Searching for Key Genes in Ulcerative Colitis: a Bioinformatics Study

Shuangli Peng^{1,*}, Hongci Chen^{1,2,} *,Xingyu Yu¹, Jinjie Li¹,Yamei Ge¹,Xiong Zhang^{1,2},Dengqin Song¹,Huiyun Pu¹ and Ping Yang¹,

¹Hubei University of traditional Chinese Medicine, Wuhan 430000, China

²Hubei Provincial Hospital of Traditional Chinese Medicine, Wuhan 430000, China

*981078456@qq.com

Abstract. Objective: The incidence of ulcerative colitis (UC) is increasing year by year, which leads to increasing cost of medical resources investment, and the exact pathogenesis of the disease is still unclear, leading to difficulties in clinical diagnosis and treatment. This study is based on a bioinformatics approach to search for possible key genes leading to the development of this disease, hoping to provide clinicians with new biomarkers that can be used for reference, so that patients can receive timely diagnosis and treatment.Methods: Peripheral blood-related information was first obtained from the Gene Expression Omnibus (GEO) database for UC patients and non-UC populations. Then differentially expressed genes were identified using R language (based on "Limma" software). Functional enrichment analysis was then performed, followed by screening of central genes, and finally correlations between the central genes were investigated using Spearman correlation analysis.Results: Fifty-nine up-regulated and 39 down-regulated DEGs were screened. These genes were involved in pathways such as viral protein-cytokine receptor interaction, cytokine-cytokine receptor interaction, biosynthesis of pantothenic acid and coenzyme A, IL-17 signaling pathway and chemokine signaling pathway. And growth regulatory alpha protein, C-X-C chemokine receptor 1, C-C patterned chemokine 2, and trigger receptor 1 expressed in bone marrow cells were found to be the top ten hub genes. The final correlation analysis showed a positive correlation between these ten hub genes. Conclusion The present study identified 10 DEGs as possible biomarkers for the diagnosis of patients with ulcerative colitis. Experiments are needed to validate the present study.

Keywords: ulcerative colitis, bioinformatics, biomarkers, differentially expressed gene

1. Introduction

Among inflammatory bowel diseases, ulcerative colitis accounts for a relatively high proportion compared to other diseases, with lesions occurring in the colonic mucosa. It is a chronic idiopathic disease [1]. Epidemiological surveys have shown that the prevalence of UC is showing a yearly increase worldwide, which means that the medical costs for ulcerative colitis increase year by year [2]. Researchers have found that UC pathogenesis can be linked to immune dysfunction, environmental factors, genetic susceptibility, microbiome, and other factors. Over time, understanding has been deepened, but the precise mechanism of the disease has not been fully elucidated[3].

Non-surgical treatment and surgical treatment are two common methods currently used to treat UC. Pharmacological and non-pharmacological treatments fall under the scope of non-surgical treatment. Aminosalicylic acid preparations, immunosuppressants, glucocorticoids, etc. are commonly used drugs for the treatment of UC. Compared with other drugs, the clinical efficacy of immunosuppressants is more prominent. Such drugs are commonly used as hyperbaric oxygen therapy, anti-TNF- α monoclonal antibodies[4,5] and leukocyte adsorption therapy. Fecal bacterial transplantation is a commonly used non-pharmacological treatment[6-8]. It can be seen that the choice of treatment options for UC is diverse and has a certain curative effect. Therefore, a simpler and faster method should be found to identify and diagnose this disease in a timely manner, so that patients can receive treatment as soon as possible.

DOI: 10.56028/aetr.3.1.581

In recent years, studies have shown that the differences in the levels of fecal biomarkers and serum biomarkers between UC patients and non-UC patients are objective and have high research value. Serum biomarkers include B-cell activator, matrix metalloproteinase, α -1 antitrypsin, serum trefoil factor 3, advanced glycation end products, leucine-rich α -2 glycoprotein, soluble receptor, and soluble ST2 protein, and fecal biomarkers include fecal lactoferrin, fecal calprotectin, and fecal myeloperoxidase [9,10]. Symptoms are subjective, and the clinical symptoms of some patients are not obvious. These factors are not conducive to the diagnosis of diseases. Biomarkers can objectively reflect changes in the body and have a certain auxiliary role in disease diagnosis. If the ideal UC biomarkers can be found, it will help gastroenterologists diagnose this disease and provide greater help, which may lead to better clinical outcomes.

Therefore, we hope to discover new ideal markers for diagnosing UC through bioinformatics methods. In the present study, We used GSE3365 microarray data to look for differentially expressed genes (DEGs) in peripheral blood serum from UC patients and non-UC populations. Then, the main biological functions regulated by DEGs were explored by means of enrichment analysis. At the same time, in order to identify key genes related to UC diagnosis and treatment, many analytical methods such as correlation analysis, protein interaction analysis and modularity analysis were used in this study.

2. Materials and Methods

2.1 Gene Expression Profiling Data

The GSE3365 gene expression profile data was downloaded in its entirety from the GEO database, which was used to screen out DEGs in UC. The expression profile data included 26 non-UC samples and 42 UC samples.Patients with ulcerative colitis were then molecularly classified (based on peripheral blood mononuclear cell transcriptomes). Finally, Gene expression was accurately detected using the U133A array contained in the human genome (GPL96 [HG-U133A] Affymetrix).

2.2 Screening for Differentially Expressed Genes and Association Analysis

The data of the GSE3365 matrix are first normalized, then log2 transformed and finally the DEGs are selected. The above processing is implemented based on the "Limma" package [11]. For DEGs samples, the threshold level of statistical significance was set to log2FC greater than 1, p-value (after adjustment) was 0.05. To gain more insight into DEGs, we then performed correlation analysis (Spearman analysis) and visualization through the "ggplot2" and "ComplexHeatmap" R packages [12].

2.3 DEG Functional Enrichment Research

Further analytics of statistically significant Differentially expressed genes was done using the version 3.6.3 of the R language (based on the "clusterProfiler" and "ggplot2" packages), The Kyoto Gene and Genome Encyclopedia (KEGG) belongs to the common databases for bioinformatics analysis, and we used not only this database but also combined it with the Gene Ontology (GO) database in order to obtain a more complete functional pathway enrichment analysis..Analysis was terminated when the p-value (after adjustment) was <0.05.

2.4 Construction of Protein Interaction Network and Module Analysis

The STRING database (this study is based on version 11.5) was used to obtain proteins that interact directly or indirectly with DEGs. Meanwhile, Homo sapiens was selected as the study subject and the confidence score was set to be greater than 0.4. Finally, in order to efficiently find out the key modules in the protein interaction network, Cytoscape software (version 3.8.2) was used and then the modules that met the requirements were saved (score ≥ 4.0)

2.5 Expression Level Analysis of Hub Genes

To explore the differences in gene expression levels in non-UC populations of UC patients, we used the "ggplot2" package based on the R language in version 3.6.3.

3. Results

3.1 Results of Screening for Differentially Expressed Genes

Based on the results of the R software analysis, and a total of 98 differentially expressed genes were identified, of which 59 were up-regulated and 39 were down-regulated ($|\log 2FC|$ greater than 1, adjusted P value less than 0.05). This result is graphically illustrated in Figure 1A below, the red and blue panels represent up-and down-regulated genes, while the gray panels represent other genes. Figure 1B shows the expression levels of these genes well, and it can be seen that the genes in the UC group and the non-UC group are well clustered.

3.2 Correlational Research

Correlation analysis was performed using the top 10 up- or down-regulated genes(| log2FC | greater than 1, adjusted P value less than 0.05). Figure 1C shows the bitmap of differentially expressed genes (based on correlation analysis). Red indicates a positive correlation, green indicates a negative correlation. The darker the color, the higher the correlation between the differentially expressed genes. The correlation coefficient (ρ) is taken as the absolute value, and the magnitude of the absolute value is proportional to the level of correlation. 0-0.10 indicates a "negligible" correlation, An absolute value between 0.10 and 0.39 indicates the weakest correlation, while an absolute value between 0.90 and 1.00 indicates the strongest correlation. Second, the absolute value of medium correlation is 0.40-0.69, and the absolute value of strong correlation is 0.70-0.89[13].

Advances in Engineering Technology Research

ISCTA 2022 DOI: 10.56028/aetr.3.1.581



Figure 1. (A) Volcano map of differentially expressed genes in ulcerative colitis. Up-regulated genes are represented in red, down-regulated genes are represented in blue, and the remaining genes are represented in gray. (B) Heat map of differentially expressed genes in ulcerative colitis. Columns indicate data sets and rows indicate genes. Colors represent meanings as before. (C) Bitmap of correlation analysis between the top 10 key DEGs. Red indicates positive correlation, blue indicates negative correlation.

3.3 Functional Enrichment Analysis

GO and KEGG analysis were performed to gain a more in-depth look at the biological functions of DEGs. Molecular function (MF) refers to the activities of gene products at the molecular level, cell composition (CC) is a kind of anatomical structure of cells, which does not refer to any function, and biological process (BP) is a biological process completed by various

ISSN:2790-1688

DOI: 10.56028/aetr.3.1.581

molecular activities. These are the three parts of GO terms[14], as shown in Figures 2 and 3 below. Cellular component (CC) analysis suggests that significant enrichment of DEGs in blood particles, outside of the plasma membrane, secretory granule membranes, membrane-anchored components and tertiary granules. Molecular function (MF) analysis showed significant enrichment of the analyzed term for chemokine activity, and biological process (BP) analysis showed that DEGs are involved in the process of acute inflammation and cell migration, involving cells such as leukocytes and granulocytes. It also involves neutrophil degranulation. As shown in Figure 4 below, KEGG analysis indicated that many pathways were involved, such as interleukin-17 signaling pathway and chemokine signaling pathway, the interaction of viral proteins with cytokines and cytokine receptors, and the biosynthesis of pantothenic acid and coenzyme A.



Figure 2. GO enrichment analysis of DEGs (P-value<.01 and q-value<0.05). DEGs: differentially expressed genes; GO: Gene Ontology.





Figure 3. The top 5 enriched GO terms. GO: Gene Ontology.



3.4 Construct Protein-Protein Interaction Network and Conduct Module Analysis

PPI (protein-protein interaction) pairs were identified using the STRING database. After PPI network analysis, it is found that these DEGs can form 51 nodes, and 218 edges can be formed between these 51 nodes. The top 10 hub DEGs were then identified based on degree values. (shown in Figure 5A). Based on the analysis, the most important genes were CXCL1 and CCL2, which had a degree value of 26. the next most important genes were CXCR1 and FCGR3B, which had a degree value of 24, and the CXCR2, which had a degree value of 22. the results of these studies are clearly shown in the following Table 1.

Three important modules (scoring > 4.0) were also taken out of the PPI network. Six gene nodes with 28 edges were present in Module 1, including SELENBP1, SNCA, ALAS2, CA1, and SLC4A1 (Figure 5B). The four gene nodes and 12 edges that make up module 2 are interleukin-1 receptor type 2 (IL1R2), C-C pattern chemokine 2 (CCL2), interleukin-1 receptor type 1 (IL1R1), and C-X-C chemokine receptor type 2 (CXCR2). (See Figure 5C.) The seven gene nodes and 24 edges in module 3 are PTGS2, vascular non-inflammatory molecule 2 (VNN2), low affinity immunoglobulin gamma Fc region receptor III-B (FCGR3B), C-C pattern chemokine 7 (CCL7), CXCL1, CXCR1, and Aquaporin-9 (AQP9). (See Figure 5D.) It's noteworthy that Module 1 included just one hub DEGs of SLC4A1. Three hub DEGs for CCL2, CXCR2, and FPR1 were enriched in module 2. Three hub DEGs for CXCL1, CXCR1, and PTGS2 were enriched in module 3. However, no detectable modules of TREM1, IL1R1, or FCGR3B were found.

Advances in Engineering	Technology	Research
ISSN:2790-1688		

ISCTA 2022 DOI: 10.56028/aetr.3.1.581

Gene name	Protein name	Expression level	Degree	Enriched significant modules
CXCL1	Growth-regulated alpha protein	UP	26	module 3
CCL2	C-C motif chemokine 2	UP	26	module 2
CXCR1	C-X-C chemokine receptor type 1	UP	24	module 3
FCGR3B	Low affinity immunoglobulin gamma Fc region receptor III-B	UP	24	None
CXCR2	C-X-C chemokine receptor type 2	UP	22	module 2
PTGS2	Prostaglandin G/H synthase 2	UP	20	module 3
TREM1	Triggering receptor expressed on myeloid cells 1	UP	18	None
IL1R1	Interleukin-1 receptor type 1	UP	18	None
FPR1	fMet-Leu-Phe receptor	UP	16	module 2
SLC4A1	Band 3 anion transport protein	UP	14	module 1

Table 1 .The information of top 10 hub genes based on their degree value.



Figure 5. (A) PPI network diagram with up-regulated genes indicated in yellow and downregulated genes indicated in blue. (B) Module 1 analyzed 6 gene nodes and 28 edges with an MCODE (molecular complex detection) score of 5.6. (C) Module 2 analyzed 4 gene nodes and 12 edges, with an MCODE score of 4.0. (D) Module 3 analyzed 7 gene nodes as well as 24 edges, with an MCODE score of 4.0.

3.5 Expression Level Analysis of Hub Genes

Use the Cytoscape version 3.8.2 plug-in "cytoHubba" to construct an interactive network of 10 central DEGs, as shown in Figure 6A below. As shown in Figure 6B, the expression levels of FPR1, CCL2, CXCR2, CXCL1, CXCR1, SLC4A1, FCGR3B, IL1R1, TREM1 and PTGS2 in UC peripheral blood were significantly up-regulated compared to non-UC peripheral blood. Spearman

ISSN:2790-1688

DOI: 10.56028/aetr.3.1.581

correlation analysis was performed on the 10 hub DEGs, and the results showed that the expressions of 10 hub genes were positively correlated.



Figure 6. (A) Network diagram of the 10 hub genes. (B) Expression level analysis of the 10 hub genes. (C) Correlation analysis bitmap between the 10 hub genes. Positive and negative correlations are shown in red and blue, respectively

4. Discussion

In recent years, new breakthroughs have been made in the diagnosis and treatment of UC. Finding a simpler and faster diagnosis method can provide great help for the early identification of the disease, and early intervention in patients can also benefit patients more [15]. In this study, 98 differentially expressed genes (DEGs) in UC were successfully screened based on bioinformatics methods, of which 59 were up-regulated DEGs and 39 were down-regulated DEGs, and then 10 hub DEGs were screened. After that, we further used GO and KEGG analysis methods to analyze and explore the specific biological processes of these DEGs. DEGs are involved in the process of acute inflammation and cell migration, involving cells such as leukocytes and granulocytes. It also involves neutrophil degranulation. KEGG analysis indicated that the main related pathways were numerous,Such as interleukin-17 signaling pathway and chemokine signaling pathway, the interaction of viral proteins with cytokines and cytokine receptors, and the biosynthesis of pantothenic acid and coenzyme A.The results of PPI network analysis showed that targets such as FPR1, CCL2, CXCR2, CXCL1, CXCR1, SLC4A1, FCGR3B, IL1R1, TREM1 and PTGS2 belong

ISCTA 2022

ISSN:2790-1688

DOI: 10.56028/aetr.3.1.581

to the core nodes in the network, and it can be speculated that these genes are key genes in the pathogenesis of UC.

Interleukin 17 (IL-17) is a highly adaptable pro-inflammatory cytokine important for a number of functions, such as host defense, tissue healing, the pathogenesis of inflammatory diseases, and the development of cancer[16]. The IL-17 signaling pathway is enriched in three hub DEGs, namely CXCL1, PTGS2, and CCL2. Based on research with animals, it may be deduced that the CXCL gene target is somehow related to the development of ulcerative colitis in rats because the expression of CXC chemokines is higher in UC rats than it is in non-UC rats [17]. At the same time, CXCL1 It has also been shown to be a key chemokine for neutrophils induced by IL-17 [18]. Some studies have found that trigger receptor 1 (TREM1) expressed by myeloid cells is found to be significantly up-regulated in the gut of patients with inflammatory bowel disease. In addition, Chinese scholars have found that TREM-1 in human peripheral blood mononuclear cells. The expression level was significantly correlated with disease activity [19]. Based on experimental animal models, FPR1 deficiency was found to reduce the incidence of colitis, which means that the presence of FPR1 is correlated with the development of colitis in rats, and it has been found that we can consider FPR1 as a marker for the diagnosis of ulcerative colitis. [20]

Of course, this bioinformatics analysis also has some shortcomings. Firstly, the sample included in this study is small, this may lead to some influence on the completeness and accuracy of the results, secondly, the sample selected for this study was human peripheral blood, and the sensitivity of peripheral blood serum biomarkers is not as high as that of fecal markers, and finally, this study was not validated experimentally due to the limitations of the team's scientific conditions.

References

- [1] Ordás I, Eckmann L, Talamini M, et al. Ulcerative colitis. Lancet (London, England) 2012; 380: 1606-1619.
- [2] Du L, Ha C. Epidemiology and Pathogenesis of Ulcerative Colitis. Gastroenterol Clin N 2020; 49: 643-654.
- [3] Ryzhakov G, West NR, Franchini F, et al. Alpha kinase 1 controls intestinal inflammation by suppressing the IL-12/Th1 axis. Nat Commun 2018; 9: 3797.
- [4] Ungaro RC, Limketkai BN, Jensen CB, et al. Stopping 5-aminosalicylates in patients with ulcerative colitis starting biologic therapy does not increase the risk of adverse clinical outcomes: analysis of two nationwide population-based cohorts. Gut 2019; 68: 977-984.
- [5] Wehkamp J, Stange EF. Recent advances and emerging therapies in the non-surgical management of ulcerative. F1000Research 2018; 7.
- [6] Sood A, Mahajan R, Singh A, et al. Role of Faecal Microbiota Transplantation for Maintenance of Remission in Patients With Ulcerative Colitis: A Pilot Study. Journal of Crohn's & colitis 2019; 13: 1311-1317.
- [7] Yamasaki S, Sakata Y, Yoshida H, et al. Shorter Relapse-Free Period after Leukocyte Removal Therapy in Younger than Older. Digestion 2019; 100: 247-253.
- [8] Jairath VB. Hyperbaric Oxygen for Hospitalized patients with Ulcerative Colitis.. 2018; 113: 1432-1434.
- [9] Kourkoulis P, Kapizioni C, Michalopoulos G, et al. Novel potential biomarkers for the diagnosis and monitoring of patients with ulcerative colitis. Eur J Gastroen Hepat 2019; 31: 1173-1183.
- [10] Eisenstein M. Ulcerative colitis: towards remission.. 2018; 563: S33.
- [11] Smyth GK. limma: Linear Models for Microarray Data. Springer New York: New York, NY. 2005. 397-420.
- [12] Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016; 32: 2847-2849.
- [13] Schober P, Boer C, Schwarte LA. Correlation Coefficients. Anesthesia & Analgesia 2018; 126: 1763-1768.

ISSN:2790-1688

DOI: 10.56028/aetr.3.1.581

- [14] Di Lena P, Domeniconi G, Margara L, et al. GOTA: GO term annotation of biomedical literature. Bmc Bioinformatics 2015; 16: 346.
- [15] Kucharzik T, Koletzko S, Kannengiesser K, et al. Ulcerative Colitis-Diagnostic and Therapeutic Algorithms. Dtsch Arztebl Int 2020; 117: 564-574.
- [16] Li X, Bechara R, Zhao J, et al. IL-17 receptor-based signaling and implications for disease. Nat Immunol 2019; 20: 1594-1602.
- [17] Boshagh MA, Foroutan P, Moloudi MR, et al. ELR positive CXCL chemokines are highly expressed in an animal model of ulcerative colitis. Journal of inflammation research 2019; 12: 167-174.
- [18] Catar RA, Chen L, Cuff SM, et al. Control of neutrophil influx during peritonitis by transcriptional cross-regulation. The Journal of pathology 2020; 251: 175-186.
- [19] Zhang S, Luo J, Shen B, et al. Correlation between triggering receptor expressed on myeloid cells-1 and clinical disease activity in Chinese patients with ulcerative colitis. Int J Clin Exp Med 2015; 8: 2147-2155.
- [20] Manuc M, Ionescu EM, Milanesi E, et al. Molecular Signature of Persistent Histological Inflammation in Ulcerative Colitis with Mucosal Healing. Journal of gastrointestinal and liver diseases : JGLD 2020; 29: 159-166.