Comprehensive bioinformatics analyses of Crohn's disease

YI ZHOU1*, CHENG ZHAN2*, YINGYU HUANG1 and HONGCHUN LIU1

Departments of 1Gastroenterology and 2Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai 200032, P.R. China

Received January 26, 2016; Accepted January 17, 2017

DOI: 10.3892/mmr.2017.6250

Abstract. Crohn’s disease (CD) is a chronic, relapsing inflammatory disease with increasing incidence and prevalence worldwide. In previous years, the accumulation of microarray data has provided us an approach to obtain further insight into CD. In the present study, the microarray data of CD was comprehensively analyzed using multiple bioinformatics methods, and the pathobiological process of the disease was examined. Gene expression data from colon tissues of patients with CD were obtained from the Gene Expression Omnibus database; following which differentially expressed genes were identified between CD and control sample groups. Subsequently, Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analyses were performed to investigate which functions and pathways in which the differentially expressed genes enriched. TargetScan and miRDB databases were then used to predict which microRNAs (miRNAs) regulated the differentially expressed genes. As a result, a total of 432 differentially expressed genes, including 229 upregulated and 203 downregulated genes, including matrix metalloproteinase 3 and glutathione S-transferase α1, were identified in CD samples. These differentially expressed genes were significantly involved in regulation of the inflammatory response, innate immune response, cell migration, extracellular matrix organization, Janus kinase/signal transducers and activators of transcription signaling pathway, and cytokine-cytokine receptor interaction. The miRNA-gene network showed that miR-149-3p and miR-4447 regulated the most differentially expressed genes. These findings extend current understanding of the mechanisms underlying CD, and the differentially expressed genes and regulator miRNAs identified may be used as potential biomarkers and therapeutic targets for CD.

Introduction

Crohn's disease (CD) is a chronic and relapsing inflammatory disease, which can affect any part of the gastrointestinal tract. Over the last few decades, CD has become progressively common in several developed countries, affecting almost one in 200 individuals, and its incidence and prevalence are increasing rapidly in developing countries (1). Patients with CD usually present with abdominal pain and severe diarrhea, accompanied with fever and serious weight loss. The fluctuating course of CD readily leads to patients experiencing stress and other psychosocial problems, significantly affecting quality of life in terms of health. Despite the wide use of immunosuppressive and anti-tumor necrosis factor α (anti-TNFα) therapies, >50% of patients require surgery within 10 years of diagnosis (2). Therefore, additional investigations into the etiology and molecular pathogenesis are required to improve clinical management of this disease.

The rapid development of microarray technology has provided an innovative approach for examining the biological mechanism of diseases. Previous studies based on genome-wide expression analyses have found novel biomarkers of inflammatory bowel disease and several other diseases (3-5). As this high-throughput technology can assist in revealing the etiology and pathogenesis underlying the disease course, the present study screened differentially expressed genes between colon tissues of patients with CD and control samples with microarray data obtained from the Gene Expression Omnibus (GEO) database. The significant functions and signaling pathways in which these genes are enriched were then identified, and the microRNAs (miRNAs) regulating these genes were predicted. The aim of these investigations was to obtain an improved understanding of the pathogenesis, and provide novel diagnostic and therapeutic approaches for CD.

Materials and methods

Data resources and preprocessing. A total of seven microarray profiles of CD, which were constructed within the last 10 years, were obtained from the GEO (http://www.ncbi.nlm.nih.gov/geo/) database (accession nos. GSE6731, GSE9686, GSE10616, GSE20881, GSE26305, GSE36807 and GSE52746) (6). A total of 267 samples, including 116 normal control colon samples and 151 colon tissue samples from patients with CD without TNFα therapy were included for the analyses. Subsequently, the original data were converted into probe expression measurements. The average value was

* Correspondence to: Dr Hongchun Liu, Department of Gastroenterology, Zhongshan Hospital, Fudan University, 180 Fenglin Road, Xuhui, Shanghai 200032, P.R. China
E-mail: liuhch.aliyun.com

*Contributed equally

Key words: Crohn's disease, differentially expressed genes, Gene Ontology analysis, Kyoto Encyclopedia of Genes and Genomes analysis, microRNA-gene network
applied for probes matching one gene from the same profile, and an intersection of the seven profiles was then performed, which obtained 7,793 common genes. Normal distribution was used to standardize each profile, as reported previously (7).

**Screening of differentially expressed genes.** The differently expressed genes between the control and CD sample groups were identified as previously reported (8). In brief, a random variance model t-test was performed using SPSS software (version, 20.0; IBM SPSS, Armonk, NY, USA) to filter the differently expressed genes, following which the false discovery rate (FDR) and significance were calculated. Genes with P<0.05 and FDR<0.05 values were considered to be significantly different.

**Gene ontology (GO) analysis.** Using data from the GO database (http://www.geneontology.org/), GO analysis was performed to determine the function of the significantly enriched differentially expressed genes, as reported previously (9). For classifying the GO categories, the χ² test and Fisher's exact test were performed using SPSS software (version, 22.0; IBM SPSS). FDR<0.05 was used to correct for multiple comparisons, and the GO terms with P<0.05 and FDR<0.05 were considered significant. Subsequently, enrichment values were calculated to select significant GO terms with the most concrete description of function. A GO network was then formed using Cytoscape v3.2.0 (http://cytoscape.org/) to summarize associated interactions on a GO map.

**Pathway enrichment analysis.** To investigate the significant pathways in which the differentially expressed genes were enriched, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/) was used to perform a KEGG analysis (10). As above, the χ² test and Fisher's exact test were used to identify significant pathways. In addition, the FDR, P-value and enrichment values were calculated. P<0.05 and FDR<0.05 were regarded as the cutoff for selecting significantly enriched pathways. Finally, Cytoscape v3.2.0 was used to establish a path-net to outline the association among these significant pathways.

**miRNA-gene network.** To investigate the regulatory miRNAs of these differentially expressed genes, the miRNAs were predicted using TargetScan (http://www.targetscan.org/) and miRDB (http://www.mirdb.org/), and the results of these two databases were intersected to guarantee the accuracy of the prediction (11,12). To elucidate the possible links between the differentially expressed genes, a miRNA-gene network was constructed using Cytoscape v3.2.0.

**Results**

**Differentially expressed genes in CD, compared with normal colon tissue.** A total of 432 differentially expressed genes, including 229 upregulated genes and 203 downregulated genes were screened out in the CD samples, compared with the normal colon tissues (Fig. 1). As shown in Fig. 1, the top five upregulated differentially expressed genes were matrix metallopeptidase (MMP)3, chemokine (C-X-C motif) ligand 1 (CXCL1), MMP1, regenerating family member 3α (REG3A) and CXCL9, whereas the downregulated genes were glutathione S-transferase α1 (GSTA1), carbonic anhydrase I (CA1), ATP synthase, H+ transporting, mitochondrial Fo complex subunit B1 (ATP5F1), heat shock protein family B (small) member 3 (HSPB3) and interleukin 2 (IL-2).

**GO enrichment analysis.** GO analysis was performed on these differentially expressed genes to provide a preliminarily outlook on their biological functions. As shown in Fig. 2, the upregulated genes were predominantly involved in regulation of inflammatory response, innate immune response, mast cell degranulation, cellular response to TNF and defense response to bacterium, indicating that innate immune system and microbial factors play an important role in inducing inflammation in CD. They were also enriched in proteolysis, chemotaxis, cytoskeleton organization and cell adhesion, and these functions enhance cell migration. These genes were also associated with extracellular matrix (ECM) organization, collagen fibril organization and mitogen-activated protein kinase cascade. However, the upregulated genes were found to be involved in the negative regulation of endothelial cell proliferation and angiogenesis, which was in contrast to the majority of previous studies and may require further investigation (13,14).

The downregulated genes were significantly enriched in the regulation of transcription, transport, response to stress, cellular respiration, chromatin modification and several other metabolic processes. In addition, the two groups of differentially expressed genes were involved in regulating the tyrosine phosphorylation of signal transducers and activators of transcription 3 (STAT3) protein, the mitotic cell cycle, smooth muscle cell proliferation and lipid metabolic process.

**Signaling network analysis.** As demonstrated in Fig. 3, the upregulated genes were primarily involved in focal adhesion, cell adhesion molecules B cell receptor signaling pathway, leukocyte transendothelial migration and natural killer cell mediated cytotoxicity, consistent with the results of the GO enrichment analyses. They were also associated with the Janus kinase/STAT (JAK/STAT) signaling pathway, apoptosis, calcium signaling pathway, cytokine-cytokine receptor interaction, Toll-like receptor signaling pathway and the peroxisome proliferator activated receptor. Other signaling pathways, including phosphatidylinositol 3-kinase/Akt and nuclear factor kB, were also upregulated, however, these were not statistically significant and thus not shown.

The downregulated genes were enriched in the transforming growth factor-β signaling pathway, tight junctions, basal transcription factors, oxidative phosphorylation and several other metabolic pathways.

**miRNA-gene network.** Based on the regulated association between the differentially expressed genes and miRNAs, a network of miRNA-genes was constructed. As presented in Fig. 4, miR-3148 and miR-6756-5p regulated the majority of genes, including JAK2 and collagen, type I, alpha 1 (COL1A1). miR-149-3p and miR-4447 regulated upregulated genes only, whereas the remaining miRNAs regulated upregulated and downregulated genes. These miRNAs may be central in the development and progression of CD.
Discussion

In the present study, the gene expression profiles of CD and normal colon tissue samples were examined using bioinformatics methods. It was found that 229 genes were upregulated and 203 genes were downregulated in the colon tissues of patients with CD, compared with those of normal controls. These differentially expressed genes may serve as characteristic genes closely associated with the diagnosis and treatment of CD, particularly those ranked at the top of the list. Among these genes, MMPs, including MMP1 and MMP3, are known to be markedly upregulated in the inflamed intestinal mucosa of patients with CD (15), which was in accordance with the results of the present study. As the imbalance of ECM components between the synthesis and breakdown in IBD can result in progressive tissue destruction or the over-deposition...
of collagens, finally causing ulcers, fistulas and fibrosis (16), MMPs are key in the intestinal tissue remodeling process (17). In addition, their upregulation may lead to patients with CD not responding to anti-TNFα therapy (18). In terms of the downregulated genes, GSTA1 may protect the cells from reactive oxygen species and the products of peroxidation through...
glutathione peroxidase activity. Restoring the expression of GSAT1 attenuates the inflammatory response in the intestinal mucosa of CD (19). Therefore, GSAT1 may serve as a novel therapeutic target and biomarker of prognosis in CD.

In the present study, GO analysis was performed to further interpret which functions these differentially expressed genes enriched. The results indicated that the functions included regulation of inflammatory response, immune response, defense response to bacterium and cellular response to tumor necrosis factor, which was consistent with a previous study, which identified CD is a complex inflammatory disease with the immune response and microbiota involved in its pathogenesis (20). In addition, these differentially expressed genes were significantly enriched in proteolysis, ECM organization and collagen fibril organization, the functions of which are important in the development of intestinal fibrosis and CD-associated strictures (17). As MMPs are involved in the degradation of the ECM, their potential role in the pathogenesis of CD warrants further investigation. The present study also showed that the functions of downregulated genes comprised the response to stress and cellular respiration, suggesting that oxidative stress is also involved in the etiology of CD.

Signaling pathways of the target genes were assessed using KEGG analysis. The results showed that several signaling pathways, including the cytokine-cytokine receptor interaction and the JAK/STAT signaling pathway, were associated with the occurrence of CD. Previous studies have indicated that various cytokines, including interleukins, regulate intracellular signaling through inducing the JAK/STAT pathway, and then convert extracellular stimuli into several cellular processes, for example, cell growth, proliferation and differentiation (21,22). In addition, the numerous pro-inflammatory cytokines secreted, including IL-6 and TNF-α, by immune cells, including T cells and neutrophils, may cause an inflammatory cascade by activating the JAK/STAT pathway, resulting in a mucosal inflammatory response (23). The findings indicated that the JAK/STAT pathway may be essential in regulation of the immune response and the pathogenesis of mucosal inflammation in CD, and present a promising novel therapeutic strategy for CD.

Finally, a miRNA-gene network was constructed in the present study to illustrate the internal association between the differentially expressed genes and the miRNAs of the target genes. Several miRNAs were found to be involved in the regulation of differentially expressed genes and may be considered as crucial regulators in the pathogenesis of CD. For example, miR-149 is downregulated in multiple types of tumor (24-26), and the overexpression of miR-149 can suppress cell migration, cell proliferation and the cell cycle (25,27). The downregulation of miR-149 can also lead to the high expression of IL-6 (28), a pro-inflammatory cytokine, which has been implicated in the pathogenesis of CD (29). Therefore, the role of miR-149 in CD requires further investigation, particularly as the function of miR-149 in CD has not been investigated.

In conclusion, the present study identified a number of differentially expressed genes, some of which may be important in the diagnosis and treatment of CD. Furthermore, using GO and KEGG analyses, combined with the construction of an miRNA-gene network, important gene functions, pathways and miRNAs, were identified, which may provide novel insights on the molecular mechanism and treatment of CD. However, further evidence from independent experimental data is required to confirm these results.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (grant nos. 81401875 and 81100263; http://www.nsfc.gov.cn/).

References

11. Agarwal V, Bell GW, Nam JW and Bartel DP: Predicting effective microRNA target sites in mammalian mRNAs. ELIFE 4, 2015.