



OPEN ACCESS

EDITED BY

Mohamed Ezzat Abdin,
Agricultural Research Center, Egypt

REVIEWED BY

Alejandro Sanchez-Flores,
National Autonomous University of Mexico,
Mexico
Georgina Hernandez-Montes,
National Autonomous University of Mexico,
Mexico

*CORRESPONDENCE

Haiyan Lu
✉ 320272@hnucm.edu.cn

RECEIVED 09 May 2025

ACCEPTED 13 August 2025

PUBLISHED 26 August 2025

CITATION

Li X and Lu H (2025) Enhanced metagenomic
strategies for elucidating the complexities of
gut microbiota: a review.
Front. Microbiol. 16:1626002.
doi: 10.3389/fmicb.2025.1626002

COPYRIGHT

© 2025 Li and Lu. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Enhanced metagenomic strategies for elucidating the complexities of gut microbiota: a review

Xinru Li¹ and Haiyan Lu^{2*}

¹Department of Anorectal Surgery, Changsha Hospital of Traditional Chinese Medicine (Changsha Eighth Hospital), Changsha, China, ²Department of Anorectal Surgery, The Second Affiliated Hospital of Hunan University of Chinese Medicine, Changsha, China

The human gastrointestinal tract (GIT) is inhabited by a heterogeneous and dynamic microbial community that influences host health at multiple levels both metabolically, immunologically and via neurological pathways. Though the gut microbiota—overwhelmingly *Bacteroidetes* and *Firmicutes*—has essential functions in nutrient metabolism, immune regulation, and resistance to pathogens, its dysbiosis is likewise associated with pathologies, such as inflammatory bowel disease (IBD), obesity, type 2 diabetes (T2D), and neurodegenerative diseases. While conventional metagenomic techniques laid the groundwork for understanding microbial composition, next-generation enhanced metagenomic techniques permit an unprecedented resolution in exploring the functional and spatial complexity of gut communities. Advanced frameworks such as high-throughput sequencing, bioinformatic and multi-omics technologies are expanding the understanding of microbial gene regulation, metagenomic pathways, and host-microbe communication. Beyond taxonomic profiling, they map niche-specific activities of gut microbiota along a dichotomy of facultative mutualism, evidenced by relations of beneficial symbionts, represented here by *Enterobacteriaceae*. In this review, we critically consider the latest approaches (e.g., long-read sequencing, single-cell metagenomics and AI-guided annotation) that mitigate biases stemming from DNA extraction, sequencing depth and functional inference.

KEYWORDS

gut microbiome, metagenomics, beneficial and harmful microbes, long-read sequencing, multi-omics integration

1 Introduction

The human gastrointestinal tract (GIT) hosts one of the most intricate microbial ecosystems known to science, comprising bacteria, archaea, fungi, and viruses that collectively influence host health through metabolic, immunological, and neurological pathways (Thursby and Juge, 2017). The gut microbiota dominated by the phyla *Bacteroidetes* and *Firmicutes* which plays indispensable roles in nutrient metabolism, immune regulation, and pathogen resistance (Thursby and Juge, 2017). These microbial communities ferment dietary fibers into short-chain fatty acids (SCFAs) like butyrate and acetate, which regulate intestinal epithelial integrity and systemic immune responses (Shin et al., 2023). However, disruptions in this delicate balance termed dysbiosis are increasingly linked to pathologies such as IBD, obesity, type 2 diabetes (T2D), and neurodegenerative disorders (Mousa and Ali, 2024). The dualistic nature of gut microbiota, wherein symbionts like *Akkermansia muciniphila* promote metabolic health while pathobionts such as *Enterobacteriaceae* drive inflammation, underscores the need

for advanced methodologies to decode microbial dynamics (Mousa and Ali, 2024; Wang et al., 2015).

Traditional metagenomic approaches, pioneered by initiatives like the MetaHIT consortium and the Human Microbiome Project, laid the foundation for cataloging microbial diversity by sequencing 16S rRNA genes and shotgun metagenomes. These efforts revealed over 3.3 million non-redundant genes in the human gut, far exceeding the human genome. Yet, these methods faced limitations: short-read sequencing often fragmented complex genomic regions, while DNA extraction biases skewed taxonomic profiles toward abundant species (Bai et al., 2021). Functional insights remained inferential, relying on homology-based predictions rather than direct measurements of gene expression or metabolic activity (Cozzetto et al., 2016). For instance, early metagenomic studies associated *Faecalibacterium prausnitzii* depletion with IBD but could not clarify whether this reflected causation or correlation (Cozzetto et al., 2016). Similarly, while antibiotic resistance genes (ARGs) were identified in fecal metagenomes, their plasmid-borne mobility and strain-specific distribution required deeper investigation (Xuan et al., 2023).

Emerging enhanced metagenomic strategies now transcend these limitations by integrating high-throughput sequencing, single-cell resolution, and multi-omics frameworks (Forster et al., 2019). Long-read sequencing technologies, such as Oxford Nanopore and PacBio, resolve repetitive genomic elements and structural variations, enabling complete assembly of microbial genomes from complex samples (Forster et al., 2019). This advancement is critical for studying mobile genetic elements like plasmids, which facilitate horizontal gene transfer of ARGs and virulence factors. Complementing this, single-cell metagenomics isolates individual microbial cells, bypassing cultivation biases and revealing genomic blueprints of uncultured taxa (Forster et al., 2019; Tokuda and Shintani, 2024). The Human Gastrointestinal Bacteria Culture Collection (HBC), encompassing 737 whole-genome-sequenced isolates, exemplifies how reference databases enhance taxonomic and functional annotation in metagenomic studies (Mangoma et al., 2024). By mid-2025, such resources have improved subspecies-level classification for nearly 50% of gut microbial sequences, a leap from the 37% genome coverage achieved by earlier projects (Forster et al., 2019; Sugimoto et al., 2019).

These advancements are reshaping translational research. By correlating microbial signatures with clinical outcomes, enhanced metagenomics identifies diagnostic biomarkers and therapeutic targets (Feng et al., 2020). For instance, *Streptococcus anginosus* and *Rothia mucilaginosa* enrichments in HIV-1 patients on antiretroviral therapy correlate with immunodeficiency severity, suggesting microbial markers for treatment monitoring (Forster et al., 2019). Meanwhile, precision editing of gut microbiota through phage therapy or engineered probiotics tailors interventions to individual microbiomes, mitigating adverse drug reactions (Feng et al., 2020).

As the field transitions from observational studies to mechanistic exploration, enhanced metagenomics bridges the gap between microbial taxonomy and host pathophysiology. By resolving strain-level variations, functional pathways, and ecological interactions, these strategies illuminate the gut microbiome's role in health and disease, paving the way for personalized therapies.

2 Gut microbiota: a nexus of health and disease

2.1 Gut microbiota in homeostasis

The human gastrointestinal tract harbors a highly diverse and dynamic microbial ecosystem, predominantly composed of over 1,000 bacterial species, with the phyla *Firmicutes* and *Bacteroidetes* being the most dominant (Bull and Plummer, 2014). This microbiota functions as a critical interface between the host and its environment, playing indispensable roles in nutrient metabolism, immune system modulation, and pathogen resistance. Dietary fibers fermented by these microbes generate short-chain fatty acids (SCFAs) notably acetate, propionate, and butyrate—which reinforce intestinal barrier integrity, modulate systemic immune responses, and suppress inflammation by inducing regulatory T-cell differentiation (Yang and Cong, 2021; Rooks and Garrett, 2016).

Beneficial symbionts such as *Faecalibacterium prausnitzii* and *Akkermansia muciniphila* enhance mucosal immunity by producing anti-inflammatory metabolites like indole derivatives and conjugated linoleic acid (Yang and Cong, 2021). Additionally, commensal microbes maintain ecological balance by competitively excluding pathogens and secreting antimicrobial compounds such as bacteriocins, thereby ensuring gastrointestinal homeostasis (Zeng et al., 2016). The complex interplay between specific microbial taxa, their metabolic byproducts, and their roles in health and disease is summarized in Tables 1, 2.

2.2 Dysbiosis and metabolic-immune dysfunction

Disruptions in microbial equilibrium, termed dysbiosis, can be triggered by factors like high-sugar diets, antibiotic overuse, and reduced fiber intake. Dysbiosis favors pro-inflammatory taxa such as *Enterobacteriaceae* and *Fusobacterium nucleatum*, while depleting protective microbes (Forster et al., 2019; Zeng et al., 2016). These alterations impair the intestinal barrier, allowing microbial products like lipopolysaccharide (LPS) and flagellin to translocate into systemic circulation. Such pathogen-associated molecular patterns (PAMPs) activate Toll-like receptors (TLRs) and NOD-like receptors (NLRs), triggering NF- κ B signaling and systemic low-grade inflammation that underlies metabolic diseases such as obesity and T2D (Potrykus et al., 2021).

In IBD for instance, blooms of *Enterobacteriaceae* are associated with elevated IL-17 production and mucosal damage (Zeng et al., 2016). Similarly, in colorectal cancer (CRC), overabundance of *Bacteroides fragilis* promotes oncogenic Wnt/ β -catenin signaling through polysaccharide A, emphasizing how pathobionts exploit dysbiosis to drive disease progression (Bull and Plummer, 2014; de Vos et al., 2022).

2.3 Systemic impacts: gut-liver, gut-joint, and gut-brain axes

The impact of gut dysbiosis extends to extraintestinal sites via multiple host-microbe interaction axes.

TABLE 1 Role of microorganisms in the health and disease.

Microorganism	Role in the health and disease link	Mechanism	References
<i>Faecalibacterium prausnitzii</i>	<i>Faecalibacterium prausnitzii</i> produces SCFAs, particularly butyrate, which exert anti-inflammatory effects in IBD by enhancing regulatory T cell differentiation and maintaining intestinal barrier integrity	Butyrate enhances T cell differentiation and strengthens intestinal barrier	Louis et al. (2014)
<i>Enterobacteriaceae</i>	Dysbiosis characterized by the overgrowth of <i>Enterobacteriaceae</i> exacerbates IBD by promoting inflammation through lipopolysaccharide (LPS)-mediated immune activation	LPS activates TLR4/NF- κ B, driving IL-17 and tissue damage	Costa et al. (2020)
<i>Clostridium scindens</i>	Development of non-alcoholic fatty liver disease (NAFLD) by producing deoxycholic acid, which inhibits hepatic FXR signaling and promotes lipid accumulation in the liver	Deoxycholic acid inhibits hepatic FXR signaling, promoting steatosis	Arab et al. (2016)
<i>Akkermansia muciniphila</i>	<i>Akkermansia muciniphila</i> contributes to the improvement of metabolic syndrome by degrading mucin, thereby strengthening the gut barrier and enhancing insulin sensitivity	Mucin degradation enhances gut barrier function and insulin sensitivity	Vogt et al. (2018) and Plovier et al. (2016)
<i>Bacteroides fragilis</i>	Gut microbiota-linked colorectal cancer (CRC) refers to the development or progression of colorectal cancer influenced by specific microbial species	Polysaccharide A activates Wnt/ β -catenin signaling in epithelial cells	Dejea et al. (2018)
<i>Lactobacillus rhamnosus</i>	Reduction in anxiety- and depression-like behaviors	GABA synthesis and vagal nerve stimulation regulate serotonin availability	Bravo et al. (2011)
<i>Bifidobacterium longum</i>	Alleviation of symptoms in irritable bowel syndrome (IBS)	Competitive exclusion of pathogens and reinforcement of mucosal defenses	Chen et al. (2023)
<i>Bacteroidetes</i> -enriched communities	Fecal microbiota transplantation-induced remission of <i>C. difficile</i> infection	Restores bile acid metabolism and niche competition against pathogens	van Nood et al. (2013)

TABLE 2 Microbial metabolites and their role in the human health and disease.

Metabolites	Role in health and disease	Mechanism	References
Trimethylamine N-oxide (TMAO)	Neuroinflammation and play important role in Alzheimer's disease	TMAO crosses BBB, triggers microglial activation, and promotes A β aggregation	Vogt et al. (2018)
Resistant starch fermentation	Diet-microbiota crosstalk and improved glycemic control in T2D	Acetate stimulates GPR43, enhancing insulin secretion and β -cell function	Chen et al. (2024)

2.3.1 Gut-liver axis

The gut-liver axis represents a critical communication network influenced by gut microbiota and their metabolites. Altered microbial composition, particularly the enrichment of *Clostridium scindens*, leads to increased production of secondary bile acids such as deoxycholic acid, which disrupts farnesoid X receptor (FXR) signaling in the liver. Impaired FXR activity affects lipid and glucose metabolism, bile acid homeostasis, and promotes hepatic

inflammation and steatosis. These disruptions contribute significantly to the onset and progression of NAFLD. Moreover, gut-derived endotoxins entering the portal circulation exacerbate liver injury by activating inflammatory pathways and fibrogenesis ([Hrncir, 2022](#)).

2.3.2 Gut-joint axis

The gut-joint axis highlights the interplay between gut microbiota and autoimmune joint diseases. In rheumatoid arthritis (RA), an

increased abundance of *Prevotella copri* promotes T-helper 17 (Th17) cell differentiation, leading to heightened systemic inflammation and joint destruction. These immune alterations are driven by microbial antigens that trigger proinflammatory cytokine release, including IL-17 and TNF- α . Similar gut-immune interactions are observed in systemic lupus erythematosus (SLE), where dysbiosis alters mucosal tolerance and promotes autoantibody production. These findings suggest that gut microbiota play a pivotal role in initiating and perpetuating autoimmune responses, making the gut a potential therapeutic target in RA and SLE (Yang and Cong, 2021).

2.3.3 Gut-brain axis

The gut-brain axis represents a bidirectional communication network between the gut microbiota and the central nervous system, mediated through immune, neuroendocrine, and metabolic pathways. Dysbiosis can reduce the availability of serotonin precursors and impair γ -aminobutyric acid (GABA) synthesis, contributing to anxiety and depression. Moreover, microbial metabolites such as TMAO and diminished short-chain fatty acids (SCFAs) exacerbate neuroinflammation and compromise blood-brain barrier integrity, factors implicated in Alzheimer's disease. These disruptions highlight the critical role of gut microbiota in maintaining neurological health and suggest that microbial modulation may offer therapeutic avenues for various neuropsychiatric disorders (Yang and Cong, 2021).

2.4 Cardiovascular and multisystemic consequences

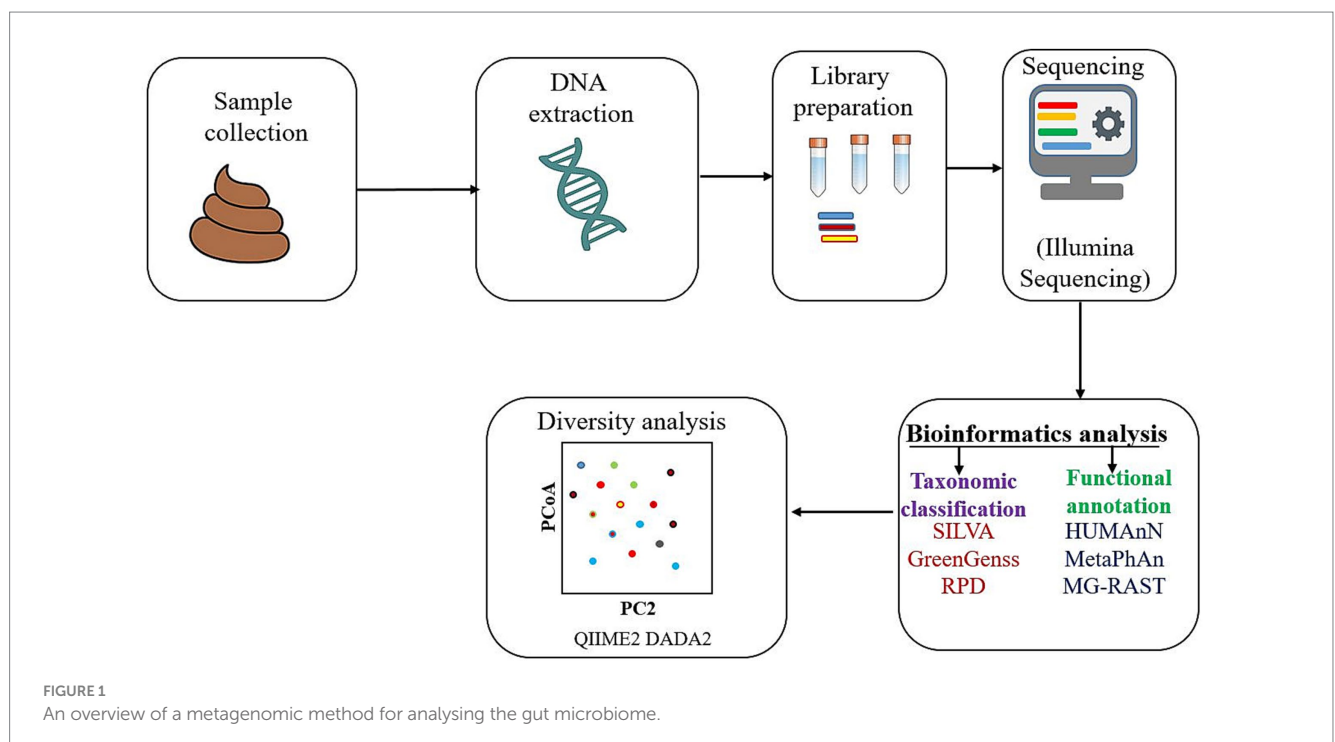
Beyond metabolic and neurological impacts, gut microbial metabolites significantly influence cardiovascular health and multisystemic processes. One such metabolite, phenylacetylglutamine (PAGln), produced by gut microbes from dietary phenylalanine, has

been shown to enhance platelet reactivity, thereby increasing the risk of atherosclerosis and thrombotic events. Additionally, metabolites like TMAO contribute to endothelial dysfunction, inflammation, and lipid metabolism disturbances. These microbial products not only affect cardiovascular physiology but also have systemic effects, linking gut dysbiosis to broader health issues, including chronic kidney disease and metabolic syndrome, underscoring the gut microbiota's central role in maintaining systemic homeostasis and disease susceptibility (Rooks and Garrett, 2016).

3 Evolution of metagenomic methodologies

Before investigating the functional involvement of microbial communities in disease pathophysiology, it is essential to accurately identify and characterize these communities with high sensitivity and specificity. Metagenomics enables such identification through two major culture-independent approaches: 16S rRNA gene amplicon sequencing and shotgun metagenomic sequencing (Jovel et al., 2016; Ranjan et al., 2016; Ghaisas et al., 2016; Bastiaanssen et al., 2018; Bastiaanssen et al., 2018). The 16S rRNA approach is mainly used for taxonomic profiling by amplifying conserved bacterial gene regions, while shotgun metagenomics provides comprehensive insights into both taxonomic composition and functional potential by sequencing total genomic DNA from a sample (Bastiaanssen et al., 2018) (Figure 1).

The analysis workflow generally begins with sample collection, DNA extraction, and library preparation. Sequencing is typically performed on high-throughput platforms such as Illumina due to its greater efficiency and accuracy. The resulting DNA sequences are then analyzed using bioinformatics pipelines. Taxonomic classification is performed by comparing sequence reads to curated databases such as



SILVA, GreenGenes, or RDP. Functional annotations are obtained using tools like HUMAnN, MetaPhlAn, or MG-RAST, which map sequences to gene and pathway databases (Reck et al., 2015). To evaluate differences in microbial composition across conditions or cohorts, diversity metrics such as alpha (within-sample) and beta diversity (between-sample) are calculated. Tools like QIIME2 and DADA2 enable comprehensive statistical analyses, including principal coordinate analysis (PCoA) and differential abundance testing (Reck et al., 2015; Balvočiūtė and Huson, 2017).

3.1 16S rRNA gene sequencing

The 16S rRNA gene is a highly conserved and universally present genetic marker in bacterial and archaeal genomes, making it one of the most widely used tools for microbial taxonomic profiling (Bukin et al., 2019). Cryan et al. (2019) highlighted that 16S rRNA sequencing is a cost-effective, widely accessible method for studying microbial communities, particularly in the gut microbiome of humans, mice, and insects. This method allows for the assessment of microbial diversity and relative abundance using next-generation sequencing technologies.

In this approach, PCR is used to amplify conserved regions of the 16S rRNA gene, and the variable regions are sequenced to distinguish different taxa. The resulting sequences, or amplicons, are grouped into operational taxonomic units (OTUs), typically at 97% sequence similarity (Bukin et al., 2019). Alternatively, amplicon sequence variants (ASVs), generated through denoising algorithms such as DADA2 or Deblur, offer higher resolution and reduced error rates (Prodan et al., 2020).

However, while 16S rRNA is the most commonly used genetic marker for bacterial identification, it is not without limitations. Critically, the 16S rRNA gene is not always present as a single copy within bacterial genomes—some bacteria carry multiple copies with varying sequences. This gene copy number variability can distort estimates of microbial abundance and skew taxonomic profiles, especially when comparing species with different 16S copy numbers. Moreover, the resolution of 16S rRNA sequencing is generally limited to the genus level, often lacking the specificity to accurately identify species (Knight et al., 2018; Callahan et al., 2017).

Different strategies for OTU clustering (e.g., *de novo*, closed-reference) further influence the accuracy and comparability of results (Callahan et al., 2017). Despite these challenges, 16S rRNA sequencing remains a widely adopted approach for large-scale microbiome studies, especially when focusing on bacterial and archaeal populations across various health and disease conditions (Prodan et al., 2020).

3.2 Shotgun metagenomic sequencing

Shotgun metagenomic sequencing provides a comprehensive and accurate depiction of microbial communities by sequencing all DNA present in a sample, surpassing 16S rRNA in species-level resolution and functional potential (Dovrolis et al., 2017; Ranjan et al., 2016). However, this approach is expensive, data-intensive, and technically complex. Crucially, the accuracy and reproducibility of results heavily depend on sample handling, conservation, and preparation

procedures—critical elements often overlooked. Temperature control, cell-size filtration, and DNA extraction methods directly influence microbial representation and community structure, introducing variability across studies.

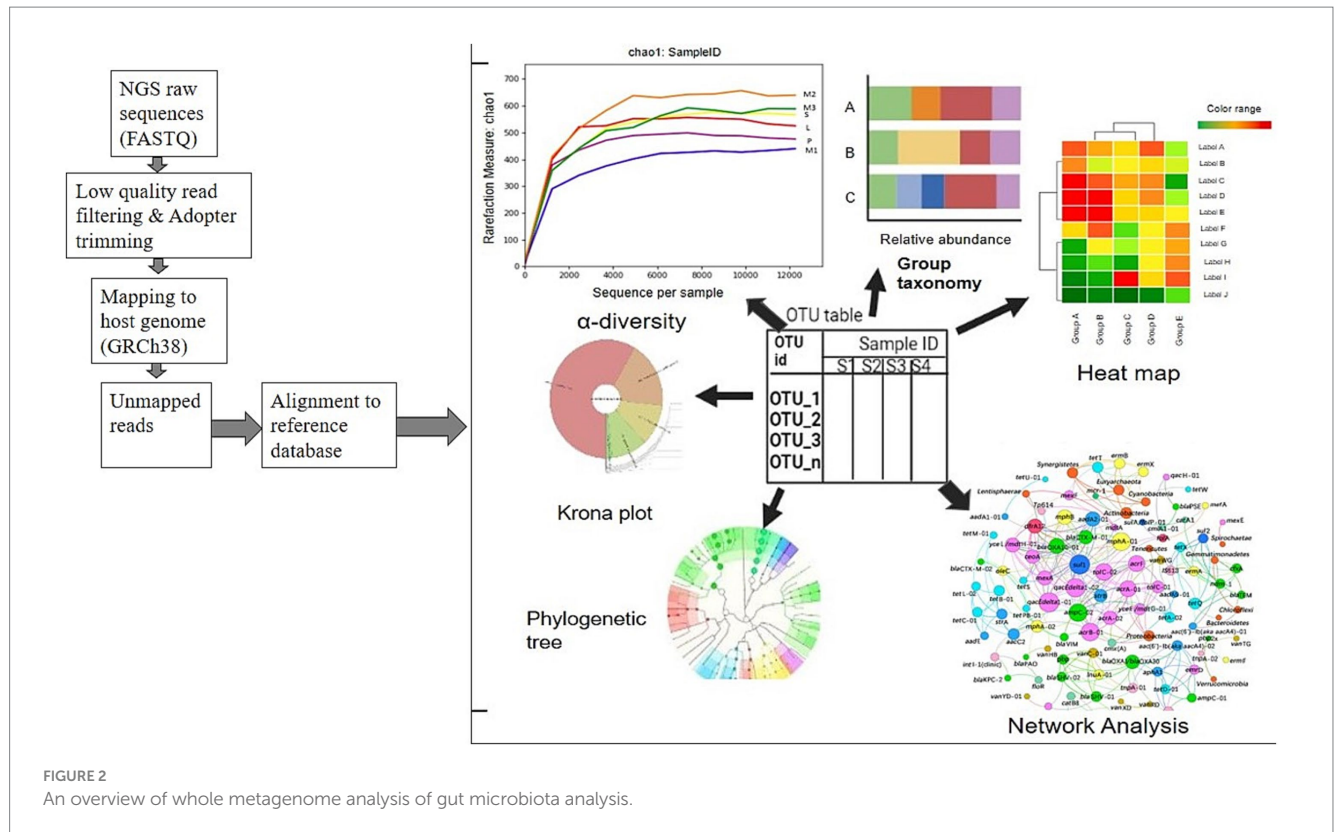
Inconsistent sample processing, such as improper freezing or delayed processing, may degrade DNA, affect microbial viability, or bias detection of certain taxa. Moreover, variations in DNA extraction protocols can result in differential lysis efficiency, leading to underrepresentation of key microbial groups, particularly Gram-positive species. Filtration by cell size, if improperly performed, may skew microbial diversity by excluding small-sized microbes or including host DNA (Ranjan et al., 2016).

The NGS library construction for RNA or DNA uses a procedural methodology that results in variations between studies. Regardless of whether a complete sample is being sequenced, this methodology entails the generation of infinitesimal reads ranging from 25 to 500 base pairs. This allows the identification of microorganisms that are either unknown or exist in minute quantities. Chiu and Miller (2019) reported that extensive bioinformatics preprocessing tools are required, including pruning, merging, assembly, scaffolding, and mapping tools. Following the sequencing procedure, distinct sequences of the microbial components of the samples will be produced in fasta or fastq files, along with a mapping file that contains all the necessary metadata associated with the sample. It is said that these files serve as inputs to the subsequent identification of the species to which the sequences belong and the assignment of taxonomy to the sequences (Caporaso et al., 2010). Using the term OTU as shown in Figure 2, it is possible to identify groups of similar sequences that have the potential to represent a distinct taxonomic classification based on these similarities (Caporaso et al., 2010).

Although this method is not without faults, it is remarkably effective for clustering sequences with 97% similarity. Using phylogenetic alignment, one sequence is selected per OTU to represent its corresponding taxa. Numerous bioinformatics techniques and algorithms have been devised in the field of shotgun and 16S rRNA metagenomics, either as independent homology- and prediction-based methods or as components of more comprehensive workflows (Singh et al., 2022; Edgar, 2018).

Multiple investigations, utilising 16S rRNA analysis, have demonstrated a connection between the gut microbiota and general well-being (Kinross et al., 2011; Schippa and Conte, 2014; Ganesan et al., 2018). An extensive examination of the gut metagenome, including WGS, can enhance our understanding of the development of illnesses and allow for the discovery of new therapeutic targets. This occurs because there is a chance of discovering small genetic differences among species that cause changes in physical traits, ultimately resulting in the development of diseases. For instance, WGS investigations carried out with *Citrobacter* spp. have shown that genetic differences within the species lead to changes in their observable characteristics and ability to adapt to different environments (Karlsson et al., 2013).

At present, the use of Illumina shotgun sequencing of stool samples is widely prevalent in the field of whole genome WGS studies of the gut microbiome. This is primarily because the gut harbours a multitude of diverse microbial species, thereby necessitating a thorough sequencing process with a coverage of at least 20 times. The purpose of such extensive sequencing is to investigate and analyze individual communities within the gut microbiome that possess low abundance



(Karlsson et al., 2013). However, it is important to analyse the substantial amount of WGS data, particularly in the form of short reads, which poses significant challenges. This is needful because the gut microbiome is home to a wide range of bacterial species, ranging from hundreds to thousands, all of which exhibit varying levels of abundance. Complicating matters further, there is a lack of taxonomic identification available for the majority of these species, further exacerbating the analytical complexities (Karlsson et al., 2013; Caporaso et al., 2010).

3.3 Bioinformatics pipelines for gut microbial analysis

Bioinformatics pipelines are essential for analyzing gut microbiota, allowing researchers to process and interpret complex metagenomic datasets efficiently. They support taxonomic, functional, and strain-level analyses, providing insights into microbial diversity and health associations. Table 3 highlights key, updated bioinformatics tools commonly used in gut microbiota studies for comprehensive and accurate data analysis.

These pipelines streamline data processing from raw sequences to biological insights, enabling researchers to study gut microbiota's role in health and disease. By integrating these tools, researchers can uncover novel therapeutic targets and biomarkers, advancing our understanding of microbiome-disease interactions.

4 Enhanced metagenomic strategies: overcoming biases and gaps

Enhanced metagenomic strategies address limitations in traditional methods by integrating advanced technologies. Long-read

sequencing resolves genomic complexities, while single-cell metagenomics bypasses culturability biases. AI-driven annotation improves functional inference, overcoming biases in DNA extraction and sequencing depth (Forbes et al., 2017). These advancements enable precise characterization of microbial communities, revealing strain-level variations and niche-specific activities. By transcending taxonomic profiling, they illuminate the dualistic nature of gut microbiota, distinguishing beneficial symbionts from pathobionts. These strategies unlock microbial roles in health and disease, offering biomarkers for diagnostics and therapeutic targets (Yan et al., 2020).

4.1 Long-read sequencing for genomic resolution and structural variation analysis

Long-read sequencing technologies, such as Oxford Nanopore and PacBio SMRT, have transformed metagenomic analyses by enabling the resolution of complex genomic regions and structural variations that evade detection by short-read methods. These platforms generate reads spanning thousands of base pairs, allowing for the assembly of complete microbial genomes, including repetitive elements, plasmids, and mobile genetic elements critical for antibiotic resistance and virulence (Panahi et al., 2024; Satam et al., 2023). Unlike short-read sequencing, which fragments these regions, long-read data preserves genomic context, enhancing functional annotation accuracy and strain-level resolution. For instance, long-read sequencing has resolved strain-specific adaptations in *Citrobacter* spp., linking genetic variations to phenotypic changes in environmental adaptability and pathogenicity (Logsdon et al., 2020). This capability is vital for studying horizontal gene transfer dynamics in dysbiotic gut communities, where plasmid-borne resistance genes proliferate. Additionally, long-read sequencing mitigates biases in taxonomic

TABLE 3 Bioinformatics pipelines for gut microbiota analysis.

Tools	Main function	Advantage	Platform	References
mothur	Microbial ecology analysis of gut microbiota	This tool has the ability of fast process of large sequence data set	Linux, macOS, Windows	Schloss et al. (2009)
MEGAN	Taxonomic analysis and visualization	This tool is applicable for the analysis of large metagenomic shotgun sequencing data	Windows, macOS	Huson et al. (2011)
MetaPhinder	Identification of phages in metagenomic data	Phage identification in metagenomic datasets	Linux	Namiki et al. (2012)
MOCAT	Metagenomic data analysis	This tool has important application in the generation and assembly of taxonomic profiles and assemble metagenomes	Linux	Kultima et al. (2012)
QIIME	Microbiome analysis pipeline	QIIME is use for Net-work analysis of meta-genomic data and histograms between microbial sample diversity or histograms within the sample	Linux, macOS, Windows	Caporaso et al. (2010) and Li (2009)
MetaPhlAn	Microbial composition profiling	This computational tool is applicable for the faster profiling of the microbial composition of microbial communities by clade-specific marker genes	Linux, macOS	Segata et al. (2012)
ConStrains	Strain-level analysis of metagenomic data	High-resolution strain identification from metagenomes	Linux	Luo et al. (2015)
MEBS	Metagenomic data processing and analysis	Efficient analysis of metagenomic sequencing data	Linux	Tarracchini et al. (2023)
Kraken2/Bracken	Taxonomic classification from shotgun data	Ultrafast and highly accurate with abundance estimation (Bracken)	Linux	Lu et al. (2024)
HUMAnN 3	Functional profiling of metagenomes	Improved accuracy in pathway reconstruction and gene-family abundance	Linux	Beghini et al. (2021)
StrainPhlAn 4	Strain-level resolution of metagenomic data	Accurate strain tracking based on MetaPhlAn markers	Linux	Blanco-Míguez et al. (2023)

profiling by capturing low-abundance taxa and uncultured species, which are often underrepresented in traditional workflows ([Ruscheweyh et al., 2022](#)). Long-read sequencing technologies overcome limitations of short-read methods by resolving repetitive genomic regions, structural variations, and microbial mobile elements. Below is a comparative analysis of key platforms and their applications in gut microbiota research (Table 4).

The application of long-read sequencing in gut microbiota research has unveiled niche-specific microbial activities and structural variations driving disease. For example, it has identified mucosal-associated *Escherichia coli* strains in IBD patients with intact virulence operons, correlating with NF- κ B-mediated inflammation ([Schirmer et al., 2019](#)). Similarly, long-read assemblies of *Akkermansia muciniphila* genomes have revealed strain-specific mucin degradation pathways critical for metabolic health ([Ouwkerk et al., 2022](#)). Despite its advantages, challenges persist, including higher costs, computational demands for data processing, and the need for high-quality DNA input. Platforms like PacBio HiFi address error rate

limitations, offering >99% accuracy for clinical-grade analyses ([Oehler et al., 2023](#)). As these technologies mature, they will bridge gaps in functional and structural metagenomics, enabling precision interventions targeting microbial contributions to diseases like T2D and colorectal cancer ([Oehler et al., 2023](#)).

4.2 Single-cell metagenomics: decoding uncultured taxa and strain heterogeneity

Single-cell metagenomics has emerged as a transformative strategy to resolve uncultured microbial taxa and strain-level heterogeneity, overcoming limitations of bulk sequencing methods that obscure rare or low-abundance species ([Xu and Zhao, 2018](#)). By isolating individual microbial cells via microfluidics or fluorescence-activated cell sorting, this approach bypasses PCR amplification biases and culturability challenges, enabling direct sequencing of genomes from “microbial dark matter.” For example, single-cell genomics has

TABLE 4 Comparative analysis of key platforms and their applications in gut microbiota.

Sequencing platform	Read length	Error rate	Advantages	Applications in gut microbiota	References
Oxford Nanopore	Up to 2 Mb	~5–10% (recent <5% with duplex reads)	Real-time sequencing, detects plasmids	Structural variation analysis, mobile elements, antibiotic resistance profiling	Cantalapiedra et al. (2021)
PacBio SMRT	10–25 kb (CLR)	~10–13% (CLR), <1% (HiFi)	High accuracy, complete genome assembly	Strain-level resolution, functional pathway annotation	Ma et al. (2024)
PacBio HiFi	15–25 kb	<0.1%	Ultra-high accuracy, hybrid assemblies	Precision metagenomics, rare taxa and variant detection	Salmaso et al. (2022)
Illumina Short-Read	100–300 bp	<0.1%	High accuracy, cost-effective, high throughput	16S rRNA profiling, taxonomic/functional profiling and diversity analysis	Ma et al. (2024) and DSouza et al. (2020)

expanded the Human Gastrointestinal Bacteria Culture Collection (HBC), providing genomic blueprints for novel species like *Saccharimonadia* and uncultured *Clostridiales*, which evade traditional cultivation ([Yu et al., 2022](#)). This technique resolves strain-specific genetic variations, such as antibiotic resistance gene clusters in *Escherichia coli* subpopulations or mucin-degrading adaptations in *Akkermansia muciniphila* strains, critical for understanding niche-specific functionalities ([Ouwkerk et al., 2022](#)). Coupled with AI-driven annotation pipelines, single-cell data enhances reference databases, improving taxonomic classification accuracy by 30% compared to conventional methods ([Erfanian et al., 2023](#)). Furthermore, it elucidates horizontal gene transfer dynamics, revealing plasmid-mediated virulence factor exchange in dysbiotic gut communities. By decoding strain-specific metabolic capabilities and host-interaction genes, single-cell metagenomics bridges gaps in functional annotation, offering insights into microbial contributions to diseases like IBD and T2D, while guiding precision probiotics and phage therapies ([Ott and Mellata, 2022](#)).

4.3 Integrative multi-omics frameworks: bridging genomics, transcriptomics, and metabolomics

Integrative multi-omics frameworks synergize genomic, transcriptomic, and metabolomic datasets to unravel the functional and spatial dynamics of gut microbiota. By coupling shotgun metagenomics with metatranscriptomics, researchers can map microbial gene expression to metabolic pathways, revealing how taxa like *Akkermansia muciniphila* modulate mucin degradation or *Faecalibacterium prausnitzii* regulate butyrate synthesis in health and disease ([Chetty and Blekhan, 2024](#)). For instance, metatranscriptomic profiling in T2D patients identified upregulated carbohydrate metabolism genes in *Muribaculaceae*, linking microbial activity to glycemic dysregulation ([Zhu and Goodarzi, 2020](#)). Tools like HUMAnN3 quantify pathway contributions across taxa, while KEGG and EggNOG databases annotate gene functions, enabling systems-level insights into host-microbe crosstalk ([Hernández-Plaza et al., 2022](#)). These frameworks resolve strain-specific adaptations, such as *Citrobacter* spp. genetic variations influencing environmental

adaptability, and track horizontal gene transfer of antibiotic resistance genes via plasmids. There are few important integrative multi-omics methods for gut microbiota analysis are listed in [Table 5](#).

Metabolomic integration further contextualizes microbial activity by identifying metabolites like SCFAs or TMAO that mediate host physiology. For example, NMR-based metabolomics paired with metagenomics revealed reduced butyrate and elevated LPS in T2D, correlating with *Bifidobacterium* depletion and *Enterobacteriaceae* blooms ([Zou et al., 2022](#)). Multi-omics platforms like MetaboAnalyst and XCMS align metabolite profiles with microbial gene expression, clarifying mechanisms like bile acid transformations by *Clostridium scindens* in non-alcoholic fatty liver disease ([Eicher et al., 2020](#)). However, challenges persist in data harmonization, as batch effects and platform-specific biases require advanced normalization algorithms ([Adamer et al., 2022](#)). Emerging AI-driven pipelines, such as MetaBGC, predict biosynthetic gene clusters from metagenomic reads, accelerating therapeutic discovery ([Sahayashela et al., 2022](#)). By bridging omics layers, these frameworks decode microbial contributions to disease, paving the way for precision probiotics and microbiome-editing therapies ([Sahayashela et al., 2022](#)).

4.4 Artificial intelligence and machine learning in functional annotation and pathway prediction

The integration of artificial intelligence (AI) and machine learning (ML) into metagenomics has transformed functional annotation and pathway prediction, addressing limitations of traditional homology-based methods such as database bias and incomplete reference genomes. AI-driven tools like MetaBGC leverage probabilistic models to identify biosynthetic gene clusters (BGCs) directly from metagenomic reads, enabling the discovery of geographically stratified metabolites with therapeutic potential, such as type II polyketides with antimicrobial properties ([Vaccaro et al., 2024](#); [Tsouka and Masoodi, 2023](#)). For instance, MetaBGC identified BGCs in gut microbiota linked to dietary adaptations, offering insights into evolutionary strategies of uncultured taxa like *Clostridiales* ([Vaccaro et al., 2024](#)). Similarly, DeepARG, a deep learning framework, predicts ARGs from raw sequencing data with

TABLE 5 Integrative multi-omics approach: application and mechanism.

Framework component	Application	Mechanism	Example	References
Shotgun metagenomics	Species- and strain-level microbial identification	High-throughput sequencing of all microbial DNA in a sample, enabling functional pathway annotation	Identified <i>Citrobacter</i> spp. strain variations linked to environmental adaptability	Gallegos et al. (2020) and Sequeira et al. (2022)
Metatranscriptomics	Gene expression profiling of active microbial pathways	RNA sequencing to map microbial transcripts, linking gene activity to metabolic outputs	Upregulated carbohydrate metabolism genes in <i>Muribaculaceae</i> in T2D patients	Ojala et al. (2023) and Vaccaro et al. (2024)
Multi-omics platforms (HUMAnN3)	Integrating genomic, transcriptomic, and metabolomic data streams	Hierarchical alignment of reads to KEGG/MetaCyc pathways, quantifying taxonomic contributions	Mapped <i>Akkermansia muciniphila</i> mucin degradation pathways to host metabolic health	Sequeira et al. (2022) and Vaccaro et al. (2024)
Metabolomics integration	Linking microbial activity to host physiology	NMR/LC-MS detects metabolites (e.g., SCFAs, TMAO) correlated with microbial gene expression	Reduced butyrate and elevated LPS in T2D linked to <i>Bifidobacterium</i> depletion	Gallegos et al. (2020) and Tsouka and Masoodi (2023)
AI-driven annotation (MetaBGC)	Functional pathway prediction and biosynthetic gene cluster identification	Machine learning models predict gene clusters from metagenomic reads	Identified type II polyketide BGCs in gut microbiota with antimicrobial properties	Vaccaro et al. (2024) and Tsouka and Masoodi (2023)

TABLE 6 AI/ML tools for functional annotation and pathway prediction.

Tool	Function	Mechanism	Application example	References
MetaBGC	BGC identification	Probabilistic model detects biosynthetic pathways from metagenomic reads	Type II polyketide discovery in gut microbiota	Vaccaro et al. (2024) and Tsouka and Masoodi (2023)
DeepARG	Antibiotic resistance prediction	Deep neural network classifies ARGs from sequencing data	Plasmid-borne β -lactamase detection in <i>E. coli</i>	Tsouka and Masoodi (2023) and Quainoo et al. (2017)
HUMAnN3	Pathway quantification	Hierarchical alignment to KEGG/MetaCyc databases	<i>A. muciniphila</i> mucin degradation pathway mapping	Vaccaro et al. (2024) and Kanehisa et al. (2015)
eggNOG-mapper	Orthologous group assignment	DIAMOND aligner matches sequences to evolutionary gene clusters	Functional annotation of uncultured <i>Clostridiales</i>	Tsouka and Masoodi (2023) and Huerta-Cepas et al. (2017)
MMvec	Microbial metabolite interaction prediction	Neural network models microbe-metabolite covariation	Linking <i>Faecalibacterium</i> butyrate synthesis to IBD remission	Vaccaro et al. (2024) and Huerta-Cepas et al. (2017)

20% higher accuracy than BLAST-based methods, critical for tracking plasmid-borne resistance genes in dysbiotic communities ([Tsouka and Masoodi, 2023](#); [Quainoo et al., 2017](#)). These models bypass reliance on reference databases, enabling annotation of “microbial dark matter” that lacks representation in public repositories ([Kanehisa et al., 2015](#)).

ML frameworks also enhance pathway prediction by integrating multi-omics data. HUMAnN3 employs hierarchical alignment to KEGG and MetaCyc databases, quantifying taxonomic contributions to metabolic pathways. For example, it mapped *Akkermansia muciniphila* mucin degradation pathways to improved insulin sensitivity in metabolic syndrome ([Vaccaro et al., 2024](#); [Kanehisa et al., 2015](#)). AI models trained on metatranscriptomic data have linked upregulated carbohydrate metabolism genes in *Muribaculaceae* to

glycemic dysregulation in T2D ([Shen et al., 2015](#)). Tools like eggNOG-mapper use DIAMOND aligners to assign orthologous groups, improving functional annotation of genes from fragmented metagenomic assemblies ([Tsouka and Masoodi, 2023](#)). Meanwhile, MMvec, a neural network, predicts microbe-metabolite interactions, such as *Faecalibacterium prausnitzii*-derived butyrate synthesis correlating with IBD remission ([Vaccaro et al., 2024](#); [Shen et al., 2015](#)). The key AI/ML tools for functional annotation and pathways prediction are listed in the [Table 6](#).

AI/ML frameworks bridge gaps in strain-specific gene function and host-microbe crosstalk. For example, *Citrobacter* spp. strain variations influencing environmental adaptability were resolved using PacBio HiFi sequencing and ML-driven annotation ([Quainoo et al., 2017](#)). These tools are pivotal for identifying therapeutic targets, such

as *Clostridium scindens*-mediated bile acid metabolism in NAFLD (Kanehisa et al., 2015). As the field advances, AI-driven metagenomics will underpin precision probiotics and microbiome-editing therapies, translating microbial ecology into clinical innovation (Vaccaro et al., 2024; Huerta-Cepas et al., 2017).

5 Conclusion

Enhanced metagenomic strategies have revolutionized our understanding of gut microbiota by transcending the limitations of traditional approaches. Long-read sequencing resolves structural variations and plasmids, enabling complete genome assemblies of uncultured taxa and strain-level insights into pathobionts like *Citrobacter* spp. Single-cell metagenomics deciphers “microbial dark matter,” while AI-driven tools (MetaBGC, HUMAnN3) predict biosynthetic pathways and antibiotic resistance genes with unprecedented accuracy. Multi-omics frameworks integrate genomic, transcriptomic, and metabolomic data, linking microbial activity to host phenotypes—such as *Akkermansia muciniphila*’s mucin degradation in metabolic health or *Muribaculaceae*’s upregulated carbohydrate metabolism in T2D. These strategies identify biomarkers (e.g., butyrate-producing *Bifidobacterium* depletion in T2D) and therapeutic targets, such as *Clostridium scindens*-mediated bile acid metabolism in NAFLD. Despite challenges in standardization and computational demands, enhanced metagenomics bridges observational and mechanistic research, paving the way for precision probiotics, microbiota identification and microbiome-editing interventions. As the field advances, these tools will be pivotal in translating microbial ecology into actionable clinical strategies, transforming our approach to managing chronic diseases.

Author contributions

XL: Visualization, Formal analysis, Validation, Writing – review & editing, Writing – original draft, Conceptualization, Methodology. HL: Funding acquisition, Investigation, Writing – review & editing, Resources, Methodology, Formal analysis, Visualization, Validation, Conceptualization, Supervision.

References

- Adamer, M. F., Brüningk, S. C., Tejada-Arranz, A., Estermann, F., Basler, M., and Borgwardt, K. (2022). Recombat: batch-effect removal in large-scale multi-source gene-expression data integration. *Bioinformatics Adv.* 2:vbac071. doi: 10.1093/bioadv/vbac071
- Arab, J. P., Karpen, S. J., Dawson, P. A., Arrese, M., and Trauner, M. (2016). Bile acids and nonalcoholic fatty liver disease: molecular insights and therapeutic perspectives. *Hepatology* 65, 350–362. doi: 10.1002/hep.28709
- Bai, X., Narayanan, A., Nowak, P., Ray, S., Neogi, U., and Sönnnerborg, A. (2021). Whole-genome metagenomic analysis of the gut microbiome in HIV-1-infected individuals on antiretroviral therapy. *Front. Microbiol.* 12:667718. doi: 10.3389/fmicb.2021.667718
- Balvočiūtė, M., and Huson, D. H. (2017). SILVA, rdp, Greengenes, NCBI and OTT—how do these taxonomies compare? *BMC Genomics* 18:114. doi: 10.1186/s12864-017-3501-4
- Bastiaansen, T. F. S., Cowan, C. S. M., Claesson, M. J., Dinan, T. G., and Cryan, J. F. (2018). Making sense of ... the microbiome in psychiatry. *Int. J. Neuropsychopharmacol.* 22, 37–52. doi: 10.1093/ijnp/pyy067
- Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., et al. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* 10:e65088. doi: 10.7554/eLife.65088
- Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., et al. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* 41, 1633–1644. doi: 10.1038/s41587-023-01688-w
- Bravo, J. A., Forsythe, P., Chew, M. V., Escaravage, E., Savignac, H. M., Dinan, T. G., et al. (2011). Ingestion of *Lactobacillus* strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16050–16055. doi: 10.1073/pnas.1102999108
- Bukin, Y. S., Galachyants, Y. P., Morozov, I. V., Bukin, S. V., Zakharenko, A. S., and Zemskaia, T. I. (2019). The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Sci. Data* 6:190007. doi: 10.1038/sdata.2019.7
- Bull, M. J., and Plummer, N. T. (2014). Part 1: the human gut microbiome in health and disease. *Integr. Med* 13, 17–22.
- Callahan, B. J., McMurdie, P. J., and Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* 11, 2639–2643. doi: 10.1038/ismej.2017.119
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, orthology assignments, and domain

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Department of Anorectal Surgery, The Second Affiliated Hospital of Hunan University of Chinese Medicine, Changsha, 410005, China provides funding for this research.

Acknowledgments

The authors thank the Changsha Hospital of Traditional Chinese Medicine and The Second Affiliated Hospital of Hunan University of Chinese Medicine, Changsha, China for necessary research support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- prediction at the metagenomic scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Chen, Z., Liang, N., Zhang, H., Li, H., Guo, J., Zhang, Y., et al. (2024). Resistant starch and the gut microbiome: exploring beneficial interactions and dietary impacts. *Food Chem.: X* 21:101118. doi: 10.1016/j.fochx.2024.101118
- Chen, M., Yuan, L., Xie, C. R., Wang, X. Y., Feng, S. J., Xiao, X. Y., et al. (2023). Probiotics for the management of irritable bowel syndrome: a systematic review and three-level meta-analysis. *Int. J. Surg.* 109, 3631–3647. doi: 10.1097/JS9.0000000000000658
- Chetty, A., and Blekhan, R. (2024). Multi-omic approaches for host-microbiome data integration. *Gut Microbes* 16:2297860. doi: 10.1080/19490976.2023.2297860
- Chiu, C. Y., and Miller, S. A. (2019). Clinical Metagenomics. *Nat. Rev. Genet.* 20, 341–355. doi: 10.1038/s41576-019-0113-7
- Costa, R. F. A., Ferrari, M. L. A., Bringer, M., Darfeuille-Michaud, A., Martins, F. S., and Barnich, N. (2020). Characterization of mucosa-associated *Escherichia coli* strains isolated from Crohn's disease patients in Brazil. *BMC Microbiol.* 20:178. doi: 10.1186/s12866-020-01856-x
- Cozzetto, D., Minneci, F., Currant, H., and Jones, D. T. (2016). FFPred 3: feature-based function prediction for all gene ontology domains. *Sci. Rep.* 6:31865. doi: 10.1038/srep31865
- Cryan, J. F., O'Riordan, K. J., Cowan, C. S. M., Sandhu, K. V., Bastiaansen, T. F. S., Boehme, M., et al. (2019). The microbiota-gut-brain axis. *Physiol. Rev.* 99, 1877–2013. doi: 10.1152/physrev.00018.2018
- de Vos, W. M., Tilg, H., Van Hul, M., and Cani, P. D. (2022). Gut microbiome and health: mechanistic insights. *Gut* 71, 1020–1032. doi: 10.1136/gutjnl-2021-326789
- Dejea, C. M., Fathi, P., Craig, J. M., Boleij, A., Taddese, R., Geis, A. L., et al. (2018). Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* 359, 592–597. doi: 10.1126/science.aah3648
- Dovrolis, N., Kolios, G., Spyrou, G. M., and Maroulakou, I. (2017). Computational profiling of the gut-brain axis: microflora dysbiosis insights to neurological disorders. *Briefings Bioinform.* 20, 825–841. doi: 10.1093/bib/bbx154
- DSouza, S., Ponnanna, K., Chokkanna, A., and Ramachandra, N. (2020). Illumina short-read sequencing data, de novo assembly and annotations of the *Drosophila nasuta* genome. *Data Brief* 34:106674. doi: 10.1016/j.dib.2020.106674
- Edgar, R. C. (2018). Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics* 34, 2371–2375. doi: 10.1093/bioinformatics/bty113
- Eicher, T., Kinnebrew, G., Patt, A., Spencer, K., Ying, K., Ma, Q., et al. (2020). Metabolomics and multi-omics integration: a survey of computational methods and resources. *Metabolites* 10:202. doi: 10.3390/metabo10050202
- Erfanian, N., Heydari, A. A., Feriz, A. M., Iañez, P., Derakhshani, A., Ghasemigol, M., et al. (2023). Deep learning applications in single-cell genomics and transcriptomics data analysis. *Biomed. Pharmacother.* 165:115077. doi: 10.1016/j.biopha.2023.115077
- Feng, W., Liu, J., Ao, H., Yue, S., and Peng, C. (2020). Targeting gut microbiota for precision medicine: focusing on the efficacy and toxicity of drugs. *Theranostics* 10, 11278–11301. doi: 10.7150/thno.47289
- Forbes, J. D., Knox, N. C., Ronholm, J., Pagotto, F., and Reimer, A. (2017). Metagenomics: the next culture-independent game changer. *Front. Microbiol.* 8:1069. doi: 10.3389/fmicb.2017.01069
- Forster, S. C., Kumar, N., Anonye, B. O., Almeida, A., Viciani, E., Stares, M. D., et al. (2019). A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nat. Biotechnol.* 37, 186–192. doi: 10.1038/s41587-018-0009-7
- Gallegos, J. E., Hayrynen, S., Adames, N. R., and Peccoud, J. (2020). Challenges and opportunities for strain verification by whole-genome sequencing. *Sci. Rep.* 10:5873. doi: 10.1038/s41598-020-62364-6
- Ganesan, K., Chung, S. K., Vanamala, J., and Xu, B. (2018). Causal relationship between diet-induced gut microbiota changes and diabetes: a novel strategy to transplant *Faecalibacterium prausnitzii* in preventing diabetes. *Int. J. Mol. Sci.* 19:3720. doi: 10.3390/ijms19123720
- Ghaisas, S., Maher, J., and Kanthasamy, A. (2016). Gut microbiome in health and disease: linking the microbiome-gut-brain axis and environmental factors in the pathogenesis of systemic and neurodegenerative diseases. *Pharmacol. Ther.* 158, 52–62. doi: 10.1016/j.pharmthera.2015.11.012
- Hernández-Plaza, A., Szklarczyk, D., Botas, J., Cantalapiedra, C. P., Giner-Lamia, J., Mende, D. R., et al. (2022). eggNOG 6.0: enabling comparative genomics across 12535 organisms. *Nucleic Acids Res.* 51, D389–D394. doi: 10.1093/nar/gkac1022
- Hrncir, T. (2022). Gut microbiota dysbiosis: triggers, consequences, diagnostic and therapeutic options. *Microorganisms* 10:578. doi: 10.3390/microorganisms10030578
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148
- Huson, D. H., Mitra, S., Ruscheweyh, H.-J., Weber, N., and Schuster, S. C. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* 21, 1552–1560. doi: 10.1101/gr.120618.111
- Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., et al. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. *Front. Microbiol.* 7:459. doi: 10.3389/fmicb.2016.00459
- Kanehisa, M., Sato, Y., and Morishima, K. (2015). Blastkoala and ghostkoala: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006
- Karlsson, F., Tremaroli, V., Nielsen, J., and Bäckhed, F. (2013). Assessing the human gut microbiota in metabolic diseases. *Diabetes* 62, 3341–3349. doi: 10.2337/db13-0844
- Kinross, J. M., Darzi, A. W., and Nicholson, J. K. (2011). Gut microbiome-host interactions in health and disease. *Genome Med.* 3:14. doi: 10.1186/gm228
- Knight, R., Vrbanc, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., et al. (2018). Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* 16, 410–422. doi: 10.1038/s41579-018-0029-9
- Kultima, J. R., Sunagawa, S., Li, J., Chen, W., Chen, H., Mende, D. R., et al. (2012). MOCAT: a metagenomics assembly and gene prediction toolkit. *PLoS One* 7:e47656. doi: 10.1371/journal.pone.0047656
- Li, W. (2009). Analysis and comparison of very large metagenomes with fast clustering and functional annotation. *BMC Bioinformatics* 10:359. doi: 10.1186/1471-2105-10-359
- Logsdon, G. A., Vollger, M. R., and Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* 21, 597–614. doi: 10.1038/s41576-020-0236-x
- Louis, P., Hold, G. L., and Flint, H. J. (2014). The gut microbiota, bacterial metabolites and colorectal cancer. *Nat. Rev. Microbiol.* 12, 661–672. doi: 10.1038/nrmicro3344
- Lu, J., Rincon, N., Wood, D. E., Breitwieser, F. P., Pockrandt, C., Langmead, B., et al. (2022). Author Correction: Metagenome analysis using the kraken software suite. *Nat. Protoc.* 17, 2815–2839. doi: 10.1038/s41596-024-01064-1
- Luo, C., Knight, R., Siljander, H., Knip, M., Xavier, R. J., and Gevers, D. (2015). ConStrains identifies microbial strains in metagenomic datasets. *Nat. Biotechnol.* 33, 1045–1052. doi: 10.1038/nbt.3319
- Ma, J., Xu, R., Li, W., Liu, M., and Ding, X. (2024). Whole-genome sequencing of clinical isolates of *Citrobacter europaeus* in China carrying bla_{OXA-48} and bla_{NDM-1}. *Ann. Clin. Microbiol. Antimicrob.* 23:38. doi: 10.1186/s12941-024-00699-y
- Mangoma, N., Zhou, N., and Ncube, T. (2024). Metagenome-assembled genomes provide insight into the microbial taxonomy and ecology of the Buhera soda pans, Zimbabwe. *PLoS One* 19:e0299620. doi: 10.1371/journal.pone.0299620
- Mousa, W. K., and Ali, A. A. (2024). The gut microbiome advances precision medicine and diagnostics for inflammatory bowel diseases. *Int. J. Mol. Sci.* 25:11259. doi: 10.3390/ijms252011259
- Namiki, T., Hachiya, T., Tanaka, H., and Sakakibara, Y. (2012). MetaVelvet: an extension of velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res.* 40:e155. doi: 10.1093/nar/gks678
- Oehler, J. B., Wright, H., Stark, Z., Mallett, A. J., and Schmitz, U. (2023). The application of long-read sequencing in clinical settings. *Hum. Genomics* 17:73. doi: 10.1186/s40246-023-00522-3
- Ojala, T., Kankuri, E., and Kankainen, M. (2023). Understanding human health through metatranscriptomics. *Trends Mol. Med.* 29, 376–389. doi: 10.1016/j.molmed.2023.02.002
- Ott, L. C., and Mellata, M. (2022). Models for gut-mediated horizontal gene transfer by bacterial plasmid conjugation. *Front. Microbiol.* 13:891548. doi: 10.3389/fmicb.2022.891548
- Ouwkerk, J. P., Tytgat, H. L. P., Elzinga, J., Koehorst, J., van den Abbeele, P., Henrissat, B., et al. (2022). Comparative genomics and physiology of *Akkermansia muciniphila* isolates from human intestine reveal specialized mucosal adaptation. *Microorganisms* 10:1605. doi: 10.3390/microorganisms10081605
- Panahi, B., Jalaly, H. M., and Hamid, R. (2024). Using next-generation sequencing approach for discovery and characterization of plant molecular markers. *Curr. Plant Biol.* 40:100412. doi: 10.1016/j.cpb.2024.100412
- Plovier, H., Everard, A., Druart, C., Depommier, C., Van Hul, M., Geurts, L., et al. (2016). A purified membrane protein from *Akkermansia muciniphila* or the pasteurized bacterium improves metabolism in obese and diabetic mice. *Nat. Med.* 23, 107–113. doi: 10.1038/nm.4236
- Potrykus, M., Czaja-Stolc, S., Stankiewicz, M., Kaska, Ł., and Małgorzewicz, S. (2021). Intestinal microbiota as a contributor to chronic inflammation and its potential modifications. *Nutrients* 13:3839. doi: 10.3390/nu13113839
- Prodan, A., Tremaroli, V., Brolin, H., Zwinderman, A. H., Nieuwdorp, M., and Levin, E. (2020). Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLoS One* 15:e0227434. doi: 10.1371/journal.pone.0227434
- Quainoo, S., Coolen, J. P. M., van Hijum, S. A. F. T., Huynen, M. A., Melchers, W. J. G., van Schaik, W., et al. (2017). Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. *Clin. Microbiol. Rev.* 30, 1015–1063. doi: 10.1128/cmr.00016-17

- Ranjan, R., Rani, A., Metwally, A., McGee, H. S., and Perkins, D. L. (2016). Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem. Biophys. Res. Commun.* 469, 967–977. doi: 10.1016/j.bbrc.2015.12.083
- Reck, M., Tomasch, J., Deng, Z., Jarek, M., Husemann, P., and Wagner-Döbler, I. (2015). Stool metatranscriptomics: a technical guideline for mRNA stabilisation and isolation. *BMC Genomics* 16:494. doi: 10.1186/s12864-015-1694-y
- Rooks, M. G., and Garrett, W. S. (2016). Gut microbiota, metabolites and host immunity. *Nat. Rev. Immunol.* 16, 341–352. doi: 10.1038/nri.2016.42
- Ruscheweyh, H.-J., Milanese, A., Paoli, L., Karcher, N., Clayssen, Q., Keller, M. I., et al. (2022). Cultivation-independent genomes greatly expand taxonomic-profiling capabilities of mOTUs across various environments. *Microbiome* 10:212. doi: 10.1186/s40168-022-01410-z
- Sahayasheela, V. J., Lankadasari, M. B., Dan, V. M., Dastager, S. G., Pandian, G. N., and Sugiyama, H. (2022). Artificial intelligence in microbial natural product drug discovery: current and emerging role. *Nat. Prod. Rep.* 39, 2215–2230. doi: 10.1039/d2np00035k
- Salmaso, N., Vasselon, V., Rimet, F., Vautier, M., Elerssek, T., Boscaini, A., et al. (2022). DNA sequence and taxonomic gap analyses to quantify the coverage of aquatic cyanobacteria and eukaryotic microalgae in reference databases: results of a survey in the Alpine region. *Sci. Total Environ.* 834:155175. doi: 10.1016/j.scitotenv.2022.155175
- Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., et al. (2023). Next-generation sequencing technology: current trends and advancements. *Biology* 12:997. doi: 10.3390/biology12070997
- Schipa, S., and Conte, M. (2014). Dysbiotic events in gut microbiota: impact on human health. *Nutrients* 6, 5786–5805. doi: 10.3390/nu6125786
- Schirmer, M., Garner, A., Vlamakis, H., and Xavier, R. J. (2019). Microbial genes and pathways in inflammatory bowel disease. *Nat. Rev. Microbiol.* 17, 497–511. doi: 10.1038/s41579-019-0213-6
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/aem.01541-09
- Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* 9, 811–814. doi: 10.1038/nmeth.2066
- Sequeira, J. C., Rocha, M., Alves, M. M., and Salvador, A. F. (2022). UPIMAPI, reCOGNizer and KEGGCharter: bioinformatics tools for functional annotation and visualization of (meta)-omics datasets. *Comput. Struct. Biotechnol. J.* 20, 1798–1810. doi: 10.1016/j.csbj.2022.03.042
- Shen, N., Dimitrova, N., Ho, C. H., Torres, P. J., Camacho, F. R., Cai, Y., et al. (2015). Gut microbiome activity predicts risk of type 2 diabetes and metformin control in a large human cohort. *medRxiv*. Available online at: <https://doi.org/10.1101/2021.08.13.21262051>. [Epub ahead of preprint]
- Shin, Y., Han, S., Kwon, J., Ju, S., Choi, T., Kang, I., et al. (2023). Roles of short-chain fatty acids in inflammatory bowel disease. *Nutrients* 15:4466. doi: 10.3390/nu15204466
- Singh, N., Singh, V., Rai, S. N., Mishra, V., Vamanu, E., and Singh, M. P. (2022). Deciphering the gut microbiome in neurodegenerative diseases and metagenomic approaches for characterization of gut microbes. *Biomed. Pharmacother.* 156:113958. doi: 10.1016/j.biopha.2022.113958
- Sugimoto, Y., Camacho, F. R., Wang, S., Chankhamjon, P., Odabas, A., Biswas, A., et al. (2019). A metagenomic strategy for harnessing the chemical repertoire of the human microbiome. *Science* 366:eaax9176:366. doi: 10.1126/science.aax9176
- Tarracchini, C., Alessandri, G., Fontana, F., Rizzo, S. M., Lugli, G. A., Bianchi, M. G., et al. (2023). Genetic strategies for sex-biased persistence of gut microbes across human life. *Nat. Commun.* 14:4220. doi: 10.1038/s41467-023-39931-2
- Thursby, E., and Juge, N. (2017). Introduction to the human gut microbiota. *Biochem. J.* 474, 1823–1836. doi: 10.1042/bcj20160510
- Tokuda, M., and Shintani, M. (2024). Microbial evolution through horizontal gene transfer by mobile genetic elements. *Microb. Biotechnol.* 17:e14408. doi: 10.1111/1751-7915.14408
- Tsouka, S., and Masoodi, M. (2023). Metabolic pathway analysis: advantages and pitfalls for the functional interpretation of metabolomics and lipidomics data. *Biomolecules* 13:244. doi: 10.3390/biom13020244
- Vaccaro, M., Almaatouq, A., and Malone, T. (2024). When combinations of humans and AI are useful: a systematic review and meta-analysis. *Nat. Hum. Behav.* 8, 2293–2303. doi: 10.1038/s41562-024-02024-1
- van Nood, E., Vrieze, A., Nieuwdorp, M., Fuentes, S., Zoetendal, E. G., de Vos, W. M., et al. (2013). Duodenal infusion of donor feces for recurrent *Clostridium difficile*. *N. Engl. J. Med.* 368, 407–415. doi: 10.1056/nejmoa1205037
- Vogt, N. M., Romano, K. A., Darst, B. F., Engelman, C. D., Johnson, S. C., Carlsson, C. M., et al. (2018). The gut microbiota-derived metabolite trimethylamine N-oxide is elevated in Alzheimer's disease. *Alzheimers Res Ther* 10:124. doi: 10.1186/s13195-018-0451-2
- Wang, W.-L., Xu, S.-Y., Ren, Z.-G., Tao, L., Jiang, J.-W., and Zheng, S.-S. (2015). Application of metagenomics in the human gut microbiome. *World J. Gastroenterol.* 21, 803–814. doi: 10.3748/wjg.v21.i3.803
- Xu, Y., and Zhao, F. (2018). Single-cell metagenomics: challenges and applications. *Protein Cell* 9, 501–510. doi: 10.1007/s13238-018-0544-5
- Xuan, W., Ou, Y., Chen, W., Huang, L., Wen, C., Huang, G., et al. (2023). *Faecalibacterium prausnitzii* improves lipid metabolism disorder and insulin resistance in type 2 diabetic mice. *Br. J. Biomed. Sci.* 80:10794. doi: 10.3389/bjbs.2023.10794
- Yan, Y., Nguyen, L. H., Franzosa, E. A., and Huttenhower, C. (2020). Strain-level epidemiology of microbial communities and the human microbiome. *Genome Med.* 12:71. doi: 10.1186/s13073-020-00765-y
- Yang, W., and Cong, Y. (2021). Gut microbiota-derived metabolites in the regulation of host immune responses and immune-related inflammatory diseases. *Cell. Mol. Immunol.* 18, 866–877. doi: 10.1038/s41423-021-00661-4
- Yu, Y., Wen, H., Li, S., Cao, H., Li, X., Ma, Z., et al. (2022). Emerging microfluidic technologies for microbiome research. *Front. Microbiol.* 13:906979. doi: 10.3389/fmicb.2022.906979
- Zeng, M. Y., Inohara, N., and Núñez, G. (2016). Mechanisms of inflammation-driven bacterial dysbiosis in the gut. *Mucosal Immunol.* 10, 18–26. doi: 10.1038/mi.2016.75
- Zhu, T., and Goodarzi, M. O. (2020). Metabolites linking the gut microbiome with risk for type 2 diabetes. *Curr. Nutr. Rep.* 9, 83–93. doi: 10.1007/s13668-020-00307-3
- Zou, H., Huang, C., Zhou, L., Lu, R., Zhang, Y., and Lin, D. (2022). NMR-based metabolomic analysis for the effects of trimethylamine N-oxide treatment on C2C12 myoblasts under oxidative stress. *Biomolecules* 12:1288. doi: 10.3390/biom12091288