



OPEN

Identification of immune-associated genes and single-cell sequencing analysis in diagnosing Behçet's disease

Qiyi Yu^{1,5}✉, Yidong Zhang^{2,5}, Yifan Wu^{3,5}, Meinuo Li⁴ & Tianyue Pan³

Behçet's Disease (BD) is diagnosed as continuing systematic inflammation, lacking clarified etiology and pertinent detection of clinical usage. We screened bulk-seq and single-cell seq data of BD from GEO database, then investigated immune infiltration landscape in BD and further explored main immune cells in a single-cell resolution. It was inferred that monocytes might be the trigger cell of BD occurrence and development. On the other hand, we performed differentially expression analysis and weighted correlation network analysis, and intersected the selected genes with immune-related genes from ImmPort database. After KEGG/GO-enrichment, Protein-Protein Interaction network and miRNA-mRNA regulatory network were constructed. Meanwhile, we utilized the Least Absolute Shrinkage and Selection Operator to filtrate 5 significant intersected genes for BD's diagnosis (NRTN, TRDJ1, IGLV4-69, PDIA2 and AVPR1A), then drew Operating Characteristic curve to analyze diagnosis value. For experimental validation, the BD mice models were constructed and qRT-PCR experiment was performed to validate the expression level of selected genes. In conclusion, we investigated the immune landscape in BD and explored the role of monocyte in BD occurrence and development. Through gene selection and machine learning, five new potential diagnostic biomarkers were discovered and validated in both external dataset and experiment assay levels.

Keywords Biomarker, Immunity, Single-cell sequencing analysis, Behçet's disease

Behçet's Disease (BD), diagnosed as a systemic vasculitis with recurrent oral and genital ulcers, skin lesions, and uveitis, has been in prevalence in Asia and the Middle East^{1,2}. This multisystem disease can affect various organs, including the eyes, skin, joints, gastrointestinal tract, and central nervous system³. While relevant researches have been conducted widely in BD field, the etiology remains to be clarified due to the sophisticated mechanism⁴. Apart from diagnosing via typical symptoms of BD, clinical experience also plays a significant role⁵.

Inflammation occurs in blood vessel walls leads to vascular damage and the development of vasculitis⁶. Typical pathological changes include collagen fiber degeneration as well as endothelial and smooth muscle cell necrosis⁷. When vascular inflammatory lesions invade various types of vessels, the condition becomes relatively serious, which is called BD vasculitis. During this progress, immune cells, especially monocytes, play a critical role in orchestrating the inflammatory response⁸. Monocytes are rapidly recruited to sites of vascular injury where they adhere to activated endothelial cells and transigrate into the vessel wall after the inflammation occurs⁹. Upon infiltrating into vessels, these cells secrete a variety of cytokines, chemokines to functioning as the antigen presenter and regulator of inflammation response¹⁰.

Since BD invades the vessels and triggers continuous inflammation, accurate and early diagnosis is crucial for effective disease management. Currently, the International Criteria for Behçet's Disease (ICBD) has been established for clinical diagnosis¹¹, which is largely empirically based^{12,13}. Additionally, despite advanced imaging modalities such as contrast-enhanced multislice helical computed tomography (MSCT) and magnetic resonance imaging (MRI), contributes to the diagnosis and identification of typical vasculitis¹⁴, it still remains only a minor role in BD diagnosis¹⁵. Therefore, novel and precise diagnostic approaches are urgently needed to improve BD detection and management¹⁶.

Here, through integration of single-cell analysis and bulk-seq analysis, we identified the immune infiltration landscape of BD, and identified the triggering role of monocytes in BD occurrence. Additionally, we employed

¹Carnegie Mellon University, Pittsburgh, USA. ²Queen's Belfast University, Northern Ireland, UK. ³China Pharmaceutical University, Nanjing, Jiangsu, China. ⁴The University of Hongkong, Hongkong, China. ⁵Qiyi Yu, Yidong Zhang and Yifan Wu contributed equally. ✉email: qiyiy@andrew.cmu.edu

WGCNA and machine learning methods, discovered 5 novel gene biomarkers for BD diagnosis. BD models constructed by tropomyosin injection validated the bioinformatic results. In summary, this research shed a new light on monocytes in BD, and provided 5 new biomarkers for detection, which introduced novel ideas for BD mechanism elucidation and diagnostic practice.

Materials and methods

Data sources

We screened GEO (<https://www.ncbi.nlm.nih.gov/geo/>) database for mRNA expression profiles of BD patients, which were GSE198533, GSE17114 and GSE209567. The BD-related miRNA genes which expressed differently were extracted from the GSE145191 dataset. ImmPort (<https://www.immport.org/home>) database was used to find immune-related genes. For external validation, GSE69063, GSE205962 and GSE66407 were acquired for exploration of gene expression in sepsis, rheumatoid arthritis and inflammatory bowel diseases. Besides, we collected BD single cell seq-data from GEO database, which was GSE198616, including 8 peripheral blood samples (4 BD patients and 4 healthy controls) uploaded by Zheng W et al.^{17–21}. The detailed information of applied datasets was listed in Table 1.

Immune infiltration landscape of BD

Batch effect might introduce bias to results, so data quality checking targeting bulk-seq data GSE198533 was first conducted. CIBERSORT method was then adopted to estimate immune cell proportion in BD, which uses the specific reference matrix to evaluate the proportions of diverse immune cells²². Via “CIBERSORT” R package (ver 0.1.0), immune cell infiltration evaluation was performed in BD and HC samples for construction of BD immune infiltration landscape.

Single-cell quality control and dimension reduction

We collected BD single cell seq-data GSE198616 from GEO database, and the single-cell analysis was conducted by “Seurat” package (ver 4.3.0)²³. To screen cells that express an appropriate number of genes, we set the cell identification criteria as over 300 genes but less than 4,000 genes. Mitochondrial genes were filtrated with thresholds of 10%. After quality control, the identification of hypervariable genes was performed via FindVariableFeatures in Seurat package, where the criterion is top 2,000 high-variability genes. Subsequently, the Principal Components Analysis (PCA) was conducted and we adopted top 10 PCs. Eventually, we utilized “Uniform manifold approximation and projection (Umap)” method of dimensionality reduction to perform the visualization of cells. According to cell markers from previous provided by Zheng. et al. and common cell markers, the annotation of cell clusters was completed manually (Table 2)²⁴.

Perturbation of cells during BD development

Cells may exhibit different cellular states during the progression of BD, which is potentially correlated to their functions²⁵. To identify the most interfered cell types in BD development, “scDist” (ver 1.1.4) was introduced. scDist provides a statistically rigorous and computationally fast method for detecting cell-type specific differences across multiple groups or conditions. By using a mixed-effects model, scDist estimates the difference of cell cluster between BD and HC while quantifying the statistical uncertainty due to individual-to-individual variation and other sources of variability. Here, in our research, we removed the low-proportion cells with a threshold of $\leq 5\%$ of the total cells, for the potential bias introduced.

Dataset	Categorizes	Author	Platform	Group	Number
GSE198533	Bulk-seq	Zhang Y. et al.	GPL24676	BD	9
				Control	10
GSE17114	Bulk-seq	Krug T. et al.	GPL570	BD	15
				Control	14
GSE209567	Bulk-seq	Oğuz AK. et al.	GPL570	BD	29
				Control	15
GSE145191	Non-coding RNA-seq	Emmi G. et al.	GPL9081	BD	16
				Control	18
GSE69063	Bulk-seq	Bosco A. et al.	GPL19983	Sepsis	68
				Control	42
GSE205962	Bulk-seq	Zhang Y. et al.	GPL16043	RA	16
				Control	4
GSE66407	Bulk-seq	Folkersen L. et al.	GPL19833	UC	161
				CD	103
				Control	99
GSE198616	scRNA-seq	Zhang Y. et al.	GPL24676	BD	4
				Control	4

Table 1. Clinical information of datasets applied in manuscript.

Cell	Marker
B cells	CD79 A, MS4 A1
NK cells	KLRF1, FCGR3 A, GZMB, NCAM1
CD4 + T cells	CD3D, CCR7, IL7R, GZMK
CD8 + T cells	CD3D, CD8 A, GZMB
Monocytes	LY, CD14, ISG15
Plasma cells	JCHAIN, IGHG1, IGKC
DCs	CD1 C, CLEC10 As, CLEC4 C, LILRA4
Platelets	PPBP
Erythrocytes	HBA1

Table 2. Markers for cell cluster annotation.

Cellular trajectory discovery in BD

scDist helped to identify the most impacted cell clusters during BD progress, and we subsequently needed to explore that the differentiation timepoint of these perturbed cells. To identify the differentiation trajectory of cells in BD, we conducted pseudotime analysis. Pseudotime analysis was performed via “monocle” package (ver 2.26.0)²⁶. The screening criteria for annotated cell clusters were set as expressions >0.1 and dispersion empirical >1*dispersion fit cells. The dimension reduction was performed by “DDRTree” method and specific cell differentiation trajectories were identified with “reduceDimension” function. Eventually, the cell differentiation states were then identified and visualized. Also, we employed “Slingshot” package for further illustration of cellular trajectory²⁷. Slingshot is developed for inference of cellular lineages. It consists of two main steps: global lineage structure inference based on MST construction, and inference of pseudotime variable for cells in each lineage by fitting smooth branching curves to these lineages.

Phenotype correlation of cell clusters in BD

The Scissor R package was used to integrate BD bulk and single-cell data and BD/HC group information to identify cells associated with BD phenotype²⁸. Briefly, the Scissor algorithm first quantifies the similarity between single-cell data and bulk data. Then Scissor optimized the regression model between the correlation matrix and the phenotype. Next, Scissor calculated the cells associated with the phenotype by imposing a sparsity penalty and graph regularization. Scissor’s built-in classification models identified cells that are positively or negatively correlated with the assigned phenotype. For the classification mode, Scissor returned Scissor + cells (assigned as 1) that were positively correlated to BD, Scissor – cells that negatively correlated to BD (assigned as 2), or background cells with no significant association (assigned as 0).

Cell–Cell communication analysis for bd’s subclusters

We utilized “CellChat” (ver 1.6.1), the R package that possesses ligand-receptor (L-R) interaction databases to evaluate cell-cell communication networks based on annotated RNA sequence data²⁹. CellChat provided a method to infer probability of cell-cell communications between specific types of cells in samples, based on prior knowledge of communications between various signaling ligand-receptor pairs and cofactors.

Identification of differentially expressed genes (DEGs)

Immune infiltration landscape and single-cell analysis provided insights into key immune cell subpopulations and their functional states. To further explore gene expression alterations associated with BD and their potential clinical relevance, we conducted differential expression analysis of genes. DEGs in original data between the BD and control group were filtered in employment of “DESeq2” package in R software (ver 1.38.3)³⁰. With a threshold of adjusted P-value <0.05, for statistical significance, a volcano map was drawn of DEGs via R software.

Weighted gene co-Expression network analysis (WGCNA) and module gene selection

We employed WGCNA package (ver 1.72-1) to create a scale-free mRNA co-expression network in dataset on the basis of BD and HC groups³¹. First, we identified the pearson correlation coefficient of gene pairs to construct a similarity matrix and transformed the similarity matrix into an adjacency matrix. Subsequently, we picked the appropriate value β with “pickSoftThreshold” function in WGCNA to emphasize strong associations and penalize weak associations between given groups and expressed genes. By analyzing gene significance and module membership, the co-expressed genes were clustered into modules and correlations between groups and modules were calculated.

Functional enrichment analysis and PPI network

After performing WGCNA, we intersected genes from DEG analysis, WGCNA modules related to BD, and Immune-related genes extracted from ImmPort Database. To characterize the function of these genes, we conducted Kyoto Encyclopedia of Genes and Genomes (KEGG)/Gene Ontology (GO) functional enrichment^{32–35}, by package “clusterProfiler” in R software (ver 4.6.2)³⁶.

Meanwhile, using STRING database (<https://cn.string-db.org/>), we established a PPI network for the discovery of interactions between gene-coding proteins. Parameters were set as: parameters of meaning of network edges (“evidence”), minimum required interaction score [“Medium confidence (0.400)”] and max

number of interactors to show (“no more than 10 interactors” in 1 st shell). Genes that could function as junctions in PPI network were chosen to conduct further analysis.

Construction of the miRNA-mRNA regulatory network

miRNAs bind to complementary sequences on target mRNAs at the post-transcriptional level, leading to gene expression suppression through translational inhibition or mRNA degradation³⁷. We employed TargetScanHuman 7.2 (https://www.targetscan.org/vert_72/) to predict miRNAs of PPI-related genes³⁸. GSE145191 provided the BD-related miRNA expression profile, from which we identified differentially expressed miRNAs (DEmiRNAs). The regulatory network between miRNAs and mRNAs was then visualized with the software Cytoscape³⁹.

Machine learning and diagnostic value evaluation

To further locate the most significant diagnostic genes for BD, we employed Least Absolute Shrinkage and Selection Operator (LASSO), a regression method⁴⁰. We utilized “glmnet” package (ver 4.1–7) to perform the selection process of LASSO regression⁴¹. Genes identified through regression-based filtering were considered candidate hub genes for BD diagnosis. Moreover, to associate the selected genes with infiltrated immune cells, we also calculated the correlations between candidate hub genes and immune cells with significant differential proportions in BD.

Subsequently, the Receiver Operating Characteristic (ROC) curve was right after LASSO process using pROC package (ver 1.18.0)⁴², which was established to analyze diagnosis values of candidate hub genes or immune cells with significant differential proportions in BD diagnosis. The calculation of area under the curve (AUC) was plotted to measure diagnostic value, and 95% CI or p.value were also calculated. AUC = 0.7 was the threshold of the value of diagnosis. We adopted the array dataset GSE17114 and GSE209567, and before ROC validation, we performed normalization using “NormalizeBetweenArrays” function in “limma” package (ver 3.54.0)⁴³.

Reagents preparation

Shrimp meat was placed in buffer A (50 mmol/L KCl, 2 mmol/L NaHCO₃) solution, homogenized thoroughly, and centrifuged at 4 °C. Collecting the precipitate, then we repeated homogenization and centrifugation for 4 times. Next, the centrifuged precipitate was then placed in pre-cooled acetone and stirred thoroughly for a powdery precipitate. Placing precipitate in buffer B (20 mmol/L Tris-HCl, 1 mol/L KCl, 0.1 mmol/L DTT) overnight, then centrifuged at 4 °C to collect the precipitate. We adjusted the pH of the supernatant to 4.5, and slowly added 40% ammonium sulfate solution until precipitation deposited. We collected the precipitate at 4 °C centrifuge, 12,000 rpm. dissolved the above precipitate with buffer B respectively, heated in a boiling water bath for 10 min and then quickly cooled the reagents on ice, centrifuged at 4 °C to collect the supernatant. The shrimp tropomyosin was dissolved in the supernatant, and then we dialyzed the supernatant overnight. Using BCA method to detect protein concentration. The shrimp tropomyosin solution was diluted into 12 mg/mL, then an equal volume of complete Freund's adjuvant (Sigma-Aldrich, F5881) was added and the system was mixed thoroughly for the water-in-oil emulsion. The final concentration of tropomyosin emulsion was 6 mg/mL, and the emulsion should be prepared right before for use.

BD model construction

The BD mouse model construction was constructed according to previous research^{44,45}. Mice were maintained under specific pathogen-free facilities. All mice used were of the Balb/c type. Mice were acclimatized for 7 days and then randomly assigned into the model group and control group. We administered 100 µl of the tropomyosin emulsion and PBS via intraperitoneal infection to construct the model group and control group respectively. The day of first injection was identified as day 1. Apply 200 µl intraperitoneal injection to both groups at day 8, 14 and 21. After model construction, the blood samples were collected for the qRT-PCR experiment. For the purpose of histological evaluation, the colon tissues exhibiting the highest degree of inflammation were preserved in 4% paraformaldehyde, followed by embedding in paraffin. The sections underwent staining with Hematoxylin and Eosin, after which they were evaluated and scored according to their pathological characteristics.

HE staining & HAI

The tissue sections were sequentially immersed in xylene for 10 min (twice), followed by anhydrous ethanol for 5 min, 95% ethanol for 5 min, 80% ethanol for 5 min, and 75% ethanol for 5 min, and then rinsed with tap water. Hematoxylin staining was performed by immersing the sections in hematoxylin solution for 5 min, followed by rinsing with tap water to remove excess stain and restore the blue color. Subsequently, eosin staining was carried out by immersing the sections in eosin solution for 1 min, followed by rinsing with tap water. The sections were then dehydrated through three changes of anhydrous ethanol (2 min each) and cleared in xylene (5 min, twice). Finally, the sections were mounted with neutral gum and coverslipped before being scanned using a 20× objective lens.

Histological activity index (HAI) was calculated as the sum of the epithelium and infiltration score, resulting in the total HAI score ranging from 0 (unaffected) to 8 (severe colitis). Histological scores were blindly determined as per Obermeier et al.⁴⁶. The following parameters were used for calculation: (a) Epithelial damage (0 point: none, 1 point: minimal loss of goblet cells, 2 points: extensive loss of goblet cells, 3 points: minimal loss of crypts and extensive loss of goblet cells, and 4 points: extensive loss of crypts) (b) Infiltration (0 point: none, 1 point: infiltrate around crypt bases, 2 points: infiltrate in muscularis mucosa, 3 points: extensive infiltrate in muscularis mucosa with edema, and 4 points: infiltration of submucosa).

Expression validation

The expression of selected genes was validated both in experimental level and external dataset level. First, we adopted BD dataset GSE209567 to explore the expression of inflammation factors in BD and normal group. To investigate the difference of expression pattern of selected genes between BD and immune-related diseases, external datasets including sepsis, rheumatoid arthritis (RA) as well as inflammatory bowel diseases involving ulcerative colitis (UC) and Crohn's disease (CD) were acquired and the expression of selected genes in different datasets were presented (detailed information listed in Table 1)^{47–49}. RNA from cells was isolated using Trizol reagent (Invitrogen) and reverse-transcribed using Hiscript III Reverse Transcriptase (Vazyme, R302-01) according to the manufacturer's instruction. Real-time quantitative PCR (RT-qPCR) was performed in a QuantStudio 3 using Taq Pro Universal SYBR qPCR Master Mix reagents (Vazyme, Q712-02).

Immunohistochemical staining methods

Antigen retrieval was performed by immersing the slide in 0.01 M citrate buffer (pH 6.0) and heating it in a microwave at 100 °C for 10 min or boiling for 7 min, followed by an additional 7 min of heating. The solution was then cooled to room temperature (~ 20 min), with optional ice to accelerate cooling. Once cooled, the slide was rinsed with PBS buffer for 5 min, repeated three times. Endogenous peroxidase activity was blocked by drawing circles around the tissue specimen with an oil-based immunohistochemistry pen to prevent reagent overflow. After dewaxing, 3% hydrogen peroxide (H₂O₂) was applied to the specimen, followed by three 5-minute PBS rinses. The slide was gently dried and placed in an incubation box. Normal sheep serum (Solution B) was applied as a blocking agent and incubated at room temperature for 10 min, then left to dry without washing. Finally, the primary antibody was added, the slide was covered, and incubated overnight at 4 °C.

Statistical analysis

R software v.4.2.3 was used to perform bioinformatic analysis. Wilcoxon.test was applied in external dataset validation. Statistical analysis of experiments was performed using SPSS 22.0 software, Origin 2022. Two-tailed and unpaired Student's t-tests were performed to compare differences. Data were presented as the means ± standard deviation (SD). Statistical significance was defined as a P-value less than 0.05. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001; ns, no significance.

Results

Immune cell landscape in BD

Our workflow was illustrated as Fig. 1A. Until now, numerous experiments, as well as clinical practice, have evidenced that immune cell activities are closely related to BD⁵⁰. To optimize the elucidation of immune-regulated regarding to BD, the immune cell infiltration analysis was performed. 22 sorts of immune cells' proportions in each sample of BD and control groups are shown in bar plot (Fig. 1B). Here, the results demonstrated a significant higher level of monocytes and a lower level of B memory cells, neutrophils and Natural Killer (NK) resting cells (Fig. 1C) in BD patients.

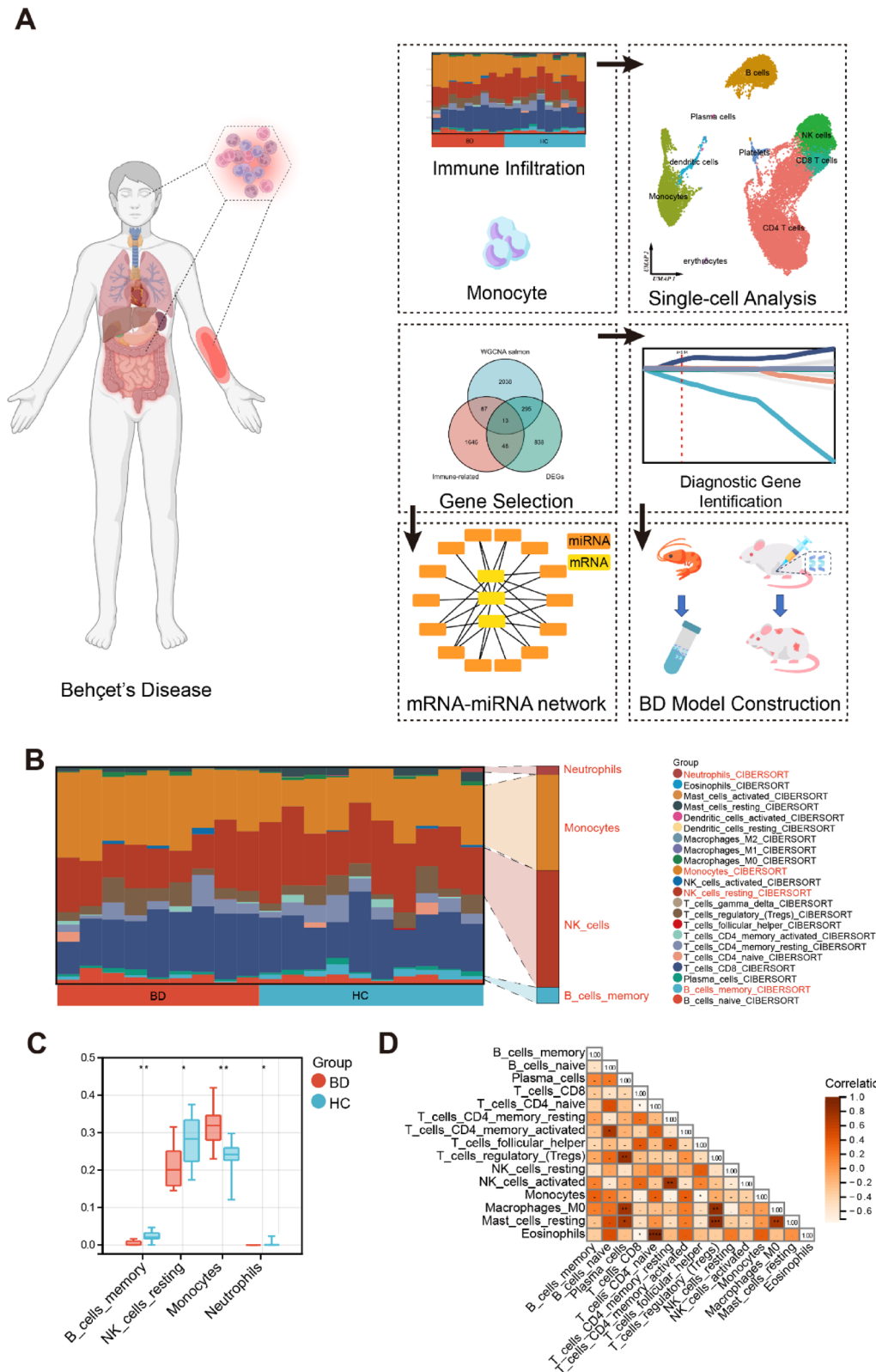
We used package “corrplot” in R software (ver 0.92) to plot correlation heatmaps⁵¹. The correlation of 15 types of immune cells revealed that there are positive associations between B memory cells-T CD4 memory activated cells, T regulatory cells-Plasma cells, Macrophages M0-Plasma cells, Mast resting cells-Plasma cells, Eosinophils-T CD4 naive cells, NK activated cells-T CD4 memory resting cells, Mast resting cells-T regulatory cells, Macrophages M0-T regulatory cells, and Mast resting cells-Macrophages M0. On the contrary, negative associations occurred between T CD4 naive cells-T CD8 cells, Eosinophils-T CD8 cells, and Monocytes-T follicular helper cells. Generally speaking, the differentially infiltrated immune cells in BD may reveal key regulatory points, providing insights for further mechanistic exploration (Fig. 1D).

Single-cell quality control and dimension reduction clustering

Quality Control (QC) was processed upon the single-cell seq-data (Fig. S1A and S1B). 2,000 high-variability genes were selected as the threshold, with top 10 PCs selected after PCA (Fig. S1C). Subsequently, we classified cells into 18 clusters (Fig. S1D) with “Umap” dimensional reduction and performed cell annotation with the marker genes provided by Zheng et al.²⁴. The annotated cell clusters were illustrated in Fig. 2A, including B cells, plasma cells, monocytes, NK cells, platelets, erythrocytes, dendritic cells (DC), CD4 + T cells and CD8 + T cells. By comparing cell clusters of BD patients with those of healthy control, we observed an obvious increase in Monocytes, and the decrease in B cells and CD8 + T cells (Fig. 2B). The expression of genes for manual annotation in different cell clusters were also visualized in Fig. 2C. During the development of BD, the cellular state of cells might be perturbed (Fig. 2D). scDist revealed the most perturbed cell clusters during the process of BD development, which was monocytes. These results suggested the remarkable alternation of cellular state occurred in monocytes during the BD progression.

Characteristics of immune-related cells in BD

Immune-related cells may play different roles at various stages of BD progression. Thus, we conducted the pseudotime analysis to explore the cell differentiation trajectory of the immune cells in BD. Here, we randomly sampled 10,000 cells from single cell samples of BD, and all downstream single-cell analysis is based on these cells. The inferred cell trajectories both by monocle and Slingshot revealed that monocytes represented the earliest immune, cell population during differentiation, being the initial lineage (Fig. 2E). Subsequently, T cells developed into two distinct branches, suggesting functional heterogeneity within this type of cells. Additionally, NK cells and B cells were primarily located on the terminal stage, indicating their involvement in later phases of immune response regulation.



To correlate the cells with BD phenotypes, we utilized “Scissors” package. We applied GSE198533 as the bulk-seq dataset, and identified the associations between cells in single-cell dataset and BD phenotypes (Fig. 2F). Here, we assigned the state of cells “0” as positive correlation with HC, “1” as positive correlation with “BD” phenotype, “2” as negative correlation with “BD” phenotype. Compared to most of immune cells in BD, CD4 + T cells, CD8 + T cells and monocytes showed higher proportion of cells with positive correlation to BD phenotype. On contrary, B cells, plasma and DC cells were negatively correlated with BD. These results might indicate that monocytes, CD4 + and CD8 + T cells were activated during the BD development, while B cells, plasma and DC were inhibited. Given that monocytes were in the initial stage of differentiation trajectory, we infer that upon

◀ **Fig. 1.** Immune-related infiltration of BD patients and controls. (A) Schematic workflow of the research methodology, illustrating the integration of bulk RNA-seq, single-cell analysis, and experimental validations. (B) Stacked bar plot visualization of 22 different immune cell proportions in individual BD and control samples, as determined by CIBERSORT deconvolution algorithm. (C) Box plots comparing immune cell types with statistically significant differential proportions between BD patients and healthy controls, highlighting increased monocytes and decreased B memory cells, neutrophils, and NK resting cells in BD. (D) Correlation heatmap of 15 immune cell type compositions, with red indicating positive correlations and blue indicating negative correlations between cell populations. Statistical significance: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. Horizontal and vertical axes represent immune cell subtypes. The flowchart was created by Biorender webtool (<https://www.biorender.com/>).

the BD occurrence, monocytes were firstly stimulated, subsequently functioning as the antigen presenter and triggering further immune response.

BD's inflammation process involved the modification of cell-cell communication. "CellChat" package was adopted to conduct intercellular communication analysis on annotated cell clusters (Fig. 2G and H). Here, we have identified some noticeable pathways, in cellular communications related to monocytes, including BAFF, cholesterol, SEAM4 and PECAM. BAFF activates B cells in inflammation response, while cholesterol triggers NLRP3 inflammasomes, leading to IL-1 β and IL-18 secretion. Additionally, PECAM and SEAM4 affect vascular permeability and monocyte migration. Generally, the secretion of cytokines and migration of monocyte enhance leukocyte recruitment and inflammatory cascades, which might aggravate the inflammation process and development of BD.

Identifications of DEGs

To further investigate gene expression alterations in BD and their potential clinical significance, we performed differential expression analysis on BD bulk-seq data GSE198533. Via DESeq2 method, 2,526 DEGs were screened out, including 1,969 up-regulated genes and 557 down-regulated genes. Relevant results were displayed in volcano plot Fig. S2 A. Setting the criterion of $|\log_2(FC)| > 1.5$, we filtered out 1,217 DEGs, of which 1,050 were upregulated and 167 were downregulated.

WGCNA and key module identification

Clustering dendrogram of BD and control samples was illustrated in Fig. 3A. The most correlated modules in MS were identified through WGCNA. We selected 9 (scale-free, $R^2 = 0.78$) as the soft threshold, on account of average connectivity and scale independence (Fig. 3B). 24 co-expression modules of genes were generated in different colors and correlations between groups and co-expression modules were presented in Fig. 3C. The salmon module (2,433 genes) showed a correlation coefficient of 0.86 and p-value of 2.7×10^{-6} with BD group, which was identified to be a critical module for the following analysis. Additionally, after calculating the correlations between module membership and gene significance in salmon module for BD, we discovered a remarkable positive correlation ($r = 0.92$) illustrated in Fig. S2B. Thus, we identified the salmon module as the gene module that were most significantly connected to BD.

Functional enrichment analysis of BD

Based on 13 common genes (CGs) from the intersection of the DEGs, salmon module genes of BD, and immune-related genes from ImmPort, we performed the functional enrichment analysis (Fig. 3D). KEGG/GO analysis reveals that the "Neuroactive ligand-receptor interaction" (Fig. 3E), the "adaptive immune response" and "antigen-binding" (Fig. 3F) were the pathways that CGs primarily enriched respectively.

PPI network construction

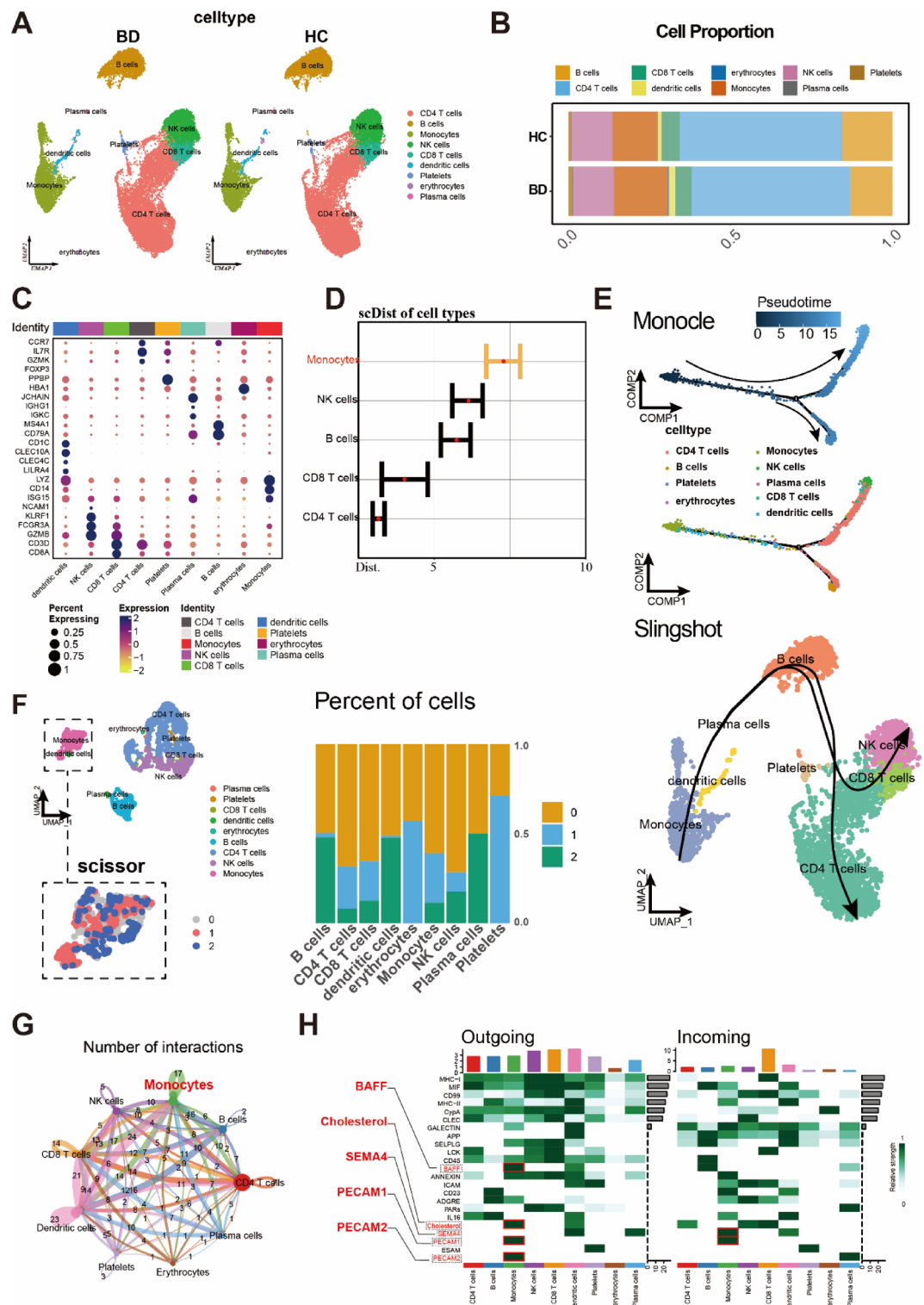
After gene intersection, a PPI network was constructed to locate protein-coding genes which might possess interplay with each other for the following miRNA-mRNA regulatory network construction. Figure 4A shows the PPI network which illustrates 18 gene-encoding proteins could interact with each other.

Construction of interaction regulatory network

As the potential regulatory mechanism of interaction still needed to be further demonstrated, a miRNA regulatory network of integrated analysis was conducted and visualized. TargetScanHuman 7.2 miRNA database was employed to predict target miRNAs of hub genes. Furthermore, we analyzed GSE145191 to identify BD-related DE miRNAs. The intersection of DE miRNAs and miRNAs related to PPI genes was visualized using Cytoscape software. We obtained 14 target miRNAs of 9 specifically expressed genes (Fig. 4B). Finally, we used the cytoHubba MCC to select out hub genes. From the network constructed, those identified genes including GNAQ, GFRA1 and AVPR1 A were regulated by 14 miRNAs (Fig. 4C).

Identification of candidate diagnostic genes and value assessment

LASSO regression was utilized to select candidate genes for BD diagnosis. Figure 5A and B showed that LASSO method picked out 5 prospective biomarkers (NRTN, TRDJ1, IGLV4-69, PDIA2 and AVPR1 A). Next, the corresponding ROC curve illustrated each gene's specificity and sensitivity of diagnosis, followed by calculation of the AUC and 95% CI for each item, which were presented as follows: AVPR1 A (AUC 1.000, CI 1.000–1.000), IGLV4-69 (AUC 0.989, CI 0.958–1.000), NRTN (AUC 0.922, CI 0.806–1.000), PDIA2 (AUC 0.956, CI 0.861–1.000), TRDJ1 (AUC 1.000, CI 1.000–1.000) (Fig. 5C–G). All selected genes expressed high diagnostic



significance for BD. Then the mRNA expression profile microarray GSE17114 was used for validation (Fig. 5H). The expressions of *NRTN*, *PDIA2* and *AVPR1 A* were combined to increase the prediction efficiency of BD. The result showed that the combined group (AUC 0.710, CI 0.512–0.907) still had a high diagnostic value for BD in the validation set.

Moreover, since we have identified the immune cells that have significant differential proportions in BD, we have also explored their diagnostic significances in clinical practice. In GSE198533, we first calculated the correlations between diagnostic genes and immune cell proportions (Fig. S3 A). *TRDJ1*, *IGLV4-69* and *AVPR1 A* showed the significant positive correlations with B cells, plasma cells and NK resting cells, while negatively correlated with monocytes and macrophages. *NRTN* showed the opposite trend, with high positive correlation with M0 macrophage. Particularly, monocyte showed the significant negative correlation with *TRDJ1* and *AVPR1 A*. Subsequently, we applied the proportions of immune cells on the diagnosis of BD (Fig. S3B and S3

◀ **Fig. 2.** Single-cell analysis of BD patient samples from GSE198616. **(A)** UMAP visualization of reduced dimensionality and cluster analysis, with cells classified into nine major cell populations based on marker gene expression. **(B)** Comparative bar chart showing cell type proportions in BD patients versus healthy controls (HC), demonstrating significant increases in monocytes and decreases in B cells and CD8 + T cells in BD. **(C)** Dotplot visualization of key marker genes used for cell type annotation, with dot size representing percentage of cells expressing the gene and color intensity indicating expression level. **(D)** scDist analysis showing the perturbation magnitude of different cell types from healthy state to BD state, with monocytes exhibiting the highest perturbation score. **(E)** Pseudotime trajectory analysis of immune cells in BD, with differentiation development order. For monocle method, the order was indicated by color gradient (darker blue to lighter blue) and directional arrows. For slingshot, the lines and arrows suggested the direction. Monocytes appear at the earliest pseudotime point. **(F)** Scissor analysis integrating single-cell data with bulk RNA-seq dataset GSE198533, showing cells positively correlated with BD phenotype (red), negatively correlated (blue), or not significantly associated (gray). **(G and H)** Cell-cell communication network analysis showing interaction strength **(G)** and specific signaling pathways **(H)** among different cell types, with BAFF, cholesterol, SEMA4, and PECAM pathways most strongly activated in monocytes.

C). ROC curves revealed that among the immune cells with significant differential proportion between BD and HC, monocytes showed the highest accuracy of diagnosis (AUC 0.833, *p*-value 0.048). Moreover, the prediction ability was further optimized after integrating the correlated diagnostic genes, with an AUC of 0.9. These results indicated the diagnostic value of monocytes in BD clinical practice, which also indicated the early lesion of monocytes in BD progress.

BD model construction

Modeling procedure of BD was presented in Fig. 6A. In the BD model mice, disease progression is characterized by significant weight loss and arthritis (Fig. 6B and C). Additionally, colitis in BD mice is marked by a significant thickening and diffused inflammation of the colon. The Histological Activity Index (HAI) of the rectum shows a significant increase in the model group compared to the control group. This model is characterized by intestinal gland damage, visible adhesions, and fibrosis, leading to a significant increase in HAI (Fig. 6C). We observed that patients with BD typically exhibit both colorectal inflammation and joint inflammation. In the model group mice, Safranin O-Fast Green staining of the joint regions showed thinner cartilage tissue with unclear cartilage boundaries (Fig. 6D), indicating inflammatory features. Furthermore, immunohistochemical analysis revealed a higher level of TNF infiltration in the articular cartilage tissue, indicating that the tissue was in an inflammatory state (Fig. 6E). Similarly, the immunohistochemical results of TNF in the colorectal tissue demonstrated that the colorectal region was also in an inflamed condition (Fig. 6F).

Cytokines are the primary mediators of innate and adaptive immune responses in mucosal inflammation. To validate the immunodysregulation characteristics of the BD mouse model, we conducted an analysis of serum cytokine expression in a α -tropomyosin treated mouse model. Compared to the control group, the BD model mice exhibited cytotoxicity and chemotaxis, characterized by (intrinsic) immune dysregulation, with significantly elevated levels of Tnf, Il-1 β , Il-6 and Tgfb1 (*P* < 0.05) (Fig. 6G). BD mice modeled with α -tropomyosin exhibit significant intestinal and joint inflammation, along with symptoms of immune dysregulation. To comprehensively evaluate the expression of inflammation-related factors, we adopted GSE209567 dataset (Fig. 6H and S3D). In BD group, key inflammation factors showed the trend of upregulation. The dataset validation is corresponded to the experiment results, which evidenced the ever-increasing extent of inflammation symptoms.

Moreover, since cholesterol communication in monocyte triggers NLRP3 inflammasomes, leading to IL-1 β and IL-18 secretion, we also investigated NLRP3 inflammasome-related genes in BD dataset in biological experiments and dataset GSE198533. In BD mice model, we validated the upregulation and expression of Nlrp3 in blood and tissues, using qPCR and IHC method. Besides, NLRP3 and CASP1 showed upregulation trend in BD groups in bulk-seq data (Fig. 6I-K).

The expression levels of diagnostic genes in external datasets

IGLV4-69, TRDJ1, AVPR1 A, NRTN, and PDIA2 are expressed in human samples; However, there are no homologous gene sequences for IGLV4-69 and TRDJ1 in mice. To validate the expression of AVPR1 A, NRTN, and PDIA2 in blood cells, RT-qPCR analysis was conducted on blood samples from both BD model mice and the control group, the primers of selected genes were presented in Table 3. The expression levels of NRTN and PDIA2 were significantly elevated in the model group compared to the control group, whereas AVPR1 A exhibited a significant decrease (Fig. 6L). The expression patterns of 5 diagnostic genes in BD validation set GSE209567 and different diseases were presented in Fig. 6M (BD validation set) and S4 (different diseases). The results indicated that the TRDJ1 and IGLV4-69 decreased in BD and sepsis, NRTN and PDIA2 expressed upregulated trend in BD, sepsis, arthritis and inflammatory bowel diseases. AVPR1 A was downregulated in sepsis, BD and inflammatory bowel disease, while upregulated in arthritis.

Discussion

BD has become a global health concern with increasing prevalence worldwide⁵². Despite the development of several targeted therapies, further elucidation is still needed regarding BD pathogenesis. In our research, we employed bioinformatics approaches to gain new insights into BD through comprehensive transcriptomic analysis. By integrating immune infiltration and single-cell analyses, we identified monocytes as the primary

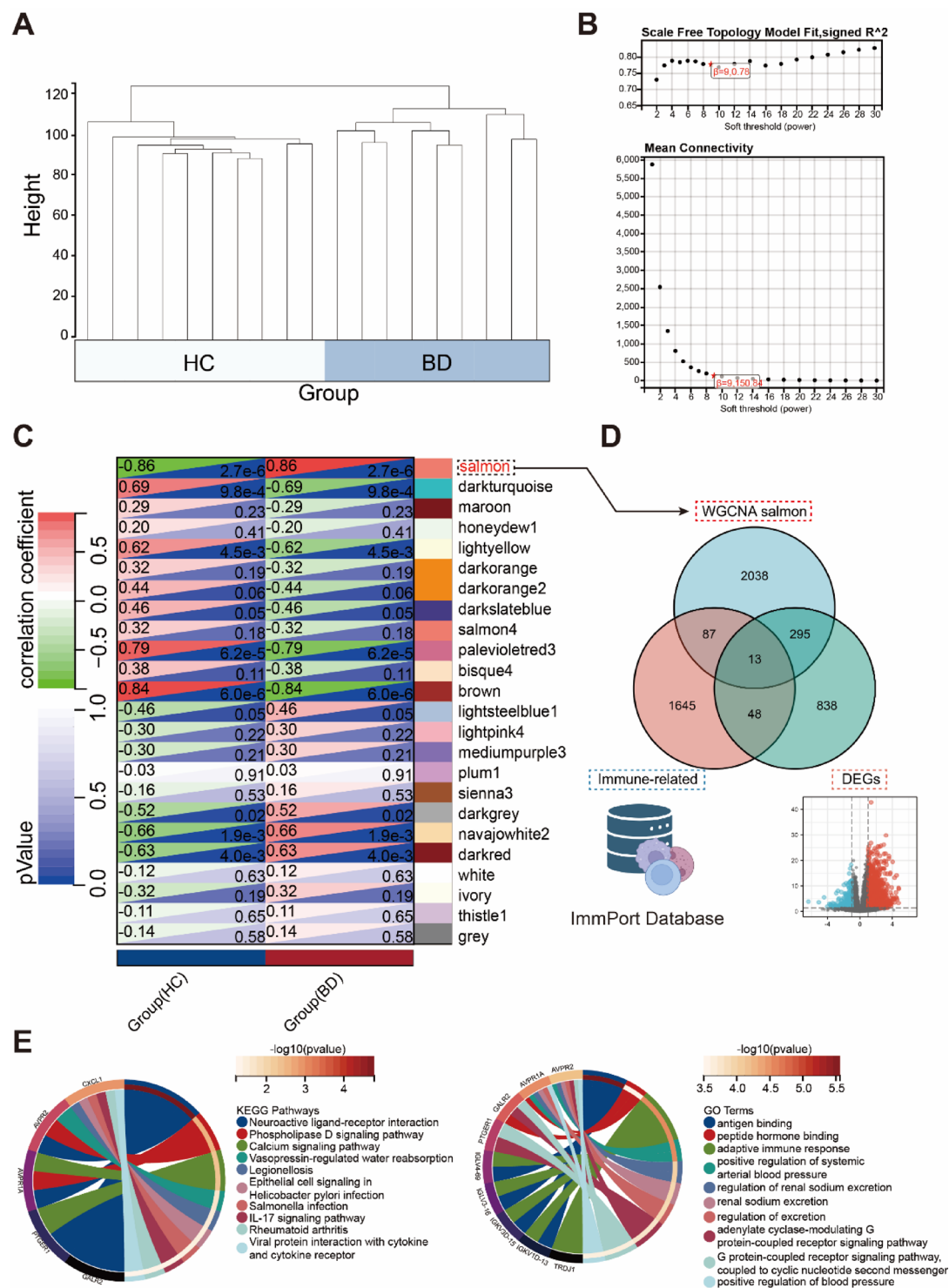


Fig. 3. Identification of module genes through WGCNA in BD. **(A)** Clustering dendrogram of the BD and control samples. **(B)** $\beta = 9$ is chosen to be the soft threshold based on scale independence and average connectivity. **(C)** Heatmap of the association between modules and samples. Values at the top and bottom edges stand for the correlation coefficient and p-value, respectively. **(D)** Venn diagram illustrated 13 identified genes from the intersection of the DEGs, the screened module genes of BD and immune-related genes. Illustrations in Fig. 3D was created using BioRender webtool (<https://www.biorender.com/>). **(E)** The CG's KEGG and **(F)** GO enrichment results. Different colors in diagrams represent various significant pathways and related enriched genes. The usage of KEGG image was under the guideline of KEGG database (<https://www.kegg.jp/>), and obtained the permission from Kanehisa Lab (Ref: 251500).

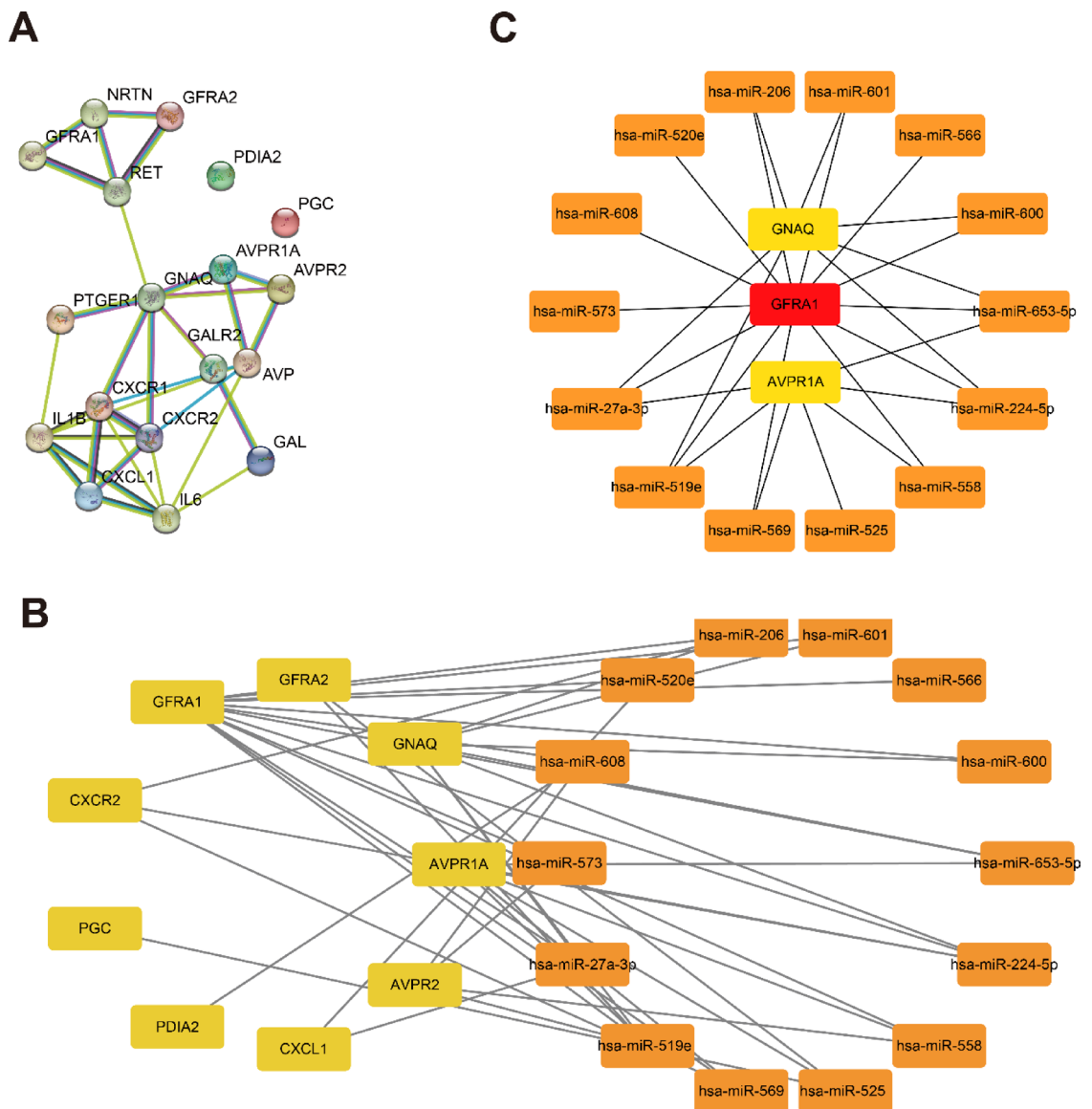


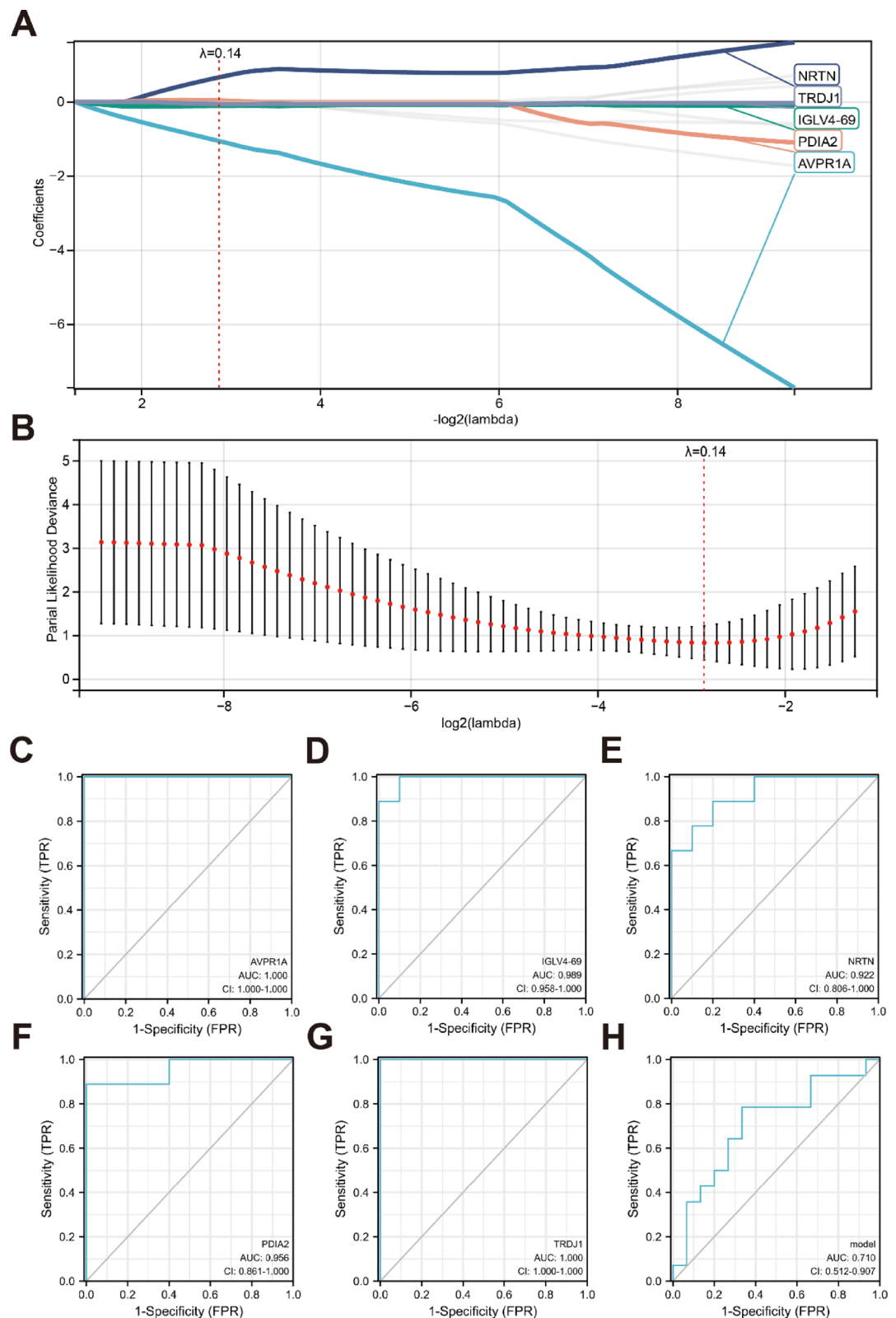
Fig. 4. Construction of the PPI Network and miRNA-mRNA regulatory network. (A) PPI network of common genes. (B and C) The miRNA-mRNA regulatory network extracted by MCC.

drivers of inflammatory responses in BD. Additionally, our bulk RNA sequencing analysis revealed potential novel biomarkers with significant diagnostic value for BD.

BD is diagnosed as a systemic vasculitis with invasive vascular inflammatory lesions, thus the variations of immune-related cells might suggest the driving factors of BD⁵³. Our immune infiltration analysis of BD and healthy control bulk-seq samples revealed significant changes in immune cell proportions. Specifically, BD patients showed increased monocyte proportion, and decreased B memory cells and NK resting cells. Correlation analysis identified both positive and negative associations between specific immune cell populations, highlighting the complex immune dysregulation in BD. These differential proportions of infiltrated immune cell might suggest potential novel therapeutic targets. To further characterize these immune alterations at higher resolution, we further performed comprehensive single-cell analysis.

We have conducted single-cell analysis via R package “Seurat”. After implementing quality control and filtering cells, we applied unbiased clustering of cells. Clustering process identified several main subclusters, which are CD4 + T cells, CD8 + T cells, NK cells, B cells, plasma cells, monocytes, DC, platelets and erythrocytes. Compared with the control group, BD patients obviously expressed a higher proportion of monocytes and a lower proportion of B cells, which is corresponding to “Cibersort” package analysis results. scDist revealed that monocytes received the highest perturbation impact, suggesting that monocyte alterations may be a primary driver of BD pathogenesis.

Subsequently, we conducted the trajectory analysis and phenotype correlation analysis. The results showed that monocytes showed the earliest differentiation order and relative high correlation with BD, compared to other types of cells. Monocytes participate in the early immune response after phagocytosis of antigens, as



they will transfer the antigen determinant to lymphocytes, to induce the subsequent specific immune response of other cells⁵⁴. Generally, single-cell analysis results indicated the monocyte might be the triggering cells of immune response in early BD.

Cell-cell communication reveals the possible interactions among selected cell types on basis of annotated ligand-receptor expression of each other. Monocytes expressed high communication intensity, which has established various interaction approaches to other cells such as NK cells, T cells and B cells. Especially, monocytes showed high communication intensities in BAFF, cholesterol, SEMA4, and PECAM pathways. BAFF (B-cell Activating Factor) was primarily released by myeloids, which is essential for B cell development, survival, and

Fig. 5. Machine learning in screening candidate diagnostic biomarkers for BD. **(A)** LASSO coefficient profiles of the 13 candidate genes, with each colored line representing a gene's coefficient path as the regularization parameter λ increases. **(B)** Cross-validation error curve for LASSO model selection, showing mean squared error with standard deviation. The optimal model with lowest error identified five genes ($n = 5$) as the most suitable combination for BD diagnosis. **(C–G)** Receiver Operating Characteristic (ROC) curves demonstrating the diagnostic performance of each candidate gene: **(C)** AVPR1 A (AUC 1.000, CI 1.000–1.000), **(D)** IGLV4-69 (AUC 0.989, CI 0.958–1.000), **(E)** NRTN (AUC 0.922, CI 0.806–1.000), **(F)** PDIA2 (AUC 0.956, CI 0.861–1.000), and **(G)** TRDJ1 (AUC 1.000, CI 1.000–1.000). **(H)** ROC curve showing the combined diagnostic performance of NRTN, PDIA2, and AVPR1 A in the independent validation dataset GSE17114 (AUC 0.710, CI 0.512–0.907). AVPR1 A, Arginine Vasopressin Receptor 1 A; IGLV4-69, Immunoglobulin Lambda Variable 4–69; NRTN, Neurturin; PDIA2, Protein Disulfide Isomerase Family A Member 2; TRDJ1, T Cell Receptor Delta Joining 1.

differentiation⁵⁵. BAFF adaptive immune response to sustain inflammation; The cholesterol pathway activates NLRP3 inflammasomes, triggering the release of pro-inflammatory cytokines IL-1 β and IL-18 that exacerbate chronic inflammation⁵⁶; Additionally, SEMA4 and PECAM pathways facilitate monocyte migration through vascular endothelium, enhancing leukocyte recruitment and tissue infiltration^{57,58}. These results revealed the specific molecular mechanisms of the monocytes to trigger immune response and regulate other immune cells in BD, discovered the role of initial responders of monocytes in BD occurrence.

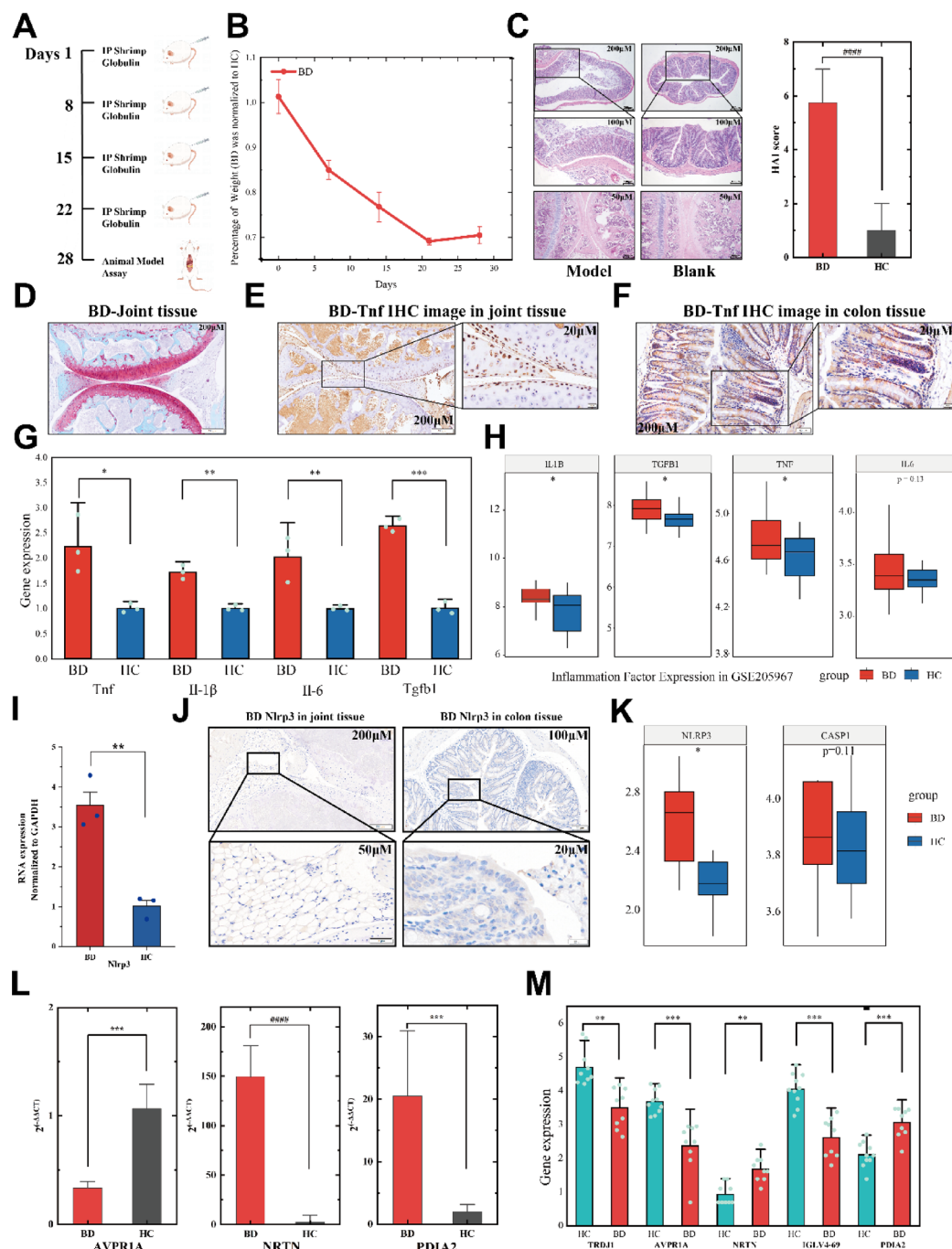
Subsequently, we focused on novel diagnostic biomarkers of BD. From GEO database, we have obtained datasets that contain peripheral blood from BD patients and controls, and after intersecting DEGs, WGCNA salmon module genes and immune-related genes, 13 common hub genes were screened out. KEGG and GO functional enrichment analyses revealed that these genes were predominantly involved in “neuroactive ligand-receptor interaction,” “adaptive immune response,” and “antigen-binding” pathways. These findings align with current pathogenic theories of BD, which propose that activation of immune responses leads to the release of inflammatory cytokines and chemokines that perpetuate the inflammatory cascade characteristic of BD⁵⁹.

To explore the potential regulatory mechanism between these genes, PPI network and miRNA regulatory network were conducted. We noticed that the enrollment of CXCR1/2 and IL-18/IL-1 β in protein regulation network revealed their integration with broader inflammatory signaling networks in BD. CXCR1 and CXCR2, primarily responsive to IL-8, promote myeloid chemotaxis and degranulation downstream of NF- κ B activation⁶⁰. Similarly, IL-1 β and IL-18, produced through NLRP3 inflammasome activation, amplify inflammatory cascades by inducing additional cytokines and enhancing T cell responses⁵⁶. Our network analysis demonstrated that these molecules form functional clusters with significant cross-talk mediated primarily through the inflammation pathways. The identification of these integrated inflammatory networks highlights potential targets for therapeutic intervention aimed at disrupting these self-amplifying inflammatory loops. mRNA-miRNA regulatory network indicated that the hub genes of GNAQ, GFRA1 and AVPR1 A were regulated by 14 miRNAs. Then through LASSO method, 5 potential candidate biomarkers out of 13 screened DEGs were selected, which are NRTN, TRDJ1, IGLV4-69, PDIA2 and AVPR1 A.

The five identified biomarkers have distinct biological functions relevant to BD pathogenesis. Neurturin (NRTN), a neurotrophic factor expressed on both innate and adaptive immune cells^{61–63}, likely modulates immune responses during BD inflammation. T Cell Receptor Delta Joining 1 (TRDJ1) participates in antigen recognition and triggers acute, innate-like immune responses essential for pathogen clearance and tissue recovery^{64–66}. Its downregulation in BD suggests compromised immune function as the disease progresses, supporting its potential as a diagnostic biomarker. Immunoglobulin Lambda Variable 4–69 (IGLV4-69), a plasma membrane protein involved in immunoglobulin complex formation⁶⁷, shows decreased expression in BD patients, potentially reflecting damage to immunoglobulin complexes during disease progression. Protein Disulfide Isomerase Family A Member 2 (PDIA2) functions as an intracellular estrogen-binding protein that modulates cellular functions and prevents protein misfolding. PDIA2 also participates in antigen presentation and autoimmunity^{68–70}, explaining its upregulation in BD where antigen presentation processes are heightened during inflammation. Arginine Vasopressin Receptor 1 A (AVPR1 A), an arginine vasopressin receptor, is mediated by G proteins which activate a phosphatidylinositol-calcium second messenger system. It has been reported that in some immune-related responses, AVPR1 A is most commonly down-regulated⁷¹, suggesting that the very receptor may impact the metabolism and immune regulation.

To evaluate each biomarker's specificity and sensitivity of diagnosis, we established the ROC curves and through calculating AUC and 95% CI of each curve, we discovered that all selected genes possessed a high diagnostic value for BD and the combination of NRTN, PDIA2 and AVPR1 A increased the prediction efficiency, which guaranteed biomarkers' value of clinical application. Moreover, we also explored the diagnostic value of immune cells with differential proportions in BD. Here, the monocytes showed highest AUC of 0.833, and the prediction ability was further optimized after combining with TRDJ1 and AVPR1 A, two genes with significant correlation with monocytes. These results revealed the diagnostic significance of identified biomarkers and monocyte, which provided new perspective for BD diagnosis in clinical practice. Noticeably, although our training and validation datasets are derived from different ethnic groups, we still obtained identical results, indicating that our findings are universally applicable across different ethnics.

To validate the results in experimental level, the BD mice model was constructed via injecting α -tropomyosin^{45,72}. α -tropomyosin is a self-antigen in BD autoimmune progress, which has been used for BD model construction⁷³. α -tropomyosin has been identified as a potential self-antigen in BD through several clinical investigations, with studies reporting enhanced lymphoproliferative responses and elevated autoantibody titers



against α -tropomyosin in BD patients⁴⁵. Immunoblot analyses using BD patient sera have revealed antibodies targeting α -tropomyosin in tissues commonly affected in BD⁷⁴. The autoimmunity against α -tropomyosin appears pathogenic, as demonstrated in animal models where immunization with this protein induces BD-like symptoms. The model can reproduce key clinical, histological, and molecular features of human BD, including characteristic cytokine profiles and inflammatory responses, validates its utility for studying disease mechanisms and evaluating potential therapeutic interventions. qRT-PCR experiment validated the expression levels of NRTN, PDIA2 and AVPR1 A, which significantly differentially expressed in BD. The expression levels of 5 diagnostic genes were explored in immune-related diseases including sepsis and arthritis, as well as inflammatory bowel diseases involving UC and CD. The secretion profiles of key inflammatory cytokines in human samples and the colon tissues of BD model mice show certain similarities, indicating that the model partially reflects the pathological features of BD. Based on this, we propose that the animal model used provides meaningful insights into the immune mechanisms of BD and lays a foundation for future drug development and target validation.

The expression trends of TRDJ1, IGLV4-69, NRTN and PDIA2 in different datasets were in correspondence with the trends in BD, which further supported our analysis results. AVPR1 A in BD showed similar expression pattern with inflammatory bowel disease, which might be attributed to the similarity between inflammatory bowel disease and BD bowel complications. Inflammatory bowel disease was demonstrated as a critical risk factor of colorectal cancer clinically, and the AVPR1 A has been reported to downregulate in colorectal cancer^{75,76}. The

◀**Fig. 6.** BD mouse model characterization and gene expression analysis. **(A)** Schematic timeline of α -tropomyosin-induced BD mouse model development, showing injection schedule and experimental endpoints. **(B)** Line graph showing body weight ratio of BD model mice relative to healthy control mice over 28 days post-induction, with significant weight loss in BD mice starting at day 14. **(C)** Representative H&E staining images (200 \times magnification) and corresponding Histological Activity Index (HAI) scores of colorectal tissue and joints from normal and BD model mice. HAI scores range from 0 (healthy) to 8 (severe colitis), with BD mice showing significantly higher scores. **(D)** Safranin O-Fast Green-stained joint Sect. (200 \times magnification) from control and BD mice, showing thinner cartilage with unclear boundaries in BD model. **(E)** Immunohistochemical detection of TNF expression (brown staining, 400 \times magnification) in joint tissues of control versus BD mice, with quantification of positive cell percentage. **(F)** Immunohistochemical visualization of TNF expression in colon tissues of control versus BD mice (400 \times magnification), with significantly increased TNF-positive cells in inflamed regions. **(G)** qRT-PCR analysis of pro-inflammatory cytokines Tnf, Il-1 β , Il-6, and Tgfb1 in blood samples from control and BD mice, showing significant upregulation in BD model. **(H)** Heatmap showing expression patterns of key inflammation factors in BD patients versus healthy controls from dataset GSE205967. **(I)** qRT-PCR analysis of Nlrp3 expression in control versus BD mice. **(J)** Immunohistochemical detection of Nlrp3 protein in colon tissues of control versus BD mice (400 \times magnification). **(K)** Expression levels of NLRP3-related inflammatory genes in human BD dataset GSE198533. **(L)** qRT-PCR validation of candidate diagnostic genes AVPR1 A, NRTN, and PDIA2 in blood samples from control and BD model mice. **(M)** Expression levels of the five diagnostic genes in human BD dataset GSE198533. Statistical significance: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

Gene	Sequence (5' to 3')
Avpr1a -F	ATTCGCCAAGGATGACTCGG
Avpr1a -R	TGGATATGCGGCTCAAGTGG
Pdia2 -F	CGCTTCCTCGTCATACATAGCA
Pdia2 -R	AGCAGGTGGTTGAGGATCTTG
Nrtn -F	TCTGTCTGGATGTGCCAGGA
Nrtn -R	GAAAGTTCTCGAAGCTCCACC
Il-6 -F	TACCACTTCACAAGTCGGAGGC
Il-6 -R	CTGCAAGTGATCATCGTTGTTC
mTnf-alpha -F	CCACCATCAAGGACTCAA
mTnf-alpha -R	ATCTTATCCAGCCTCATTCT
Il1b	QM03422S
Nlrp3	QM55990S
Tgfb1	QM05438S
Tnf-antibody	ER65189, Huabio
Nlrp3-antibody	ab263899
Safranin O and Fast Green Cartilage Staining	C0621S

Table 3. Sequences of primers and antibodies for detected genes.

dysregulation of AVPR1 A in BD, especially in bowel complications might indicate the risk of intestinal cancer. In another aspect, the different expression patterns of diagnostic genes might help distinguish inflammatory bowel disease and BD, which improved the diagnostic accuracy in clinical use.

However, limitations in research still exist. Though we have performed analysis on bulk-seq level and single-cell seq level, more external datasets are needed for further validation of results we have obtained. Also, in vivo experiments could provide stronger evidence to support our analysis results. Though bioinformatics analysis provides a new perspective on the internal mechanism of BD, the external validation and experiments are still significant.

Conclusion

This study provides compelling evidence for the central role of monocytes in Behçet’s Disease pathogenesis. Through comprehensive integration of bulk RNA sequencing and single-cell analyses, we demonstrated that monocytes represent the earliest activated cell population in BD, exhibiting enhanced stemness characteristics and extensive communication with other immune cells. Our integrative bioinformatics approach identified five novel diagnostic biomarkers (NRTN, TRDJ1, IGLV4-69, PDIA2, and AVPR1 A) with high sensitivity and specificity for BD detection. Particularly, AVPR1 A expression combined with monocyte proportion assessment demonstrated exceptional diagnostic accuracy. Experimental validation in our α -tropomyosin-induced BD mouse model recapitulated key inflammatory features of human BD and confirmed the dysregulated expression of these biomarkers. These findings significantly advance our understanding of BD immunopathology by establishing monocytes as primary drivers of disease initiation and progression. The identified biomarkers offer promising tools for precision diagnostics, while the elucidated signaling pathways highlight potential targets for

developing novel immune-modulating therapies. Future studies should focus on validating these biomarkers in larger clinical cohorts and exploring therapeutic strategies targeting monocyte-mediated inflammatory cascades in BD.

Data availability

The datasets selected in our research can be found and downloaded for free online, while the databases we searched and accession number(s) are all provided in article or supplement materials. GEO Database accession number: bulk-seq: GSE198533, GSE17114, GSE209567, GSE69063, GSE205962, GSE66407; Non-coding: GSE145191; single-cell: GSE198616. The usage of KEGG image was permitted by Kanehisa Lab, Ref: 251500.

Received: 24 February 2025; Accepted: 30 May 2025

Published online: 06 June 2025

References

1. Davatchi, F. et al. Behçet's disease: epidemiology, clinical manifestations, and diagnosis. *Expert Rev. Clin. Immunol.* **13** (1), 57–65 (2017).
2. Dalvi, S. R., Yildirim, R. & Yazici, Y. Behçet's syndrome. *Drugs* **72** (17), 2223–2241 (2012).
3. Lavalle, S. et al. Behçet's disease, pathogenesis, clinical features, and treatment approaches: A comprehensive review. *Medicina* **60** (4), 562 (2024).
4. Yazici, Y. et al. Behçet syndrome. *Nat. Reviews Disease Primers.* **7** (1), 67 (2021).
5. Bolek, E. C. et al. Clinical features and disease course of neurological involvement in behçet's disease: HUVAC experience. *Multiple Scler. Relat. Disorders.* **38**, 101512 (2020).
6. Watts, R. A., Hatemi, G., Burns, J. C. & Mohammad, A. J. Global epidemiology of vasculitis. *Nat. Rev. Rheumatol.* **18** (1), 22–34 (2022).
7. Adams, T. N., Zhang, D., Batra, K. & Fitzgerald, J. E. Pulmonary manifestations of large, medium, and variable vessel vasculitis. *Respir. Med.* **145**, 182–191 (2018).
8. Vegting, Y. et al. Monocytes and macrophages in ANCA-associated vasculitis. *Autoimmun. Rev.* **20** (10), 102911 (2021).
9. Gerhardt, T. & Ley, K. Monocyte trafficking across the vessel wall. *Cardiovascular. Res.* **107** (3), 321–330 (2015).
10. Borish, L. C. & Steinke, J. W. 2. Cytokines and chemokines. *J. Allergy Clin. Immunol.* **111** (2), S460–S475 (2003).
11. The International Criteria for Behçet's Disease (ICBD). A collaborative study of 27 countries on the sensitivity and specificity of the new criteria. *J. Eur. Acad. Dermatology Venereology: JEADV.* **28** (3), 338–347 (2014).
12. Kiafar, M. et al. Diagnosis of behçet's disease: clinical characteristics, diagnostic criteria, and differential diagnoses. *BMC Rheumatol.* **5** (1), 2 (2021).
13. Hatemi, I., Hatemi, G. & Çelik, A. F. Gastrointestinal involvement in behçet disease. *Rheum. Dis. Clin. North Am.* **44** (1), 45–64 (2018).
14. Zhou, J. et al. The clinical features, image findings and risk factors of Vena Cava syndrome in behçet's syndrome. *Clin. Exp. Rheumatol.* **40** (8), 1526–1534 (2022).
15. Giannessi, C. et al. Behçet's Disease: A Radiological Review of Vascular and Parenchymal Pulmonary Involvement. *Diagnostics (Basel Switzerland)* **12**(11). (2022).
16. Amin, M. M. & Abdel Latif, O. M. Anti-tubulin-alpha-1c antibody as a marker of value in behçet syndrome. *Clin. Rheumatol.* **41** (6), 1759–1767 (2022).
17. Zheng, W. et al. Single-cell analyses highlight the Proinflammatory contribution of C1q-high monocytes to behçet's disease. *Proc. Natl. Acad. Sci. U.S.A.* **119** (26), e2204289119 (2022).
18. Emmi, G. et al. A unique Circulating MiRNA profile highlights thrombo-inflammation in behçet's syndrome. *Ann. Rheum. Dis.* **81** (3), 386–397 (2022).
19. Bhattacharya, S. et al. ImmPort: disseminating data to the public for the future of immunology. *Immunol. Res.* **58** (2–3), 234–239 (2014).
20. Ding, Z. et al. Integrative network fusion-based multi-omics study for biomarker identification and patient classification of rheumatoid arthritis. *Chin. Med.* **18** (1), 48 (2023).
21. Oğuz, A. K., Oygür, Ç., Taşır, S., Özdağ, H. & Akar, M. N. Behçet syndrome: the disturbed balance between anti- (CLEC12A, CLC) and Proinflammatory (IFI27) gene expressions. *Immun. Inflamm. Dis.* **11** (4), e836 (2023).
22. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods.* **12** (5), 453–457 (2015).
23. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184** (13), 3573–3587e3529 (2021).
24. Zheng, W. et al. Single-cell analyses highlight the proinflammatory contribution of C1q-high monocytes to Behçet's disease. *Proceedings of the National Academy of Sciences* 119(26):e2204289119. (2022).
25. Nicol, P. B. et al. Robust identification of perturbed cell types in single-cell RNA-seq data. *Nat. Commun.* **15** (1), 7610 (2024).
26. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32** (4), 381–386 (2014).
27. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genom.* **19**, 1–16 (2018).
28. Sun, D. et al. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. *Nat. Biotechnol.* **40** (4), 527–538 (2022).
29. Jin, S. et al. Inference and analysis of cell-cell communication using cellchat. *Nat. Commun.* **12** (1), 1088 (2021).
30. Love, M. I., Huber, W. & Anders, S. Moderated Estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15** (12), 550 (2014).
31. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9** (1), 559 (2008).
32. The Gene Ontology Resource. 20 years and still going strong. *Nucleic Acids Res.* **47** (D1), D330–d338 (2019).
33. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28** (1), 27–30 (2000).
34. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53** (D1), D672–D677 (2025).
35. Kanehisa, M. Toward Understanding the origin and evolution of cellular organisms. *Protein Sci.* **28** (11), 1947–1951 (2019).
36. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J. Integr. Biology.* **16** (5), 284–287 (2012).
37. Ameres, S. L. & Zamore, P. D. Diversifying MicroRNA sequence and function. *Nat. Rev. Mol. Cell Biol.* **14** (8), 475–488 (2013).
38. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**, (2015).
39. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13** (11), 2498–2504 (2003).

40. Xu, J., Zhou, H., Cheng, Y. & Xiang, G. Identifying potential signatures for atherosclerosis in the context of predictive, preventive, and personalized medicine using integrative bioinformatics approaches and machine-learning strategies. *EPMA J.* **13** (3), 433–449 (2022).
41. Hastie, T. & Qian, J. Glmnet vignette. Retrieved June. **9** (2016), 1–30 (2014).
42. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **12** (1), 77 (2011).
43. Smyth, G. K. Limma: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer; : 397–420. (2005).
44. Goldberg, I., Baharav, E., Weinberger, A. & Krause, I. Animal models of behçet syndrome. *Behçet Syndrome* **2020**:235–242.
45. Mahesh, S. et al. Alpha Tropomyosin as a self-antigen in patients with behçet's disease. *Clin. Experimental Immunol.* **140** (2), 368–375 (2005).
46. Boutsikou, E. et al. Tumour necrosis factor, interferon-gamma and interleukins as predictive markers of antiprogrammed cell-death protein-1 treatment in advanced non-small cell lung cancer: a pragmatic approach in clinical practice. *Therapeutic Adv. Med. Oncol.* **10**, 1758835918768238 (2018).
47. Song, C. et al. Alterations in the immune landscape characterized by inflammatory activation and immune escape within 12 h after trauma. *Immunobiology* **229** (3), 152801 (2024).
48. Han, Z., Zhang, J., Guo, H. & Zheng, G. The Genes Associated with Rheumatoid Arthritis Patients with Blood-Stasis Syndrome. In: *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM): 2023*; IEEE; : 4741–4746. (2023).
49. Ayididaer, A. Comparative analysis of Over-Representation analysis and gene set enrichment analysis methods in KEGG pathway enrichment: insights from inflammatory bowel disease studies. In.; (2024).
50. Hu, D. & Guan, J.-L. The roles of immune cells in behçet's disease. *Adv. Rheumatol.* **63**, 49 (2023).
51. Hu, K. Become competent within one day in generating boxplots and violin plots for a novice without prior R experience. *Methods Protocols* **3**(4). (2020).
52. Sugimura, N. et al. Real-world efficacy of adalimumab and Infliximab for refractory intestinal behçet's disease. *Dig. Liver Disease: Official J. Italian Soc. Gastroenterol. Italian Association Study Liver.* **51** 7, 967–971 (2019).
53. Zhong, Z., Su, G. & Yang, P. Risk factors, clinical features and treatment of Behçet's disease uveitis. *Progress Retinal Eye Research* :101216. (2023).
54. Shi, C. & Pamer, E. G. Monocyte recruitment during infection and inflammation. *Nat. Rev. Immunol.* **11** (11), 762–774 (2011).
55. Zhang, Y. et al. B cell-activating factor and its targeted therapy in autoimmune diseases. *Cytokine Growth Factor Rev.* **64**, 57–70 (2022).
56. Jiang, C., Xie, S., Yang, G. & Wang, N. Spotlight on NLRP3 inflammasome: role in pathogenesis and therapies of atherosclerosis. *Journal Inflamm. Research* **2021**:7143–7172.
57. Nojima, S. Class IV semaphorins in disease pathogenesis. *Pathol. Int.* **72** (10), 471–487 (2022).
58. Lan-min, W. & Dong, S. Clinical significance of ICAM-1, VCAM-1 and PECAM-1 in vascular endothelial injury of end-stage renal disease patients. *J. Clin. Nephrol.* **21** (8), 643–647 (2021).
59. Tong, B., Liu, X., Xiao, J. & Su, G. Immunopathogenesis of behçet's disease. *Front. Immunol.* **10**, 665 (2019).
60. Liu, P., Li, X., Lv, W. & Xu, Z. Inhibition of CXCL1-CXCR2 axis ameliorates cisplatin-induced acute kidney injury by mediating inflammatory response. *Biomed. Pharmacother.* **122**, 109693 (2020).
61. Neves, F. S. & Spiller, F. Possible mechanisms of neutrophil activation in behçet's disease. *Int. Immunopharmacol.* **17** (4), 1206–1210 (2013).
62. Perazzio, S. F., Andrade, L. E. C. & de Souza, A. W. S. Understanding behçet's disease in the context of innate immunity activation. *Front. Immunol.* **11**, 586558 (2020).
63. Morel, L., Domingues, O., Zimmer, J. & Michel, T. Revisiting the role of neurotrophic factors in inflammation. *Cells* **9**(4). (2020).
64. Ravens, S. et al. Microbial exposure drives polyclonal expansion of innate $\gamma\delta$ T cells immediately after birth. *Proc. Natl. Acad. Sci. U.S.A.* **117** (31), 18649–18660 (2020).
65. Lefranc, M. P. Immunoglobulin and T cell receptor genes: IMGT(®) and the birth and rise of immunoinformatics. *Front. Immunol.* **5**, 22 (2014).
66. André, S. et al. Highly restricted diversity of TCR delta chains of the amphibian Mexican Axolotl (*Ambystoma mexicanum*) in peripheral tissues. *European J. Immunology* **37**. (2007).
67. Tardif, G. et al. Mass spectrometry-based proteomics identify novel serum osteoarthritis biomarkers. *Arthritis Res. Therapy.* **24** (1), 120 (2022).
68. Walker, A. K., Soo, K. Y., Levina, V., Talbo, G. H. & Atkin, J. D. N-linked glycosylation modulates dimerization of protein disulfide isomerase family A member 2 (PDIA2). *FEBS J.* **280** (1), 233–243 (2013).
69. Pay, S., Simşek, I., Erdem, H. & Dinç, A. Immunopathogenesis of behçet's disease with special emphasize on the possible role of antigen presenting cells. *Rheumatol. Int.* **27** (5), 417–424 (2007).
70. Wooten, E. C. et al. Application of gene network analysis techniques identifies AXIN1/PDIA2 and endoglin haplotypes associated with bicuspid aortic valve. *PLoS One.* **5** (1), e8830 (2010).
71. Ripoll, G. V., Pifano, M., Garona, J. & Alonso, D. F. Commentary: arginine vasopressin receptor 1a is a therapeutic target for castration-resistant prostate cancer. *Frontiers Oncology* **9**. (2020).
72. Yüceer, H. & Tüzün, E. Animal Models of Behçet's Disease. In: *Neuro-Behçet's Disease: Pathogenesis, Clinical Aspects, Treatment*. Edited by Tüzün E, Kürtüncü M. Cham: Springer International Publishing; : 41–48. (2021).
73. Piao, Y.-Z. *Effect of Sanwu Huangqin Decoction on IFN- γ and IL-17A levels in mice with Behçet's disease* (Editorial Board of Journal of Hainan Medical University, 2020).
74. Mor, F., Weinberger, A. & Cohen, I. R. Identification of alpha-tropomyosin as a target self-antigen in behçet's syndrome. *Eur. J. Immunol.* **32** (2), 356–365 (2002).
75. Annese, V. et al. European Evidence-based consensus: inflammatory bowel disease and malignancies. *J. Crohn's Colitis.* **9** (11), 945–965 (2015).
76. Centrone, M. et al. dDAVP downregulates the AQP3-Mediated glycerol transport via V1aR in human Colon HCT8 cells. *Front. Cell. Dev. Biol.* **10**, 919438 (2022).

Acknowledgements

Thanks to all participants who have contributed to the study.

Author contributions

All the authors drafted the manuscript and approved its submission. Conceptualization, Qiyi Yu; Methodology, Qiyi Yu; Software, Yidong Zhang; Data Curation, Tianyue Pan; Experiment, Yifan Wu; Writing – Original Draft Preparation, Meinuo Li; Writing – Review & Editing, Qiyi Yu.

Funding

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

Declarations

Competing interests

The authors declare no competing interests.

Consent for publication

All authors have given consent for publication.

Ethnic approval

All experiments were performed according to the guidelines of the Association for Assessment and Accreditation of Laboratory Animals Care International. All of the procedures were carried out strictly in accordance with the animal care and use protocol (No: LL-202401220001) approved by the Institutional Animal Care and Use Committee (IACUC) at Jiangxi Zhonghongboyuan Institute.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-05086-x>.

Correspondence and requests for materials should be addressed to Q.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025