The Common Gut Microbiome Structure Cross Inflammatory Bowel Disease and Autoimmune Arthritis

Junxian Chen *

Jiangnan University, Wuxi 214122, Jiangsu, China

Abstract. Objective: Inflammatory bowel disease (IBD) and autoimmune arthritis frequently co-occur and share genetic susceptibility, yet their common gut microbial alterations remain poorly characterized.
Methods: We conducted a meta-analysis of two independent 16S rRNA sequencing datasets from the Microbiome HD database, encompassing patients with Crohn's disease (CD), ulcerative colitis (UC), rheumatoid arthritis (RA) or psoriatic arthritis (PsA). Data were normalized using log-CPM and supervised normalization to remove batch effects. Differential abundance analysis was performed using Wilcoxon tests with FDR correction. A gradient boosting machine (GBM) model was trained to evaluate the diagnostic potential of microbial features.
Results: We identified consistent microbial alterations across all four diseases, including significant enrichment of *Veillonella* and depletion of SCFA-producing genera such as *Ethanoligenens, Oscillibacter, Phascolarctobacterium,* and *Clostridium_IV*. The GBM models demonstrated strong diagnostic performance for CD, RA, and PsA (AUC > 0.8), though cross-disease classification performance was variable, indicating both shared and disease-specific microbial features.
Conclusion: Our findings reveal a common gut dysbiosis pattern in IBD and arthritis, characterized by the rise of pro-inflammatory taxa and loss of protective bacteria. These shared microbial signatures suggest a potential common etiology involving impaired SCFA metabolism and mucosal immunity, supporting the "gut–joint axis" hypothesis. Microbial biomarkers may serve as diagnostic tools, though disease-specific variations warrant further investigation.

1. Introduction

Inflammatory Bowel Disease (IBD) is a chronic inflammatory disorder of the gastrointestinal tract, primarily comprising two subtypes: Crohn's Disease (CD) and Ulcerative Colitis (UC). Patients with IBD often present with symptoms such as diarrhea, abdominal pain, and weight loss. Furthermore, over 50% of patients may develop extraintestinal manifestations, including anemia, osteoporosis, and arthritis, indicating a close association with various autoimmune conditions. Among these, arthritis is the most prevalent [1], including conditions such as Rheumatoid Arthritis (RA) and Psoriatic Arthritis (PsA). RA is an autoimmune disease characterized by synovitis, typically presenting with symmetric polyarticular swelling, pain, morning stiffness, and potentially leading to joint deformity and functional loss in advanced stages. PsA is an inflammatory arthropathy associated with psoriasis, featuring psoriatic skin lesions accompanied by pain, swelling, tenderness, stiffness, and impaired movement in the joints and surrounding soft tissues. Epidemiological studies have demonstrated that the incidence of both RA and PsA is significantly elevated in IBD patients compared to the general population [2, 3]. Moreover, both IBD and these arthritic conditions are heritable and demonstrate a significant genetic correlation at the genomic level [4].

Recent research advances indicate that disease pathogenesis cannot be fully explained by genetic factors alone, with environmental elements playing a pivotal role in disease progression. Specifically, disruptions in the structure and function of the gut microbiota are recognized as a critical bridge connecting genetic susceptibility and environmental exposures, with aberrant changes in the abundance of specific microbial taxa being a key initiating factor. Importantly, such microbial changes can provoke immune dysregulation, a central mechanism underlying the pathogenesis of autoimmune diseases [5]. Recently, the "gut-joint axis" hypothesis has gained increasing attention, providing a novel perspective for understanding the relationship between gut microbial dysbiosis and arthritis. The hypothesis proposes that the local immune disturbances and microbial imbalances in the gut may remotely influence the onset and progression of arthritis via immune regulation. Previous studies have shown that, compared to other disease categories, autoimmune diseases exhibit a higher consistency in gut microbiome dysbiosis patterns and disease severity and clinical heterogeneity are more significantly influenced by the gut microbiota [6]. Although numerous studies have individually described the microbial characteristics of

* Corresponding author: 1282230422@stu.jiangnan.edu.cn

patients with CD, UC, RA, and PsA, there remains a lack of research on the shared patterns of microbial dysbiosis across these diseases.

Therefore, this study aims to analyze the shared patterns of gut microbial composition across Crohn's disease, ulcerative colitis, rheumatoid arthritis, and psoriatic arthritis by integrating gut microbiota data from patients with these four diseases. We hypothesize that although these disorders exhibit distinct clinical manifestations, they may share common features of gut microbiome dysbiosis, which are closely associated with immune-mediated pathogenesis. To test this hypothesis, we will utilize publicly available fecal 16S rRNA sequencing datasets to compare the microbial structures between disease groups and healthy controls, and to identify cross-disease microbial markers. This work is expected to provide novel evidence for elucidating the common microbial mechanisms underlying the "gut–joint axis" in autoimmune diseases.

2. Methods

2.1. Data acquisition and processing

The 16S rRNA microbiome sequencing data required for this study were sourced from the Microbiome HD database [7], which constitutes a comprehensive repository of systematically integrated human gut microbiome studies. We collected two independent datasets for analysis from the database: dataset from a study on IBD patients [8] and dataset from a study on arthritis patients [9].

2.2. Normalization of the variation between datasets for meta-analyses

Since variations exist in both subject population characteristics and experimental protocols across the different datasets, we performed standardized preprocessing of the data to remove batch effects prior to analysis. In brief, we first normalized the raw species abundance counts to counts per million (CPM) to adjust for variations in sequencing depth. This was followed by log-transformation (log-CPM) to approximate a normal distribution and stabilize the variance. Finally, We performed supervised normalization (SNM) to remove significant batch effects while preserving biologically relevant variations [10].

2.3. Univariate analysis of disease-associated microbial species

To identify microbial species significantly associated with disease status, we performed a univariate analysis approach. Non-parametric Wilcoxon rank-sum tests were conducted to assess abundance differences between disease and healthy groups for each species within individual studies, and the $\log_{10}$-transformed median fold-change (FC) was calculated as the effect size. A blocked Wilcoxon test was further applied to integrate results across all studies while controlling for inter-study variation. The false discovery rate (FDR) method was used to adjust p-values, with a significance threshold set at FDR < 0.05. Finally, a heatmap was generated to visualize species-specific abundance differences and statistical significance across studies, highlighting species that reached FDR < 0.05 in at least one study.

2.4. Gradient Boosting Machine (GBM)–based model for regression analysis of bacterial signatures for disease diagnosis

GBM-based models were constructed to differentiate between the diseased group and the healthy control group (or other disease groups) using the caret package in R.

To ensure a non-overlapping and reproducible allocation of healthy control samples across different disease models, all healthy samples were first isolated and then randomly partitioned. Given the limited overall sample size in this study, the allocation of 70% of the healthy controls to the training set and the remaining 30% reserved for other models was primarily designed to ensure that the model had sufficient data to learn the microbial signature differences between healthy and diseased states, thereby preventing underfitting due to inadequate training data. The final dataset for model training and testing was assembled by combining samples from the target disease group with the predefined subset of healthy controls. The dataset was randomly partitioned into a 70% training set and a 30% hold-out test set. During model training and tuning, a 4-fold cross-validation repeated three times was employed. This process involved randomly dividing the training data into four subsets in each repetition. The model was tuned over a grid of hyperparameters, including tree depth, number of boosting iterations, and learning rate.

The performance of the final tuned model was evaluated on the held-out test set by calculating the AUROC and generating a confusion matrix. The relative importance of bacterial features in the final model was extracted and ranked for further analysis.

3. Results

3.1. Collected datasets and normalization

Meta-analysis data of the 242 participants from two cohorts (IBD128 [8], Arthritis 114 [9]) was processed using the same pipeline. To mitigate batch effects between different studies, raw species abundance counts were first normalized using CPM to account for differences in sequencing depth (Figure 1A). Then, log-transformation and scaling were applied to convert the data into approximately normally distributed log-CPM, thereby reducing heteroscedasticity (Figure 1B). Finally, supervised normalization of microarrays (SNM) was employed to process the data, effectively removing batch effects while preserving biologically relevant variations (Figure 1C).

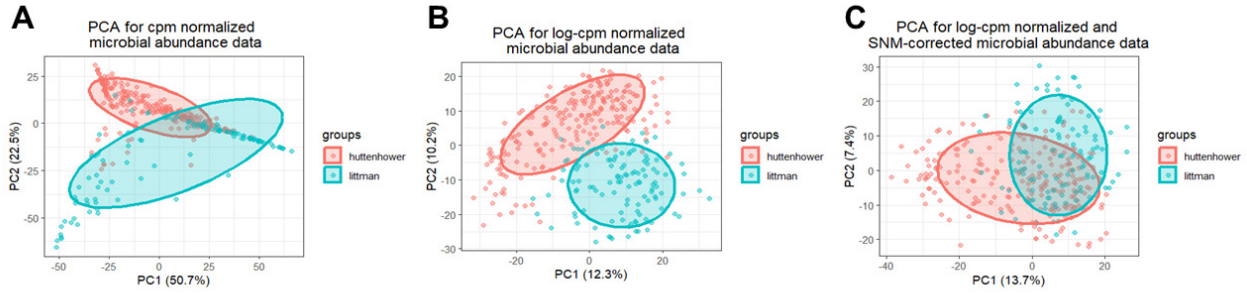

Figure 1. Normalization of individual microbiome groups at the genus level with samples colored by studies. (A) PCA of CPM normalized data. (B) PCA of log-CPM normalized data (C) PCA of log-CPM normalized data after SNM.

3.2. Consistency of gut microbiota alterations in patients with UC, CD, RA, and PsA

To investigate whether shared patterns of microbial alteration exist among CD, UC, RA, and PsA, we performed a meta-analysis of gut microbiota sequencing data from two independent datasets. The results demonstrated a consistent microbial shift across all four diseases, characterized by a significant enrichment of *Veillonella* (FDR < 0.05) and a marked reduction in several general, including *Ethanoligenens, Clostridium_IV, Oscillibacter, Phascolarctobacterium*. Furthermore, in CD, UC, RA, and PsA, the abundances of *Veillonella, Sutterella, Haemophilus, Clostridium_XIVa,* and *Lactobacillus* consistently exhibited an increasing trend. Among these, *Veillonella* showed a significant increase in CD, RA, and PsA (FDR < 0.05), but not in UC, which may be attributable to insufficient sample size and high inter-individual variability. Each of the other elevated taxa, except for *Haemophilus*, reached statistical significance (FDR < 0.05) in at least one dataset. On the other hand, Lactococcus, Megamonas, Faecalibacterium, Coprobacillus, Alistipes, Akkermansia, Ruminococcus2, Dorea, Coprococcus, Ruminococcus, Ethanoligenens, Oscillibacter, Phascolarctobacterium, and *Clostridium_IV* mostly demonstrated a decreasing abundance trend across the four diseases. However, likely due to limited sample size or substantial interpersonal variation, only a minority of these declines reached statistical significance at FDR < 0.05.

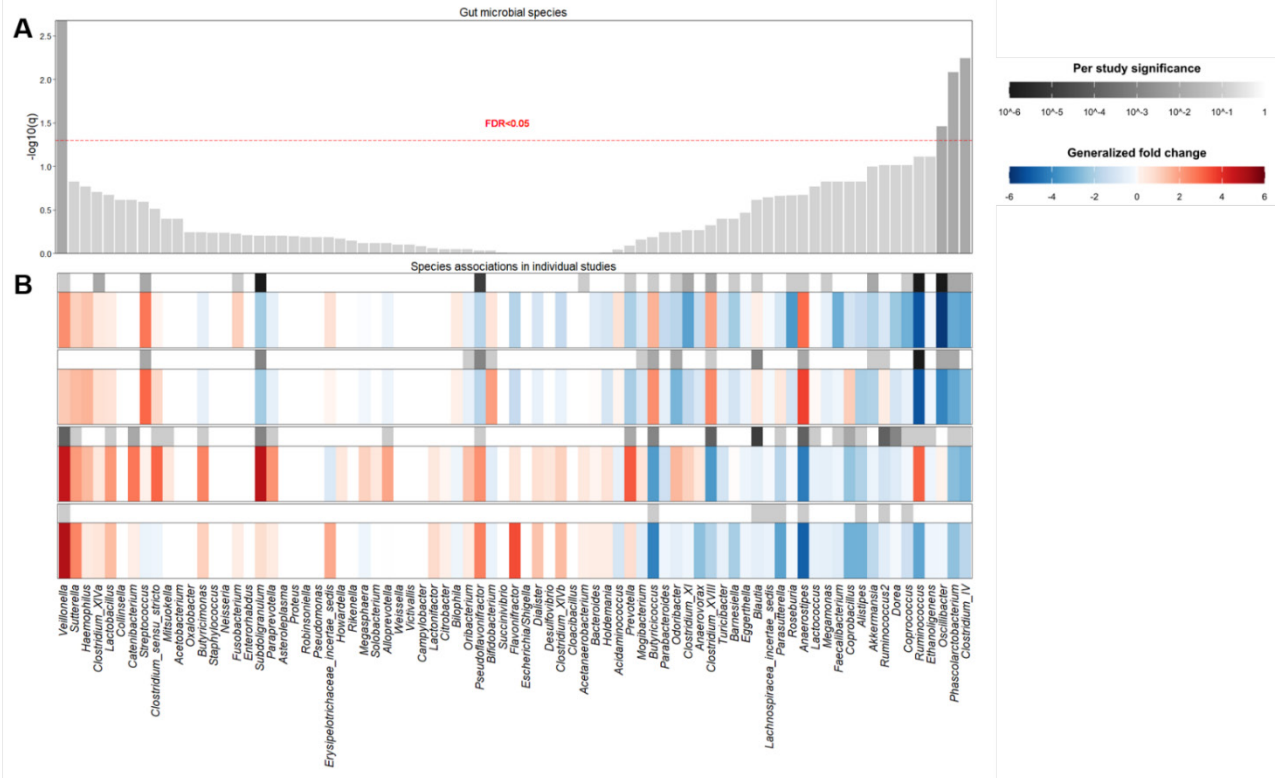

Figure 2. Alterations in the abundance of the gut microbiota in CD, UC, RA, and PsA. (A) Meta-analysis identified a core set of gut microbiota associated with the four diseases. Species are ordered by the significance and direction of change (left: enriched in patients; right: depleted in patients). (B) Genera-level generalized fold change within individual studies.

3.3. Gut microbiota alterations in UC, CD, RA, and PsA from individual studies

To further investigate whether similar patterns of microbial alterations exist among these autoimmune diseases, independent analyses were conducted for each dataset. The results revealed consistent gut microbial alterations across patients with CD, UC, RA, and PsA (Figure 3). Specifically, genera such as *Veillonella*, Lactobacillus, and *Clostridium_XIVa* were enriched, while Faecalibacterium, Ethanoligenens, Dorea, Coprococcus, Akkermansia, Alistipes, Oscillibacter, Phascolarctobacterium, and *Clostridium_IV* were reduced. Notably, most of these differentially abundant genera belong to the order *Clostridiales*, which is critically involved in short-chain fatty acid (SCFA) metabolism. These findings suggest a potential dysregulation of SCFA metabolic pathways in individuals with CD, UC, RA, and PsA, possibly contributing to disease pathogenesis.

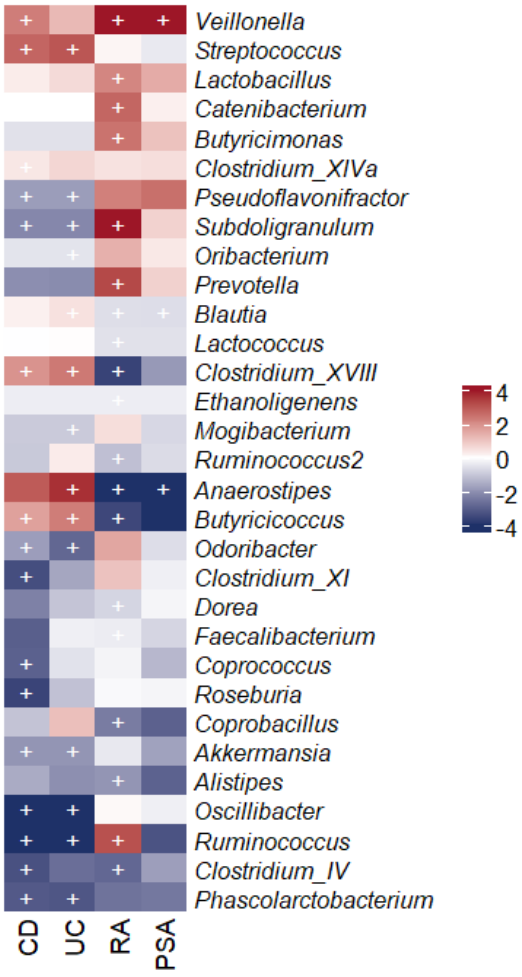

Figure 3. Common microbial signatures across the four diseases. (two-sided Wilcoxon test, FDR-corrected P-value <0.05); columns represent datasets. Generalized fold change is colored by direction of the effect, where red indicates enriched abundance in cases and blue indicates depletion. White indicates that the genus was not present in that data set. +: P<0.05

3.4. Diagnostic microbial signatures across disease types

To evaluate whether gut microbiota characteristics share common alteration trends in IBD and arthritis, we trained GBM learning models using four-fold cross-validation and assessed their diagnostic accuracy in discriminating between patients and healthy controls. We first evaluated the intrinsic discriminative ability of the constructed models. In patient–control validation using species-level taxonomically normalized abundances, we observed performance metrics ranging in the area under the receiver operating characteristic curve (AUROC) from 0.7769 to 0.9455 (Figure 4A, diagonal section). Models trained on CD, RA, and PsA datasets all demonstrated strong discriminative power (Figure 4B, diagonal section, AUC > 0.8), whereas the model trained on UC data showed lower performance (AUC = 0.7769) in self-discrimination compared to the other three models. Furthermore, key genera with significantly altered abundances—such as *Ruminococcus, Clostridium_XIV,* and *Phascolarctobacterium*—played critical roles in the discriminative performance of the models, ranking among the top in feature importance (Figure 4A). To more comprehensively evaluate model accuracy, we subsequently systematically examined the models' discriminatory efficacy across different disease types, thereby reassessing whether relatively consistent microbial signatures exist among CD, UC, RA, and PsA. The cross-disease discriminative ability of our constructed models was unstable. Although the model trained on UC data demonstrated reasonably good discriminative performance for UC, RA, and PsA (AUC > 0.7), models trained on RA and PsA data exhibited relatively poor performance in distinguishing IBD. The model trained on CD data exhibited poor discriminative efficacy for UC and RA, while the model trained on PsA data performed poorly in distinguishing CD but showed relatively better discriminative capability for UC and RA (AUC > 0.7). The model trained on RA data demonstrated the weakest performance in discriminating both CD and UC, with AUC values below 0.6. The instability in model discriminative performance may be attributed to the limited sample size or substantial inter-individual variability within the datasets.

4. Conclusion

This study was based on two publicly available 16s rRNA sequencing datasets, including four autoimmune diseases: Ulcerative Colitis, Crohn's Disease, Rheumatoid Arthritis, and Psoriatic Arthritis. We applied standardization procedures to mitigate batch effects (Figure 1A, B, C). After processing with log-CPM and SNM, the distribution of sample points in the PCA plot no longer showed separation by batch, with healthy control samples from different studies intermixing (Figure 1C), demonstrating that batch effects were effectively controlled. This laid a solid foundation for reliable cross-disease microbiome analysis on the integrated dataset.

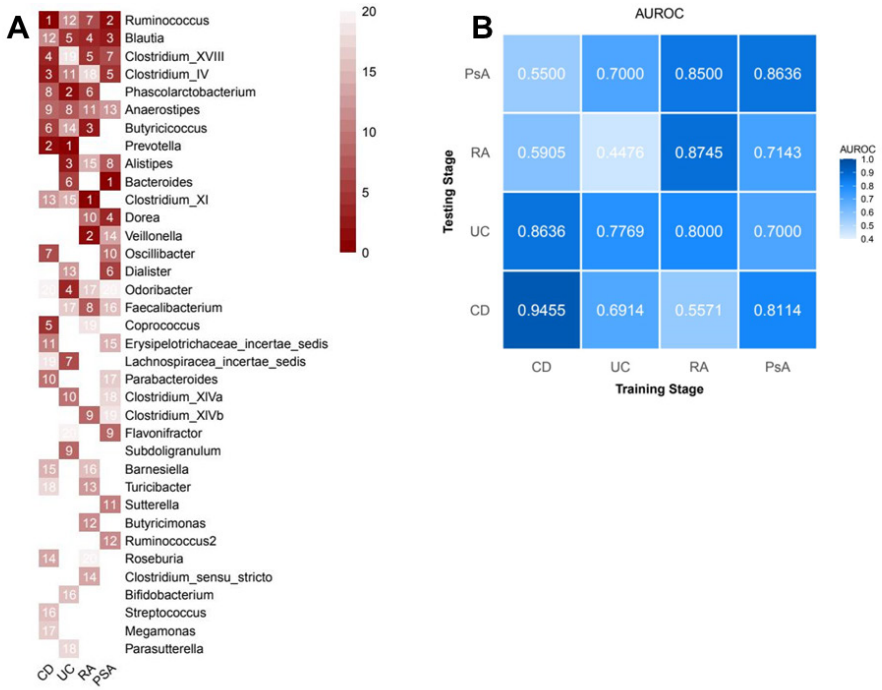

Figure 4. Ranking the feature importance of each bacterial genus in each trained model and A GBM-based model for inter-disease classification. (A) Ranking species by feature importance in each trained model based on GBM. (B) The model classification performance, assessed by the AUROC, includes internal validation (diagonal) and inter-model transfer validation (off-diagonal).

First, we conducted a meta-analysis of two independent datasets, which revealed consistent patterns of gut microbiota alterations in patients with IBD and arthritis. Notably, *Veillonella* showed significant enrichment across all these disease groups. This discovery aligns with previous literature studies [11, 12], further confirming the potential importance of this genus in autoimmune diseases. From a mechanistic perspective, *Veillonella*, as a predominant oral microbiota, may translocate through the "oral-gut" and "oral-joint" biological axes [13-15]. Particularly noteworthy is that the enrichment level of this bacterium correlates positively with disease inflammatory activity, suggesting its direct involvement in the inflammatory process. Recent studies indicate that increased nitrate production in the inflammatory microenvironment may provide favorable conditions for the abnormal colonization of *Veillonella,* thereby forming a vicious cycle that promotes persistent inflammation. On the other hand, our study also found a widespread reduction in several protective bacterial genera, including *Ethanoligenens, Oscillibacter, Phascolarctobacterium,* and *Clostridium_IV* (Figure 2). This phenomenon was consistently verified in subsequent independent dataset analyses (Figure 3), enhancing the reliability of the finding. In-depth analysis revealed that among these reduced microbiota, such as *Phascolarctobacterium* and *Roseburia* (Figure 3), are key producers of SCFA. Their reduction may lead to decreased levels of anti-inflammatory metabolites like butyrate and propionate, thereby impairing intestinal epithelial barrier function [16]. It is particularly important to note that *Roseburia* not only participates in SCFA synthesis but has also been proven to promote the differentiation and proliferation of anti-inflammatory regulatory T cells (Tregs) in the gut [17]. Therefore, the reduction of this bacterium may exacerbate intestinal inflammation by weakening immune regulatory functions. These microbial changes collectively form a vicious cycle: the reduction of protective microbiota leads to impaired intestinal barrier integrity and weakened anti-inflammatory signals, while the abnormal proliferation of pro-inflammatory bacteria further activates the immune system, ultimately resulting in the persistence and exacerbation of local and systemic inflammatory responses.

Meanwhile, these genera with significantly altered abundances predominantly play crucial roles in the GBM discrimination model we constructed based on gut microbiota characteristics, with most of these genera - such as *Ruminococcus, Clostridium_XIV,* and *Phascolarctobacterium* - ranking among the top features in importance (Figure 4A). The GBM model we developed effectively distinguished between CD, UC, RA, PsA, and healthy controls. Specifically, models trained on CD, RA, and PsA datasets demonstrated strong discriminative power (AUC > 0.8), while the model trained on UC data showed relatively weaker performance (AUC > 0.7). However, the cross-disease discriminative performance of these trained models varied significantly. Only the model trained on UC data demonstrated relatively good discriminatory ability for the other three diseases (AUC > 0.7). In contrast, the model trained on PsA data showed poor performance in distinguishing CD (AUC = 0.55), while maintaining better efficacy for UC and RA (AUC > 0.7). The model trained on RA data exhibited the weakest cross-disease discriminative capability, with AUC values for both CD and UC below 0.6. This may be attributed to the limited sample size in

both datasets, particularly the insufficient number of healthy controls leading to inadequate model training, coupled with substantial inter-individual variability in the sample data (including undefined disease subtypes, wide age ranges among patients, and imbalanced sample sizes across different disease groups). These Factors collectively result in unstable discriminative performance across different diseases.

In summary, we identified numerous common microbial signatures associated with IBD and autoimmune arthritis, implying a potential common microbial function.

References

1. M. Harbord et al., "The First European Evidence-based Consensus on Extra-intestinal Manifestations in Inflammatory Bowel Disease," (in eng), J Crohns Colitis, vol. 10, no. 3, pp. 239-54, Mar 2016, doi: 10.1093/ecco-jcc/jjv213.
2. Y. Chen et al., "The risk of rheumatoid arthritis among patients with inflammatory bowel disease: a systematic review and meta-analysis," (in eng), BMC Gastroenterol, vol. 20, no. 1, p. 192, Jun 17 2020, doi: 10.1186/s12876-020-01339-3.
3. Y. Sun, Y. Li, and J. Zhang, "The causal relationship between psoriasis, psoriatic arthritis, and inflammatory bowel diseases," (in eng), Sci Rep, vol. 12, no. 1, p. 20526, Nov 28 2022, doi: 10.1038/s41598-022-24872-5.
4. G. Cao, Q. Luo, Y. Wu, and G. Chen, "Inflammatory bowel disease and rheumatoid arthritis share a common genetic structure," (in eng), Front Immunol, vol. 15, p. 1359857, 2024, doi: 10.3389/fimmu.2024.1359857.
5. H. J. Wu and E. Wu, "The role of gut microbiota in immune homeostasis and autoimmunity," (in eng), Gut Microbes, vol. 3, no. 1, pp. 4-14, Jan-Feb 2012, doi: 10.4161/gmic.19320.
6. T. Wang, P. R. Sternes, X. K. Guo, H. Zhao, C. Xu, and H. Xu, "Autoimmune diseases exhibit shared alterations in the gut microbiota," (in eng), Rheumatology (Oxford), vol. 63, no. 3, pp. 856-865, Mar 1 2024, doi: 10.1093/rheumatology/kead364.
7. C. Duvallet, S. M. Gibbons, T. Gurry, R. A. Irizarry, and E. J. Alm, "Meta-analysis of gut microbiome studies identifies disease-specific and shared responses," (in eng), Nat Commun, vol. 8, no. 1, p. 1784, Dec 5 2017, doi: 10.1038/s41467-017-01973-8.
8. X. C. Morgan et al., "Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment," Genome Biology, vol. 13, no. 9, p. R79, 2012/09/26 2012, doi: 10.1186/gb-2012-13-9-r79.
9. J. U. Scher et al., "Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis," (in eng), Elife, vol. 2, p. e01202, Nov 5 2013, doi: 10.7554/eLife.01202.
10. B. H. Mecham, P. S. Nelson, and J. D. Storey, "Supervised normalization of microarrays," (in eng), Bioinformatics, vol. 26, no. 10, pp. 1308-15, May 15 2010, doi: 10.1093/bioinformatics/btq118.
11. A. Mucientes et al., "Intestinal Dysbiosis, Tight Junction Proteins, and Inflammation in Rheumatoid Arthritis Patients: A Cross-Sectional Study," (in eng), Int J Mol Sci, vol. 25, no. 16, Aug 8 2024, doi: 10.3390/ijms25168649.
12. P. Qiu, T. Ishimoto, L. Fu, J. Zhang, Z. Zhang, and Y. Liu, "The Gut Microbiota in Inflammatory Bowel Disease," (in eng), Front Cell Infect Microbiol, vol. 12, p. 733992, 2022, doi: 10.3389/fcimb.2022.733992.
13. M. I. Hammad, G. Conrads, and M. M. H. Abdelbary, "Isolation, identification, and significance of salivary Veillonella spp., Prevotella spp., and Prevotella salivae in patients with inflammatory bowel disease," (in eng), Front Cell Infect Microbiol, vol. 13, p. 1278582, 2023, doi: 10.3389/fcimb.2023.1278582.
14. K. Eriksson et al., "Salivary Microbiota and Host-Inflammatory Responses in Periodontitis Affected Individuals With and Without Rheumatoid Arthritis," (in eng), Front Cell Infect Microbiol, vol. 12, p. 841139, 2022, doi: 10.3389/fcimb.2022.841139.
15. D. F. Rojas-Tapias et al., "Inflammation-associated nitrate facilitates ectopic colonization of oral bacterium Veillonella parvula in the intestine," (in eng), Nat Microbiol, vol. 7, no. 10, pp. 1673-1685, Oct 2022, doi: 10.1038/s41564-022-01224-7.
16. H. Yan and K. M. Ajuwon, "Butyrate modifies intestinal barrier function in IPEC-J2 cells through a selective upregulation of tight junction proteins and activation of the Akt signaling pathway," (in eng), PLoS One, vol. 12, no. 6, p. e0179586, 2017, doi: 10.1371/journal.pone.0179586.
17. Y. Haneishi, Y. Furuya, M. Hasegawa, A. Picarelli, M. Rossi, and J. Miyamoto, "Inflammatory Bowel Diseases and Gut Microbiota," (in eng), Int J Mol Sci, vol. 24, no. 4, Feb 14 2023, doi: 10.3390/ijms24043817.