



## Research article

# Integrated single-cell and bulk RNA sequencing reveals CREM is involved in the pathogenesis of ulcerative colitis

Zongqi He<sup>a,1</sup>, Qing Zhou<sup>b,1</sup>, Jun Du<sup>a</sup>, Yuyu Huang<sup>a</sup>, Bensheng Wu<sup>a</sup>, Zhizhong Xu<sup>a</sup>, Chao Wang<sup>a,\*\*</sup>, Xudong Cheng<sup>a,\*</sup><sup>a</sup> Suzhou TCM Hospital Affiliated to Nanjing University of Chinese Medicine, Suzhou, 215009, PR China<sup>b</sup> Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, 210004, PR China

## ARTICLE INFO

## Keywords:

Ulcerative colitis  
Single-cell  
RNA sequencing  
Immune  
Transcription factors

## ABSTRACT

**Background:** Ulcerative colitis (UC) is an inflammatory bowel disease characterized by persistent colonic inflammation. Here, we performed a systematic analysis to gain better insights into UC pathogenesis.

**Methods:** We analyzed two UC-related datasets extracted from the gene expression omnibus database using several bioinformatics tools. The primary cell types and key subgroups of primary cells associated with UC and differentially expressed genes (DEGs) between UC and control samples were identified. The molecular regulation of the key genes was also predicted. The gene ontology and Kyoto encyclopedia of genes and genomes enrichment analyses of marker genes of key cell subgroups and model genes were performed. The expression of key enriched genes was validated in 10 clinical samples using real-time quantitative polymerase chain reaction (RT-qPCR).

**Results:** Monocytes were identified as the major cell type. Ten differentially expressed marker genes were obtained by intersecting the 3121 DEGs, 38 marker genes in major cell types, and 104 marker genes in key cell subgroups. Four essential genes, associated with immune response, were obtained using support vector machine recursive feature elimination and least absolute shrinkage and selection operator analyses. The four essential genes were highly expressed in Cluster 0 during differentiation. Validation of the four key genes in colonic mucosal biopsy specimens from 10 normal and 10 UC patients revealed that *CREM* was highly expressed in both the lesion-free sites and lesion sites colonic mucosa of UC patients compared with normal adults.

**Conclusions:** We identified *CREM* involved in UC pathogenesis, which is expected to provide a new therapeutic target for UC.

\* Corresponding author. Pharmacy Department, Suzhou TCM Hospital Affiliated to Nanjing University of Chinese Medicine, No. 18, Yang Su Road, Suzhou, Jiangsu Province, 215009, PR China.

\*\* Corresponding author. Department of Pathology, Suzhou TCM Hospital Affiliated to Nanjing University of Chinese Medicine, No. 18, Yang Su Road, Suzhou, Jiangsu Province, 215009, PR China.

E-mail addresses: [hezongqi@njucm.edu.cn](mailto:hezongqi@njucm.edu.cn) (Z. He), [13601401869@163.com](mailto:13601401869@163.com) (Q. Zhou), [1004261444@qq.com](mailto:1004261444@qq.com) (J. Du), [yuyuhuang-sz@163.com](mailto:yuyuhuang-sz@163.com) (Y. Huang), [472044066@qq.com](mailto:472044066@qq.com) (B. Wu), [306735761@qq.com](mailto:306735761@qq.com) (Z. Xu), [419924452@qq.com](mailto:419924452@qq.com) (C. Wang), [chengxudong@njucm.edu.cn](mailto:chengxudong@njucm.edu.cn) (X. Cheng).

<sup>1</sup> Zongqi He and Qing Zhou contributed equally to this article.

<https://doi.org/10.1016/j.heliyon.2024.e27805>

Received 2 September 2023; Received in revised form 22 February 2024; Accepted 6 March 2024

Available online 8 March 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Ulcerative colitis (UC) is a chronic, recurrent, nonspecific inflammatory intestine disease involving the colon and rectum. The etiology of UC remains unclear and may be related to immune dysregulation, genetic susceptibility, environmental influences, dietary stimuli, and dysbiosis of the gut microbes [1]. UC was previously considered a Western disease but is now becoming increasingly prevalent in Asia, Africa, and the Middle East and poses a serious risk to human health, generating a huge disease burden [2,3]. There is presently no cure for UC and medication only relieves the condition. The primary drugs used to treat UC include 5-aminosalicylic acid, steroid hormones, immunosuppressants, and biological agents [4,5]. Despite advances in treatment, only ~40% of patients achieve clinical remission by the end of one year, warranting the need to explore new treatment modalities [4]. Identifying key cells and genes involved in the pathogenesis of UC using bioinformatics tools (single-cell and bulk RNA sequencing) may provide new targets for treating UC.

Although thousands of disease-associated variants have been identified using whole-genome sequencing, the molecular mechanisms by which these variants drive specific diseases remain unknown. One of the main reasons for this is the limited knowledge of specific cellular and functional programs in which these genes are involved. Single-cell RNA sequencing (scRNA-seq) data provide a unique opportunity to address this problem [6]. scRNA-seq can provide RNA expression profiles of each cell independently and can help in identifying rare cells in heterogeneous cell populations. scRNA-seq has been used to study Crohn's disease and colorectal cancer [7], Alzheimer's disease [8], multiple sclerosis, asthma, idiopathic pulmonary fibrosis, coronavirus disease (COVID-19), and UC [6]. As of date, scRNA-seq has been used to identify the amplification of M-like cells, inflammatory monocytes, fibroblasts, and CD8+IL-17+ T cells in UC [9]. The scRNA-seq technology reveals subpopulations of cells that contribute to the pathogenesis of UC and provides new insights for linking genomes to pathology [10–13]. However, data integrating single cell and bulk RNA sequencing to explore UC pathology are still limited.

Inspired by existing studies on immune-related biomarkers for UC, we conducted the first joint analysis of immune-related single-cell data and transcriptomic data to discover novel biomarkers for UC. In this study, we identified the major cell types and marker genes associated with UC pathogenesis and investigated the patterns of intercellular communication in UC. Furthermore, we validated the expression of the key genes in specimens from patients with UC and normal subjects. Based on our analysis of the functions of key genes, their expression during cell differentiation, and the manner in which differentially expressed transcription factors bind, we provide a theoretical basis that should help in obtaining further insights into the mechanisms driving UC.

## 2. Materials and methods

### 2.1. Data extraction

UC-related datasets (GSE134649, GSE87466, GSE9686, and GSE10616) were extracted from the gene expression omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). Datasets included in this study have to meet the following criteria: (1) The organism of samples is *Homo sapiens*. (2) The sample type is colon tissue. (3) dataset including UC and normal (negative control) samples (4) The dataset contains at least 3 samples per group. The GSE134649 dataset includes single-cell data for CD8+ cells from three UC and three normal samples, and the GSE87466 dataset includes data for 87 UC and 21 normal colon tissue samples. The GSE9686 dataset including 8 normal and 5 UC samples and the GSE10616 dataset including 11 normal and 10 UC samples were utilized to validate the expression and diagnostic value of key genes.

### 2.2. Analysis of cell types and heterogeneity

Quality control of the single-cell dataset (GSE134649) was performed using the “Seurat” R package (version 4.10). The exclusion criteria were shown as follows: (1) genes that could only be detected in 3 and fewer low-quality cells; (2) percent.mt  $\geq$  10%; (3) nCount  $<$  200. After screening for core cells based on the above conditions, ANOVA was used to analyze the top 2000 highly variable genes for subsequent identification of cell types. In order to observe whether there were obvious outlier samples, we performed dimensionality reduction on the samples. After normalizing the gene expression of core cells using a linear regression model, the available dimensions were filtered by the JackStraw function and the ScoreJackStraw function. The principal component analysis (PCA) analysis was implemented and the ElbowPlot method was used to further confirm the correctness of the selection of components. After that, the principal component was clustered using “tSNE” and cell clusters were annotated using the “SingleR” R package (version 1.6.1). The marker genes were screened using “FindAllMarkers” (min. pct = 0.2, only. pos = TRUE).

The major cell types in the UC and control groups were compared using the “t-test,” and the cell subgrouping was re-performed based on gene expression of the major cell types using “UMAP.” The key cell subgroups were identified using correlation analysis, and the gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) enrichment analyses of marker genes of key cell subgroups were performed using the “clusterProfiler” R package (version 4.0.2) ( $p < 0.05$ , count  $>$  1) [14].

### 2.3. Analysis of the intercellular communication

Cell communication was analyzed using the CellPhoneDB, which is a database of receptors, ligands, and their interactions. The ligand–receptor and polysomes were screened with the thresholds set as  $p \leq 0.05$ , log2 mean (Molecule 1, Molecule 2)  $\geq 0.1$ .

#### 2.4. Screening of differentially expressed marker genes

The differentially expressed genes (DEGs) between UC and control samples in the GSE87466 dataset were identified using the “limma” R package (version 3.48.3), with the threshold set as  $p < 0.05$  and  $|\log_2\text{fold change}| > 0.5$  [15]. The DEGs were shown in the volcano map and heat map which were plotted by the “ggplot2” R package (version 3.3.5) and the “heat map” R package (version 1.0.12), respectively [16].

The single sample gene set enrichment analysis (ssGSEA) score of marker genes in major cell types and key cell subgroups in UC and normal samples (GSE87466) were calculated using “gsva” (version 1.40.1) and compared using the “rank-sum” test [17]. Differentially expressed marker genes were identified by intersecting the DEGs and marker genes in major cell types and key cell subgroups.

#### 2.5. Functional enrichment analysis of differential expression marker-related genes (model genes)

Weighted gene co-expression network analysis (WGCNA) is commonly used to analyze associations between modules and phenotypes. In this study, the ssGSEA scores of the differentially expressed marker genes in each sample of the GSE87466 dataset were calculated. These scores were used to construct the coexpression network with the “WGCNAR” R package (version 1.70–3) [18]. First, the samples were clustered to ensure that there were no outliers and the soft threshold of the data was determined. Then, the gene module was set and a module clustering tree was drawn to calculate the correlation between the module and grouping traits according to the hybrid dynamic tree-cutting algorithm. A gene module with a significant correlation was identified as a key gene module with model genes. Besides, GO and KEGG pathways analyses of these model genes were performed using the “clusterprofiler” R package.

#### 2.6. Screening of the key genes involved in UC

Target genes were obtained by intersecting the differentially expressed marker and model genes. The least absolute shrinkage and selection operator (LASSO) analysis (Version 4.1–3) was performed to screen the candidate genes using the “glmnet” R package [19]. The support vector machine recursive feature elimination (SVM-RFE) method was used to obtain the importance ranking of each gene and the error rate and accuracy of each iterative combination. The best combination with the lowest error rate was selected, and the corresponding genes were selected as characteristic genes. Finally, key genes were obtained by crossing two sets of genes.

#### 2.7. Expression of the key genes

The expression of the key genes in UC and normal samples, cell types, and key cell subgroups in the GSE87466 dataset was studied. In addition, the pROC package (Version 1.18.0) was used to plot receiver operating characteristic (ROC) curves to assess the ability of key genes to distinguish between normal and UC samples [20]. The single-cell trajectory plots were generated using the “monocle” R package (Version 2.24.1).

#### 2.8. Gene set enrichment analysis (GSEA) of key genes

The GSEA files were obtained from <http://www.gsea-msigdb.org/gsea/msigdb>. The median of the single key gene expression value was used to categorize the samples into high and low expression groups, and GSEA was performed on all genes in the high and low expression groups, with the threshold set at  $|\text{NES}| > 1$ ,  $\text{NOMP} < 0.05$ , and  $q < 0.25$  [15].

#### 2.9. Immune infiltration analysis

Based on a set of 24 immune cells, the percentage abundance of infiltrating immune cells in each sample was calculated using the ssGSEA algorithm. The proportions of immune infiltrating cells were compared and visualized by Wilcoxon test and ggplot2 package [21] (Version 3.3.5), respectively. Spearman correlation analysis was used to explore the relevance of key genes to immune cells.

#### 2.10. Analysis of the mechanisms of the molecular regulation of the key genes

Transcription factors (TF) bind to specific DNA sequences and regulate gene expression, thus affecting the normal life activities of the organism. Therefore, we initially explored the regulatory mechanisms by predicting TFs targeting key genes. TFs of key genes were predicted using the CistromeDB database (cistrome.org/db). TFs within 1 KB of the promoters of key genes were selected as target TFs using  $\text{RP\_score} > 0.7$  as the threshold.

#### 2.11. Experimental validation

Colon mucosal biopsies were obtained from ten cases of normal adult volunteers and adult UC patients by an experienced endoscopist. During endoscopic procedures, 3 mucosal biopsies were obtained from different parts of the colon in normal adult volunteers and placed in Liquid Nitrogen. As quickly as possible, the biopsies were then stored at  $-80^\circ\text{C}$  until RNA extraction was performed. In UC patients, biopsies need to be taken at 3 sites each in the normal colonic mucosa and the inflammation colonic mucosa. This study was conducted with the approval of the Institutional Review Board at Suzhou TCM Hospital (Approval Number: 2022 Ethics Research

Approval 043). Informed consent was obtained from patients and volunteers for all human colonic mucosal biopsy specimen collections. Basic information on UC patients and normal adult volunteers is shown in Supplementary Materials (Table S1).

The RNA of UC and normal samples was extracted from tissues with TRIzol (Servicebio Technology CO., LTD, Wuhan, China). Then, the RNA concentration was detected with a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Total RNA was subjected to reverse transcription using SweScript All-in-One First-Strand cDNA Synthesis SuperMix for qPCR kit (Servicebio Technology CO., LTD, Wuhan, China). The real-time quantitative polymerase chain reaction (RT-qPCR) reaction system was made up of SYBR Green qPCR Master Mix (Servicebio technology CO., LTD, Wuhan, China). The amplification reactions were performed using CFX Connect (Bio Rad, Hercules, CA, USA) programmed as follows: initial denaturation at 95 °C for 30 s, followed by 40 cycles of denaturation at 95 °C for 15 s, annealing at 60 °C for 30 s. The relative expression of genes was calculated by the  $2^{-\Delta\Delta CT}$  method using GAPDH as the internal reference gene. The gene sequences are shown in Table 1 below.

## 2.12. Statistical analyses

Experimental results are expressed as mean  $\pm$  standard deviation. Statistical analysis of RT-qPCR was performed using unpaired, 2-tailed Student's t-test and one-way ANOVA to compare all groups (GraphPad Prism software, version 9.0).  $P < 0.05$  was considered statistically significant.

## 3. Results

### 3.1. Expression of marker genes in different cell types

After quality control of single-cell samples was completed, 8738 core cells were identified (Supplementary Fig. 1A). A total of 2000 genes with highly variable expression in 8738 core cells were identified for subsequent identification of cell types involved in UC in the GSE134649 dataset (Supplementary Fig. 1B). Based on the results of PCA, all core cells and the top 20 PCs ( $p < 0.05$ ) could be incorporated into subsequent analyses (Supplementary Figs. 1C–E). Supplementary Fig. 2 also showed that the overall decline tended to be relatively gentle after the top 20 PCs, so it was reasonable to select the first 20 PCs. The core cells were classified into 14 clusters and identified as CD8<sup>+</sup> T cells, CD4<sup>+</sup> T cells, and monocytes by annotating different clusters in the Cellmarker database (Fig. 1A&B, Table S2). Fig. 1C revealed that the number of CD8<sup>+</sup> T cells was largest compared to the other two types of cells. The expression of marker genes in the different cell types was shown in Fig. 1D. *CREM*, *CD8A*, *CLEC2B*, *HLA-A*, and *CTSW* were identified as the marker genes for CD8<sup>+</sup> T cells. *AIF1*, *S1PR1*, *RPS3A*, *LYPD3*, and *ACTN1* were identified as the marker genes for CD4<sup>+</sup> T cells, and 38 genes, including *PSAP*, *TUBB*, *GSTP1*, *HLA-DRB1*, and *TXN*, were identified as the marker genes for monocytes.

### 3.2. Functional analysis of marker genes in the key subgroups

To recognize key cell subgroups from three cell types, we implemented statistical analysis and comparison for cell types of different samples. The proportion and distribution of three cells in UC and normal groups were shown in Fig. 2A–B. Meanwhile, Table 2 demonstrated that the proportion of monocytes was significantly different ( $p < 0.05$ ) between UC and normal samples, and monocyte was defined as the major cell type in the two groups. After that, six cell subgroups were identified in monocytes based on the UMAP clustering analysis (Fig. 2C). The expression of marker genes in the different subgroups is shown in Supplementary Fig. 3A&B. Clusters 0 and 1 showed the highest correlation and were defined as key subgroups (Fig. 2D).

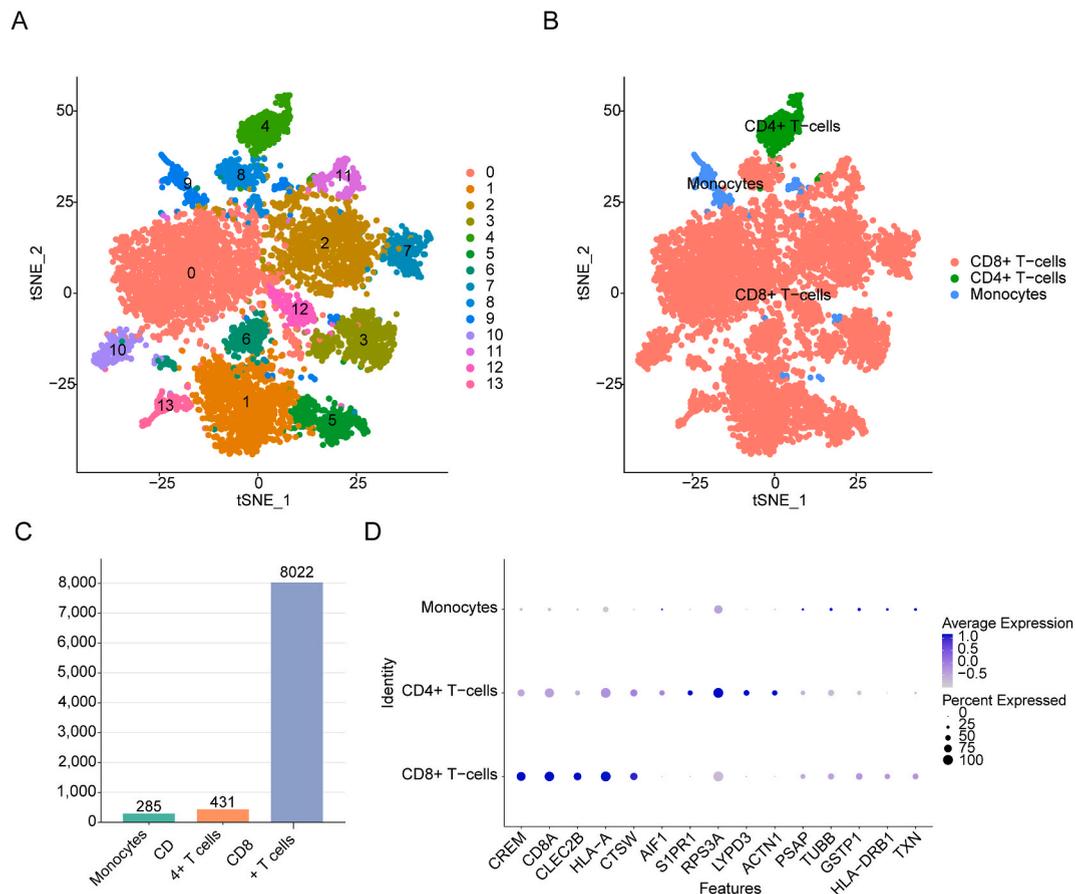
A total of 104 marker genes in the key subgroups were enriched in mononuclear cell differentiation, ribosomes, and structural constituents of ribosomes (Fig. 2E). Among the identified KEGG terms, these genes were associated with coronavirus disease (COVID-19), ribosome, and differentiation of Th1, Th2, and Th17 cells (Fig. 2F).

### 3.3. Analysis of intercellular communication

Eighteen ligand–receptors were screened. The communication between CD8<sup>+</sup> and CD8<sup>+</sup> T cells was the closest, and all 18

**Table 1**  
Nucleotide sequence of human primers used for RT-qPCR.

Gene	Nucleotide Sequence (5'-3')	Primer	Gene Bank
GAPDH	GGAAGCTTGTCAATGGAATC	Forward	NM 002046
GAPDH	TGATGACCCCTTTGGCTCCC	Reverse	
PTPRC	TAAGACAACAGTGGAGAAAGGACG	Forward	NM 001267798.2
PTPRC	CAAATGCCAAGAGTTTAAGCCAC	Reverse	
RGS1	TATTGAGTTCTGGCTGGCTTGT	Forward	NM 002922.4
RGS1	CTGATTTGAGGAACCTGGGATA	Reverse	
CREM	ATGCCAACTTACCAGATCCGA	Forward	NM 001267562.2
CREM	TCGTTTGGTGTGCTTCTTC	Reverse	
GNLY	ACACTTCTGGAAGGGAGAGTGGA	Forward	NM 001302758.2
GNLY	ATCACGCAGGTGGGCTCTT	Reverse	



**Fig. 1.** Identification of cell types in the data of single-cell RNA sequencing (scRNA-seq). (A) T-distributed stochastic Neighbor Embedding (t-SNE) plot for the 14 clusters across 8738 core cells. (B) Presentation of clustering results for three cell types annotated. (C) Boxplot for the numbers of marker genes in the different cell types. (D) Bubble diagram of marker genes expressed in different cell types.

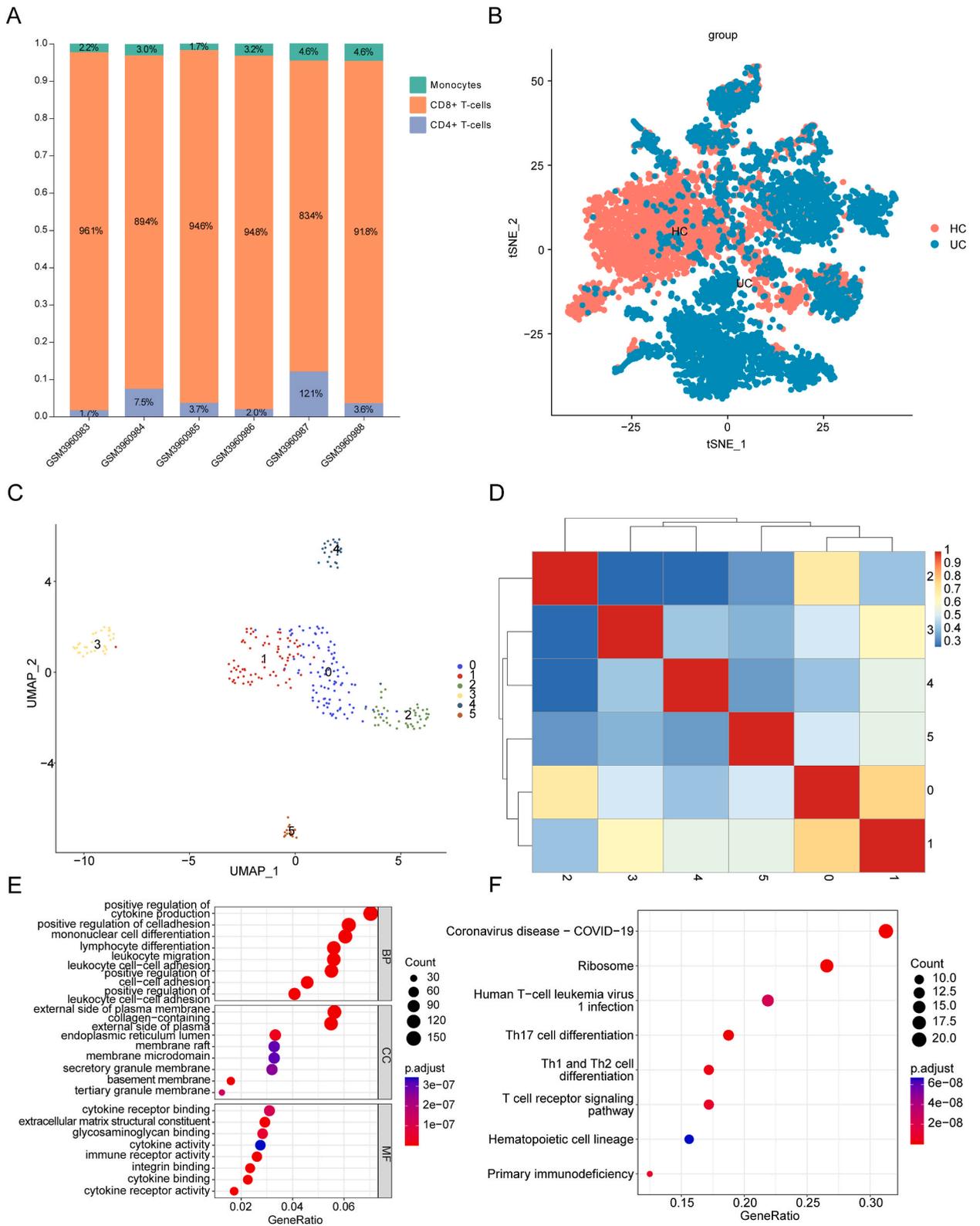
ligand–receptors were found to play important roles in communication between CD8<sup>+</sup> and CD8<sup>+</sup> T cells. In addition, there were negative correlations between the CD8 receptor LCK and CD4<sup>+</sup> T cells|CD4<sup>+</sup> T cells, NCR3\_BAG6, and CD8<sup>+</sup> T cells|CD4<sup>+</sup> T cells (Fig. 3A). The mutual communication networks between the different cell types are shown in Fig. 3B. Additionally, we performed cell communication analysis in normal and UC groups to show the receptor–ligand profiles between the cells of different groups. In total, 32 receptor–ligands in the normal group played a role in cell communication, while 52 receptor–ligands in UC group. Interestingly, we found that UC group contained all the receptor–ligands of the normal group and had 20 unique receptor–ligands (Fig. 3C–D). Therefore, we suggested that the communication between cells in the UC group was more complex.

### 3.4. Identification of differentially expressed marker genes in UC

We identified 3121 DEGs (1798 upregulated and 1323 downregulated mRNAs) between 21 normal control and 87 UC samples in the GSE87466 dataset (Fig. 4A&B). As shown in Fig. 4C, the ssGSEA scores of marker genes in major cell types and key cell subgroups were significantly different between UC and normal samples. Finally, 10 differentially expressed marker genes, viz., *CST7*, *PTPRC*, *RGS1*, *CLEC2B*, *CREM*, *CD3D*, *SRGN*, *GNLY*, *CD8A*, and *MYADM*, were screened by crossing 38 marker genes in the major cell types, 104 marker genes in the key cell subgroups, and 3121 DEGs (Fig. 4D–F).

### 3.5. Functional enrichment analysis of the model genes

The analysis of clustering results showed that there were no outlier samples in the GSE87466 dataset (Fig. 5A), and the mean connectivity gradually approached 0 with a gentle trend indicating that the network approximated a scale-free distribution (Fig. 5B). A module-clustering tree was constructed with eight modules based on the optimal soft threshold and the hybrid dynamic tree-cutting algorithm (Fig. 5C). We analyzed the correlation between modules and grouping traits and found that the light cyan module ( $R^2 = 0.85$ ,  $p = 4e-31$ ) had the highest significant positive correlation with differentially expressed marker genes (Fig. 5D). Therefore, the light-cyan module was identified as the key module.



(caption on next page)

**Fig. 2. Identification and analysis of the major cell type and functional enrichment analysis.** (A) Boxplot for the distributions of three cell types in single-cell samples in GSE134649. (B) The t-SNE plot for the ulcerative colitis (UC) and control (HC) groups. (C) Uniform manifold approximation and projection (UMAP) plot of six subtypes generated based on the expression of monocytes in three UC samples within GSE134649. (D) Correlation heatmap of different monocyte-related subtypes. (E) The most enriched Gene Ontology (GO) terms of 104 marker genes in the two key subgroups (subgroups 0 and subgroups 1). (F) The mostly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of 104 marker genes.

**Table 2**

The results of *t*-test for three cell types between UC and normal groups.

	Group (mean ± SD)		t	p
	HC (n = 3)	UC (n = 3)		
CD4 T-cells	57.00 ± 40.15	86.67 ± 63.88	-0.681	0.533
CD8 T-cells	1176.33 ± 345.65	1497.67 ± 359.32	-1.116	0.327
Monocytes	28.33 ± 9.07	66.67 ± 11.55	-4.521	0.011*

Furthermore, these module genes were enriched in 1394 “biological process,” 50 “cellular component,” and 92 “molecular function” sub-ontologies of GO. These genes are significantly associated with lymphocyte and mononuclear cell differentiation, positive regulation of cytokine production, and cell adhesion, among other functions. In KEGG analysis, these genes were enriched in cytokine–cytokine receptor interaction, chemokine signaling pathway, and PI3K-Akt signaling pathway, among other KEGG terms (Fig. 5E&F).

### 3.6. Identification of the key genes in UC

Six genes (*CST7*, *PTPRC*, *RGS1*, *CREM*, *GNLV*, and *CD8A*) were identified by intersecting 10 differentially expressed marker genes and 2973 model genes. After that, LASSO and SVM-RFE models were constructed to select characteristic genes based on the samples in the GSE87466 dataset. The results of the LASSO analysis showed that four candidate genes, namely *PTPRC*, *RGS1*, *CREM*, and *GNLY*, were associated with UC (Supplementary Fig. 4A & B). The best combination with the lowest error rate and the corresponding genes were obtained using the SVM-RFE method. As shown in Supplementary Fig. 4C&D, the error rate for predicting the best point of disease and control samples was 0.11, and the accuracy rate was 0.89, when the number of genes changed from one to six. Six key characteristic genes, namely *CST7*, *PTPRC*, *RGS1*, *CREM*, *GNLY*, and *CD8A*, were screened by 5-fold cross-validation. Finally, we took the intersection of genes obtained using these two algorithms and identified four key genes—*PTPRC*, *RGS1*, *CREM*, and *GNLY*—in UC (Fig. 5G).

### 3.7. Expression of the key genes

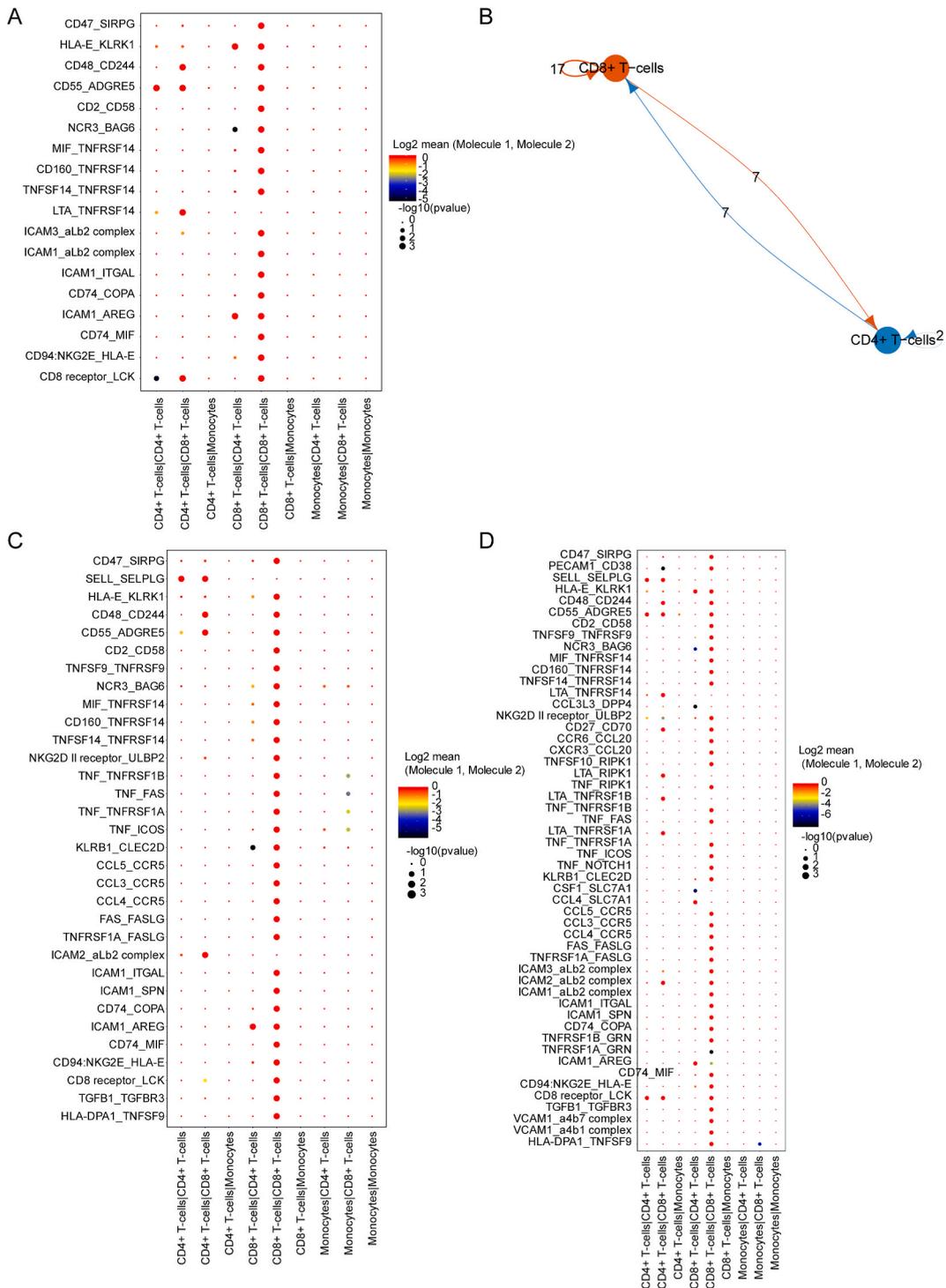
A comparison of the expression of the key genes in GSE87466 between UC and normal samples showed that all four key genes were significantly upregulated in patients with UC (Fig. 6A). Meanwhile, the expression trends of key genes in the validation sets (GSE9686 and GSE10616) were consistent with those in the training set (Fig. 6B–C). Subsequently, the ROC curves of key genes were plotted in the training and validation sets. Fig. 6D–F showed the AUC values of key genes were greater than 0.7 in three datasets, indicating the decent abilities to distinguish between normal and UC samples. The expression of these genes in the three key cell types is shown in Fig. 7A&B; all the key genes were highly expressed in CD8<sup>+</sup> T cells and monocytes, and *PTPRC* and *CREM* were simultaneously highly expressed in all three cell types. Expression analysis of the key genes in the key cell subgroups showed that four key genes were highly expressed in cluster 0, indicating that this cluster might be the main cell type involved in the pathogenesis of UC (Fig. 7C&D). Furthermore, the single-cell trajectory plots and the differential expression of four key genes showed that three key cell types coexisted in the process of cell differentiation, and the four key genes were highly expressed in cluster 0 during differentiation (Fig. 7E–G).

### 3.8. Four key genes are associated with the immune response

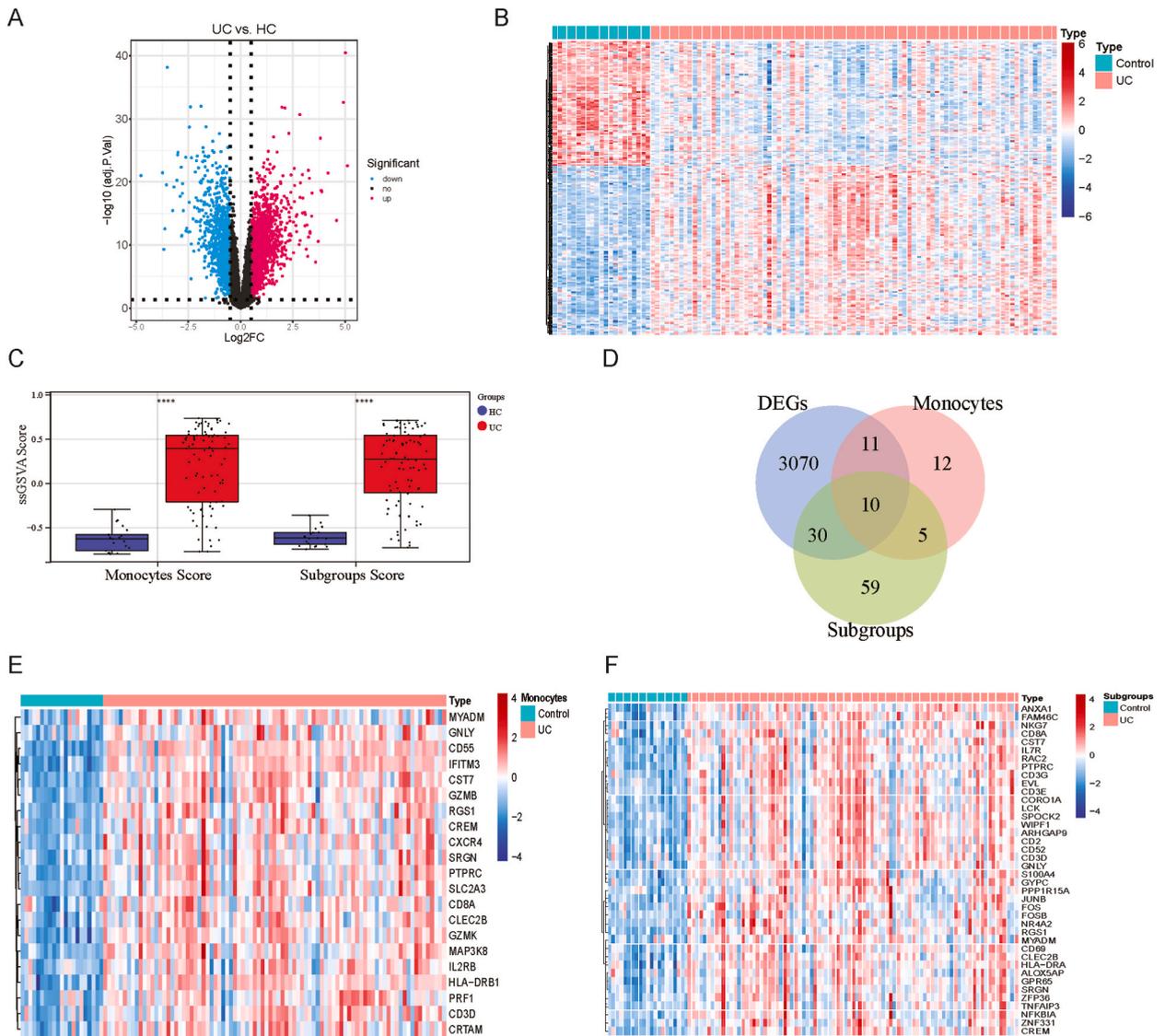
The results of the GSEA analysis of the four key genes are shown in Fig. 8A–H. Each key gene was highly enriched in adaptive immune response, cellular response to molecules of bacterial origin, B cell-mediated immunity, humoral immune response, and lymphocyte-mediated immunity. More than half of the key genes were involved in humoral immune response, lymphocyte-mediated immunity, and positive regulation of leukocyte cell adhesion. Moreover, all the key genes were involved in metabolic pathways, such as the chemokine signaling pathway and receptor interaction, hematopoietic cell lineage, STAT signaling pathway, Leishmania infection, and systemic lupus erythematosus. In addition, some genes were highly enriched in the intestinal immune network for IgA production.

### 3.9. Changes in the immune microenvironment in patients with UC

Based on the GSEA enrichment analysis, four key genes were associated with immune-related pathways. The immune infiltration



**Fig. 3. Analysis of intercellular communication.** (A) Interaction bubble diagram of the ligand-receptor and polysome inside different cell types. The horizontal axis is the cell type that interacts, and the vertical axis is the interacting ligand receptors and polymers. (B) Schematic diagram of cell-cell communication between CD8<sup>+</sup> T cells and CD4<sup>+</sup> T cells. Different colors indicate different cell types, and the width of the lines indicates the interaction score. (C, D) The receptor-ligand interactions between cells in normal and UC groups. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

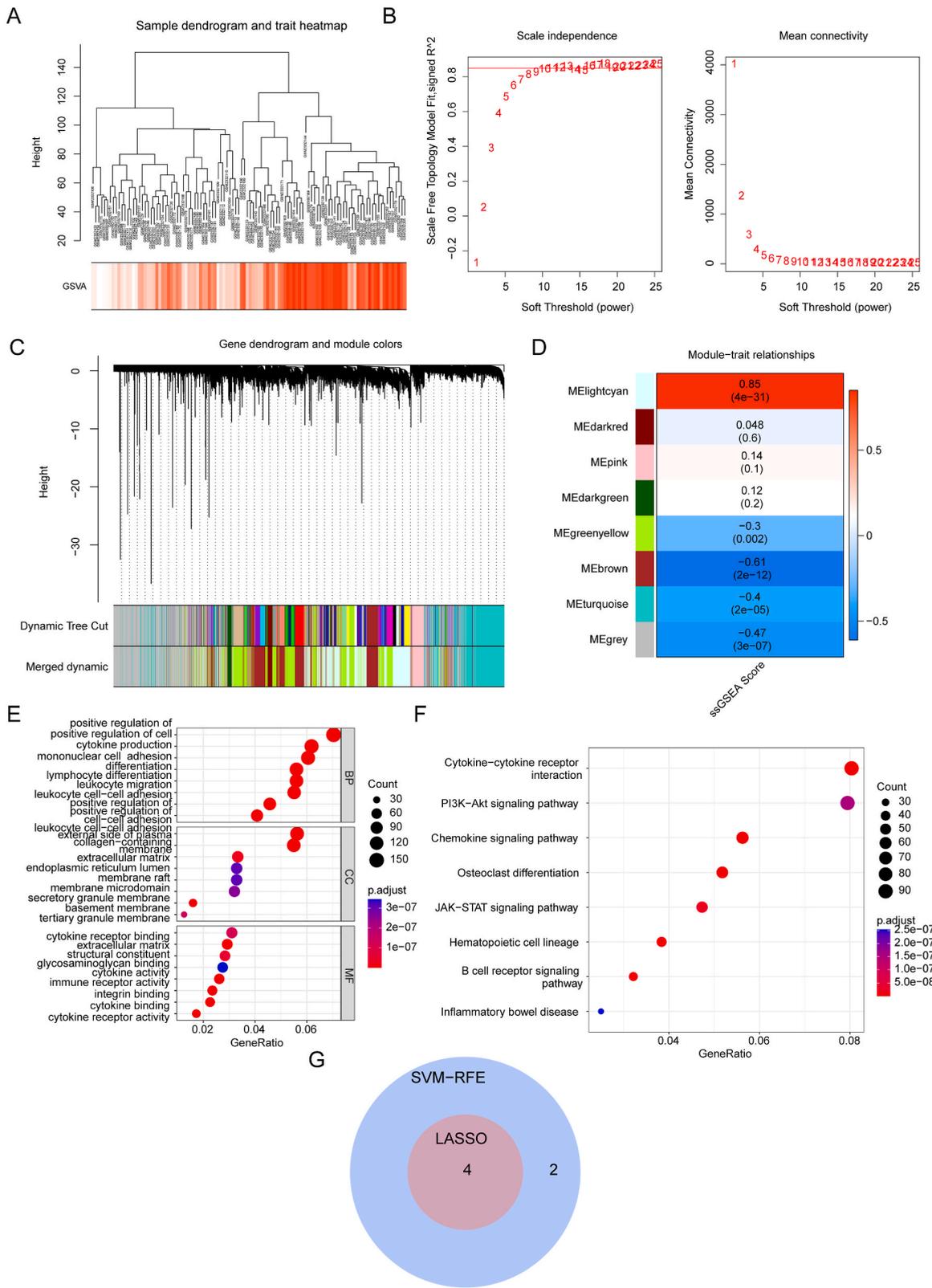


**Fig. 4. Identification of differentially expressed marker genes in GSE87466.** (A–B) Volcano plot and heatmap of the 3121 differentially expressed genes (DEGs) between UC and HC groups with  $p < 0.05$  and  $|\log_2\text{fold change}| > 0.5$ . (C) Boxplot for the single sample gene set enrichment analysis (ssGSEA) score of individuals between UC and HC groups based on the marker genes expression of major cell types (monocytes) ( $n = 38$ ) and key cell subgroups (subgroups 0 and subgroups 1) ( $n = 104$ ), respectively. (D) Venn diagram of 10 common differentially expressed marker genes. (E–F) Heatmap of 38 marker genes in major cell types and 104 marker genes in key cell subgroups.

analysis was completed to further mine the relevance of the immune microenvironment to UC. Among the 24 kinds of immune cells, except for eosinophils, iDC, NK cells, Tgd, and TReg, there were significant differences in the proportion of the other 19 kinds of immune cells between UC and normal samples, and the proportion of most immune cells in UC was significantly higher than normal samples (Supplementary Fig. 5A). Supplementary Fig. 5B indicated that key genes were significantly negatively correlated with Th17 cells and positively correlated with all other differential immune cells. Thus, we hypothesized that the immune microenvironment might play an important role in the developmental process of UC. Meanwhile, the expression levels of key genes might also influence the reactivity and infiltration of immune cells, which in turn might influence the immune microenvironment's impact on disease.

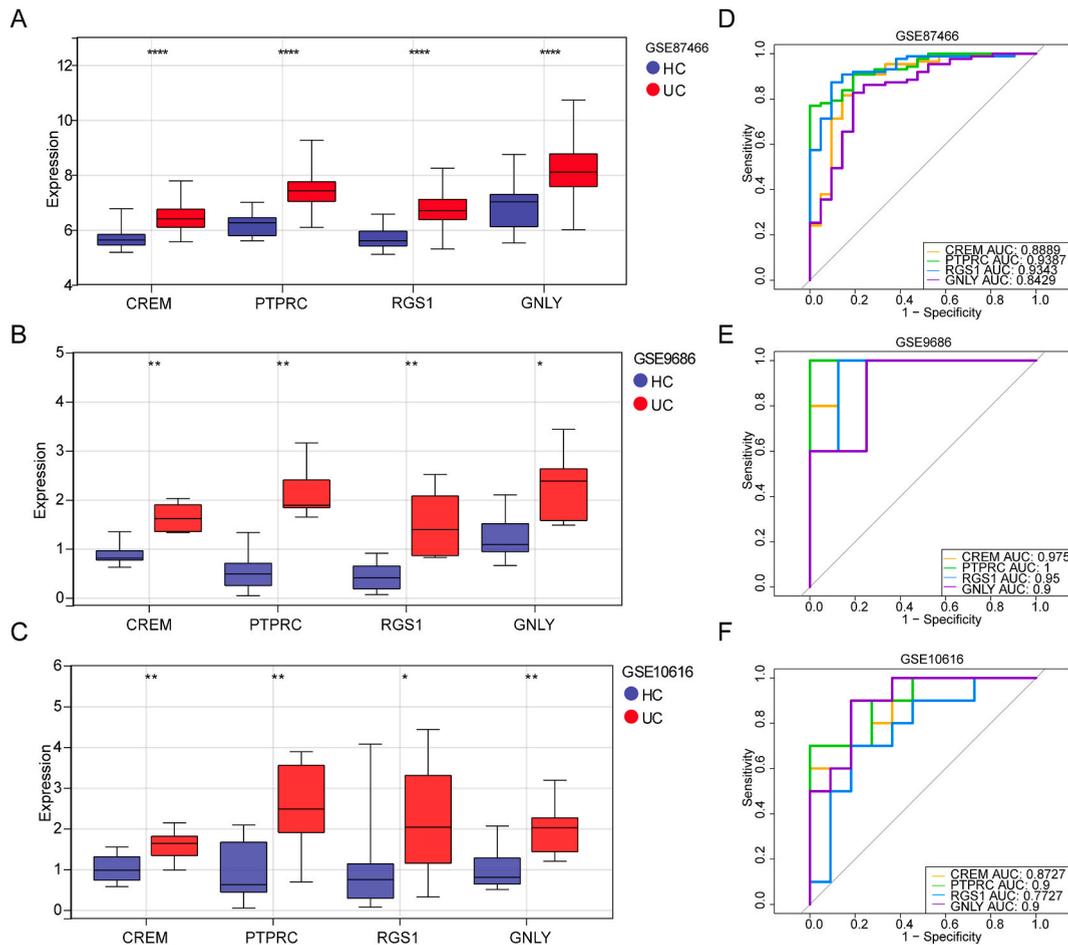
### 3.10. Analysis of the mechanisms for molecular regulation of the key genes

To initially investigate the reasons for the altered expression of key genes in UC samples, we conducted a transcription regulation analysis. In total, 54 TFs targeting CREM, 1 TFs targeting GNLY, and 3 TFs targeting PTPRC were predicted by the CistromeDB database ( $RP\_score > 0.7$ ). The gene-TF regulatory network revealed that ESR1 could regulate the expression of GNLY and CREM (Supplementary Fig. 6).



(caption on next page)

**Fig. 5.** Weighted gene co-expression network analysis (WGCNA) to screen the key model genes in GSE87466. (A) Sample clustering dendrogram and trait heatmap. (B) Analysis of the scale-free fit index (left) and the mean connectivity (right) for various soft-thresholding powers. (C) Cluster dendrogram of all DEGs clustered based on a dissimilarity measure. (D) Heatmap of the correlation between module eigengenes and clinical traits (ssGSEA score calculated based on the expression of 10 common differentially expressed marker genes). Each cell contains the correlation coefficient and  $p$ -value. (E) GO and (F) KEGG functional enrichment analysis of 2973 key model genes in the light cyan module. (G) The Venn diagram of four key genes is identified by two algorithms. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



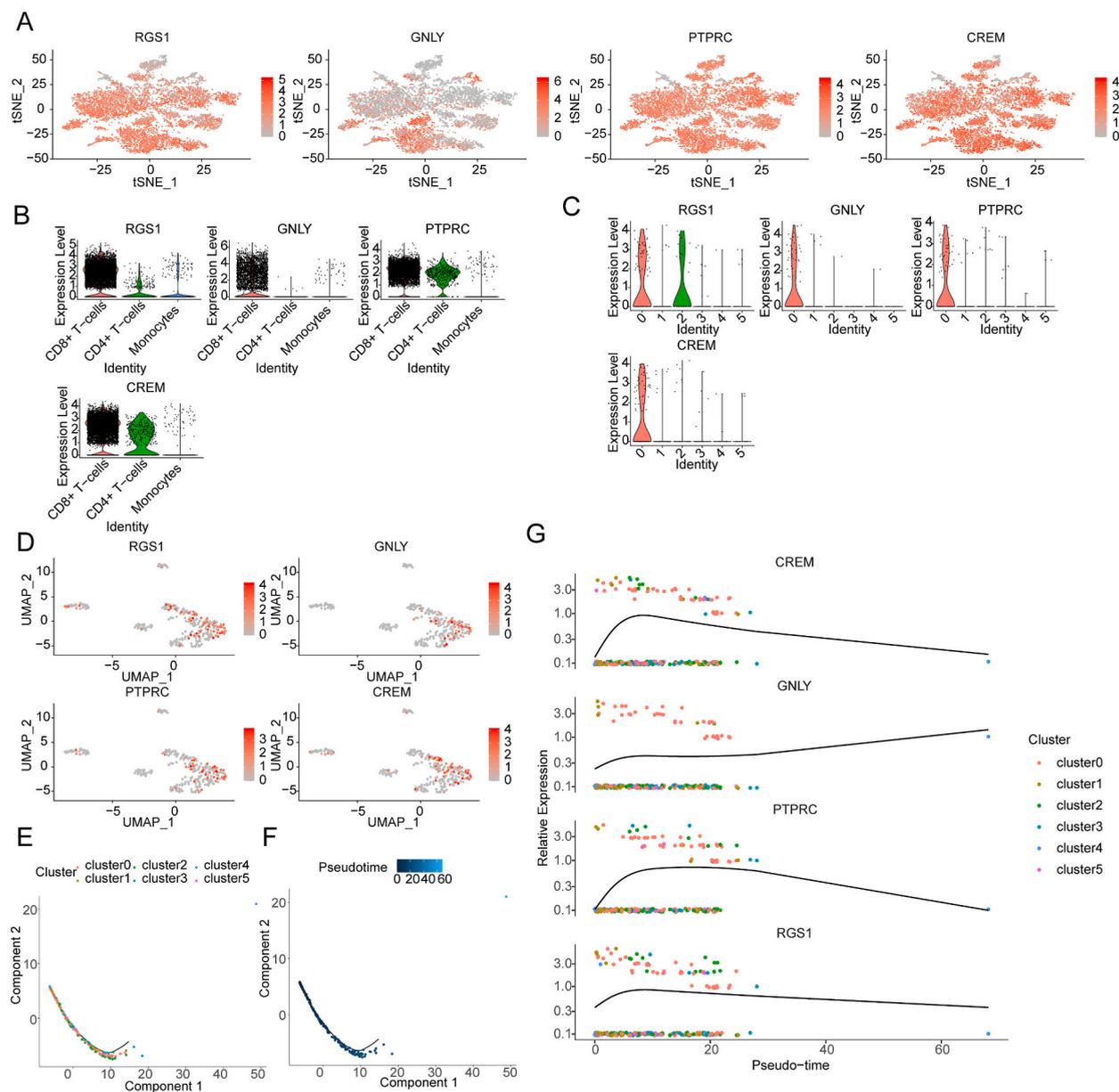
**Fig. 6.** Analysis of the expression discrepancies and diagnostic value of key genes. (A–C) The boxplots of expression levels of four key genes in the normal and UC groups in the GSE87466 (A), GSE9686 (B), and GSE10616 (C) datasets. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\*\* $p < 0.0001$ . (D–F) The receiver operating characteristic (ROC) curves of key genes in the GSE87466 (D), GSE9686 (E), and GSE10616 (F) datasets. AUC, area under the curve.

### 3.11. Validation of key differentially expressed genes

The expression of genes *RGS1*, *GNLY*, *PTPRC*, and *CREM*, was analyzed by RT-qPCR assay in normal volunteers, lesion-free sites of UC patients, and lesion sites of UC patients. Among all four genes analyzed, the expression of *RGS1*, *GNLY*, and *PTPRC* genes was not significantly altered between UC patients and normal controls ( $p > 0.05$ ) (Fig. 9A–C). The expression of *CREM* was significantly higher in lesion-free sites and lesion sites in UC patients than in normal controls ( $p < 0.001$ ), and there was no significant difference in *CREM* expression between lesion-free sites and lesion sites in UC patients ( $p > 0.05$ ) (Fig. 9D).

## 4. Discussion

Ulcerative colitis is an IBD characterized by persistent colonic inflammation. The main features of UC are manifested continuously from the rectum to the proximal colon and involve the colonic mucosa and submucosa. The disease is characterized by ulceration and bleeding and can lead to bursting colitis and colorectal cancer. Current therapeutic options for UC are inadequate for its clinical



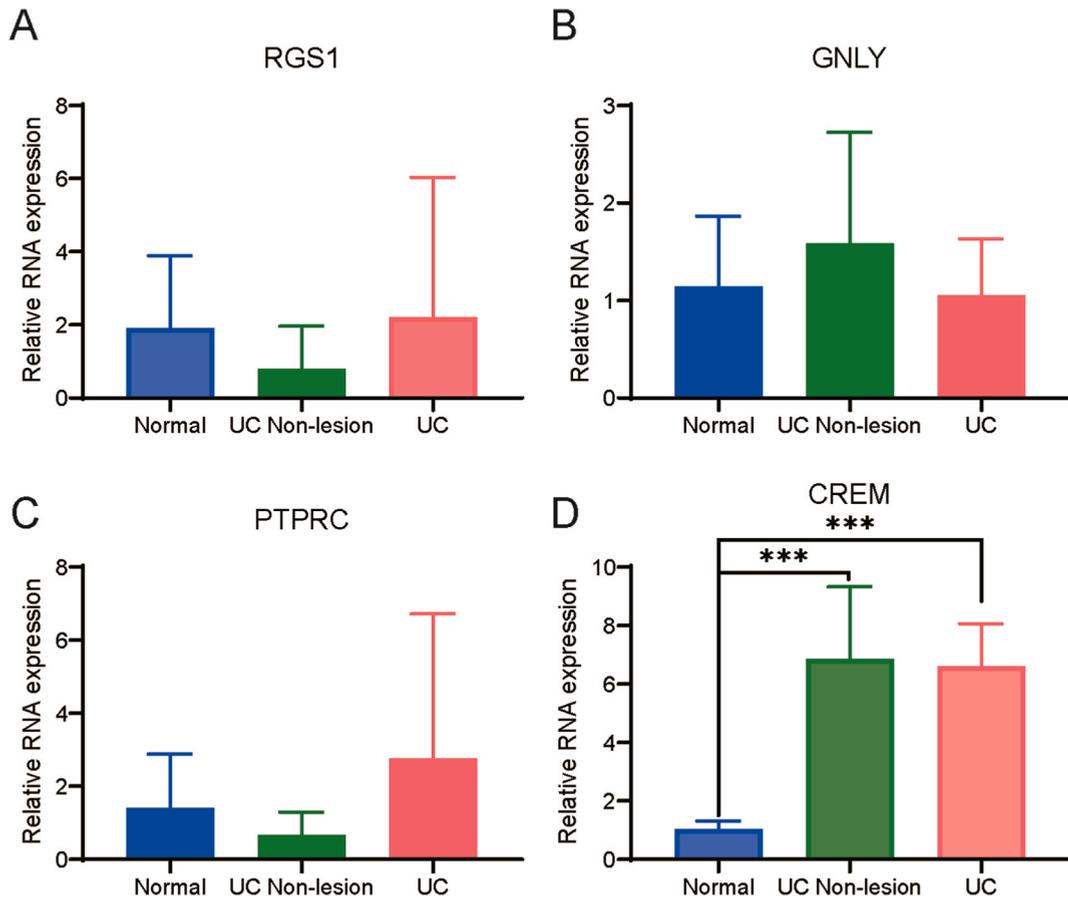
**Fig. 7.** Analysis of the four key characteristic gene expression. (A) t-SNE and (B) violin plots of four key characteristic gene expressions in three cell types. (C) Violin plots and (D) t-SNE plots of four key characteristic genes expression in key cell subgroups. (E) Pseudotime analysis of different monocyte-related subtypes. Line indicating a continuum of hypothetical differentiation from least differentiated to several branches of distinct differentiated cell types, where the differentiation of cluster 0 is the earliest. Cells in the same branch are in the same differentiation state, and cells in different branches are in different differentiation states, differentiation into different trajectories proves intracellular heterogeneity. (F) Differences in the time sequence of cell differentiation. Darker blue indicates earlier differentiation and lighter blue indicates later differentiation. (G) Correlation analysis between four key characteristic genes and monocyte differentiation. The expression of four key genes was highest in cluster 0. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

management. There is, therefore, an urgent need to comprehensively analyze the pathogenesis of UC and develop effective therapeutic strategies.

In this study, we first performed t-distributed stochastic neighbor embedding (tSNE) clustering on the GSE134649 dataset to identify three main cell types related to UC, namely CD8<sup>+</sup> T cells, CD4<sup>+</sup> T cells, and monocytes. Monocyte counts showed significant differences between patients with UC and healthy controls. Therefore, we selected monocyte marker genes as a candidate gene set for our prediction model. Monocytes mostly express genes associated with bacterial infection and inflammation. The proinflammatory cytokines, IL-1 $\beta$  and TNF- $\alpha$ , are mainly secreted by classical monocytes. Proinflammatory monocytes and macrophages predominate in the inflammatory mucosa of patients with UC. Circulating monocytes express high levels of proinflammatory cytokines. These findings



**Fig. 8. Single-gene gene set enrichment analysis (GSEA) based on four key characteristic genes.** (A–B) Results of GO and KEGG enrichment of CREM single gene GSEA (top 10). (C–D) Results of GO and KEGG enrichment of RGS1 single gene GSEA (top 10). (E–F) Results of GO and KEGG enrichment of GNLY single gene GSEA (top 10). (G–H) Results of GO and KEGG enrichment of PTPRC single gene GSEA (top 10). In this plot, the vertical coordinates of the top part represent the running enrichment scores (ES), and the peak of the line plot is the ES of the enriched pathway, with positive ES indicating that a functional gene set is enriched in the front of the sequenced sequence and negative ES indicating that a functional gene set is enriched in the back of the sequenced sequence. The horizontal coordinates represent genes labeled with small vertical lines. The bottom part is the distribution map of rank values for all genes, with the Signal2Noise algorithm by default.



**Fig. 9. Quantitative Analysis of key genes by RT-qPCR in clinical samples.** (A–D) RGS1, GNLY, PTPRC, and CREM mRNA expression in colonic mucosal biopsy specimens from normal volunteers (Normal), lesion-free sites of UC patients (UC No-lesion), and lesioned sites of UC patients (UC). (n = 10, \*\*\*p < 0.001).

indicate that the majority of monocytes in patients with UC exhibit proinflammatory properties. Studies have found that the proportion of classical monocytes in the peripheral blood of patients with UC is reduced during inflammatory activity, suggesting that these cells may migrate to the intestine and promote inflammatory processes [9,10]. Monocytes are involved in the upstream inflammatory process of the immune response to IBD, and patients with high monocyte counts experience a recurrence of inflammation after discontinuation of biologics [22]. TIM4-expressing monocytes are novel indicators of the disease activity and severity of UC; CD14<sup>+</sup>TIM4<sup>+</sup> cells were reported to be elevated in patients with UC, and the percentage of TIM4-expressing monocytes was significantly lower in such patients treated with mesalazine [23]. CD30L levels in classical monocytes in the peripheral circulation of patients were significantly elevated and correlated positively with the severity of UC. CD30L may be involved in monocyte-mediated inflammation in patients with UC through the activation of circulating classical monocytes [24]. High monocyte counts were significantly and negatively correlated with clinical remission, mucosal healing, and complete mucosal healing [25].

The KEGG pathway analysis showed that common DEGs were significantly enriched in the PI3K-AKT, chemokine, and JAK-STAT signaling pathways. PI3K-dependent AKT phosphorylation is a key event in the progression of UC through the activation of inflammatory signaling. AKT is highly phosphorylated in the colon biopsies of patients with IBD and in experimental *in vivo* models of colitis. Additionally, the severity of colitis was reduced by treatment with PI3K inhibitors [26]. CXCL8 is one of the most important proinflammatory factors that play a crucial role in many inflammatory diseases, including UC. The CXCL8-CXCR1/2 axis is involved in the

pathogenesis of UC through various signaling pathways, such as the PI3k/Akt, MAPK, and NF- $\kappa$ B signaling pathways [27]. Janus kinase (JAK), a tyrosine kinase, and signaling transducer and activator of transcription (STAT), which are DNA-binding proteins, mediate cytokine receptor-binding, signaling, and downstream biological effects, many of which are involved in the pathology of IBD. Small molecule JAK inhibitors can potentially affect multiple cytokine-dependent pathways and are effective in the treatment of IBD [28]. Blocking JAK-mediated inflammatory pathways can alter the innate and adaptive immune responses involved in IBD, and thereby, reduce chronic gastrointestinal inflammation [29].

We noted an enrichment of four key genes that are closely related to the pathogenesis of UC using bioinformatic methods. Protein tyrosine phosphatase receptor type C (PTPRC), also called cluster of differentiation 45 (CD45), was previously identified as the common antigen in leukocytes. PTPRC is a member of the protein tyrosine phosphatase (PTP) family, which is a signaling molecule that regulates various cellular processes, including cell growth, differentiation, mitosis, and oncogenic transformation [30]. This PTP is an important regulator of antigen receptor signaling in T and B cells [30]. Transgenic expression of tumor necrosis factor (TNF) superfamily member 14, produced by *Ptprc* on mouse T cells, has been reported to promote inflammation in several organs, including the intestine [31]. In another study, most apoptotic cells in the submucosa of nuclear receptor binding factor 2 knockdown mice with dextran sodium sulfate-induced colitis colocalized with *Ptprc*/Cd45 expression, which suggests that *Ptprc*/Cd45 is involved in the activation of intestinal inflammation in mice [32]. In yet another study, epithelial cells from the mouse small intestine and organoids were analyzed and new subtypes and their genetic profiles were characterized; two tuft cell subtypes were distinguished from epithelial cells of the small intestine, one of which expressed the epithelial cytokines, *Tslp* and *Cd45* (*Ptprc*), which are pan-immune markers previously believed to be unrelated to non-hematopoietic cells [33].

G protein-coupled receptors (GPCRs) play important roles in both innate and adaptive immunity. The primary function of the regulator of G protein signaling (RGS) family is to limit the intensity and duration of GPCR signaling. RGS1 is primarily expressed in B and T lymphocytes, monocytes, and macrophages. Its major immune functions include antibody response to immunization, macrophage localization in atherosclerotic plaques, and T cell-mediated inflammatory colitis [34]. The purported first study of RGS1 biology in human T cells found that RGS1 is significantly overexpressed in the intestine and in some cases of IBD. Moreover, RGS1 normally inhibits the efflux of T cells from the intestine and promotes immunopathology of the intestine [35]. In humans and mice, RGS1 is highly expressed in CD4<sup>+</sup> and CD8<sup>+</sup> T lymphocytes isolated from the gastrointestinal mucosa [36]. By combining whole-exome and scRNA-seq results, we noted that the expression of RGS1 was significantly upregulated in patients with chronic mucocutaneous candidiasis [37].

Cyclic adenosine monophosphate (cAMP)-responsive element modulator (*CREM*) encodes a bZIP transcription factor that binds to the cAMP-responsive element found in many viral and cellular promoters. *CREM* is significantly more highly expressed in UC patients than in the normal population, and in UC patients with cancer than in UC patients without cancer [38]. Several studies suggest that *CREM* is highly expressed in IBD [39–42]. New evidence on the role of *CREM* in IBD susceptibility suggests that *CREM* is a key regulator of intestinal inflammation and may have broad therapeutic potential as a drug target for inflammatory diseases of the gut [41].

Granulysin (GNLY) is a cytotoxic molecule associated with bacterial lysis and monocyte chemotaxis. A recent study showed an increase in the number of CD127+CD94<sup>+</sup> innate lymphocytes expressing GNLY and perforin in patients with Crohn's disease [43]. scRNA-seq analysis of peripheral blood mononuclear cells in acute myocardial infarction revealed that CD8<sup>+</sup> effector T cells express more GNLY in patients without plaque rupture, which may contribute to the progression of plaque erosion [44].

The identification of susceptibility genes for IBD is key to understanding its pathogenic mechanisms. Four genes with significant differences in expression were enriched in our study and were validated in human specimens. The validation results revealed that *CREM* was highly expressed in both the normal colonic mucosa and the diseased site mucosa of patients with UC compared with that in normal adults, whereas the expression of *PTPRC*, *RGS1*, and *GNLY* did not differ significantly. A perusal of the literature revealed an association between these four DEGs and IBD. However, such studies have been scarce, and in only a few of them, a direct association of these four DEGs with IBD has been observed. The available literature suggests that *CREM* has the strongest correlation with IBD, which is consistent with the results of our study.

## 5. Conclusion

Our study provides a guide for understanding the pathogenesis of IBD, which should help in formulating effective therapeutic strategies. A shortcoming of this study is that we failed to explore the involvement of the four identified key genes in depth. Furthermore, to ensure the applicability of our research findings, expanding the number of clinical samples is a pressing concern for our next steps. While we validated gene expression levels through qRT-PCR, a substantial volume of experiments remains crucial for validating our study results. Simultaneously, within the PCR results, three genes showed insignificant differences, possibly attributed to factors such as tissue specificity, sample size, experimental conditions, and methodologies. In essence, continued focus on the functionality of the four key genes in UC and exploring their molecular mechanisms will be integral to our upcoming research. We intend to focus on *CREM* in our future study, which should provide further insights into the pathogenesis of UC and lead to the development of effective drugs.

## Ethics approval and consent to participate

This study was conducted with the approval of the Institutional Review Board at Suzhou TCM Hospital (Approval Number: 2022 Ethics Research Approval 043). Informed consent was obtained from all individual participants included in the study.

## Consent for publication

The authors confirmed that consent to publish has been received from all participants.

## Funding

This work was supported by the National Natural Science Foundation of China (grant number 82104859), the Natural Science Foundation of Jiangsu Province (grant number BK20200212), and the Suzhou Science and Technology Program Project (grant number SYS2020185, SYS2020184, SKY2021012). This work was also supported by the Jiangsu Province Chinese Medicine Key Talent Project.

## Data availability statement

The article contains the data supporting its conclusions. The data utilized in this study can be obtained from the corresponding author upon request.

## CRediT authorship contribution statement

**Zongqi He:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization. **Qing Zhou:** Writing – review & editing, Validation, Methodology, Investigation, Data curation. **Jun Du:** Validation, Methodology, Investigation. **Yuyu Huang:** Data curation. **Bensheng Wu:** Validation, Methodology. **Zhizhong Xu:** Visualization. **Chao Wang:** Validation, Supervision, Methodology. **Xudong Cheng:** Writing – review & editing, Methodology, Formal analysis, Data curation.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We thank the Editage ([www.editage.cn](http://www.editage.cn)) for editing the English language.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e27805>.

## References

- [1] M. Gajendran, P. Loganathan, G. Jimenez, A.P. Catinella, N. Ng, C. Umapathy, N. Ziade, J.G. Hashash, A comprehensive review and update on ulcerative colitis, *Dis Mon* 65 (2019) 100851, <https://doi.org/10.1016/j.disamonth.2019.02.004>.
- [2] A.I. Sharara, S. Al Awadhi, O. Alharbi, H. Al Dhahab, M. Mounir, L. Salese, E. Singh, N. Sunna, N. Tarcha, M. Mosli, Epidemiology, disease burden, and treatment challenges of ulcerative colitis in Africa and the Middle East, *Exp Rev. Gastroenterol. Hepatol.* 12 (2018) 883–897, <https://doi.org/10.1080/17474124.2018.1503052>.
- [3] S.-C. Wei, J. Sollano, Y.T. Hui, W. Yu, P.V. Santos Estrella, L.J.Q. Llamado, N. Koram, Epidemiology, burden of disease, and unmet needs in the treatment of ulcerative colitis in Asia, *Exp Rev. Gastroenterol. Hepatol.* 15 (2021) 275–289, <https://doi.org/10.1080/17474124.2021.1840976>.
- [4] R.P. Hirten, B.E. Sands, New therapeutics for ulcerative colitis, *Annu. Rev. Med.* 72 (2021) 199–213, <https://doi.org/10.1146/annurev-med-052919-120048>.
- [5] T. Raine, S. Bonovas, J. Burisch, T. Kucharzik, M. Adamina, V. Annese, O. Bachmann, D. Bettenworth, M. Chaparro, W. Czuber-Dochan, P. Eder, P. Ellul, C. Fidalgo, G. Fiorino, P. Gionchetti, J.P. Gisbert, H. Gordon, C. Hedin, S. Holubar, M. Iacucci, K. Karmiris, K. Katsanos, U. Kopylov, P.L. Lakatos, T. Lytras, I. Lyutakov, N. Noor, G. Pellino, D. Piovani, E. Savarino, F. Selvaggi, B. Verstockt, A. Spinelli, Y. Panis, G. Doherty, ECCO guidelines on therapeutics in ulcerative colitis: medical treatment, *J. Crohn's and Colitis* 16 (2022) 2–17, <https://doi.org/10.1093/ecco-jcc/ijab178>.
- [6] K.A. Jagadeesh, K.K. Dey, D.T. Montoro, R. Mohan, S. Gazal, J.M. Engreitz, R.J. Xavier, A.L. Price, A. Regev, Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics, *Nat. Genet.* 54 (2022) 1479–1492, <https://doi.org/10.1038/s41588-022-01187-9>.
- [7] D. Saul, L. Leite Barros, A.Q. Wixom, B. Gellhaus, H.R. Gibbons, W.A. Faubion, R.L. Kosinsky, Cell type-specific induction of inflammation-associated genes in Crohn's disease and colorectal cancer, *Indian J. Manag. Sci.* 23 (2022) 3082, <https://doi.org/10.3390/ijms23063082>.
- [8] Y. Lu, K. Li, Y. Hu, X. Wang, Expression of immune related genes and possible regulatory mechanisms in Alzheimer's disease, *Front. Immunol.* 12 (2021) 768966, <https://doi.org/10.3389/fimmu.2021.768966>.
- [9] C.S. Smillie, M. Biton, J. Ordovas-Montanes, K.M. Sullivan, G. Burgin, D.B. Graham, R.H. Herbst, N. Rogel, M. Slyper, J. Waldman, M. Sud, E. Andrews, G. Velonias, A.L. Haber, K. Jagadeesh, S. Vickovic, J. Yao, C. Stevens, D. Dionne, L.T. Nguyen, A.-C. Villani, M. Hofree, E.A. Creasey, H. Huang, O. Rozenblatt-Rosen, J.J. Garber, H. Khalili, A.N. Desch, M.J. Daly, A.N. Ananthakrishnan, A.K. Shalek, R.J. Xavier, A. Regev, Intra- and inter-cellular rewiring of the human colon during ulcerative colitis, *Cell* 178 (2019) 714–730.e22, <https://doi.org/10.1016/j.cell.2019.06.029>.
- [10] V. Mitsialis, S. Wall, P. Liu, J. Ordovas-Montanes, T. Parmet, M. Vukovic, D. Spencer, M. Field, C. McCourt, J. Toothaker, A. Bousvaros, A.K. Shalek, L. Kean, B. Horwitz, J. Goldsmith, G. Tseng, S.B. Snapper, L. Konnikova, S. Ballal, S. Bonilla, R. Fawaz, L.N. Fishman, A. Flores, V. Fox, A.S. Grover, L. Higuchi, S. Huh, S. Kahn, C. Lee, M. Mobassaleh, J. Ouahed, R.G. Pleskow, B. Regan, P.A. Rufo, S. Sabharwal, J. Silverstein, M. Verhave, A. Wolf, L. Zimmerman, N. Zitomersky, J.R. Allegretti, P. De Silva, S. Friedman, M. Hamilton, J. Korzenik, F. Makrauer, B.-A. Norton, R.W. Winter, Single-cell analyses of colon and blood reveal distinct immune cell signatures of ulcerative colitis and Crohn's disease, *Gastroenterology* 159 (2020) 591–608.e10, <https://doi.org/10.1053/j.gastro.2020.04.074>.

- [11] M. Uzzan, J.C. Martin, L. Mesin, A.E. Livanos, T. Castro-Dopico, R. Huang, F. Petralia, G. Magri, S. Kumar, Q. Zhao, A.K. Rosenstein, M. Tokuyama, K. Sharma, R. Ungaro, R. Kosoy, D. Jha, J. Fischer, H. Singh, M.E. Keir, N. Ramamoorthi, W.E.O. Gorman, B.L. Cohen, A. Rahman, F. Cossarini, A. Seki, L. Leyre, S. T. Vaquero, S. Gurunathan, E.K. Grasset, B. Losic, M. Dubinsky, A.J. Greenstein, Z. Gottlieb, P. Legnani, J. George, H. Irizar, A. Stojmivovic, C. Brodmerkel, A. Kasarkis, B.E. Sands, G. Furtado, S.A. Lira, Z.K. Tuong, H.M. Ko, A. Cerutti, C.O. Elson, M.R. Clatworthy, M. Merad, M. Suárez-Fariñas, C. Argmann, J. A. Hackney, G.D. Victora, G.J. Randolph, E. Kenigsberg, J.F. Colombel, S. Mehandru, Ulcerative colitis is characterized by a plasmablast-skewed humoral response associated with disease activity, *Nat. Med.* 28 (2022) 766–779, <https://doi.org/10.1038/s41591-022-01680-y>.
- [12] D. Corridoni, A. Antanaviciute, V. Gupta, D. Fawcner-Corbett, A. Aulicino, M. Jagielowicz, K. Parikh, E. Repapi, S. Taylor, D. Ishikawa, R. Hatano, T. Yamada, W. Xin, H. Slawinski, R. Bowden, G. Napolitani, O. Brain, C. Morimoto, H. Koohy, A. Simmons, Single-cell atlas of colonic CD8+ T cells in ulcerative colitis, *Nat. Med.* 26 (2020) 1480–1490, <https://doi.org/10.1038/s41591-020-1003-4>.
- [13] J.M. Serigado, J. Foulke-Abel, W.C. Hines, J.A. Hanson, J. In, O. Kovbasnjuk, Ulcerative Colitis, Novel epithelial insights provided by single cell RNA sequencing, *Front. Med.* 9 (2022) 868508, <https://doi.org/10.3389/fmed.2022.868508>.
- [14] T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, clusterProfiler 4.0: a universal enrichment tool for interpreting omics data, *Innovation* 2 (2021) 100141, <https://doi.org/10.1016/j.xinn.2021.100141>.
- [15] A. Colaprico, T.C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T.S. Sabedot, T.M. Malta, S.M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, H. Noushmehr, TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data, *Nucleic Acids Res.* 44 (2016) e71, <https://doi.org/10.1093/nar/gkv1507>.
- [16] K. Hu, Become competent in generating RNA-seq heat maps in one day for novices without prior R experience, *Methods Mol. Biol.* 2239 (2021) 269–303, [https://doi.org/10.1007/978-1-0716-1084-8\\_17](https://doi.org/10.1007/978-1-0716-1084-8_17).
- [17] S. Hänzelmann, R. Castelo, J. Guinney, GSEA: gene set variation analysis for microarray and RNA-seq data, *BMC Bioinf.* 14 (2013) 7, <https://doi.org/10.1186/1471-2105-14-7>.
- [18] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, *BMC Bioinf.* 9 (2008) 559, <https://doi.org/10.1186/1471-2105-9-559>.
- [19] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, *J. Stat. Software* 33 (2010) 1–22.
- [20] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, M. Müller, pROC: an open-source package for R and S+ to analyze and compare ROC curves, *BMC Bioinf.* 12 (2011) 77, <https://doi.org/10.1186/1471-2105-12-77>.
- [21] K. Ito, D. Murphy, Application of ggplot2 to pharmacometric graphics, *CPT Pharmacometrics Syst. Pharmacol.* 2 (2013) e79, <https://doi.org/10.1038/psp.2013.56>.
- [22] Y. Kwon, Y.Z. Kim, Y.H. Choe, M.J. Kim, Increased monocyte abundance as a marker for relapse after discontinuation of biologics in inflammatory bowel disease with deep remission, *Front. Immunol.* 13 (2022) 996875, <https://doi.org/10.3389/fimmu.2022.996875>.
- [23] G. Xue, L. Hua, D. Liu, M. Zhong, Y. Chen, B. Zhou, Y. Xie, J. Li, Tim-4 expressing monocytes as a novel indicator to assess disease activity and severity of ulcerative colitis, *Life Sci.* 269 (2021) 119077, <https://doi.org/10.1016/j.lfs.2021.119077>.
- [24] C. Mei, X. Wang, F. Meng, X. Zhang, L. Gan, Y. Wang, X. Sun, CD30L+ classical monocytes play a pro-inflammatory role in the development of ulcerative colitis in patients, *Mol. Immunol.* 138 (2021) 10–19, <https://doi.org/10.1016/j.molimm.2021.06.016>.
- [25] S. Furukawa, Y. Ikeda, S. Yagi, T. Miyake, K. Shiraishi, K. Tange, Y. Hashimoto, K. Mori, T. Ninomiya, S. Suzuki, N. Shibata, H. Murakami, K. Ohashi, A. Hasebe, H. Tomida, Y. Yamamoto, E. Takeshita, Y. Hiata, Association between peripheral blood monocyte count and mucosal healing in Japanese patients with ulcerative colitis, *Clin. Transl. Gastroenterol.* 12 (2021) e00429, <https://doi.org/10.14309/ctg.00000000000000429>.
- [26] H. Lee, Y.S. Son, M.-O. Lee, J.-W. Ryu, K. Park, O. Kwon, K.B. Jung, K. Kim, T.Y. Ryu, A. Baek, J. Kim, C.-R. Jung, C.-M. Ryu, Y.-J. Park, T.-S. Han, D.-S. Kim, H.-S. Cho, M.-Y. Son, Low-dose interleukin-2 alleviates dextran sodium sulfate-induced colitis in mice by recovering intestinal integrity and inhibiting AKT-dependent pathways, *Theranostics* 10 (2020) 5048–5063, <https://doi.org/10.7150/thno.41534>.
- [27] Y. Zhu, S. Yang, N. Zhao, C. Liu, F. Zhang, Y. Guo, H. Liu, CXCL8 chemokine in ulcerative colitis, *Biomed. Pharmacother.* 138 (2021) 111427, <https://doi.org/10.1016/j.biopha.2021.111427>.
- [28] A. Salas, C. Hernandez-Rocha, M. Duijvestein, W. Faubion, D. McGovern, S. Vermeire, S. Vetrano, N. Vande Casteele, JAK-STAT pathway targeting for the treatment of inflammatory bowel disease, *Nat. Rev. Gastroenterol. Hepatol.* 17 (2020) 323–337, <https://doi.org/10.1038/s41575-020-0273-0>.
- [29] C. Harris, J.R.F. Cummings, JAK1 inhibition and inflammatory bowel disease, *Rheumatology* (2021) 60, <https://doi.org/10.1093/rheumatology/keaa896>, ii45–ii51.
- [30] M.A. Al Barashdi, A. Ali, M.F. McMullin, K. Mills, Protein tyrosine phosphatase receptor type C (PTPRC or CD45), *J. Clin. Pathol.* 74 (2021) 548–552, <https://doi.org/10.1136/jclinpath-2020-206927>.
- [31] P. Krause, S.P. Zahner, G. Kim, R.B. Shaikh, M.W. Steinberg, M. Kronenberg, The tumor necrosis factor family member TNFSF14 (LIGHT) is required for resolution of intestinal inflammation in mice, *Gastroenterology* 146 (2014) 1752–1762.e4, <https://doi.org/10.1053/j.gastro.2014.02.010>.
- [32] M.-Y. Wu, L. Liu, E.-J. Wang, H.-T. Xiao, C.-Z. Cai, J. Wang, H. Su, Y. Wang, J. Tan, Z. Zhang, J. Wang, M. Yao, D.-F. Ouyang, Z. Yue, M. Li, Y. Chen, Z.-X. Bian, J.-H. Lu, PI3KC3 complex subunit NRBF2 is required for apoptotic cell clearance to restrict intestinal inflammation, *Autophagy* 17 (2021) 1096–1111, <https://doi.org/10.1080/15548627.2020.1741332>.
- [33] A.L. Haber, M. Biton, N. Rogel, R.H. Herbst, K. Shekhar, C. Smillie, G. Burgin, T.M. Delorey, M.R. Howitt, Y. Katz, I. Tirosh, S. Beyaz, D. Dionne, M. Zhang, R. Raychowdhury, W.S. Garrett, O. Rozenblatt-Rosen, H.N. Shi, O. Yilmaz, R.J. Xavier, A. Regev, A single-cell survey of the small intestinal epithelium, *Nature* 551 (2017) 333–339, <https://doi.org/10.1038/nature24489>.
- [34] Z. Xie, E.C. Chan, K.M. Druey, R4 regulator of G protein signaling (RGS) proteins in inflammation and immunity, *AAPS J.* 18 (2016) 294–304, <https://doi.org/10.1208/s12248-015-9847-0>.
- [35] D.L. Gibbons, L. Abeler-Dörner, T. Raine, I.-Y. Hwang, M. Wencker, L. Deban, C.E. Rudd, P.M. Irving, H. Kehrl, A.C. Hayday, Regulator of G Protein Signalling-1 (RGS1) Selectively Regulates Gut T Cell Trafficking and Colitic Potential, 2012.
- [36] D.L. Gibbons, L. Abeler-Dörner, T. Raine, I.-Y. Hwang, A. Jandke, M. Wencker, L. Deban, C.E. Rudd, P.M. Irving, J.H. Kehrl, A.C. Hayday, Cutting edge: regulator of G protein signaling-1 selectively regulates gut T cell trafficking and colitic potential, *J. Immunol.* 187 (2011) 2067–2071, <https://doi.org/10.4049/jimmunol.1100833>.
- [37] X. Lu, K. Zhang, W. Jiang, H. Li, Y. Huang, M. Du, J. Wan, Y. Cao, L. Du, X. Liu, W. Pan, Single-cell RNA sequencing combined with whole exome sequencing reveals the landscape of the immune pathogenic response to chronic mucocutaneous candidiasis with STAT1 GOF mutation, *Front. Immunol.* 13 (2022) 988766, <https://doi.org/10.3389/fimmu.2022.988766>.
- [38] J. Pekow, U. Dougherty, Y. Huang, E. Gometz, J. Nathanson, G. Cohen, S. Levy, M. Kocherginsky, N. Venu, M. Westerhoff, J. Hart, A.E. Noffsinger, S.B. Hanauer, R.D. Hurst, A. Fichera, L.J. Joseph, Q. Liu, M. Bissonnette, Gene signature distinguishes patients with chronic ulcerative colitis harboring remote neoplastic lesions, *Inflamm. Bowel Dis.* 19 (2013) 461–470, <https://doi.org/10.1097/MIB.0b013e3182802bac>.
- [39] C. Lasconi, M.C. Pahl, D.L. Cousminer, C.A. Doege, A. Chesi, K.M. Hodge, M.E. Leonard, S. Lu, M.E. Johnson, C. Su, R.K. Hammond, J.A. Pippin, N.A. Terry, L. R. Ghanem, R.L. Leibel, A.D. Wells, S.F.A. Grant, Variant-to-Gene-Mapping analyses reveal a role for the hypothalamus in genetic susceptibility to inflammatory bowel disease, *Cellular and Molecular Gastroenterology and Hepatology* 11 (2021) 667–682, <https://doi.org/10.1016/j.jcmgh.2020.10.004>.
- [40] A. D'Addabbo, O. Palmieri, R. Maglietta, A. Latiano, S. Mukherjee, V. Annesse, N. Ancona, Discovering genetic variants in Crohn's disease by exploring genomic regions enriched of weak association signals, *Dig. Liver Dis.* 43 (2011) 623–631, <https://doi.org/10.1016/j.dld.2011.02.010>.
- [41] G.L. Wojcik, C. Marie, M.M. Abhyankar, N. Yoshida, K. Watanabe, A.J. Mentzer, T. Carstensen, J. Mychaleckyj, B.D. Kirkpatrick, S.S. Rich, P. Concannon, R. Haque, G.C. Tsokos, W.A. Petri, P. Duggal, Genome-wide association study reveals genetic link between diarrhea-associated entamoeba histolytica infection and inflammatory bowel disease, *mBio* 9 (2018) e01668-18, <https://doi.org/10.1128/mBio.01668-18>.

- [42] D. Bouzid, H. Fourati, A. Amouri, I. Marques, O. Abida, S. Haddouk, M.B. Ayed, N. Tahri, C. Penha-Gonçalves, H. Masmoudi, The CREM gene is involved in genetic predisposition to inflammatory bowel disease in the Tunisian population, *Hum. Immunol.* 72 (2011) 1204–1209, <https://doi.org/10.1016/j.humimm.2011.10.002>.
- [43] L. Krabbendam, B.A. Heesters, C.M.A. Kradolfer, N.J.E. Haverkate, M.A.J. Becker, C.J. Buskens, W.A. Bemelman, J.H. Bernink, H. Spits, CD127+ CD94+ innate lymphoid cells expressing granulysin and perforin are expanded in patients with Crohn's disease, *Nat. Commun.* 12 (2021) 5841, <https://doi.org/10.1038/s41467-021-26187-x>.
- [44] J. Qian, Y. Gao, Y. Lai, Z. Ye, Y. Yao, K. Ding, J. Tong, H. Lin, G. Zhu, Y. Yu, H. Ding, D. Yuan, J. Chu, F. Chen, X. Liu, Single-cell RNA sequencing of peripheral blood mononuclear cells from acute myocardial infarction, *Front. Immunol.* 13 (2022) 908815, <https://doi.org/10.3389/fimmu.2022.908815>.