Regular Article

Bioinformatics Analysis of Key Candidate Genes and Pathways in Ulcerative Colitis

Guangya Xu,^{a,#} Xueling Yan,^{a,#} Jie Chen,^b Xiaoheng Guo,^a Xiaolan Guo,^a Yong Tang,^{*,a,c} and Zheng Shi^{*,a}

^a College of Basic Medicine & Sichuan Industrial Institute of Antibiotics, Chengdu University; Chengdu 610106, China; ^b Central Laboratory of Clinical Medicine, Sichuan Academy of Medical Science and Sichuan Provincial People's Hospital; Chengdu 610000, China; and ^c Acupuncture and Tuina School, Chengdu University of Traditional Chinese Medicine; Chengdu 610075, China.

Received June 9, 2020; accepted August 5, 2020

Ulcerative colitis (UC) is chronic, idiopathic disease that affects the colon and the rectum and the underlying pathogenesis of UC remains to be known. The clinical drugs are mainly work based on anti-inflammation and immune system. However, most of them are expensive and have severe side effects. Therefore, identification of novel targets and exploring new drugs are urgently needed. In this study, several bioinformatics approaches were used to discover key genes and further in order to explore the pathogenesis of UC. Two microarray datasets, GSE38713 and GSE9452 were selected from NCBI-Gene Expression Omnibus database. Differentially expression genes (DEGs) were identified by using LIMMA Package of R. Then, we filtered clustered candidate genes into Gene Ontology (GO) and pathway enrichment analysis with the Database for Annotation, Visualization and Integrated Discovery (DAVID), KEGG pathway based on functions and signaling pathways with significant enrichment analysis. The protein–protein interaction (PPI) network was constructed by the Search Tool for the Retrieval of Interacting Genes/ Proteins (STRING) analysis, and visualized by Cytoscape and further analyzed by Molecular Complex Detection. Lastly, 353 up-regulated and 145 down-regulated genes were than recognized. After consulting a number of references and network degree analysis, four hub genes, namely FCGR2A, C3, INPP5A, and ACAA1 were identified, and these genes were mainly enriched in complement and coagulation cascades, mineral absorption, and Peroxisome Proliferator-Activated Receptor (PPAR) signaling pathways. In conclusion, this study would provide new clues for the pathogenesis and identification of drug targets of UC in the near future.

Key words ulcerative colitis; differentially expression gene; gene chip; network analysis

INTRODUCTION

Ulcerative colitis (UC), an inflammatory bowel disease, causes inflammation and ulcer of rectum and colon. And, it affects millions of people in the world and increases the risk of cancer. It mainly manifests as abdominal pain and diarrhea mixed with blood, weight loss, fever, and anemia may also occur. The complications may include ocular inflammation, joints or liver, and colon cancer.¹⁾ It is reported that the pathogenesis of UC involves the interaction of the immune system, environment, and susceptibility genes, but its exact mechanism is not yet known.²⁾ Various factors induce UC, such as genes, environment, and the infections. The current treatments of UC have serious effects because of the high doses and long-term administration. Therefore, exploration of the pathogenesis and discovering effective drugs with high efficiency and low side effects for UC is urgent demanded.

Gene chip, a gene detection technology, is utilized for drug target discovery for years. Gene chips can detect all expression information of all genes from one sample and is suitable for screening differentially expressed genes (DEGs).³⁾ There are so many core data with the wide application of gene chips, and most of them have been stored in public databases. The integration and reanalysis of these data can bring valuable clues for new research.⁴⁾ Notably, systems biology is a new in-

terdisciplinary study combining molecular biology and information technology. Integrating the microarray data and combining expression profiling techniques will provide a new clue for drug target discovery. To the best of our knowledge, bulks of studies successfully recognized appropriate targets for therapeutic intervention based on the microarray analysis.^{5,6)}

The raw gene chips in NCBI performed to help understand the basic molecular and genetic processes. Then, by using R software (version 3.5.2, <https://www.r-project.org/>) and packages of Bioconductor (<http://www.bioconductor.org/>) were applied to get the DEGs. Database for Annotation, Visualization, and Integrated Discovery (DAVID) is an on-line bioinformatics analysis tool that enables the systematic synthesis of genes or proteins in biological function analysis and annotation.⁷⁾ We take advantage of Gene Ontology (GO) and Genomes (KEGG) pathway to perform enrichment analysis of DEGs.⁸⁾ And, the Search Tool for the Retrieval of Interacting Genes database (STRING) can provide information regarding both direct and indirect protein interactions.⁹⁾

In this study, we selected and analyzed UC gene chips through various bioinformatics approaches, further clustered candidate targets based on functions and signaling pathways with significant enrichment analysis throughout DAVID and STRING online databases. Genes with high degree often play more vital roles in this sub-network¹⁰⁾ and finally there were four hub genes that were identified, namely FCGR2A, C3, INPP5A, and ACAA1. They were enriched in several signal

[#]These authors contributed equally to this work.

*To whom correspondence should be addressed. e-mail: tangyong@cducm.edu.cn; drshiz1002@hotmail.com

pathways, such as complement, coagulation cascades, mineral absorption, and peroxisome Proliferator-Activated Receptor (PPAR) signaling pathways. To sum up, our findings would provide a theoretical basis in biological function and molecular mechanisms for the occurrence and development of UC drug development.

MATERIALS AND METHODS

Microarray Data Proteins that connect with other proteins often exert similar gene expression patterns. In this study, in order identify genes whether co-expressed or not, microarray data was selected to identify genes to measure pair-wise co-expression gene in UC. We selected the cell chip and gene chip of the differential gene of UC. Two microarray data with different text conditions were selected. The UC and normal or adjacent mucosa tissue gene expression profiles of GSE38713 and GSE9452 were obtained from GEO-Datasets. GSE31056 is a cell chip, and the GSE23558 is a tissue chip. These genes were expressed both in cell and tissue level, which might be quite helpful in the clinical study. The microarray data of GSE38713 included a total of 43 biopsies: 13 healthy controls, 8 inactive UC, 7 non-involved active UC and 15 involved active UC respectively.¹¹⁾ The microarray data of GSE9452 was from GPL570 Platforms and include a total of 26 biopsies: 8 UC samples with macroscopic signs of inflammation, 13 UC samples without macroscopic signs of inflammation and five control subjects.¹²⁾ Then the R ggplot2 packages were applied to paint the Volcano to visualize those datasets.

Identification of DEGs in UC Firstly, the downloaded data in the series matrix file was preprocessed using the affy package (version 1.50.0) in software R (version 3.5.3, <https://www.r-project.org/>),¹³⁾ including normalization and expression calculation. Then, empirical Bayes method was used to select significant DEGs between UC samples and normal samples based on “limma” package of Bioconductor.¹⁴⁾ The RMA algorithm was used to calculate the background correction, standardization, and calculation of expression values. We identify the DEGs with the classical *t*-test, and we defined the $p < 0.05$ and $|\log FC| > 1$ as the cut-off criterion.

Gene Ontology and Pathway Enrichment Analysis Can-

didate genes functions and pathways enrichment were analyzed throughout multiple online databases, including DAVID and STRING analysis. Then, the construction of the protein interaction network was utilized by STRING and Cytoscape. DAVID (version 6.8; <https://david.ncifcrf.gov/>),¹⁵⁾ which is an online database with gene function, visualize, and integrated discovery function, was used for enrichment analysis of GO and KEGG pathways of the candidate DGEs, and identified the $p < 0.05$ as the cut-off criterion, and the most significant result was selected according to the value of the p value.

Modular Analysis and Significant Candidate Genes, PPI Network and Pathway Identification We used the online Database STRING 10.0 (<https://string-db.org/>) to develop DEGs-encode proteins and protein-protein interaction network (PPI), and the effective binding fraction of >0.4 was further analyzed. The Cytoscape 3.4.0 was used to construct a protein interaction relationship network and analyze the interaction relationship of the candidate DEGs encoding proteins in UC. The Molecular Complex Detection (MCODE) was used to analyze PPI network modules,¹⁶⁾ and MCODE scores >3 and the number of nodes >5 were set as cutoff criteria with the default parameters (Degree cutoff ≥ 2 , Node score cutoff ≥ 2 , K-core ≥ 2 and Max depth = 100). Finally, the node degree that means the numbers of inter-connections was used to filter hub genes. The core proteins and key candidate genes were which have the most node that may have important physiological regulatory functions.

RESULTS

Identification of DEGs in UC The microarray data of GSE38713 includes a total of 43 biopsies: 13 healthy controls, 8 inactive UC, 7 non-involved active UC and 15 involved active UC. The microarray data of GSE9452 includes a total of 26 biopsies: 8 UC samples with macroscopic signs of inflammation, 13 UC samples without macroscopic signs of inflammation, and five control subjects. We identify the DEGs with the classical *t*-test, and we defined the $p < 0.05$ and $|\log FC| > 1$ as the cut-off criterion. There are 1646 and 1284 of the two datasets respectively (Fig. 1A). Totally, there are 498 DEGs in total (Fig. 1B), and there are 353 up-regulated genes and 145 down-regulated genes respectively (Table 1).

Fig. 1. Identification of Differentially Expressed Genes in Two Cohort Profile Datasets (GSE38713 and GSE9452)

(A). Respective volcano plot of the two datasets. The red data-points represent genes with $|\log FC| > 1$ and $p < 0.05$. The black data-points represent genes with no significant difference in expression. (B). Venn plot of DEGs from GSE38713 and GSE9452. Blue color areas represented GSE38713 datasets. Red color areas represented GSE9452 datasets. The cross area meant the commonly changed DEGs. (Color figure can be accessed in the online version.)

Table 1. Four Hundred and Ninety Eight DEGs Were Identified from Two Profile Datasets, Including 353 Up-Regulated Genes and 145 Down-Regulated Genes in the UC Tissues Compared to Normal Tissues

Category	Genes name
Up-regulated	REG1A, REG3A, REG1B, SLC6A14, DEFA5, DUOX2, CHI3L1, CXCL1, S100A8, CXCL3, KYNU, PI3, LCN2, CXCL2, MMP3, TNIP3, VNN1, CXCL8, DUOX2, IDO1, IGHM, ALDOB, DEFA6, IGLV1-44, FOXQ1, IGHM, IGLJ3, MMP12, FCGR3B, CXCL9, IGKC, TCN1, NOS2, CD55, CXCL6, CXCL13, MMP1, FCRL5, MMP10, CCL11, NPTX2, BMS1P20, REG4, IGLC1, PDZK1IP1, ANXA1, CXCL5, LOC100293211, AZGP1, CXCL11, SERPINB5, IGHD, C4BPA, IGLJ3, MZB1, CD274, CCL20, SPINK4, KCND3, C4BPB, DMBT1, CD38, VSIG1, CFB, KLK10, FAM26F, OLFM4, IL1RN, FDCSP, CTHRC1, ERO1A, MLIP, SRGN, ANKRD36BP2, LY96, FKBP11, WISP1, DERL3, IGLL5, MMP9, CCL18, BCL2A1, MNDA, AGT, IGKV10R2-108, NCF2, PHLDA1, S100A9, IL1B, PLA2, MMP7, ZBP1, FCGR2A, IGKV10R2-118, CFI, PECAM1, IGK, PNOC, GBP1, LOXL2, COL4A1, MGP, COL1A2, IGFBP5, CXCL10, FAM30A, SLAMF7, SELL, DAPP1, FPR1, FCER1G, IL13RA2, ENTPD1, TIMP1, MALAT1, POU2AF1, SPP1, SPAG4, STAT1, RGCC, SCD, ADGRG6, SLC7A11, TNFRSF17, SAMSNI, CHST15, CD74, HLA-DRA, CEMIP, GLCC1, CTSK, PLEK, ST3GAL1, COL6A3, EAF2, LAX1, PLA2G7, KLHL5, CLDN2, CASP1, F2R, SLC6A6, HCAR3, IFIT3, COL1A1, ITGB2, VWF, S100P, CADM1, EVI2B, RBPMS, GREM1, NFKBIZ, LOC101929272, CD79A, VCAN, COL5A2, NAMPT, WNT5A, PFKFB3, TRIM29, RGS5, SRD5A3, CSTA, KCNN3, PDE4B, TDO2, RAC2, ELL2, TRIB2, C2CD4A, HLA-DMA, SEC14L1, BASP1, LPCAT1, ME1, TPK1, TGFB1, KLHL6, TAGAP, HLA-QB1, C2, THY1, SAMD9L, GZMB, FUT8, CTSC, SOCS1, PLA2G2A, CXCR2, BACE2, HLA-DPA1, PSMB9, PTGDS, MS4A1, APOL1, SPARC, CD163, PCSK1, TNC, CLEC4A, CDH11, C1S, HCLS1, LPGAT1, SPCS3, TLR8, SERPINA1, COL12A1, CYR61, PIM2, OSBP3, CCL4, F2RL2, TGM2, CD86, GBP5, PITPNC1, IKBIP, ZG16B, STS, CHST2, C3, CLDN1, IGFBP7, RAB31, TNFSF13B, CPA3, SERPINA3, ARNTL2, IFI16, IL7R, LYN, ISG20, WARS, BIRC3, LIPG, SLA, GPX8, CSGALNACT1, STOM, CSF2RB, NUCB2, THEMIS2, KDELR3, GMFG, PSAT1, ADAMTS2, CALU, IFITM2, SLAMF8, CD53, BHLHA15, CCDC69, FSTL1, NID1, CD27, LOC643733, CSRP2, ASRGL1, IRF4, P2RY8, ROBO1, CD44, FAM92A1, VNN2, TFPI, GLIPR1, LCP2, LUM, UBE2L6, ADGRL4, TMEM158, LPIN1, CPEB4, CLEC7A, IFITM3, CECR1, SELP, CDH3, LOC100996809//HLA-DRB4//HLA-DRB1, MYEOV, DOK3
Down-regulated	CLDN8, PCK1, AQP8, HMGCS2, ABCB1, TRPM6, UGT2A3, GUCA2B, GUCA2A, CA1, MT1M, SLC16A9, SLC26A2, CHP2, ABCG2, SIAE, SGK2, SLC51A, DEPDC7, RUNDC3B, GHR, CDKN2B, GBA3, SLC4A4, PADI2, ANPEP, SLC30A10, ADH1C, ABCA8, CKB, SLC3A1, HEPACAM2, CD177, DHRS11, CWH43, SLC16A1, CDHR1, LAMA1, PHLPP2, ADIRF, SLC22A5, EXPH5, HIGD1A, RHOU, APOBEC3B, ENTPD5, MCOLN2, SATB2-AS1, TUBAL3, BEST2, ABAT, PPID, EDN3, TMEM38B, VIPR1, CNTN3, SELENBP1, HSD17B2, NEDD4L, PRAP1, CAPN13, AMN, SLC17A4, TEX11, TSPAN7, SOWAHA, CNM2, CES2, SEMA5A, NK3, ACSF2, PKIB, LOC101929340, SLC35G1, NXPE4, THRB, MTIF, THRB, PRLR, AQR5, FAM213A, SCNN1B, CAMK2N1, ISX, P2RY1, MAOA, SRI, C2orf88, DPP4, MTMR11, RAVR2, SEMA6D, FAM162A, SLC38A4, SLC25A34, MUC12, PBLD, MEP1B, ARHGAP44, SLC13A2, MEP1A, PPARGC1A, ICF, PRDX6, FGFR2, MST1L, CLYBL, NAAA, DPP10-AS1, TMEM72, MS4A12, HLA-DRB4, AHCYL2, MIER3, AIFM3, SLC51B, ZNF704

The up-regulated genes were listed from the largest to the smallest of fold changes, and down-regulated genes were listed from the smallest to largest of fold changes.

Fig. 2. Gene Ontology Analysis and Significant Enriched GO Terms of DEGs in UC

GO analysis classified the DEGs into 3 groups (molecular function, biological process and cellular analysis classified the DEGs into 3 groups (molecular function, biological process and cellular component); (B) Significant Enriched Go Terms of DEGs in UC based on their functions. (A is up-regulated genes, and the B is down-regulated genes) (Color figure can be accessed in the online version.)

Fig. 3. Significantly Enriched Pathway Terms of DEGs in UC

DEGs functional and signaling pathway enrichment were conducted using online websites of KEGG PATHWAY and Gene Ontology analysis. (Color figure can be accessed in the online version.)

Table 2. Signal Pathway Enrichment Analysis of Differentially Expressed Genes

Pathway	Name	Gene count	p-Value
Up-regulated			
hsa05150	<i>Staphylococcus aureus</i> infection	14	5.62005E-11
hsa05323	Rheumatoid arthritis	15	3.47943E-09
hsa04610	Complement and coagulation cascades	13	1.70461E-08
hsa05133	Pertussis	13	4.5278E-08
hsa05140	Leishmaniasis	12	2.37893E-07
hsa04062	Chemokine signaling pathway	18	3.4668E-07
Down-regulated			
hsa04978	Mineral absorption	136	7.71844E-06
hsa03320	PPAR signaling pathway	15	7.7749E-06
hsa00280	Valine, leucine and isoleucine degradation	14	1.00144E-05
hsa00920	Sulfur metabolism	17	1.65221E-05
hsa04146	Peroxisome	14	3.54913E-05
hsa04530	Tight junction	7	4.73766E-05

DEGs Gene Ontology Analysis in UC DAVID is an online bioinformatics analysis tool. The DEGs utilize David 6.8 for enrichment analysis of GO and KEGG pathways (<http://www.genome.jp/kegg>) and we make the cut-off criterion as $p < 0.05$. The GO annotation of DEGs was divided into three parts: Biological process (BP), Cellular component (CC), and Molecular function (MF). GO analysis showed that the up-regulated genes were mainly enriched in biological processes such as extracellular exosome, extracellular regions, and extracellular space, (Fig. 2A) and the down-regulated genes were mainly enriched in extracellular exosome and mitochondrion (Fig. 2B).

Signaling Pathway Enrichment Analysis After the KEGG PATHWAY analysis, the up-regulation of genes mainly concentrated on signaling pathways such as chemokine signaling pathway, rheumatoid arthritis and *Staphylococcus aureus* infection; The down-regulated genes were mainly en-

Table 3. The Degrees of Candidate Genes in the Core PPI Network

Gene name	Degree
FCGR2A	35
C3	34
INPP5A	33
ACAA1	32
CD79A	31
CXCL6	31
HCAR3	31
ETFB	30
FCGR3B	30
IL8	30
GZMB	30
IL7R	29
CXCL5	28
PNOC	28
CD86	28
CD44	26
PLCE1	26
CXCL3	26
CXCL2	25
AGT	25
ANXA1	25
FPR1	24
ALDH5A1	24
CD38	23
CD27	23
PLCD3	21
MINPP1	21
ITPKA	21
ACOX1	21
PLCB4	21
MUT	20
ACADSB	20
ACADS	20
ACADM	20
PTEN	20
PLCD1	20

Fig. 4. Identification of Key Candidate Genes (A) Module 1 Protein Interactions and Related Gene Pathway Analysis Results; (B) Module 2 Protein Interactions and Related Gene Pathway Analysis Results

(Color figure can be accessed in the online version.)

riched in signaling pathways such as mineral absorption, sulfur metabolism and PPAR signaling pathway (Fig. 3, Table 2).

Key Candidate Genes and Pathways Identification with DEGs Protein-Protein Interaction Network and Modular Analysis STRING 10.0 (<https://string-db.org/>) was performed to develop DEGs-encode proteins and PPI, setting an effective binding fraction of >0.4 . There are 88 pathways in total and including 40 up-regulated and 48 down-regulated genes respectively. There were four hub genes with high degree scores that were screened through the String database, namely, FCGR2A, C3, INPP5A, and ACAA1 (Table 3). Cytoscape software MCODE plug-in screened a total of 2 significant modules; module 1 involved genes mainly enriched in signal pathways, including *S. aureus* infection, complement and coagulation cascades (Fig. 4A), and module 2 involved genes mainly Enriched in mineral absorption, PPAR signaling pathways (Fig. 4B).

DISCUSSION

UC is one of the main forms of inflammatory bowel disease in humans, and it becomes a heavy social burden due to its high incident rates and rise.¹⁷⁾ Currently, the clinical therapy of UC is amino salicylic acid (such as Mesalazine, willow *N*-Methyl Pyridine) and glucocorticoid drugs (Prednisone acetate, acetic acid and Dexamethasone). Accumulating evidence have proved that these drugs showed the effective mainly

by inhibiting nuclear factor-kappaB (NF- κ B) and transforming growth factor- β (TGF- β) signaling pathways, but these drugs presented poisonous effects after long term using and recurrence problems.¹⁸⁾ Therefore, fundamental breakthroughs are urgently needed for the development of novel therapeutic techniques in UC.

To the best of our knowledge, numbers of studies were utilized to uncover key genes and biological processes by network-based biological analysis. The systems biology approaches might provide new direction for identifying new targets and drugs in future UC therapeutics.¹⁹⁾ In current study, a series of bioinformatics analysis were applied to recognize the key genes in UC. Several classic hub genes and novel hub genes were identified in this core PPI network, and this would provide a high priority list of potential drug targets. According the network analysis and previous reference, FCGR2A, C3, INPP5A, and ACAA1 were finally recognized throughout microarray data and several online databases. Furthermore, we found that these genes mainly have an impact on biological processes such as immune responses, inflammatory responses, and kinase-mediated signaling pathways. More importantly, most studies have suggested that UC is associated with abnormal immune responses,²⁰⁾ oxidative damage,²¹⁾ infection²²⁾ and *etc.*, Such results are consistent with our analysis. After the KEGG pathway analysis, we found the up-regulation of genes were mainly concentrated on signaling pathways, such as chemokine signaling pathway, rheumatoid arthritis, and

Staphylococcus aureus infection; The down-regulated genes were mainly enriched in signaling pathways such as mineral absorption, sulfur metabolism, and PPAR signaling pathways.

The chemokine is a kind of small protein family, and it can trigger a variety of inflammation responses, and more and more chemokines have been proved to be involved in the pathogenesis of ulcerative, blocking the interactions between chemokines can be used as a means to treat UC.²³⁾ The PPAR signaling pathway takes part in the immune mechanism, and numerous studies have improved the immune mechanism plays a vital role in the incidence of UC.²⁴⁾ The colonic epithelium is an important constituent of the colon mucosal immune system. Previous studies have found that the expression of PPAR- γ decreased and correlated with the severity of the disease in the patients with UC and relative normal mucosa around inflammation, and the more severe the inflammation, the lower the expression.²⁵⁾

Among them, FCGR2A is a member of the immunoglobulin Fc receptor gene family which found on the surface of many immune response cells.²⁶⁾ FCGR2A is participating in phagocytosis and clearance of immune complexes. Besides, it selectively splicing produces multiple transcript variants and participates in phagocytic processes such as transport, catabolism of somatic cell, development of osteoclast differentiation organic systems, Fc γ R-mediated phagocytosis, immune system signaling pathways.^{27,28)} Previous studies have shown that the FCGR2A is the susceptibility gene in Japan, Korea, and Caucasian people. When the T>C mutation occurs in rs1801274 located in the fourth exon region, it can reduce the ability of recognition immune complex in the FCGR2A receptor, and thus affecting the activity of a variety of immune cells.²⁹⁾ C3 plays a core role in the complement system, and its activation is essential for classical and alternative complement activation pathways. Moreover, previous studies have argued that the immunomodulatory dysfunction is the direct pathogenesis of UC. Additionally, the intestinal flora is an important activation factor in such immune damage, and the man who has a defective C3 is more susceptible to bacterial infections.³⁰⁾

INPP5A is mainly involved in inositol phosphate metabolism and phosphatidylinositol signaling pathways and plays a key role in phosphoinositide dephosphorylating, protein dephosphorylating, and inositol phosphate metabolism. INPP5A has a low level of skin squamous cell carcinoma.³¹⁾ Previous studies have reported that the INPP5A in the top 50 marginal interactions related to UC. Besides, Kotaro Kigal *et al.* found that INPP5A can be suppressed by the miR-210 expression which increases the proliferation of gastric epithelium during chronic *Helicobacter pylori* infection. Those findings suggested that INPP5A is a hub gene in inflammatory bowel disease.³²⁾ Furthermore, ACAA1 is an effective enzyme encoded in the peroxisomal β -oxidation system and it can selective splicing yield multiple transcript variants and reduce Zellweger syndrome when it defected. And, ACAA1 mainly is involved in fatty acid, Valine, Leucine, and Isoleucine degradation, and PPAR signal pathways.³³⁾ To our knowledge, few studies were shown on the relationship between abnormal expression of ACAA1, therefore it could become a promising target for the treatment of UC.

CONCLUSION

In this study, multiple bioinformatics approaches, including microarray data, GO annotation and network analysis were integrated together to identify hub genes and according to the network characteristics, four key genes, namely FCGR2A, C3, INPP5A, and ACAA1 were identified in UC. Taken together, our findings would bring a new perspective for understanding the pathogenesis and pave a new road for drug development of UC.

Acknowledgments This work was supported in part by National Natural Science Foundation of China (81803561); the Science Project of the Health planning Committee of Sichuan Province (19PJ001); and supported by Sichuan Science and Technology Program (2020YJ0487).

Conflict of Interest The authors declare no conflict of interest.

REFERENCES

- 1) Gajendran M, Loganathan P, Jimenez G, Catinella AP, Ng N, Umaphathy C, Ziade N, Hashash JG. A comprehensive review and update on ulcerative colitis. *Dis. Mon.*, **65**, 100851 (2019).
- 2) Wanderås MH, Moum BA, Høivik ML, Hovde Ø. Predictive factors for a severe clinical course in ulcerative colitis: Results from population-based studies. *World J. Gastrointest. Pharmacol. Ther.*, **7**, 235–241 (2016).
- 3) Hanai T, Hamada H, Okamoto M. Application of bioinformatics for DNA microarray data to bioscience, bioengineering and medical fields. *J. Biosci. Bioeng.*, **101**, 377–384 (2006).
- 4) Dudoit S. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Stat. Sin.*, **12**, 111–139 (2002).
- 5) Hackl H, Sanchez Cabo F, Sturn A, Wolkenhauer O, Trajanoski Z. Analysis of DNA microarray data. *Curr. Top. Med. Chem.*, **4**, 1357–1370 (2004).
- 6) Petryszak R, Burdett T, Fiorelli B, *et al.* Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic Acids Res.*, **42** (D1), D926–D932 (2014).
- 7) Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57 (2009).
- 8) Chen L, Zhang YH, Wang S, Zhang Y, Huang T, Cai YD. Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways. *PLOS ONE*, **12**, e0184129 (2017).
- 9) Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.*, **39** (Database), D561–D568 (2011).
- 10) Shi Z, Sun R, Yu T, Liu R, Cheng LJ, Bao JK, Zou L, Tang Y. Identification of novel pathways in plant lectin-induced cancer cell apoptosis. *Int. J. Mol. Sci.*, **17**, 228 (2016).
- 11) Planell N, Lozano JJ, Mora-Buch R, Masamunt MC, Jimeno M, Ordás I, Esteller M, Ricart E, Pique JM, Panes J, Salas A. Transcriptional analysis of the intestinal mucosa of patients with ulcerative colitis in remission reveals lasting epithelial cell alterations. *Gut*, **62**, 967–976 (2013).
- 12) Olsen J, Gerds TA, Seidelin JB, Csillag C, Bjerrum JT, Troelsen JT, Nielsen OH. Diagnosis of ulcerative colitis before onset of inflammation by multivariate modeling of genome-wide gene expression

- data. *Inflamm. Bowel Dis.*, **15**, 1032–1038 (2009).
- 13) Gautier L, Cope L, Bolstad BM, Irizarry RA. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, **20**, 307–315 (2004).
 - 14) Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, **43**, e47 (2015).
 - 15) Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57 (2009).
 - 16) Bandettini WP, Kellman P, Mancini C, Booker OJ, Vasu S, Leung SW, Wilson JR, Shanbhag SM, Chen MY, Arai AE. MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. *J. Cardiovasc. Magn. Reson.*, **14**, 83 (2012).
 - 17) Ananthakrishnan AN. Epidemiology and risk factors for IBD. *Nature Rev. Gastroenterol. Hepatol.*, **12**, 205–217 (2015).
 - 18) Feuerstein JD, Moss AC, Farraye FA. Ulcerative colitis. *Mayo Clin. Proc.*, **94**, 1357–1373 (2019).
 - 19) Liu J, Lian X, Liu F, Yan X, Cheng C, Cheng L, Sun X, Shi Z. Identification of novel key targets and candidate drugs in oral squamous cell carcinoma. *Curr. Bioinform.*, **14**, 1–10 (2019).
 - 20) Lee SH, Kwon JE, Cho ML. Immunological pathogenesis of inflammatory bowel disease. *Intest. Res.*, **16**, 26–42 (2018).
 - 21) Pereira C, Grácio D, Teixeira JP, Magro F. Oxidative stress and DNA damage: implications in inflammatory bowel disease. *Inflamm. Bowel Dis.*, **21**, 2403–2417 (2015).
 - 22) Pillet S, Pozzetto B, Roblin X. Cytomegalovirus and ulcerative colitis: place of antiviral therapy. *World J. Gastroenterol.*, **22**, 2030–2045 (2016).
 - 23) Werner L, Elad H, Brazowski E, Tulchinsky H, Vigodman S, Kopylov U, Halpern Z, Guznergur H, Dotan I. Reciprocal regulation of CXCR4 and CXCR7 in intestinal mucosal homeostasis and inflammatory bowel disease. *J. Leukoc. Biol.*, **90**, 583–590 (2011).
 - 24) Tian X, Peng Z, Luo S, Zhang S, Li B, Zhou C, Fan H. Aesculin protects against DSS-induced colitis through activating PPAR γ and inhibiting NF- κ B pathway. *Eur. J. Pharmacol.*, **857**, 172453 (2019).
 - 25) Orchard TR, Thiyagaraja S, Welsh KI, Wordsworth BP, Gaston JSH, Jewell DP. Clinical phenotype is related to HLA genotype in the peripheral arthropathies of inflammatory bowel disease. *Gastroenterology*, **118**, 274–278 (2000).
 - 26) Morgan AW, Robinson J, Barrett JH, Martin J, Walker A, Babbage SJ, Ollier WER, Gonzalezgay MA, Isaacs JD. Association of FCGR2A and FCGR2A-FCGR3A haplotypes with susceptibility to giant cell arteritis. *Arthritis Res. Ther.*, **8**, R109 (2006).
 - 27) McGovern D, Gardet A, Torkvist L, Goyette P, Essers J, Taylor KD, Neale BM, Ong RT, Lagace C, Li C. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat. Genet.*, **42**, 332–337 (2010).
 - 28) Asano K, Matsushita T, Umeno J, Hosono N, Takahashi A, Kawaguchi T, Matsumoto T, Matsui T, Kakuta Y, Kinouchi Y, Shimosegawa T, Hosokawa M, Arimura Y, Shinomura Y, Kiyohara Y, Tsunoda T, Kamatani N, Iida M, Nakamura Y, Kubo M. A genome-wide association study identifies three new susceptibility loci for ulcerative colitis in the Japanese population. *Nat. Genet.*, **41**, 1325–1329 (2009).
 - 29) Vogelpoel LTC, Baeten D, De Jong EC, Den Dunnen J. Control of Cytokine Production by Human Fc Gamma Receptors: Implications for Pathogen Defense and Autoimmunity. *Front. Immunol.*, **6**, 79 (2015).
 - 30) Marques da Costa P, Correia L, Correia LA. The complexity of renal involvement in IBD—C3 glomerulopathy in ulcerative colitis. *Inflamm. Bowel Dis.*, **24**, e4–e5 (2018).
 - 31) Sekulic A, Kim SY, Hostetter G, Savage S, Einspahr JG, Prasad A, Sagerman P, Curiellewandrowski C, Krouse RS, Bowden GT, Warneke J, Alberts DS, Pittelkow MR, DiCaudo D, Nickoloff BJ, Trent JM, Bittner M. Loss of inositol polyphosphate 5-phosphatase is an early event in development of cutaneous squamous cell carcinoma. *Cancer Prev. Res.*, **3**, 1277–1283 (2010).
 - 32) Kiga K, Mimuro H, Suzuki M, Shinozakiushiku A, Kobayashi T, Sanada T, Kim M, Ogawa M, Iwasaki YW, Kayo H, Fukuda-Yuzawa Y, Yashiro M, Fukayama M, Fukao T, Sasakawa C. Epigenetic silencing of miR-210 increases the proliferation of gastric epithelium during chronic *Helicobacter pylori* infection. *Nat. Commun.*, **5**, 4497 (2014).
 - 33) Sordillo JE, Sharma S, Poon A, Lasky-Su J, Belanger K, Milton DK, Bracken MB, Triche EW, Leaderer BP, Gold DR, Litonjua AA. Effects of endotoxin exposure on childhood asthma risk are modified by a genetic polymorphism in ACAA1. *BMC Med. Genet.*, **12**, 158 (2011).