Identification and Validation of Potential Hub Genes Correlated With Ulcerative Colitis Through Integrated Bioinformatics Analysis

Ting Li
Xinjiang Medical University https://orcid.org/0000-0001-9470-0451
Wenjia Hui
People's Hospital of Xinjiang Uygur Autonomous Region
Halina Halike
People's Hospital of Xinjiang Uygur Autonomous Region
Feng Gao (xjgf@sina.com)
People's Hospital of Xinjiang Uygur Autonomous Region https://orcid.org/0000-0001-7355-2672

Research

Keywords: Ulcerative colitis, differentially expressed genes, bioinformatics, hub genes

Posted Date: October 8th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-86637/v1

Abstract

Background: Ulcerative colitis (UC) is a chronic nonspecific intestinal inflammatory disorder associated with continuous, diffuse inflammatory alterations in the colonic mucosa of unknown etiology, Increasing evidence has showed aberrant expression of gene plays a vital function in the pathophysiological mechanisms of ulcerative colitis, Herein, we employed bioinformatics to investigate the core of the pathogenesis and provide potential markers for UC.

Results: We downloaded the GSE36807, GSE65114, GSE59071 datasets from the Gene Expression Omnibus(GEO), then the differentially expressed genes (DEGs) were determined using adjusted P□0. 05 and $|\log_2 FC>2|$ between normal samples and UC samples. Intersection analysis among three datasets showed 12 DEGs were found to be significantly dysregulated in UC. Results indicated that the DEGs were primarily associated with functions like the humoral immune response, antimicrobial humoral response, and CXCR chemokine receptor binding, and they were primarily enriched in KEGG pathways, including the IL−17 signaling pathway, and Toll−like receptor signaling pathway . Cytoscape software calculated that CXCL8, DMBT1, REG3A, S100A8, DUOX2, and MMP1 were hub genes of UC. In addition, We collected samples of 8 UC tissues and 8 normal Colonic tissues to validate the selected genes by *mRNA microarray*.

Conclusions: These results may provide potential biomarkers for UC, and our data and methodology provides new ideas that may be helpful in the understanding of the vital mechanisms underlying UC development.

Background

Ulcerative colitis (UC) is a chronic inflammatory-related condition with unknown etiology. The mechanism of UC is considered to be complicated. Studies have indicated that environment, genes, autoimmune factors, and intestinal microorganisms are valuable reference to the etiology of UC[1]. The morbidity of Ulcerative colitis is increasing in Western countries ,and expanding at an increasing rate in developing countries[2]. Moreover, due to repeated outbreak and protracted course of the disease, people with these disorders are often subjected to poor quality of life and additional financial burdens[3].Hence the need for expanding the pathogenetic recognition of UC.

Herein, we obtained three UC gene expression data from the GEO, a group of potentially hub gene was identified between normal samples and UC samples after conducting bioinformatics analysis. Also, the genes that were substantially upregulated were conducted for preliminary validation by mRNA microarray in our center. the present study provided molecular targets for UC and predicted the molecular mechanism underlying UC.

Results

1. DEGs in UC samples versus Normal Samples

We identified 76 DEGs from the GSE36807 dataset, and 95 DEGs in GSE59071, 28 DEGs were screened in GSE65114, a total of 12 DEGs were expressed in all the three datasets. (Fig 1) Compared with the normal tissues, there are 10 overexpressed and 2 under-expressed genes in UC tissues.

2. GO and pathway enrichment analysis of DEGs

To annotate mechanisms and function of screened overlapping genes, GO and KEGG pathway analyses were conducted on the 12 DEGs. Fig 2 shows the top ten terms enriched in biological process (BP), cellular component (CC). and molecular function (MF). For BP analysis, the enriched terms were those associated with humoral immune response, antimicrobial humoral response, and antimicrobial humoral immune response induced by antimicrobial peptide; CC analysis was mainly contained zymogen granule, apical part of cell, and cell leading edge; MF analysis mainly involved in CXCR chemokine receptor binding, glycosaminoglycan binding, and carbohydrate binding. According to the KEGG pathways, the genes were primarily enriched in the IL-17 signaling pathway. Moreover, to offer a comprehensible graphic depiction of the intricate association between genes and the relative KEGG pathway, the “pathway-gene” network was shown in Fig 2.

3. PPI network and module analysis

PPI network was established using Cytoscape software to investigate further the functions of DEGs in UC. The network included 12 nodes and 28 edges according to the data from the STRING database (Fig 3A). The top 6 hub nodes with a high degree of interaction in the PPI network included CXCL8, DMBT1, REG3A, S100A8, DUOX2, and MMP1(Fig 3B).

4. Verification of hub genes

We selected 6 hub genes to validate in colon tissues using microarray. Results of microarrays indicated that the hub genes were remarkably upregulated ($P<0.05$. Fig 4) in UC tissues, relative to the controls, and this was agreed with the previous bioinformatics analysis.

Discussion

Ulcerative colitis is a kind of common chronic inflammatory disease of colon, its cause is complicated, various kinds of internal and external factors participate in its pathogenesis[6]. There are currently no reliable markers for diagnosis, therefore, there is an urgent need for effective detection biomarkers. Herein, 12 DEGs were selected from three GEO datasets, consisting of 10 upregulated genes and 2 downregulated genes. Abnormal expression of 12 genes was mainly enriched during inflammation processes, including humoral immune response, response to chemokine, CXCR chemokine receptor binding. Based on the KEGG analysis, the main signaling pathway might focus on IL − 17 signaling pathway. subsequently, we identified 6 hub genes with high degrees in the PPI network. Following combined with our microarray data, CXCL8, DMBT1, REG3A, S100A8, DUOX2, and MMP1 could be useful diagnostic markers of UC.

According to the results of GO analyses, biological functions associated with UC pathogenesis included inflammatory response, zymogen granule, and chemotaxis, suggesting that the pathogenesis of UC could be facilitated by several factors, such as genetics, environment, as well as dysregulated immune responses. Based on the KEGG pathway analysis, one of the signaling pathway was enriched in the IL – 17 pathway. Several clinical trials suggest that the IL-23/IL-17 pathway might play a vital function in chronic inflammation, like ulcerative colitis and Crohn's disease.[7–9] Studies that involved mouse colitis models demonstrated that IL-17A might play a crucial function in protecting the integrity of the barrier of the intestinal epithelium, in spite of its potent proinflammatory properties[10, 11]. As such, we speculate that the IL-17 pathway could play a vital function in UC pathogenesis.

Out of the identified hub genes, CXCL8 was underscored as the gene with the highest degree of connectivity. CXCL8 ranked the first in the core gene. Based on the above analysis results, we hypothesized that CXCL8 might play an essential function in UC progression and its malignant complications. CXCL8, also known as neutrophil factor, is a member of the CXC chemokine family, which is an important mediator of inflammatory response, Increased expression of CXCL8 has been characterized in endothelial cells, cancer cells, and tumor-associated macrophages, indicating that CXCL8 may function as a significant regulatory factor within the tumor microenvironment[12]. IL-8 is vital in the induction of colonic inflammation and has been implicated in the IBD disease activity[13]. The previous studies showed that in UC patients, CXCL8 was particularly upregulated in the inflamed mucosa, relative to the non-inflamed mucosa [14–16].The gene Deleted in Malignant Brain Tumors 1 (DMBT1), which is found in chromosome 10q25.3-q26.1, is regarded as a tumor suppressor because of its homologous deletions and downregulation in lung cancer, medulloblastoma, gastrointestinal cancers, and glioblastoma multiforme[17, 18] The existing researches have confirmed the relationship between DMBT1 and UC. Treatment with IL-22 promoted DMBT1 expression by inducing the phosphorylation of STAT3 tyrosine, as well as activation of NF-κB, suggesting the IL-22/DMBT1 axis might play a vital function in UC pathophysiology[19]. The NOX/DUOX superfamily comprises dual oxidase 1 and 2 (DUOX1 and DUOX2), NOX1, 2, 3, 4 and 5, [20]. One of the major functions of DUOX is to maintain mucosal homeostasis[21]. Several NOX family members, such as DUOX2 and NOX1 are expressed in intestinal epithelial cells and are believed to be essential factors in the host mucosal surface defence[22].Evidence indicates that DUOX2 is upregulated in inflamed tissue, and play a vital role in facilitating the progression of IBDs.[23] The REG gene family comprises four important members: REG1A, REG1B, REG3A, and REG4, all of which are expressed in the small intestine and pancreas, but only REG4 is expressed constitutively in colon epithelia [24]. Previous studies have reported that REG3A is upregulated during IBD-related colon inflammation [25, 26]. Matrix Metalloproteinase-1 (MMP-1) is vital in the degradation and remodeling of extracellular matrix, MMP-1 is associated with acute injury of colonic mucosal tissue, formation of vessels, mucosal destruction, and initial steps of ulcer formation in UC patients[27, 28]. Enhanced generation of matrix metalloproteinases (MMPs) promotes tissue damage in IBD.[29] As an important pro-inflammatory cytokine in the S100 protein family, S100A8 is a key factor in inflammation and immunity, and it’s up regulation has been observed in IBD patients. A previous study reported a consistent and significant overexpression of S100A8 mRNA in ulcerative colitis, relative to the

control [30]. Moreover, S100A8 were up-regulated in the peripheral blood leucocytes of IBD patients.[31] Therefore, We supposed that these hub genes aberrantly expressed could facilitate the occurrence of UC.and their connection and significance in UC should be explored further. Finally, we validated this hub genes in our patients cohort by microarray technique, the expression of these hub genes were remarkably enhanced in the UC patients. which further suggested its convincing value.

Collectively, we performed all-inclusive bioinformatics analysis and selected 6 hub genes highly associated with UC and verified the robustness of their diagnostic significance. Meanwhile, all of them were preliminarily confirmed on our patients cohort and may act as possible biomarkers of UC. Besides, we exposed many vital gene functions and pathways that could boost our knowledge of the pathogenesis of UC. However, there is a need to conduct further studies to elucidate the precise mechanisms of UC.

Conclusion

In conclusion, we comprehensively analyzed associations between mRNA expressions and the development of UC. Among the DEGs, the abnormal expressions of CXCL8, DMBT1, REG3A, S100A8, DUOX2, and MMP1 were found to be significantly related to the UC, and were identified as key regulators. This study provide important guidance for selecting therapeutic strategies.

Methods

1. Microarray data

Three gene expression datasets (GSE36807, GSE65114, GSE59071) were obtained from the GEO database[4] (http://www.ncbi.nlm.nih.gov/geo/). In total, 158 mucosal biopsy samples are included, which consisted of 30 control subjects and 128 UC patients. The platform for GSE36807 was based on the GPL570 ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array), comprising 15 UC and 7 controls samples. GSE65114 dataset was obtained using the platform GPL16686 ([HuGene-2_0-st] Affymetrix Human Gene 2.0 ST Array [transcript (gene) version]), and comprised 12 control and16 UC samples. The GSE59071 platform involved the GPL6244 ([HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]), which comprised 97 UC and 11 controls samples. All UC samples were taken from pathological tissues of the colon biopsy.

2. Identification of DEGs and Data Preprocessing

Differential gene expression was analyzed using the ‘limma’ R package [5] between UC and control samples. The adjusted $p<0.05$ and $\log_2(FC) >2$ were regarded as the cutoff values. Statistical analysis was carried out for each dataset. The Venn diagram was used to identify the intersecting part and visualizes the DEGs. The screened DEGs were preserved and for the next bioinformatics analysis.

3. Functional enrichment analysis of DEGs.

We then investigated the functions of identified DEGs via by GO and KEGG pathway analyses, All KEGG pathway and GO analyses were carried out with the org.Hs.eg.db (version 3.7.0) and clusterProfiler (version 3.10.1) packages. Both q-value and p-value were <0.05.

4. Protein-protein interaction (PPI) network analysis and module selection

The STRING database (http://string-db.org) was utilized for analyzing the protein-protein interaction (PPI) among the DEGs. Cytoscape 3.7.0 software was performed to construct and visualize the network. Nodes with a higher degree of connectivity were regarded as more important in the entire network. And then calculate the degree of each protein node.

5. Validation of the selected gene using microarrays

The Ethics Committee of People's Hospital of Xin jiang Uygur Autonomous Region approved this study. A total of 8 UC tissues and 8 normal Colonic tissues were obtained from patients at the institution. All patients provided informed consent. Colonic mucosal RNA was extracted by QIAGEN RNeasy Kit. Qualitative and quantitative analyses of RNA were performed using NanoDrop ND-1000. cDNA amplification was conducted with MessageAmp™ Premier RNA Amplification Kit. Affymetrix Human Primeview microarrays were used to analyze microarray gene expression. GeneChip Fluidics Station 450 plus GeneChip Scanner 3000 were employed to scan the arrays as per the manufacturer's protocol after they were processed, hybridized, and then washed.

6. Statistics analysis

Data analysis was completed using GraphPad Prism 6.0. or R statistical package (R version 4.0.0). Differences between various groups were determined using the *t*-test. Probability value $p < 0.05$ defined statistical significance.

Abbreviations

CXCL8
interleuin-8; DEGs:differentially expressed genes; DMBT1:Deleted in Malignant Brain Tumors 1; DUOX2:Dual oxidase 2; FC:fold change; GO:Gene Ontology; IBD:Inflammatory bowel disease; KEGG:Kyoto Encyclopedia of Genes and Genomes; MMP-1:Matrix Metalloproteinase-1;PPI:Protein-protein interaction; REG:Regenerating gene protein; UC:Ulcerative colitis.

Declarations

Ethics approval and consent to participate

Ethic approval was provided by the Ethics Committee of People's Hospital of Xin jiang Uygur Autonomous Region.

Consent for publication

All the authors have consented for the publication.

Availability of data and materials

The data sets used in this study can be obtained from the corresponding author upon reasonable request. The public data source: Gene Expression Omnibus database with the accession GSE36807,GSE65114,and GSE59071 (https://www.ncbi.nlm.nih.gov/geo/)

Competing interests

There are no competing interests to declare.

Funding

This study was funded by the National Natural Science Foundation of China (No. 81560091).

Authors' contributions

TL conceptualized and planned the study, collected samples for microarray and wrote the manuscript. WJ.H and HH analyzed the interpreted the results and microarray datasets. FG designed the study, revised it critically for significant intellectual information, and approved the final version for publication. The final manuscript was read and approved by all authors.

Acknowledgments

The results of this study are derived from the GEO database. We are grateful to authors whose data is used in this study.

References

1. Chu H, Khosravi A, Kusumawardhani IP, Kwon AH, Vasconcelos AC, Cunha LD, et al. Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. Science (New York, NY). 2016;352(6289):1116-20. doi:10.1126/science.aad9948.
2. Kaplan GG, Ng SC. Understanding and Preventing the Global Increase of Inflammatory Bowel Disease. Gastroenterology. 2017;152(2):313-21.e2. doi:10.1053/j.gastro.2016.10.020.
3. Hoivik ML, Bernklev T, Moum B. Need for standardization in population-based quality of life studies: a review of the current literature. Inflammatory bowel diseases. 2010;16(3):525-36. doi:10.1002/ibd.21032.
4. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, et al. NCBI GEO: mining tens of millions of expression profiles--database and tools update. Nucleic acids research. 2007;35(Database issue):D760-5. doi:10.1093/nar/gkl887.

5. Diboun I, Wernisch L, Orengo CA, Koltzenburg M. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. BMC genomics. 2006;7:252. doi:10.1186/1471-2164-7-252.
6. Danese S, Fiocchi C. Ulcerative colitis. The New England journal of medicine. 2011;365(18):1713-25. doi:10.1056/NEJMra1102942.
7. Geremia A, Arancibia-Cárcamo CV, Fleming MP, Rust N, Singh B, Mortensen NJ, et al. IL-23-responsive innate lymphoid cells are increased in inflammatory bowel disease. The Journal of experimental medicine. 2011;208(6):1127-33. doi:10.1084/jem.20101712.
8. Sutton CE, Lalor SJ, Sweeney CM, Brereton CF, Lavelle EC, Mills KH. Interleukin-1 and IL-23 induce innate IL-17 production from gammadelta T cells, amplifying Th17 responses and autoimmunity. Immunity. 2009;31(2):331-41. doi:10.1016/j.immuni.2009.08.001.
9. Kenna TJ, Davidson SI, Duan R, Bradbury LA, McFarlane J, Smith M, et al. Enrichment of circulating interleukin-17-secreting interleukin-23 receptor-positive γ/δ T cells in patients with active ankylosing spondylitis. Arthritis and rheumatism. 2012;64(5):1420-9. doi:10.1002/art.33507.
10. Lee JS, Tato CM, Joyce-Shaikh B, Gulen MF, Cayatte C, Chen Y, et al. Interleukin-23-Independent IL-17 Production Regulates Intestinal Epithelial Permeability. Immunity. 2015;43(4):727-38. doi:10.1016/j.immuni.2015.09.003.
11. Maxwell JR, Zhang Y, Brown WA, Smith CL, Byrne FR, Fiorino M, et al. Differential Roles for Interleukin-23 and Interleukin-17 in Intestinal Immunoregulation. Immunity. 2015;43(4):739-50. doi:10.1016/j.immuni.2015.08.019.
12. Waugh DJ, Wilson C. The interleukin-8 pathway in cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2008;14(21):6735-41. doi:10.1158/1078-0432.ccr-07-4843.
13. Skov L, Beurskens FJ, Zachariae CO, Reitamo S, Teeling J, Satijn D, et al. IL-8 as antibody therapeutic target in inflammatory diseases: reduction of clinical activity in palmoplantar pustulosis. Journal of immunology (Baltimore, Md : 1950). 2008;181(1):669-79. doi:10.4049/jimmunol.181.1.669.
14. Bruno ME, Rogier EW, Arsenescu RI, Flomenhoft DR, Kurkjian CJ, Ellis GI, et al. Correlation of Biomarker Expression in Colonic Mucosa with Disease Phenotype in Crohn's Disease and Ulcerative Colitis. Digestive diseases and sciences. 2015;60(10):2976-84. doi:10.1007/s10620-015-3700-2.
15. Lin X, Li J, Zhao Q, Feng JR, Gao Q, Nie JY. WGCNA Reveals Key Roles of IL8 and MMP-9 in Progression of Involvement Area in Colon of Patients with Ulcerative Colitis. Current medical science. 2018;38(2):252-8. doi:10.1007/s11596-018-1873-6.
16. Korolkova OY, Myers JN, Pellom ST, Wang L, M'Koma AE. Characterization of Serum Cytokine Profile in Predominantly Colonic Inflammatory Bowel Disease to Delineate Ulcerative and Crohn's Colitides. Clinical medicine insights Gastroenterology. 2015;8:29-44. doi:10.4137/CGast.S20612.
17. Mollenhauer J, Wiemann S, Scheurlen W, Korn B, Hayashi Y, Wilgenbus KK, et al. DMBT1, a new member of the SRCR superfamily, on chromosome 10q25.3-26.1 is deleted in malignant brain tumours. Nature genetics. 1997;17(1):32-9. doi:10.1038/ng0997-32.

18. Somerville RP, Shoshan Y, Eng C, Barnett G, Miller D, Cowell JK. Molecular analysis of two putative tumour suppressor genes, PTEN and DMBT, which have been implicated in glioblastoma multiforme disease progression. Oncogene. 1998;17(13):1755-7. doi:10.1038/sj.onc.1202066.
19. Fukui H, Sekikawa A, Tanaka H, Fujimori Y, Katake Y, Fujii S, et al. DMBT1 is a novel gene induced by IL-22 in ulcerative colitis. Inflammatory bowel diseases. 2011;17(5):1177-88. doi:10.1002/ibd.21473.
20. Bokoch GM, Knaus UG. NADPH oxidases: not just for leukocytes anymore! Trends in biochemical sciences. 2003;28(9):502-8. doi:10.1016/s0968-0004(03)00194-4.
21. Allaoui A, Botteaux A, Dumont JE, Hoste C, De Deken X. Dual oxidases and hydrogen peroxide in a complex dialogue between host mucosae and bacteria. Trends in molecular medicine. 2009;15(12):571-9. doi:10.1016/j.molmed.2009.10.003.
22. El Hassani RA, Benfares N, Caillou B, Talbot M, Sabourin JC, Belotte V, et al. Dual oxidase2 is expressed all along the digestive tract. American journal of physiology Gastrointestinal and liver physiology. 2005;288(5):G933-42. doi:10.1152/ajpgi.00198.2004.
23. Lipinski S, Till A, Sina C, Arlt A, Grasberger H, Schreiber S, et al. DUOX2-derived reactive oxygen species are effectors of NOD2-mediated antibacterial responses. Journal of cell science. 2009;122(Pt 19):3522-30. doi:10.1242/jcs.050690.
24. Zhang YW, Ding LS, Lai MD. Reg gene family and human diseases. World journal of gastroenterology. 2003;9(12):2635-41. doi:10.3748/wjg.v9.i12.2635.
25. Granlund A, Beisvag V, Torp SH, Flatberg A, Kleveland PM, Ostvik AE, et al. Activation of REG family proteins in colitis. Scandinavian journal of gastroenterology. 2011;46(11):1316-23. doi:10.3109/00365521.2011.605463.
26. van Beelen Granlund A, Østvik AE, Brenna Ø, Torp SH, Gustafsson BI, Sandvik AK. REG gene expression in inflamed and healthy colon mucosa explored by in situ hybridisation. Cell and tissue research. 2013;352(3):639-46. doi:10.1007/s00441-013-1592-z.
27. von Lampe B, Barthel B, Coupland SE, Riecken EO, Rosewicz S. Differential expression of matrix metalloproteinases and their tissue inhibitors in colon mucosa of patients with inflammatory bowel disease. Gut. 2000;47(1):63-73. doi:10.1136/gut.47.1.63.
28. Wang YD, Yan PY. Expression of matrix metalloproteinase-1 and tissue inhibitor of metalloproteinase-1 in ulcerative colitis. World journal of gastroenterology. 2006;12(37):6050-3. doi:10.3748/wjg.v12.i37.6050.
29. Morgan AR, Han DY, Lam WJ, Triggs CM, Fraser AG, Barclay M, et al. Genetic variations in matrix metalloproteinases may be associated with increased risk of ulcerative colitis. Human immunology. 2011;72(11):1117-27. doi:10.1016/j.humimm.2011.08.011.
30. Zhang X, Ai F, Li X, She X, Li N, Tang A, et al. Inflammation-induced S100A8 activates Id3 and promotes colorectal tumorigenesis. International journal of cancer. 2015;137(12):2803-14. doi:10.1002/ijc.29671.
31. Vavricka SR, Schoepfer A, Scharl M, Lakatos PL, Navarini A, Rogler G. Extraintestinal Manifestations of Inflammatory Bowel Disease. Inflammatory bowel diseases. 2015;21(8):1982-92.

doi:10.1097/mib.0000000000000392.

Figures

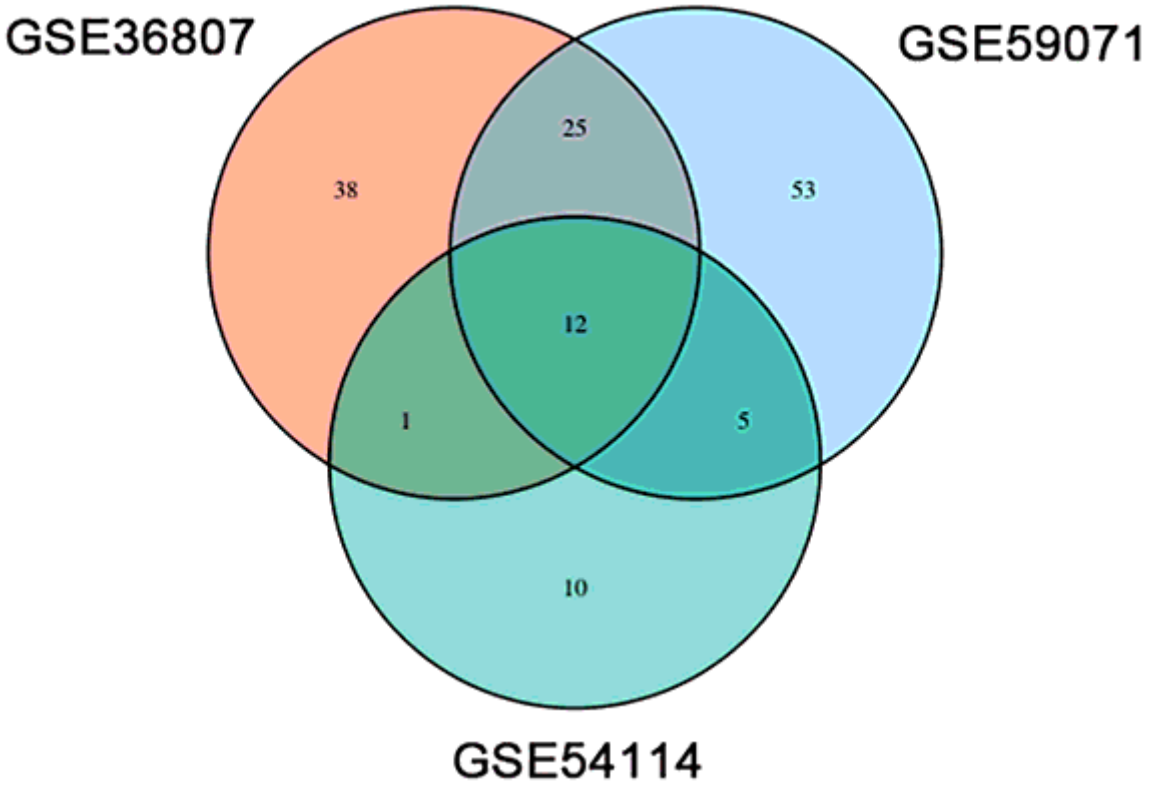

Figure 1

Identification of common DEGs from the three datasets GSE36807, GSE65114, GSE59071

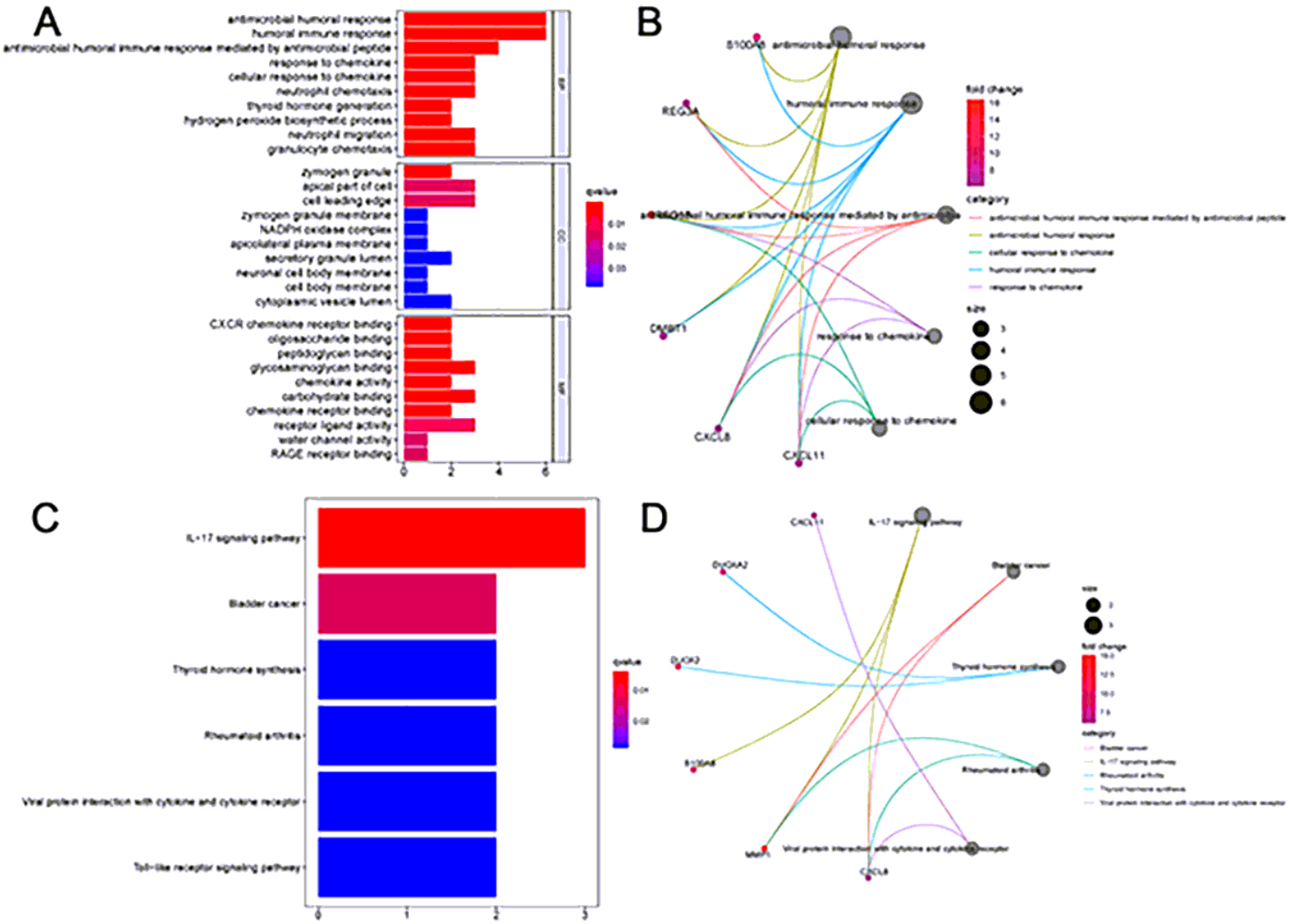

Figure 2

Functional enrichment analyses of gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). (A) The bubble plot of enriched GO terms. (B) The circos plot of enriched GO terms. (C) The bubble plot of enriched KEGG pathways. (D) The circos plot of enriched KEGG pathways.

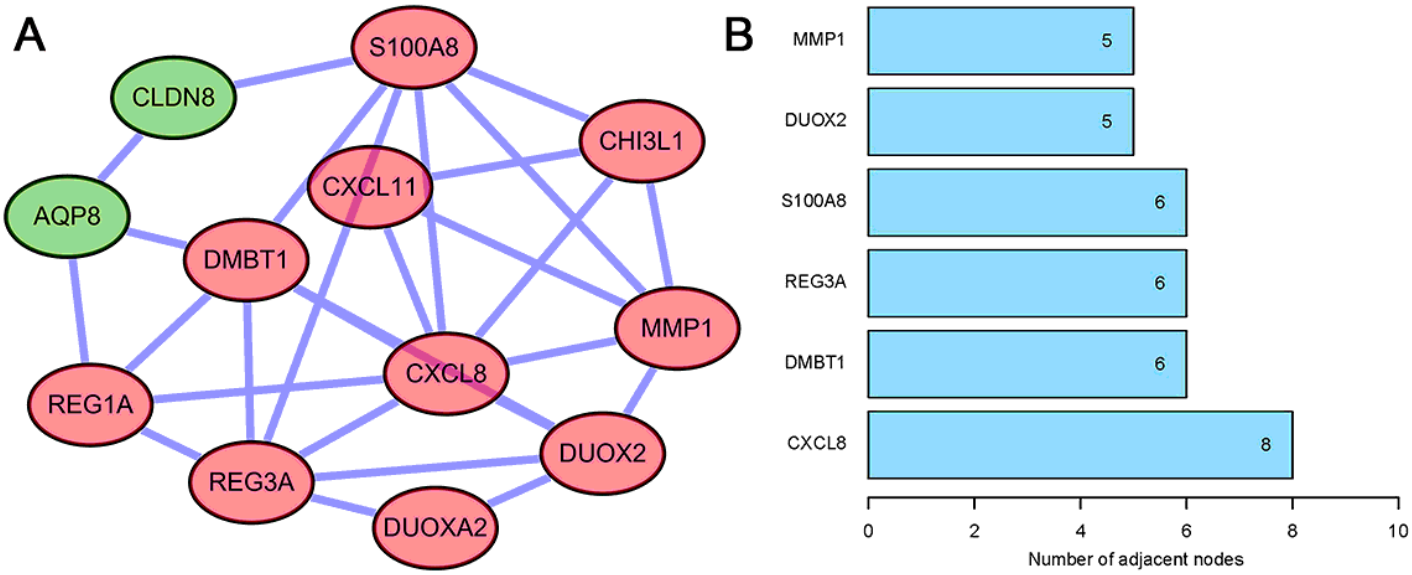

Figure 3

Protein-protein interactions network and hub genes. (A) PPI network visualized with Cytoscape. (B) Hub genes with degree beyond 5.

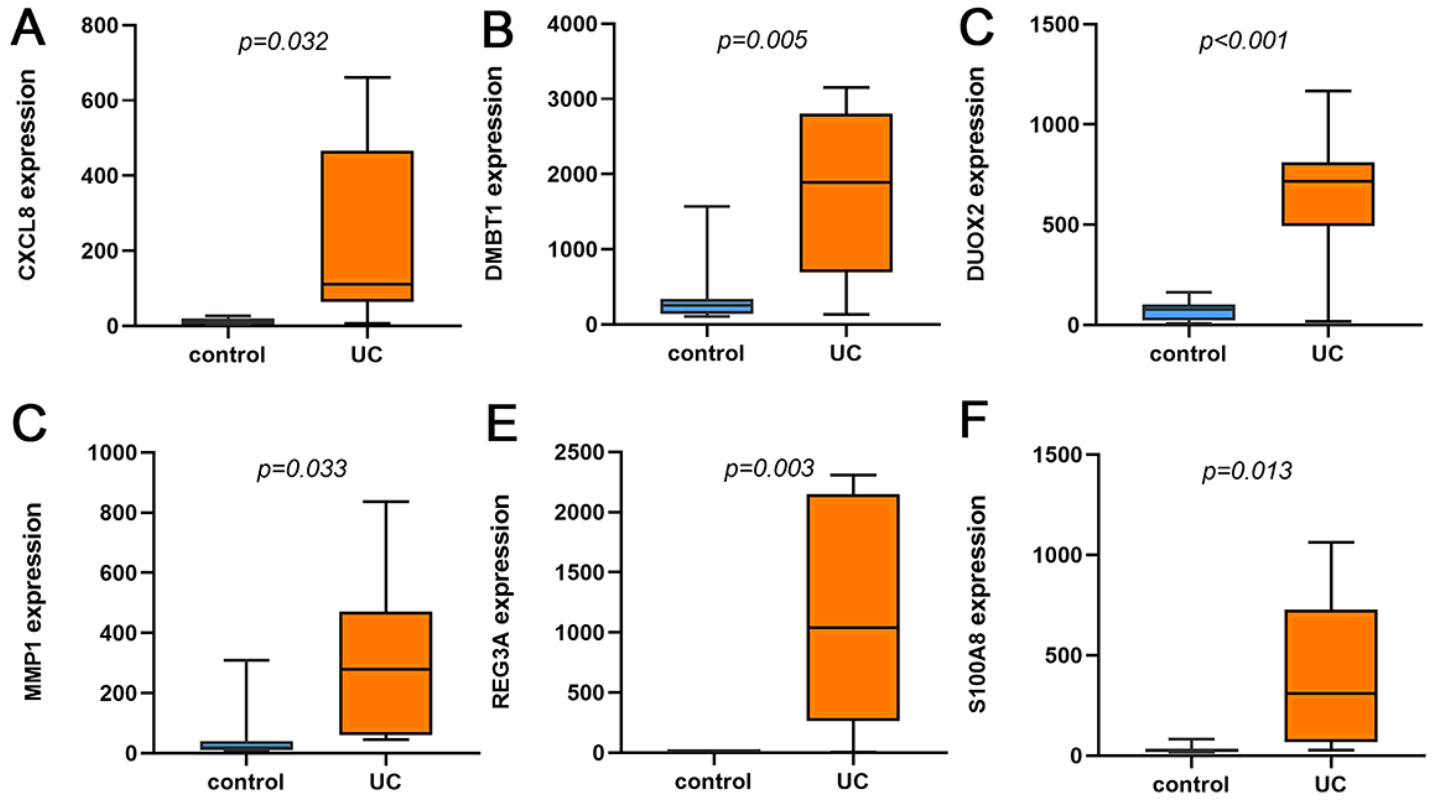

Figure 4

Validation of the 6 hub genes' expression levels between UC samples and normal samples by microarray. (A) CXCL8, (B) DMBT1, (C) DUOX2, (D) MMP1, (E) REG3A, (F)S100A8