Identification of Biomarkers for Inflammatory Bowel Disease Based on Single-Cell RNA Sequencing Data

Wenbo Qin[1,2], Xuezhen Chen[2], Jinbo Zhang[2], Xiao Sun[1,*], Jue Fan[2,*]

[1]School of Biological Science and Medical Engineering, Southeast University, 211189 Nanjing, China
[2]Singleron Biotech Co., Ltd., 211189 Nanjing, China

Abstract: Inflammatory bowel disease (IBD) is a chronic gastrointestinal disease that imposes a severe health burden globally. Single-cell RNA sequencing (scRNA-seq) technology helps to elucidate the molecular characteristics and functional states of individual cells, thereby advancing the study of IBD's intricate mechanisms. Most current studies focus on small-scale samples from a few cell subclusters, lacking large-scale, systematic analyses of IBD. By collecting and integrating multiple datasets, large-scale sample datasets were analyzed for cell type identification and reclassification, addressing this gap. In this study, we collected and merged several public IBD-related scRNA-seq datasets, creating a dataset with millions of cells. Using data mining tools and bioinformatics techniques, we have identified Treg cells, Th17 cells, the macrophage C12 subcluster, and the fibroblast C5 subcluster as being significantly associated with IBD. These cell types play crucial roles in inflammation, fibrosis, and immune regulation. Differential gene expression analysis revealed several potential biomarkers, including IL2RA, HSF1, and TNFSF8 in Treg cells; PIM3, RIPK2, and TLR2 in macrophage C12 subcluster; and CDH11, OSMR, and BRD4 in fibroblast C5 subcluster. These biomarkers could serve as potential therapeutic targets, contributing to a deeper understanding and more effective treatment of IBD.

1 Introduction

IBD is a complex chronic intestinal disease, mainly classified into ulcerative colitis (UC) and Crohn's disease (CD)[1], both of which impose an immense global health burden[2]. The most used treatment in clinical practice is monoclonal antibodies targeting tumor necrosis factor (TNF), but this approach is not universally effective for all IBD patients[3]. Although recent studies indicate that gut microbiome significantly influences IBD[4], treatment strategies that can be widely applied to all IBD patients are still lacking due to the variability in microbiota composition across different individuals. Moreover, the intricate interplay between genetic background and environmental factors plays a significant role in contributing to the heterogeneity of the disease, making it essential to understand the complexity of IBD, where exploring key cells and target genes as biomarkers is pivotal to understanding its heterogeneity and complexity.

With scRNA-seq technology, researchers can explore the dynamic networks between cells and gene expression changes, offering a unique perspective for revealing the cellular composition and molecular mechanisms of IBD-associated tissues[5]. Previous studies using scRNA-seq have mapped the single-cell profiles of IBD-related tissues[6] and identified differentially expressed risk-related genes under disease conditions[7]. Parikh et al.[8] used scRNA-seq to analyze colon epithelial cells in IBD patients, revealing that multiple cell subtypes and populations in inflamed tissues undergo remodeling and dysregulation, and identified 1147 genes with dysregulated expression in ulcerative colitis. Kinchen et al.[9] further elucidated the presence of mesenchymal cell populations near epithelial crypts associated with colon epithelial function. Dysfunction of these cells leads to epithelial damage and drives inflammatory states. These findings suggest that the pathogenesis of IBD is the result of multifactorial interactions involving complex cellular networks and molecular mechanisms. However, most studies currently focus on specific cell populations and have yet to provide a systematic analysis of IBD. For example, Uzzan et al.[10] applied scRNA-seq to map the composition, transcriptional features, and clonal profiles of mucosal and circulating B cells in ulcerative colitis, revealing how B cells promote inflammation through immunoglobulin production during the disease.

With the rapid advancement of scRNA-seq technology, a large volume of single-cell data has been accumulated, providing rich resources for the systematic study of IBD. Krzak et al.[11] constructed a large scRNA-seq dataset that included 50 CD samples and 71 healthy controls, comprising over 440,000 cells. Their transcriptomic analysis revealed affected genes, cell types, and mechanisms in CD. However, single datasets often face limitations in terms of experimental design, technical platforms, and sample sources, which may

* Xiao Sun: xsun@seu.edu.cn, *Jue Fan: fanjue@singleronbio.com

hinder the generalizability of findings to broader contexts. The application of scRNA-seq has not only enabled the identification of rare cell types but also provided technical support for the integrative, systemic analysis of large datasets. Therefore, cross-dataset studies are now feasible and crucial for gaining deeper insights into biological principles and enhancing the universality of research outcomes.

Given the increasing demand for molecular-level studies on pathogenic cells and target genes in IBD, recent studies that combine genome-wide association studies (GWAS) with scRNA-seq are becoming mainstream[12]. GWAS analyzes the correlation between genotype and phenotype in large populations, identifying candidate genes or genomic regions associated with diseases[13]. However, GWAS does not directly reveal how these genes act in specific cell types or microenvironments. By leveraging scRNA-seq's precise expression data at single-cell resolution, researchers can delve into the cell-specific functions of candidate genes, thus more accurately identifying risk genes in key cell types. The Single-Cell Disease Relevance Score (scDRS)[14] is a tool that combines scRNA-seq with GWAS to score each cell's enrichment for the gene modules related to a specific disease. This tool helps identify critical cell types or pathways involved in the disease. By integrating the strengths of both techniques, scDRS provides valuable clues for further target gene discovery and precision treatment strategies.

This study aims to identify key disease-associated cell types in IBD and analyze the target genes within these cell types using the scDRS tool by integrating large-scale public datasets. The strategy of this study is outlined in Figure 1. First, multiple scRNA-seq datasets related to IBD were collected from authoritative public databases and integrated to form high-quality datasets. Then, cell subtypes were clustered to identify major cell types. Next, GWAS data were collected, and IBD risk genes were extracted. Using the scDRS tool, disease relevance scores were calculated based on the GWAS data, identifying the cell types most associated with IBD. Finally, differential gene expression analysis was performed on key cell types, and risk genes were selected based on clinical progression and other standards to identify IBD-associated risk genes in these critical cell types.

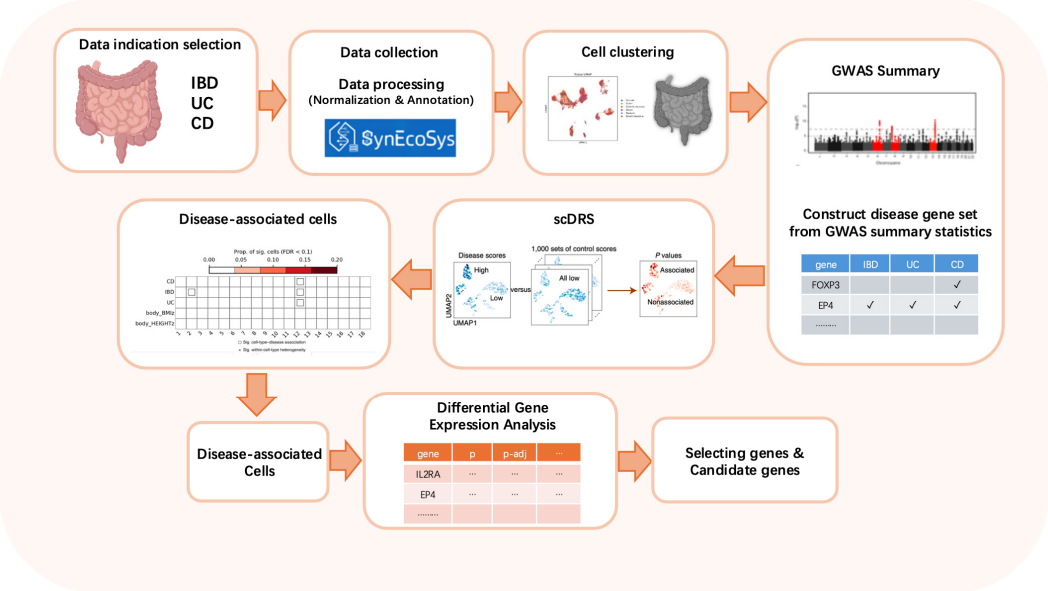

Figure 1. The Overall Computational and Analytical Framework of This Study.

2 Methods

2.1 Data Collection and Integration

A total of 14 IBD-related datasets, including SCP1884 and GSE144469, were downloaded from the Single Cell Portal (SCP) database[15] and the GEO database[16]. The collected scRNA-seq datasets were integrated using *scanpy v1.8.1*[17] to construct a unified large-scale single-cell dataset. Quality control (QC) was performed for each dataset according to the following criteria: 1) Cells with fewer than 200 detected genes or those in the bottom 2% of gene counts were excluded; 2) Cells with unique molecular identifier (UMI) counts in the top 2% were removed; 3) Cells with mitochondrial gene expression exceeding 20% and genes expressed in fewer than 5 cells were also excluded. After QC, each dataset was normalized, and the top 2000 highly variable genes were selected as feature genes for subsequent analyses. Dimensionality reduction and clustering were performed to facilitate cell annotation. A KNN graph was constructed in the reduced principal component space to identify high-dimensional proximity relationships between cells, followed by clustering algorithms to categorize the cells. Known marker gene expression information and dataset backgrounds were used to annotate the clustered cell types. For data integration, an anchor-based strategy leveraging shared variance

features was employed, and different datasets were integrated by identifying cross-dataset anchors to generate a consistent expression matrix. This study integrated multiple datasets using the anndata [17] toolkit and applied the Harmony algorithm within scanpy to correct for batch effects.

2.2 Data Preprocessing

To ensure data stability, normalization and standardization of the integrated dataset were conducted, followed by re-selection of highly variable genes. Since data integration may cause shifts in the distribution of cell types, further dimensionality reduction and clustering were applied. The clustering results were visualized using Uniform Manifold Approximation and Projection (UMAP)[18] to capture the high-dimensional relationships between cells. Finally, cell annotation was performed based on known marker gene expression and the original dataset information. The resulting dataset comprehensively covered the pathogenic tissues and potential disease-associated cell clusters in IBD, providing critical support for the study of IBD mechanisms and data analysis.

2.3 Cell-Disease Relevance Scores

The Single-Cell Disease Relevance Score (scDRS)[14] is a method for integrating scRNA-seq data with genome-wide association study (GWAS) data. scDRS evaluates the association of each cell type with a disease by performing a weighted summation of the expression of IBD risk genes identified in GWAS studies for each cell and comparing this with the weighted expression of a random gene set. This approach enables the precise identification of cell types and subclusters significantly associated with the disease at single-cell resolution, while also revealing cellular heterogeneity. First, the top 1000 IBD-related genes were selected from the GWAS data. For each cell, the expression of these genes was weighted and summed, and a disease score for each cell was calculated based on the formula. A weighted expression of random gene sets was used to compute the control score for each cell, which was then compared with the disease score. Both the disease and control scores were normalized to eliminate systematic bias. Finally, the normalized disease scores for each cell were compared with the empirical distribution of all control scores, and the single-cell level p-value for disease association was computed. A heatmap representing the association of each cell type with IBD was generated based on the disease scores and p-values.

2.4 Cell Proportion Analysis

To reveal the cell types or subclusters that undergo significant changes during the disease process, cell proportion analysis was conducted by dividing the samples into disease and healthy control groups. The average and median proportions of each cell type were calculated for both groups, followed by differential analysis to assess changes in cell proportions between the disease and control groups. The results of the differential cell proportion analysis were visualized to show the dynamic changes in cell types across the disease process, highlighting those cell types with significant changes. Statistical tests such as the Wilcoxon rank-sum test were used to calculate the p-values for each cell type, ensuring the reliability and scientific rigor of the analysis.

2.5 Differential Gene Expression Analysis

To identify differentially expressed genes, we conducted differential expression analysis across different cell types. The gene selection criteria were as follows: at least 10% of cells in each comparison group must express the gene, and the average log-fold change (log FC) should exceed 1. To control for multiple comparisons, we applied the Benjamini-Hochberg correction method to calculate adjusted p-values, with a significance threshold set at 0.05. Through this filtering process, we identified multiple sets of differentially expressed genes, which exhibited significant expression changes between the comparison groups and provided a reliable set of candidate genes for further analysis of functional differences among cell types.

3 Results and Discussion

The processed dataset comprised a total of 1,167,689 cells from 174 individuals, with the major sampling tissues including the colon, ileum, rectum, and cecum. Among these, 110 were healthy samples, and 235 were disease samples, which included 114 UC and 121 CD samples. The dataset identified 22 cell clusters, further subdivided into 63 cell subclusters, encompassing three major cell types: epithelial cells, stromal cells, and immune cells. Additionally, the dataset included important metadata associated with patients, such as treatment status, medication types, gender, smoking history, and tissue origins.

3.1 Identification of Key IBD Cell Types

Using the scDRS tool, cell types in the dataset were scored and visualized. The results are shown in Figure 2, where the vertical axis represents disease types, and the horizontal axis represents cell types. Cells significantly associated with a disease type are marked with boxes at the intersections. In UC, enterocytes, platelets, regulatory T cells (Tregs), and T-helper 17 cells (Th17 cells) exhibited significant associations. In CD, a broader range of cell types showed significant associations. Further analysis revealed that Tregs and Th17 cells were significantly correlated with UC, CD, and IBD. Consequently, subsequent research focused on these cell types. Although macrophages and fibroblasts were not significantly highlighted in the overall scoring results, extensive research has demonstrated their critical roles in IBD[19]. Macrophages can be classified further into pro-inflammatory and anti-inflammatory subtypes[20].

While macrophages and fibroblasts as overarching categories lack a strong association with IBD, their subclusters may be disease-relevant. Therefore, this study delves deeper into the subclusters of macrophages and fibroblasts to precisely identify key IBD-related subclusters.

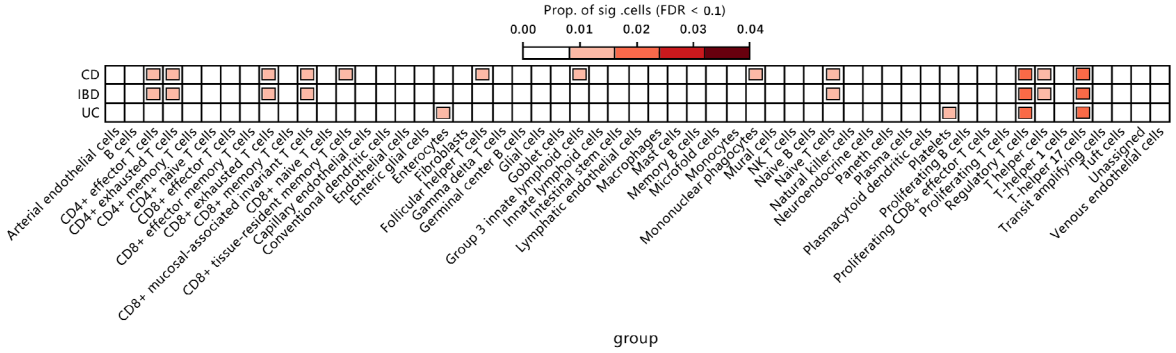

Figure 2. Disease Relevance Scores (scDRS) for Various Cell Types Associated with IBD. The color gradient at the top indicates the proportion of disease-associated cells, with darker colors representing a higher proportion of the total cells. Boxes indicate cell types associated with the disease.

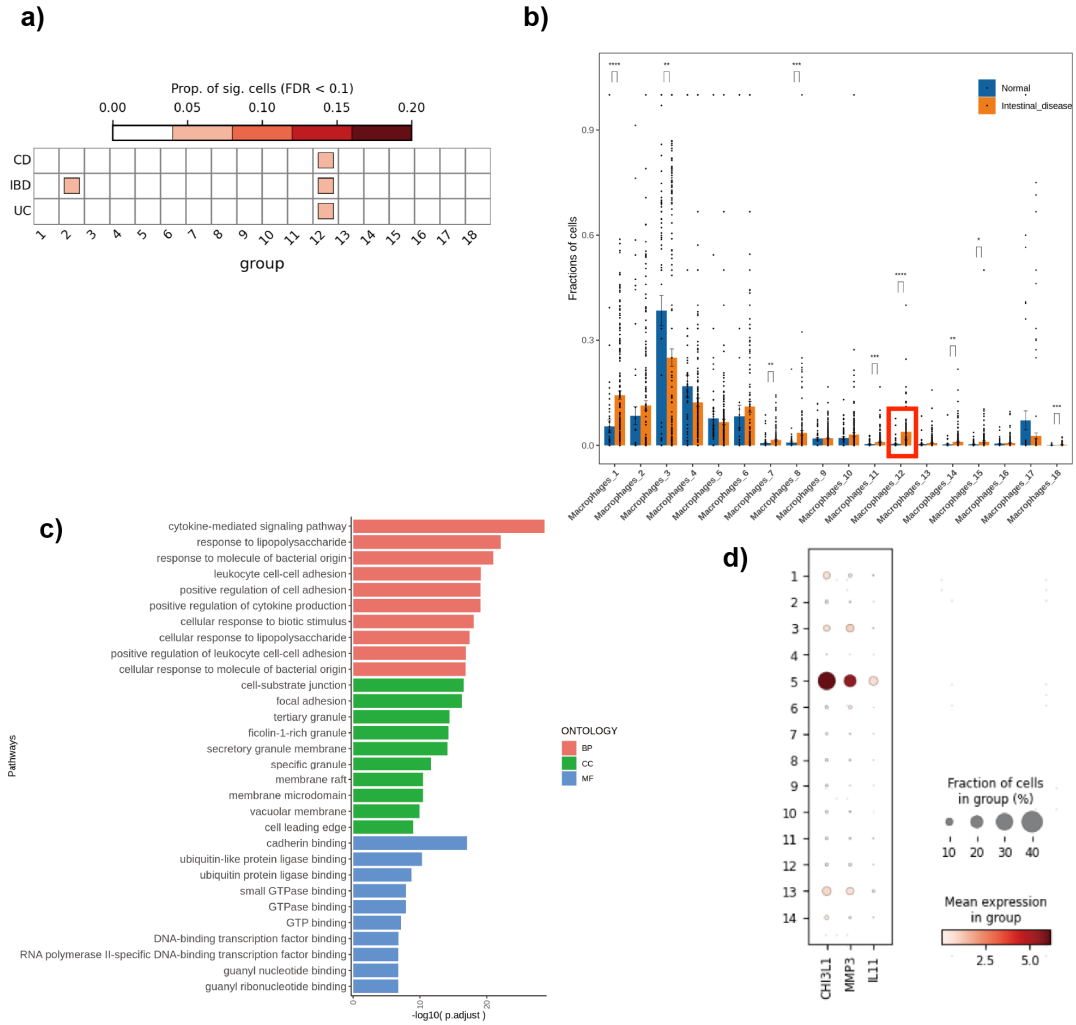

Figure 3. Disease Relevance Scores and Proportion Analysis of Macrophage Subclusters. (a) scDRS scores for macrophage subclusters. (b) Proportion analysis of macrophage subclusters, showing significant differences in the C12 subcluster between the disease and healthy groups. (c) Enrichment analysis of characteristic genes in the macrophage C12 subcluster. (d) Expression of characteristic genes of three types of inflammatory fibroblasts across fibroblast subclusters.

3.2 Identification of Key Subclusters in Macrophages and Fibroblasts

To explore the association of macrophage and fibroblast subclusters with IBD, these two cell types were isolated from the dataset and subjected to standardized preprocessing. Mechanical clustering was employed to divide these cells into subclusters, which were then analyzed for IBD disease relevance using scDRS scoring.

Macrophages were divided into 18 subclusters via mechanical clustering. Among them, the C12 subcluster showed significant associations with IBD, UC, and CD (as indicated by scDRS scores, Figure 3a). Proportion analysis of all macrophage subclusters further revealed a significant increase in the C12 subcluster in the disease group (Figure 3b). Functional analysis and pathway enrichment of upregulated characteristic genes in the C12 subcluster indicated significant enrichment in immune-regulatory pathways such as lymphocyte adhesion and cytokine secretion (Figure 3c). These findings suggest that the C12 subcluster of macrophages is closely related to disease status, warranting further exploration.

Fibroblasts were similarly divided into 14 subclusters. scDRS analysis revealed that the C8 subcluster was significantly associated with IBD, UC, and CD, but no significant difference in its proportion was observed between disease and healthy groups. Since GWAS often focuses on disease-related phenotypes such as inflammation markers and immune dysregulation, non-immune functions of fibroblasts may not be directly captured. However, inflammatory fibroblast subclusters, known for their roles in tissue repair and fibrosis, have been identified, characterized by specific markers such as CHI3L1, MMP3, and IL-11[21]. Analysis of these biomarkers in fibroblast subclusters revealed that the C5 subcluster exhibited high expression of these inflammatory markers in the disease group (Figure 3d). Proportion analysis also showed a higher proportion of the C5 subcluster in the disease group compared to controls. These results indicate that the C5 subcluster of fibroblasts is closely associated with disease status, and further research will focus on this subcluster.

3.3 Identification of IBD-Related Target Genes in Key Cell Types

Using scDRS and other analyses, this study identified key cell types and subclusters significantly associated with IBD. Tregs and Th17 cells demonstrated strong correlations with UC, CD, and IBD, indicating their potential roles in regulating intestinal inflammation and immune balance. Additionally, the macrophage C12 subcluster exhibited significant associations with IBD, with functional enrichment analyses showing involvement in immune-regulatory pathways such as lymphocyte adhesion and cytokine secretion. Although GWAS data failed to capture fibrosis-related characteristics in fibroblasts, clustering studies identified the C5 subcluster of fibroblasts as representative of inflammatory fibroblasts, showing high expression of inflammation-related markers such as CHI3L1, MMP3, and IL-11. Based on the above findings, this study will focus on Tregs, Th17 cells, the macrophage C12 subcluster, and the fibroblast C5 subcluster, conducting an in-depth analysis of the potential biomarkers of these key cell types and subclusters in IBD.

3.3.1 Candidate Biomarkers in Treg Cells for IBD

Treg cells were extracted from the dataset and processed. Differential gene expression analysis between disease and healthy control groups identified 112 upregulated genes in UC compared to the healthy controls and 877 upregulated genes in CD compared to the healthy controls. Among these, 39 genes were commonly upregulated in both UC and CD. From these 39 genes, those that were both listed as IBD risk genes in GWAS studies and appeared in the differential expression gene list of the IBD Tamma bulk data database[22] were prioritized. Further analysis of the drug development progress for these genes indicated that drug targets already in Phase II clinical trials should be prioritized to shorten drug development timelines. Ultimately, only IL2RA met these criteria. Similarly, the downregulated differential genes were filtered, resulting in the identification of HSF1, TNFSF8, AREG, and LIME1 as potential IBD biomarkers.

In this study, IL2RA was found to be significantly upregulated in T cells from the disease group (Figure 4). As a subunit of the IL-2 receptor, IL2RA plays a crucial role in the proliferation, differentiation, and survival of Tregs. Its upregulation may represent a regulatory mechanism aimed at enhancing Treg function to suppress excessive immune responses in IBD[23]. Based on this, targeted interventions in the IL-2/IL2RA signaling pathway hold potential as a novel therapeutic strategy for IBD, though further clinical validation is required. HSF1 is a key transcription factor primarily regulating the expression of heat shock proteins. It plays vital roles in protein homeostasis, damage repair, and cell survival. Some studies suggest it may promote the progression from colitis to colorectal cancer[24], though its specific functions in IBD require further exploration. TNFSF8 (CD153), a member of the tumor necrosis factor (TNF) family, enhances Treg functionality by binding to its receptor, maintaining immune tolerance, and preventing excessive inflammatory responses from the immune system. This suggests that TNFSF8 may play a regulatory role in IBD[25]. AREG, a growth factor, was found to be significantly upregulated in the healthy control group (Figure 4). While AREG is known for its pro-cancer activity in various tumor tissues, studies in inflammation models have shown that AREG administration can restore epithelial barrier function and alleviate colitis[26]. This indicates a potential role in barrier repair in IBD. LIME1 has been relatively less studied, but it was identified as a differential gene in this study. Although its specific mechanisms in IBD remain unclear, its potential as a target gene warrants further investigation to explore its therapeutic significance.

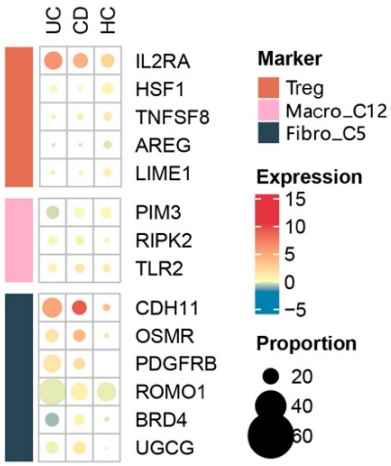

Figure 4. Expression of candidate target genes in key cell types. The figure is divided into three sections: the upper, middle, and lower panels represent Treg cells, macrophage C12 subcluster, and fibroblast C5 subcluster, respectively (Th17 cells were not further explored as they have already been extensively studied). The three columns represent UC, CD, and HC disease conditions. The color of the dots indicates the expression intensity of the gene in the corresponding cell type, while the size of the dot represents the proportion of gene expression.

3.3.2 Candidate Biomarkers in Th17 Cells for IBD

Th17 cells are known to contribute to the pathogenesis of various common autoimmune diseases, including IBD, psoriasis, and rheumatoid arthritis, primarily through the secretion of pro-inflammatory cytokines such as IL-17 and IL-22. In this study, differential gene expression analysis was conducted on Th17 cells between disease and healthy control groups. For CD versus healthy controls, 295 genes were upregulated, and 357 genes were downregulated. Among these, genes that were identified as IBD risk genes in GWAS studies and were also present in the differential expression gene list of the IBD Tamma bulk data database were prioritized. Further investigation into the drug development progress of these genes prioritized targets associated with autoimmune diseases or tissue fibrosis. This process identified six candidate target genes: TNFSF8, AHR, BCL2, IL7R, GATA3, and IL2, all of which were downregulated. Notably, only drugs targeting AHR were agonists, while the other targets were associated with inhibitors. This observation contradicts the well-established pro-inflammatory role of Th17 cells in IBD[27], potentially reflecting sample-specific biases in expression-level analysis. Given the extensive prior research on Th17 cells, which has identified many candidate biomarkers for IBD, this study will not conduct further analysis on this cell type.

3.3.3 Candidate Biomarkers in Macrophage C12 Subcluster for IBD

The macrophage C12 subcluster was identified as an IBD-associated cell subcluster. Differential gene expression analysis between disease and healthy control groups revealed 687 commonly upregulated genes in both UC and CD. Genes that were identified as IBD risk genes in GWAS studies and were also present in the IBD Tamma bulk data database were prioritized. Drug development progress was then investigated, focusing on inhibitory target genes with indications for IBD. This process identified three candidate biomarkers: PIM3, RIPK2, and TLR2. Downregulated genes from the analysis did not meet the clinical progress criteria.

This study investigated these biomarkers and found that PIM3, a member of the PIM kinase family, plays a role in T cell proliferation and differentiation, promoting the production of inflammatory cytokines such as IL-6, IL-17, TNF-α, and IFN-γ by Th1 and Th17 effector cells[28]. Studies have shown that PIM kinase inhibitors can modulate the immune environment and promote the differentiation of T cells into anti-inflammatory Tregs. These inhibitors are currently in preclinical development for various inflammatory diseases, including rheumatoid arthritis and IBD[28]. RIPK2 is a critical downstream molecule in the NOD signaling pathway, regulating NOD1- and NOD2-mediated inflammatory responses. In the affected tissues of IBD patients, RIPK2 expression correlates positively with inflammatory cytokines such as TNF-α and IL-6[29]. Experimental data indicate that RIPK2 inhibitors, such as GSK2983559, exhibit significant efficacy in mouse models of IBD. Although clinical development halted at Phase I, these findings highlight RIPK2's potential as a therapeutic target[29]. TLR2, a member of the Toll-like receptor (TLR) family in innate immunity, recognizes pathogens and activates the NF-κB pathway, inducing the expression of inflammatory cytokines and chemokines[30]. Research has shown that TLR2-targeting drugs, such as VB-201, can reduce gastrointestinal inflammation in mouse models of IBD[30], although their clinical development was discontinued. These findings confirm TLR2 as an important therapeutic target in IBD. Based on this analysis, PIM3, RIPK2, and TLR2 were identified as candidate biomarkers for IBD in the macrophage C12 subcluster.

3.3.4 Candidate Biomarkers in Fibroblast C5 Subcluster for IBD

Differential gene expression analysis of the fibroblast C5 subcluster between disease and healthy control groups identified 949 commonly upregulated genes in both UC and CD. Since GWAS studies have identified relatively few IBD risk genes directly associated with fibroblasts, this study collected additional fibroblast-related IBD datasets from public databases. Differential expression analysis of these datasets was performed, and the results were intersected with the 949 previously identified genes. This process yielded 64 consistently upregulated genes in the disease group. After reviewing the clinical progress of these genes, six candidate target genes with anti-fibrotic clinical trials were identified: CDH11, OSMR, PDGFRB, ROMO1, BRD4, and UGCG. Downregulated genes did not meet the criteria for further analysis.

CDH11 was significantly upregulated in fibroblasts in the disease group (Figure 4). By mediating cell adhesion, CDH11 promotes tissue development and maintains barrier integrity, suggesting a key role in IBD[31]. OSMR, also highly expressed in fibroblasts from the disease group, increases the production of inflammatory cytokines through downstream signaling pathways, contributing to inflammatory responses[32]. PDGFRB, elevated in disease-associated fibroblasts, has been implicated in fibrosis in various diseases, suggesting its involvement in IBD-related fibrotic processes[33]. ROMO1, which regulates reactive oxygen species (ROS) production, is associated with inflammation and oxidative stress, making it a potential key factor in IBD[34]. BRD4, significantly upregulated in fibroblasts from the disease group, acts as an inflammatory gene regulator and may influence IBD through the regulation of IL-1β-induced inflammatory pathways[35]. UGCG, also highly expressed in fibroblasts from the disease group, is linked to tissue fibrosis, suggesting its involvement in IBD-related fibrotic mechanisms[36]. Collectively, these candidate biomarkers participate in pathological processes such as inflammation, fibrosis, and oxidative stress, providing valuable insights into the molecular mechanisms of IBD and potential targets for therapeutic development.

4 Conclusions

By integrating multiple IBD-related scRNA-seq datasets from public databases, this study constructed a high-quality, large-scale dataset containing 1,167,689 cells. Using bioinformatics tools, key IBD-associated cell types and their target genes were systematically analyzed. The study identified Tregs, Th17 cells, macrophage C12 subcluster, and fibroblast C5 subcluster as significantly associated with IBD. This work not only elucidate the critical roles of these cells in inflammatory responses, fibrosis, and immune regulation at the single-cell level, but also provides novel insights for further investigation into their specific regulatory networks, the identification of potential therapeutic targets, and the exploration of personalized treatment strategies. Differential gene expression analysis and drug development investigations identified several IBD-related candidate biomarkers. In Tregs, IL2RA, HSF1, and TNFSF8 showed regulatory potential. In the macrophage C12 subcluster, PIM3, RIPK2, and TLR2 emerged as key target genes involved in the regulation of inflammatory factors and immune responses. In the fibroblast C5 subcluster, CDH11, OSMR, and PDGFRB were strongly linked to fibrosis and inflammation, potentially contributing to tissue damage and repair in IBD. These findings providing significant insights for further research into pathogenesis and the development of targeted therapies. Future studies could incorporate larger clinical datasets and experimental validations to identify specific biomarkers and explore their potential in precision medicine for IBD.

References

1. G. P. Ramos and K. A. Papadakis, Mayo Clinic Proceedings **94** (1), 155 (2019).
2. G. G. Kaplan, Nature Reviews Gastroenterology & Hepatology **12** (12), 720 (2015).
3. Q. Guan, Journal of Immunology Research **2019**, 1 (2019).
4. S. T. Radhakrishnan, J. L. Alexander, B. H. Mullish, K. I. Gallagher, N. Powell, L. C. Hicks, A. L. Hart, J. V. Li, J. R. Marchesi, and H. R. Williams, Alimentary pharmacology & therapeutics **55** (1), 26 (2022).
5. E. Hedlund and Q. Deng, Molecular Aspects of Medicine **59**, 36 (2018).
6. J. C. Martin, C. Chang, G. Boschetti, R. Ungaro, M. Giri, J. A. Grout, K. Gettler, L.-S. Chuang, S. Nayar, A. J. Greenstein, M. Dubinsky, L. Walker, A. Leader, J. S. Fine, C. E. Whitehurst, M. L. Mbow, S. Kugathasan, L. A. Denson, J. S. Hyams, J. R. Friedman, P. T. Desai, H. M. Ko, I. Laface, G. Akturk, E. E. Schadt, H. Salmon, S. Gnjatic, A. H. Rahman, M. Merad, J. H. Cho, and E. Kenigsberg, Cell **178** (6), 1493 (2019).
7. C. S. Smillie, M. Biton, J. Ordovas-Montanes, K. M. Sullivan, G. Burgin, D. B. Graham, R. H. Herbst, N. Rogel, M. Slyper, J. Waldman, M. Sud, E. Andrews, G. Velonias, A. L. Haber, K. Jagadeesh, S. Vickovic, J. Yao, C. Stevens, D. Dionne, L. T. Nguyen, A. C. Villani, M. Hofree, E. A. Creasey, H. Huang, O. Rozenblatt-Rosen, J. J. Garber, H. Khalili, A. N. Desch, M. J. Daly, A. N. Ananthakrishnan, A. K. Shalek, R. J. Xavier, and A. Regev, Cell **178** (3), 714 (2019).
8. K. Parikh, A. Antanaviciute, D. Fawkner-Corbett, M. Jagielowicz, A. Aulicino, C. Lagerholm, S. Davis, J. Kinchen, H. H. Chen, N. K. Alham, N. Ashley, E. Johnson, P. Hublitz, L. Bao, J. Lukomska, R. S. Andev, E. Björklund, B. M. Kessler, R. Fischer, R. Goldin, H. Koohy, and A. Simmons, Nature **567** (7746), 49 (2019).
9. J. Kinchen, H. H. Chen, K. Parikh, A. Antanaviciute, M. Jagielowicz, D. Fawkner-Corbett, N. Ashley, L. Cubitt, E. Mellado-Gomez, M. Attar, E. Sharma, Q. Wills, R. Bowden, F. C. Richter, D. Ahern, K. D. Puri, J. Henault, F. Gervais, H. Koohy, and A. Simmons, Cell **175** (2), 372 (2018).
10. M. Uzzan, J. C. Martin, L. Mesin, A. E. Livanos, T. Castro-Dopico, R. Huang, F. Petralia, G. Magri, S. Kumar, Q. Zhao, A. K. Rosenstein, M. Tokuyama, K. Sharma, R. Ungaro, R. Kosoy, D. Jha, J. Fischer, H. Singh, M. E. Keir, N. Ramamoorthi, W. E. O'Gorman, B. L. Cohen, A. Rahman, F. Cossarini, A. Seki, L. Leyre, S. T. Vaquero, S. Gurunathan, E. K. Grasset, B. Losic, M. Dubinsky, A. J. Greenstein, Z. Gottlieb, P. Legnani, J. George, H. Irizar, A. Stojmirovic, C. Brodmerkel, A. Kasarkis, B. E. Sands, G. Furtado, S. A. Lira, Z. K. Tuong, H. M. Ko, A. Cerutti, C. O. Elson, M. R. Clatworthy, M.

Merad, M. Suárez-Fariñas, C. Argmann, J. A. Hackney, G. D. Victora, G. J. Randolph, E. Kenigsberg, J. F. Colombel, and S. Mehandru, Nature Medicine **28** (4), 766 (2022).

11. M. Krzak, T. Alegbe, D. L. Taylor, M. Ghouraba, M. Strickland, R. Satti, T. Thompson, K. Arestang, M. J. Przybilla, and L. Ramirez-Navarro, medRxiv, 2023.09. 06.23295056 (2023).
12. X. Hao, K. Wang, C. Dai, Z. Ding, W. Yang, C. Wang, and S. Cheng, Human Molecular Genetics **29** (18), 3145 (2020).
13. Y. Ma, F. Qiu, C. Deng, J. Li, Y. Huang, Z. Wu, Y. Zhou, Y. Zhang, Y. Xiong, and Y. Yao, Genome Medicine **14** (1), 16 (2022).
14. M. J. Zhang, K. Hou, K. K. Dey, S. Sakaue, K. A. Jagadeesh, K. Weinand, A. Taychameekiatchai, P. Rao, A. O. Pisco, and J. Zou, Nature genetics **54** (10), 1572 (2022).
15. L. Tarhan, J. Bistline, J. Chang, B. Galloway, E. Hanna, and E. Weitz, BioRxiv (2023).
16. R. Edgar, M. Domrachev, and A. E. Lash, Nucleic acids research **30** (1), 207 (2002).
17. F. A. Wolf, P. Angerer, and F. J. Theis, Genome Biology **19** (1), 15 (2018).
18. L. McInnes, J. Healy, and J. Melville, arXiv preprint arXiv:1802.03426 (2018).
19. L. Kong, V. Pokatayev, A. Lefkovith, G. T. Carter, E. A. Creasey, C. Krishna, S. Subramanian, B. Kochar, O. Ashenberg, H. Lau, A. N. Ananthakrishnan, D. B. Graham, J. Deguine, and R. J. Xavier, Immunity **56** (2), 444 (2023).
20. S. Dharmasiri, E. M. Garrido-Martin, R. J. Harris, A. C. Bateman, J. E. Collins, J. R. F. Cummings, and T. Sanchez-Elsner, Inflammatory Bowel Diseases **27** (10), 1641 (2021).
21. A. Garrido-Trigo, A. M. Corraliza, M. Veny, I. Dotti, E. Melón-Ardanaz, A. Rill, H. L. Crowell, Á. Corbí, V. Gudiño, M. Esteller, I. Álvarez-Teubel, D. Aguilar, M. C. Masamunt, E. Killingbeck, Y. Kim, M. Leon, S. Visvanathan, D. Marchese, G. Caratù, A. Martin-Cardona, M. Esteve, J. Panés, E. Ricart, E. Mereu, H. Heyn, and A. Salas, Nature Communications **14** (1) (2023).
22. L. Massimino, L. A. Lamparelli, Y. Houshyar, S. D'Alessio, L. Peyrin-Biroulet, S. Vetrano, S. Danese, and F. Ungaro, Nature Computational Science **1** (8), 511 (2021).
23. A. Gerussi, M. Carbone, R. Asselta, and P. Invernizzi, Liver Immunology: Principles and Practice, 69 (2020).
24. O. Levi-Galibov, H. Lavon, R. Wassermann-Dozorets, M. Pevsner-Fischer, S. Mayer, E. Wershof, Y. Stein, L. Brown, W. Zhang, and G. Friedman, (2020).
25. E. Chang, UMass Chan Medical School, 2024.
26. M. Mahapatro, L. Erkert, and C. Becker, Cells **10** (1), 111 (2021).
27. C.-S. Cătană, I. B. Neagoe, V. Cozma, C. Magdaş, F. Tăbăran, and D. L. Dumitraşcu, World journal of gastroenterology: WJG **21** (19), 5823 (2015).
28. T. Buchacher, A. Shetty, S. A. Koskela, J. Smolander, R. Kaukonen, A. G. Sousa, S. Junttila, A. Laiho, O. Rundquist, and T. Lönnberg, Cell Reports **42** (12) (2023).
29. Y. Lai, X. Wang, X. Sun, S. Wu, X. Chen, C. Yang, W. Zhang, X. Yu, Y. Tong, and F. Ma, Biochemical Pharmacology **214**, 115647 (2023).
30. D. Corridoni, T. Chapman, T. Ambrose, and A. Simmons, Frontiers in medicine **5**, 32 (2018).
31. J. Liu, W. Gong, P. Liu, Y. Li, H. Jiang, X. Wu, Y. Zhao, and J. Ren, Expert Opinion on Therapeutic Targets **26** (11), 1011 (2022).
32. C. L. Wolf, C. Pruett, D. Lighter, and C. L. Jorcyk, Frontiers in Immunology **14**, 1239732 (2023).
33. V. Valatas, E. Filidou, I. Drygiannakis, and G. Kolios, Annals of gastroenterology **30** (4), 393 (2017).
34. M. F. Olaniyan, M. A. Muhibi, and T. B. Olaniyan, Journal of Preventive, Diagnostic and Treatment Strategies in Medicine **2** (2), 94 (2023).
35. J. T.-C. Chen, X. Hu, I. U. Otto, C. Schürger, B. R. von Bieberstein, K. Doppler, S. M. Krug, M. K. Hankir, R. Blasig, and C. Sommer, Neurobiology of Disease **185**, 106246 (2023).
36. J. C. Hartel, N. Merz, and S. Grösch, Frontiers in Pharmacology **13**, 1002915 (2022).