Gut and Liver
https://doi.org/10.5009/gnl250360
pISSN 1976-2283 eISSN 2005-1212

Review Article

Translating Gut Microbiota into Diagnostics: A Multidimensional Approach for the Diagnosis of Inflammatory Bowel Disease

June-Young Lee[1,2], Ji-Ho Yoo[3], Ji Eun Kim[4], Jin-Woo Bae[1], Chang Kyun Lee[4]

[1]Department of Biology, Kyung Hee University, Seoul, Korea; [2]InnoCORE AI-CRED Institute, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea; [3]Department of Biomedical and Pharmaceutical Sciences, Kyung Hee University, Seoul, Korea; [4]Department of Gastroenterology, Center for Crohn's and Colitis, Kyung Hee University Hospital, College of Medicine, Kyung Hee University, Seoul, Korea

Article Info
Received July 29, 2025
Revised September 22, 2025
Accepted October 1, 2025
Published online November 12, 2025

Corresponding Author
Chang Kyun Lee
ORCID https://orcid.org/0000-0002-4279-3825
E-mail changkyun.lee@khu.ac.kr

The gut microbiota has emerged as a key factor in the pathophysiology of inflammatory bowel disease (IBD), providing novel opportunities for diagnostic innovation. Traditional biomarkers, such as C-reactive protein and fecal calprotectin, are widely used in clinical practice; however, their ability to reflect disease complexity and microbial dysregulation remains limited. Recent advances in metagenomics and multi-omics integration have enabled high-resolution profiling of microbial communities and their functional capacities and associated metabolites. Differential abundance analysis and machine learning models have been used to identify microbial biomarkers that can distinguish patients with IBD from healthy individuals. Multicohort studies integrating microbiome and metabolomic data have further improved diagnostic accuracy and generalizability. Transcriptomic and proteomic analyses provide complementary insights into host–microbe interactions and disease mechanisms. In this review, we explored the potential of metagenomic biodata as diagnostic markers for IBD, with an emphasis on a multidimensional analytical approach. We highlight the recent developments in sequencing technologies, computational pipelines for microbial feature selection, and machine learning strategies applied to biomarker discovery. The integration of multi-omics data deepens our understanding of host–microbe interactions and facilitates the development of microbiota-informed diagnostic tools. As multidimensional microbial profiling evolves, its clinical utility for the diagnosis and stratification of IBD requires further investigation. **(Gut Liver, 2026;20:199-212)**

Key Words: Inflammatory bowel diseases; Gastrointestinal microbiome; Metagenomics; Biomarkers; Multiomics

INTRODUCTION

The human gut microbiota represents a dynamic and complex ecosystem that plays a fundamental role in maintaining gastrointestinal and systemic health. Disruptions in this ecosystem, referred to as dysbiosis, have been implicated as both a driver and consequence of inflammatory bowel disease (IBD), which encompasses Crohn's disease (CD) and ulcerative colitis (UC).[1-3] Although the pathogenesis of IBD involves a combination of genetic susceptibility, immune dysregulation, and environmental triggers, alterations in the gut microbiota have emerged as a central component of disease development and progression, gaining increasing attention as indicators of disease activity and potential contributors to the diagnosis of IBD.[4,5]

Currently, the diagnosis and monitoring of IBD are contingent upon endoscopic examination (with or without histological analysis) and/or radiological evaluation, supported by noninvasive biomarkers such as serological markers, including C-reactive protein; autoantibodies such as anti-*Saccharomyces cerevisiae* antibodies (ASCAs) and perinuclear anti-neutrophil cytoplasmic antibodies; and stool-based indicators such as fecal calprotectin and lactoferrin.[6,7] Although these markers offer practical advantages and are routinely used in clinical settings, they are limited in their ability to capture the diverse immune pathways

and inflammatory phenotypes associated with IBD or differentiate between disease subtypes.[8] Moreover, these conventional tools do not represent the contribution of the gut microbiota to disease pathogenesis or provide mechanistic insights into host–microbe interactions, which are increasingly recognized as integral to disease progression.

In recent years, developments in high-throughput sequencing and computational biology have significantly advanced our understanding of gut microbiota and its clinical applications in IBD. Metagenomic approaches, including 16S rRNA gene sequencing and shotgun metagenomics, allow for detailed identification of gut microbial communities at both the taxonomic and functional levels.[9,10] The introduction of long-read sequencing technologies has enhanced the resolution of microbial species and strains, facilitating the detection of subtle microbial shifts associated with disease phenotypes.[11] Beyond taxonomic profiling, recent studies have emphasized the importance of integrating multidimensional datasets, including transcriptomics and metabolomics, to better capture the functional state of the gut ecosystem. These multi-omics strategies identify the microbial components of the gut ecosystem and elucidate their functional roles and interactions with host biology.[9] These integrated approaches have facilitated the discovery of biomarkers with diagnostic and prognostic relevance. Furthermore, computational frameworks that leverage machine learning (ML) algorithms have been developed to systematically mine multi-omics data for the identification of robust microbial biomarkers.[12]

In this review, we summarized the latest progress in gut microbiome research relevant to IBD diagnosis. We emphasized the shortcomings of traditional diagnostic methods, reviewed recent developments in multi-omics technologies, and investigated new computational approaches for identifying biomarkers. Finally, we discuss how multidimensional analyses can guide the development of diagnostic frameworks for microbiome-oriented applications in IBD.

ADVANCES IN MULTIDIMENSIONAL MICROBIOME ANALYSIS FOR IBD

In recent years, gut microbiome studies on IBD have evolved beyond simple taxonomic profiling to include more comprehensive and mechanistic investigations.[13] Although fecal microbiome analysis remains a cornerstone of IBD microbiome studies and is widely used, many current studies are still limited to short-read 16S rRNA sequencing and often lack downstream validation following taxonomic profiling. Furthermore, due to the predominantly cross-sectional design of current host-microbiome studies,[14] a growing need exists to adopt high-resolution sequencing technologies, advanced statistical frameworks, and integrated multi-omics approaches to address these limitations.[9] These methodological advances are expected to enhance the identification of microbial signatures that are functionally relevant and potentially applicable in clinical diagnostics. To understand how this precision has been achieved, it is helpful to first examine advances in sequencing technologies used to study the gut microbiome.

1. Transition from 16S rRNA to high-resolution metagenomics

1) Evolution of 16S rRNA gene sequencing

16S rRNA gene sequencing has long been the standard approach for characterizing bacterial communities in fecal samples.[15] Initial applications relied on Sanger sequencing, which, despite its high precision, had limited throughput. The emergence of second-generation sequencing (next-generation sequencing [NGS]) platforms such as 454 pyrosequencing, Ion Torrent, and Illumina MiSeq has enabled more efficient and cost-effective community profiling.[16]

Among these, 454 pyrosequencing was one of the earliest NGS platforms applied to microbiome research and provided longer reads (400 to 500 bp) for improved genus-level classification.[17,18] It contributed to early IBD studies by demonstrating characteristic dysbiosis in patients, most notably the depletion of Firmicutes and the enrichment of Proteobacteria.[19] However, the platform was discontinued in 2016 due to high per-base costs, relatively low throughput, and the tendency to produce insertion–deletion errors in homopolymer regions. Thus, it now serves mainly as a historical reference and is unlikely to be integrated into future diagnostic applications.

Ion Torrent, introduced shortly after 454 pyrosequencing, presents an alternative approach based on semiconductor detection, enabling faster run times and lower reagent costs with read lengths comparable to those of its predecessor.[20] The platform provided read lengths similar to those of 454, offering faster running times and lower reagent costs. This platform has been applied in regional IBD studies. A study conducted in Russia used Ion Torrent sequencing to investigate microbial changes in patients with UC, reporting an increased abundance of pro-inflammatory genera, such as *Haemophilus*, *Olsenella*, and *Butyrivibrio*.[21] In contrast, the relative abundance of several beneficial taxa, including the butyrate-producing genera *Butyricimonas*, *Leuconostoc*, *Lactococcus*, and *Fusicatenibacter*, was significantly reduced in patients with UC compared to that in healthy controls. However, the platform exhibits technical limitations, including reduced

accuracy in homopolymer-rich sequences and batch variability, which limit its adoption in comparative studies requiring high precision.[22,23]

The launch of the Illumina MiSeq platform represented a pivotal advancement in 16S rRNA sequencing. Using sequencing-by-synthesis with fluorescently labeled reversible terminators, MiSeq enables high-fidelity paired-end reads of up to 2×300 bp, allowing nearly full coverage of the V3–V4 region, which is currently the most widely targeted region in gut microbiome studies. The high accuracy, scalability, and cost-efficiency of MiSeq, combined with strong support from reference databases such as Greengenes2[24] and SILVA,[25] and widely adopted analysis pipelines such as QIIME2.[26] In a recent large-scale Korean study, Kim *et al.* used Illumina-based 16S rRNA sequencing to investigate microbial alterations in patients with IBD, identifying distinct taxonomic shifts between CD and UC.[27] Their analysis revealed enrichment of *Escherichia-Shigella* in CD, whereas genera such as *Faecalibacterium*, *Bacteroides*, and *Dialister* were significantly reduced in patients with UC. Furthermore, He *et al.*[28] characterized the gut microbiome in Chinese IBD cohorts and reported that microbial alpha diversity declined with increasing disease activity. Their analysis further revealed a depletion of Lachnospiraceae and Ruminococcaceae, accompanied by an enrichment of *Escherichia-Shigella* in IBD patients.

Third-generation sequencing technologies have recently extended the utility of 16S rRNA gene profiling by facilitating full-length sequencing. Platforms such as Pacific Biosciences (PacBio) Single Molecule Real-Time sequencing and Oxford Nanopore Technologies produce kilobase-length reads, surpassing region-specific resolution and allowing for the precise classification of species and even strains.[11,29] These capabilities are particularly relevant in IBD, where strain-level shifts and functionally divergent lineages within the same species may play important roles in disease heterogeneity. However, to date, no studies have directly applied full-length 16S rRNA long-read sequencing using PacBio or Nanopore in large-scale IBD cohorts. This represents the current gap in the field. Considering the increased resolution and gene-level coverage provided by long-read technologies, their integration into IBD microbiome workflows is expected to significantly advance biomarker discovery and provide a more accurate picture of dysbiosis at the functional and strain-specific levels. Although the cost and computational requirements remain to be considered, the integration of long-read sequencing into microbiome workflows represents a significant advancement toward more precise and informative microbial biomarker discovery.

Table 1 provides a structured comparison of widely used sequencing platforms, including second- and third-generation technologies. The table highlights their technical characteristics (read length, resolution, and cost), as well as their reported utilization in IBD studies. To facilitate clinical interpretation, we also indicate whether each platform has been validated in large-scale cohorts and its current clinical readiness for diagnostic applications.

2) Metagenome shotgun sequencing: microbial biomarker

Although 16S rRNA gene sequencing has served as the cornerstone for characterizing gut microbial communities, limitations in taxonomic resolution and a lack of functional information have prompted a growing shift toward shotgun metagenomic sequencing.[30] Unlike amplicon-based methods, which target specific regions of the bacterial 16S rRNA

Table 1. Comparison of Sequencing Platforms Used in Microbiome and IBD Studies

Feature	Illumina	PacBio	Oxford Nanopore	Ion Torrent
Technology	Sequencing by synthesis (SBS)	Single molecule, real-time (SMRT) sequencing	Nanopore sequencing	Semiconductor sequencing (pH change detection)
Read length	50–300 bp	10–25 kb	10–100 kb	100–600 bp
Cost (per Gbp)	Low	High	Medium	Low
Resolution (16S)	Genus and species-level	Species and strain-level	Species and strain-level	Genus-level
IBD utilization (16S)	Widely used in large-scale studies to identify taxonomic shifts	No direct large-scale IBD applications to date	No direct large-scale IBD applications to date	Used in regional IBD studies to investigate microbial changes
Resolution (Shotgun)	Strain-level	Strain-level	Strain-level	Not typically applicable
IBD utilization (Shotgun)	Used to explore microbial signatures and find functional dysbiosis	No direct large-scale IBD applications to date	No direct large-scale IBD applications to date	Not typically applicable
Key applications	Amplicon sequencing, WGS, RNA-Seq, resequencing	*De novo* assembly, structural variation, full-length gene analysis	On-site analysis, real-time detection, epigenomics	Gene panels, microbial analysis, clinical diagnostics

IBD, inflammatory bowel disease; WGS, whole genome sequencing.

gene, shotgun metagenomics enables untargeted sequencing of the entire pool of microbial DNA within a sample. This comprehensive approach enables the simultaneous identification of bacteria, archaea, fungi, and viruses, thereby providing a broader view of the intestinal ecosystem.

One of the major advantages of metagenomic sequencing is its ability to characterize microbial communities at the species or strain level while simultaneously capturing the functional gene content.[31] This high resolution is particularly valuable in the context of IBD, in which subtle microbial shifts and species/strain-level differences may have clinical relevance.[32,33] Moreover, metagenomic data can be used to identify genes involved in microbial metabolic capacity, antibiotic resistance, and virulence.[34] These features are increasingly being recognized as modulators of host–microbe interactions.[35]

Several IBD cohort studies have used shotgun metagenomics to explore gut microbial signatures. Recently, Orejudo *et al.*[36] reported a comprehensive metagenomic profiling study that revealed consistent depletion of short-chain fatty acid (SCFA)–producing taxa, including *Faecalibacterium prausnitzii* and *Roseburia*, along with the enrichment of pro-inflammatory lineages, such as *Escherichia coli,* in newly diagnosed patients with IBD. These alterations were more pronounced during disease flares and were observed in patients naive to immunomodulatory therapy, revealing their potential role in disease initiation.[36] Furthermore, the study by Vich Vila *et al.*[32] reported distinct microbiome signatures between patients with IBD and those with IBS, using high-resolution metagenomic sequencing of 1,792 individuals. Patients with IBD showed more pronounced dysbiosis than those with IBS, characterized by a reduced abundance and strain-level diversity of beneficial taxa, such as *F. prausnitzii* and *Roseburia intestinalis*, and increased levels of *Bacteroides fragilis*, *Bacteroides vulgatus*, and *E. coli*. In addition, a recent multicohort study by Ning *et al.*[35] applied metagenomic analysis across several international IBD cohorts, consistently identifying the depletion of anti-inflammatory commensals and the enrichment of pro-inflammatory strains, including *Ruminococcus gnavus*, *B. fragilis*, and *E. coli*. Although the study incorporated metabolomic data, microbiome-specific findings indicated that compositional and functional dysbiosis are conserved across geographically and clinically diverse patient populations.

Metagenomics has facilitated the development of microbial biomarker panels that are increasingly integrated with clinical and omics data for building diagnostic models. Despite these advantages, shotgun metagenomic sequencing presents technical challenges that must be carefully managed. High levels of host DNA contamination, especially in mucosal or low-biomass samples, can obscure microbial signals and reduce the effective sequencing depth.[37] Furthermore, accurate taxonomic classification depends on the completeness and uneven coverage of sequences, particularly in low-biomass samples.[38] However, with continued improvements in sequencing technologies, bioinformatics pipelines, and reference databases, shotgun metagenomics is emerging as a powerful tool for high-resolution, functionally beneficial microbiome studies in IBD.

3) Beyond the bacteria: virome and mycobiome signatures in IBD

The gut ecosystem harbors not only bacteria but also a diverse range of microorganisms, including eukaryotic viruses, prokaryotic viruses (bacteriophages), and fungi.[39,40] Advances in sequencing technologies have enabled systematic characterization of these communities, making it possible to detect taxa that exist at relatively lower abundance compared with bacteria and to uncover their potential contributions to intestinal homeostasis and disease. In the context of IBD, dysbiosis extends beyond bacterial taxa, and altered virome and mycobiome profiles are increasingly recognized as important disease-associated signatures.[41,42] These domains may provide additional layers of disease-specific biomarkers when analyzed in conjunction with bacterial communities.

Several studies have demonstrated that the gut virome is perturbed in IBD patients compared with healthy controls. Norman *et al.*[42] performed metagenomic sequencing and reported increased diversity and richness of the enteric virome, with a notable expansion of Caudovirales bacteriophages in both CD and UC, suggesting that phage proliferation may represent a hallmark of IBD-associated dysbiosis. These findings indicate that virome signatures, particularly shifts in bacteriophage communities, may serve as non-bacterial biomarkers for IBD diagnosis and mechanistic studies.

Parallel research has highlighted the role of the gut mycobiome in IBD pathogenesis. Sokol *et al.*[41] observed an increased abundance of *Candida albicans* and altered fungal diversity in CD, along with reduced levels of *S. cerevisiae.* These changes were associated with immune activation, including ASCA responses commonly used as a serological marker in CD. In another study, Hoarau *et al.* demonstrated that *C. albicans* strains isolated from IBD patients exhibited enhanced virulence and promoted pro-inflammatory cytokine production, further implicating the mycobiome in disease exacerbation.[43] Together, these studies suggest that fungal dysbiosis may complement bacterial markers in differentiating disease states.

Considering virome and mycobiome data alongside bacterial microbiome profiles enables a more holistic ecological analysis of gut dysbiosis in IBD. Network-based approaches

that incorporate viral, fungal, and bacterial interactions may reveal key ecological drivers of disease. Importantly, such integrative analyses not only improve ecological interpretation but also provide a foundation for the systematic development of robust, multidimensional biomarkers that better capture the complexity of host–microbe dynamics in IBD.

2. Development of biomarker analysis

Advances in these sequencing platforms have enhanced the resolution of microbiome profiling as well as changed the way microbial features are analyzed and interpreted. Early analytical strategies focused primarily on taxonomic composition.[44] In addition, differential abundance (DA) analysis was used to identify the discriminant taxa between the groups using only the relative abundance of microbes without considering inter-individual differences and multivariable metadata.[45-47] In response to these limitations, newer DA methods and advanced computational tools have been developed, shifting the research focus toward identifying biologically meaningful microbial biomarkers that distinguish between disease states and healthy controls.[46,48] In the following sections, we outline the analytical strategies used for microbiome-based biomarker discovery.

1) DA analysis

Linear Discriminant Analysis Effect Size (LEfSe) is one of the earliest and most widely used tools for DA analysis.[49] LEfSe combines nonparametric statistical testing with linear discriminant analysis to identify microbial features that differ significantly between groups and have consistent effect sizes. It has been widely adopted owing to its interpretability and simplicity, especially in exploratory microbiome studies. Zhou *et al.*[50] applied LEfSe analysis to different intestinal sites and identified site-specific bacterial biomarkers associated with the IBD subtypes. *Fusobacterium* was enriched in CD, whereas *Bifidobacteriaceae* and *Enterococcus* were more abundant in UC. Furthermore, this study provided a detailed comparison of microbial signatures across lesion sites, revealing that microbial biomarkers vary according to anatomical location in both UC and CD, thereby highlighting the importance of spatial resolution in microbiome-based diagnostics. However, LEfSe does not adjust for covariates or the compositional structure of microbiome data, which limits its robustness in clinical applications.[45]

To address these limitations, more advanced models such as Microbiome Multivariable Association with Linear Models (MaAsLin2) have been developed. MaAsLin2 was developed as a generalized linear modeling framework that accommodates both fixed and random effects, allowing for the adjustment of key confounding variables, such as age, medications, and sequencing batch.[48] MaAsLin3 further extends this functionality to multi-omics integration, supporting both continuous and categorical outcomes and handling relative or absolute abundance data.[51] In a landmark study published in *Nature Medicine*, Zheng *et al.*[52] applied MaAsLin2 to identify microbial features from metagenomic data that distinguished patients with IBD from healthy controls, yielding biomarkers such as *R. gnavus*, *E. coli*, and *Clostridium innocuum*. These were subsequently validated using ML classifiers, as discussed below. Despite its strengths, MaAsLin2 requires careful normalization and model specifications to avoid overfitting or false discoveries, particularly in high-dimensional or compositional datasets.[48,53]

Analysis of Composition of Microbiomes with Bias Correction (ANCOM-BC) offers a robust alternative by directly addressing the compositional constraints inherent in microbiome data.[54] By estimating and correcting for sampling bias, ANCOM-BC produces bias-adjusted abundance estimates and provides false discovery rate (FDR)-controlled results. This method exhibits improved sensitivity in detecting differentially abundant features, particularly those present in low abundance. In a recent study using ANCOM-BC, 77 genera were identified as differentially abundant in patients with CD and 64 genera in patients with UC.[55] *C. innocuum* and *Enterococcus* were significantly enriched in both CD and UC, whereas *Dialister* and *Fusobacterium* showed higher abundances, specifically in CD. In contrast, *Bifidobacterium* was enriched in the UC group. Genera that were depleted in both IBD subtypes included *Akkermansia* and *Roseburia*. *Faecalibacterium* was significantly reduced only in patients with CD, whereas *Bacteroides* and *Odoribacter* were specifically decreased in patients with UC. These findings highlight both shared and distinct microbial signatures across IBD subtypes.

Despite its advantages, ANCOM-BC has limitations.[54] Although this effectively controls FDR and corrects for compositional bias, its computational complexity increases substantially with large datasets, particularly those containing thousands of features, making it challenging in resource-limited environments. Moreover, ANCOM-BC does not support flexible multivariate modeling for complex confounders in the same manner as tools such as MaAsLin2.[56] Although extended versions, such as ANCOM-BC2, have improved this limitation,[57] the need for method-aware applications remains critical.

Given the strengths and limitations of each DA approach, it is essential to select tools that align with the specific characteristics of the study design, dataset, and analytical objectives. Moreover, rather than relying on a single method, using multiple complementary DA tools and prioritizing microbial features that are consistently identified across analyses can enhance the robustness and generalizability of biomarker

discovery. This multi-angle strategy mitigated method-specific bias and strengthened confidence in the clinical relevance of the resulting microbiome-based signatures.[58]

2) Computational strategies for biomarker discovery (ML)

As microbiome datasets increase in size and complexity, traditional univariate approaches are insufficient for robust biomarker discovery. ML techniques have emerged as powerful alternatives for classification and feature selection, particularly in disease diagnosis.[59] These methods excel at modeling high-dimensional, nonlinear data and can integrate heterogeneous features, including microbial, metabolomic, and clinical variables.[60]

Random forest (RF) classifiers are among the most frequently used ML algorithms for microbiome research. They built ensembles of decision trees to improve the classification performance and are often evaluated using metrics such as the area under the receiver operating characteristic (ROC) curve with area under the curve (AUC) value, specificity, and sensitivity.[35,61] In the context of IBD, several studies have successfully applied ML models trained on microbial profiles to distinguish patients from healthy individuals and differentiate UC from CD with high accuracy. For example, Halfvarson *et al.*[61] used RF models to analyze microbiome and clinical data collected from patients with IBD over a 2-year period to assess subtype classification performance. The models were trained using operational taxonomic units (OTUs) abundances, clinical metadata (body mass index, sex, fecal calprotectin), and a unique "distance to healthy plane" metric derived from ordination space. The classifier achieved an overall prediction accuracy of 66.6% for distinguishing IBD subtypes from healthy controls, and the "distance to healthy plane" was identified as a key predictive feature. In another study, Manandhar *et al.*[62] applied LEfSe analysis to identify significantly different microbial taxa between IBD and non-IBD samples using 16S rRNA gene sequencing data, followed by the implementation of various supervised ML algorithms, including RF, decision tree, elastic net, support vector machines with radial kernel, and neural networks, for disease classification. Among these, the RF model exhibited the best performance, achieving AUCs of >0.80 and consistently identifying microbial taxa of diagnostic relevance. Furthermore, to distinguish between CD and UC, LEfSe was used to select 117 bacterial taxa. When evaluated using a trained RF classifier, this feature set yielded high classification accuracy for CD and UC, with AUC values exceeding 0.91. In their multicohort study, Zheng *et al.*[52] validated MaAsLin2-based microbial features using ML modeling to construct a noninvasive diagnostic classifier for IBD. The classifier showed strong performance in differentiating patients with UC and CD from non-IBD controls in the discovery cohort (AUCs >0.9). When applied to an independent validation cohort, the models maintained AUCs above 0.80 for both UC and CD versus non-IBD. In addition, when evaluated on a geographically and clinically diverse international multi-disease dataset, the models achieved AUCs of 0.7830 and 0.7206 for CD and UC, respectively, supporting their generalizability. These results highlight the clinical potential of combining statistical feature selection with the robust ML-based validation of microbiome-based diagnostics.

However, the performance and stability of ML-based biomarker discovery can be strongly influenced by cohort-specific effects and study heterogeneity, underscoring the need for cross-cohort integration strategies. The MMUPHin framework provides a statistical methodology for cross-cohort data integration and meta-analysis of microbiome studies.[63] By correcting for batch effects and study-specific heterogeneity, MMUPHin facilitates the identification of consistent biological patterns that are reproducible across independent datasets, thereby enhancing the reliability of microbiome-based biomarkers. Together with cross-cohort integration strategies such as MMUPHin, methodological rigor at the level of feature selection and validation remains equally important.

Although these studies revealed the utility of ML in microbiome-based diagnostics, many relied on a single DA method or ML-driven feature selection without comparative validation. As discussed above, integrating the results from multiple DA tools to identify robust biomarkers before ML modeling may enhance prediction accuracy and biological relevance. Moreover, unsupervised approaches such as microbial co-occurrence network analysis can highlight key hub taxa that may serve as important diagnostic features when validated through supervised learning models.[64] Thus, multistep pipelines combining DA, network inference, and ML represent promising directions for robust biomarker discovery and clinical translation in IBD (Fig. 1).

3. Integration of multi-omics analysis

Taxonomic profiling of the gut microbiota has provided valuable insights into the microbial alterations associated with IBD; however, a compositional perspective alone cannot fully capture the functional dynamics of the gut ecosystem. To address these limitations, multi-omics approaches have emerged as powerful frameworks for linking microbial presence and activity to host responses.[65] By integrating multiple disciplines, including metabolomics, transcriptomics, and proteomics, researchers are now able to elucidate what microbes are as well as what they do and how they interact with the host.[66] This multilevel perspective is critical for developing mechanistic and clinically

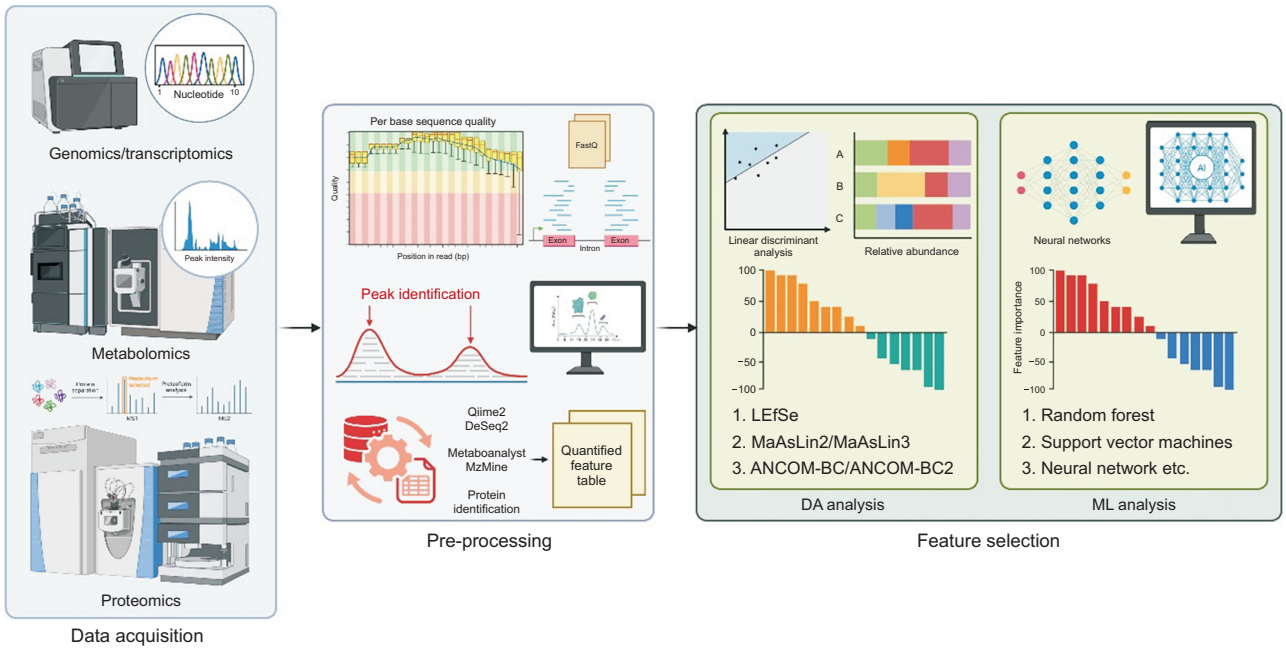

Fig. 1. Workflow for biomarker identification using a multi-omics dataset through differential abundance (DA) and machine learning (ML) approaches. Multi-omics datasets, including genomic, transcriptomic, metabolomic, and proteomic profiles, were first acquired and preprocessed to generate quantified feature tables. DA analysis methods (e.g., LEfSe, MaAsLin2/3, ANCOM-BC2) were then applied to identify candidate features. In parallel, ML approaches (e.g., random forest, support vector machines, neural networks) were used for feature selection. LEfSe, Linear Discriminant Analysis Effect Size; MaAsLin, Microbiome Multivariable Association with Linear Models; ANCOM-BC, Analysis of Composition of Microbiomes with Bias Correction. The figure was created in BioRender.

meaningful microbiota-based diagnostics.

Beyond single-omics analyses, recent frameworks have been designed to integrate multiple data layers. The MintTea framework combines canonical correlation analysis with consensus-based integration to extract a small number of disease-relevant core modules from multi-omics datasets.[67] This approach achieves comparable predictive performance with fewer features, thereby improving interpretability and robustness. As such, MintTea represents a promising strategy for translating complex multi-omics signatures into clinically meaningful biomarkers and further highlights the value of systematic integration strategies.

1) Metabolomic analysis

Metabolomics enables the profiling of small-molecule metabolites produced by both microbial and host metabolisms. In IBD, fecal and serum metabolomic analyses have revealed consistent alterations in pathways related to SCFAs, bile acids, tryptophan metabolism, and lipid signaling.[68] Reduced levels of SCFAs such as butyrate, which are known to support epithelial integrity and anti-inflammatory responses, have been frequently reported in patients with IBD. Similarly, the dysregulation of bile acid composition, particularly the ratio of primary to secondary bile acids, reflects disrupted microbial enzymatic activity and contributes to mucosal inflammation. These metabolites reflect microbial function as well as serve as potential functional biomarkers that may precede or parallel shifts in taxonomic composition. Their measurement provides a more immediate readout of microbiome activity and may be particularly useful for identifying subclinical disease states or monitoring therapeutic responses. Moreover, combining metabolomic data with metagenomic profiles enables the reconstruction of microbe–metabolite interaction networks that may be disease-specific.

Recent multicohort integration studies have used metabolomic data to enhance disease classification. For instance, Kim *et al.*[69] conducted untargeted and targeted serum metabolomics in a large Korean IBD cohort (n=346) and identified diagnostic panels capable of distinguishing patients with IBD from healthy controls and differentiating CD from UC. Their analysis revealed distinct alterations in tryptophan and bile acid metabolism, including elevated indole-3-acetic acid levels and reduced primary-to-secondary bile acid ratios. Using Boruta-based feature selection and ROC curve analysis, they reported high diagnostic accuracy (AUC >0.97 for CD and UC vs control, and AUC=0.714 for UC vs CD), highlighting the utility of serum metabolomic markers in both diagnosis and molecular subtyping of IBD. Ning *et al.*[35] showed that combin-

ing microbiome and metabolome features across cohorts improves the robustness of IBD-specific signatures and facilitates cross-study reproducibility. They reported that various organic acids and carnitine compounds were enriched in patients with IBD and that ML models based on these metabolomic features achieved high diagnostic performance in distinguishing patients with IBD from controls. These findings allowed the identification of robust diagnostic markers that generalized well across independent datasets. Integrating these features with metagenomic data enables the construction of microbe–metabolite interaction networks, thereby providing a mechanistic basis for identifying functionally relevant microbial biomarkers.

2) Transcriptomic analysis

Transcriptome analysis provides a dynamic snapshot of gene expression, allowing researchers to capture the active biological processes that occur in both the host and microbial components of the gut ecosystem. Metatranscriptomics, which involves sequencing total RNA from microbial communities, enables the identification of functionally active microbes and their expressed pathways.[70] For instance, metatranscriptomic analyses have revealed elevated microbial expression of pro-inflammatory genes related to oxidative stress responses, nutrient acquisition, and mucin degradation during active disease phases.[71,72] The integration of microbial and host transcriptomes enables the exploration of cross-kingdom regulatory dynamics, such as how microbiota-derived metabolites modulate host immune gene expression, thus providing a comprehensive understanding of the functional landscape of IBD pathogenesis and aiding in the development of composite molecular biomarkers.[73,74]

3) Proteomic analysis

Although technically more demanding than other omics approaches, proteomics provides a direct measure of protein abundance and post-translational modifications, complementing transcriptomic insights and enabling the functional validation of candidate biomarkers.[75] Recent advancements in mass spectrometry, such as label-free quantification and Matrix-Assisted Laser Desorption/Ionization Time-of-Flight mass spectrometry, have significantly improved proteome depth and quantification accuracy in both host and microbial samples.[76] Host proteomic studies have revealed dysregulated proteins involved in mucosal healing, immune modulation, and extracellular matrix remodeling in IBD.[77-79]

On the microbial side, proteomics, especially metaproteomics, has gained traction for linking microbial functions to disease phenotypes. Zhang *et al.*[80] performed metaproteomic profiling of pediatric IBD samples and revealed associations between microbiome composition and extracellular vesicle protein content in the gut, revealing that microbial proteins may influence host signaling pathways. However, they did not include a follow-up validation, highlighting the gap in translational utility.

Despite its promise, proteomics remains underutilized in microbiome-integrated IBD research compared with other omics layers. However, its role in validating candidate biomarkers and uncovering post-transcriptional regulation is increasingly being appreciated. Protein-level markers are particularly amenable to clinical translation, as they are more accessible for diagnostic development using antibody-based assays. A recent study by Lee *et al.*[81] integrated proteomic, metagenomic, and clinical datasets to identify microbial signatures that predict responses to anti-tumor necrosis factor therapy. The authors showed that baseline differences in microbial protein expression profiles were associated with treatment outcomes, indicating that proteomic data could improve patient stratification and therapeutic decision-making for IBD.

As reviewed above, previous studies have used diverse sequencing platforms, biomarker analysis methods, and strategies to integrate multi-omics data. To provide a structured comparison, we summarize and categorize the key features of the representative studies in Table 2.[82-94] This comparative table enables the examination of the methodological strategies used in each study to identify robust and generalizable biomarkers, along with the validation efforts and the extent to which multi-omics approaches were used.

DISCUSSION

Recent multi-omics studies have provided important lessons for multidimensional biomarker discovery. The integration of metagenomic, metabolomic, transcriptomic, and proteomic data has transformed our understanding of the gut microbiota in IBD, enabling the identification of biomarkers with both mechanistic and diagnostic significance. As reported by recent studies, multi-omics integration enhances the resolution of microbial profiling as well as enables the functional interpretation of disease-associated shifts and their impact on host biology.

A perspective study by Metwaly and Haller[95] emphasized that "missing links" in IBD biomarker discovery could be bridged by combining multi-omics datasets within a biologically informed framework. Recent advances in ML have made it feasible to construct composite models that integrate microbial taxa, microbial metabolites (such as SCFAs and bile acids), host gene expression profiles, and protein-level biomarkers. These models have shown improved performance over single-omics approaches, yield-

Table 2. Overview of Methodological Characteristics in Selected IBD Microbiome Studies

Reference	Country	No. of cohorts	Sequencing strategy	Taxa comparison method	Biomarker validation	Integration of multi-omics
Naive relative abundance comparison						
Rojas-Feria *et al.*[82]	Spain	16 Control, 13 IBD (with CD)	454 pyrosequencing using 16S rRNA gene (V1–V3 region)	Relative abundance	NA	Integration with miRNA data
Altomare *et al.*[83]	Italy	11 Control, 14 IBD (10 CD, 4 UC)	454 pyrosequencing using 16S rRNA gene (V1–V3 region)	Relative abundance	Swets classification	NA
Shaw *et al.*[84]	USA	18 Control, 49 IBD (19 CD, 30 UC)	Illumina MiSeq using 16S rRNA gene (V3–V4 region)	Relative abundance	Random forest	NA
Scaldaferri *et al.*[85]	Italy	37 Control, 39 IBD (with UC)	Illumina MiSeq using 16S rRNA gene (V3–V4 region)	Relative abundance	NA	NA
Ding *et al.*[86]	UK	86 IBD (76 CD, 10 UC)	Illumina MiSeq using 16S rRNA gene (V3 region)	Relative abundance	NA	Integration with a metabolomic data set
Chen *et al.*[87]	China	44 Control, 52 IBD (27 CD-A, 25 CD-R)	Illumina MiSeq using 16S rRNA gene (V5–V6 region)	Relative abundance	Random forest	Investigate short-chain fatty acids profiles
Lo Sasso *et al.*[88]	Russia	40 Control, 71 IBD (35 CD, 36 UC)	Metagenome shotgun sequencing	Relative abundance	NA	Investigate short-chain fatty acids profiles
DA analysis						
Scanu *et al.*[89]	Italy	37 Control, 53 IBD (with UC)	Illumina MiSeq using 16S rRNA gene (V3–V4 region)	LEfSe	NA	Integration with mycobiome and metabolomic data
Han *et al.*[90]	China	24 Control, 80 IBD (67 CD, 13 UC)	Illumina MiSeq using 16S rRNA gene (V3–V4 region)	LEfSe	NA	NA
Toto *et al.*[91]	Italy	Control (NA), 35 IBD (21 CD, 14 UC)	Illumina MiSeq using 16S rRNA gene (V3–V4 region)	LEfSe	NA	NA
DeSantis *et al.*[92]	USA	4 Control, 22 IBD (10 CD, 12 UC)	Illumina MiSeq using 16S rRNA gene (V4 region)	MaAsLin3	NA	Integration with transcriptomic and metabolomic data
Machine learning-based feature selection						
Hodgkiss and Acharjee[93]	American cohort [PRISM]	56 Control, 164 IBD	Metagenome shotgun sequencing	XGBoost, random forest, and LASSO	Random forest	Integration with metabolomic data
Bushman *et al.*[94]	USA	18 Control, 27 IBD	Metagenome shotgun sequencing	Random forest	NA	Integration with metabolomic data

IBD, inflammatory bowel disease; CD, Crohn's disease; NA, not applicable; UC, ulcerative colitis; CD-A, active CD; CD-R, quiescent CD; LEfSe, Linear discriminant analysis Effect Size; MaAsLin, Microbiome Multivariable Associations with Linear Models; PRISM, the Prospective Registry in IBD Study at MGH); LASSO, least absolute shrinkage and selection operator.

ing higher AUCs and greater sensitivity and specificity in distinguishing between disease states and subtypes.

However, the potential of multidimensional diagnostics poses several challenges.[45] Microbiome datasets are inherently sparse, with a large proportion of zero values, and are constrained by their compositional nature, which can generate spurious correlations if not appropriately handled. Furthermore, results are highly sensitive to batch effects and technical variability, requiring coordinated preprocessing, normalization, and validation strategies across omics platforms. Variability introduced by study design and population differences further complicates reproducibility.[9,96,97] Distinct microbial signatures can also emerge depending on whether samples are collected from stool, mucosal biopsies, or oral sites, reflecting both local and systemic host–microbe interactions. Laboratory protocols, sequencing platforms, and analytic pipelines may contribute additional heterogeneity, sometimes exceeding disease-associated signals. Moreover, ethnic and geographic factors have been shown to influence microbial signatures, with divergent biomarkers reported across Western and Asian IBD cohorts.[98,99] In addition, integration of multi-omics layers introduces further complexity, necessitating strategies for appropriate weighting across datasets, handling of missing data, and ensuring computational scalability for clinical deployment.[97] Addressing these sources of variability through standardized sampling strategies, harmonized workflows, and cross-population validation will be essential for translating multi-omics insights into clinically reliable tools.

Recent large-scale studies have begun to demonstrate

the clinical utility of microbiome-based diagnostics. Zheng *et al.*[52] provided the first real-world evidence that a microbiome-informed classifier outperformed fecal calprotectin for distinguishing patients of IBD from controls, validating the multidimensional approach as clinically transformative. Such findings underscore the potential of microbiome signatures to extend beyond proof-of-concept and into practical applications. In the clinical setting, stool-derived microbial biomarkers could support the early detection of IBD, including the identification of at-risk individuals or patients with ambiguous symptoms prior to definitive diagnosis. They may also facilitate more accurate disease subtyping, differentiating UC from CD and enabling patient stratification for optimized therapeutic choices. In addition, longitudinal profiling of microbial signatures can be leveraged to monitor treatment response and predict relapse, potentially reducing the frequency of follow-up endoscopies. Finally, stool-based multidimensional biomarkers offer a noninvasive alternative to colonoscopy or tissue biopsy, lowering patient burden while improving feasibility for large-scale monitoring. Collectively, these advances illustrate how microbiome-informed diagnostics may complement existing clinical tools and accelerate their integration into precision medicine frameworks.

Furthermore, the clinical implementation of microbiome-based diagnostics presents both opportunities and challenges. On the one hand, noninvasive stool-derived biomarkers may complement or even reduce invasive procedures such as colonoscopy, thereby improving patient comfort and enabling longitudinal monitoring. On the other hand, incidental detection of potentially pathogenic species such as enteropathogenic/enterotoxigenic *E. coli* (EPEC and ETEC), *Clostridioides difficile*, or *Fusobacterium nucleatum* can create interpretive dilemmas if not carefully contextualized with patient symptoms and clinical history. In routine 16S rRNA gene–based biomarker discovery, species-level resolution is generally limited and only marginally achievable with full-length sequencing, and even then, additional confirmation through detection of toxin or virulence genes is required. While shotgun metagenomic sequencing provides higher resolution and enables more precise pathotype identification, its substantially higher cost and computational burden currently limit widespread clinical application. This trade-off highlights the need for careful selection of sequencing approaches depending on the clinical context and intended diagnostic use. Therefore, successful clinical translation will require not only robust microbiome signatures but also their careful integration with clinical metadata and patient outcomes. Such an approach will help maximize the benefits

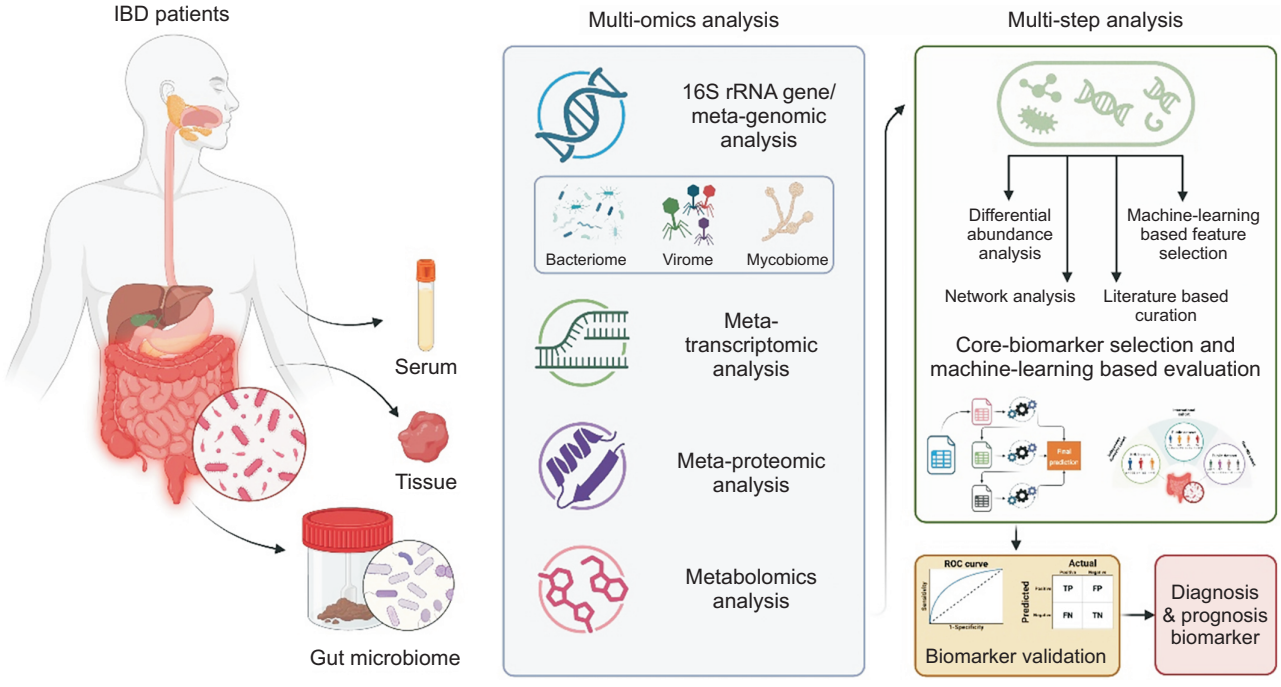

Fig. 2. Overview of a multidimensional biomarker discovery framework for inflammatory bowel disease (IBD) diagnosis and prognosis. Samples from patients with IBD, including oral, serum, tissue, and fecal specimens, were subjected to multi-omics analyses such as 16S/metagenomic profiling, metatranscriptomics, metaproteomics, and metabolomics. These data were integrated using a multistep analytical strategy involving differential abundance testing, network analysis, machine learning-based feature selection, and literature-based curation. The selected core biomarkers were evaluated using supervised modeling (such as receiver operating characteristic and accuracy) for diagnostic and prognostic utility in IBD. The figure was created in BioRender.

of noninvasive monitoring while minimizing unintended harms, ultimately aligning microbiome-informed diagnostics with the goals of precision medicine in IBD care.

CONCLUSIONS

In summary, multidimensional microbiome studies based on integrated omics and computational analyses have revolutionized the diagnosis of IBD (Fig. 2). Continuous interdisciplinary collaboration is essential to advance these tools from the laboratory to the clinic, enabling precision medicine approaches tailored to the microbial and molecular environments of individual patients.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2023-KH135855) and the InnoCORE program of the Ministry of Science and ICT (N10250153).

AUTHOR CONTRIBUTIONS

Study concept and design: J.Y.L. Data acquisition: J.Y.L., J.H.Y., J.E.K. Data analysis and interpretation: J.Y.L., J.H.Y. Drafting of the manuscript: J.Y.L., J.H.Y. Critical revision of the manuscript for important intellectual content: all authors. Statistical analysis: J.Y.L. Obtained funding: J.Y.L., C.K.L. Administrative, technical, or material support: J.W.B., C.K.L. Study supervision: J.W.B., C.K.L. Approval of final manuscript: All authors.

ORCID

June-Young Lee https://orcid.org/0000-0002-6793-6108
Ji-Ho Yoo https://orcid.org/0000-0001-5507-3707
Ji Eun Kim https://orcid.org/0009-0003-9877-6543
Jin-Woo Bae https://orcid.org/0000-0001-6433-5270
Chang Kyun Lee https://orcid.org/0000-0002-4279-3825

DATA AVAILABILITY STATEMENT

Data sharing is not applicable as no new data were created or analyzed in this study.

REFERENCES

1. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A 2007;104:13780-13785.
2. Gevers D, Kugathasan S, Denson LA, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe 2014;15:382-392.
3. Nishida A, Inoue R, Inatomi O, Bamba S, Naito Y, Andoh A. Gut microbiota in the pathogenesis of inflammatory bowel disease. Clin J Gastroenterol 2018;11:1-10.
4. Paramsothy S, Nielsen S, Kamm MA, et al. Specific bacteria and metabolites associated with response to fecal microbiota transplantation in patients with ulcerative colitis. Gastroenterology 2019;156:1440-1454.e2.
5. Zuo T, Kamm MA, Colombel JF, Ng SC. Urbanization and the gut microbiota in health and inflammatory bowel disease. Nat Rev Gastroenterol Hepatol 2018;15:440-452.
6. Vermeire S, Van Assche G, Rutgeerts P. C-reactive protein as a marker for inflammatory bowel disease. Inflamm Bowel Dis 2004;10:661-665.
7. Bjarnason I. The use of fecal calprotectin in inflammatory bowel disease. Gastroenterol Hepatol (N Y) 2017;13:53-56.
8. Clough J, Colwill M, Poullis A, Pollok R, Patel K, Honap S. Biomarkers in inflammatory bowel disease: a practical guide. Therap Adv Gastroenterol 2024;17:17562848241251600.
9. Lloyd-Price J, Arze C, Ananthakrishnan AN, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 2019;569:655-662.
10. Eckburg PB, Bik EM, Bernstein CN, et al. Diversity of the human intestinal microbial flora. Science 2005;308:1635-1638.
11. Buetas E, Jordán-López M, López-Roldán A, et al. Full-length 16S rRNA gene sequencing by PacBio improves taxonomic resolution in human microbiome samples. BMC Genomics 2024;25:310.
12. Rozera T, Pasolli E, Segata N, Ianiro G. Machine learning and artificial intelligence in the multi-omics approach to gut microbiota. Gastroenterology 2025;169:487-501.
13. Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology 2014;146:1489-1499.
14. Scherz V, Greub G, Bertelli C. Building up a clinical micro-

biota profiling: a quality framework proposal. Crit Rev Microbiol 2022;48:356-375.
15. Yang MQ, Wang ZJ, Zhai CB, Chen LQ. Research progress on the application of 16S rRNA gene sequencing and machine learning in forensic microbiome individual identification. Front Microbiol 2024;15:1360457.
16. Allali I, Arnold JW, Roach J, et al. A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. BMC Microbiol 2017;17:194.
17. Hamady M, Knight R. Microbial community profiling for human microbiome projects: tools, techniques, and challenges. Genome Res 2009;19:1141-1152.
18. Andersson AF, Lindberg M, Jakobsson H, Bäckhed F, Nyrén P, Engstrand L. Comparative analysis of human gut microbiota by barcoded pyrosequencing. PLoS One 2008;3:e2836.
19. Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 2012;13:R79.
20. Malapelle U, Vigliar E, Sgariglia R, et al. Ion Torrent next-generation sequencing for routine identification of clinically relevant mutations in colorectal cancer patients. J Clin Pathol 2015;68:64-68.
21. Gryaznova MV, Solodskikh SA, Panevina AV, et al. Study of microbiome changes in patients with ulcerative colitis in the Central European part of Russia. Heliyon 2021;7:e06432.
22. Loman NJ, Misra RV, Dallman TJ, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 2012;30:434-439.
23. Salipante SJ, Kawashima T, Rosenthal C, et al. Performance comparison of Illumina and Ion Torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. Appl Environ Microbiol 2014;80:7583-7591.
24. McDonald D, Jiang Y, Balaban M, et al. Greengenes2 unifies microbial data in a single reference tree. Nat Biotechnol 2024;42:715-718.
25. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013;41:D590-D596.
26. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 2019;37:852-857.
27. Kim HS, Oh SJ, Kim BK, et al. Dysbiotic signatures and diagnostic potential of gut microbial markers for inflammatory bowel disease in Korean population. Sci Rep 2024;14:23701.
28. He XX, Li YH, Yan PG, et al. Relationship between clinical features and intestinal microbiota in Chinese patients with ulcerative colitis. World J Gastroenterol 2021;27:4722-4737.
29. Jeong J, Yun K, Mun S, et al. The effect of taxonomic classification by full-length 16S rRNA sequencing with a synthetic long-read technology. Sci Rep 2021;11:1727.
30. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 2017;35:833-844.
31. Nishiwaki H, Ueyama J, Ito M, et al. Meta-analysis of shotgun sequencing of gut microbiota in Parkinson's disease. NPJ Parkinsons Dis 2024;10:106.
32. Vich Vila A, Imhann F, Collij V, et al. Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome. Sci Transl Med 2018;10:eaap8914.
33. Fang X, Monk JM, Nurk S, et al. Metagenomics-based, strain-level analysis of Escherichia coli from a time-series of microbiome samples from a Crohn's disease patient. Front Microbiol 2018;9:2559.
34. Hu Y, Yang X, Qin J, et al. Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. Nat Commun 2013;4:2151.
35. Ning L, Zhou YL, Sun H, et al. Microbiome and metabolome features in inflammatory bowel disease via multi-omics integration analyses across cohorts. Nat Commun 2023;14:7135.
36. Orejudo M, Gómez MJ, Riestra S, et al. Exploration of fecal microbiota in newly diagnosed patients with inflammatory bowel disease using shotgun metagenomics. Front Cell Infect Microbiol 2025;15:1595884.
37. Cheng WY, Liu WX, Ding Y, et al. High sensitivity of shotgun metagenomic sequencing in colon tissue biopsy by host DNA depletion. Genomics Proteomics Bioinformatics 2023;21:1195-1205.
38. Cattonaro F, Spadotto A, Radovic S, Marroni F. Do you cov me? Effect of coverage reduction on metagenome shotgun sequencing studies. F1000Res 2018;7:1767.
39. Ungaro F, Massimino L, D'Alessio S, Danese S. The gut virome in inflammatory bowel disease pathogenesis: from metagenomics to novel therapeutic approaches. United European Gastroenterol J 2019;7:999-1007.
40. Zhang F, Aschenbrenner D, Yoo JY, Zuo T. The gut mycobiome in health, disease, and clinical applications in association with the gut bacterial microbiome assembly. Lancet Microbe 2022;3:e969-e983.
41. Sokol H, Leducq V, Aschard H, et al. Fungal microbiota dysbiosis in IBD. Gut 2017;66:1039-1048.
42. Norman JM, Handley SA, Baldridge MT, et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell 2015;160:447-460.
43. Hoarau G, Mukherjee PK, Gower-Rousseau C, et al. Bacteriome and mycobiome interactions underscore microbial dysbiosis in familial Crohn's disease. mBio 2016;7:e01250-16.
44. Higashi S, Barreto Ada M, Cantão ME, de Vasconcelos AT. Analysis of composition-based metagenomic classification. BMC Genomics 2012;13(Suppl 5):S1.
45. Nearing JT, Douglas GM, Hayes MG, et al. Microbiome dif-

ferential abundance methods produce different results across 38 datasets. Nat Commun 2022;13:342.
46. Weiss S, Xu ZZ, Peddada S, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 2017;5:27.
47. Wallen ZD. Comparison study of differential abundance testing methods using two large Parkinson disease gut microbiome datasets derived from 16S amplicon sequencing. BMC Bioinformatics 2021;22:265.
48. Mallick H, Rahnavard A, McIver LJ, et al. Multivariable association discovery in population-scale meta-omics studies. PLoS Comput Biol 2021;17:e1009442.
49. Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. Genome Biol 2011;12:R60.
50. Zhou Y, He Y, Liu L, et al. Alterations in gut microbial communities across anatomical locations in inflammatory bowel diseases. Front Nutr 2021;8:615064.
51. Nickols WA, Kuntz T, Shen J, et al. MaAsLin 3: refining and extending generalized multivariable linear models for meta-omic association discovery. bioRxiv [Preprint]. 2024 [cited 2025 Jul 29]. Available from: https://doi.org/10.1101/2024.12.13.628459
52. Zheng J, Sun Q, Zhang M, et al. Noninvasive, microbiome-based diagnosis of inflammatory bowel disease. Nat Med 2024;30:3555-3567.
53. Yang L, Chen J. Benchmarking differential abundance analysis methods for correlated microbiome sequencing data. Brief Bioinform 2023;24:bbac607.
54. Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. Nat Commun 2020;11:3514.
55. Vestergaard MV, Allin KH, Eriksen C, et al. Gut microbiota signatures in inflammatory bowel disease. United European Gastroenterol J 2024;12:22-33.
56. Yang L, Chen J. A comprehensive evaluation of microbial differential abundance analysis methods: current status and potential solutions. Microbiome 2022;10:130.
57. Lin H, Peddada SD. Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures. Nat Methods 2024;21:83-91.
58. Ananthakrishnan AN. Precision medicine in inflammatory bowel diseases. Intest Res 2024;22:8-14.
59. Pasolli E, Truong DT, Malik F, Waldron L, Segata N. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. PLoS Comput Biol 2016;12:e1004977.
60. Knights D, Costello EK, Knight R. Supervised classification of human microbiota. FEMS Microbiol Rev 2011;35:343-359.
61. Halfvarson J, Brislawn CJ, Lamendella R, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol 2017;2:17004.
62. Manandhar I, Alimadadi A, Aryal S, Munroe PB, Joe B, Cheng X. Gut microbiome-based supervised machine learning for clinical diagnosis of inflammatory bowel diseases. Am J Physiol Gastrointest Liver Physiol 2021;320:G328-G337.
63. Ma S, Shungin D, Mallick H, et al. Population structure discovery in meta-analyzed microbial communities and inflammatory bowel disease using MMUPHin. Genome Biol 2022;23:208.
64. Kim S, Thapa I, Ali H. A novel computational approach for the mining of signature pathways using species co-occurrence networks in gut microbiomes. BMC Microbiol 2024;24(Suppl 1):490.
65. Zhou W, Sailani MR, Contrepois K, et al. Longitudinal multi-omics of host-microbe dynamics in prediabetes. Nature 2019;569:663-671.
66. Chetty A, Blekhman R. Multi-omic approaches for host-microbiome data integration. Gut Microbes 2024;16:2297860.
67. Muller E, Shiryan I, Borenstein E. Multi-omic integration of microbiome data for identifying disease-associated modules. Nat Commun 2024;15:2621.
68. Guggeis MA, Harris DM, Welz L, Rosenstiel P, Aden K. Microbiota-derived metabolites in inflammatory bowel disease. Semin Immunopathol 2025;47:19.
69. Kim JE, Suh DH, Park YJ, et al. Identifying robust biomarkers for the diagnosis and subtype distinction of inflammatory bowel disease through comprehensive serum metabolomic profiling. Sci Rep 2025;15:5661.
70. Schirmer M, Garner A, Vlamakis H, Xavier RJ. Microbial genes and pathways in inflammatory bowel disease. Nat Rev Microbiol 2019;17:497-511.
71. Zhang Y, Bhosle A, Bae S, et al. Discovery of bioactive microbial gene products in inflammatory bowel disease. Nature 2022;606:754-760.
72. Valles-Colomer M, Darzi Y, Vieira-Silva S, Falony G, Raes J, Joossens M. Meta-omics in inflammatory bowel disease research: applications, challenges, and guidelines. J Crohns Colitis 2016;10:735-746.
73. Pavlidis P, Tsakmaki A, Treveil A, et al. Cytokine responsive networks in human colonic epithelial organoids unveil a molecular classification of inflammatory bowel disease. Cell Rep 2022;40:111439.
74. Smillie CS, Biton M, Ordovas-Montanes J, et al. Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. Cell 2019;178:714-730.e22.
75. Yau Y, Leong RW, Zeng M, Wasinger VC. Proteomics and metabolomics in inflammatory bowel disease. J Gastroenterol Hepatol 2013;28:1076-1086.
76. Park JM, Han NY, Han YM, et al. Predictive proteomic biomarkers for inflammatory bowel disease-associated cancer: where are we now in the era of the next generation pro-

teomics? World J Gastroenterol 2014;20:13466-13476.
77. Bennike T, Birkelund S, Stensballe A, Andersen V. Biomarkers in inflammatory bowel diseases: current status and proteomics identification strategies. World J Gastroenterol 2014;20:3231-3244.
78. Meuwis MA, Fillet M, Geurts P, et al. Biomarker discovery for inflammatory bowel disease, using proteomic serum profiling. Biochem Pharmacol 2007;73:1422-1433.
79. M'Koma AE, Seeley EH, Washington MK, et al. Proteomic profiling of mucosal and submucosal colonic tissues yields protein signatures that differentiate the inflammatory colitides. Inflamm Bowel Dis 2011;17:875-883.
80. Zhang X, Deeke SA, Ning Z, et al. Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease. Nat Commun 2018;9:2873.
81. Lee JWJ, Plichta D, Hogstrom L, et al. Multi-omics reveal microbial determinants impacting responses to biologic therapies in inflammatory bowel disease. Cell Host Microbe 2021;29:1294-1304.e4.
82. Rojas-Feria M, Romero-García T, Fernández Caballero-Rico JÁ, et al. Modulation of faecal metagenome in Crohn's disease: role of microRNAs as biomarkers. World J Gastroenterol 2018;24:5223-5233.
83. Altomare A, Putignani L, Del Chierico F, et al. Gut mucosal-associated microbiota better discloses inflammatory bowel disease differential patterns than faecal microbiota. Dig Liver Dis 2019;51:648-656.
84. Shaw KA, Bertha M, Hofmekler T, et al. Dysbiosis, inflammation, and response to treatment: a longitudinal study of pediatric subjects with newly diagnosed inflammatory bowel disease. Genome Med 2016;8:75.
85. Scaldaferri F, D'Onofrio AM, Calia R, et al. Gut microbiota signatures are associated with psychopathological profiles in patients with ulcerative colitis: results from an Italian tertiary IBD center. Inflamm Bowel Dis 2023;29:1805-1818.
86. Ding NS, McDonald JAK, Perdones-Montero A, et al. Metabonomics and the gut microbiome associated with primary response to anti-TNF therapy in Crohn's disease. J Crohns Colitis 2020;14:1090-1102.
87. Chen W, Li Y, Wang W, et al. Enhanced microbiota profiling in patients with quiescent Crohn's disease through comparison with paired healthy first-degree relatives. Cell Rep Med 2024;5:101624.
88. Lo Sasso G, Khachatryan L, Kondylis A, et al. Inflammatory bowel disease-associated changes in the gut: focus on Kazan patients. Inflamm Bowel Dis 2021;27:418-433.
89. Scanu M, Toto F, Petito V, et al. An integrative multi-omic analysis defines gut microbiota, mycobiota, and metabolic fingerprints in ulcerative colitis patients. Front Cell Infect Microbiol 2024;14:1366192.
90. Han A, Yang M, Chen B, et al. Microbiome and its relevance to indigenous inflammatory bowel diseases in China. Gene 2024;909:148257.
91. Toto F, Marangelo C, Scanu M, et al. A novel microbial dysbiosis index and intestinal microbiota-associated markers as tools of precision medicine in inflammatory bowel disease paediatric patients. Int J Mol Sci 2024;25:9618.
92. DeSantis AH, Buss K, Coker KM, et al. Multiomics-based profiling of the fecal microbiome reveals potential disease-specific signatures in pediatric IBD (PIBD). Biomolecules 2025;15:746.
93. Hodgkiss R, Acharjee A. Unravelling metabolite-microbiome interactions in inflammatory bowel disease through AI and interaction-based modelling. Biochim Biophys Acta Mol Basis Dis 2025;1871:167618.
94. Bushman FD, Conrad M, Ren Y, et al. Multi-omic analysis of the interaction between Clostridioides difficile infection and pediatric inflammatory bowel disease. Cell Host Microbe 2020;28:422-433.e7.
95. Metwaly A, Haller D. Multi-omics in IBD biomarker discovery: the missing links. Nat Rev Gastroenterol Hepatol 2019;16:587-588.
96. Kim D, Jung JY, Oh HS, et al. Comparison of sampling methods in assessing the microbiome from patients with ulcerative colitis. BMC Gastroenterol 2021;21:396.
97. Hernández Medina R, Kutuzova S, Nielsen KN, et al. Machine learning and deep learning applications in microbiome research. ISME Commun 2022;2:98.
98. Hao F, Zhu W, Zheng J, Ng SC, Zhang J. Geographic diversity in inflammatory bowel disease genetics and microbiome. Trends Microbiol. Microbiol 2025;33:1196-1211.
99. Misra R, Sarafian M, Pechlivanis A, et al. Ethnicity associated microbial and metabonomic profiling in newly diagnosed ulcerative colitis. Clin Exp Gastroenterol 2022;15:199-212.