Quality Improvement Study

Medicine®

OPEN

Identification and analysis of key genes associated with ulcerative colitis based on DNA microarray data

Ruifeng Song, MD, Ya Li, MD, Weiwei Hao, MM, Bingxue Wang, MM, Lei Yang, MM, Feng Xu, MD*

Abstract

This study aimed to identify genes that may play a role in development of ulcerative colitis (UC) and gain insight into its pathogenesis.

Gene expression profiling data, including samples collected from 13 early-stage UC (EUC), 8 advanced-stage UC (AUC), and 5 control subjects, were downloaded from the Gene Expression Omnibus database under the accession number of GSE9452. Differentially expressed genes (DEGs) were identified in EUC and AUC compared with controls. DEGs for EUC and AUC, as well as AUC-specific DEGs were subjected to pathway enrichment analysis. Random Walk with Restart (RWR) was used to identify DEGs that are critical in UC based on a protein-to-protein interaction (PPI) network and the inflammatory bowel disease (IBD) pathway downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. IL17 and transforming growth factor beta TGF-β) expression levels in colonic tissue from patients with UC and normal colonic mucosa from healthy adults were analyzed by immunohistochemistry (IHC).

A total of 3511 and 911 DEGs were identified in AUC and EUC, respectively. The overlapping DEGs and the AUC-specific DEGs were both enriched in pathways related to immunity, such as antigen processing and presentation. AUC-specific DEGs were related to cell migration, such as ECM-receptor interaction. Following DEG prioritization, *TLR4* and *STAT1* were linked with EUC, AUC, and CD. The upregulated gene *TGFB* increased the number of Th17 cells, as verified by IHC. Furthermore, *PIK3R1*, *CREBBP*, and *STAT1* were part of high-degree nodes in the PPI sub-network.

The upregulated gene *TGFB* may regulate IL17 expression in UC. *PIK3R1* may participate in immunity and *CREBBP* may interact with *STAT1* in the development and progression of UC.

Abbreviations: AUC = advanced-stage UC, CAM = cell adhesion molecule, CD = Crohn disease, DEG = differentially expressed gene, ECM = extracellular matrix, EUC = early-stage UC, IBD = inflammatory bowel disease, IHC = immunohistochemical, KEGG = Kyoto Encyclopedia of Genes and Genomes, PPI = protein-to-protein interaction, RWR = Random Walk with Restart, STRING = Search Tool for the Retrieval of Interacting Genes, TGF-β = transforming growth factor beta, UC = ulcerative colitis.

Keywords: data verification, differentially expressed genes, inflammatory bowel disease, protein-to-protein interaction

Highlights

Cell migration ability may increase in AUC.
The upregulated gene *TGFB* may regulate IL17 expression in UC.
PIK3R1 may be involved in the onset and progression of IBD via immunity.
CREBBP may interact with *STAT1* in the development and progression of IBD.

1. Introduction

Inflammatory bowel disease (IBD), which comprises ulcerative colitis (UC) and Crohn disease (CD), is a chronic and idiopathic inflammatory disorder of the gastrointestinal tract.[1] CD may affect any part of the gastrointestinal tract, but mainly affects the distal ileum and the colon. In contrast, UC usually starts in the rectum and extends through part of or the entire colon.[2] Because of its early onset, irregular course of disease, unpredictable prognosis, and lack of a definitive cure, IBD imposes a considerable global public health burden.[3] In Asia, the incidence of IBD, especially UC, is increasing in parallel with rapid socioeconomic development and increasingly westernized lifestyle.[4,5] Therefore, it is essential to gain further insight into the pathogenesis of IBD, in order to develop effective drugs and therapies for the treatment of IBD.

Although the precise etiology of IBD is not yet fully understood, evidence has demonstrated that the etiology of this disease is multifactorial.[6] A widely accepted hypothesis is that both CD and UC result from the interaction of genetic and environmental factors including an overactive, poorly controlled mucosal immune response, and defects in mucosal barrier function.[7] Numerous studies have shown that the dynamic balance between microbes, especially commensal flora, and host defense responses at the mucosal barrier have a crucial role in the development and pathogenesis of chronic IBD.[6,8] Strober et al[9] reviewed the cytokine responses mediating intestinal inflammation in IBD and

Editor: Haseeb Zubair.

RS and YL regarded as first coauthors.

The authors have no funding and conflicts of interest to disclose.

Department of Gastroenterology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China.

* Correspondence: Feng Xu, Department of Gastroenterology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, No. 1 East Jianshe Road, Henan Province 450052, China (e-mail: xufengmd@163.com).

Medicine (2018) 97:21(e10658)

Received: 14 July 2017 / Accepted: 13 April 2018

http://dx.doi.org/10.1097/MD.0000000000010658

demonstrated that anticytokine agents, such as antitumor necrosis factor (TNF)-α therapy, are likely to be useful in the treatment of IBD. Genetic factors play an important role in IBD pathogenesis. Fujimoto et al[10] revealed that the UC-associated gene *RNF186* maintained gut homeostasis through controlling protein homeostasis in the colonic epithelia and may be a candidate marker for the diagnosis of UC. In addition, Majumdar et al[11] showed that altered expression of *TNFAIP3* was associated with disease severity in UC. Furthermore, a study recommended the use of p53 and p21 in the large bowel mucosa as valuable biomarkers of inflammatory-related carcinogenesis in UC.[12] However, the molecular mechanism leading to development and progression of IBD remains to be elucidated.

Olsen et al[13] reported a diagnosable preinflammatory state in patients diagnosed with UC and noted the usefulness of random forest modeling of genome-wide gene expression in distinguishing quiescent and active UC colonic mucosa versus CD colonic mucosa and controls. Another study using the same microarray data for gene selection revealed that changes in the expression of a select set of genes could serve as an early marker of UC.[14] In this study, we downloaded the same microarray data and reanalyzed it using different methods. Differentially expressed genes (DEGs) were identified by comparing the expression profiles in early-stage UC (EUC) and advanced-stage UC (AUC) with controls. We also performed pathway enrichment analysis on overlapping DEGs in EUC and AUC as well as AUC-specific DEGs, and the algorithm Random Walk with Restart (RWR) was applied to prioritize critical DEGs in UC based on a protein-to-protein interaction (PPI) network and the IBD pathway downloaded from a public database. Immunohistochemical (IHC) analysis was used to detect expression of IL17 and transforming growth factor beta (TGF-β) in colonic tissues from UC patients and normal colonic mucosa from healthy volunteers. We aimed to identify critical genes involved in UC and to gain insight into the molecular mechanisms responsible for UC onset and progression. Understanding the molecular mechanisms of UC will be an important step in gaining further insight into its pathogenesis and guiding future in-depth functional studies as well as treatment of UC.

2. Materials and methods

2.1. Microarray data

The gene expression dataset GSE9452[13] was obtained from the NCBI Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/), which is based on a GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array platform. A total of 26 samples are included in this dataset,[13] including 13 UC samples without macroscopic signs of inflammation defined as the EUC group, 8 UC samples with macroscopic signs of inflammation defined as the AUC group, and 5 non-IBD control subjects.

2.2. Data preprocessing and screening of DEGs

Data preprocessing was performed using robust multiarray analysis[15] implemented in R. DEGs in the EUC and AUC groups were identified by comparison with the normal controls using the Linear Models for Microarray data (Limma) package.[16] The threshold for the identification of DEGs was set at $|\log_2$ fold change $| \geq 0.5$ and a P-value $< .01$.

2.3. Pathway enrichment analysis of DEGs

Overlapping DEGs in the EUC and AUC groups as well as AUC-specific DEGs were selected for pathway enrichment analysis. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was implemented for the overlapping DEGs and the AUC-specific DEGs using the online tool, the Database for Annotation, Visualization, and Integrated Discovery.[17] The threshold for significance was set at $P < .01$. KEGG network analysis was performed with the plugin Enrichment Map (http://baderlab.org/Software/EnrichmentMap), a Cytoscape[18] plugin for functional enrichment visualization.

2.4. PPI network construction

The online database resource Search Tool for the Retrieval of Interacting Genes (STRING) can provide comprehensive coverage and ease of access to both predicted and experimental PPI data.[19] Moreover, interactions identified in STRING include a confidence score, allowing calibration of accuracy and coverage.[19] Genes with documented interactions with the DEGs were extracted from the STRING database. Subsequently, the PPI network was constructed based on DEG–DEG, gene–gene, and gene–DEG interactions. The interaction pairs with a PPI combined score >0.7 were selected for subsequent construction of a new network.

2.5. Random walk with restart to prioritize DEGs

A random walk is defined as an iterative walker's transition from a single node to a randomly chosen neighbor starting at a given source node, s.[20] In this study, we applied the RWR algorithm in the dent package,[21] which calculated a priority score for every node based on the steady state probability. The equation for RWR is defined as:

$$p^{t+1} = (1 - r) W p^t + r p^0$$

where r is the restart probability, W the column-normalized adjacency matrix of the graph, and p^t is the vector of size equivalent to the number of nodes in this graph where the ith element holds the probability of being at node i at time step t. In the present study, RWR analysis was performed using a custom R program in the total DEG PPI network with 48 IBD genes set as the seed node. These 48 IBD genes were extracted from the IBD pathway (pathway ID, hsa05321) which was downloaded from the KEGG pathway database.[22] We set the restart probability parameter at 0.7, as recommended by Wu et al.[23] DEGs were ranked according to the values (probability scores) in the steady-state probability vector. The probability scores of DEGs were $\log_{10}$ transformed and regarded as node attributes and displayed by Cytoscape.[19] Higher scores indicated the nodes were more closely connected with IBD genes. The top 100 genes with the highest scores were used to construct the PPI subnetwork.

2.6. Tissue collection and IHC detection

Colonic mucosal biopsies were obtained from 4 patients diagnosed with UC who visited the Department of Gastroenterology at the First Affiliated Hospital of Zhengzhou University, along with normal colonic mucosa from 6 healthy adult volunteers. All experimental procedures were authorized by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University and performed in accordance with the appropriate ethical standards. Written informed consent was obtained from all patients prior to obtaining samples.

Paraffin-embedded specimens were sectioned at 6-μm thickness, and then used for IHC analysis. Samples were dewaxed in

xylene and dehydrated in ethanol. Sections were sequentially subjected to antigen retrieval in citrate buffer and heated in a microwave oven for 13 minutes and then cooled at room temperature. The endogenous peroxidase activity was quenched with 3% H_2O_2 for 10 minutes and the samples were washed 3 times in PBS (3 minutes each wash). The samples were flooded with 10% bovine serum albumin (BSA) for 60 minutes at room temperature to reduce background staining. After washing, sections were incubated with the rabbit-derived monoclonal antibody anti-IL17 or anti-TGF-β (1:100 dilution, Abgent, San Diego, CA) overnight at 4 °C. Samples were washed 3 times in PBS (3 minutes each wash) and incubated with antirabbit IgG biotinylated secondary antibody for 60 minutes at 37 °C. Samples were washed 3 times with PBS (3 minutes each wash). Antibody staining was visualized in 3,3′-diaminobenzidine tetrahydrochloride (DAB) according to the manufacturer's instructions. Counterstaining was performed with hematoxylin. The sections were dehydrated through a graded alcohol series for 30 seconds, cleared with xylene, and then mounted with neutral balsam. Images at 200× magnification were obtained, and the integrated optical density (IOD) was determined by Image Pro Plus (IPP) software to quantify protein expression. Mean density (IOD/area) was calculated.

2.7. Statistical analysis

The data were analyzed using SPSS 17.0 software (SPSS Inc., Chicago, IL). The Student *t* test was used to assess the significance of differences among groups. A *P*-value of $<.05$ was defined as statistically significant.

3. Results

3.1. DEG screening

As shown in Fig. 1, 3511 DEGs were identified in AUC compared with controls, including 1512 upregulated genes and 1999 downregulated genes. Comparison of EUC with healthy controls identified 911 DEGs, including 433 upregulated genes and 478 downregulated genes. Most of the DEGs identified in EUC were also differentially expressed in AUC. We found that 93.3% (404/433) of genes upregulated in EUC were also upregulated in AUC and 95.8% (458/478) of genes downregulated in EUC were also downregulated in AUC.

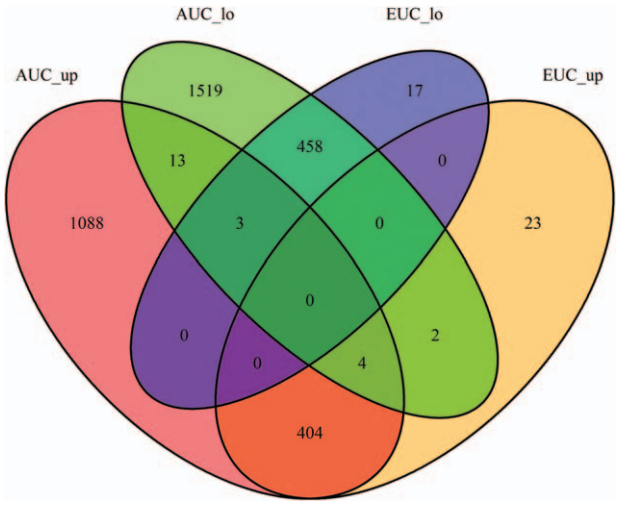

Figure 1. DEGs. Red and orange represent genes upregulated in AUC and EUC, respectively. Green and blue represent genes downregulated in AUC and EUC, respectively. AUC = advanced-stage ulcerative colitis, DEG = differentially expressed gene, EUC = early-stage ulcerative colitis.

3.2. KEGG enrichment analysis

The results of KEGG enrichment analysis of the overlapping DEGs and the AUC-specific DEGs using Database for Annotation, Visualization and Integrated Discovery are shown in Fig. 2. The overlapping DEGs and the AUC-specific DEGs were both enriched in the pathways related to immunity. For example, the overlapping DEGs were related to the intestinal immune network for IgA production. The AUC-specific DEGs were enriched in cytokine–cytokine receptor interactions, natural killer (NK) cell mediated cytotoxicity, and the B cell receptor signaling pathway. Furthermore, the AUC-specific DEGs were enriched in the pathways related to cell migration, such as extracellular matrix (ECM)–receptor interaction, and focal adhesion. The overlapping DEGs and the AUC-specific DEGs were simultaneously enriched in cell adhesion molecules (CAMs) and antigen processing and presentation pathways.

3.3. DEG prioritization and IBD pathways

The probability scores of DEGs ranged from −Inf to −1.79 after $\log_{10}$ transformation. Higher scores indicate closer connection with IBD genes; the DEGs with the top 15 scores were connected to the IBD pathway, as shown in Fig. 3.[24] All 15 genes were found to be differentially expressed in AUC, and 4 of the 15 were also differentially expressed in EUC, including signal transducer and activator of transcription 4 (*STAT4*), *STAT1*, toll-like receptor 4 (*TLR4*), and RAR-related orphan receptor C (*RORC*). Most of the 15 genes could be found in CD branches. Furthermore, these genes were linked with branches related to increasing the frequency of Th17 and regulatory T (Treg) cells, such as the upregulated gene *TGFB*.

3.4. PPI subnetwork of DEGs

In order to gain more insight into key genes in UC, we performed a full screen of the top 100 genes with higher probability scores and other DEGs in the PPI subnetwork to better understand their functions. As shown in Fig. 4, the subnetwork had 665 nodes and 1892 edges. Furthermore, the top 10 DEGs with higher node degrees were extracted and are presented in Table 1, for example, phosphoinositide-3-kinase, regulatory subunit 1 (*PIK3R1*) (degree = 104), CREB binding protein (*CREBBP*) (degree = 92), *STAT1* (degree = 87), CD44 molecule (*CD44*) (degree = 81), and Janus kinase 1 (*JAK1*) (degree = 80).

3.5. IHC detection analysis

The mean age of the 4 UC patients was 48 years (range, 20–60 years) and the ratio of males to females was 1:1. The results of IHC analysis of IL17 and TGF-β expression levels are shown in Fig. 5A and B. Positive expression of TGF-β protein was detected in IBD samples and its expression was rare in normal controls. High levels of expression of IL17 were identified in IBD samples. TGF-β expression was comparatively higher in disease samples than in healthy samples (Fig. 5B). Overall, the expression levels of

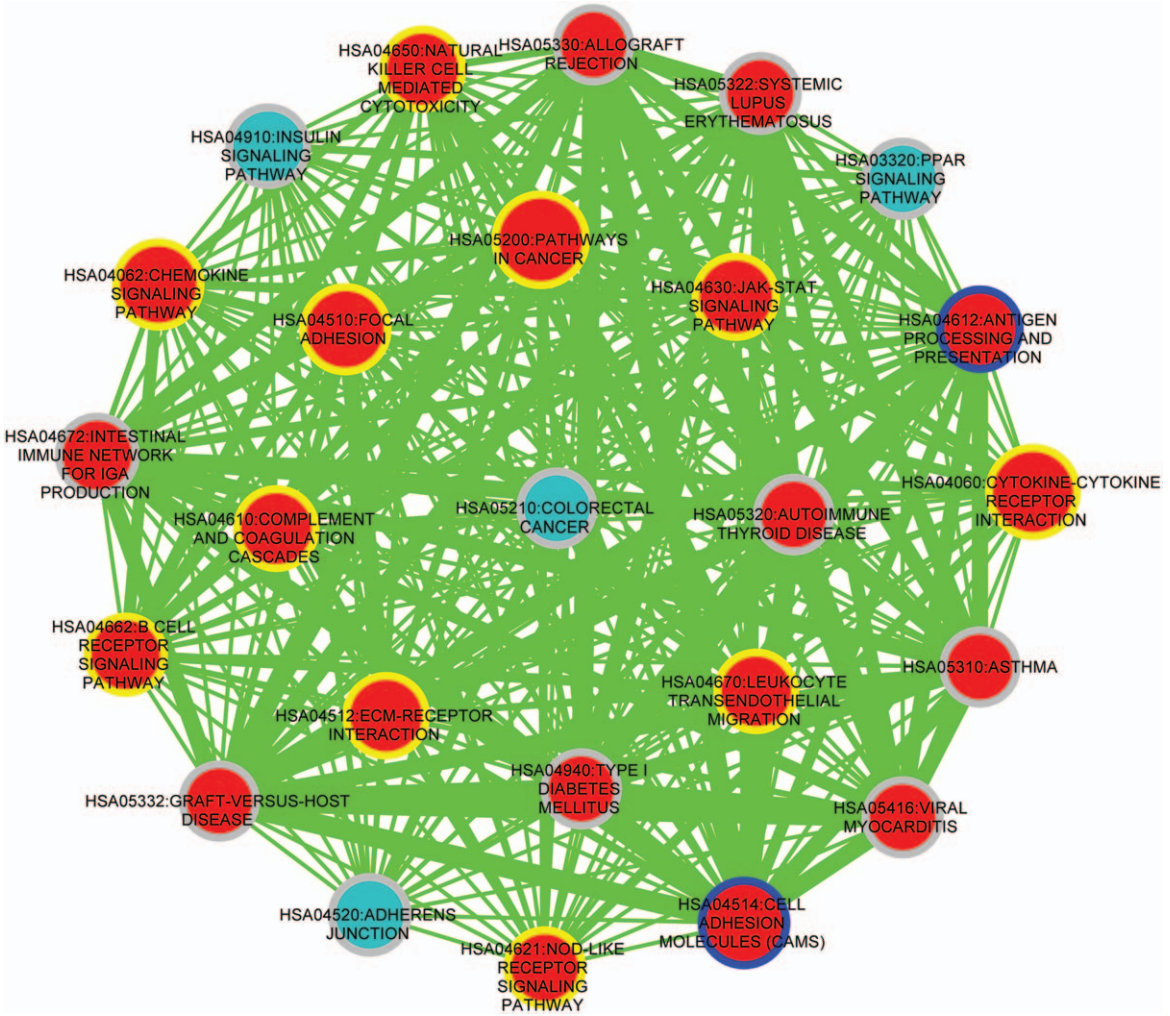

Figure 2. Pathway enrichment analysis of DEGs. Red nodes indicate the pathways enriched in upregulated genes. Blue nodes indicate the pathways enriched in downregulated genes. Grey borders represent the pathways enriched in overlapping DEGs in AUC and EUC. Yellow borders represent the pathways enriched in AUC-specific DEGs. Dark blue borders represent the pathways shared by 2 groups of enrichment results. An edge between 2 pathways indicates that they share the same genes, and the wider lines indicate larger overlaps. AUC = advanced-stage ulcerative colitis, DEG = differentially expressed gene, EUC = early-stage ulcerative colitis.

IL17 and TGF-β in the IBD specimens were found to be upregulated compared with normal controls ($P < .05$).

4. Discussion

Gene expression profiling in disease can uncover the underlying gene changes contributing to the disease and promote the identification of targets for therapeutic intervention. In the current study, we analyzed the DEGs using gene expression profiles for EUC and AUC in comparison with healthy controls. The results showed that 93.3% (404/433) of the genes upregulated in EUC were also upregulated in AUC and 95.8% (458/478) of the genes downregulated in EUC were also downregulated in AUC. The overlapping DEGs in AUC and EUC and the AUC-specific DEGs were both enriched in pathways related to immunity, such as NK cell mediated cytotoxicity, and antigen processing and presentation. In addition, the AUC-specific DEGs were also enriched in pathways related to cell migration, such as ECM-receptor interaction. Following DEG prioritization, we found that the DEGs *TLR4*, *STAT4*, *STAT1*, and *RORC* were linked with EUC, AUC, and CD. Moreover, the upregulated gene *TGFB* was involved in increasing the frequency of Th17 cells, as verified by IHC analysis. Furthermore, *PIK3R1*, *CREBBP*, and *STAT1* were part of higher-degree nodes in the PPI subnetwork.

Available evidence suggests that the adaptive immune response plays a major role in the pathogenesis of IBD.[25,26] An abnormal immune response against the microorganisms of the intestinal flora is considered to be responsible for IBD in genetically susceptible individuals.[6] Not only T helper cell type (Th) 1 and Th2 immune responses, but also other subsets of T cells, namely Th17 and Treg cells, are likely to play an important role in IBD.[25,27] Yu et al[28] demonstrated that NK cells, collectively referred to as innate lymphoid cells, provide a crucial link between the innate and adaptive immune responses to infection. Additionally, the work of Bär et al[29] demonstrated that CD and UC appear to modulate the major histocompatibility complex class I- and II-related pathways for the presentation of exogenous antigens in intestinal epithelial cells. In the present study, we found that the overlapping DEGs and the AUC-specific DEGs were both enriched in pathways related to immunity, such as antigen processing and presentation pathways. In addition,

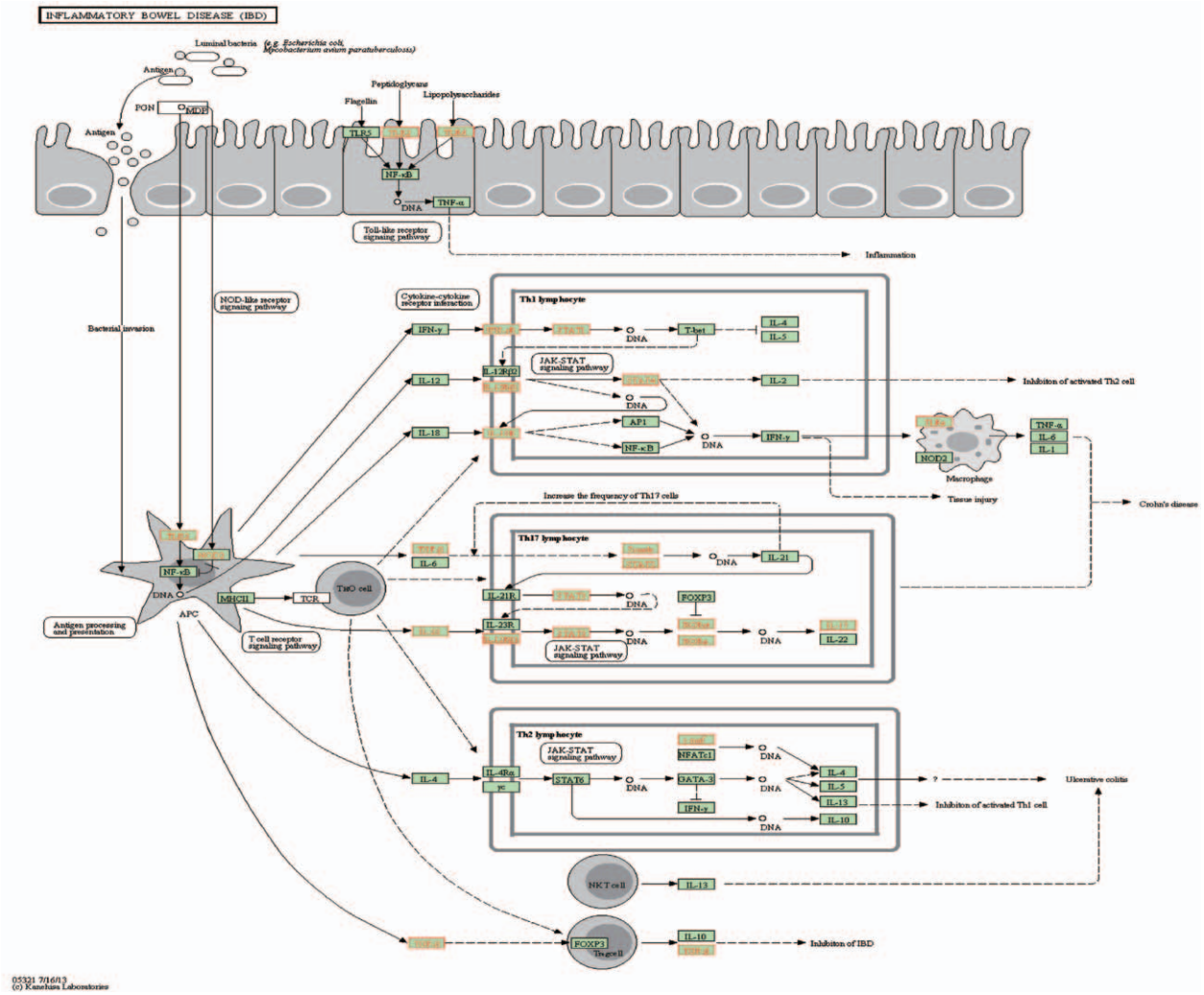

Figure 3. IBD pathway, red nodes indicate genes differentially expressed in AUC or EUC. AUC = advanced-stage ulcerative colitis, EUC = early-stage ulcerative colitis, IBD = inflammatory bowel disease.

AUC-specific DEGs were related to the NK cell-mediated cytotoxicity pathway, in accord with previous studies. However, Tanida et al[30] had reported that mucosal addressin CAM-1 plays an essential role in IBD development by mediating the migration as well as accumulation of lymphocytes into the gut interstitium, causing mucosal injury. A study indicated ECM remodeling as a key event and an active participant in sustaining intestinal inflammation in IBD.[31] Evidence shows that the involvement of adhesion molecules play an important role in the adherence of lymphocytes and macrophages to endothelial cells maintaining the chronic inflammation in UC patients.[32] In our study, both the overlapping DEGs and the AUC-specific DEGs were enriched in CAMs. The AUC-specific DEGs were enriched in pathways related to cell migration, such as ECM–receptor interaction and focal adhesion. In this context, our observations indicate that immunity and focal adhesion and ECM–receptor interaction are key components of UC which may increase in the AUCs, maintaining inflammation and tissue damage.

The recent discovery and characterization of Th17 cells and their signature cytokines (IL-17) represents a hallmark in T-cell immunobiology by providing a novel pathway connecting adaptive and innate immunity.[25] High transcript levels of IL-17A have been found in both CD and UC mucosa, compared with healthy mucosa.[33,34] In addition, Hundorfean et al[35] had revealed that IL17 is pathogenic in IBD (especially in CD) and that cytokines can directly contribute to long-term intestinal inflammation through complex mechanisms, such as proinflammatory effects on the IEC barrier. However, the TGFB family of cytokines are multifunctional proteins that control proliferation, differentiation, adhesion, and other functions in many cell types.[36] Treg cells are defined as T cells able to suppress the proliferation of Th0 cells which can differentiate into Th1, Th2, or Th17 cells. Treg cells can exert their function via producing the antiinflammatory cytokines TGF-β and IL-10, preventing both the activation and function of T cells that have escaped other mechanisms.[25] Marafini et al[37] demonstrated that TGF-β1, which can inhibit pathogenic responses in the gut, is highly expressed in both CD and UC. In accordance with these findings, our study clearly showed that the upregulated gene *TGFB* was linked with branches related to increasing the frequency of Th17 and Treg cells. Furthermore, the expression levels of IL17 and TGF-β were found to be upregulated in UC specimens compared with healthy specimens, by IHC analysis. A recent study showed that the joint expression of IL-17A, IL-17F, IL-21, RORC, and

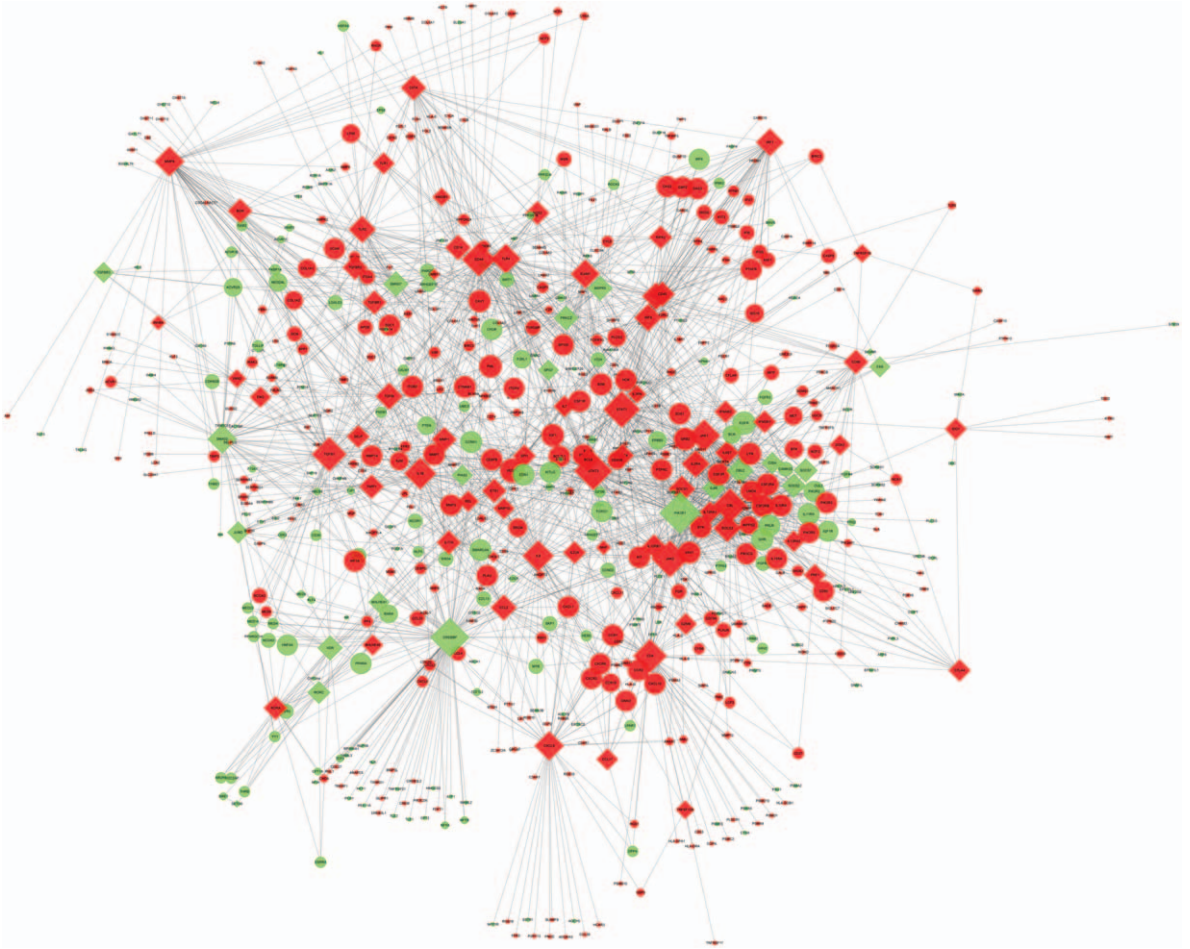

Figure 4. PPI subnetwork, genes upregulated in AUC or EUC are shown in red and genes downregulated in AUC or EUC are shown in green. Diamonds indicate the top 100 genes with higher probability scores. The node size shows the node degree. AUC = advanced-stage ulcerative colitis, EUC = early-stage ulcerative colitis, PPI = protein-to-protein interaction.

TGF-β was significantly predictive of Rachmilewitz endoscopic index in UC patients.[38] IL17 is increased in active UC and serum IL-17A levels of treatment-naive patients with UC is shown to reflect clinical disease severity at the onset of the disease.[39,40] Additionally, Feng et al[41] reported that IL-17 increased accompany with TGF-β in UC patients. On this basis, we postulate that the upregulated gene TGF-β may regulate the function of Th17 cells, affecting the expression of in the progression of IBD. IL-17 and TGF-β levels may be valuable in the clinical management of patients with UC; the underlying mechanisms warrant further investigation using larger sample sizes.

Table 1

The top 10 genes with higher degree in PPI subnetwork.

Gene	Degree	AUC_DEstate	EUC_DEstate
PIK3R1	104	−1	−1
CREBBP	92	−1	−1
STAT3	87	1	0
JAK2	81	1	0
STAT1	80	1	1
TGFB1	70	1	0
CBL	70	1	0
CD44	67	1	1
CD4	65	1	0
JAK1	60	1	1

AUC_DEstate and EUC_DEstate represent the gene differential expression state in AUC and EUC samples, respectively; 1 represents for upregulated, −1 for downregulated, and 0 for no-change. AUC = advanced-stage ulcerative colitis, EUC = early-stage ulcerative colitis, PPI = protein-to-protein interaction.

PIK3R1 can phosphorylate the inositol ring of phosphatidylinositol and is responsible for regulating a diverse range of cellular functions, such as proliferation and migration.[42] Cahill et al[43] demonstrated that the important isoform PI3-Kγ plays a crucial role in intestinal inflammation and dysregulation of the PI3K signaling pathway in innate and adaptive immune cells as well as in the intestinal epithelium, perhaps leading to inflammatory disorders including IBD. Moreover, in the current study, *PIK3R1* had the highest node degree in the PPI subnetwork. In this context, we suggest that *PIK3R1* might play a critical role in the development and progression of UC via immune response involvement.

In addition, CREBBP has histone acetyltransferase activity and is involved in transcriptional coactivation. STAT1 is a signal transducer and transcriptional activator which mediates cellular responses to interferons and other cytokines. Zhang et al[44] showed that the promoter-bound STAT1 dimer could interact with the histone acetyltransferase CREBBP and that acetylation

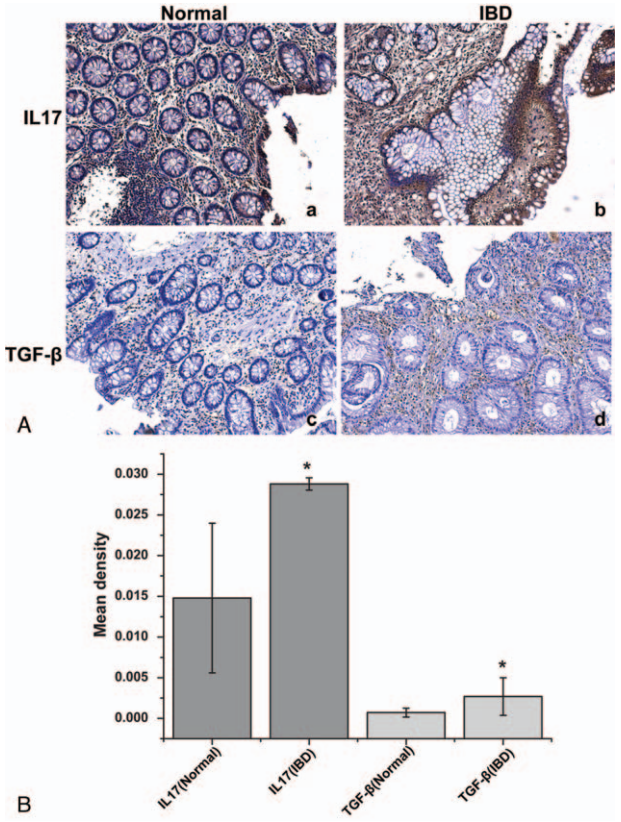

Figure 5. Immunohistochemical analysis of IL17 and TGF-β protein expression levels in IBD colonic tissues and normal colonic mucosa. Representative images are shown (200×). (A) a, IL17 expression in normal control; b, IL17 expression in IBD sample; c, TGF-β expression in normal control; and d, TGF-β expression in IBD sample; (B) the mean density of IL17 and TGF-β protein expression in IBD and normal samples. IBD = inflammatory bowel disease, IL = interleukin, TGF-β = transforming growth factor beta.

of histones is crucial for STAT1-mediated transcriptional activation. Moreover, Neurath showed that several IBD susceptibility loci contain genes that encode proteins involved in cytokine induction, chemokine receptor signaling, and Th cell responses including STAT1.[45] Furthermore, in our study, we found that *STAT1* expression was altered in both EUC and AUC, as well as CD, and *CREBBP* had a higher node degree in the PPI subnetwork. On this basis, we suggest that *CREBBP* might play an essential role in the onset and progression of IBD by interacting with *STAT1*.

This study has some limitations. Notably, the number of samples included in the microarray data was relatively small. Additionally, data from only 4 UC cases were used for validation of the expression of IL17 and TGF-β, which reduced the power of this study. Further studies using a larger sample size are required. Moreover, the molecular functions of other potentially significant genes were analyzed mostly using bioinformatic methods. Thus, the biological roles of these genes should be investigated and determined by further experimental studies.

In conclusion, we performed gene expression profile analysis to identify critical genes in IBD. Immunity and cell movement appear to be key components of UC and cell migration ability may increase in the AUC, causing lesion metastases. The upregulated gene *TGFB* may regulate IL17 expression in UC. *PIK3R1* may be involved in the onset and progression of UC by participating in immunity. *CREBBP* may interact with *STAT1* in the development and progression of IBD. However, extensive further work is necessary. Advances in understanding the molecular mechanism of IBD will aid in the development of novel targeted therapies with greater efficacy.

Acknowledgements

This work was supported by the Key Project of Scientific Research Foundation of the Education Department of Henan Province for Colleges and Universities in Henan Province (Program No. 16A320081).

Author contributions

Conceptualization: Ya Li.
Data curation: Weiwei Hao.
Investigation: Bingxue Wang, Lei Yang.
Writing – original draft: Ruifeng Song.
Writing – review & editing: Feng Xu.

References

[1] Cleynen I, Boucher G, Jostins L, et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. Lancet (Lond, Engl) 2016;387:156–67.
[2] Zhang Y-Z, Li Y-Y. Inflammatory bowel disease: pathogenesis. World J Gastroenterol 2014;20:91.
[3] Rencz F, Pentek M, Bortlik M, et al. Biological therapy in inflammatory bowel diseases: access in Central and Eastern Europe. World J Gastroenterol 2015;21:1728–37.
[4] Zhu H, Wan X, Li J, et al. Computational prediction and validation of BAHD1 as a novel molecule for ulcerative colitis. Sci Rep 2015;5:12227.
[5] Ng SC. Epidemiology of inflammatory bowel disease: focus on Asia. Best Pract Res Clin Gastroenterol 2014;28:363–72.
[6] Xavier R, Podolsky D. Unravelling the pathogenesis of inflammatory bowel disease. Nature 2007;448:427–34.
[7] Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012;491:119–24.
[8] Tlaskalová-Hogenová H, Štěpánková R, Kozáková H, et al. The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases. Cell Mol Immunol 2011;8:110–20.
[9] Strober W, Fuss IJ. Proinflammatory cytokines in the pathogenesis of inflammatory bowel diseases. Gastroenterology 2011;140:1756–67. e1751.
[10] Fujimoto K, Kinoshita M, Tanaka H, et al. Regulation of intestinal homeostasis by the ulcerative colitis-associated gene RNF186. Mucosal Immunol 2017;10:446–59.
[11] Majumdar I, Ahuja V, Paul J. Altered expression of tumor necrosis factor alpha-induced protein 3 correlates with disease severity in ulcerative colitis. Sci Rep 2017;7:9420.
[12] Popp C, Nichita L, Voiosu T, et al. Expression profile of p53 and p21 in large bowel mucosa as biomarkers of inflammatory-related carcinogenesis in ulcerative colitis. Dis Markers 2016;2016:3625279.
[13] Olsen J, Gerds TA, Seidelin JB, et al. Diagnosis of ulcerative colitis before onset of inflammation by multivariate modeling of genome-wide gene expression data. Inflamm Bowel Dis 2009;15:1032–8.
[14] Roman J, Planell N, Lozano JJ, et al. Evaluation of responsive gene expression as a sensitive and specific biomarker in patients with ulcerative colitis. Inflamm Bowel Dis 2013;19:221–9.
[15] Bolstad BM, Irizarry RA, Åstrand M, et al. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003;19:185–93.
[16] Smyth GK. Limma: linear models for microarray data. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer, New York; 2005: 397–420.
[17] Da Wei Huang BTS, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2008;4:44–57.

[18] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.
[19] Szklarczyk D, Franceschini A, Kuhn M, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011;39(Database issue):D561–568.
[20] Can T, Çamoğlu O, Singh AK. Analysis of protein-protein interaction networks using randomwalks. Paper presented at: Proceedings of the 5th international workshop on Bioinformatics, Chicago, 2005.
[21] Fang H, Gough J. The 'dnet'approach promotes emerging research on cancer patient survival. Genome Med 2014;6:1–6.
[22] Kanehisa M, Goto S, Sato Y, et al. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 2014;42(D1): D199–205.
[23] Wu B, Xie J, Du Z, et al. PPI network analysis of mRNA expression profile of ezrin knockdown in esophageal squamous cell carcinoma. Biomed Res Int 2014;2014:651954.
[24] Kanehisa M, Sato Y, Kawashima M, et al. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 2016;44(D1):D457–462.
[25] Geremia A, Biancheri P, Allan P, et al. Innate and adaptive immunity in inflammatory bowel disease. Autoimmun Rev 2014;13:3–10.
[26] Foersch S, Waldner MJ, Neurath MF. Innate and adaptive immunity in inflammatory bowel diseases. Dig Dis 2012;31:317–20.
[27] Zhou L, Ivanov II, Spolski R, et al. IL-6 programs TH-17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways. Nat Immunol 2007;8:967–74.
[28] Yu J, Freud AG, Caligiuri MA. Location and cellular stages of natural killer cell development. Trends Immunol 2013;34:573–82.
[29] Bär F, Sina C, Hundorfean G, et al. Inflammatory bowel diseases influence major histocompatibility complex class I (MHC I) and II compartments in intestinal epithelial cells. Clin Exp Immunol 2013; 172:280–9.
[30] Tanida S, Mizoshita T, Mizushima T, et al. Involvement of oxidative stress and mucosal addressin cell adhesion molecule-1 (MAdCAM-1) in inflammatory bowel disease. J Clin Biochem Nutr 2011;48:112.
[31] Shimshoni E, Yablecovitch D, Baram L, et al. ECM remodelling in IBD: innocent bystander or partner in crime? The emerging role of extracellular molecular events in sustaining intestinal inflammation. Gut 2015;64:367–72.
[32] Gulubova MV, Manolova IM, Vlaykova TI, et al. Adhesion molecules in chronic ulcerative colitis. Int J Colorectal Dis 2007;22:581–9.
[33] Sugihara T, Kobori A, Imaeda H, et al. The increased mucosal mRNA expressions of complement C3 and interleukin-17 in inflammatory bowel disease. Clin Exp Immunol 2010;160:386–93.
[34] Kobayashi T, Okamoto S, Hisamatsu T, et al. IL23 differentially regulates the Th1/Th17 balance in ulcerative colitis and Crohn's disease. Gut 2008;57:1682–9.
[35] Hundorfean G, Neurath MF, Mudter J. Functional relevance of T helper 17 (Th17) cells and the IL-17 cytokine family in inflammatory bowel disease. Inflamm Bowel Dis 2012;18:180–6.
[36] Cunha SI, Pietras K. TGF-β Signaling in Physiological and Pathological Angiogenesis. TGF-β in Human Disease: Springer, Tokyo; 2013: 323–348.
[37] Marafini I, Zorzi F, Codazza S, et al. TGF-beta signaling manipulation as potential therapy for IBD. Curr Drug Targets 2013;14:1400–4.
[38] Iboshi Y, Nakamura K, Fukaura K, et al. Increased IL-17A/IL-17F expression ratio represents the key mucosal T helper/regulatory cell-related gene signature paralleling disease activity in ulcerative colitis. J Gastroenterol 2017;52:315–26.
[39] D'Ambrosio A, Cossu A, Amendola A, et al. Lamina propria CD4+LAP+ regulatory T cells are increased in active ulcerative colitis but show increased IL-17 expression and reduced suppressor activity. J Crohns Colitis 2016;10:346–53.
[40] Ohman L, Dahlen R, Isaksson S, et al. Serum IL-17A in newly diagnosed treatment-naive patients with ulcerative colitis reflects clinical disease severity and predicts the course of disease. Inflamm Bowel Dis 2013;19:2433–9.
[41] Feng JS, Yang Z, Zhu YZ, et al. Serum IL-17 and IL-6 increased accompany with TGF-beta and IL-13 respectively in ulcerative colitis patients. Int J Clin Exp Med 2014;7:5498–504.
[42] Quayle SN, Lee JY, Cheung LWT, et al. Somatic mutations of PIK3R1 promote gliomagenesis. PloS One 2012;7:e49466.
[43] Cahill CM, Rogers JT, Walker WA. The role of phosphoinositide 3-kinase signaling in intestinal inflammation. J Signal Transduct 2012; 2012:358476.
[44] Zhang JJ, Vinkemeier U, Gu W, et al. Two contact regions between Stat1 and CBP/p300 in interferon gamma signaling. Proc Natl Acad Sci U S A 1996;93:15092–6.
[45] Neurath MF. Cytokines in inflammatory bowel disease. Nat Rev Immunol 2014;14:329–42.