BMC Genomics

RESEARCH

Open Access

Gene expression regulation and polyadenylation in ulcerative colitis via long-chain RNA sequencing

Zhe Zhang[1], Dan Li[2], Shihang Zheng[1], Changqing Zheng[1], Hao Xu[1] and Xueqing Wang[1*]

Abstract

Background Ulcerative colitis (UC) is an immune-mediated chronic intestinal disease, with a pathogenesis that remains incompletely understood. The purpose of this study is to analyze the difference of gene expression between UC patients and healthy controls using Oxford Nanopore Technology's long-read RNA sequencing (ONT-RNA-seq) and to explore how alternative polyadenylation (APA) site selection contributes to UC pathogenesis.

Methods Colon tissue samples from UC and normal controls (NC) were collected, and total RNA was extracted and sequenced using ONT-RNA-seq technology. Various bioinformatics analyses were performed, including differential expression gene (DEG) analysis, functional enrichment analysis, APA site analysis, and prediction miRNAs and RNA binding proteins (RBPs) targets, to explore the molecular mechanism underlying UC.

Results ONT-RNA-seq analysis revealed that the expression levels of ACSF2, NPY, SLC26A3, BRINP3, and PKLPP2 were significantly lower in UC patients compared to the NC group, while the expression levels of CCL20, CCL21, CD55, IDO1, LCN2, NOS2, CCL11, OLFM4, ANXA1, REG1A, S100A9, SLPI, SPINK1, and AGR2 were significantly higher. Functional enrichment analysis showed that DEGs were closely related to immune and inflammatory responses, which in turn are related to many challenges in the diagnosis and treatment of UC. Mechanistically, APA site selection was found to contribute to the regulation of gene expression in UC, and some APA genes were identified as potential regulators of miRNAs and RBPs. Vene diagram revealed significant overlap between miRNA- and RBP-targeted genes and DEGs, suggesting that APA genes may modulate genes expression in UC through miRNA and RBP targeting. Additionally, five key APA genes——CD38, NCALD, SMIM31, GPX7, and SWAP70——were identified as potentially playing crucial role in UC pathogenesis.

Conclusions This study provides new insights into the molecular mechanisms of UC through ONT-RNA-seq technology, especially in gene expression regulation and APA site selection.

Keywords ONT-RNA-seq, Ulcerative colitis, APA, RBPs

Background

Ulcerative colitis (UC) is a chronic inflammatory bowel disease primarily affecting the colon and rectum, characterized by a persistent or recurrent inflammation that leads to intestinal dysfunction [1]. Although the exact cause of UC remains unclear, it is believed to result from a combination of genetic, environmental, microbial, or immune factors [2]. The diagnostic process of UC relies on a combination of clinical, laboratory, imaging,

*Correspondence:
Xueqing Wang
280869661@qq.com
[1] The Second Department of Gastroenterology, Shengjing Hospital of China Medical University, Shenyang, Liaoning 110004, China
[2] Department of Gastroenterology, The First Affiliated Hospital of China Medical University, Shenyang, Liaoning 110001, China

endoscopic, and histological criteria [3]. Unfortunately, even experienced clinicians may struggle to make definitive diagnosis. Conventional treatments include 5-aminosalicylic acid, corticosteroids, and immunomodulators. Since the early twenty-first century, various biotherapies and small molecule drugs have shown efficacy in patients with moderate to severe UC. Despite these advances, by the end of this year, only about 40% of patients are expected to achieved clinical remission, and 20–30% of patients still need surgery at some point during the course of the disease [4]. Therefore, it is crucial to explore the pathogenesis of UC to better control disease progression, optimize the targeted treatments, and improve patient outcomes.

RNA sequencing is a high-throughput sequencing technique used to study RNA expression in cells or tissues [4]. It allows for the analysis of gene expression levels, splicing variations, non-coding RNA expression, and other critical molecular information [5, 6]. In disease research, RNA sequencing is widely utilized to explore pathogenesis, identify diagnostic markers, and uncover therapeutic targets. For example, a recent study has used full-length sequencing to characterize the HCoV-229E virus, a member of the coronavirus family that contains the largest known RNA genome. This technique leverages defective interfering RNA (DI-RNA) to analyze transcripts in vitro using full-length cDNA [7]. RNA-seq has also been applied in cancers such as pancreatic ductal adenocarcinoma. where it can measure tumor mutational burden (TMB), a potential biomarker for immune checkpoint therapy and cancer prognosis [8]. Additionally, RNA-seq has enable the detection of mutation in the MET proto-oncogene and the isocitrate dehydrogenase 1 (IDH1) gene, paving the way for improved treatments for lung adenocarcinoma and chondrosarcoma [9, 10]. In monogenic neuromuscular disorders like congenital muscular dystrophy (CMD), RNA-seq has identified variants in regulatory upstream gene regions [11]. Although long-chain RNA sequencing (ONT-RNA-seq) technology is rarely applied in UC research, it holds significant potential due to its ability to capture longer reads and directly direct detection of RNA modifications and unedited transcripts, offering new insights into molecular mechanisms underlying diseases like UC.

In recent years, alternative polyadenylation (APA) has emerged as a key mechanism in the regulation of gene expression [12]. APA can affect the formation of the 3' end of RNA, thereby affecting the stability of RNA, ribosome loading, and translation efficiency, all of which can impact disease development [12]. It has been demonstrated that influenza A virus (IAV) strains, such as PR8 and Cal09, can induce alternative polyadenylation (APA) in host cells, contributing to the antiviral response. The loss of APA leads to increased inflammatory responses and reduced virulence in mice [13], suggesting that APA involved in the regulation of immune response. Therefore, this study systematically analyzes gene expression and APA site selection in UC patients and healthy controls using ONT-RNA-seq technology, aiming uncover the molecular pathological mechanisms of UC and provide a scientific foundation for identifying new biomarkers and therapeutic targets.

Methods

Sample collection and processing

From March 2023 to May 2023, mucosal specimens were collected from patients diagnosed with UC (UC = 7; ages: 30,34,62,32,53,31,35; gender:3 male,4 female; region: China, Liaoning Province;Severity: active moderate ulcerative colitis patients, the modified Mayo score is 5–7 points;Sampling locations: 3 cases in left colon and 4 cases in extensive colon;Health status: Chronic relapsing type patients, the course of disease is 2–5 years. All patients were ineffective in oral treatment with 5-aminosalicylic acid) and a normal control group (NC = 7; ages: 55,35,58,32,57,60,32; gender: 5 male,2 female; region: China, Liaoning Province;Health status: non-inflammatory, non-neoplastic and pedunculated colon polyps;

Sampling locations: Normal intestinal mucosal tissue adjacent to the polyps which was removed after polypectomy surgery in patients with colon polyps) at our hospital. The sample were obtained from lesion sites during colonoscopy, while the NC samples came from normal mucosa surrounding resected colon polyps. All specimens were immediately frozen in liquid nitrogen tank for 30 min and then stored at −80 °C. Inclusion criteria: (1) UC diagnosis confirmed pathology and (2) active UC mucosa. Exclusion criteria encompassed: (1) incomplete clinical data, (2) severe comorbidities or other malignancies, or (3) patients in UC remission. This study was conducted in accordance with the ethical principles of the Declaration of Helsinki. The study protocol was reviewed and approved by the Medical Ethics Committee of the Sheng jing Hospital of China Medical University (approval number: 2023PS602K, on the 10th of March 2023), and all patients who participated in this study or their families understood its approach. Informed consent for participation in the study has been obtained from the patients.

RNA extraction

Total RNA was extracted from colon tissue using TRIzol reagent (Invitrogen) followingthe manufacturer's instructions. The RNA purity and concentration of RNA were assessed with a Nanodrop 2000 spectrophotometer (Thermo Scientific), and integrity was verifiedwith an

Agilent 2100 bioanalyzer (Agilent Technologies). Only samples with an RNA integrity number (RIN) above 7were used for subsequent ONT-RNA-seq sequencing.

ONT-RNA-seq sequence

One microgram of total RNA was used to construct cDNA libraries with a cDNA-PCR sequencing kit (SQK-PCS109; ONT). Products were ligated with ONT adapters using T4 DNA ligase (NEB) and then sequenced on the PromethION platform at Biomarker Technology Company (Beijing, China). Raw reads were filtered (minimum read quality score = 7, minimum length = 500 bp), with ribosomal RNAs removed through rRNA database mapping. Full-length non-chimeric (FLNC) transcripts were mapped to the reference genome (GRCh38_release95) with minimap2 (version 2.7-r654). Mapped reads were further processed using the cDNA-Cupcake package (min-coverage = 85%, min-identity = 90%) and fusion candidates were identified by multi-locus mapping. Transcripts were compared against known annotations (hg38) using Gffcompare (version 0.1.26). Alternative splicing events, such as intron retention (IR), exon skipping (ES), alternative acceptor (AA), and mutually exclusive exon (MEE), were identified via AStalavista (version 3.0). Simple sequence repeats (SSR) were detected using MIcroSAtellite (version 2.1), while APA analysis was performed using TAPIS (version 1.2.1). Differential expression analysis was conducted using the DESeq2 R package (version 1.6.3), with differentially expressed genes (DEGs) defined as those with an adjusted $P < 0.01$ and fold change (FC) ≤ -1.5 or ≥ 1.5.

Functional enrichment analysis

Gene ontology (GO) enrichment analysis for DEGswas implemented using the DAVID Bioinformatics Resources (version 6.8) tool and and results were visualized using GO plot R packages (version 1.0.2). GO terms included biological process (BP), cellular component (CC), and molecular function (MF). The top eight BP terms were statistically analyzed at $P < 0.05$ and FC ≤ -1.5 or ≥ 1.5.

Targeted miRNA and RBP prediction

Potential miRNA targets and RNA binding proteins (RBPs) interacting with APA genes were predicted using the TargetScan and RBPDB databases, respectively.

qRT-PCR

Total RNA was extracted using the TRIzol method (Invitrogen, Carlsbad, CA, USA) then reverse-transcribed into cDNA using the PrimeScrip™ RT reagent kit (Takara Biotechnology, Dalian, China). qRT-PCR was conducted on anABI7900 real-time PCR system (Applied Biosystems, Carlsbad, California, USA) using a Sybr Premix ex Taq Kit (Takara Biotechnology, Dalian, China). Reaction conditions were as follows: pre-denaturation at 75 °C for 120 s, denaturation at 90 °C for 5 min,annealing at 60 °C for 60 s, extension at 72 °C for 30 s for 40 cycles. Relative expression was calculated using the 2-ΔΔCT method.

Statistical methods

Statistical software was performed using SPSS v27.0.Normally distributed data were expressed as mean ± SD. For intergroup comparisons, one-way ANOVA was used for normally distributed data, while nonparametric tests were used otherwise. A P-value < 0.05 was considered statistically significant.

Results

Long-read sequencing reveals the regulation of gene expression in UC

The difference in gene expression between UC patients and NC patients was analyzed using ONT-RNA-seq technology. A total of 67,192 genes (62,710 known and 252,798 novel) and 278,629 transcripts (10,884 known and 25,831 novel) were identified (Figure S1A). Thermographic analysis revealed that the Spearman correlation coefficient between the UC and NC samples was inversely proportional to the original gene (Fig. 1A) and the original transcripts (Figure S1B). The number of DEGs (Fig. 1B) and transcripts (Figure S1C) is displayed in the bar chart. The 10 most enriched GO terms (biological process) and KEGG pathways for DEGs in the UC and NC groups were illustrated (Fig. 1C–D). The key enriched GO terms (biological process) in the UC and NC groups were "inflammatory response," "immune response," and "innate immune response" (Fig. 1E). Compared to the NC group, the UC group showed lower expression of ACSF2, NPY, SLC26A3, BRINP3, and PKLPP2, while the expression of CCL20, CCL21, CD55, IDO1, LCN2, NOS2, CCL11, OLFM4, ANXA1, REG1A, S100A9, SLPI, SPINK1, and AGR2 were higher (Fig. 1F). Results from the Q-PCR experiment confirmed that ACSF2, NPY, SLC26A3, BRINP3, and PKLPP2 were significantly downregulated in the UC group, whereas CCL20, CCL21, CD55, IDO1, LCN2, NOS2, CCL11, OLFM4, ANXA1, REG1A, S100A9, SLPI, SPINK1, and AGR2 were significantly upregulated (Figure S2).

Long-read sequencing recapitulates APA site selection in UC

Analysis of APA selection sites revealed significant differences between the UC and NC samples. The bar graph shows a significantly higher number of APA sites in UC samples compared to NC samples, indicating a shift in APA site selection (Fig. 2A).In addition, the sequence dynamics diagram at around the 50 nt region of the

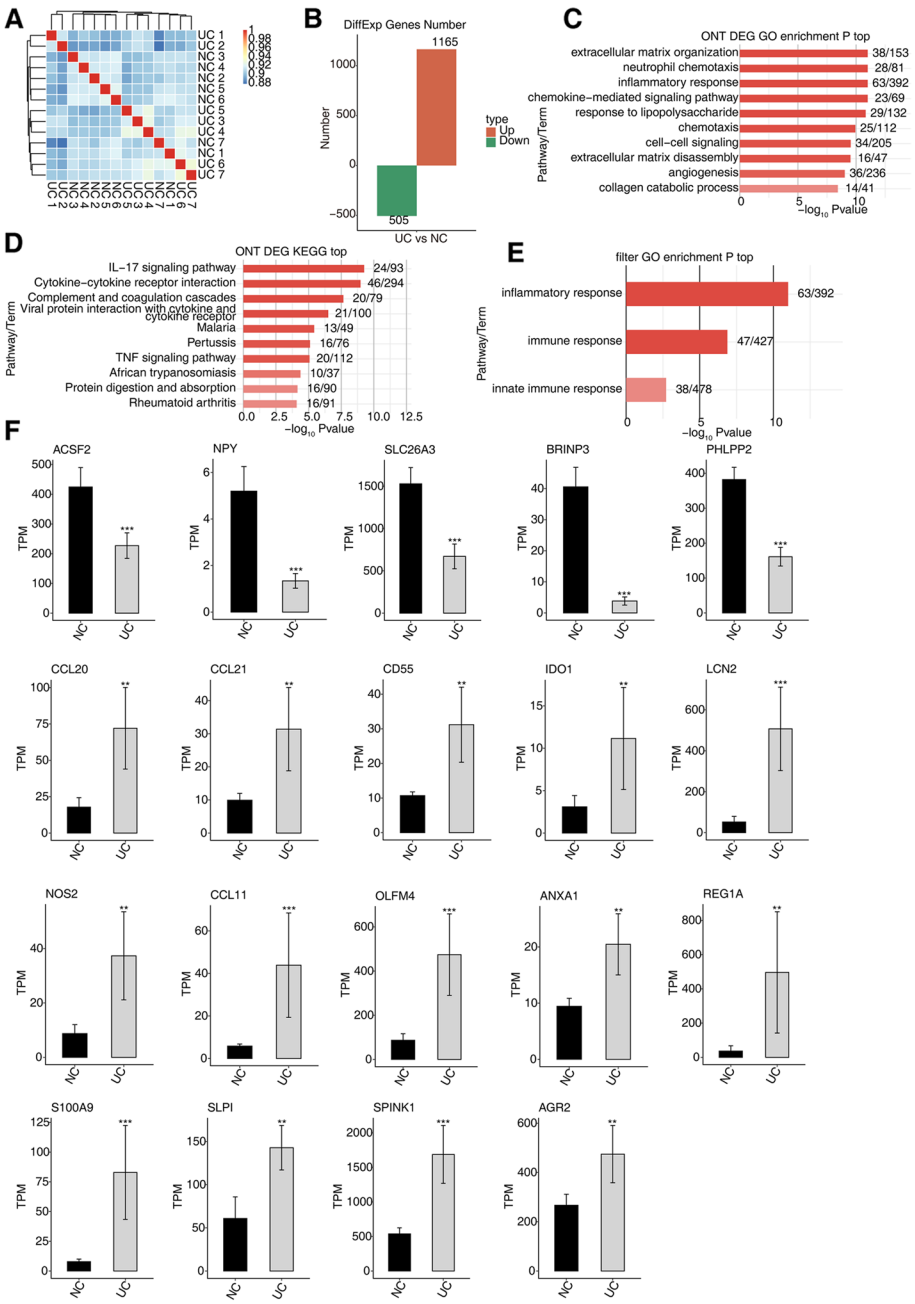

Fig. 1 Long-read sequencing reveals the regulation of gene expression in ulcerative colitis. **A** Heat map of Spearman correlation coefficients of raw gene counts for UC and NC samples. **B** Bar plot showing the number of differentially expressed genes. **C** The 10 most enriched GO terms (biological process) illustrated for DEGs in the UC and NC groups. **D** The 10 most enriched KEGG pathways illustrated for DEGs in the UC and NC groups. **E** The key enriched GO terms (biological processes) illustrated for DEGs in the UC and NC groups. **F** Bar plot showing the expression pattern and statistical difference of the DEGs. Error bars represent the mean ± SEM. $^{*}p \le 0.05$, ** $p \le 0.01$,*** $p \le 0.001$

poly(A) site reveals the potential sequence characteristics of APA site selection (Fig. 2B).The scatter plot further highlights the differences in APA sites between UC and NC samples, suggesting that these shifts may contribute to the differential regulation of gene expression associated with UC pathogenesis.(Fig. 2C). The top 10 most enriched GO terms (biological process) for genes with shorter and longer APA genes were illustrated in Fig. 2D and 2E. For genes with shorter APA sites (Fig. 2D), the most enriched GO terms include "protein phosphorylation," "cell division," "peptidyl-serine phosphorylation," "protein-containing complex localization," "endoplasmic

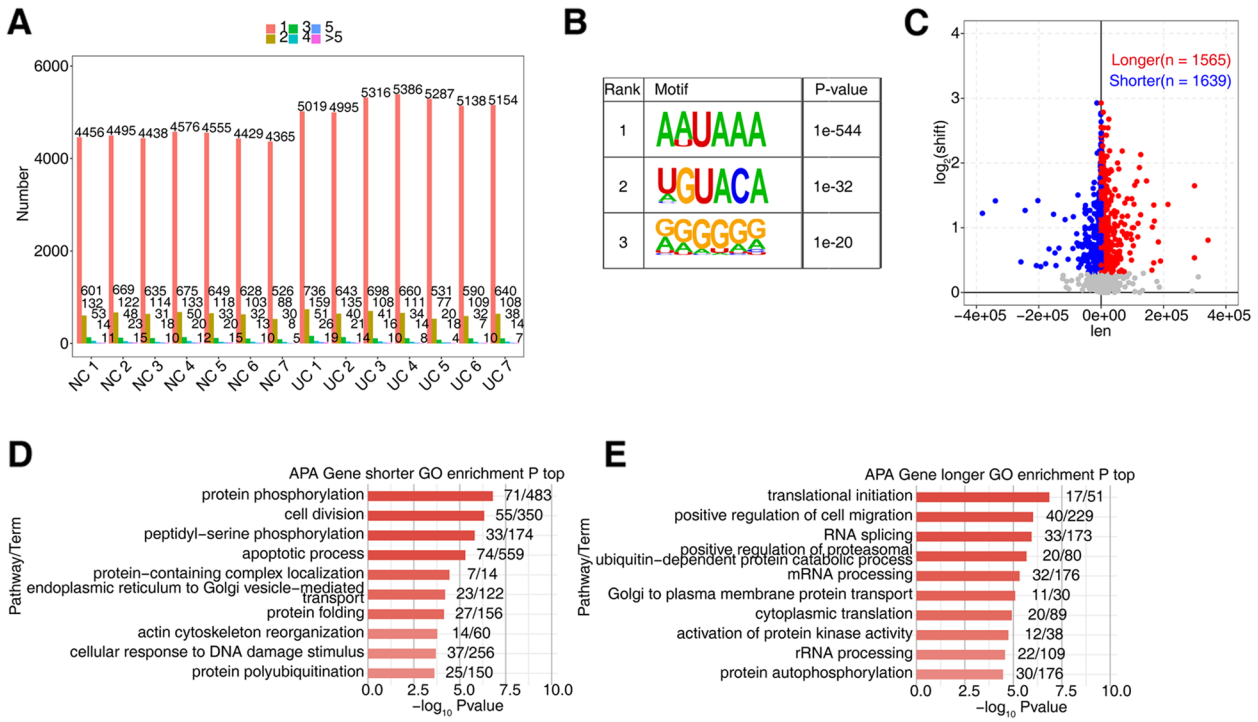

Fig. 2 Long-read sequencing recapitulates alternative polyadenylation (APA) site selection in ulcerative colitis. **A** Bar plot showing the polyadenylation sites in the UC and NC samples (in duplicate). **B** Motif map around 50 nts of the poly(A) site. **C** Scatter plots showing APA sites of difference. **D** The 10 most enriched GO terms (biological process) illustrated for the shorter APA gene. **E** The 10 most enriched GO terms (biological process) illustrated for the longer APA gene

reticulum to Golgi vesicle-mediated transport". These processes are crucial for regulating cell cycle progression, apoptosis, protein transport. In contrast, genes with longer APA sites (Fig. 2E) are primarily enriched in biological processes such as "translational initiation," "positive regulation of cell migration," "RNA splicing," "positive regulation of proteasomal ubiquitin-dependent protein catabolic process," and "mRNA processing". These processes indicate an involvement in protein synthesis, RNA regulation, and protein degradation, which could influence gene expression dynamics. The 10 most enriched KEGG pathways were illustrated for shorter and longer APA genes (Figure S3A–B). These findings suggest that both shortening and lengthening APA sites are involved in UC, impacting various biological processes that could be relevant to UC progression.

APA site selection and gene expression regulation in UC

A Venn diagram illustrates the overlap between APA genes corresponding to upregulated DEGs (466) and downregulated DEGs (229) (Fig. 3A). Enrichment analysis of these overlapping genes identified the10 most enriched Goers (biological process) (Fig. 3B-C) and KEGG Pathways (Figure S4A–B). For upregulated DEGs, the most enriched GO terms include "positive regulation of cell migration," "signal transduction," "protein autophosphorylation," "response to hydrogen peroxide," "platelet-derived growth factor receptor signaling pathway," and "DNA damage checkpoint signaling." These processes are crucial for regulating cellular signaling, migration, oxidative stress response, DNA repair mechanisms and apoptosis, all of which play important roles in maintaining cellular homeostasis and responding to environmental stressors. For downregulated DEGs, the most enriched GO terms include "protein phosphorylation," "epidermal growth factor receptor signaling pathway," "cellular response to cAMP," "positive regulation of protein localization to plasma membrane," "cellular response to insulin stimulus," "intracellular signal transduction." These processes play a critical role in regulating intracellular signaling, cytoskeletal organization, metabolic activity, and transport mechanisms, all of which are essential for cellular communication, metabolic homeostasis, and maintaining the integrity of key physiological systems. In addition, the bar graphs depicting the read distribution and expression patterns of the GPX7, SWAP70, CD38, NCALD, and SMIM31 transcripts further validated that APA site selection could regulate gene expression in UC (Fig. 3D; Figure S4C).

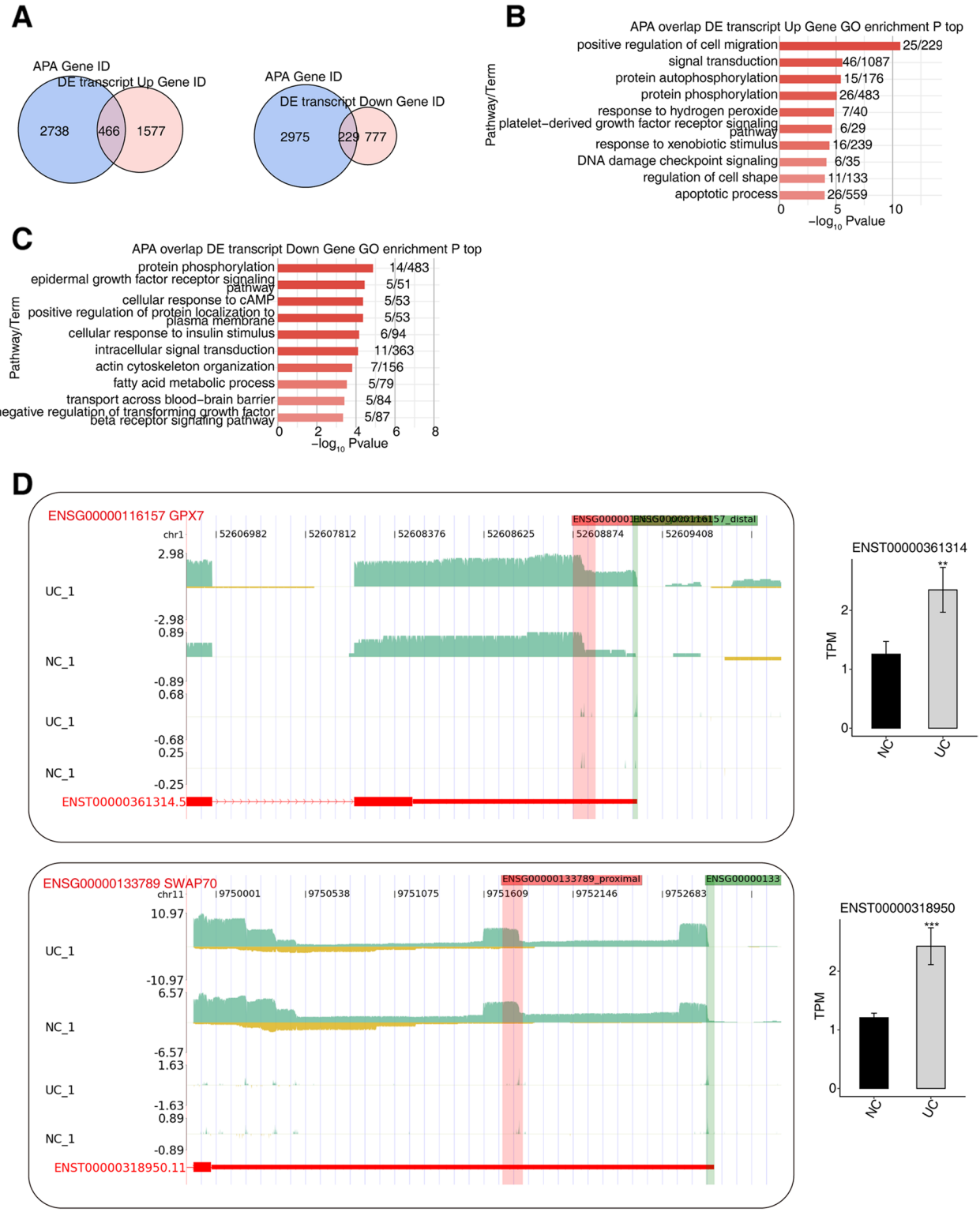

Fig. 3 Ulcerative colitis affects alternative polyadenylation (APA) site selection and the regulation of gene expression. **A** Venn diagram showing the number of overlapping APA genes and genes corresponding to the upregulated differentially expressed transcripts. **B** The 10 most enriched GO terms (biological process) illustrated for overlapping genes corresponding to the upregulated DE transcripts. **C** The 10 most enriched GO terms (biological process) illustrated for overlapping genes corresponding to the downregulated DE transcripts. **D** Reads distribution showing GPX7, transcript ENST00000361314, and SWAP70, transcript ENST00000318950. Bar plot showing the expression of transcript. * $p \leq 0.05$, ** $p \leq 0.01$,*** $p \leq 0.001$

Prediction of targeted miRNAs

To further explore the how APA site selection regulates gene expression and influence biological processes, we predicted the target miRNAs of APA genes. Figure 4A presents a network diagram showing that miRNA-140-3a is targeted by APA gene CD38, while miR-145-5p is targeted by APA gene SWAP70. These miRNAs can affect the expression of numerous other genes, indicating that APA genes may regulate gene expression indirectly by targeting miRNAs. In Fig. 4B, the Venn diagram shows that 35 mRNAs targeted by these miRNAs overlap with DE transcripts. To confirm whether APA genes regulate UC pathogenesis through miRNA targeting, we analyzed the expression of the target mRNAs in UC group compared to NC group. Figure 4C displays a heatmap of the expression levels of overlapping transcripts, and revealing that expression levels of SWAP70 and CD38 in UC were significantly higher than in the NC group. These findings confirmed that APA genes may regulating gene expression through miRNA in pathogenesis of UC.

Prediction of targeted RBPs

In addition to miRNAs, RBPs also play a role in regulating gene expression. To explore whether APA genes can influence gene expression by targeting RBPs, we conduct further investigations. Figure 5A presents a network diagram illustrating that a large number of RBPs are targeted by multiple APA genes (CD38, NCALD, SMIM31, GPX7, and SWAP70), highlighting the complex interaction between APA genes and RBP. Figure 5B features an upset diagram showing the intersection between the gene corresponding to the DE transcript and RBP regulated by the APA genes, revealing that 24 DE transcripts overlap with APA genes targeted RBPs. Figure 5C displays a heatmap of the expression level of these overlap transcripts, showing that, compared with the control group, four transcripts were downregulated and 20 were upregulated.

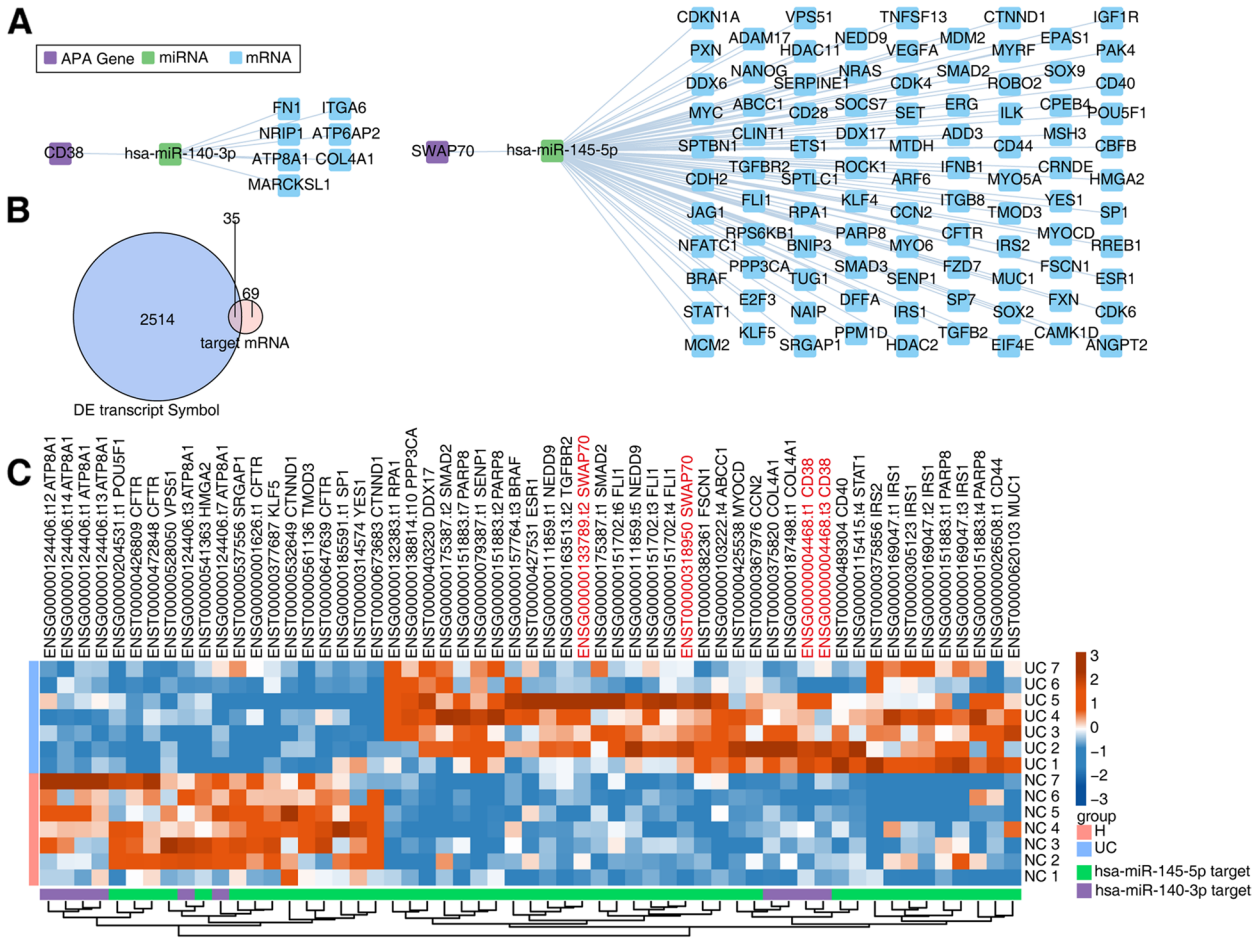

Fig. 4 Prediction of targeted miRNAs. **A** Network diagram showing all miRNAs regulated by the APA gene and the targeted mRNAs of miRNAs. **B** Venn diagram showing the overlap number of the targeted mRNAs of miRNAs and genes corresponding to the DE transcripts. **C** Heat map showing the expression level of overlap transcripts

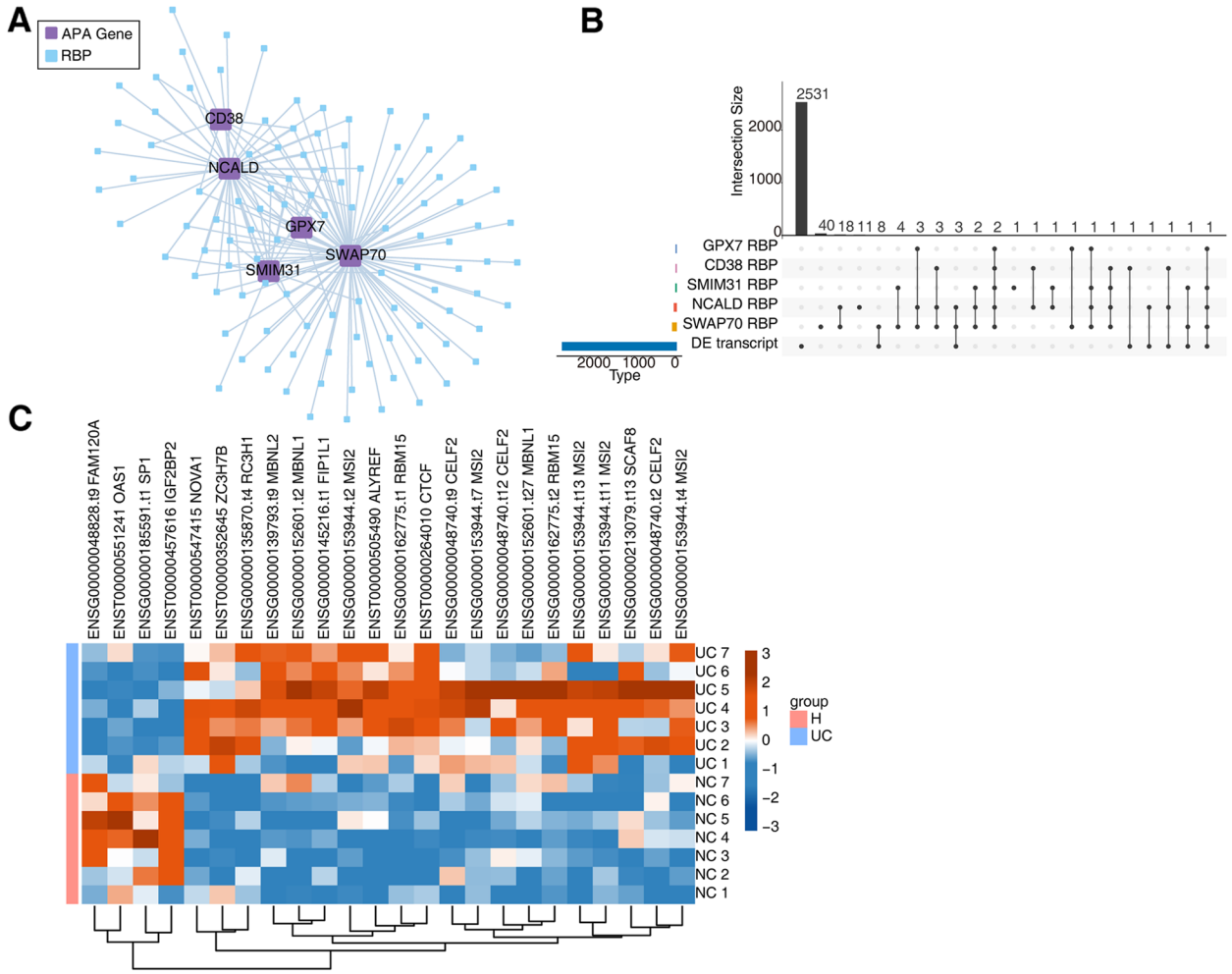

Fig. 5 Prediction of targeted RBPs. **A** Network diagram showing all the RBP regulated by the APA gene. **B** Upset plot showing the intersection of genes corresponding to the DE transcripts and RBPs regulated by the APA gene. **C** Heat map showing the expression level of overlap transcripts

These findings further confirm that APA genes may regulating gene expression through RBPs in UC.

Discussion

UC is a chronic condition with a prolonged course that is difficult to cure. The World Health Organization has classified it as a refractory disease. UC treatment faces challenges such as interindividual heterogeneity and the presence of complex complications. In this study, we identified DEGs in UC compared to NC using ONT-RNA-seq. We also analyzed APA site selection in UC and found that APA selection may contribute to UC progression by regulating gene expression.

ONT-RNA-seq is the most important part of the human cell mapping project [14]. It overcomes the limitation of traditional sequencing technology in studying cell specificity within tissues, allowing for a more precise exploration of gene expression [15]. Our findings revealed that DEGs are mainly involved in inflammatory response, immune response, as well as signal transduction in the intestinal mucosa of UC patients, which aligns with previous research.

For instance, RNA sequencing analyses have identified key immune cell genes in UC, with DEGs closely associated with immune and inflammatory responses immune and inflammatory response [16, 17]. We discovered more than a dozen up-regulated and down-regulated genes related to UC. We take SLC26A3(down-regulated), CCL20, CD55, CCL11 and ANXA1 for example. SLC26A3 is a transmembrane glycoprotein that transports chloride ions across the cell membrane in exchange for bicarbonate ions. Down-regulation of chloride transporter SLC26A3 or down-regulated in adenoma (DRA) in colonocytes has recently been linked to the pathogenesis of UC. Because exaggerated immune responses are one of the hallmarks of UC, many previous studies were

undertaken to define the mechanisms by SLC26A3 loss of DRA relays signals to immune cells to increase susceptibility to inflammation [18–21]. CCL20 induced release of cytokines from PBMCs. Stimulation with CCL20 combined with TNF increased IL-1β release from PBMCs. By attracting additional immune cells, as well as inducing proinflammatory IL-1β release from immune cells, CCL20 may protract the inflammatory response in ulcerative colitis [22–24]. Expression of CD55 is enhanced on colonic epithelial cells of patients with UC, and stool CD55 concentrations are increased in patients with active disease. Cytokines are known to modulate CD55 expression in various human cells, and lesions of UC reveal altered profiles of cytokine production [25]. CCL11, also known as eotaxin-1, is described as an eosinophil chemoattractant, which has been implicated in allergic and Th2 inflammatory diseases. We have reported that CCL11 is significantly increased in the serum of inflammatory bowel disease (IBD) patients, colonic eosinophils are increased and correlate with tissue CCL11 levels in ulcerative colitis patients [26–29]. Accumulating evidence supports an important role of ANXA1 in facilitating resolution of UC inflammation and mucosal wound repair. Up-regulation of Annexin-A1(ANXA1) in individuals with UC may promote mucosal homeostasis. Importantly, ANXA1 expression was not limited to cells infiltrating the lamina propria but was also detected in epithelial cells lining the intestinal crypts [30–34]. Our study results related to the above genes are consistent with the conclusions of previous literature, and once again confirmed the above conclusions.

Recent findings have shown that in nearly 70% of human genes, cleavage and polyadenylation can occur at multiple sites through a process called APA [35]. As a form of co-transcriptional gene regulation, APA significantly enhances the diversity of mRNA transcripts [36]. previous studies have highlighted the role of APA in various diseases. For example, Singh et al. [37] shown that intron APA subtypes are widely expressed in immune cells and participate in the development of B cells, resulting in the production of truncated proteins lacking a functional C-terminal domain. In addition, the number of intron APA subtypes in multiple myeloma cells decreased. This may lead to the progression of multiple myelomas and is a factor related to the short progression-free survival [38]. However, there has been relatively little research on the role of APA in UC. In this study, we observed significant changes in APA site selection in UC patients compared to controls. These changes may affect mRNA stability and translation efficiency, influencing protein expression and cell function. We also found that APA sites could regulate gene expression by targeting miRNAs or RBPs. Our research fills this gap, suggesting that altered APA site selection may be an important factor in UC pathogenesis.

In addition, we identified five key genes—CD38, NCALD, SMIM 31, GPX 7, and SWAP 70 —— that may play crucial role in pathogenesis of UC. NCALD is a protein involved in calcium signal transmission. Although it mainly plays a role in the nervous system, it may also do so in the pathogenesis of UC, especially related to the function of intestinal smooth muscle and the regulation of the intestinal nervous system [39]. The function of SMIM31 in UC is unclear. However, considering that it is a membrane protein, it may be related to cell signal transmission, cell adhesion, or cell barrier function [40, 41]. GPX7 is an antioxidant enzyme that is related to resisting oxidative stress and may play a role in protecting intestinal cells from oxidative damage [42]. CD38 is a surface molecule that is usually associated with cell activation and signal transmission in the immune system. Previous studies have pointed out that the increase in CD38 expression on the surface of immune cells in patients with inflammatory bowel disease may be related to the severity of the disease [43]. One study shows circulating TIGITnegCD38 + TM frequencies were higher in Crohn's disease patients with high plasma IFN-γ and a more severe disease course. TIGITnegCD38 + TM were highly enriched in HLA-DR + and ex-Th17/Th1-like cells, high producers of IFN-γ. Cultures of healthy-adult-stimulated TM identified IL-12 as the only IBD-related inflammatory cytokine to drive the pathogenic ex-Th17-TIGITnegCD38 + phenotype [44]. SWAP70 is a kind of protein involved in intracellular signal transduction, which may be related to regulating the migration and activation of immune cells. In UC, the appropriate migration and response of immune cells is very important for controlling inflammation [45]. It has been speculated that CD38 and SWAP70 may be involved in regulating the inflammatory response in UC.

To improve outcome, itis essential to constantly improve biological treatment strategies, shifting focus from the short-term treatment response to the ultimate goal of disease clearance. A variety of markers can be used to predict patients' response to different treatments. Some study results showed after supplementing high-dose butyric acid, inflammation in mice was reduced, the expression of SLC26A3, ZO-1, Occludin and Claudin-1 increased, the expression of Claudin-2 decreased, and the epithelial barrier function improved in colon tissue. After using butyric acid or HDAC inhibitors (TSA, SAHA) to treat lipopolysaccharide-stimulated Caco2bbe cells in in vitro experiments, the expression of SLC26A3 and tight junction proteins increased significantly which effectively inhibited the progression of inflammation [18, 46–48]. From above

existing study of therapeutic methods, we could make efforts to further study the therapeutic drugs of molecular targets, such as CCL20, CD55, CCL11,CD38 and ANXA1. Also, paying attention to safety factors would be necessary.

To advance UC precision medicine, it's necessary to improved disease classification, a deepen understanding of the natural history of disease, optimize the design of cohort studies and clinical trials, and integrate multiomic, clinical, environmental, and other relevant data. Due to the limitation of the small sample size of our study, as well as the possible selection bias, our results may not fully represent the situation of all UC patients of different race or areas. We could only provide reference for researchers of Asian populations, especially Chinese people. Although some key genes and regulatory mechanisms have been identified, these findings need to be validated in larger sample sizes and independent cohorts.

In this study, we constructed a gene expression network involving multilevel regulation by analyzing DEGs, APA sites, and predicting miRNA and RBP interactions. This comprehensive analysis method revealed not only the role of DEGs or APA changes, but also their interactions within a broader regulatory network. Generally speaking, the latest long-chain RNA sequencing technology has provided new insights into the molecular mechanism of UC, particularly in gene expression and APA sites selection. These findings not only enhance our understanding of the pathogenesis of UC but may also provide potential molecular targets for developing new diagnostic tools and therapeutic strategies. Future research should thorough explore the specific role of these regulatory mechanisms in UC and other inflammatory diseases, as well as how to accurately regulate these mechanisms to treat UC.

Abbreviations

UC	Ulcerative colitis
APA	Alternative polyadenylation
NC	Normal controls
DEG	Differential expression gene
RBPs	RNA binding proteins

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11346-x.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Supplementary Material 5.

Supplementary Material 6.

Supplementary Material 7.

Acknowledgements
Not applicable.

Authors' contributions
XQW: Conceptualization, Methodology, Software. ZZ: Data curation, Writing- Original draft preparation. SHZ and DL: Visualization, Investigation. CQZ: Supervision. HX: Software, Validation. All authors: Writing- Reviewing and Editing.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
The study protocol was reviewed and approved by the Medical Ethics Committee of the Shengjing Hospital of China Medical University (approval number: 2023PS602K, on the 10th of March 2023).

Consent for publication
All patients who participated in this study or their families understood its approach. Informed consent for participation in the study has been obtained from the patients.

Competing interests
The authors declare no competing interests.

Received: 1 November 2024 Accepted: 10 February 2025
Published online: 15 February 2025

References

1. Guo G, Tan Z, Liu Y, Shi F, She J. The therapeutic potential of stem cell-derived exosomes in the ulcerative colitis and colorectal cancer. Stem Cell Res Ther. 2022;13(1):138.
2. Zheng J, Wang S, Zhang T, Li H, Zhu M, Wei X, et al. Nogo-B inhibition restricts ulcerative colitis via inhibiting p68/miR-155 signaling pathway. Int Immunopharmacol. 2023;120: 110378.
3. Han J, Li Y, Zhang B, Liu H, Wu M, Zhang X. lncRNA TUG1 regulates ulcerative colitis through miR-142-5p/SOCS1 axis. Microb Pathog. 2020;143: 104139.
4. Samuel S, Ingle SB, Dhillon S, Yadav S, Harmsen WS, Zinsmeister AR, et al. Cumulative incidence and risk factors for hospitalization and surgery in a population-based cohort of ulcerative colitis. Inflamm Bowel Dis. 2013;19(9):1858–66.
5. Yamada S, Nomura S. Review of Single-Cell RNA Sequencing in the Heart. Int J Mol Sci. 2020;21(21):8345.
6. Khozyainova AA, Valyaeva AA, Arbatsky MS, Isaev SV, Iamshchikov PS, Volchkov EV, et al. Complex Analysis of Single-Cell RNA Sequencing Data. Biochemistry Biokhimiia. 2023;88(2):231–52.
7. Depledge DP, Srinivas KP, Sadaoka T, Bready D, Mori Y, Placantonakis DG, et al. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. Nat Commun. 2019;10(1):754.
8. Huang XT, Li X, Qin PZ, Zhu Y, Xu SN, Chen JP. Technical Advances in Single-Cell RNA Sequencing and Applications in Normal and Malignant Hematopoiesis. Front Oncol. 2018;8:582.
9. Wang L, Ge J, Lan Y, Shi Y, Luo Y, Tan Y, et al. Tumor mutational burden is associated with poor outcomes in diffuse glioma. BMC Cancer. 2020;20(1):213.
10. Seo JS, Ju YS, Lee WC, Shin JY, Lee JK, Bleazard T, et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. Genome Res. 2012;22(11):2109–19.

11. Belaya K, Rodríguez Cruz PM, Liu WW, Maxwell S, McGowan S, Farrugia ME, et al. Mutations in GMPPB cause congenital myasthenic syndrome and bridge myasthenic disorders with dystroglycanopathies. Brain : a journal of neurology. 2015;138(Pt 9):2493–504.
12. Wang G, Xie Z, Su J, Chen M, Du Y, Gao Q, et al. Characterization of Immune-Related Alternative Polyadenylation Events in Cancer Immunotherapy. Can Res. 2022;82(19):3474–85.
13. Bergant V, Schnepf D, de Andrade KN, Hubel P, Urban C, Engleitner T, et al. mRNA 3'UTR lengthening by alternative polyadenylation attenuates inflammatory responses and correlates with virulence of Influenza A virus. Nat Commun. 2023;14(1):4906.
14. Grünberger F, Ferreira-Cerca S, Grohmann D. Nanopore sequencing of RNA and cDNA molecules in Escherichia coli. RNA (New York, NY). 2022;28(3):400–17.
15. Massaiu I, Songia P, Chiesa M, Valerio V, Moschetta D, Alfieri V, et al. Evaluation of Oxford Nanopore MinION RNA-Seq Performance for Human Primary Cells. Int J Mol Sci. 2021;22(12):6317.
16. Dai YC, Qiao D, Fang CY, Chen QQ, Que RY, Xiao TG, et al. Single-cell RNA-sequencing combined with bulk RNA-sequencing analysis of peripheral blood reveals the characteristics and key immune cell genes of ulcerative colitis. World journal of clinical cases. 2022;10(33):12116–35.
17. Kim H, Kim HK, Hong D, Kim M, Jang S, Yang CS, et al. Identification of ulcerative colitis-specific immune cell signatures from public single-cell RNA-seq data. Genes & genomics. 2023;45(7):957–67.
18. Peng K, Xiao S, Xia S, Li C, Yu H, Yu Q. Butyrate Inhibits the HDAC8/NF-κB Pathway to Enhance Slc26a3 Expression and Improve the Intestinal Epithelial Barrier to Relieve Colitis. J Agric Food Chem. 2024;72(44):24400–16.
19. Liu M, Wang Y, Xu X, Guan G, Zhang S, Zhu S, et al. p-hydroxy benzaldehyde attenuates intestinal epithelial barrier dysfunction caused by colitis via activating the HNF-1β/SLC26A3 pathway. Front Pharmacol. 2024;15:1448863.
20. Camarillo GF, Goyon EI, Zuñiga RB, Salas LAS, Escárcega AEP, Yamamoto-Furusho JK. Gene Expression Profiling of Mediators Associated with the Inflammatory Pathways in the Intestinal Tissue from Patients with Ulcerative Colitis. Mediators Inflamm. 2020;2020:9238970.
21. Fujii T, Sato M, Hosoi K, Ohbayashi N, Ikuse T, Jimbo K, et al. Assessment of the Family History of Patients With Ulcerative Colitis at a Single Center in Japan. J Pediatr Gastroenterol Nutr. 2016;63(5):512–5.
22. Hanna DN, Smith PM, Novitskiy SV, Washington MK, Zi J, Weaver CJ, et al. SMAD4 Suppresses Colitis-associated Carcinoma Through Inhibition of CCL20/CCR6-mediated Inflammation. Gastroenterology. 2022;163(5):1334–50.e14.
23. Hashimoto K, Saigusa S, Araki T, Tanaka K, Okita Y, Fujikawa H, et al. Correlation of CCL20 expression in rectal mucosa with the development of ulcerative colitis-associated neoplasia. Oncol Lett. 2013;6(5):1271–6.
24. Skovdahl HK, Damås JK, Granlund AVB, Østvik AE, Doseth B, Bruland T ,et al. C-C Motif Ligand 20 (CCL20) and C-C Motif Chemokine Receptor 6 (CCR6) in Human Peripheral Blood Mononuclear Cells: Dysregulated in Ulcerative Colitis and a Potential Role for CCL20 in IL-1β Release. International journal of molecular sciences. 2018; 19(10).
25. Kohno H, Mizuno M, Nasu J, Makidono C, Hiraoka S, Inaba T, et al. Stool decay-accelerating factor as a marker for monitoring the disease activity during leukocyte apheresis therapy in patients with refractory ulcerative colitis. J Gastroenterol Hepatol. 2005;20(1):73–8.
26. Chen J, Zhou Y, Sun Y, Yuan S, Kalla R, Sun J, et al. Bidirectional Mendelian Randomisation Analysis Provides Evidence for the Causal Involvement of Dysregulation of CXCL9, CCL11 and CASP8 in the Pathogenesis of Ulcerative Colitis. J Crohns Colitis. 2023;17(5):777–85.
27. Bergemalm D, Andersson E, Hultdin J, Eriksson C, Rush ST, Kalla R, et al. Systemic Inflammation in Preclinical Ulcerative Colitis. Gastroenterology. 2021;161(5):1526–39.e9.
28. Polosukhina D, Singh K, Asim M, Barry DP, Allaman MM, Hardbower DM, et al. CCL11 exacerbates colitis and inflammation-associated colon tumorigenesis. Oncogene. 2021;40(47):6540–6.
29. Yang L, Bian Y, Li Z, Yan Y, Li J, Li W, et al. Identification of potential biomarkers and pathways in ulcerative colitis with combined public mRNA and miRNA expression microarray data analysis. J Gastrointest Oncol. 2019;10(5):847–58.
30. James JP, Nielsen BS, Christensen IJ, Langholz E, Malham M, Poulsen TS, et al. Mucosal expression of PI3, ANXA1, and VDR discriminates Crohn's disease from ulcerative colitis. Sci Rep. 2023;13(1):18421.
31. Deng J, Zhao N, Lv LP, Ma P, Zhang YY, Xu JB, et al. Integrated analysis of multiple microarray studies to establish differential diagnostic models of Crohn's disease and ulcerative colitis based on a metalloproteinase-associated module. Front Immunol. 2022;13:1022850.
32. Kourkoulis P, Michalopoulos G, Katifelis H, Giannopoulou I, Lazaris AC, Papaconstantinou I, et al. Leucine-rich alpha-2 glycoprotein 1, high mobility group box 1, matrix metalloproteinase 3 and annexin A1 as biomarkers of ulcerative colitis endoscopic and histological activity. Eur J Gastroenterol Hepatol. 2020;32(9):1106–15.
33. Sena AA, Pedrotti LP, Barrios BE, Cejas H, Balderramo D, Diller A, et al. Lack of TNFRI signaling enhances annexin A1 biological activity in intestinal inflammation. Biochem Pharmacol. 2015;98(3):422–31.
34. Vong L, Ferraz JG, Dufton N, Panaccione R, Beck PL, Sherman PM, et al. Up-regulation of Annexin-A1 and lipoxin A(4) in individuals with ulcerative colitis may promote mucosal homeostasis. PLoS ONE. 2012;7(6): e39244.
35. Mohanan NK, Shaji F, Koshre GR, Laishram RS. Alternative polyadenylation: An enigma of transcript length variation in health and disease. Wiley interdisciplinary reviews RNA. 2022;13(1): e1692.
36. Turner RE, Pattison AD, Beilharz TH. Alternative polyadenylation in the regulation and dysregulation of gene expression. Semin Cell Dev Biol. 2018;75:61–9.
37. Singh I, Lee SH, Sperling AS, Samur MK, Tai YT, Fulciniti M, et al. Widespread intronic polyadenylation diversifies immune cell transcriptomes. Nat Commun. 2018;9(1):1716.
38. Di Giammartino DC, Li W, Ogami K, Yashinskie JJ, Hoque M, Tian B, et al. RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3' UTRs. Genes Dev. 2014;28(20):2248–60.
39. Montén C, Gudjonsdottir AH, Browaldh L, Arnell H, Nilsson S, Agardh D, et al. Genes involved in muscle contractility and nutrient signaling pathways within celiac disease risk loci show differential mRNA expression. BMC Med Genet. 2015;16:44.
40. Jiang D, Deng J, Dong C, Ma X, Xiao Q, Zhou B, et al. Knowledge-based analyses reveal new candidate genes associated with risk of hepatitis B virus related hepatocellular carcinoma. BMC Cancer. 2020;20(1):403.
41. Chen C, Jiang L, Zhang Y, Zheng W. FOXA1-induced LINC01207 facilitates head and neck squamous cell carcinoma via up-regulation of TNRC6B. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie. 2020; 128:110220.
42. Chen YI, Wei PC, Hsu JL, Su FY, Lee WH. NPGPx (GPx7): a novel oxidative stress sensor/transmitter with multiple roles in redox homeostasis. American journal of translational research. 2016;8(4):1626–40.
43. Joosse ME, Menckeberg CL, de Ruiter LF, Raatgeep HRC, van Berkel LA, Simons-Oosterhuis Y, et al. Frequencies of circulating regulatory TIGIT(+) CD38(+) effector T cells correlate with the course of inflammatory bowel disease. Mucosal Immunol. 2019;12(1):154–63.
44. Heredia M, Barendregt DMH, Tindemans I, Klomberg RCW, Aardoom MA, Calado B ,et al. Gut-homing and intestinal TIGIT(neg)CD38(+) memory T cells acquire an IL-12-induced, ex-Th17 pathogenic phenotype in a subgroup of Crohn's disease patients with a severe disease course. Mucosal immunology. 2024.
45. Jin Q, Ren F, Song P. Innovate therapeutic targets for autoimmune diseases: insights from proteome-wide mendelian randomization and Bayesian colocalization. Autoimmunity. 2024;57(1):2330392.
46. Banasiewicz T, Domagalska D, Borycka-Kiciak K, Rydzewska G. Determination of butyric acid dosage based on clinical and experimental studies - a literature review. Prz Gastroenterol. 2020;15(2):119–25.
47. Gibbs B, Brown B. Butyrate therapy for treatment-resistant ulcerative colitis: a case study. Nutr Med J. 2021;1(1):60–7.
48. Fang J, Wang H, Zhou Y, Zhang H, Zhou H, Zhang X. Slimy partners: the mucus barrier and gut microbiome in ulcerative colitis. Exp Mol Med. 2021;53(5):772–87.

Publisher's Note