Clinical and Experimental Gastroenterology

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

# Aimed at Subtype Discrimination but Yielding a Shared Marker: Integrative Analysis of Blood Transcriptomes Reveals Upregulated *TLR5* as a Potential Biomarker for IBD

Arman Mokaram Doust Delkhah

Department of Genetics, Mashhad, Iran

Correspondence: Arman Mokaram Doust Delkhah, Email arman.mokaram@gmail.com



**Background:** Despite the similarities, Crohn's disease (CD) and ulcerative colitis (UC), the two major subtypes of inflammatory bowel disease (IBD), exhibit distinctions. The increasing burden of IBD necessitates discovering novel diagnostic markers. Considering the importance of distinguishing between CD and UC in selecting therapeutic strategies in clinical settings, this investigation focused on identifying subtype-specific blood biomarkers.

**Methods:** The discovery set was formed by integrating five blood transcriptomic datasets, including GSE119600, GSE126124, GSE94648, GSE86434, and GSE71730, which incorporated samples from CD, UC, and controls. After determining DEGs in CD and UC, they were separately filtered according to WGCNA and then analyzed by LASSO and RF algorithms. Eventually, ROC analysis of the diagnostic performances was conducted independently in the datasets used for discovery. Moreover, ROC analysis was implemented in independent cohorts to assess the generalizability of findings.

**Results:** Initially, the identified subtype-specific candidate biomarkers included *INPPL1, TLR5, SLC9A8, IMPDH1*, and *GRK6* for CD, and *IL4R, ACAA1, NARF*, and *RRM2* for UC. However, external validation only accentuated the promising diagnostic potential of *TLR5* for CD. However, *TLR5* was nonspecific for CD and also acts as a potential biomarker for UC. While this study did not unveil subtype-specific diagnostic markers capable of stratifying CD from UC, the upregulated *TLR5* was identified as a shared biomarker for both subtypes.

**Conclusion:** This integrative analysis of blood transcriptomes diverged from its initial purpose, the identification of subtype-specific biomarkers, and demonstrated that *TLR5* exhibits reproducible diagnostic efficacy for IBD.

**Keywords:** biomarkers, blood, IBD, TLR5

## Introduction

Inflammatory bowel disease (IBD), encompassing a group of disorders identified by chronic inflammation in the gastrointestinal (GI) tract, is expected to affect a growing number of individuals in the approaching future, stemming from the high incidence rate in developing countries and the growing number of older adults in developed countries.[1] Recently, many efforts have been made to facilitate the diagnosis of IBD, including the investigation of novel markers[2,3] as well as the optimization of conventional diagnostic methods.[4] Considering blood transcriptomic data as a valuable diagnostic source, this study aimed to discover specific biomarkers separately for CD and UC, the most prominent subtypes of the disease.

Despite the wide range of similarities in the pathogenic signatures of CD and UC, they have marked differences in manifestations. Inflammation occurs extensively in the GI tract in CD, while it mainly affects the colon and rectum in UC.[5] Considering the variability in the efficacy of different therapeutic strategies for these two subtypes of the disorder,[6] identification of biomarkers distinguishing CD from UC is beneficial for early selection of an effective therapeutic plan.

https://doi.org/10.2147/CEG.S584999
Received: 9 December 2025
Accepted: 25 April 2026
Published: 4 May 2026

Discovering specific biomarkers for CD or UC has attracted growing interest. For instance, in a transcriptomic analysis followed by wet-lab validation, the mucosal expression of *PI3, ANXA1*, and *VDR* demonstrated diagnostic potential for differentiating CD and UC.[7] By studying peripheral blood routine parameters (PBRPs), the multi-layer perceptron artificial neural network model (MLP-ANN) model was suggested as an efficient machine learning (ML) method for discriminating CD and UC.[8] A study on serum levels of amino acid metabolites suggested that taurine, homocitrulline, and kynurenine, being increased in CD patients compared to patients with UC, hold diagnostic potential.[9] In another effort, on the basis of partial least squares discriminant analysis (PLS-DA) analysis, a supervised learning model was developed, which was suggested to hold the potential to distinguish CD from UC using RNA-seq analysis of endoscopic biopsy specimens.[10] Moreover, using dual-layer spectral detector computed tomography enterography (CTE), the quantitative energy spectrum parameters of enteric phase normalized ID (NID) and portal phase effective atomic number (Z-eff) were suggested to be effective in differentiating CD from UC.[11] Collectively, these investigations underscore the imperative of developing a diagnostic method for distinguishing between the two major subtypes of IBD. This is beneficial for treatment decision-making, and it may also lead to the identification of subtype-specific diagnostic markers that may possess greater sensitivity and specificity, which ensures their applicability.

Despite recent advancements in the management of IBD, the generalizability and reproducibility of biomarkers across cohorts have remained a challenge in IBD.[3] Given the noninvasiveness and accessibility of blood specimens,[12] they are a valuable source and a potential biological substitute for intestinal diagnostic assays. Compared to endoscopy[13] as the central diagnostic method,[14] evaluation of mRNA levels in blood is neither expensive nor invasive.[15] With this substantial value in mind, herein five datasets incorporating blood transcriptomes for both CD and UC, along with healthy controls, were integrated (GSE119600, GSE126124, GSE94648, GSE86434, GSE71730). After determining differentially expressed genes in CD and UC, weighted gene co-expression network analysis (WGCNA) was performed independently for these subtypes. The overlay of these two methods was separately subjected to random forest (RF) and least absolute shrinkage and selection operator (LASSO) regression. Lastly, ROC analysis was carried out to determine the diagnostic performance of the identified genes. Although this pipeline did not lead to the identification of promising unique biomarkers for CD and UC, upregulated *TLR5* was highlighted as a promising biomarker for CD and, to a lesser extent, for UC.

# Methods

## Ethical Considerations

This study retrospectively re-analyzed de-identified, publicly available transcriptomic datasets obtained from the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/). There was no direct contact with study participants, so individual informed consent was not applicable for this reanalysis. Furthermore, in this study no in vivo or in vitro experiments requiring ethical approval were conducted. The transcriptomic data that has been utilized for biomarker discovery and validation has obtained ethical approval.

## Data Acquisition

The GEO database[16] was analyzed for obtaining transcriptomic data. After the retrieval of 12 datasets from GEO (Table 1 and Table 2), a harmonization step was carried out using the sample information that is retrievable using the GEOquery[17] package. This included setting the source as blood and removing patients with disorders other than IBD. Furthermore, principal component analysis (PCA) was conducted to detect outliers. While one sample was excluded in each of GSE86434 and GSE112057, the outlier detection methods did not mark any obvious outlier samples in the other datasets (Supplementary Figure S1).

Five datasets, namely GSE119600, GSE126124, GSE94648, GSE86434, and GSE71730, were integrated since they included both subtypes of IBD and contained expression data for more genes than the other datasets. This allowed for a more comprehensive biomarker investigation while minimizing gene loss during integration. The details of these five datasets are presented in Table 1. To this end, the mean value of probe-level measurements was considered as gene-level expression in each dataset using corresponding platform annotation. After merging the five datasets according to the gene

**Table 1** Integrated Transcriptomic Datasets Used for Discovery Analysis

| GEO Accession | Source | Year | Control | CD | UC | Platform | Number of Genes | Reference |
|---|---|---|---|---|---|---|---|---|
| GSE119600 | Blood | 2019 | 47 | 95 | 93 | GPL10588 | 31,320 | [18] |
| GSE126124 | Blood | 2019 | 32 | 39 | 18 | GPL6244 | 23,307 | [19] |
| GSE94648 | Blood | 2017 | 22 | 50 | 25 | GPL19109 | 22,880 | [20] |
| GSE86434 | Blood | 2016 | 24 | 22 | 22 | GPL10558 | 31,171 | [21] |
| GSE71730 | PBMCs | 2015 | 10 | 22 | 15 | GPL570 | 22,880 | [22] |

**Abbreviation**: PBMCs, Peripheral blood mononuclear cells.

**Table 2** Datasets Utilized for External Validation

| GEO Accession | Source | Control | CD | UC | IBD | Platform | Number of Genes | Reference |
|---|---|---|---|---|---|---|---|---|
| GSE276395 | Blood | 51 | | | 40 | GPL23126 | 18,843 | [23] |
| GSE169568 | Blood | 30 | 52 | 58 | | GPL10588 | 11,646 | [24] |
| GSE177044 | Blood | 311 | | 481 | | GPL16791 | 24,787 | [25] |
| GSE112057 | Blood | 12 | 59 | 15 | | GPL11154 | 11,923 | [26] |
| GSE100833 | Blood | 49 | 589 | | | GPL13158 | 20,277 | [27] |
| GSE33943 | PBLs | 13 | | | 45 | GPL570 | 22,880 | [28] |
| GSE3365 | PBMCs | 42 | 59 | 26 | | GPL96 | 13,237 | [29] |

**Abbreviations**: PBMCs, Peripheral blood mononuclear cells; PBLs, Peripheral blood leukocytes.

symbols, batch effects were minimized using the ComBat method of the SVA[30] package, and PCA plots were illustrated using the scatterplot3d[31] package. Lastly, the integrated data incorporated expression data of 15,932 genes across 536 samples, including 228 CD patients, 173 UC patients, and 135 healthy controls (HCs).

## Differential Expression Analysis

To enhance the focus of the differential expression (DE) analysis on CD and UC, the integrated data were divided into two datasets, including one containing CD samples and HCs and the other containing UC samples and HCs. Using the limma[32] package, differentially expressed genes (DEGs) in CD and UC were then identified using a $|\log_2$ fold change| (logFC) threshold above 0.2 and an adjusted p-value below 0.05. The volcano plots demonstrating DEGs were illustrated using the ggplot2[33] package.

## Functional Enrichment Analysis

After reaching DEGs in CD and UC, Gene Ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway analyses were carried out to provide insights on the mechanisms through which they contribute to the pathogenesis of these two disorders using the clusterProfiler[34] package. Subsequently, the top terms in each GO category and KEGG pathway were illustrated using the ggplot2 package according to their enrichment and statistical significance (adjusted p-value).

## Mapping PPIs and Hub Gene Identification of miRNAs

The STRING[35] database was used to retrieve protein-protein interactions (PPIs) of DEGs in CD and UC using a confidence threshold of 0.7. The PPI network for DEGs in CD and UC was constructed, and hub genes were defined as nodes ranking in the top decile based on the number of interactions. Finally, PPI networks of hub genes were constructed using the Cytoscape v3.10.3 app.[36]

After identifying hub genes in CD and UC, regulatory miRNAs for these gene sets were determined using the miRDB[37] database. To include miRNAs in the miRNA-mRNA regulatory networks of hub genes, multiple criteria were considered. These included having no more than 2,000 predicted targets in the genome, a prediction score exceeding 80, and being regulators of multiple hub genes. Finally, the regulatory networks were mapped using the Cytoscape app.

## WGCNA

WGCNA was conducted separately for CD and UC to identify the subtype-specific modules having a statistically significant correlation with these two disorders. After evaluating the optimal power, block-wise construction of modules was separately carried out for these two major subtypes of IBD using the WGCNA[38] package. Modules with their correlation p-value below 0.001 were pinpointed as significant. Then, KEGG pathway enrichment analysis was performed for these gene modules, and highly enriched pathways were selected based on the combined score from the Enrichr[39] database.

## Machine Learning-Based Discovery of Potential Biomarkers

The consistent results from DE analysis and WGCNA were subjected to machine learning methods to identify candidate biomarkers for CD and UC. Upregulated genes were compared with positively correlated modules, while downregulated genes were compared with negatively correlated modules to achieve this goal. Using genes that passed this step of filtration, RF and LASSO were independently performed in parallel for CD and UC. The randomForest[40] and glmnet[41] packages were utilized for the implementation of RF and LASSO, respectively. The top 20 genes identified by each machine learning method were then compared, and those identified by both algorithms were singled out. Using the pROC package, candidate genes were then subjected to ROC analysis to evaluate their diagnostic performance in the integrated data.

## Testing Biomarker Performance in External Cohorts

The diagnostic efficacy of candidate biomarkers for CD and UC was evaluated in external datasets to assess the reproducibility of the findings. To this end, ROC analysis was conducted in three common datasets, including GSE169568, GSE112057, and GSE3365, for candidate biomarkers of CD and UC. Meanwhile, GSE100833 and GSE177044, which encompassed only CD and UC samples, respectively, were utilized to validate their corresponding conditions. Furthermore, the diagnostic efficacy of the genes was also evaluated independently in datasets used for discovery purposes. Overall, a total of nine datasets were examined for each of the two subtypes to provide insights into the reproducibility of specific biomarkers for CD and UC. The details of these datasets are presented in Table 2.

## Expression Pattern of Selected Biomarkers Across Datasets

Independent DE analysis was performed on the 12 retrieved datasets to evaluate whether the direction of alterations is consistent. Violin plots demonstrating the expression level of *TLR5* across these datasets were subsequently illustrated using the ggplot2[33] package.

# Results

## Identification of Disease-Expressed Genes in CD and UC

Following the batch effect correction of the integrated data (Figure 1a–c), a total of 746 and 504 DEGs were identified in CD and UC, respectively (Figure 1d and e). The DEGs in CD encompassed 563 upregulated and 183 downregulated genes, whereas DEGs in UC included 388 upregulated and 116 downregulated genes. The comprehensive list of DEGs in CD and UC, along with the results of DE analysis and their mean expression in case and control groups, are presented in Supplementary Tables S1–S4.

## GO and KEGG Pathways

Taking a broad view, functional enrichment analysis of DEGs in CD and UC led to the identification of a similar spectrum of terms but with different orders of enrichment and significance. This highlights the similarity of pathogenic signatures in CD and UC. GO analysis categorizes terms in three distinct classifications, namely biological processes (BP), cellular components (CC), and molecular functions (MF). Among the top 10 terms in the BP, mononuclear cell differentiation (GO:1903131) and immune response-regulating cell surface receptor signaling pathway (GO:0002768) were only enriched in CD, while positive regulation of leukocyte activation (GO:0002696) and positive regulation of cell

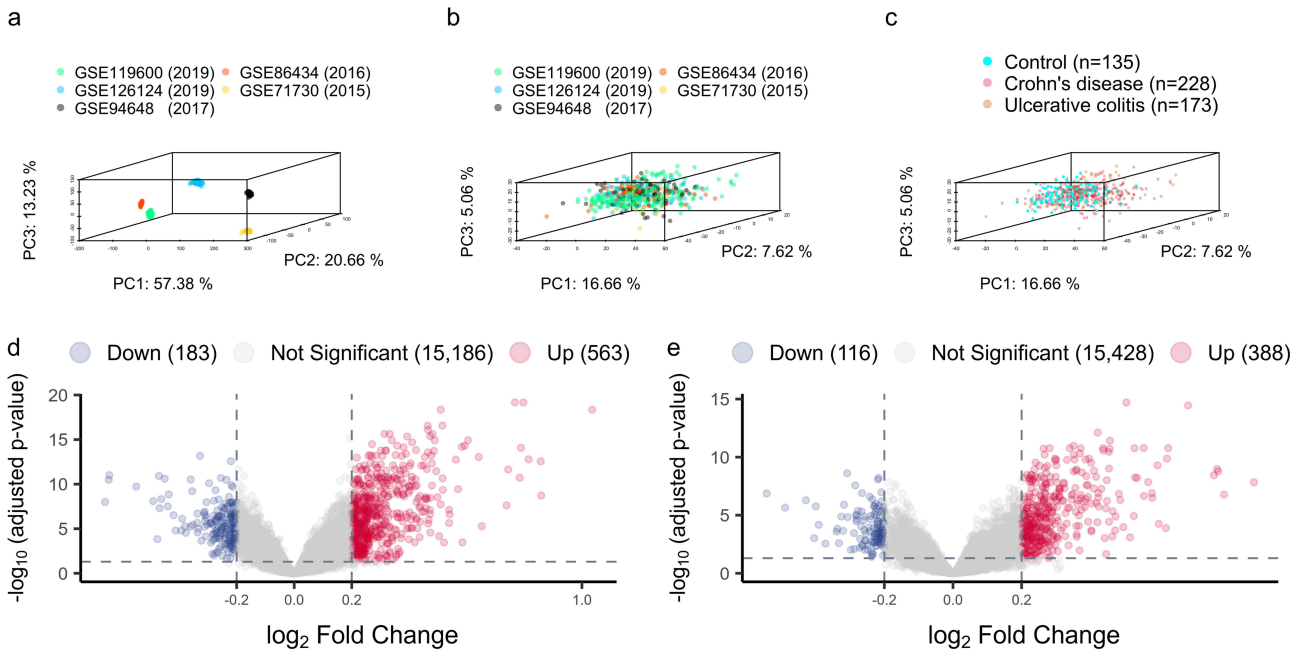


**Figure 1** DEGs in CD and UC. (**a**) PCA plot illustrating the distribution of samples in the integrated data before batch effect correction. (**b–c**) PCA plots illustrating the distribution of samples in the integrated data following batch effect correction. (**d**) Differentially expressed genes in CD. (**e**) Differentially expressed genes in UC.

activation (GO:0050867) were detected only in UC. In spite of these limited differences, the remaining terms were shared between CD and UC, notably immune response-activating signaling pathways (GO:0002757), regulation of innate immune response (GO:0045088), positive regulation of cytokine production (GO:0001819), and leukocyte-mediated immunity (GO:0002443), which were highly enriched BP terms in these two subtypes of IBD. In the CC category, vesicle lumen, cytoplasmic vesicle lumen, and secretory granule lumen were the top three terms identified in both CD and UC. A slight difference in the spectrum of CC terms was the identification of membrane rafts (GO:0045121) only in CD and membrane microdomains (GO:0098857) only in UC. Monosaccharide binding (GO:0048029), cytokine receptor activity (GO:0004896), and cytokine binding (GO:0019955) were among the top MF terms only in CD, whereas carbohydrate kinase activity (GO:0019200), NADP+ nucleosidase activity (GO:0050135), and NAD+ nucleotidase (cyclic ADP-ribose generating, GO:0061809) were highlighted only in UC. In spite of these distinctions, the top four terms in the MF category, including carboxylic acid binding (GO:0031406), organic acid binding (GO:0043177), immune receptor activity (GO:0140375), and carbohydrate binding (GO:0030246), were common to both subtypes. Hematopoietic cell lineage, neutrophil extracellular trap formation, and NOD-like receptor signaling were the three shared KEGG pathways between CD and UC among the top 10 terms. Additionally, NOD-like receptor signaling and hematopoietic cell lineage were identified as the most enriched pathways in CD and UC, respectively. Overall, functional enrichment analysis indicates that the two main types of IBD share similar pathogenic signatures (Figure 2).

## PPI and miRNA-mRNA Regulatory Networks

Using 746 DEGs in CD, a PPI network with 411 nodes and 1,483 edges was constructed. Hub genes were defined as the top 10% of nodes based on the number of interactions. The hub genes of the network included 43 genes, with interactions ranging from 59 for downregulated *CD4*, the central hub gene, to 18 for *CD69, MPO, EIF2AK2, CD274, IFIT2, MAPK14, KLRB1, HSP90AB1, H3C12*, and *ELANE* (Figure 3a). The retrieval of miRNAs for these hub genes resulted in the identification of multiple members of the miR-17, miR-548, miR-15/107, miR-130, and miR-301 families as key regulators (Figure 3b). The comprehensive list of miRNAs targeting hub genes of the PPI network of DEGs in CD is presented in Supplementary Table S5.

The PPI network for UC included 224 nodes that were connected through 581 interactions. The interaction counts for the 21 hub genes in this network ranged from 42 for downregulated *CD4* to 12 for *RPS28* (Figure 3c). Members of the


https://doi.org/10.2147/CEG.S584999

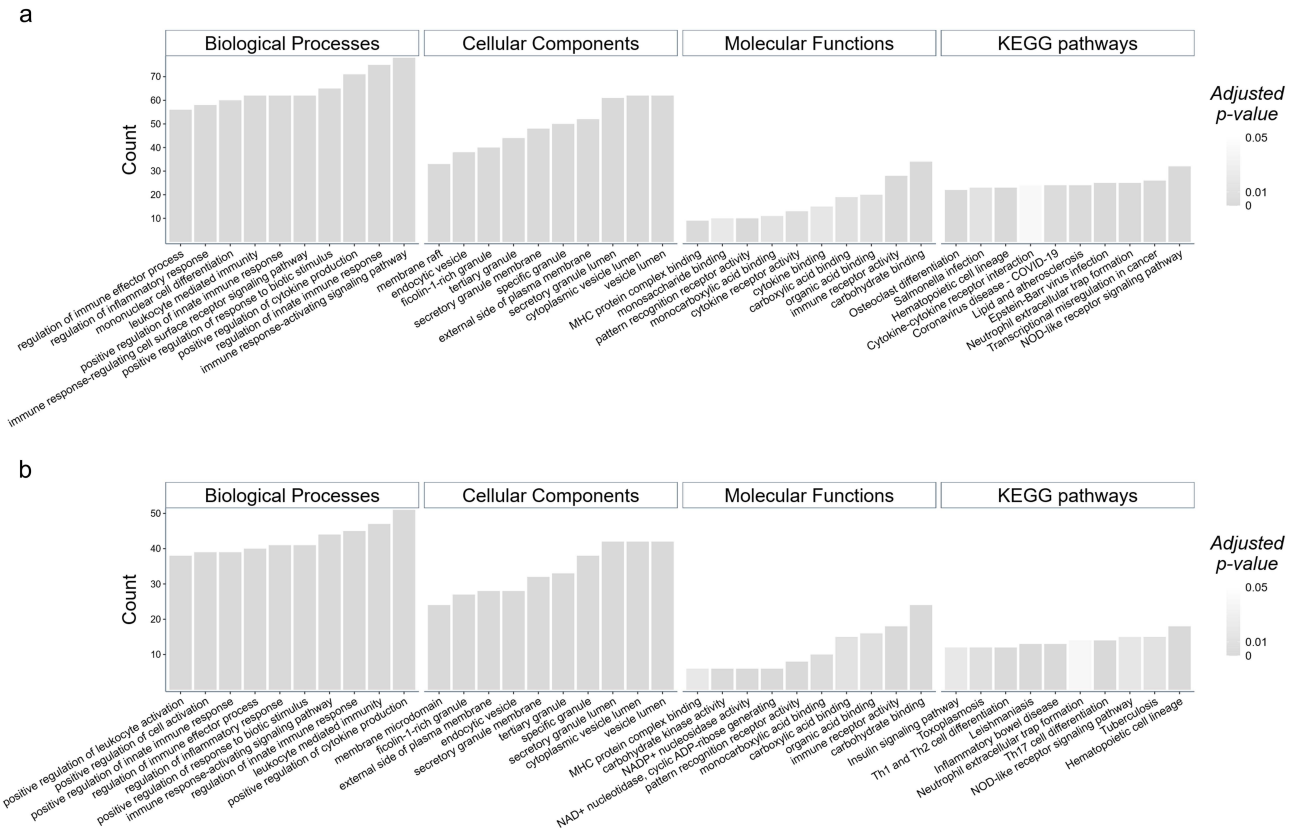


**Figure 2** Functional enrichment analysis of DEGs. (**a**) GO and KEGG pathway analysis of DEGs in CD. (**b**) GO and KEGG pathway analysis of DEGs in UC.

miR-15/107, miR-181, and miR-513 families were identified as regulators of the hub genes in UC (Figure 3d). The list of regulatory miRNAs for the hub genes in the PPI network of DEGs in UC is presented in Supplementary Table S6.

## Identification of CD-Specific and UC-Specific Modules

After ensuring no outlier samples in the integrated CD and healthy control (HC) data (Supplementary Figure S2), 8 was selected as the optimal soft thresholding power for the block-wise construction of modules (Figure 4a). The analysis assigned the 15,932 genes to 24 modules, numbered 0 to 23 (Figure 4b). Given that the statistical significance threshold of a p-value below 0.001 was utilized as the cut-off threshold for considering a module prominent, modules 2, 9, 10, 11, 18, and 20 were positively correlated modules, and modules 1 and 13 were negatively correlated modules further utilized for filtration of DEGs. Modules 2, 9, 10, 11, 18, and 20 were enriched in osteoclast differentiation, influenza A, Salmonella infection, other glycan degradation, the NOD-like receptor signaling pathway, and cell cycle, respectively. In addition, negatively correlated modules, including modules 1 and 13, were enriched in herpes simplex virus 1 infection and graft-versus-host disease (GVHD), respectively (Figure 4c).

After confirming the absence of outliers in the integrated data encompassing UC and HC samples using hierarchical clustering (Supplementary Figure S3), the soft thresholding power of 8 was selected for the block-wise module construction (Figure 5a). The analysis assigned 15,932 genes to 19 modules, numerically labeled from 0 to 18 (Figure 5b). Using a p-value threshold of 0.001 as the module selection criterion, modules 2 and 15 were identified as positively correlated, while 1, 11, and 14 were negatively correlated with UC. KEGG pathway analysis highlighted osteoclast differentiation and Fc gamma R-mediated phagocytosis as enriched pathways in module 2. Enrichment analysis of module 15 underlined the cell cycle and p53 signaling pathways. Meanwhile, highly enriched pathways in the selected negatively correlated modules included herpes simplex virus 1 infection, mismatch repair, ribosome, RNA transport, Th1 and Th2 cell differentiation, and T cell receptor signaling (Figure 5c).

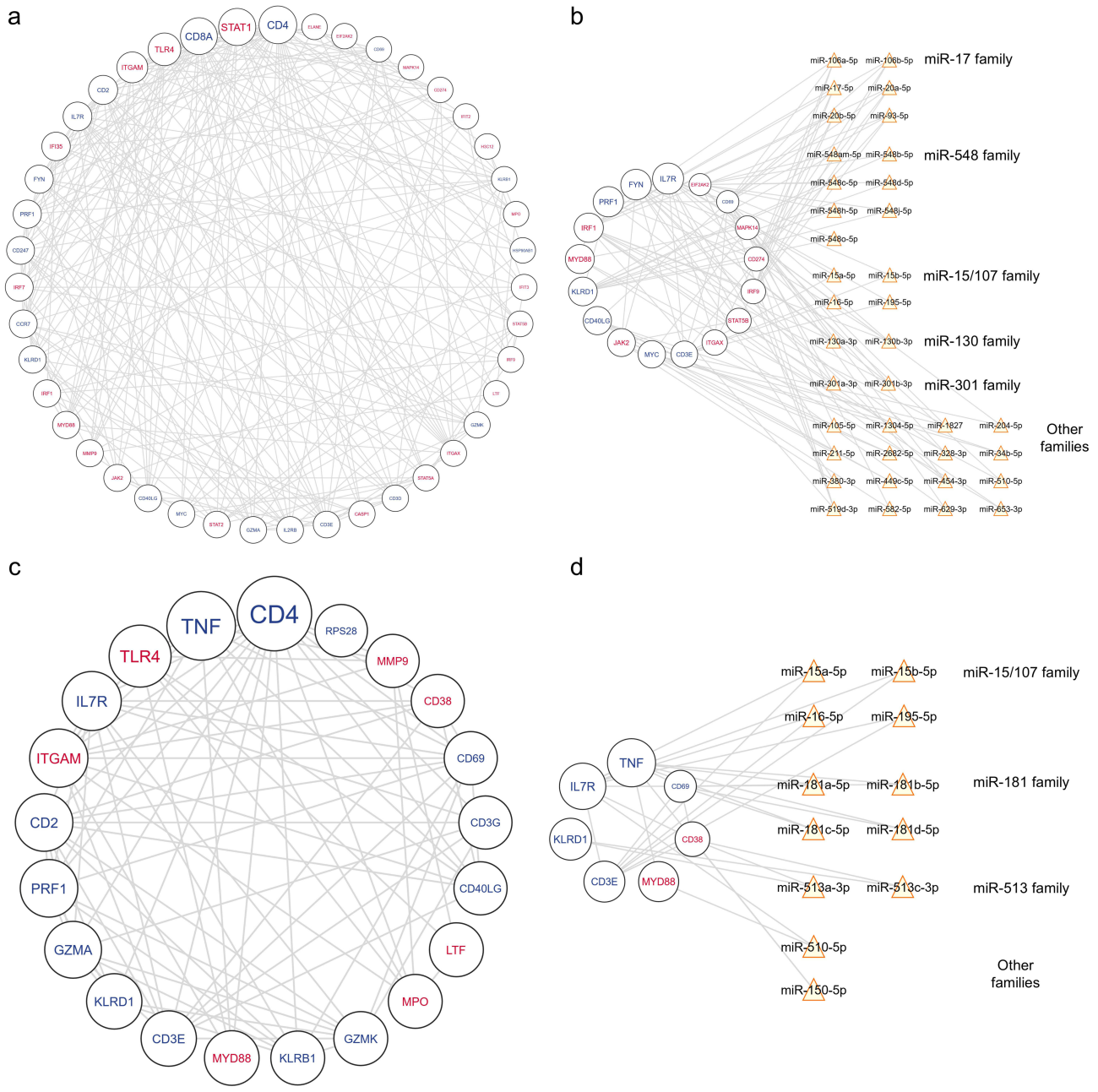


**Figure 3** Networks demonstrating PPIs and regulatory miRNAs of hub genes in CD and UC. (**a**) PPIs of hub genes in CD. (**b**) miRNA-mRNA regulatory network of hub genes in the PPI network of CD. (**c**) PPIs of hub genes in UC. (**d**) miRNA-mRNA regulatory network of hub genes in the PPI network of UC.

## Machine Learning-Based Biomarker Identification

The converging results of DE analysis and WGCNA were a total of 542 genes, including 426 upregulated and 116 downregulated genes, which were also detected in the selected positively correlated and negatively correlated gene modules, respectively (Figure 6a). After comparing the most important 20 genes by LASSO and RF, five genes, namely *INPPL1, TLR5, SLC9A8, IMPDH1*, and *GRK6*, were selected for further evaluation of their diagnostic performance by ROC analysis (Figure 6b–h). The area under the ROC curve for *TLR5, IMPDH1, SLC9A8, INPPL1*, and *GRK6* for CD in the integrated data was computed as 0.8065, 0.7735, 0.7631, 0.7611, and 0.7447, respectively (Figure 6i).

A total of 346 genes, comprising 277 upregulated and 69 downregulated genes, were consistently discerned in modules that demonstrated positive and negative correlations with UC, respectively (Figure 7a). After determination of the top 20 genes by RF and LASSO (Figure 7b–g), four genes, namely *IL4R, ACAA1, NARF*, and *RRM2*, were


https://doi.org/10.2147/CEG.S584999

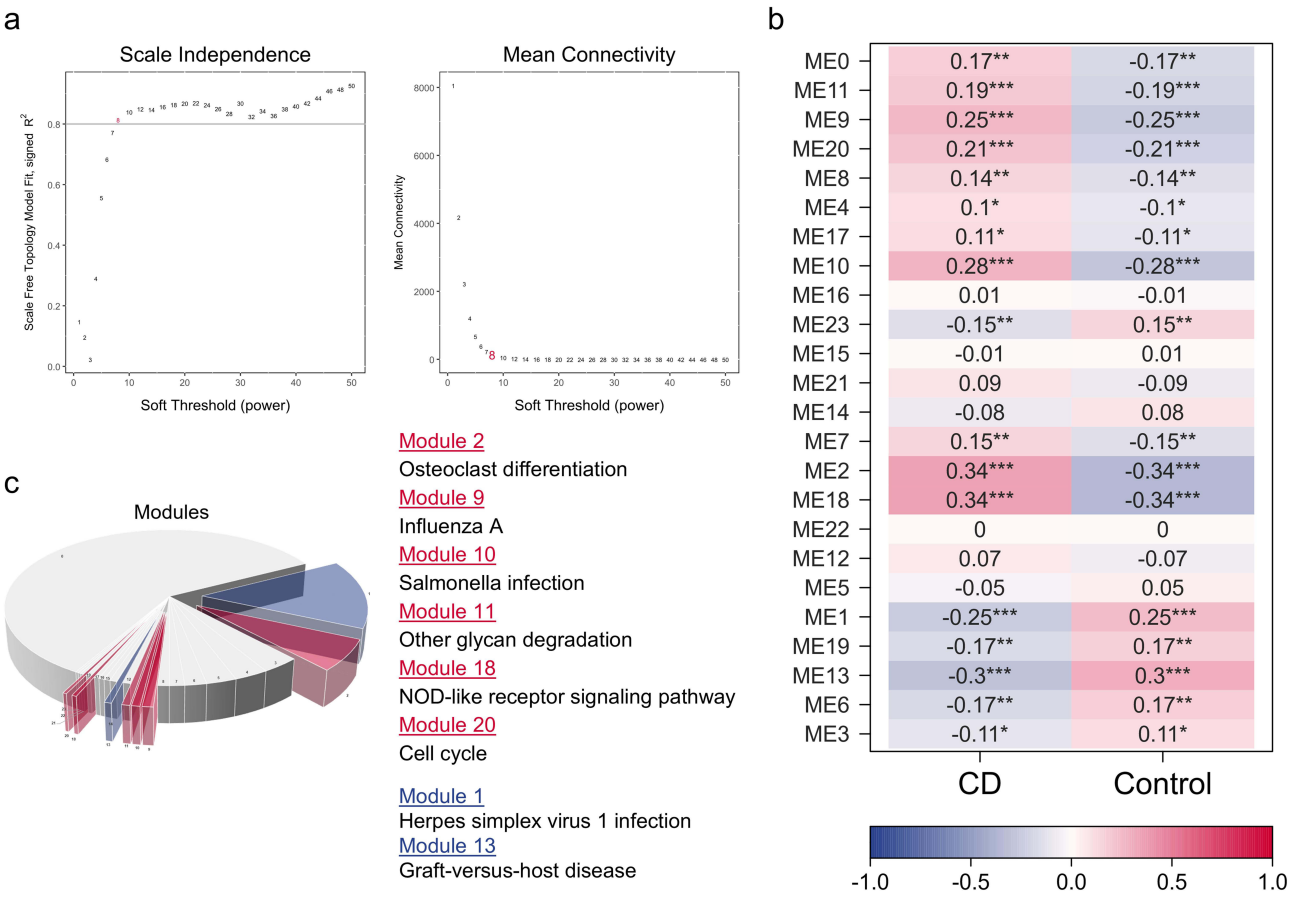


**Figure 4** The identification of prominent gene modules correlating with CD. (**a**) The selection of the optimal power based on scale Independence and mean connectivity. The red number shows the optimal soft thresholding power used for the construction of CD-specific modules. (**b**) Heatmap of module-trait relationship demonstrating modules and their correlation with CD. Asterisks indicate the statistical significance of the correlation (*p < 0.05, **p < 0.01, ***p < 0.001). (**c**) The proportion of selected positively and negatively correlated modules based on the number of genes. The KEGG pathway with the highest enrichment score in each module was noted based on the combined score provided in the Enrichr database. Red labels demonstrate modules positively correlated with CD, whereas blue labels show modules negatively correlated with CD.

highlighted as identified by both analyses (Figure 7h). ROC analysis evaluated the AUC of *IL4R, ACAA1, NARF*, and *RRM2* for UC as 0.7943, 0.7542, 0.7453, and 0.7213, respectively (Figure 7i).

## Evaluating the Discriminatory Power of Selected Genes in External Datasets

The AUC of *TLR5* for CD passed the 0.7 threshold in eight of the nine datasets, suggesting its promising potential as a biomarker for CD. Meanwhile, the diagnostic efficacy of *IMPDH1* and *SLC9A8* surpassed this threshold in six datasets. The diagnostic potential of *INPPL1* and *GRK6* had low reproducibility, passing the acceptable threshold (AUC > 0.7) in five and four datasets, respectively (Supplementary Figure S4).

The diagnostic performance of selected biomarkers for UC had moderate reproducibility across the analyzed datasets, with AUC values exceeding the 0.7 threshold in four to five datasets. Given that *TLR5* demonstrated promising diagnostic performance for CD, the ROC analysis was performed to evaluate its diagnostic potential for UC to assess whether it could differentiate between the two subtypes or not. Interestingly, the diagnostic efficacy of *TLR5* encompassed a higher AUC than *IL4R, ACAA1, NARF*, and *RRM2*, which were identified through the specific pipeline for UC (Supplementary Figure S5).

## Dysregulations of the Expression Levels of TLR5 in IBD Patients

ROC analysis, evaluating the diagnostic performance of *TLR5* for IBD, was then performed. Moreover, to ensure that *TLR5* exhibits a consistent direction of alteration across the analyzed datasets, DE analysis was separately conducted in

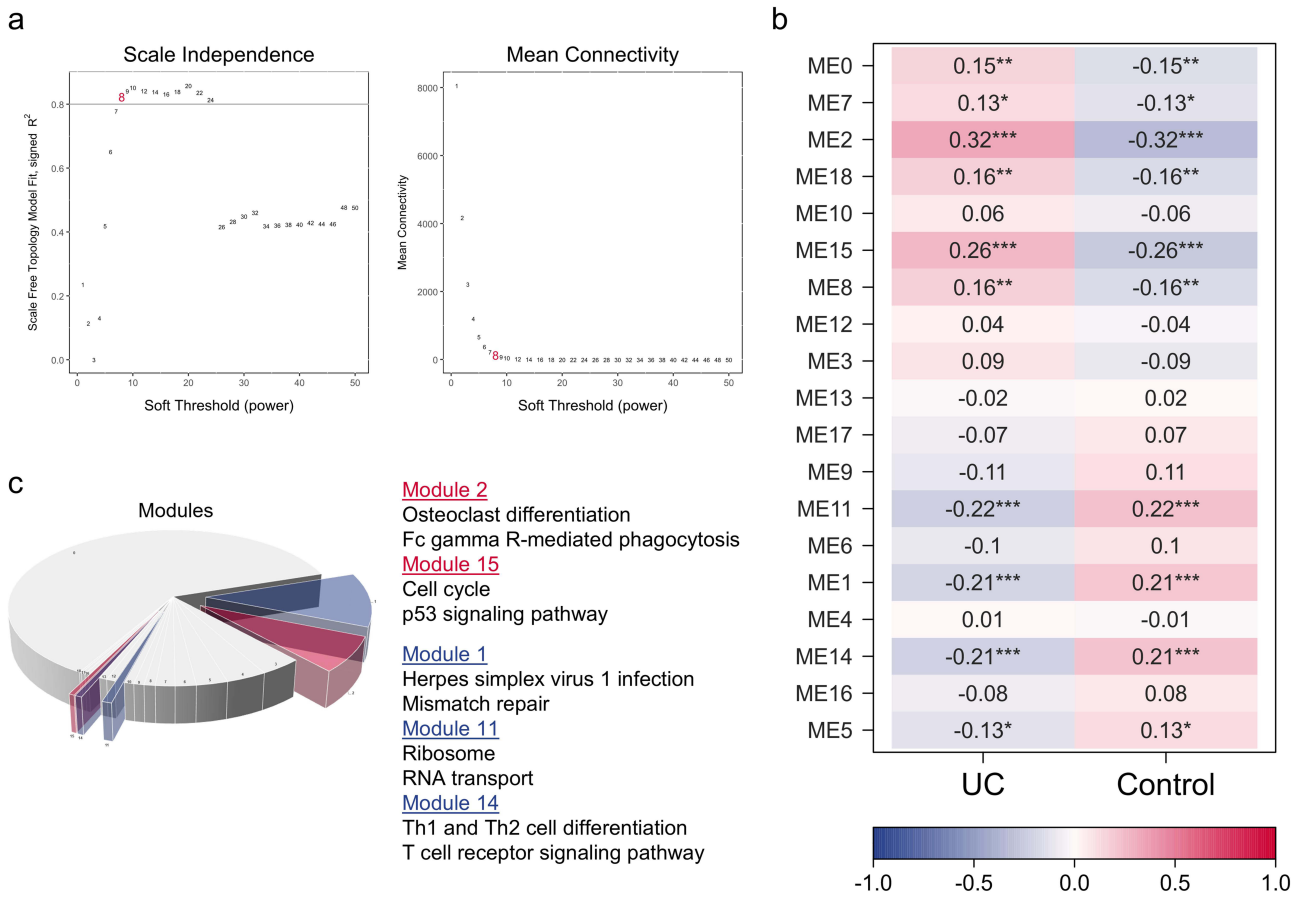


**Figure 5** The identification of prominent gene modules correlating with UC. (**a**) The selection of the optimal power based on scale Independence and mean connectivity. The red number shows the optimal soft thresholding power used for the construction of UC-specific modules. (**b**) Heatmap of module-trait relationship demonstrating modules and their correlation with UC. Asterisks demonstrate the statistical significance of the correlation (*p < 0.05, **p < 0.01, ***p < 0.001). (**c**) The proportion of selected positively and negatively correlated modules based on the number of genes. The KEGG pathway with the highest enrichment score in each module was reported based on the combined score provided in the Enrichr database. Red labels demonstrate modules positively correlated with UC, whereas blue labels show modules negatively correlated with UC.

each of the datasets. As shown in Figure 8, *TLR5* surpassed the AUC threshold of 0.7 in 9 out of the 12 datasets and exhibited a consistent trend of upregulation in the datasets, indicating its diagnostic potential for IBD.

## Discussion

Novel diagnostic and therapeutic strategies are essential to address the growing global burden of IBD. With the substantial value of blood transcriptomes and distinctions among the subtypes of IBD in mind, this study endeavored to identify specific biomarkers separately for CD and UC. Multiple computational analyses, including DE analysis, WGCNA, RF, and LASSO, were performed to select potential biomarkers for CD and UC, and their generalizability was evaluated by ROC analysis across the datasets. Meanwhile, no promising biomarkers were identified for UC, and *TLR5* as the most diagnostic marker for CD also demonstrated diagnostic potential for UC. Therefore, the diagnostic performance of *TLR5* was evaluated for the joint phenotype of CD and UC as IBD.

The GO analysis underlined several immune-related BP, CC, and MF terms shared between CD and UC. For instance, dysregulations of innate and adaptive immune responses in both subtypes of the IBD were reflected by the enrichment of BP terms, such as immune response-activating signaling pathways, regulation of innate immune response, positive regulation of cytokine production, and leukocyte-mediated immunity. Likewise, CC and MF terms were substantially similar and reflected different aspects of immune system responses. In accordance with the involvement of DEGs in the two subtypes of IBD in immune-related pathways, it has been previously shown that in patients with IBD, transcriptional alterations in the colon and peripheral blood are concordant, and immune pathways are enriched with the shared genes.[19]


https://doi.org/10.2147/CEG.S584999

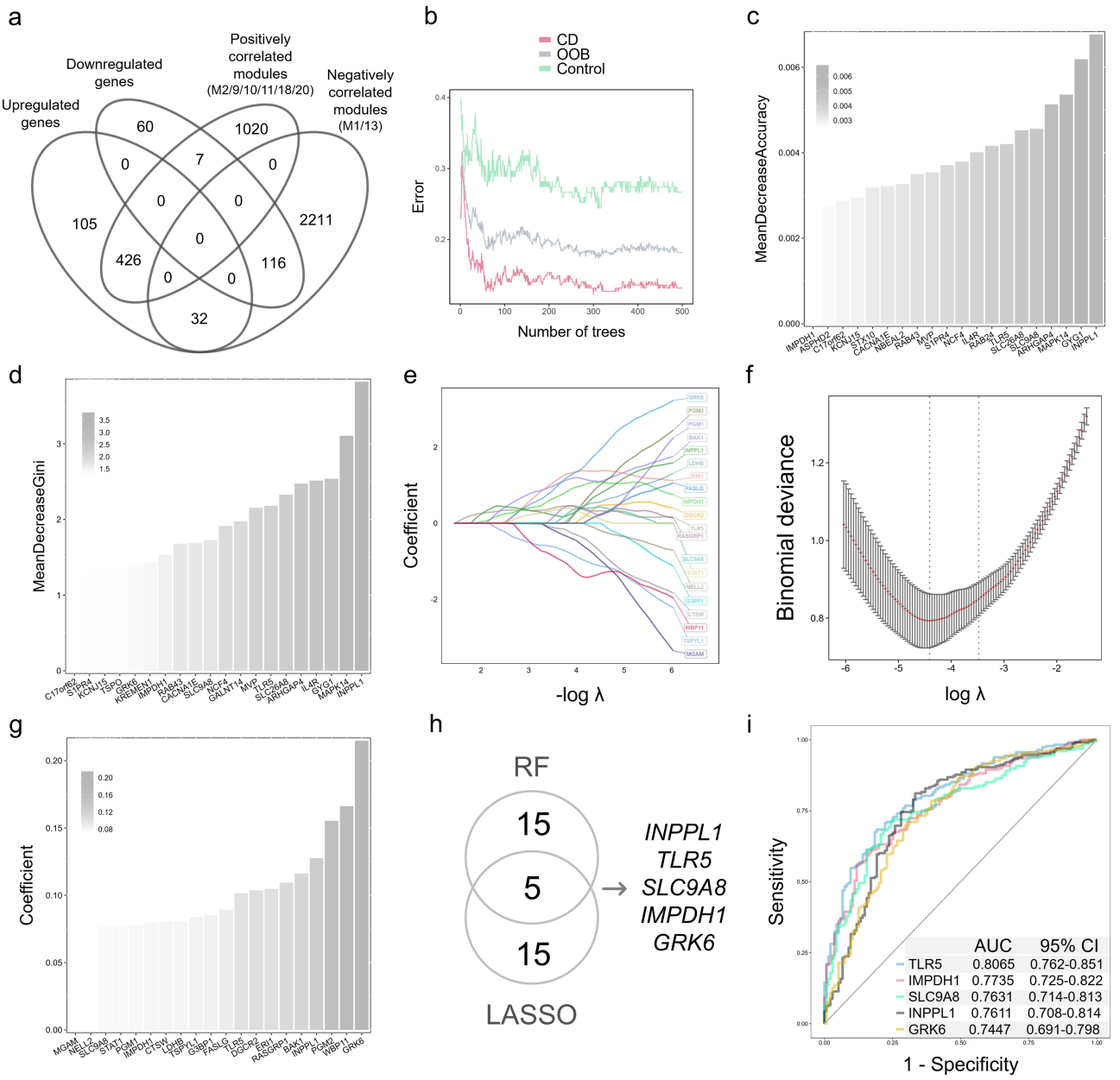


**Figure 6** Discovering potential biomarkers for CD using machine learning algorithms. (**a**) The intersection of DE analysis and WGCNA. (**b**) Error rate based on the number of decision trees in the RF method. (**c**) The importance of the top 20 genes based on the mean decrease in accuracy. (**d**) The importance of the top 20 genes based on the mean decrease in the Gini index. (**e**) LASSO coefficient trajectories for the top 20 features based on the magnitude of the absolute values of the coefficients. (**f**) Deviance plotted against the logarithm of the regularization parameter in LASSO regression. (**g**) The top 20 variables identified by LASSO. (**h**) Five overlapping genes were identified from the results of machine learning methods. (**i**) ROC analysis of the diagnostic performance of TLR5, IMPDH1, SLC9A8, INPPL1, and GRK6 for CD.

Circulating B cells exhibit increased activation in patients with CD[42] and UC.[43] Similarly, it has been reported that peripheral T cell subsets have an elevated functioning in patients with IBD.[44] Furthermore, a wide range of innate immune cells, including $CD177^{+}$ neutrophils,[45] monocytes,[46] basophils,[47] and eosinophils,[48] demonstrate dysregulation in IBD. NOD-like receptor signaling and hematopoietic cell lineage were the most prominent pathways in CD and UC, respectively. Nucleotide-binding and oligomerization domain NOD-like receptors (NLRs) are involved in the regulation of inflammation by multiple mechanisms of action,[49] including NF-κB, MAPK, and caspase-1.[50] The NLR signaling pathway has attracted growing interest as a therapeutic target for IBD.[50,51] NLRs[52] and their associated proteins[49] have been underlined as IBD susceptibility genes. NLRs are cytoplasmic receptors recognizing pathogens and damaged cells. These receptors are expressed in both innate and adaptive immune cells.[53] Hematopoietic cell lineage, the most

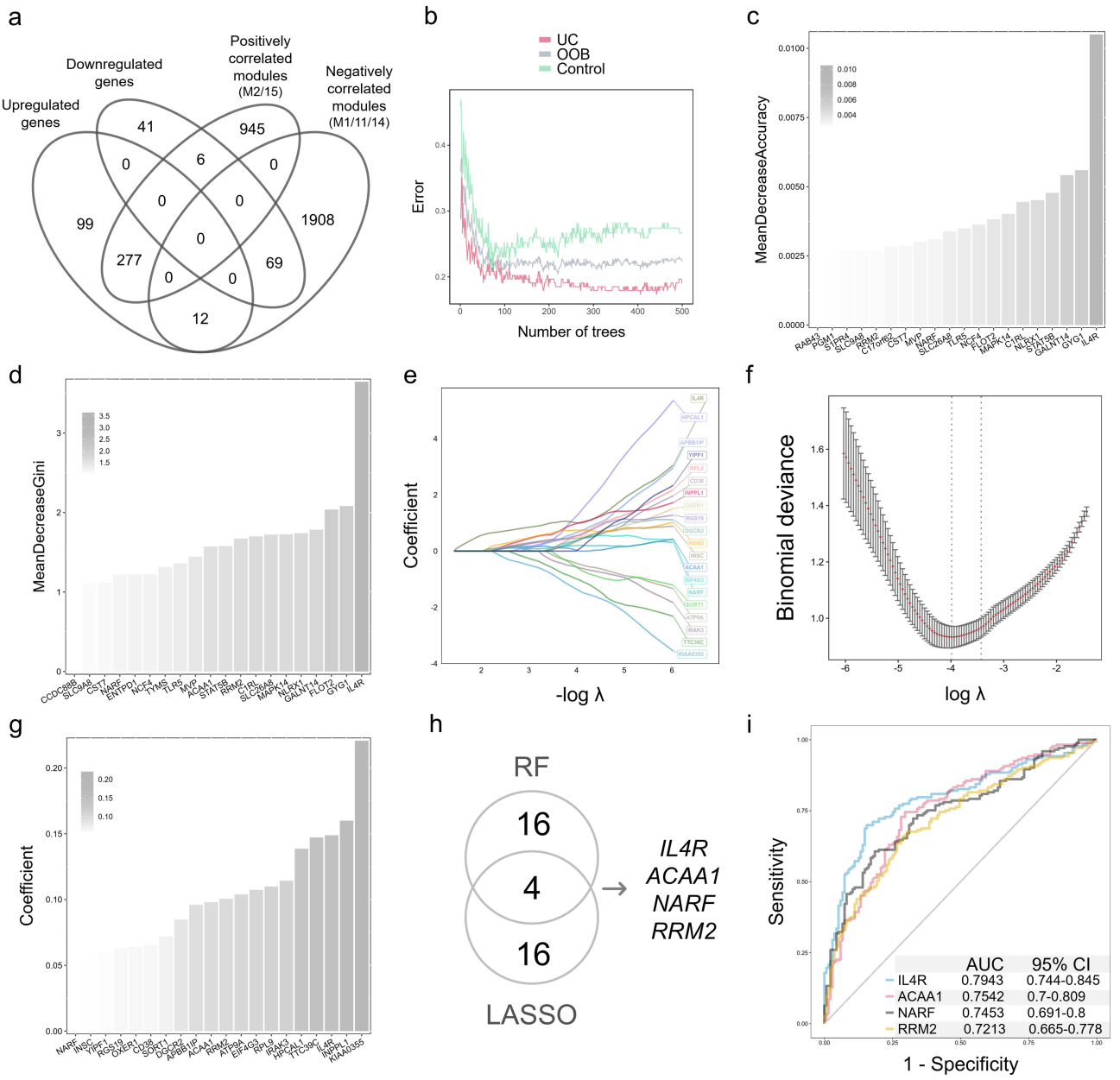


**Figure 7** Discovering potential biomarkers for UC using machine learning algorithms. (**a**) The intersection of DE analysis and WGCNA. (**b**) Error rate based on the number of decision trees in the RF method. (**c**) The importance of the top 20 genes based on the mean decrease in accuracy. (**d**) The importance of the top 20 genes based on the mean decrease in the Gini index. (**e**) LASSO coefficient trajectories for the top 20 features based on the magnitude of the absolute values of the coefficients. (**f**) Deviance plotted against the logarithm of the regularization parameter in LASSO regression. (**g**) The top 20 variables identified by LASSO. (**h**) Four overlapping genes were identified from the results of machine learning methods. (**i**) ROC analysis of the diagnostic performance of IL4R, ACAA1, NARF, and RRM2 for UC.

prominent pathway in which DEGs in UC and, to a lesser proportion, DEGs in CD were involved, reflects the differentiation of different blood cells, including immune cells, from hematopoietic stem cells (HSC). Collectively, functional enrichment analyses underscored a fundamental resemblance in the pathogenic pathways that are dysregulated in CD and UC.

Mapping PPIs of DEGs in CD and UC led to the identification of downregulated *CD4* as the central hub gene in both of the networks. Moreover, 16 other genes, including *TLR4, ITGAM, CD2, IL7R, PRF1, MYD88, KLRD1, MMP9, CD40LG, GZMA, CD3E, GZMK, LTF, MPO, KLRB1*, and *CD69*, were identified as hub genes in both of the PPI networks. Construction of miRNA-mRNA regulatory networks underscored four members of the miR-15/107 family, including miR-15a-5p, miR-15b-5p, miR-16-5p, and miR-195-5p,[54] as regulators of the hub genes in both CD and UC.


https://doi.org/10.2147/CEG.S584999

https://doi.org/10.2147/CEG.S584999





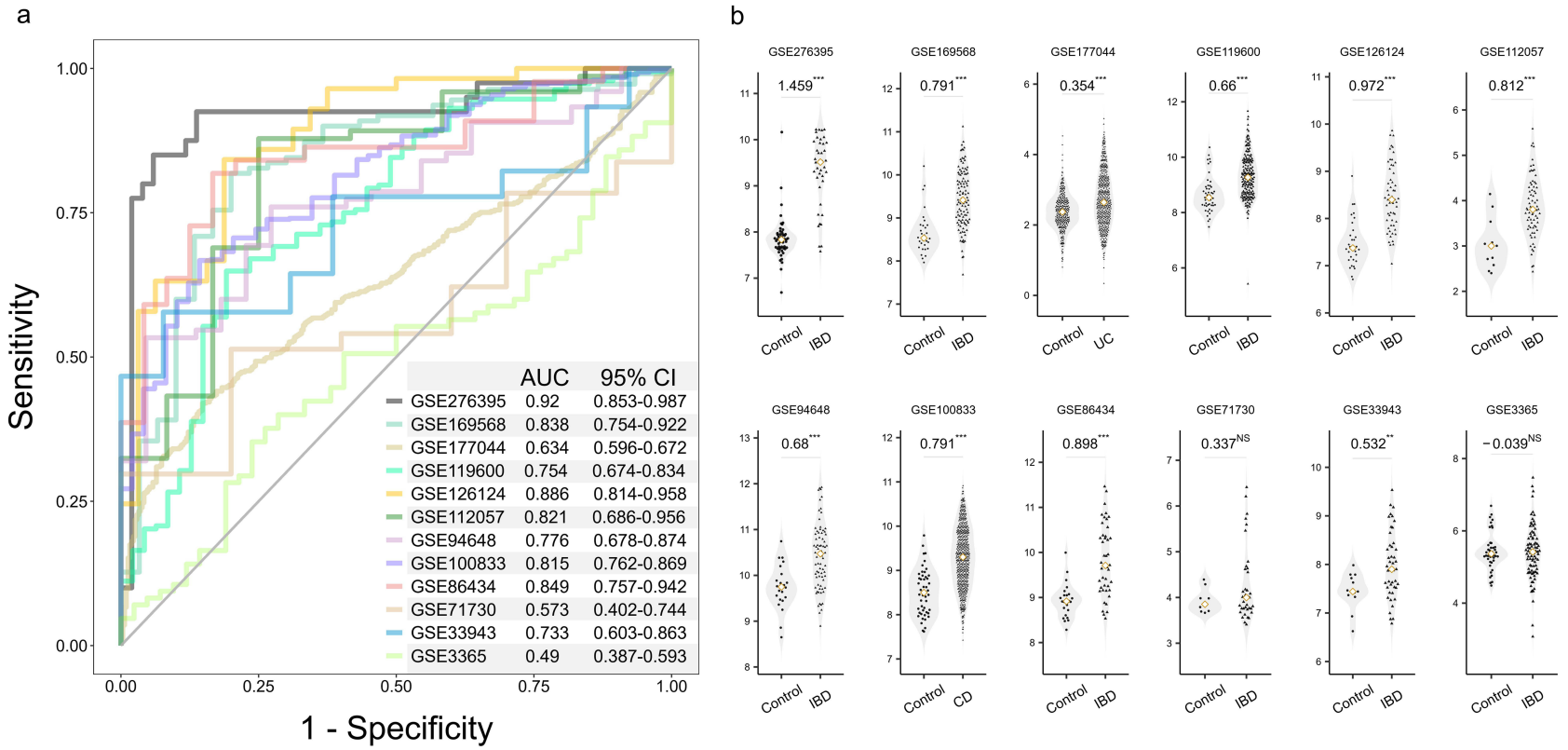


**Figure 8** The expression level of TLR5 in blood efficiently distinguished IBD from controls. (**a**) Diagnostic performance of TLR5 for IBD. (**b**) The expression alterations of TLR5 in IBD patients compared to controls. The median is indicated by a square marker with a yellow border in each study group. (NS: Not significant, **p < 0.01, ***p < 0.001).

In pediatric patients with IBD, miR-15a-5p was reported to exhibit upregulation in patients with active CD compared to the patients in the remission stage. Moreover, the level of miR-15a-5p was found to be positively correlated with the severity of the disease, and it was proposed as a potential biomarker. The study highlighted that Cdc42, an essential protein involved in the differentiation of intestinal epithelial stem cells (IESC) and barrier function of the intestine, is negatively regulated by miR-15a-5p.[55] Examination of fecal specimens revealed that the levels of miR-15a-5p demonstrate more than a three-fold increase in CD patients compared to HCs. Furthermore, miR-15a-5p demonstrated a positive correlation with the CD endoscopic index of severity (CDEIS).[56] In addition, this miRNA has also been identified as a key regulator of shared DEGs between IBD and colorectal cancer (CRC).[57] A previous study explored plasma miRNA signatures capable of predicting postoperative recurrence (POR) in patients with CD. The study highlighted miR-15b-5p as a part of a multi-miRNA biomarker panel, showing upregulated expression both when measured before surgery and at recurrence one year post-surgery. By contributing to the modulation of apoptosis, oxidative stress,[58] and autophagy,[59] this miRNA is involved in the pathogenesis of the disease.

By performing both microarray screening and qPCR measurements, elevated levels of miR-16-5p in fecal samples have been observed in CD and UC patients compared to HCs.[60] Examination of different sample sources, including serum, tissue, and feces, has consistently documented upregulation of miR-16-5p in UC patients.[61] Being capable of sponging miR-16-5p and miR-195-5p, the long non-coding RNA (LncRNA) CDKN2B-AS1 attenuates inflammation in UC.[62] A study on the colonic tissue of patients with UC underlined downregulation of the adenosine A2a receptor (A2aAR), which is an inhibitor of the NF-κB signaling pathway, and its expression is post-transcriptionally inhibited by miR-16 in UC patients. On the other hand, the upregulation of miR-16 was observed. By inhibiting A2aAR, miR-16 leads to the activation of the NF-κB signaling cascade and subsequent inflammation, which could be disrupted in vitro using miR-16 inhibitors.[63] By modulating the expression of tight junctions (TJs), it has been observed that miR-195-5p alleviates intestinal permeability,[64,65] a well-established hallmark of patients with IBD.[66] By disrupting the inflammatory signaling cascade of TLR2, miR-195-5p reduces polarization of M1 macrophages.[67] A study on the therapeutic effects of luteolin (Lut) on dextran sulfate sodium (DSS)-induced mouse models of UC revealed that alleviation in disease symptoms was associated with the suppression of the Notch signaling pathway, which was mediated by miR-195-5p as a negative feedback regulator of the pathway.[68] Collectively, these records underscore the key contribution of the four members of the miR-15/107 family to the pathogenesis of IBD. While these miRNAs can be considered for establishing miRNA-based therapies in IBD, subsequent experiments are necessary to shed light on the potential side effects and applicability of this approach.

Employing machine-learning methods subsequent to the filtration of DEGs based on the WGCNA results initially nominated *TLR5, IMPDH1, SLC9A8, INPPL1*, and *GRK6* as potential biomarkers for CD. Meanwhile, the approach underlined *IL4R, ACAA1, NARF*, and *RRM2* as potential biomarkers for UC. Evaluations of the diagnostic power of these genes in external datasets highlighted *TLR5* as a promising biomarker for CD. Furthermore, *TLR5* exhibited stronger diagnostic potential for UC than the genes identified via the UC-specific workflow. The UC-specific analyses failed to single out any promising biomarker for UC or IBD. Meanwhile, it is of paramount importance to note that the workflow for UC was not utterly inconclusive. For instance, it accentuated *IL4R*. This transcript has recently been introduced as a member of a diagnostic panel (*IL4R, EIF5A*, and *SLC9A8*), which was shown to be efficient in the diagnosis of IBD.[69] The unsuccessful analyses can also be attributed to technical issues such as dataset heterogeneity or biological similarities of CD and UC in blood samples.

Although dysregulations and involvement of *TLR5* have long been noticed in CD and UC, the results of the present study for the first time highlighted its diagnostic potential for IBD. *TLR5* encodes Toll-like receptor 5, a highly conserved member of the Toll-like receptor (TLR) family that specifically recognizes bacterial flagellin. Upon binding to its ligand, this receptor orchestrates an innate immune response via the MyD88 adaptor and subsequently activates the NF-kB signaling cascade.[70] It has been previously observed that knockout of *TLR5* in mice is associated with the increased risk of the onset of spontaneous colitis.[71] A single nucleotide polymorphism (SNP) reflecting elevated activity of *TLR5* (rs5744174) has been demonstrated to correlate with resistance to anti-TNF treatment in CD patients. Meanwhile, this association did not withstand Bonferroni correction for multiple testing.[72] In the North Indian population, it has been documented that two variants of *TLR5* (R392X and N592S) exhibit a statistically significant association with UC.[73]

*TLR5* substantially contributes to the maintenance of intestinal homeostasis between the microbiota and the immune system.[74] The downregulation of flagellin in gut microbes, along with the asymmetric distribution and basolateral localization of *TLR5*, as well as IgA-mediated neutralization of bacterial flagella, are known mechanisms that sustain homeostasis under normal conditions.[75] A study on the DSS-induced mouse model of colitis revealed that the alleviation of symptoms following treatment with Emodin, a naturally occurring compound, was associated with reduced expression of *TLR5*, as revealed by both in vivo and in vitro experiments.[76] Similarly, another study on TNBS-induced rat models of colitis underscored the therapeutic potential of shenling baizhu powder (SBP) treatment. The SBP treatment was associated with the suppression of the TLR5/MyD88/NF-κB pathway.[77] Moreover, a study on the SAMP1/YitFc (SAMP) mouse model of CD-like ileitis revealed that the elevated expression of *TLR5* in the epithelium of the ileum is implicated in the impaired intestinal barrier integrity, as reflected by the elevated permeability and attenuated resistance, as well as dysregulated expression of TJ proteins.[78] Collectively, these reports suggest the inhibition of the flagellin-TLR5 pathway as a potential therapeutic strategy for the disease.[79] Meanwhile, it is essential to conduct further experimental studies assessing the applicability of inhibiting the *TLR5* signaling pathway in light of studies showing reduced *TLR5* expression in intestinal mucus of patients with IBD, especially UC.[75,80] In accordance with these studies, observed downregulation of *TLR5* suggests its activation as a therapeutic approach. A study of the therapeutic potential of *Roseburia intestinalis* revealed that the activation of *TLR5* signaling induces the expression of thymic stromal lymphopoietin (TSLP). The expression of TSLP from IECs induces the secretion of anti-inflammatory cytokines, such as IL-10 and TGF-β, by dendritic cells. These cytokines are crucial for the differentiation of regulatory T cells (Treg), which are responsible for the suppression of overamplified inflammation, which results in the alleviation of colitis.[81] While the results of this study evidenced consistent upregulation of *TLR5* in patients with IBD in multiple cohorts, considering it as a therapeutic target requires subsequent experimental assays, particularly with respect to its involvement in innate immune and potential unwanted consequences of its inhibition.

In a previous study, by integrating four transcriptomic datasets, the investigation aimed to identify blood-based biomarkers for IBD.[82] While there are methodological similarities between this study and the previous one, including the implementation of DE analysis, WGCNA, and feature selection using RF and LASSO, they are distinct in their objectives and results. Firstly, the previous study integrated four transcriptomic datasets (GSE71730, GSE86434, GSE94648, and GSE119600),[82] whereas the present study merged five transcriptomic datasets (GSE71730, GSE86434, GSE94648, GSE119600, and GSE126124) to form the source data for biomarker discovery. By expanding the data source by adding GSE126124, the present study benefits from a larger sample size for biomarker discovery. Secondly, the previous study aimed to identify IBD biomarkers and therefore considered CD and UC collectively as IBD,[82] whereas the present study was designed to capture subtype-specific biomarkers, irrespective of the outcome. Lastly, while the previous study accentuated the diagnostic value of downregulated *FEZ1* for IBD,[82] the present study identifies upregulated *TLR5* as demonstrating promising diagnostic potential.

While this study underscored upregulated *TLR5* as a potential blood-based biomarker for IBD, it did not reveal CD-specific and UC-specific biomarkers, the prime research objective of this investigation. It is plausible to hypothesize that prospective transcriptomic analysis focusing on subtype-specific DEGs and analyzing them by ML methods may be able to underscore potential subtype-specific diagnostic biomarkers. At multiple stages, this study underscored the resemblance of blood transcriptomic alterations between CD and UC. Therefore, intestinal transcriptomics may better reflect the distinctions of these two subtypes. Moreover, multi-gene diagnostic panels may more efficiently stratify subtypes than single-gene approaches.

This study encountered multiple limitations. Firstly, single-omics-based studies do not comprehensively represent the nature of alterations in disorders. For instance, transcriptomic alterations are post-transcriptionally regulated by epigenetic modifications, leading to a distinct spectrum of disturbance in the protein level. Secondly, the heterogeneity of datasets might have implicitly attenuated the power of computational methods to unveil efficient subtype-specific biomarkers. Thirdly, the absence of functional validation, in vitro, and in vivo experiments highlights the need for further assays elucidating the applicability of *TLR5* as a diagnostic marker for IBD. In addition, the reduction in the number of analyzed genes following the integration of independent datasets was a limitation associated with employing an integrated transcriptomic approach. This might have concealed genes demonstrating a promising diagnostic performance. Therefore, it is plausible to assume that future studies

benefiting from a larger number of genes than those included in this study may unveil CD-specific and UC-specific diagnostic transcripts. Moreover, miRNAs visualized as regulators of hub genes present a predictive regulation based on the miRDB, and the lack of experimental validation necessitates interpretation of findings in light of this fact. Furthermore, differences in clinicodemographic characteristics, such as age, gender, smoking history, and comorbidities, between cases and controls might have implicitly affected the subtype-biomarker discovery. Lastly, given the involvement of Toll-like receptors (TLRs) in innate immunity, upregulation of *TLR5* may reflect the activation of innate immunity in the periphery rather than tracking IBD-specific pathogenic signatures, and it is crucial to evaluate the specificity of this transcript for IBD. Nevertheless, the reproducibility of its efficient diagnostic power suggests *TLR5* as a candidate for participating in the development of multi-gene diagnostic panels.

## Conclusion

There is an unmet demand to identify reliable noninvasive diagnostic markers for CD and UC, given the increasing global burden of these inflammatory bowel disorders. While the initial aim of this study was to identify CD-specific and UC-specific diagnostic markers, it eventually led to the discovery of the diagnostic power of *TLR5* for both subtypes of the disease and failed to discern subtype-specific biomarkers. At multiple stages, the results reflected the similarity of blood transcriptomic alterations between CD and UC. For instance, functional enrichment analysis of DEGs highlighted a highly similar spectrum of pathways in subtypes; *CD4* was identified as the central hub gene, and miR-15a-5p, miR-15b-5p, miR-16-5p, and miR-195-5p were predicted as regulators in networks. More importantly, transcripts identified through subtype-specific analyses were nonspecific for subtypes. The limitations of this study, including dataset heterogeneity, lack of experimental validations, and dependence on a single-omics approach, underscore the need for further studies illuminating the clinical applicability of findings.

## Abbreviations

IBD, Inflammatory bowel disease; CD, Crohn's disease; UC, Ulcerative colitis; GEO, Gene Expression Omnibus; WGCNA, Weighted gene co-expression network analysis; RF, Random forest; LASSO, Least absolute shrinkage and selection operator; MLP-ANN, Multi-layer perceptron artificial neural network model; PLS-DA, Partial least squares discriminant analysis; CTE, Computed tomography enterography; ROC, Receiver operating characteristic; AUC, Area under the curve; DE analysis, Differential expression analysis; DEGs, Differentially expressed genes; LogFC, $Log_2$ fold change; GO, Gene Ontology; BP, Biological processes; CC, Cellular components; MF, Molecular functions; KEGG, Kyoto Encyclopedia of Genes and Genomes; PCA, Principal component analysis; PPIs, Protein-protein interactions; PBMCs, Peripheral blood mononuclear cells; PBLs, Peripheral blood leukocytes; qPCR, Quantitative polymerase chain reaction.

## Data Sharing Statement

The datasets analyzed during the current study are available in the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) with the accession numbers GSE276395, GSE169568, GSE177044, GSE119600, GSE126124, GSE112057, GSE94648, GSE100833, GSE86434, GSE71730, GSE33943, and GSE3365.

## Ethics Approval and Consent to Participate

This study qualifies for ethics approval exemption, as it used publicly available datasets from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) that present transcriptomic data for re-analysis research purposes. No ethical approval was required by the local ethics committees. Therefore, this research does not require ethics approval.

## Acknowledgments

The author appreciates the contributions made by the authors of the GSE276395, GSE169568, GSE177044, GSE119600, GSE126124, GSE112057, GSE94648, GSE100833, GSE86434, GSE71730, GSE33943, and GSE3365.

## Funding

The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

## Disclosure

The author declares no competing interests.

## References


1. Cao L, Dayimu A, Guan X, et al. Global evolving patterns and cross-country inequalities of inflammatory bowel disease burden from 1990 to 2019: a worldwide report. *Inflamm Res*. 2024;73(2):277–287. doi:10.1007/s00011-023-01836-7
2. Galipeau HJ, Caminero A, Turpin W, et al. Novel fecal biomarkers that precede clinical diagnosis of ulcerative colitis. *Gastroenterology*. 2021;160 (5):1532–1545. doi:10.1053/j.gastro.2020.12.004
3. Clough J, Colwill M, Poullis A, Pollok R, Patel K, Honap S. Biomarkers in inflammatory bowel disease: a practical guide. *Therap Adv Gastroenterol*. 2024;17:17562848241251600. doi:10.1177/17562848241251600
4. Santacroce G, Zammarchi I, Tan CK, et al. Present and future of endoscopy precision for inflammatory bowel disease. *Digestive Endoscopy*. 2024;36(3):292–304. doi:10.1111/den.14672
5. Li L, Cheng R, Wu Y, Lin H, Gan H, Zhang H. Diagnosis and management of inflammatory bowel disease. *J Evid Based Med*. 2024;17 (2):409–433. doi:10.1111/jebm.12626
6. Li X, Lu C, Yang Y, Yu C, Rao Y. Site-specific targeted drug delivery systems for the treatment of inflammatory bowel disease. *Biomed Pharmacother*. 2020;129:110486. doi:10.1016/j.biopha.2020.110486
7. James JP, Nielsen BS, Christensen IJ, et al. Mucosal expression of PI3, ANXA1, and VDR discriminates Crohn's disease from ulcerative colitis. *Sci Rep*. 2023;13(1):18421. doi:10.1038/s41598-023-45569-3
8. Pei J, Wang G, Li Y, et al. Utility of four machine learning approaches for identifying ulcerative colitis and Crohn's disease. *Heliyon*. 2024;10(1): e23439. doi:10.1016/j.heliyon.2023.e23439
9. Zhou G, Liu H, Wei P, et al. Amino acids-targeted metabolomics reveals novel diagnostic biomarkers for ulcerative colitis and Crohn's disease. *Amino Acids*. 2023;55(3):349–358. doi:10.1007/s00726-023-03233-0
10. Park SK, Kim S, Lee GY, et al. Development of a machine learning model to distinguish between ulcerative colitis and Crohn's disease using RNA sequencing data. *Diagnostics*. 2021;11. doi:10.3390/diagnostics11122365
11. Huang M, Tu L, Li J, et al. Differentiation of Crohn's disease, ulcerative colitis, and intestinal tuberculosis by dual-layer spectral detector CT enterography. *Clin Radiol*. 2024;79(3):e482–e489. doi:10.1016/j.crad.2023.12.001
12. Mishra N, Aden K, Blase JI, et al. Longitudinal multi-omics analysis identifies early blood-based predictors of anti-TNF therapy response in inflammatory bowel disease. *Genome Med*. 2022;14(1):110. doi:10.1186/s13073-022-01112-z
13. Moran CP, Neary B, Doherty GA. Endoscopic evaluation in diagnosis and management of inflammatory bowel disease. *World J Gastrointest Endoscopy*. 2016;8(20):723–732. doi:10.4253/wjge.v8.i20.723
14. Shen B, Abreu MT, Cohen ER, et al. Endoscopic diagnosis and management of adult inflammatory bowel disease: a consensus document from the American society for gastrointestinal endoscopy IBD endoscopy consensus panel. *Gastrointest Endosc*. 2025;101(2):295–314. doi:10.1016/j.gie.2024.08.034
15. Grosgeorges M, Picque Lasorsa L, Pastor B, et al. A straightforward method to quantify circulating mRNAs as biomarkers of colorectal cancer. *Sci Rep*. 2023;13(1):2739. doi:10.1038/s41598-023-29948-4
16. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2012;41(D1):D991–5. doi:10.1093/nar/gks1193
17. Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and bioconductor. *Bioinformatics*. 2007;23(14):1846–1847. doi:10.1093/bioinformatics/btm254
18. Ostrowski J, Goryca K, Lazowska I, et al. Common functional alterations identified in blood transcriptome of autoimmune cholestatic liver and inflammatory bowel diseases. *Sci Rep*. 2019;9(1):7190. doi:10.1038/s41598-019-43699-1
19. Palmer NP, Silvester JA, Lee JJ, et al. Concordance between gene expression in peripheral whole blood and colonic tissue in children with inflammatory bowel disease. *PLoS One*. 2019;14(10):e0222952. doi:10.1371/journal.pone.0222952
20. Planell N, Masamunt MC, Leal RF, et al. Usefulness of transcriptional blood biomarkers as a non-invasive surrogate marker of mucosal healing and endoscopic response in ulcerative colitis. *J Crohns Colitis*. 2017;11(11):1335–1346. doi:10.1093/ecco-jcc/jjx091
21. Ventham NT, Kennedy NA, Adams AT, et al. Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nat Commun*. 2016;7(1):13507. doi:10.1038/ncomms13507
22. Gurram B, Salzman NH, Kaldunski ML, et al. Plasma-induced signatures reveal an extracellular milieu possessing an immunoregulatory bias in treatment-naive paediatric inflammatory bowel disease. *Clin Exp Immunol*. 2016;184(1):36–49. doi:10.1111/cei.12753
23. Tsakiroglou M, Evans A, Doce-Carracedo A, et al. Gene expression dysregulation in whole blood of patients with clostridioides difficile infection. *Int J Mole Sci*. 2024;25(23):12653. doi:10.3390/ijms252312653
24. Juzenas S, Hübenthal M, Lindqvist CM, et al. Detailed transcriptional landscape of peripheral blood points to increased neutrophil activation in treatment-naïve inflammatory bowel disease. *J Crohns Colitis*. 2022;16(7):1097–1109. doi:10.1093/ecco-jcc/jjac003
25. Wacker EM, Uellendahl-Werth F, Bej S, et al. Whole blood RNA sequencing identifies transcriptional differences between primary sclerosing cholangitis and ulcerative colitis. *JHEP Rep*. 2024;6(2):100988. doi:10.1016/j.jhepr.2023.100988
26. Mo A, Marigorta UM, Arafat D, et al. Disease-specific regulation of gene expression in a comparative analysis of juvenile idiopathic arthritis and inflammatory bowel disease. *Genome Med*. 2018;10(1). doi:10.1186/s13073-018-0558-x
27. Peters LA, Perrigoue J, Mortha A, et al. A functional genomics predictive network model identifies regulators of inflammatory bowel disease. *Nat Genet*. 2017;49(10):1437–1449. doi:10.1038/ng.3947
28. van Lierop PPE, Swagemakers SM, de Bie CI, et al. Gene expression analysis of peripheral cells for subclassification of pediatric inflammatory bowel disease in remission. *PLoS One*. 2013;8(11):e79549. doi:10.1371/journal.pone.0079549
29. Burczynski ME, Peterson RL, Twine NC, et al. Molecular classification of crohn's disease and ulcerative colitis patients using transcriptional profiles in peripheral blood mononuclear cells. *J Mol Diagn*. 2006;8(1):51–61. doi:10.2353/jmoldx.2006.050079

30. Leek JT, Storey JD, Gibson G. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet*. 2007;3 (9):1724–1735. doi:10.1371/journal.pgen.0030161
31. Ligges U, Mächler M. scatterplot3d - an R package for visualizing multivariate data. *J Statist Software*. 2003;8(11):1–20. doi:10.18637/jss.v008.i11
32. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007
33. Wickham H. *Data Analysis, in ggplot2: Elegant Graphics for Data Analysis*. Cham: Springer International Publishing; 2016:189–201.
34. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16 (5):284–287. doi:10.1089/omi.2011.0118
35. Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. 2023;51(D1):D638–d646. doi:10.1093/nar/gkac1000
36. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–2504. doi:10.1101/gr.1239303
37. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res*. 2020;48(D1):D127–d131. doi:10.1093/nar/gkz757
38. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf*. 2008;9(1):559. doi:10.1186/1471-2105-9-559
39. Xie Z, Bailey A, Kuleshov MV, et al. Gene set knowledge discovery with enrichr. *Curr Protoc*. 2021;1(3):e90. doi:10.1002/cpz1.90
40. Liaw A, Wiener M. Classification and regression by random Forest. *R news*. 2002;2(3):18–22.
41. Friedman JH, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Statist Software*. 2010;33(1):1–22. doi:10.18637/jss.v033.i01
42. Noronha AM, Liang Y, Hetzel JT, et al. Hyperactivated B cells in human inflammatory bowel disease. *J Leukoc Biol*. 2009;86(4):1007–1016. doi:10.1189/jlb.0309203
43. Uzzan M, Martin JC, Mesin L, et al. Ulcerative colitis is characterized by a plasmablast-skewed humoral response associated with disease activity. *Nat Med*. 2022;28(4):766–779. doi:10.1038/s41591-022-01680-y
44. Funderburg NT, Stubblefield Park SR, Sung HC, et al. Circulating CD 4 + and CD 8 + T cells are activated in inflammatory bowel disease and are associated with plasma markers of inflammation. *Immunology*. 2013;140(1):87–97. doi:10.1111/imm.12114
45. Zhou G, Yu L, Fang L, et al. CD177 + neutrophils as functionally activated neutrophils negatively regulate IBD. *Gut*. 2018;67(6):1052–1063. doi:10.1136/gutjnl-2016-313535
46. Anderson A, Cherfane C, Click B, et al. Monocytosis is a biomarker of severity in inflammatory bowel disease: analysis of a 6-year prospective natural history registry. *Inflamm Bowel Dis*. 2022;28(1):70–78. doi:10.1093/ibd/izab031
47. Chapuy L, Bsat M, Mehta H, et al. Basophils increase in Crohn disease and ulcerative colitis and favor mesenteric lymph node memory TH17/TH1 response. *J Allergy Clin Immunol*. 2014;134(4):978–81.e1. doi:10.1016/j.jaci.2014.05.025
48. Yerushalmy-Feler A, Lujan R, Weisband YL, et al. Peripheral blood eosinophilia at diagnosis of inflammatory bowel disease is associated with severe disease course: a nationwide study from the epi-IIRN cohort. *J Crohns Colitis*. 2025;19(2). doi:10.1093/ecco-jcc/jjae130
49. Rubino SJ, Selvanantham T, Girardin SE, Philpott DJ. Nod-like receptors in the control of intestinal inflammation. *Curr Opin Immunol*. 2012;24 (4):398–404. doi:10.1016/j.coi.2012.04.010
50. Zhou F, Zhang GD, Tan Y, Hu SA, Tang Q, Pei G. NOD-like receptors mediate homeostatic intestinal epithelial barrier function: promising therapeutic targets for inflammatory bowel disease. *Therap Adv Gastroenterol*. 2023;16:17562848231176889. doi:10.1177/17562848231176889
51. Chen Q-L, Yin H-R, He Q-Y, Wang Y. Targeting the NLRP3 inflammasome as new therapeutic avenue for inflammatory bowel disease. *Biomed Pharmacother*. 2021;138:111442. doi:10.1016/j.biopha.2021.111442
52. Davis BK, Philipson C, Hontecillas R, Eden K, Bassaganya-Riera J, Allen IC. Emerging significance of NLRs in inflammatory bowel disease. *Inflamm Bowel Dis*. 2014;20(12):2412–2432. doi:10.1097/mib.0000000000000151
53. Madahar SS, Gideon A, Abdul-Sater AA. Nod-like receptors in inflammatory arthritis. *Biomed J*. 2024;47(1):100655. doi:10.1016/j.bj.2023.100655
54. Wang S, Zhu W, Xu J, et al. Interpreting the MicroRNA-15/107 family: interaction identification by combining network based and experiment supported approach. *BMC Med Genet*. 2019;20(1):96. doi:10.1186/s12881-019-0824-9
55. Tang W-J, Peng K-Y, Tang Z-F, Wang Y-H, Xue A-J, Huang Y. MicroRNA-15a - cell division cycle 42 signaling pathway in pathogenesis of pediatric inflammatory bowel disease. *World J Gastroenterol*. 2018;24(46):5234–5245. doi:10.3748/wjg.v24.i46.5234
56. Wohnhaas CT, Schmid R, Rolser M, et al. Fecal MicroRNAs show promise as noninvasive Crohn's disease biomarkers. *Crohns Colitis*. 2020;2(1): otaa003. doi:10.1093/crocol/otaa003
57. Al-Mustanjid M, Mahmud SMH, Royel MRI, et al. Detection of molecular signatures and pathways shared in inflammatory bowel disease and colorectal cancer: a bioinformatics and systems biology approach. *Genomics*. 2020;112(5):3416–3426. doi:10.1016/j.ygeno.2020.06.001
58. Moret-Tatay I, Cerrillo E, Hervás D, et al. Specific plasma MicroRNA signatures in predicting and confirming Crohn's disease recurrence: role and pathogenic implications. *Clin Transl Gastroenterol*. 2021;12(10):e00416. doi:10.14309/ctg.0000000000000416
59. Chen H, Xu Z, Zhao H, et al. Global research states and trends of micro RNA in irritable bowel syndrome: a bibliometric analysis. *Clin Exp Med*. 2024;24(1):149. doi:10.1007/s10238-024-01396-y
60. Zhou R, Qiu P, Wang H, et al. Identification of microRNA-16-5p and microRNA-21-5p in feces as potential noninvasive biomarkers for inflammatory bowel disease. *Aging*. 2021;13(3):4634–4646. doi:10.18632/aging.202428
61. Chacon-Millan P, Lama S, Del Gaudio N, et al. A combination of microarray-based profiling and biocomputational analysis identified miR331-3p and hsa-let-7d-5p as potential biomarkers of ulcerative colitis progression to colorectal cancer. *Int J Mol Sci*. 2024;25:5699. doi:10.3390/ijms25115699
62. Tian Y, Cui L, Lin C, Wang Y, Liu Z, Miao X. LncRNA CDKN2B-AS1 relieved inflammation of ulcerative colitis via sponging miR-16 and miR-195. *Int Immunopharmacol*. 2020;88:106970. doi:10.1016/j.intimp.2020.106970
63. Tian T, Zhou Y, Feng X, et al. MicroRNA-16 is putatively involved in the NF-κB pathway regulation in ulcerative colitis through adenosine A2a receptor (A2aAR) mRNA targeting. *Sci Rep*. 2016;6(1):30824. doi:10.1038/srep30824
64. Scalavino V, Piccinno E, Bianco G, et al. The increase of miR-195-5p reduces intestinal permeability in ulcerative colitis, modulating tight junctions' expression. *Int J Mol Sci*. 2022;23:5840. doi:10.3390/ijms23105840

65. Scalavino V, Piccinno E, Lacalamita A, et al. miR-195-5p regulates tight junctions expression via claudin-2 downregulation in ulcerative colitis. *Biomedicines*. 2022;10:919. doi:10.3390/biomedicines10040919
66. Iacucci M, Santacroce G, Majumder S, et al. Opening the doors of precision medicine: novel tools to assess intestinal barrier in inflammatory bowel disease and colitis-associated neoplasia. *Gut*. 2024;73(10):1749–1762. doi:10.1136/gutjnl-2023-331579
67. Ramadan YN, Kamel AM, Medhat MA, Hetta HF. MicroRNA signatures in the pathogenesis and therapy of inflammatory bowel disease. *Clin Exp Med*. 2024;24(1):217. doi:10.1007/s10238-024-01476-z
68. Liu M, Wang Y, Liu Z, Liu S, Yang Q, Li B. Luteolin improves mitochondrial dynamics and function in ulcerative colitis via the miR-195-5p/Notch signalling pathway. *J Functional Foods*. 2025;124:106644. doi:10.1016/j.jff.2024.106644
69. Nazari MHD, Ghorbaninejad M, Shahrokh S, Meyfour A. Biomarker discovery for non-invasive diagnosis of inflammatory bowel disease using blood transcriptomics. *Front Immunol*. 2025;16:1570374. doi:10.3389/fimmu.2025.1570374
70. Yoon S-I, Kurnasov O, Natarajan V, et al. Structural basis of TLR5-flagellin recognition and signaling. *Science*. 2012;335(6070):859–864. doi:10.1126/science.1215584
71. Vijay-Kumar M, Sanders CJ, Taylor RT, et al. Deletion of TLR5 results in spontaneous colitis in mice. *J Clin Invest*. 2007;117(12):3909–3921. doi:10.1172/jci33084
72. Bank S, Andersen PS, Burisch J, et al. Genetically determined high activity of IL-12 and IL-18 in ulcerative colitis and TLR5 in Crohns disease were associated with non-response to anti-TNF therapy. *Pharmacogenomics J*. 2018;18(1):87–97. doi:10.1038/tpj.2016.84
73. Meena NK, Ahuja V, Meena K, Paul J, Speletas M. Association of TLR5 gene polymorphisms in ulcerative colitis patients of north India and their role in cytokine homeostasis. *PLoS One*. 2015;10(3):e0120697. doi:10.1371/journal.pone.0120697
74. Chen L, Zhang L, Hua H, Liu L, Mao Y, Wang R. Interactions between toll-like receptors signaling pathway and gut microbiota in host homeostasis. *Immun Inflamm Dis*. 2024;12(7):e1356. doi:10.1002/iid3.1356
75. Feng S, Zhang C, Chen S, He R, Chao G, Zhang S. TLR5 signaling in the regulation of intestinal mucosal immunity. *J Inflamm Res*. 2023;16:2491–2501. doi:10.2147/jir.S407521
76. Luo S, Deng X, Liu Q, et al. Emodin ameliorates ulcerative colitis by the flagellin-TLR5 dependent pathway in mice. *Int Immunopharmacol*. 2018;59:269–275. doi:10.1016/j.intimp.2018.04.010
77. Rao K, Qin S, Yang Y, et al. Shenling Baizhu powder alleviates TNBS-induced colitis in rats by improving intestinal epithelial permeability and inhibiting inflammation through the TLR5/MyD88/NF-κB pathway. *Front Pharmacol*. 2022;13:883918. doi:10.3389/fphar.2022.883918
78. Lopetuso LR, Jia R, Wang X-M, et al. Epithelial-specific toll-like receptor (TLR)5 activation mediates barrier dysfunction in experimental ileitis. *Inflamm Bowel Dis*. 2017;23(3):392–403. doi:10.1097/mib.0000000000001035
79. Lu Y, Li X, Liu S, Zhang Y, Zhang D. Toll-like receptors and inflammatory bowel disease. *Front Immunol*. 2018;9:72. doi:10.3389/fimmu.2018.00072
80. Kordjazy N, Haj-Mirzaian A, Haj-Mirzaian A, et al. Role of toll-like receptors in inflammatory bowel disease. *Pharmacol Res*. 2018;129:204–215. doi:10.1016/j.phrs.2017.11.017
81. Shen Z, Luo W, Tan B, et al. Roseburia intestinalis stimulates TLR5-dependent intestinal immunity against Crohn's disease. *EBioMedicine*. 2022;85:104285. doi:10.1016/j.ebiom.2022.104285
82. Delkhah AMD. Blood transcriptomics identifies FEZ1 as a potential biomarker for inflammatory bowel disease. *Comput Biol Med*. 2025;187:109742. doi:10.1016/j.compbiomed.2025.109742