

Microbiome Analysis: Tools and Techniques for Understanding Human Gut Microbiota

Sabeen Raza¹, Mirjana Menkovska²

¹Texas Children's Microbiome Center, Baylor College of Medicine, Houston, TX, USA

²Department of Nutrition and Foodstuffs Processing, University of Ss. Cyril and Methodius in Skopje, Skopje, Nord Macedonia

Email: sraza@houstonmethodist.org

How to cite this paper: Raza, S. and Menkovska, M. (2026) Microbiome Analysis: Tools and Techniques for Understanding Human Gut Microbiota. *Advances in Bioscience and Biotechnology*, 17, 219-232. <https://doi.org/10.4236/abb.2026.176015>

Received: May 13, 2026

Accepted: June 12, 2026

Published: June 15, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

A comprehensive understanding of how the human gut microbiome impacts systemic health and metabolic regulation necessitates advancing from traditional marker gene sequencing to sophisticated, data-driven systems biology approaches. The interpretation of high-dimensional datasets in nutritional microbiome research is often impeded by substantial genetic variability, statistical noise, and inherent compositional bias. This review provides an analysis of the current meta-omics technologies—covering metagenomics, metatranscriptomics, metaproteomics, and metabolomics—and their corresponding bioinformatic pipelines that convert raw biological samples into clinically meaningful nutritional information. We assess established analytical frameworks such as QIIME 2, bioBakery 3, and MaAsLin 3, together with multi-omics integration tools including DIABLO and MOFA+, which address compositional data complexities. Furthermore, this manuscript discusses innovative methodologies transforming the discipline, including single-cell metagenomics (SiC-seq), telomere-to-telomere (T2T) long-read sequencing, spatially resolved microbiomics, and the use of artificial intelligence through foundational large language models (LLMs). Through clinical case studies involving pediatric irritable bowel syndrome, epilepsy, and pharmacomicrobiomics (with a focus on GLP-1 agonists), the practical value of data-driven dietary intervention is demonstrated. Ongoing enhancement of analytical techniques and adherence to standardized reporting practices remain fundamental for resolving reproducibility challenges and progressing towards the clinical implementation of precision nutrition and the development of in silico digital twins for individualized healthcare.

Keywords

Gut Microbiota, Meta-Omics Technologies, Bioinformatics Pipelines,

Precision Nutrition, Compositional Bias, Multi-Omics Integration,
Metagenomics, Metabolomics, Spatially Resolved Microbiomics,
Pharmacomicrobiomics, Artificial Intelligence in Microbiome

1. Introduction

The well-known saying, “we are what we eat”, has become increasingly significant as scientific discoveries highlight the influence of gut microbes on human physiology. Within the human gastrointestinal tract exists a vibrant ecosystem made up of trillions of microorganisms, including bacteria, fungi, parasites, and viruses. These diverse microbes are essential for stimulating the immune system, maintaining metabolic balance, and shaping our susceptibility to chronic diseases [1].

1.1. Impact of Macronutrients on Microbiome Composition

In our previous review, *Human Gut Microbiome in Relation to Food, Nutrition, Health and Disease*, we described the influence of daily macronutrient intake on the structure of microbial communities. That work clarified the mechanisms at the host-microbiome interface: dietary fibers support saccharolytic bacteria that generate beneficial short-chain fatty acids (SCFAs), whereas high-fat diets, especially those high in saturated fats, promote intestinal dysbiosis. This pathological shift—often indicated by reduced *Bacteroidetes* and increased *Firmicutes* and Proteobacteria—contributes to insulin resistance, increases gut permeability, and intensifies adipose tissue inflammation.

Turning observations into medical therapies is difficult, largely due to the vast genetic and biochemical variety found within the human microbiome [2]. Nutrition research produces huge, complex data sets that require advanced computational tools to filter out statistical noise. As a result, scientists face the critical challenge of reliably interpreting these detailed biological responses to various diets. This complexity is further compounded by inter-individual differences in microbiome composition, environmental influences, and dietary habits, which can obscure direct cause-and-effect relationships between nutrition and health outcomes. Moreover, the dynamic nature of the gut microbiome—constantly shifting in response to both short-term dietary changes and long-term lifestyle factors—necessitates longitudinal studies and integrative analytical frameworks to capture meaningful patterns. Addressing these challenges calls for collaborative efforts encompassing multi-omics data integration, standardized protocols, and machine learning approaches to translate raw data into actionable insights for personalized nutrition therapies.

Building upon our earlier clinical overview, this manuscript explicitly centers on the analytical methodologies driving nutritional microbiome studies and their clinical translation. Rather than providing a broad dietary background, we provide a comprehensive methodological guide detailing how advanced data analysis

tools, meta-omics technologies, and robust bioinformatics pipelines transform raw biological samples into actionable, personalized nutritional insights.

1.2. Omics Science in Nutritional Microbiome Research

To accurately determine the effects of specific diets on the gut ecosystem, researchers in modern systems biology rely on a suite of “omics” sciences. These specialized fields are designed to collect and analyze extensive datasets that reflect the molecular complexity of the gut environment at multiple levels. By leveraging these omics approaches, scientists can systematically study how dietary changes influence the composition, function, and interactions of microbial communities within the gastrointestinal tract [3]. Integrating these large-scale datasets enables a comprehensive understanding of the molecular mechanisms that link nutrition to gut microbiome dynamics and, ultimately, to human health. From a nutrition perspective, these methods enable scientists to follow food as it is eaten, processed by microbes, and eventually used by cells:

- **Genomics:** Maps the structural blueprint of the community, revealing which microbes possess the genetic capacity to digest incoming food [4].
- **Proteomics:** Catalogs the actual enzymes expressed by microbes to actively break down complex dietary macronutrients [5].
- **Metabolomics:** Studies the ultimate biochemical breakdown products, differentiating between metabolites derived directly from the diet (the exposome) and those synthesized by our microbial symbionts [6].
- **Transcriptomics:** Captures how microbial genes are turned on or off in real-time following a sudden dietary shift [7].
- **Interactomics:** Investigates the complex physical and biochemical relationships between dietary molecules, microbial receptors, and host immune sensors [8].

Despite generating vast amounts of data that pose storage and interpretation challenges, these technologies are essential for advancing from broad dietary advice to truly personalized nutrition.

1.3. The End-to-End Microbiome Analytical Workflow: From Clinic to Computation

In nutrition science, researchers commonly devise and adhere to standardized protocols for sample collection, processing, and data analysis to ensure consistency and reliability. By following a structured workflow, computational analyses are better positioned to reflect authentic biological interactions among diet, the microbiome, and the host.

1) Study Design and Metadata Acquisition The foundation of robust microbiome analysis lies in strict cohort selection and comprehensive clinical phenotyping. Studies rely on careful age- and environment-matching while rigorously excluding confounding factors, such as the use of antibiotics, probiotics, or steroids in the months preceding the study. Because dietary responses are highly in-

dividualized, capturing granular metadata is critical. Researchers heavily rely on validated clinical tracking, such as two-week symptom diaries, Bristol stool form charts, structured dietary intake logs, and continuous metabolic monitoring (e.g., tracking blood glucose and ketones) to pair with the microbial data.

2) Sample Collection and Preservation Capturing a true snapshot of the gut ecosystem requires rapid and sterile collection protocols to prevent post-defecation microbial overgrowth or degradation. Depending on the cohort, non-invasive collection ranges from using sterile stool “hats” and transfer cups in older children and adults to direct diaper swabs in pediatric populations. Once collected, samples must be immediately frozen and stored at ultra-low temperatures (e.g., -80°C) to suspend biological activity and preserve both the microbial DNA and highly volatile metabolites for downstream extraction.

3) DNA Extraction and Processing Before sequencing can occur, the physical microbial cells within the fecal matrix must be broken open to release their genetic material. This is typically achieved using validated commercial DNA isolation kits that employ chemical and mechanical lysis (bead-beating). This standardized extraction is crucial for ensuring that the DNA recovered represents the true diversity of the sample, overcoming the tough cell walls of certain bacterial species without degrading the nucleic acids.

4) Library Preparation and High-Throughput Sequencing With purified DNA isolated, researchers target specific genetic markers to identify the microbial community. For broad taxonomic profiling, specific hypervariable regions of the bacterial 16S rRNA gene (such as V1-V3 or V3-V5) are amplified using barcoded primers. These uniquely tagged samples are multiplexed together and loaded onto high-throughput sequencing platforms, generating tens of thousands of high-quality sequence reads per sample that act as molecular “fingerprints” for the microbes present.

5) Bioinformatic Processing and Statistical Modeling The final and most computationally heavy stage involves translating millions of raw sequencing reads into actionable clinical insights. Advanced bioinformatics pipelines, such as QIIME, are utilized to filter out low-quality reads, align the sequences to established phylogenetic databases, and group them into defined taxonomic units (OTUs or ASVs). Once the data is normalized into relative abundance tables, data scientists calculate alpha and beta diversity metrics. Finally, powerful machine learning algorithms (such as Random Forest models) and non-parametric statistical tests are applied to the dataset. These models integrate the microbial profiles with the clinical metadata to identify robust biomarkers capable of stratifying disease subtypes, predicting dietary responsiveness, or evaluating the bidirectional relationships of clinical therapies like GLP-1 agonists.

1.4. Bioinformatics and the Mathematics of Microbiome Data

Bioinformatics bridges biology and computing, helping researchers interpret DNA from stool samples. This supports nutrition studies by revealing microbial

profiles, pinpointing health-linked genes, and shaping personalized diet plans according to each person's unique microbial DNA.

1.5. Overcoming Compositional Bias

A major mathematical hurdle in meta-omics data analysis is overcoming statistical noise, most notably *compositional bias* [9]. Next-generation sequencing (NGS) instruments yield an arbitrary number of total sequence reads per sample, determined by the sequencer's technical capacity rather than the absolute microbial cellular load of the original biological sample. As a result, microbiome sequencing data is inherently compositional, representing relative proportions rather than absolute counts. Because these proportions must always sum to 100%, a true biological increase in a single bacterial taxon mathematically forces an artificial decrease in the relative abundance of all others [10]. If data scientists naively apply standard, compositionally unaware statistical tests to this relative data, the false discovery rate inflates dramatically, leading to inaccurate correlations.

1.6. Beyond Compositionality: Addressing Additional Analytical Biases

As compositional bias poses a significant mathematical challenge in microbiome research, it is also very important for nutritional microbiome studies to address several other confounding factors to ensure reliable and reproducible results. One key consideration is the management of batch effects, which are systematic, non-biological variations that can be introduced by differences in sequencing runs, reagent lots, or extraction protocols. These batch effects can obscure or distort genuine dietary signals, ultimately leading to misinterpretation of the data.

It is also essential to rigorously control host DNA contamination during sample processing and data analysis. Without this, the presence of human genetic material can interfere with microbial profiling, potentially masking true associations between diet and the gut microbiome. Researchers must also meticulously account for relevant clinical covariates—such as participant age, baseline metabolic health, and concurrent medication use—when performing statistical modeling. Careful adjustment for these variables is critical to distinguish authentic interactions between diet and the microbiome from extraneous environmental or physiological influences.

1.7. New Disciplines in Omics Research

Advanced data analysis and nutritional science now combine to form new disciplines focused on the host-diet relationship.

- **Foodomics:** Applies multi-omics data to clearly map how food components, microbial activity, and human health are interconnected.
- **Nutrigenomics:** Serves as the foundation for personalized nutrition. Using transcriptomics, proteomics, and metabolomics, it helps predict how a person's genetics and microbiome affect their metabolic response to certain diets.

1.8. Meta-Omics Technologies for Characterizing Dietary Responses

Researchers employ meta-omics methods to gain a comprehensive understanding of how food impacts the gut, examining all molecular products from the complete microbial community all at once rather than just focusing on individual strains. Because dietary metabolism is an ongoing biological process rather than a solitary occurrence, revealing its intricacies involves investigating the ecosystem across several molecular levels.

The following subsections cover: the microbiome's genetic makeup (metagenomics), gene activity (metatranscriptomics), digestive enzyme deployment (metaproteomics), and fermentation end-products (metabolomics). Integrating these layers enables researchers to trace a nutrient's path from ingestion to host effect.

1.9. Metagenomics: Mapping the Nutritional Blueprint

Metagenomic analysis enables highly accurate assessments of the gut's functional composition. Whole-genome shotgun (WGS) sequencing offers a comprehensive examination by sequencing all fragmented DNA present in a sample, providing an in-depth look at microbial communities. By aligning these sequences to large reference databases, researchers can confidently identify specific strains and evaluate the actual genetic capacity of the gut microbiome to degrade complex carbohydrates or synthesize specific metabolites. However, WGS is both computationally demanding and financially costly, and its effectiveness is limited when reference databases are incomplete, leaving many uncultured gut microbes unidentified.

In contrast, 16S rRNA marker gene sequencing remains the primary, cost-effective methodology for broad investigations into dietary trends across larger cohorts. Historically, 16S data analysis relied on clustering sequences into Operational Taxonomic Units (OTUs) based on an arbitrary 97% similarity threshold, an approach that frequently obscured functionally distinct but genetically similar microbes. Modern bioinformatics pipelines have largely replaced OTUs with Amplicon Sequence Variants (ASVs). Utilizing advanced statistical error-correcting algorithms—such as DADA2—this approach successfully distinguishes true biological variation from sequencing errors, providing exact sequence resolution down to the single-nucleotide level [11].

However, while ASV-based analysis yields a highly refined taxonomic profile compared to traditional clustering, it presents crucial limitations that must be acknowledged in nutritional research. Because the short hypervariable regions of the 16S gene (e.g., V3-V4) are highly conserved, distinct bacterial species—or even entirely different strains with distinct dietary metabolic capabilities—may share identical 16S sequences. Furthermore, individual bacterial genomes frequently contain multiple, heterogeneous copies of the 16S gene. Therefore, while ASVs provide exact DNA sequences, true species- and strain-level taxonomic resolution

is inherently limited in these datasets. Consequently, interpreting specific bacterial shifts in response to diet using 16S data generally requires restricting confident taxonomic assignment to the genus level, reserving definitive strain-level functional mapping for WGS approaches.

1.10. Metatranscriptomics: Real-Time Dietary Reactions

Metagenomics helps us understand the potential functions of a microbial community, whereas metatranscriptomics identifies which microbes are actually reacting to what we eat. By sequencing microbial messenger RNA (mRNA), computational biologists can determine which bacterial operons become more or less active after food is consumed.

This process is technically challenging because it requires the precise removal of abundant ribosomal RNA to isolate the scarce and short-lived mRNA signals that reflect changes in dietary metabolism.

1.11. Metaproteomics: Tracking Digestive Enzymes

Metaproteomics uses liquid chromatography-tandem mass spectrometry (LC-MS/MS) to examine the actual enzymes produced by the microbiome. Analyzing this data is extremely challenging because undigested host and dietary proteins present in stool samples create significant “background noise” that can easily overwhelm the mass spectrometer’s ability to detect signals. To overcome this issue, researchers have developed advanced methods. One approach involves using artificial food particles—tiny paramagnetic glass beads coated with specific dietary fibers—that can be magnetically pulled from intestinal contents [12]. This technique cleanly isolates the bacterial enzymes working to break down specific carbohydrates.

1.12. Metabolomics: Decoding the Biochemical End-Products

Metabolomics captures a real-time snapshot of gut biochemistry by measuring all small molecules in a biological sample. In nutrition research, this technique is used to monitor metabolic products from microbial fermentation, including beneficial short-chain fatty acids (SCFAs). A major bioinformatics challenge in untargeted metabolomics is accurately distinguishing between metabolites originating from the host, those consumed directly through the diet, and compounds produced exclusively by gut bacteria—a process that requires analysts to meticulously cross-reference multiple spectral databases.

1.13. Selecting the Appropriate Meta-Omics Approach

Researchers must carefully consider different options depending on their biological goals when planning a nutritional study. 16S rRNA sequencing offers an affordable way to broadly profile taxa across large groups, but it doesn’t provide functional information or high taxonomic detail. Whole-genome shotgun (WGS) metagenomics delivers deep insights into functional potential, though it’s both

expensive and requires substantial computing power. Metatranscriptomics and metaproteomics reveal which genes are actively expressed and which enzymes are actually produced, giving genuine functional knowledge; however, these approaches struggle with problems like RNA instability and contamination from host proteins. Lastly, metabolomics can show the actual biochemical products influencing the host. However, it also introduces the challenge and complexity of pinpointing the origins of these molecules, which requires sophisticated bioinformatic techniques.

1.14. Multi-Omics Integration

Metabolites, proteins, and RNA transcripts together form the chemical language of host-microbiome interactions. Analyzing only single-omics datasets provides an incomplete view of biology; future nutritional data analysis depends on computational methods for thorough multi-omics integration.

DIABLO (**D**ata **I**ntegration **A**nalysis for **B**iomarker discovery using **L**atent **c**omponents), which uses sparse Generalized Canonical Correlation Analysis, enhances the correlation across multiple data types—such as 16S profiles, host transcriptomics, and metabolomics—to pinpoint molecular signatures strongly linked to clinical dietary responses [13]. In contrast, unsupervised models like MOFA+ (Multi-Omics Factor Analysis) use powerful variational inference techniques to uncover hidden factors driving biological differences. These approaches are particularly effective at integrating sparse datasets and revealing personalized patterns in long-term clinical dietary studies [14].

1.15. Current Bioinformatics Tools and Processing Pipelines

The raw, unstructured data generated by sequencing instruments require highly specialized software suites. The following tools represent the current standard for processing nutritional microbiome data:

- **QIIME 2:** This open-source, versatile platform is commonly used to monitor longitudinal shifts in gut microbiota within dietary intervention studies [15].
- **The bioBakery 3 Suite:** Featuring a unified analysis workflow, this suite employs *MetaPhlAn* 4.0 for precise taxonomic classification through clade-specific marker genes, and *HUMAnN* 4.0 to translate those findings into functional insights by quantifying key metabolic pathways [16].
- **MaAsLin 3:** As an advanced statistical tool for large-scale epidemiological research, MaAsLin 3 innovatively models both abundance and prevalence (presence/absence) simultaneously, effectively addressing compositional data challenges via absolute abundance benchmarks or sophisticated median-based methods [17].
- **PICRUSt2:** This tool offers an efficient way to predict microbial metabolic functions straight from 16S rRNA gene data, providing a cost-effective solution [18].
- **MicrobiomeAnalyst 2.0:** Serving as a user-friendly web resource, Microbi-

omeAnalyst 2.0 streamlines intricate analyses and incorporates the **LinDA** algorithm [19]. LinDA mitigates compositional bias with a centered log-ratio transformation and precision-weighted corrections, making it highly suitable for extensive datasets [20].

- **MMvec:** Leveraging neural networks akin to those used in natural language processing, MMvec estimates conditional probabilities of microbe-metabolite associations [21].
- **SIAMCAT:** Integrating stringent machine learning techniques, cross-validation, and predictive analytics, SIAMCAT is designed for reliable host-microbiome biomarker identification [22].

1.16. Emerging Trends in Nutritional Data Analysis

The physical technologies behind data generation and the computational models used to analyze them are rapidly evolving, offering new levels of understanding.

- **Single-cell Metagenomics (SiC-seq):** Traditional bulk metagenomics often fail to connect mobile genetic elements with their microbial hosts. SiC-seq uses droplet microfluidics to isolate single microbial cells within semi-permeable hydrogels, making it possible to map specific metabolic genes directly to unculturable bacterial hosts and avoid errors from chimeric assemblies [23].
- **Long-read Sequencing (T2T):** Short-read sequencing methods cannot adequately span highly repetitive regions in genomes. Third-generation platforms from Pacific Biosciences and Oxford Nanopore Technologies provide ultra-long reads, ushering in the Telomere-to-Telomere (T2T) era and greatly improving comprehensive and circularized genome assemblies [24].
- **Spatially Resolved Microbiomics:** The gut's structure resembles a complex biogeographical ecosystem. Advanced FISH-based spatial imaging platforms, such as 10x Xenium and NanoString CosMx, now enable researchers to examine samples at true subcellular resolution [25]. These innovations allow scientists to visualize and measure how dietary metabolites change local microenvironments at the host-microbe interface [26].
- **AI and Large Language Models (LLMs):** With meta-omics datasets growing in complexity and size, traditional modeling techniques are falling short. AI classifiers like DeepMicrobes significantly boost taxonomic prediction accuracy [27]. Microbial General Model (MGM) are foundational language models which are pre-trained on hundreds of thousands of unlabeled microbiome samples [28]. These advanced LLMs can learn intricate contextual relationships within whole ecosystems, speeding up the progression from raw data to precise nutritional insights.
- **Metagenomically Assembled Genomes (MAGs):** Reconstructing complete genomes from environmental samples requires sophisticated mathematical binning tools, including *MetaBAT 2* [29], *MaxBin 2* [30], and *CONCOCT* [31]. These are commonly combined into consensus pipelines like *metaWRAP*

[32] for improved performance.

2. Case Studies: Data-Driven Dietary Management of Disease

Applying these advanced bioinformatic theories to clinical practice strengthens the real-world impact of meta-omics data analysis.

2.1. Irritable Bowel Syndrome (IBS) in Children

Functional bowel disorders, such as pediatric IBS, are notoriously difficult to diagnose but can frequently be treated with focused carbohydrate-restricted diets like the low-FODMAP regimen. Determining which patients will benefit from this approach relies on complex algorithms that analyze patterns in the data. In a groundbreaking computational study, researchers used 16S sequencing, whole-genome sequencing (WGS), and high-throughput untargeted metabolomics, integrating these datasets into machine learning models such as Random Forest and Support Vector Machines (SVM). Their analysis was able to distinguish pediatric IBS subtypes from healthy subjects with an impressive accuracy of 98.5% [33]. The findings showed that baseline microbial profiles can reliably predict whether a child has the saccharolytic capacity needed to respond positively to these dietary therapies [34].

It is important to emphasize that these predictive models have been developed using particular pediatric cohorts, and substantial external validation is necessary before their high classification accuracy can be regarded as universally applicable in standard clinical practice.

2.2. Epilepsy and the Ketogenic Diet

The microbiota-gut-brain axis demonstrates how drastic dietary changes, like the Ketogenic Diet for pediatric epilepsy, affect systemic disease [35]. Metagenomics reveals major shifts in gut microbes, while metabolomics shows that the KD increases production of neuroactive chemicals such as SCFAs and GABA, which help reduce neuro-inflammation. Integrating these data confirms a link between dietary lipids and improved neurological stability.

Similar to the IBS models, these encouraging metabolic changes were found under strict clinical trial conditions, highlighting the importance of validating these results with larger and more diverse groups before applying them widely in clinical practice.

2.3. Pharmacomicrobiomics and GLP-1 Agonists

An emerging field in clinical bioinformatics is pharmacomicrobiomics, which explores how gut microbiota and GLP-1 receptor agonists—used to treat obesity and Type 2 Diabetes—interact with each other [36]. The body's natural GLP-1 secretion depends greatly on the individual's microbiome: certain bacteria break down dietary polysaccharides into short-chain fatty acids (SCFAs), which then activate receptors on intestinal L-cells to stimulate GLP-1 release. On the other hand,

multi-omics research shows that GLP-1 medications can alter the composition of the gut microbiome, often increasing levels of beneficial bacteria like *Akkermansia muciniphila*, known for supporting gut barrier health and reducing systemic inflammation [36].

Although these pharmacomicrobiomic interactions are very promising, further validation in larger, multi-center adult groups is still needed to confirm their long-term clinical usefulness and predictive accuracy.

3. Conclusions

In the last ten years, microbiome research has changed dramatically. The field moved beyond using basic marker gene sequencing—once mainly used to find broad connections between diet and gut bacteria—and has become highly precise and data-driven. Thanks to advances like meta-omics, single-cell sequencing, and spatial mapping, scientists now collect vast, complex datasets. This shift has pushed research from simple observation toward discovering clear, causal relationships [37]. Meta-omics technologies are now essential tools that let researchers examine the details of dietary metabolism on a molecular level.

Translating the extensive biological complexity of microbiome research into clinical application presents considerable computational challenges. The implementation of advanced bioinformatics platforms, complemented by rigorous statistical tools such as MaAsLin 3 and MicrobiomeAnalyst 2.0, has enabled researchers to effectively address compositional bias and statistical noise—factors that previously impeded reproducibility. To sustain this progress, it is essential for the scientific community to consistently follow standardized reporting guidelines, including the STORMS [38] and STREAMS [39] checklists, thereby fostering the generation of harmonized, machine-actionable data in future multinational cohort studies.

Artificial intelligence and machine learning are revolutionising precision nutrition by forecasting how individuals metabolise food, thanks to their unique microbiome data. Soon, “digital twins” of microbiomes may form the backbone of personalised healthcare, allowing virtual models to connect dietary choices with metabolic outcomes [40]. As analytical methods advance, biotherapeutic strategies are evolving—from general faecal transplants towards specifically engineered microbial communities [41]. With better computational tools and multi-omics integration, we are moving closer to customized dietary and microbial treatments for managing chronic diseases.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Menkovska, M. and Raza, S. (2025) Human Gut Microbiome in Relation to Food, Nutrition, Health and Disease. *Advances in Bioscience and Biotechnology*, **16**, 550-574. <https://doi.org/10.4236/abb.2025.1612036>

- [2] Galloway-Peña, J. and Hanson, B. (2020) Tools for Analysis of the Microbiome. *Digestive Diseases and Sciences*, **65**, 674-685. <https://doi.org/10.1007/s10620-020-06091-y>
- [3] Sirangelo, T.M. (2018) Human Gut Microbiome Analysis and Multi-Omics Approach. *International Journal of Pharma Medicine and Biological Sciences*, **7**, 52-57.
- [4] Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J. and Segata, N. (2017) Shotgun Metagenomics, from Sampling to Analysis. *Nature Biotechnology*, **35**, 833-844. <https://doi.org/10.1038/nbt.3935>
- [5] Issa Isaac, N., Philippe, D., Nicholas, A., Raoult, D. and Eric, C. (2019) Metaproteomics of the Human Gut Microbiota: Challenges and Contributions to Other Omics. *Clinical Mass Spectrometry*, **14**, 18-30. <https://doi.org/10.1016/j.clinms.2019.06.001>
- [6] Gibbons, H., O’Gorman, A. and Brennan, L. (2015) Metabolomics as a Tool in Nutritional Research. *Current Opinion in Lipidology*, **26**, 30-34. <https://doi.org/10.1097/mol.0000000000000140>
- [7] Bashiardes, S., Zilberman-Schapira, G. and Elinav, E. (2016) Use of Metatranscriptomics in Microbiome Research. *Bioinformatics and Biology Insights*, **10**, 19-25. <https://doi.org/10.4137/bbi.s34610>
- [8] Zheng, T., Ni, Y., Li, J., Chow, B.K.C. and Panagiotou, G. (2017) Designing Dietary Recommendations Using System Level Interactomics Analysis and Network-Based Inference. *Frontiers in Physiology*, **8**, Article ID: 753. <https://doi.org/10.3389/fphys.2017.00753>
- [9] Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V. and Egozcue, J.J. (2017) Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology*, **8**, Article ID: 2224. <https://doi.org/10.3389/fmicb.2017.02224>
- [10] Hickman, B. and Korpela, K. (2025) Impact of Data Compositionality on the Detection of Microbiota Responses. *Gut Microbes*, **17**, Article 2590841. <https://doi.org/10.1080/19490976.2025.2590841>
- [11] Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. and Holmes, S.P. (2016) DADA2: High-Resolution Sample Inference from Illumina Amplicon Data. *Nature Methods*, **13**, 581-583. <https://doi.org/10.1038/nmeth.3869>
- [12] Patnode, M.L., Beller, Z.W., Han, N.D., Cheng, J., Peters, S.L., Terrapon, N., *et al.* (2019) Interspecies Competition Impacts Targeted Manipulation of Human Gut Bacteria by Fiber-Derived Glycans. *Cell*, **179**, 59-73.E13. <https://doi.org/10.1016/j.cell.2019.08.011>
- [13] Singh, A., Shannon, C.P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S.J., *et al.* (2019) DIABLO: An Integrative Approach for Identifying Key Molecular Drivers from Multi-Omics Assays. *Bioinformatics*, **35**, 3055-3062. <https://doi.org/10.1093/bioinformatics/bty1054>
- [14] Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J.C., *et al.* (2020) MOFA+: A Statistical Framework for Comprehensive Integration of Multimodal Single-Cell Data. *Genome Biology*, **21**, Article No. 111. <https://doi.org/10.1186/s13059-020-02015-1>
- [15] 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. *Nature Methods*, **7**, 335-336. <https://doi.org/10.1038/nmeth.f.303>
- [16] 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. <https://doi.org/10.7554/elife.65088>

- [17] Nickols, W.A., Kuntz, T., Shen, J., Maharjan, S., Mallick, H., Franzosa, E.A., Thompson, K.N., Nearing, J.T. and Huttenhower, C. (2024) MaAsLin 3: Refining and Extending Generalized Multivariable Linear Models for Meta-Omic Association Discovery. *bioRxiv*.
- [18] Douglas, G.M., Maffei, V.J., Zaneveld, J.R., Yurgel, S.N., Brown, J.R., Taylor, C.M., *et al.* (2020) PICRUSt2 for Prediction of Metagenome Functions. *Nature Biotechnology*, **38**, 685-688. <https://doi.org/10.1038/s41587-020-0548-6>
- [19] Lu, Y., Zhou, G., Ewald, J., Pang, Z., Shiri, T. and Xia, J. (2023) Microbiomeanalyst 2.0: Comprehensive Statistical, Functional and Integrative Analysis of Microbiome Data. *Nucleic Acids Research*, **51**, W310-W318. <https://doi.org/10.1093/nar/gkad407>
- [20] Zhou, H., He, K., Chen, J. and Zhang, X. (2022) Linda: Linear Models for Differential Abundance Analysis of Microbiome Compositional Data. *Genome Biology*, **23**, Article No. 20. <https://doi.org/10.1186/s13059-022-02655-5>
- [21] Morton, J.T., Aksenov, A.A., Nothias, L.F., Foulds, J.R., Quinn, R.A., Badri, M.H., *et al.* (2019) Learning Representations of Microbe-Metabolite Interactions. *Nature Methods*, **16**, 1306-1314. <https://doi.org/10.1038/s41592-019-0616-3>
- [22] Wirbel, J., Zych, K., Essex, M., Karcher, N., Kartal, E., Salazar, G., *et al.* (2021) Microbiome Meta-Analysis and Cross-Disease Comparison Enabled by the SIAMCAT Machine Learning Toolbox. *Genome Biology*, **22**, Article No. 93. <https://doi.org/10.1186/s13059-021-02306-1>
- [23] Lan, F., Demaree, B., Ahmed, N. and Abate, A.R. (2017) Single-Cell Genome Sequencing at Ultra-High-Throughput with Microfluidic Droplet Barcoding. *Nature Biotechnology*, **35**, 640-646. <https://doi.org/10.1038/nbt.3880>
- [24] Li, H. and Durbin, R. (2024) Genome Assembly in the Telomere-to-Telomere Era. *Nature Reviews Genetics*, **25**, 658-670. <https://doi.org/10.1038/s41576-024-00718-w>
- [25] Cheng, J., Jin, X., Smyth, G.K. and Chen, Y. (2025) Benchmarking Cell Type Annotation Methods for 10x Xenium Spatial Transcriptomics Data. *BMC Bioinformatics*, **26**, Article No. 22. <https://doi.org/10.1186/s12859-025-06044-0>
- [26] Mayassi, T., Li, C., Segerstolpe, Å., Brown, E.M., Weisberg, R., Nakata, T., *et al.* (2024) Spatially Restricted Immune and Microbiota-Driven Adaptation of the Gut. *Nature*, **636**, 447-456. <https://doi.org/10.1038/s41586-024-08216-z>
- [27] Liang, Q., Bible, P.W., Liu, Y., Zou, B. and Wei, L. (2020) Deepmicrobes: Taxonomic Classification for Metagenomics with Deep Learning. *NAR Genomics and Bioinformatics*, **2**, lqaa009. <https://doi.org/10.1093/nargab/lqaa009>
- [28] Zhang, H., Zhang, Y., Kang, Z., Xiong, J., Yang, R. and Ning, K. (2026) MGM as a Large-Scale Pretrained Foundation Model for Microbiome Analyses in Diverse Contexts. *Advanced Science*, **13**, e13333. <https://doi.org/10.1002/adv.202513333>
- [29] Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., *et al.* (2019) Metabat 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction from Metagenome Assemblies. *PeerJ*, **7**, e7359. <https://doi.org/10.7717/peerj.7359>
- [30] Wu, Y., Simmons, B.A. and Singer, S.W. (2016) Maxbin 2.0: An Automated Binning Algorithm to Recover Genomes from Multiple Metagenomic Datasets. *Bioinformatics*, **32**, 605-607. <https://doi.org/10.1093/bioinformatics/btv638>
- [31] Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., *et al.* (2014) Binning Metagenomic Contigs by Coverage and Composition. *Nature Methods*, **11**, 1144-1146. <https://doi.org/10.1038/nmeth.3103>
- [32] Uritskiy, G.V., DiRuggiero, J. and Taylor, J. (2018) Metawrap—A Flexible Pipeline for Genome-Resolved Metagenomic Data Analysis. *Microbiome*, **6**, Article No. 158. <https://doi.org/10.1186/s40168-018-0541-1>

- [33] Hollister, E.B., Oezguen, N., Chumpitazi, B.P., Luna, R.A., Weidler, E.M., Rubio-Gonzales, M., *et al.* (2019) Leveraging Human Microbiome Features to Diagnose and Stratify Children with Irritable Bowel Syndrome. *The Journal of Molecular Diagnostics*, **21**, 449-461. <https://doi.org/10.1016/j.jmoldx.2019.01.006>
- [34] Chumpitazi, B.P., Cope, J.L., Hollister, E.B., Tsai, C.M., McMeans, A.R., Luna, R.A., *et al.* (2015) Randomised Clinical Trial: Gut Microbiome Biomarkers Are Associated with Clinical Response to a Low FODMAP Diet in Children with the Irritable Bowel Syndrome. *Alimentary Pharmacology & Therapeutics*, **42**, 418-427. <https://doi.org/10.1111/apt.13286>
- [35] Lindefeldt, M., Eng, A., Darban, H., Bjerckner, A., Zetterström, C.K., Allander, T., *et al.* (2019) The Ketogenic Diet Influences Taxonomic and Functional Composition of the Gut Microbiota in Children with Severe Epilepsy. *npj Biofilms and Microbiomes*, **5**, Article No. 5. <https://doi.org/10.1038/s41522-018-0073-2>
- [36] Kamath, S., Chan, N.S.L. and Joyce, P. (2026) GLP-1 Agonists and the Gut Microbiome: A Bidirectional Relationship. *British Journal of Clinical Pharmacology*, **92**, 1309-1325. <https://doi.org/10.1002/bcp.70487>
- [37] Wang, H., Huang, R., Nelson, J., Gao, C., Tran, M., Yeaton, A., *et al.* (2025) Systematic Benchmarking of Imaging Spatial Transcriptomics Platforms in FFPE Tissues. *Nature Communications*, **16**, Article No. 10215. <https://doi.org/10.1038/s41467-025-64990-y>
- [38] Mirzayi, C., Renson, A., Furlanello, C., Sansone, S., Zohra, F., Elsafoury, S., *et al.* (2021) Reporting Guidelines for Human Microbiome Research: The STORMS Checklist. *Nature Medicine*, **27**, 1885-1892.
- [39] Kelliher, J.M., Mirzayi, C., Bordenstein, S.R., Oliver, A., Kellogg, C.A., Hatcher, E.L., *et al.* (2025) STREAMS Guidelines: Standards for Technical Reporting in Environmental and Host-Associated Microbiome Studies. *Nature Microbiology*, **10**, 3059-3068. <https://doi.org/10.1038/s41564-025-02186-2>
- [40] Katsoulakis, E., Wang, Q., Wu, H., Shahriyari, L., Fletcher, R., Liu, J., *et al.* (2024) Digital Twins for Health: A Scoping Review. *npj Digital Medicine*, **7**, Article No. 77. <https://doi.org/10.1038/s41746-024-01073-0>
- [41] Pribyl, A.L., Hugenholtz, P. and Cooper, M.A. (2025) A Decade of Advances in Human Gut Microbiome-Derived Biotherapeutics. *Nature Microbiology*, **10**, 301-312. <https://doi.org/10.1038/s41564-024-01896-3>