

# Identification of Key Genes and Pathways in Ulcerative Colitis Through Bioinformatics Analysis

Mingdi Liu (✉ [liumd555@163.com](mailto:liumd555@163.com))

Jilin University <https://orcid.org/0000-0001-9978-7330>

Fa Ping Li

Jilin University First Hospital

Yong Ping Jian

Jilin University

Ge Yang

Jilin University

Dan Zhang

Jilin University

Yi Shu Wang

Jilin University

Zhi Xiang Xu

Jilin University

---

## Research article

**Keywords:** Ulcerative colitis, bioinformatics, colon cancer, drug

**Posted Date:** January 19th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-148508/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background** Ulcerative colitis (UC) is a chronic inflammatory disease whose therapy remains largely uncertain due to the lack of etiological understanding. It is also a higher risk factor for colon cancer.

**Aims** This study was designed to identify key genes correlated with intestinal epithelial repair and their associated signaling pathways in the pathogenesis and malignant transformation of UC.

**Methods** With an online database pubmed2ensemble, correlative genes were identified to be associated with both UC and its self-healing, and then were imported in Cytoscape for enrichment analysis. A protein-protein interaction network containing was established, and the most significant gene modules and top ten hub genes were selected for further enrichment analysis. Then, potential drugs that target the corresponding genes were identified with online tool in DGIdb. A map which reflected the links among genes, drugs, and their associated pathways was constructed. Finally, we searched the TCGA database and CPTAC database for testing the gene expression of identified UC-associated genes in colon cancer.

**Results** We identified FN1, GRB2, EGF, CXCL8, VEGFA, STAT3, IL6, IL4, IL10, TNF, CSF2, IL13, IL1A, CCL2, ICAM1 and IL18 had closest associations with the function of epithelial function in UC patients, based on functional enrichment analysis. Besides, 11 of them were validated in colon cancer.

**Conclusions** The corresponding signaling pathways that are altered in UC, which provide a new clue for understanding the mechanism underlying the UC pathogenesis and malignant transformation. Key signals in the process and their target drugs might serve as new treatment strategy for UC.

## Introduction

Ulcerative colitis (UC) and Crohn's disease are two forms of inflammatory bowel disease, whose main clinical manifestations are diarrhea and bloody stools. UC is an inflammatory condition affecting the colon, with an annual incidence of approximately 10 to 20 per 100,000 people. UC is associated with genetic susceptibility, composition and function of intestinal microbiota, environmental factors, and mucosal and systemic immune responses [1, 2]. These factors and their interactions make damages to the intestinal mucosal barrier and cause disorder of the immune system, finally resulting in continuous intestinal inflammation. However, the specific pathogenic mechanism for the development of UC has not been clarified. In order to improve the life quality of the patients and even cure the chronic disease, it is necessary to identify key genes and pathways in the pathogenesis of UC.

The mammalian gastrointestinal mucosa is a rapidly self-renewing tissue, which is lined by continuous epithelial cells. Epithelial cells maintain the physical and functional balance under both physiology and pathophysiology conditions [3]. Composed of intestinal epithelial cells, goblet cells, Paneth cells, endocrine cells and others, the intestinal epithelium bears multiple functions and active metabolism so that it can rapidly turn over and regenerate [4–7]. These characteristics are possibly associated with the ability of intestine self-healing.

Population-based studies showed that the risk of colorectal neoplasia in patients with IBD increased 2–5 folds as compared with age-matched controls [8–10]. Inflammation is an important risk factor for interval colorectal cancer in IBD patients [11]. Inflammation exerts a field effect on constantly re-epithelializing and chronically inflames tissue, leading to dysplasia. The dysplastic tissue then progresses to colorectal cancer depending on various molecular pathways, such as chromosomal instability and microsatellite instability [12]. However, factors associated with the transformation of UC to colon cancer are not clear.

With the improvement in bioinformatics technology, more public databases are available to biologists to find inspirations in biology and medical research. To find the relevant literature on specific genomic regions or sets of functionally related genes more easily, pubmed2ensembl offers text-based queries against PubMed and PubMed Central documents in conjunction with constraints on genomic features [13]. Although differentially expressed genes analysis has become mainstream of bioinformatics, text mining still has its significance and can also be dug out to obtain valuable information especially in diseases other than tumors.

In this study, we established the gene dataset of UC from pubmed2ensembl and obtained 137 common genes, which are possibly correlated with UC and its intestine self-healing. We then used Rstudio software with different R packages for gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. Additional analyses, such as PPI network complex construction and gene cluster findings, were also applied in the current study. Moreover, drug-gene interactions were explored using DGldb-based gene cluster digging. Connection between the two was then expanded into three elements according to the results of KEGG pathway analysis. Finally, the expression level of 16 genes identified by the above-referenced methods was evaluated in TCGA and CPTAC databases using UALCAN tool. In summary, our findings identified clusters of genes and pathways tightly associated with the pathogenesis and malignant transformation of UC and provides a clue for choosing possible drugs in the treatment of UC based on the signal interaction analysis of the genes identified.

## Methods

### Gene sets collection

Two gene lists of UC and self-healing were downloaded from pubmed2ensembl website (<http://www.pubmed2ensembl.org>).

### Go enrichment and KEGG pathway analysis

GO knowledgebase is a source for information on the functions of genes, composed of biological process, molecular function and cellular component [14]. The KEGG is a collection of manually drawn pathway maps representing the knowledge on the molecular interaction, reaction and relation networks. It links genomic information to high-level functional identification [15]. These analyses were carried out with the ClusterProfiler package in Rstudio.  $p$  value  $< 0.05$  was considered as significant.

### PPI network and gene module analysis

Protein-protein interaction (PPI) network was created using the Search Tool for the Retrieval of Interacting Genes and Proteins database (STRING version 11.0). The active interaction sources are from text-mining, experiments, databases, co-expression, neighborhood, gene fusion and co-occurrence. The PPI networks were then visualized by Cytoscape software. The plug-in Molecular Complex Detection (MCODE) built-in Cytoscape was applied to identify the significant gene modules or the PPI networks [16]. The significant module was visualized with degree cutoff = 2, node score cutoff = 0.2, k-core = 2 and max.depth = 100. The CytoHubba plugin is implemented in Java, based on the Cytoscape, which provides a user-friendly interface to explore important nodes in biological networks [17]. It explores important nodes/hubs and fragile motifs in an interactome network by several topological algorithms including Degree, Edge Percolated Component, Maximum Neighborhood Component, Density of Maximum Neighborhood Component, Maximal Clique Centrality and centralities based on shortest paths, such as Bottleneck, Eccentricity, Closeness, Radiality, Betweenness, and Stress [17]. At present, Degree is the method of choice in most research in already published articles. In the current study, the top 10 nodes ranked by Degree, and hub-network was visualized in Cytoscape.

## **Drug-Gene-Pathway Interactions**

The Drug Gene Interaction Database (DGIdb [www.dgidb.org](http://www.dgidb.org)) is a web means that consolidates disparate data sources describing drug-gene interactions and gene druggability. It is a user-friendly tool for mining the druggable genome for precision medicine hypothesis generation [18]. For DGIdb 3.0, it has been a substantial expansion through the addition of new sources and the updating of existing sources [19]. Using ggplot2 and ggalluvial packages in Rstudio, a Sankey diagram of Drug-Gene-Pathway was drawn to reflect the co-relation among three elements.

## **Validation of gene expression in TCGA**

UALCAN, a tool for in-depth analyses of TCGA data, is a publicly available web-platform for researchers to facilitate the study of gene expression variation associations across different tumors [20]. To further screen potential targets in the malignant transformation of UC into colon adenocarcinoma (COAD), candidate genes were evaluated in UALCAN. 286 tumor samples and 41 normal samples both on RNA level and protein level were evaluated.  $p$  value < 0.05 was considered as statistically significant.

## **Results**

### **Identification of relative gene**

In the current study, we obtained two gene lists from pubmed2ensembl using the keyword of UC and self-healing. The two gene lists included 1104 and 232 genes, respectively. Using venny online tool (website: <https://bioinfogp.cnb.csic.es/tools/venny/>), we identified 137 genes shared by both UC and self-healing individuals (Supplement Table 1 and Fig. 1A).

### **Function enrichment analysis**

For mutually shared genes, we analyzed the functional and signal pathway enrichment using clusterProfiler package in R with the criteria of p value < 0.05. As shown in Fig. 1, the top 10 GO enrichment terms included biological process, cellular component, and molecular function.

In the signaling pathway analysis with KEGG, these common genes were mainly enriched in cytokine – cytokine receptor interaction, JAK – STAT signaling pathway, and proteoglycans in cancer (Fig. 1B).

For those altered in biological process, genes were mainly enriched in the T cell activation, leukocyte migration and positive regulation of cytokine production (Fig. 1C). Genes that involve in cellular component were mainly enriched in the external side of plasma membrane, membrane region, and membrane raft (Fig. 1D). Genes changed in molecular function were mainly enriched in the receptor ligand activity, cytokine receptor binding, and cytokine activity (Fig. 1E).

## **PPI network and module analysis**

In order to build an interactive work, the 137 common genes were uploaded onto Cytoscape software following the instructions in the STRING online database (<http://string-db.org>). It generated 105 nodes and 399 edges (Fig. 2A). The other 32 genes didn't fall into the PPI network. Based on MCODE, the significant modules containing IL6, IL13, IL1A, IL10, CSF2, IL18, IL4, CCL2, TNF, ICAM1, and STAT3 from the PPI network were selected (Fig. 2B). The MCODE score of gene cluster is 8.8, containing 11 nodes and 44 edges which showed the closest relationship with two keywords selected in this study. Using cytoHubba software, 10 genes (GRB2, IL6, CXCL8, IL10, TNF, STAT3, IL4, FN1, EGF, and VEGFA) with a higher degree of connectivity were identified as hub genes (Fig. 2C). The network and sub-networks caught by the topological analysis with cytohubba strategy should lead to new insights on essential regulatory networks and protein drug targets for experimental biologists [21]. Notably, 50% of hub genes (5/10) were consistently included in the most significant gene cluster, which further confirmed the critical roles of these elements and also suggest the reliability and consistency of the methods. Interestingly, among the 16 highlighted genes in PPI network, 11 of them are involved in inflammation. ICAM-1 is an adhesion molecule, which plays a critical role in the maintenance of epithelial integrity. GRB2 is involved in EGF receptor-mediated signal transduction, which mainly promotes the proliferation of epithelial cells. FN1 is involved in cell adhesion, cell motility, opsonization, wound healing and maintenance of cell shape. VEGFA, an antigenic growth factor, regulates angiogenesis and vasculogenesis in normal and cancer individuals, implying a relationship between UC pathological changes and tumorigenesis.

In the signaling pathway analysis with KEGG, a total of 16 genes were mainly enriched in cytokine – cytokine receptor interaction, rheumatoid arthritis, and AGE-RAGE signaling pathway in diabetic complications (Fig. 2D). Mainly based on connection data, MCODE can reveal closely relative regions in the PPI network, some of which have been proven to be in a complex. Thus, we analyzed cellular component enrichment of 11 cluster genes, including IL13, IL1A, IL10, CSF2, IL18, IL4, CLL2, TNF, ICAM1, STAT3, and IL6 in Rstudio. We found that these genes that involve in cellular component were mainly enriched in external side of plasma membrane (Fig. 2E). As for altered hub genes containing IL6, CXCL8, IL10, TNF, STAT3, IL4, FN1, EGF, VEGFA and GRB2, those involved in molecular function were mainly

enriched in the cytokine receptor binding, receptor ligand activity, and growth factor receptor binding (Fig. 2F). In the biological pathway analysis, hub genes were mainly enriched in positive regulation of cell adhesion, leukocyte migration, and regulation of JAK-STAT cascade (Fig. 2G).

## Gene-Drug-Pathway analysis

Pharmacogenomics is a new interdisciplinary field developed in the 1990s based on genetics, genomics and genetic pharmacology. It mainly focuses on the relationship between human genomic information and drug response. DGldb mines existing resources that generate hypotheses about how mutated genes might be targeted therapeutically or prioritized for drug development [22]. It provides a platform for searching lists of drugs against a compendium of drug-gene interactions.

Based on previously obtained two gene sets, which contained 11 and 10 elements respectively (Fig. 2B, 2C), we searched the DGldb and acquired a relative drug list.

Among these 16 genes, IL4, EGF, FN1, IL13, CCL2 and STAT3 have not been currently considered as the drug targets for UC according to DGldb. The final list comprised only the drugs, which were approved by FDA (Food and Drug Administration) and used as an immunotherapy drug (Supplement Table 2). In this table, four drugs including infliximab, certolizumab pegol, adalimumab, and golimumab were previously used in UC and (or) Crohn's disease. Most of targets (reagents) identified were not to be considered as anti-UC drugs previously.

As shown in Sangquito, 32 drugs and 8 genes were relative to the top 10 KEGG pathways (Supplement Table 3 and Fig. 3), containing 768 lines which indicated correlations of three elements. IL6 and TNF were linked to multiple drugs and pathways, which indicated that they might be important therapeutic target in UC. Moreover, etanercept, interferon alfa-2b, mycophenolic acid and procarbazine are associated with more than one gene. Etanercept (Enbrel) is used for the treatment of rheumatoid arthritis and axial spondyloarthritis, two diseases with inflammation as the main phenotype [23–25]. Interferon alfa-2b is an enhancing promotor in immune system, which is used in the treatment of mucosal melanoma of the head and neck [26], recurrent conjunctival papillomatosis [27], ocular surface squamous neoplasia [28], and several types of cancers. Mycophenolic acid is an immunosuppressive drug used for the prevention of rejection in solid organ transplantation and for the treatment of some immune diseases [29]. Procarbazine is also an anti-tumor drug used for the treatment of Hodgkin's disease which usually in combination with mechlorethamine, vincristine, and prednisone [30]. Together, these data provide new clues for targeted therapy in the treatment of UC patients.

## Association between UC and colon cancer

UC is associated with an increased prevalence of colorectal cancer although the mechanisms underlying the neoplastic transformation from UC are rarely understood [31–33]. In order to identify the relationship between two pathological changes, we searched the expression of 16 genes identified above in TCGA database using an online tool on the UALCAN website. We found that the expression of IL10, IL18, STAT3, GRB2 and EGF were downregulated in COAD individuals, which possibly reflected a protective

mechanism for tumorigenesis (Fig. 4A). In contrast, IL1A, ICAM1, CSF2, IL6, FN1 and VEGFA were upregulated in COAD datasets at RNA level as compared with those in normal samples (Fig. 4B). Interestingly, TPM values of IL13 and IL4 genes were extremely low (TPM < 1). There was no statistical significance for CCL2 and TNF between COAD and normal samples.

Moreover, using data from Clinical Proteomic Tumor Analysis Consortium (CPTAC) Confirmatory/Discovery dataset, UALCAN also provides a protein expression analysis option for colon cancer. Z-values represent standard deviations from the median across samples for the given cancer type. Log<sub>2</sub> Spectral count ratio values from CPTAC were first normalized within each sample profile, then normalized across samples. Among the proteins regulated by the gene set of the cluster and hub genes, ICAM1 and CXCL8 are upregulated in colon cancer with significant differences ( $Z < 0.01$ ) than normal samples (Fig. 4C). In contrast, GRB2 has a lower expression level in colon cancer samples (Fig. 4D). All the statistical significance data of gene expression were shown in supplement table 4.

The primary purpose of the study focuses on the prediction of key potential genes in the malignant change of UC via data mining and data analysis, and these results need to be confirmed through more molecular and cellular experiments.

## Discussion

In this study, our purpose was to identify potential genes and pathways with the pathogenesis and transformation of UC. It has a progressive and remitting or relapsing course related to epithelial regeneration, as a crucial pathophysiological feature of the disease [34]. Therefore, it's paramount to explore the key factors working in these pathological processes.

The current study showed that damage of intestinal epithelial self-healing are related to immune responses in UC. With the rising prevalence in global, drugs were developed in the treatment of UC, including 5-aminosalicylic acid, steroids, immunosuppressants and some active compounds extracted from Chinese medicine. The anti-UC activities of them are mainly associated with increased level of anti-inflammatory cytokine (IL-4 and IL-10) and decreased level of pro-inflammatory cytokines (TNF- $\alpha$ , IL-1 $\beta$ , IL-6, IL-8, IL-23 and NF- $\kappa$ B) [35]. But whether these drugs could target the function of intestinal epithelial cells had never been studied. Our research supported that the therapy targeting the self-healing ability of intestinal epithelial might be a new therapeutic strategy for UC.

Moreover, the changes of these potential targets in colon cancer patients suggested that intestinal epithelial self-repairing acted as a bridge between UC and colitis-associated cancer. Researchers previously characterized factors, which share some functional overlaps with targets we identified. Deng et al. [36] showed that Yes-associated protein plays a critical role in intestinal epithelial cell self-renewal, regeneration and tumorigenesis. Zundler et al. [35] reported that JAK/STAT signaling controls secretion of cytokines and transition of inflammatory lesions to tumors leading to colitis-associated cancer via the IL6/STAT3 axis.

In our study, we found that 11 immune-related genes (IL6, IL13, IL1A, IL10, CSF2, IL18, IL4, CCL2, TNF, STAT3 and CXCL8) were evaluated as crucial targets in intestinal epithelial damage and repair in UC patients. It suggested that immune system plays an important role in the whole course of UC. Cellular component analysis showed that the proteins and RNA coded by these genes might regulate the pathogenesis of UC at the external side of plasma membrane. And it also might be the position where these potential drugs work. Moreover, clarifying the link between UC and UC-related colon cancer will provide new clues to therapeutic strategies for UC, reduce the chance of malignant transformation and improve the quality of lives for UC patients. However, more direct evidence is needed to confirm the function by various experiments.

## Declarations

### COMPLIANCE WITH ETHICAL STANDARDS

This article does not contain any studies with human participants or animals performed by any of the authors.

### CONFLICTS OF INTEREST

All authors have declared no conflicts of interest

### FUNDING

This article was supported by Graduate Innovation Fund of Jilin University.

### AUTHOR CONTRIBUTIONS

ML wrote the first draft of the manuscript and performed the most of comparative analysis. All authors participated in the data analysis and discussion. All authors revised the final version of the manuscript.

### DATA AVAILABILITY

The data that support the findings of this study are available in TCGA, CPTAC and DGIdb databases at URL, reference number: 13, 14 and 15. These data were derived from the following resources available in the public domain: TCGA database: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>; CPTAC database: <https://proteomics.cancer.gov/programs/cptac>; DGIdb database: <http://www.dgidb.org>.

## References

1. Roberts-Thomson IC, Bryant RV, Costello SP. Uncovering the cause of ulcerative colitis. *JGH Open*. 2019;3(4):274–6. doi:10.1002/jgh3.12216.
2. Steed H. Ulcerative colitis. *Medicine*. 2019;47(6):371–6. doi:10.1016/j.mpmed.2019.03.001.

3. Gao JH, Guo LJ, Huang ZY, Rao JN, Tang CW. Roles of cellular polyamines in mucosal healing in the gastrointestinal tract. *J Physiol Pharmacol*. 2013;64(6):681–93.
4. Kaur H, Moreau R. Role of mTORC1 in intestinal epithelial repair and tumorigenesis. *Cell Mol Life Sci*. 2019;76(13):2525–46. doi:10.1007/s00018-019-03085-6.
5. Ayabe T, Satchell DP, Wilson CL, Parks WC, Selsted ME, Ouellette AJ. Secretion of microbicidal alpha-defensins by intestinal Paneth cells in response to bacteria. *Nat Immunol*. 2000;1(2):113–8.
6. Pelaseyed T, Bergström JH, Gustafsson JK, Ermund A, Birchenough GMH, Schütte A, et al. The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunol Rev*. 2014;260(1). doi:10.1111/imr.12182.
7. van der Flier LG, Clevers H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu Rev Physiol*. 2009;71:241–60. doi:10.1146/annurev.physiol.010908.163145.
8. von Roon AC, Reese G, Teare J, Constantinides V, Darzi AW, Tekkis PP. The Risk of Cancer in Patients with Crohn's Disease. *Diseases of the Colon Rectum*. 2007;50(6):839–55. doi:10.1007/s10350-006-0848-z.
9. Bernstein CN, Blanchard JF, Kliwer E, Wajda A. Cancer risk in patients with inflammatory bowel disease: A population-based study. *Cancer*. 2001;91(4):854–62. doi:10.1002/1097-0142(20010215)91:4<854::AID-CNCR1073>3.0.CO;2-Z.
10. Eaden JA, Abrams KR, Mayberry JF. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut*. 2001;48(4):526–35. doi:10.1136/gut.48.4.526.
11. Burke KE, Naylor J, Campbell EJ, Ananthakrishnan AN, Khalili H, Richter JM. Interval Colorectal Cancer in Inflammatory Bowel Disease: The Role of Guideline Adherence. *Dig Dis Sci*. 2020;65(1):111–8. doi:10.1007/s10620-019-05754-9.
12. Choi C-HR, Bakir IA, Hart AL, Graham TA. Clonal evolution of colorectal cancer in IBD. *Nature Reviews Gastroenterology Hepatology*. 2017;14(4):218–29. doi:10.1038/nrgastro.2017.1.
13. Baran J, Gerner M, Haeussler M, Nenadic G, Bergman CM. pubmed2ensembl: a resource for mining the biological literature on genes. *PLoS One*. 2011;6(9):e24716. doi:10.1371/journal.pone.0024716.
14. Consortium GO. The Gene Ontology (GO) project in 2006. *Nucleic acids research*. 2006;34(Database issue):D322-D6.
15. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*. 2000;28(1):27–30.
16. Bin Zhao ZW, Zhang X, Zhao Y. Comprehensive analysis reveals a four-gene signature in colorectal cancer. *Translational Cancer Research*. 2020. doi:10.21037/tcr.2020.01.18.
17. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. 2014;8(Suppl 4):11. doi:10.1186/1752-0509-8-S4-S11.

18. Wagner AH, Coffman AC, Ainscough BJ, Spies NC, Skidmore ZL, Campbell KM, et al. DGldb 2.0: mining clinically relevant drug-gene interactions. *Nucleic Acids Res.* 2016;44(D1):D1036-44. doi:10.1093/nar/gkv1165.
19. Cotto KC, Wagner AH, Feng YY, Kiwala S, Coffman AC, Spies G, et al. DGldb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* 2018;46(D1):D1068-D73. doi:10.1093/nar/gkx1143.
20. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia.* 2017;19(8):649–58. doi:10.1016/j.neo.2017.05.002.
21. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba\_ identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol.* 2014;8(4):11.
22. Griffith M, Griffith OL, Coffman AC, Weible JV, McMichael JF, Spies NC, et al. DGldb: mining the druggable genome. *Nat Methods.* 2013;10(12):1209–10. doi:10.1038/nmeth.2689.
23. Guillot X, Prati C, Sondag M, Wendling D. Etanercept for treating axial spondyloarthritis. *Expert Opinion on Biological Therapy.* 2017;17(9):1173–81. doi:10.1080/14712598.2017.1347156.
24. Zhao S, Mysler E, Moots RJ. Etanercept for the treatment of rheumatoid arthritis. *Immunotherapy.* 2018;10(6):433–45. doi:10.2217/imt-2017-0155.
25. Lethaby A, Lethaby A, Lopez-Olivo MA, Lopez-Olivo MA, Maxwell L, Maxwell L, et al. Etanercept for the treatment of rheumatoid arthritis. *The Cochrane database of systematic reviews.* 2013;5;2013;(5):CD004525. doi:10.1002/14651858.CD004525.pub2.
26. Yang L, Ji S, Wang L, Zhang Y. Episcleritis in a patient with mucosal melanoma treated with interferon alfa-2b and radiotherapy: a case report. *J Med Case Rep.* 2018;12(1):388. doi:10.1186/s13256-018-1913-7.
27. Singh M, Gautam N, Gupta A, Kaur M. Interferon alfa-2b in the management of recurrent conjunctival papillomatosis. *Indian J Ophthalmol.* 2016;64(10):778–80. doi:10.4103/0301-4738.195013.
28. Kaliki S, Bejjanki KM, Desai A, Mohamed A. Interferon Alfa 2b for Ocular Surface Squamous Neoplasia: Factors Influencing the Treatment Response. *Semin Ophthalmol.* 2019;34(7–8):465.
29. Klepacki J, Klawitter J, Bendrick-Peart J, Schniedewind B, Heischmann S, Shokati T, et al. A high-throughput U-HPLC-MS/MS assay for the quantification of mycophenolic acid and its major metabolites mycophenolic acid glucuronide and mycophenolic acid acyl-glucuronide in human plasma and urine. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2012;883–884:113–9. doi:10.1016/j.jchromb.2011.07.021.
30. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;46(D1):D1074-D82. doi:10.1093/nar/gkx1037.
31. Isbell G, Levin B. Ulcerative colitis and colon cancer. *Gastroenterol Clin N Am.* 1988;17(4):773.
32. Levin B. Ulcerative colitis and colon cancer: Biology and surveillance. *J Cell Biochem.* 1992;50(S16G):47–50. doi:10.1002/jcb.240501109.

33. Pekow J, Meckel K, Dougherty U, Huang Y, Chen X, Almoghrabi A, et al. miR-193a-3p is a Key Tumor Suppressor in Ulcerative Colitis-Associated Colon Cancer and Promotes Carcinogenesis through Upregulation of IL17RD. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2017;23(17):5281–91. doi:10.1158/1078-0432.CCR-17-0171.
34. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel J-F. Ulcerative colitis. *The Lancet*. 2017;389(10080):1756–70. doi:10.1016/s0140-6736(16)32126-2.
35. Zundler S, Neurath MF. Integrating Immunologic Signaling Networks: The JAK/STAT Pathway in Colitis and Colitis-Associated Cancer. *Vaccines (Basel)*. 2016;4(1). doi:10.3390/vaccines4010005.
36. Deng F, Peng L, Li Z, Tan G, Liang E, Chen S, et al. YAP triggers the Wnt/beta-catenin signalling pathway and promotes enterocyte self-renewal, regeneration and tumorigenesis after DSS-induced injury. *Cell Death Dis*. 2018;9(2):153. doi:10.1038/s41419-017-0244-8.

## Figures

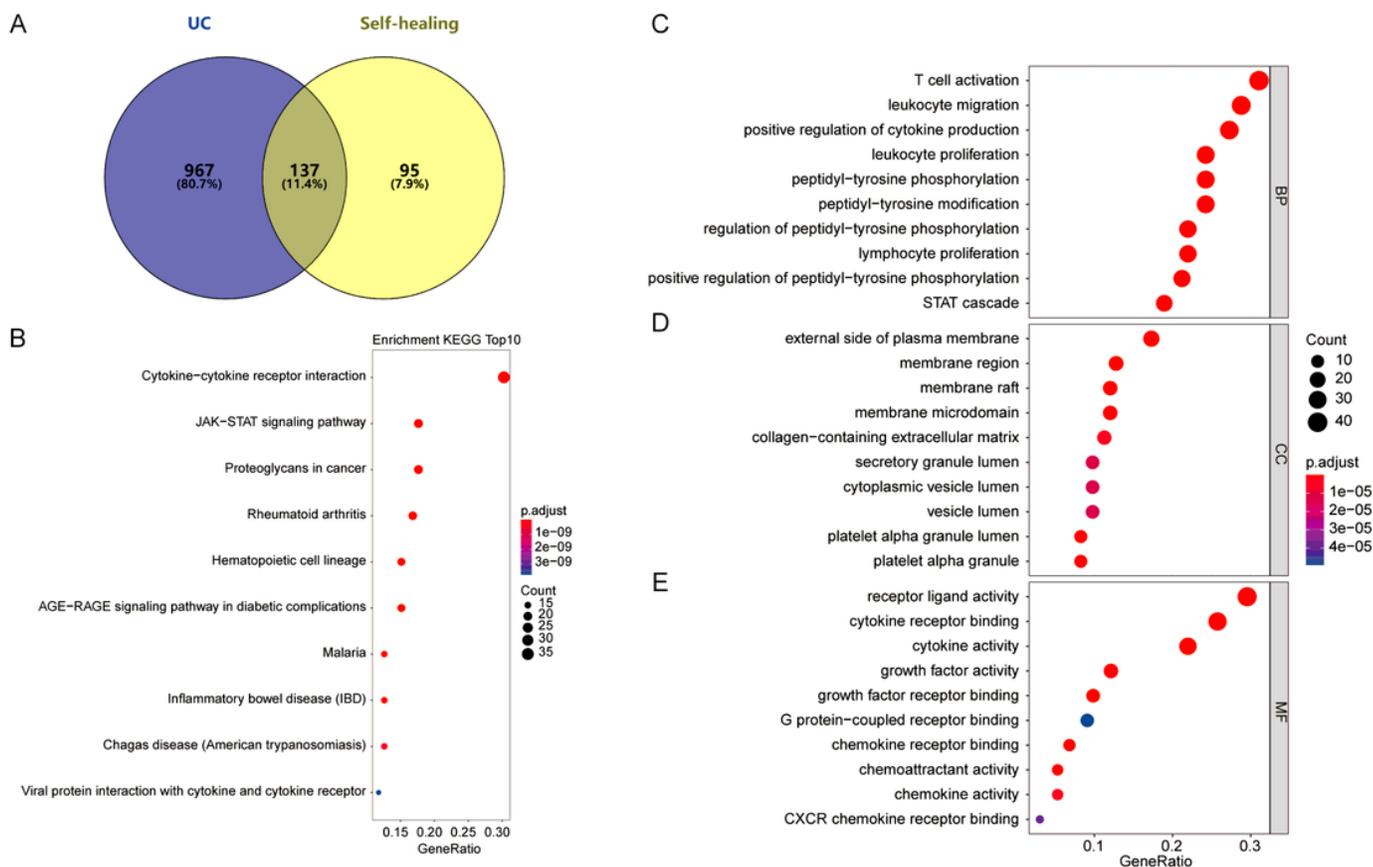


Figure 1

Identification of relative gene and Function and signal pathway enrichment analysis. (A) Venny diagram of overlapping genes between ulcerative colitis and self-healing. (B) The top ten significantly enriched signal pathways of common genes. (C-E) The top ten functional enrichment analysis of common genes in biological process (BP), cellular component (CC) and molecular function (MF), respectively.

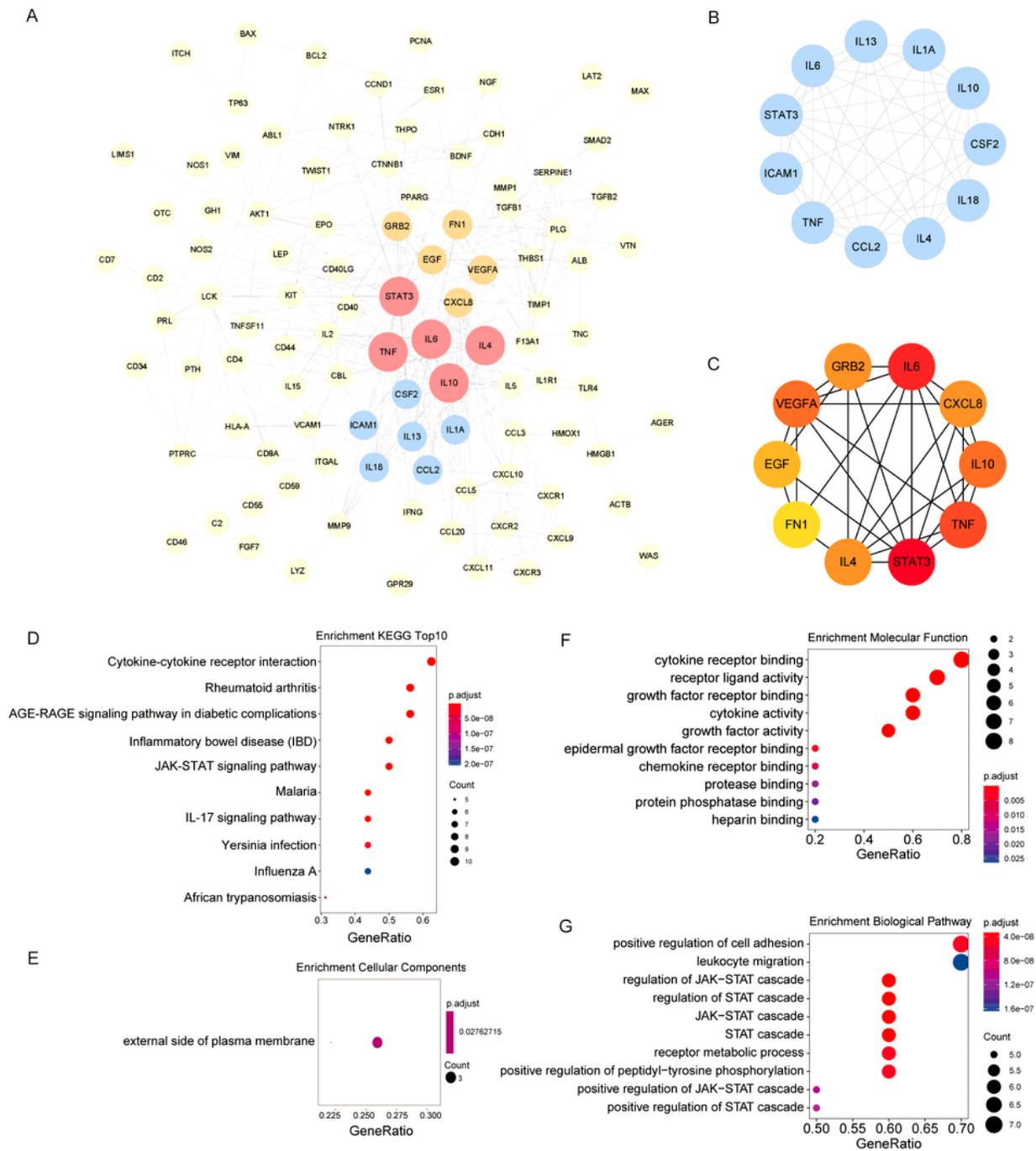


Figure 2

Protein-protein interaction (PPI) network and deeply analysis of GO and KEGG pathways. (A) Based on the STRING online database, 105 genes were filtered into PPI network. Node color: Skyblue indicates genes in cluster1; Orange indicates top 10 hub genes with the higher degree; Red indicates common genes both in cluster 1 and selected top 10 genes. (B) The most significant cluster module from the PPI network. (C) A total of 10 hub genes with higher degree selected in the PPI network. Different depth of the nodes represents the content of their degree in PPI network. (D) The top ten significantly enriched signal pathways of 16 significant clustered genes. (E) CC enrichment analysis of 11 genes in cluster module. (F-G) BP and MF enrichment analysis of 10 hub genes, respectively.

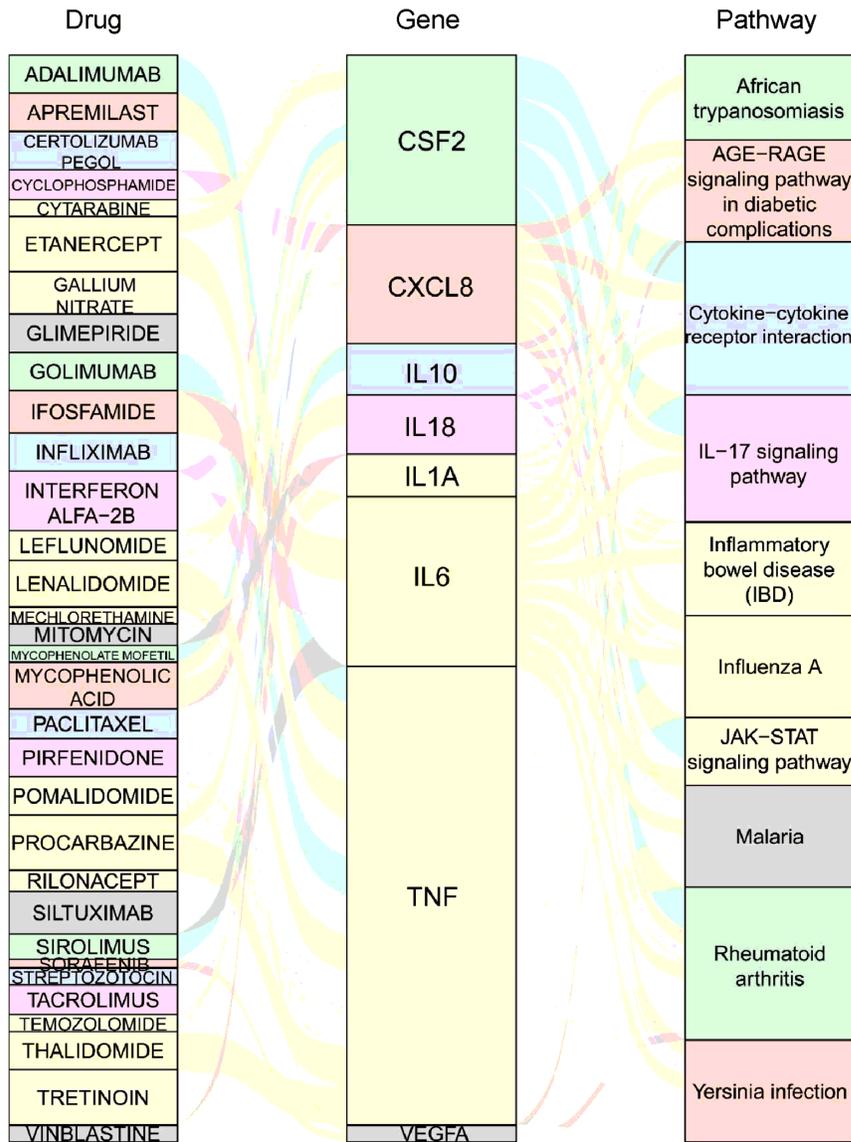
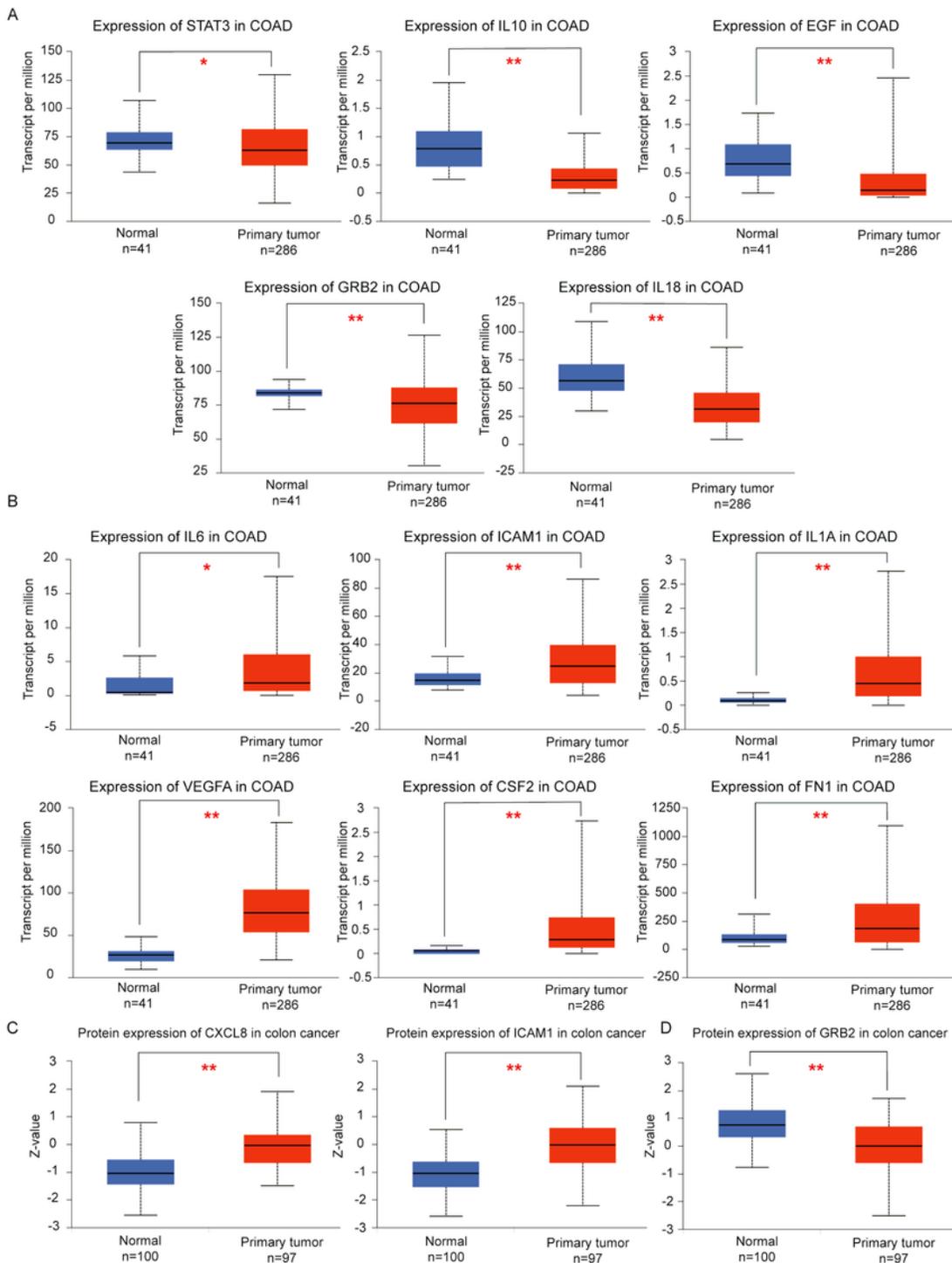


Figure 3

Sangquito of gene-drug-pathway network.



**Figure 4**

The validation of genes expression differences both on RNA and protein level. (A-B) Validation of the RNA expression of COAD datasets in TCGA. (T test; \* indicates  $p < 0.05$ ; \*\* indicates  $p < 0.01$ ) (A) STAT3, IL10, EGF, CRB2 and IL18 were upregulated in COAD compared with normal samples. (B) IL16, ICAM1, IL1A, VEGFA, CSF2 and FN1 were downregulated in COAD compared with normal samples. (C-D) Validation of the protein expression in colon cancer datasets in CPTAC. (Z-values represent standard deviations from

the median across samples for the given cancer type. Log2 Spectral count ratio values from CPTAC were first normalized within each sample profile, then normalized across samples. \*\* indicates  $Z < 0.01$ ) (C) CXCL8 and ICAM1 were upregulated in colon cancer compared with normal samples. (D) GRB2 were downregulated in colon cancer compared with normal samples.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementmaterial.docx](#)