP1B, ABCB1, ANPEP, SLC3A1, SLC26A2, PKC1. |
| CC       | Membrane fraction (GO: 0005624) | 10 | HSD3B2, SLC23A1, SLC16A1, AQP8, SLC17A4, CYP2B6, MEP1B, ABCB1, SLC3A1, SLC26A2. |
| CC       | Insoluble fraction (GO: 0005624) | 10 | HSD3B2, SLC23A1, SLC16A1, AQP8, SLC17A4, CYP2B6, MEP1B, ABCB1, SLC3A1, SLC26A2. |
| MF       | Symporter activity (GO: 0015293) | 4 | SLC38A4, SLC23A1, SLC16A1, SLC17A4. |
| MF       | Xenobiotic-transporting ATPase activity (GO: 0008559) | 2 | ABCB1, ABCG2. |
| MF       | Multidrug transporter activity (GO: 0015239) | 2 | ABCB1, ABCG2. |

MF: Molecular function; CC: Cellular component; BP: Biological process.

Table 4 Enriched terms in the Kyoto Encyclopedia of Genes and Genomes pathways for differentially expressed genes identified in ulcerative colitis (P < 0.05)

| Pathway ID | Term | Genes |
|------------|------|-------|
| bsa04062   | Chemokine signaling pathway | CXCL1, FGR, CXCL5, CXCL3, CXCL2, CXCL9, CXCR2, CXCL6, CXCL11, CCL18, CXCL10, CXCL13, JAK3. |
| bsa04060   | Cytokine-cytokine receptor interaction | CXCL1, CXCL5, CXCL3, CXCL2, CXCL9, CXCR2, CXCL6, IL24, CXCL11, CCL18, CXCL10, CXCL13, CSF3R, IL1B. |
| bsa04610   | Complement and coagulation cascades | CR1, C3, CFB, C4BP, SERPINA1, CFI. |
| bsa04514   | Cell adhesion molecules | SELP, SELD, CD274, CTLA4, CLDN2, CDH3. |
| bsa00830   | Retinol metabolism | CYP2B6, ADH1C, UGT2A3. |
| bsa00980   | Metabolism of xenobiotics by cytochrome P450 | CYP2B6, ADH1C, UGT2A3. |
| bsa04620   | Toll-like receptor signaling pathway | CXCL8, IL1B, CXCL11, SPP1, CXCL10. |
| bsa00982   | Drug metabolism | CYP2B6, ADH1C, UGT2A3. |
| bsa03320   | PPAR signaling pathway | HMGCS2, AQP7, PKC1. |

PPAR: Peroxisome proliferator-activated receptor.

model group compared with the control group (Figure 3A-D). Histological lesions indicated the successful establishment of the colitis model (Figure 3E). The CXCL13 and CXCR2 mRNAs were expressed at higher levels in the colon tissues from the colitis model mice than in the mice from the control group (Figure 4), and the difference in CXCR2 expression was significant (P < 0.01). These manifestations were consistent with the results of the bioinformatics analysis in the present study. Interestingly, higher expression of the NPY1R mRNA was also observed in the colitis model mice (Figure 4), which differed from our bioinformatics results.

**DISCUSSION**

The incidence of UC in China and other Asian countries is gradually increasing[^19], but due to the complicated pathogenesis, accurate molecular biomarkers for diagnosing UC are still lacking. In the past, biomarkers such as the C-reactive protein level and...
Table 5 Repeated Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis of 18 core genes in the first cluster (P < 0.05)

| Category     | Term                                                                 | Count, % | Genes                      |
|--------------|----------------------------------------------------------------------|----------|-----------------------------|
| KEGG pathway | hsa04060, cytokine-cytokine receptor interaction.                     | 12, 6.06 | CXCL1, CXCL5, CXCL13, CXCL3, CXCL2, CXCL9, IL1B, CXCR2, CXCL6, CXCL11, CCL18, CXCL10. |
| KEGG pathway | hsa04062, chemokine signaling pathway.                                | 11, 5.56 | CXCL1, CXCL5, CXCL13, CXCL3, CXCL2, CXCL9, CXCR2, CXCL6, CXCL11, CCL18, CXCL10. |
| KEGG pathway | hsa04620, toll-like receptor signaling pathway.                       | 4, 2.02  | CXCL19, IL1B, CXCL11, CXCL10. |
| KEGG pathway | hsa04621, NOD-like receptor signaling pathway.                        | 3, 1.52  | CXCL1, CXCL2, IL1B.          |

KEGG: Kyoto Encyclopedia of Genes and Genomes.

Chemokines and cytokines play pivotal roles in regulating mucosal inflammation and the immune system by promoting neutrophil migration to sites of inflammation, ultimately leading to tissue damage and destruction. Carlsen et al. reported the expression of CXCL13 in both healthy individuals and patients with UC, in addition to reports in rodents from published studies; however, differences in the expression of CXCL13 between patients with UC and healthy individuals were unclear. Singh et al. reported significantly increased expression of CXCL13 in patients with IBD, including patients with UC and Crohn’s disease, compared to healthy controls.

On the other hand, CXCR2 has been reported to regulate the migration and recruitment of neutrophils to the site of inflammation. A review also suggested that in addition to calprotectin, CXCR2, a neutrophil-related protein, may have potential roles in diagnosis and treatment. As shown in the study by Farooq et al., CXCR2-positive mice have more severe symptoms, such as the infiltration of polymorphonuclear neutrophils (PMNs), than CXCR2-deficient mice with DSS-induced colitis. The mechanism of increased CXCR2 expression in colitis has not been conclusively determined, as the infiltration of increased numbers of PMNs in the mucosa or submucosa is a feature of DSS-induced colitis. CXCR2 functions as a PMNchemokine receptor, which would lead to a significantly higher level in individuals with colitis. Our study verified this hypothesis by showing neutrophil infiltration and a high level of the CXCR2 mRNA.

Notably, the expression of the NPY1R mRNA in the colitis model mice differed from our bioinformatics results. NPY is a 36-amino acid peptide with a wide distribution in the central and peripheral nervous system that evokes numerous physiological...
Figure 1 Identification of differentially expressed genes in the two databases (GSE92415 and GSE87466) and Gene Ontology analysis of differentially expressed genes in ulcerative colitis. A: Upregulated differentially expressed genes (DEGs); B: Downregulated DEGs. DEGs were identified with a t-test, and statistically significant DEGs were defined by the GEO2R online tool with a |logFC| > 2 and adjusted P value < 0.05; C: The Gene Ontology analysis classified the DEGs into three groups: Molecular function, biological process and cellular component. Terms were selected with > 15 genes and arrayed in ascending order from top to bottom according to the count. GO: Gene Ontology.

responses by activating different receptors (Y1, Y2, Y3)\(^3\). NPY is considered the most potent orexigenic neuropeptide and may be involved in the stress response, anxiety and mood-related disorders as well as the regulation of the immune system and cancer\(^3\).

NPY1R expression was decreased in patients with UC based on the results of the bioinformatics analysis; however, its expression was increased in the colitis model mice as evidenced by real-time PCR, which might be related to the four aspects described below. First, our bioinformatics results were derived from human colon specimens, which were fundamentally heterologous compared to the mice. Second, microarray data were obtained from the human colonic mucosa, while RT-PCR was performed on mouse colon tissues, which contained the mucosa, submucosa, muscle layer and subserosa. Thus, differences in expression were observed between the bioinformatics results and in vivo analysis. Third, because of the current shortage of microarrays containing samples from patients with UC in the same platforms, we chose the GSE92415 microarray dataset, including patients with UC who were treated with a TNF-α inhibitor and placebos, and thus the result potentially did not accurately reflect the detailed scientific features of UC. Additionally, a 3.5% DSS solution was
Figure 2 Protein-protein interactions of differentially expressed genes and the most significant module cluster identified by Molecular Complex Detection in the protein-protein interaction network of ulcerative colitis. A: Protein-protein interaction network of differentially expressed genes determined using Cytoscape. One hundred and seventy-seven differentially expressed genes from the Search Tool for the Retrieval of Interacting Genes online database were screened using Cytoscape, including 130 nodes and 639 edges. Upregulated genes were shown in red, and downregulated genes were shown in blue; B: The most significant cluster was analyzed with the Molecular Complex Detection app in Cytoscape. Seventeen core genes were upregulated, and one gene was downregulated; a larger node indicated more interactions with a gene or a protein.

used to establish the colitis model and combined with the histological lesions, the model was much more severe. Further studies are needed to confirm whether dysplasia of the intestinal epithelium occurred, and the NPY/NPY1R system was activated.

CONCLUSION

In conclusion, the core DEGs identified in patients with UC were CXCL13, NPY1R, CXCR2, CXCL9, CXCL5, CCL18, IL1B, MMP9, CXCL3, formyl peptide receptor 1, C3, CXCL8, CXCL1, CXCL10, CXCL2, CXCL6, CXCL11 and HCAR3. These DEGs are related to inflammation and immune-inflammatory reactions, indicating that inflammation and abnormal activation of the immune system may represent the core features of the pathogenesis of UC. Based on the current data, we propose that CXCR2 may represent a new biomarker for the degree of inflammation or a treatment target, and our study may provide new insights into the diagnosis and treatment of UC.
Figure 3 Hematoxylin-eosin staining and histological lesion score of colon tissues. Hematoxylin-eosin staining of colon tissues from the control and dextran sulfate sodium-induced colitis model mice. A: Control mice (× 100); B: Control mice (× 200); C: Colitis mice (× 100); D: Colitis mice (× 200); E: Histological lesion scores of colon tissues. Numerous neutrophils infiltrated and the crypts, goblet cells and normal four-layer structure of colon disappeared in the colitis model mice. Compared to the score of the control group (n = 4, 12 pieces), the score of the model group (n = 6, 18 pieces) increased significantly (P < 0.01).
Figure 4 Real-time polymerase chain reaction of the top three core genes from the first cluster in colon tissues from the colitis model and control mice. A, C, E, and G: Amplification curves for β-actin, C-X-C motif chemokine ligand 13, neuropeptide Y receptor Y1 and C-X-C motif chemokine receptor 2 (CXCR2); B, D, F and H: Melting peaks for β-actin, C-X-C motif chemokine ligand 13, neuropeptide Y receptor Y1 and CXCR2; I: Relative expression obtained using real-time polymerase chain reaction. C-X-C motif chemokine ligand 13, neuropeptide Y receptor Y1 and CXCR2 expression, particularly CXCR2 expression (P < 0.01), increased in the colitis model mice. The other two genes were expressed at higher levels in the colitis mice than in the control mice, but the difference was not significant. CXCL13: C-X-C motif chemokine ligand 13; NPY1R: Neuropeptide Y receptor Y1; CXCR2: C-X-C motif chemokine receptor 2.

ARTICLE HIGHLIGHTS

Research background
Ulcerative colitis (UC) tends to occur in young and middle-aged people. It substantially affects the patient’s quality of life because it is difficult to cure, readily relapses and poses a high risk of colon cancer. However, the pathogenesis of UC is complex and multifaceted, and specific biomarkers for UC are currently unavailable.

Research motivation
In recent years, with the optimization of gene sequencing platforms, differentially expressed genes (DEGs) have been identified through bioinformatics analyses by comparing microarrays. To date, several studies have reported the results of bioinformatics analyses of samples from patients with inflammatory bowel disease using arrays or chips, but the analysis of patients with UC is still lacking. The specific molecules or biomarkers of UC are insufficient. Thus, we will apply bioinformatics methods to more clearly elucidate the underlying biomarkers and mechanisms of UC.

Research objectives
To identify UC-related DEGs by performing a bioinformatics analysis and verify them in vivo and to identify novel biomarkers and the underlying mechanisms of UC.
Research methods
Two microarray datasets from the National Center for Biotechnology Information-Gene Expression Omnibus database were used, and DEGs were analyzed using GEO2R and Venn diagrams. We annotated these genes based on functions and signaling pathways. Then protein-protein interaction (PPI) were constructed using the Search Tool for the Retrieval of Interacting Genes. The data were further analyzed with Cytoscape software and the Molecular Complex Detection (MCODE) app. The core genes were selected, and the Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis was repeated. Finally, colitis model mice were established by administering dextran sulfate sodium, and the top three core genes were verified in colitis mice using real-time polymerase chain reaction.

Research results
One hundred and seventy-seven DEGs (118 upregulated genes and 59 downregulated genes) predominately participated in inflammation-related pathways. Seventeen core genes were upregulated, and one gene was downregulated in the first cluster according to the PPI and MCODE analyses in Cytoscape. These genes were markedly enriched in the cytokine-cytokine receptor interaction and chemokine signaling pathways. The top three core genes showed increased expression compared with the control mice, but only the difference in C-C motif chemokine receptor 2 (CXCR2) expression was statistically significant. CXCR2 may reflect the degree of inflammation in patients with UC and serve as an underlying treatment target.

Research conclusions
Core DEGs identified in patients with UC are related to inflammation and immune inflammatory reactions, indicating that these reactions are core features of the pathogenesis of UC. CXCR2 may reflect the degree of inflammation in patients with UC.

Research perspectives
CXCR2 may represent a new biomarker to determine the degree of inflammation or a treatment target in UC. In the future, the combination of CXCR2 with other biomarkers will potentially improve the ability to diagnose and dynamically monitor UC.

REFERENCES
1. da Silva BC, Lyra AC, Rocha R, Santana GO. Epidemiology, demographic characteristics and prognostic predictors of ulcerative colitis. World J Gastroenterol 2014; 20: 9458-9467. [PMID: 25071340 DOI: 10.3748/wjg.v20.i28.9458]
2. Yu YR, Rodriguez JR. Clinical presentation of Crohn's, ulcerative colitis, and indeterminate colitis: Symptoms, extraintestinal manifestations, and disease phenotypes. Semin Pediatr Surg 2017; 26: 349-355. [PMID: 29126502 DOI: 10.1053/j.sempedsurg.2017.10.003]
3. Shivashankar R, Tremaine WJ, Harmsen WS, Loftus EV Jr. Incidence and Prevalence of Crohn's Disease and Ulcerative Colitis in Olmsted County, Minnesota From 1970 Through 2010. Clin Gastroenterol Hepatol 2017; 15: 857-863. [PMID: 27856364 DOI: 10.1016/j.cgh.2016.10.039]
4. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, Panaccione R, Ghosh S, Wu JCY, Chan FKCL, Sung JYF, Kaplan GG. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet 2018; 390: 2769-2778. [PMID: 29050646 DOI: 10.1016/S0140-6736(17)32448-0]
5. Torres J, Mehandru S, Colombel JF, Peyrin-Biroulet L. Crohn's disease. Lancet 2017; 389: 1741-1755. [PMID: 27914655 DOI: 10.1016/S0140-6736(16)31711-1]
6. Sykora J, Pomahačová R, Kreslová M, Cvalínová D, Štych P, Schwarz J. Current global trends in the incidence of pediatric-onset inflammatory bowel disease. World J Gastroenterol 2018; 24: 2741-2763. [PMID: 29991879 DOI: 10.3748/wjg.v24.i25.2741]
7. Uagaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. Lancet 2017; 389: 1756-1770. [PMID: 27914657 DOI: 10.1016/S0140-6736(16)32126-2]
8. Parikh K, Antanaviciute A, Fawker-Corbett D, Jagielowicz M, Aulicino A, Lagerholm C, Davis S, Kinchen J, Chen HH, Alham NK, Ashley N, Johnson E, Humpl J, Liu J, Andev RS, Bartscherer K, Jacob A, Hoheisel JD, Angenendt P. Microarray technology as a universal tool for high-throughput analysis of biological systems. Comb Chem High Throughput Screen 2006; 9: 365-380.
Y1 receptor targeting in health and disease.

Wan CP, neurodegenerative diseases.

Duarte-Neves J, CXCR2 on neutrophil recruitment and dextran sodium sulfate-induced colitis.

Farooq SM, MacDonald TT, Danese S, Berner Hansen M. Neutrophils in ulcerative colitis: a review of selected

Muthas D, Adenocarcinoma.

Inhibition Profoundly Suppresses Metastases and Augments Immunotherapy in Pancreatic Ductal

Chang DK, Evans TRJ, Strathdee D, Biankin AV, Nibbs RJB, Barry ST, Sansom OJ, Morton JP. CXCR2

Wilson Z, Eberlein C, Candido JB, Clarke M, Nixon C, Connelly J, Jamieson N, Carter CR, Balkwill F,

2002; (CXCL13) and its receptor CXCR5 are expressed in normal and aberrant gut associated lymphoid tissue.

Carlsen HS, 10.1093/ecco-jcc/jjx145 Bowel Disease; Pitfalls and Promise.

10.1016/j.cyto.2015.10.008 cytokine levels in inflammatory bowel disease patients.

Singh UP, GEO: archive for functional genomics data sets--update.

Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. NCBI

Barrett T, Feigenbaum LZ, Walsh AJ, increasing incidence and prevalence of the inflammatory bowel diseases with

Molodecky NA, Barkema HW, Kaplan GG. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology 2012; 142: 46-54.e2; quiz e30 [PMID: 22001864 DOI: 10.1053/j.gastro.2011.10.001]

Walsh AJ, Bryant RV, Travis SP. Current best practice for disease activity assessment in IBD. Nat Rev Gastroenterol Hepatol 2016; 13: 567-579 [PMID: 27580084 DOI: 10.1038/nrgastro.2016.128]

Feigenbaum LZ, Lee D, Ho J. Routine testing of folate levels in geriatric assessment for dementia. J Am Geriatr Soc 1988, 36: 755 [PMID: 3136199 DOI: 10.1016/cjc.2019.07.033]

Barrett T, Wilhite SE, Lodoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillips H, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res 2013; 41: D991-D995 [PMID: 23193258 DOI: 10.1093/nar/gks1193]

Singh UP, Singh NP, Murphy EA, Price RL, Fayad R, Nagarkatti M, Nagarkatti PS. Chemokine and cytokine levels in inflammatory bowel disease patients. Cytokine 2016; 77: 44-49 [PMID: 26520877 DOI: 10.1016/j.cytokine.2015.10.009]

Trivedi PJ, Adams DH. Chemokines and Chemokine Receptors as Therapeutic Targets in Inflammatory Bowel Disease; Pitfalls and Promise. J Crohns Colitis 2018, 12: S641-S652 [PMID: 30137309 DOI: 10.1093/ecco-jcc/jxs145]

Carlsen HS, Baekkveold ES, Johansen FE, Haraldsen G, Brandtzaeg P. B cell attracting chemokine 1 (CXCCL13) and its receptor CXCR5 are expressed in normal and aberrant gut associated lymphoid tissue. Gut 2002; 51: 364-371 [PMID: 12171958 DOI: 10.1136/gut.51.3.364]

Steele CW, Karin SA, Leach JG, Bailey P, Upsittil-Goddard R, Rishi L, Foth M, Bryson S, McDaid K, Wilson Z, Eberlin C, Candido JB, Clarke M, Nixon C, Connelly J, Jamieson N, Carter CR, Balkwill F, Chang DK, Evans TRJ, Strathean D, Binkanin AV, Nibbs RJ, Barry ST, Sansom OJ, Morton JP. CXCR2 Inhibition Profoundly Supresses Metastases and Augments Immunotherapy in Pancreatic Ductal Adenocarcinoma. Cancer Cell 2016; 29: 832-845 [PMID: 27265504 DOI: 10.1016/j.ccell.2016.04.014]

Muthas D, Reznichenko A, Balendran CA, Böttcher G, Clausen IG, Kjærman Mårh C, Ottosson T, Uddin M, MacDonald TT, Danese S, Berner Hansen M. Neutrophils in ulcerative colitis: a review of selected biomarkers and their potential therapeutic implications. Scand J Gastroenterol 2017; 52: 125-135 [PMID: 27610713 DOI: 10.1080/00365521.2016.1235224]

Farooq SM, Stillle R, Svensson M, Svansborg C, Strierer RM, Stadnyk AW. Therapeutic effect of blocking CXCR2 on neutrophil recruitment and dextran sodium sulfate-induced colitis. J Pharmacol Exp Ther 2009; 329: 123-129 [PMID: 19131582 DOI: 10.1124/jpet.108.145862]

Duarte-Neves J, Pereira de Almeida L, Cavadas C. Neuropeptide Y (NPY) as a therapeutic target for neurodegenerative diseases. Neurobiol Dis 2016, 95: 210-224 [PMID: 27461050 DOI: 10.1016/j.nbd.2016.07.022]

Wan CP, Lau BH. Neuropeptide Y receptor subtypes. Life Sci 1995; 56: 1055-1064 [PMID: 9001438 DOI: 10.1016/0025-2273(95)00041-4]

Hofmann S, Bellmann-Sickert K, Beck-Sickinger AG. Chemical modification of neuropeptide Y for human Y1 receptor targeting in health and disease. Biochim Biophys Acta 2019; 400: 299-311 [PMID: 30653463 DOI: 10.1515/haz-2018-0364]
