nature reviews genetics

https://doi.org/10.1038/s41576-025-00849-8

Review article

Check for updates

Genomics of host–microbiome interactions in humans

Pamela Ferretti[1], Kelsey Johnson[2,3], Sambhawa Priya[1] & Ran Blekhman[1]

Abstract

The human microbiome is a complex ecosystem of microorganisms that inhabit the human body and have a crucial role in human health. Microbiome composition is shaped by its interaction with many factors, including human genetics. Advances in genomic technologies are improving the ability to quantify the effect of human genetics on the microbiome through improved heritability studies and microbiome genome-wide association studies (GWAS). Complementary studies using transcriptomic analyses are providing a more comprehensive view of the bidirectional relationship between host gene expression and the microbiome. The resulting insights into the genetic mechanisms driving host–microbiome interactions will ultimately contribute to the development of personalized medicine and targeted therapies.

Sections

[1]Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA. [2]Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, MN, USA. [3]Center for Genetic Epidemiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.
e-mail: blekhman@uchicago.edu

Introduction

The human microbiome is a vast community of microorganisms, including bacteria, viruses, archaea and fungi, that inhabit our bodies, particularly the gut. This complex ecosystem has a crucial role in human health and well-being. The microbiome aids in digestion, helps regulate our immune system and produces essential vitamins[1–3]. Importantly, alterations in the microbiome are linked to various health issues, including obesity, autoimmune disorders and cancer[4–8]. Thus, understanding the factors that shape the composition of the microbiome is a key focus of current biomedical research.

Microbiome composition is influenced by a myriad of environmental factors, such as diet, medication use and the host's social context[9–12]. Another factor that can shape microbiome composition is host genetics[12]. Since the early days of microbiome research, considerable effort has been dedicated to exploring the relationship between genetic factors and microbiome composition. Early family-based studies (Box 1) revealed that there is a relationship between host relatedness and the human microbiome. Specifically, studies demonstrated that individuals within the same family have more similar microbiomes than unrelated individuals, and that similarity in gut microbial communities is significantly higher in monozygotic twins than in dizygotic twins[13]. As twin pairs raised together are assumed to share a common environment, this difference is attributed to the greater genetic similarity between monozygotic twins than dizygotic twins. That the microbiome is at least partially heritable and is influenced by host genetic variation was further supported by quantitative trait locus mapping in mice, which identified genetic loci associated with the abundance of specific gut bacterial taxa[14].

More recently, advances in sequencing approaches and computational tools have improved and expanded the ways in which the heritability of the microbiome can be assessed. The reduction in sequencing costs and advances in computing power have significantly increased the number of samples included in twin studies and microbiome genome-wide association studies (GWAS), enabling large-scale analyses to be conducted in much shorter time frames. Twin studies have allowed researchers to estimate heritability by leveraging known genetic similarities[15] (Box 1). In addition, microbiome GWAS have grown in both number and sample size (Fig. 1 and Box 1). These studies not only measure heritability but also identify host genetic variation associated with microbiome composition and other microbiome quantitative traits, such as community-level diversity measures, taxa abundances, single-nucleotide polymorphisms (SNPs) and structural variants (Fig. 1 and Box 2).

Here, we review recent progress in our understanding of how human genetics is linked to the microbiome. We first discuss microbiome heritability studies and microbiome GWAS, synthesizing recent results and insights into the biological mechanisms driving validated associations. We then discuss microbial regulation of host gene expression, reviewing the latest studies in this emerging field and the molecular mechanisms underlying host gene–microbiome cross-talk. Lastly, we present an overview of current challenges in associating host genetics with variation in the microbiome and discuss the new genomic technologies and computational approaches that will shape future research.

Microbiome heritability

Heritability of complex traits is typically assessed using twin studies, GWAS and family-based designs[16] (Box 1). In this section, we discuss studies that have assessed the heritability of the gut microbiome in humans and animals. We focus exclusively on the gut microbiome, as the gut has been the primary target of microbiome heritability studies because of its accessibility, high microbial density and established links to human health and disease. We review key findings from twin and family-based studies, along with recent large-scale analyses that provide insights into the genetic component of microbiome variation. These studies collectively reveal the extent to which host genetics shapes microbiome composition across different populations and environmental contexts.

Microbiome heritability in humans

Early investigations of human microbiome heritability primarily utilized 16S rRNA gene amplicon sequencing in twin-based or family-based study designs. These pioneering studies established methodological frameworks for quantifying genetic contributions to microbiome composition and identified initial taxa showing significant heritability patterns. A landmark study that estimated the heritability of several gut bacterial taxa in 416 twin pairs from the UK found that the family Christensenellaceae was the most highly heritable taxon ($h^2 = 0.39$) but also reported significant heritability for other taxa, including *Methanobrevibacter* and *Blautia*[15]. A follow-up study with an expanded cohort of twins found that of 945 shared taxa, 8.8% were significantly heritable, with heritability estimates greater than 0.2 (ref. 17). Moreover, longitudinal sampling showed a positive correlation between heritability and temporal stability[17]. A GWAS in the Hutterite population identified several bacterial taxa with significant heritability, including *Akkermansia*, a genus related to obesity, and *Bifidobacterium*, a genus with known health-promoting properties[18,19]. These early investigations consistently revealed moderate heritability for various gut microbiome features, typically ranging from 0.1 to 0.4 for individual taxa or community metrics. Notably, heritability estimates varied considerably among different bacterial taxa, suggesting that host genetic influences are not uniform across the microbiome. Some taxa with potential health benefits, such as Christensenellaceae and *Methanobrevibacter*, consistently showed high heritability, indicating robust host genetic control over these microorganisms[20,21].

More recent microbiome heritability studies used larger sample sizes, diverse population cohorts and extensive metadata to quantify microbiome heritability, which increased their statistical power (Fig. 2a). Moreover, use of shotgun metagenomic sequencing enabled species-level profiling and functional analysis of the entire microbial community. A large-scale GWAS of more than 1,000 Israeli individuals from different self-reported ancestries that combined gut microbiome data with host genotyping found that the environment had a dominant role in shaping the microbiome, with only about 2% of microbial taxa showing significant heritability[22]. Another GWAS, which used a multi-ethnic cohort to explore the interaction between host genetics, ancestry and the gut microbiome, identified several ancestry-specific microbial associations, highlighting the importance of considering population diversity in microbiome research[23]. Similarly, analysis of microbiome and host genetic data in two twin cohorts that were integrated as part of the MiBioGen consortium[24] found 19 taxa with significant heritability values[25]. Importantly, the heritable taxa were enriched among the taxa associated with host genetic variation. In another large GWAS, gut microbiome heritability was found to vary across ethnicities[26]. Despite identifying a large number of microbial taxa as significantly heritable in the full cohort, only two taxa had heritability greater than 0.2: *Bacteroides uniformis* and Prevotellaceae. A study using gut microbiome, genetic, lifestyle and diet data from

Box 1 | Methodological approaches for studying host–microbiome interactions

Host–microbiome interactions can be studied through various approaches, each with distinct advantages and limitations for revealing different aspects of this complex relationship.

Family-based and twin studies. These study designs leverage known genetic relationships to estimate heritability of microbiome traits. In twin studies, heritability is typically calculated by comparing trait concordance between monozygotic twins (who have 100% genetic similarity) and dizygotic twins (50% genetic similarity on average), with higher concordance in monozygotic twins indicating genetic influence[15]. Family-based designs examine microbiome similarity across related individuals with varying degrees of genetic relatedness[18]. Both approaches help distinguish genetic from environmental influences but require careful control of shared environmental factors, such as diet and household exposures, which can confound the estimation of genetic effects.

Microbiome GWAS. Genome-wide association studies (GWAS) identify specific host genetic variants associated with microbiome features by testing millions of host genetic variants against microbial traits[23,42]. The microbiome is characterized using either 16S rRNA gene amplicon sequencing, providing taxonomic composition at a broad level, or shotgun metagenomic sequencing, which offers species-level resolution and information on the functional potential and genetic variation within microbial genomes. Microbial traits in GWAS can be defined in multiple ways: as binary traits (presence or absence of taxa), quantitative measures (relative abundance), diversity indices or functional pathways (Box 2). Statistical approaches must account for the non-normal distribution of microbiome data, often employing linear models with appropriate transformations, zero-inflated models for sparse data or non-parametric tests[120,152]. In addition to identifying specific loci, GWAS data can also be used to estimate heritability by quantifying the proportion of phenotypic variance explained by all measured genetic variants, providing insights into the overall genetic architecture of microbiome traits[119]. These heritability estimates are often lower than those from twin studies, as they typically capture only the contribution of common genetic variants and may miss rare variants or structural genomic elements that influence the microbiome.

Transcriptome-based approaches. These methods investigate how the microbiome influences host gene expression and vice versa. Bulk RNA sequencing (RNA-seq) measures average gene expression across a population of cells within a tissue in association with microbiome features, providing a broad view of host transcriptional changes[75,85]. For higher resolution, single-cell RNA sequencing (scRNA-seq) reveals host cell type-specific correlations to microbiome signals, allowing researchers to understand which host cell populations are most affected by microbial presence. Spatial transcriptomics takes this further by mapping host gene expression changes in specific tissue regions with distinct microbial colonization patterns, preserving the spatial context of host–microbiome interactions[93,114]. Additionally, expression quantitative trait locus analysis can identify host genetic variants that influence both host gene expression and microbiome traits, potentially uncovering mechanistic links between host genetics, gene regulation and microbial communities[153].

Animal models and experimental validation. Animal models provide useful systems for both identifying and validating host–microbiome interactions. Quantitative trait locus mapping in model organisms identifies host genetic loci associated with microbial traits by leveraging controlled breeding and greater environmental standardization than is possible in human studies[14,35]. These approaches use crosses between inbred lines or diverse animal populations to map specific genomic regions that influence microbiome composition or function. In addition to genetic mapping, animal models enable functional validation of microbiome–host interactions to establish causality beyond associations. Gnotobiotic animal models with defined microbial communities allow testing of specific hypotheses about how microorganisms influence host physiology in a controlled setting[66]. Organoid co-culture systems provide an alternative approach using 3D tissue cultures exposed to specific microorganisms or communities, enabling the study of epithelial responses to microbial signals[118]. CRISPR-based screens offer systematic perturbation of host genes to identify those affecting microbiome composition or response to microbial presence.

Multi-omic integration. Combining multiple host genomic and microbiome data types can provide a comprehensive view of host–microbiome relationships. These approaches integrate various omic layers, such as host transcriptomics (including bulk RNA-seq or scRNA-seq and spatial transcriptomics), metagenomics, metatranscriptomics, metaproteomics and metabolomics, and use machine-learning approaches to capture complex patterns across high-dimensional and heterogeneous datasets, offering a more complete picture of host–microbiome dynamics[124].

thousands of families found that only around 6.6% of gut microbial taxa are heritable, and that the most heritable taxon, Proteobacteria, had a heritability estimate of 0.3 (ref. 27). More recently, studies have started investigating the heritability of traits other than taxa abundance; for example, quantification of the heritability of structural variation in the gut microbiome showed that dozens of structural variants are potentially heritable microbial traits[28].

Together, these recent studies show that microbiome heritability is overall moderate, and lower than that found in earlier studies with smaller sample sizes. Although there is some variability in the taxa identified as heritable across studies, the heritability values are mostly in the same range (Fig. 2a). Moreover, some taxa show high heritability across studies; for example, Christensenellaceae, which was also found to be heritable in earlier studies[15,17,29], and Bifidobacteria, which was found to be heritable across several studies and populations, with a specific association near the human *LCT* gene (see 'The *LCT* locus, diet and gut *Bifidobacterium*' section)[25,30]. In addition, the overall distribution of heritability estimates across microbiome traits is relatively similar across studies, and comparable with the distribution of heritability estimates for hundreds of non-microbiome complex traits in the UK

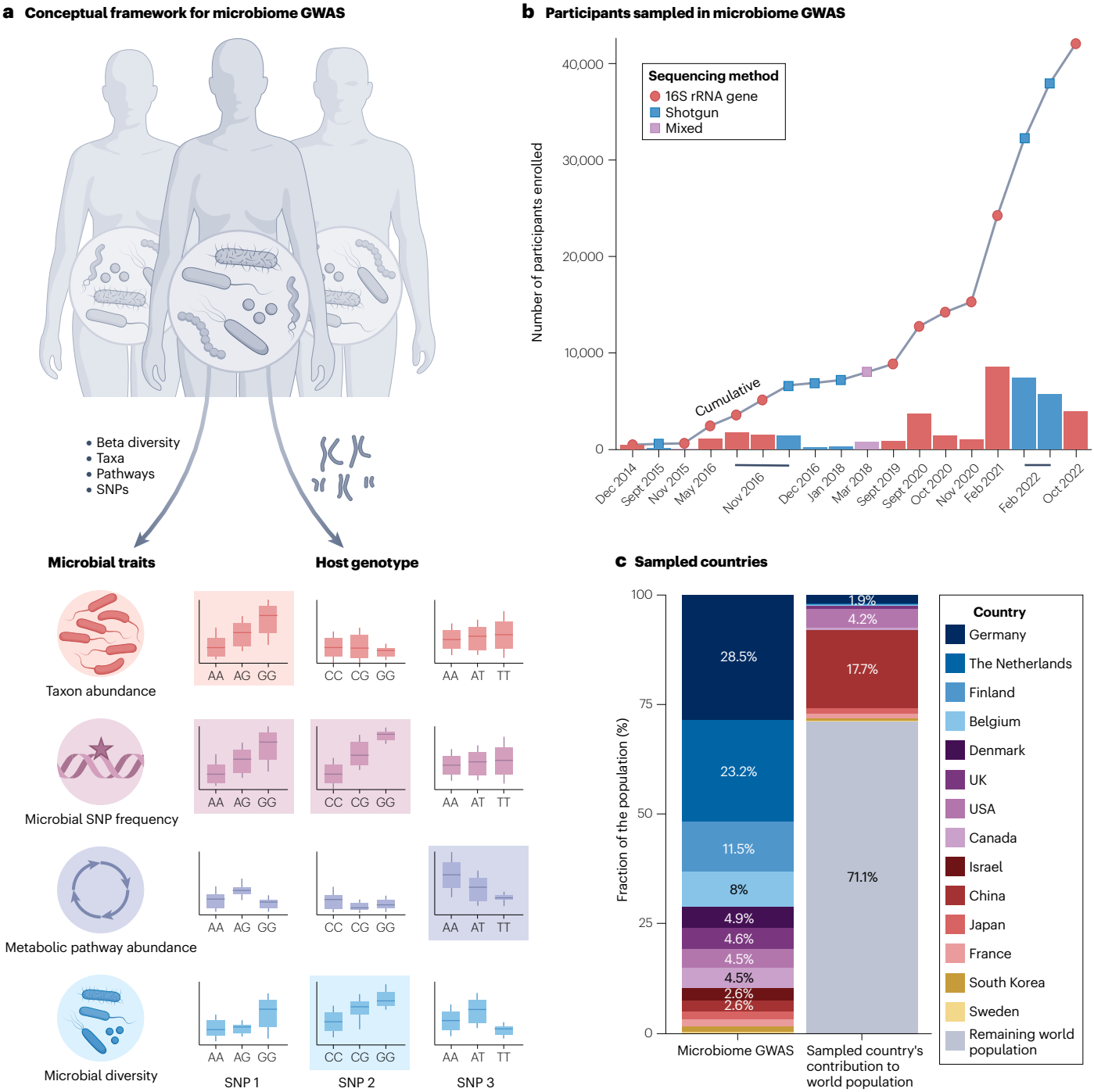


Fig. 1 | Overview of microbiome GWAS and their conceptual framework. **a**, Conceptual framework for identifying host–microbiome genomic associations in microbiome genome-wide association studies (GWAS). Microbiome GWAS aim to identify associations between genetic variations in the host's genome and specific microbial traits (such as taxa, pathways or microbial single-nucleotide polymorphisms (SNPs)). **b**, The number of participants sampled in microbiome GWAS, per study (bars) and cumulatively (line). Studies based on 16S rRNA gene amplicon sequencing (red circles), studies based on shotgun metagenomic sequencing (blue squares) and the study relying on both methodologies (purple) are highlighted. The number of sampled participants reflects what was reported in the original studies, therefore overlap between participants across different studies is possible and is not considered in this representation. The number of participants enrolled in microbiome GWAS has grown over time, with several studies now including more than 5,000 participants. **c**, Countries sampled by the publicly available human gut microbiome GWAS (left) and their representation in relation to the global human population (right).

Box 2 | Defining microbiome-based traits for heritability and GWAS analysis

Similar to any complex trait, such as height or gene expression, the heritability of microbiome-based phenotypes can be quantified. Similarly, genome-wide association studies (GWAS) can be used to characterize the genetic bases of microbiome traits. However, defining microbiome phenotypes presents unique challenges because of the complex nature of microbial communities; unlike many other complex traits studied in humans, the microbiome is not a single entity but, rather, a diverse ecosystem of phylogenetically linked microorganisms, each with its own genome and functions, existing in varying abundances[151].

Despite these complexities, researchers have developed several strategies to define microbiome traits for heritability analysis. One common approach involves using community summaries derived from dimensionality reduction techniques. For example, scientists often use principal coordinate analysis to identify major axes of variation within microbial communities. These axes can then be treated as quantitative traits for heritability studies[15]. Another frequently used method focuses on diversity measures, particularly alpha diversity. These metrics quantify the complexity of a microbial community by considering the number of different taxa present. More sophisticated measures also take into account the relative abundances of these taxa and their phylogenetic relationships (for example, the Shannon diversity index and Faith's phylogenetic diversity[154]).

Researchers also examine the abundance of individual taxa as separate complex traits. This approach allows for a more granular analysis of heritability, as each taxon's relative abundance can be treated as a quantitative trait. However, it is crucial to note that when working with relative abundance data, compositional data analysis techniques may be necessary to account for the inherent dependencies in such data[155].

In some cases, scientists opt for a simpler binary approach, looking at the presence or absence of specific taxa. This method can be particularly useful for rare taxa or when dealing with less detailed taxonomic classifications. However, caution is advised when applying this approach to taxa with very low abundances, as the depth of sequencing coverage can significantly influence whether a taxon is detected or not, potentially leading to false negatives[152].

For more advanced studies, particularly those utilizing shotgun sequencing data, researchers can investigate the heritability of structural variants in the microbial genome[28] or the heritability of functional profiles. The latter involves examining the abundance or presence of specific genes or entire metabolic pathways within the microbiome[156].

Biobank (Fig. 2b). Interestingly, heritable taxa do not share consistent phylogenetic or functional characteristics. This pattern suggests that heritability may be driven by host genetic factors that influence specific microenvironmental conditions in the gut rather than direct selection for particular taxonomic groups, potentially explaining why diverse and functionally distinct taxa can exhibit similar heritability patterns. Lastly, all studies emphasize the substantial role of environmental factors in shaping the microbiome, consistent with earlier research.

Microbiome heritability in animals

Microbiome heritability studies also reveal some inconsistencies and contradictions, and the magnitude of genetic influence on the microbiome has been a notable point of contention[22,31]. For example, the identity of significantly heritable taxa varies across studies (Fig. 2a), with most taxa fluctuating in their apparent genetic influence depending on the population studied. These inconsistencies reflect some of the common challenges associated with studying microbiome heritability in humans. The microbiome is influenced by environmental factors that can sometimes also be associated with genetic variation. For example, diet has a strong effect on the microbiome[32], but dietary preference is itself a heritable trait[33]. Additionally, it is challenging to accurately and comprehensively measure dietary composition in humans. Moreover, the microbiome is a dynamic trait, which can change dramatically day to day[34]. Accounting for this through longitudinal sampling is logistically challenging in humans, and most microbiome heritability studies have a single sample per individual. Similarly, it is challenging to account for other effects, such as transmission of microbial strains between individuals or the environment.

Non-human study systems, in which it is possible to quantify and account for environmental effects, have been useful in overcoming these challenges and improving our understanding of the heritability of microbiome traits. For example, inbred mice reared in a controlled environment show a high degree of microbiome heritability, with host genetics explaining more than 0.5 of the variation in abundance of many common taxa[35]. Although such controlled environments limit environmental variance, these experiments show that, under the right circumstances, genetic variation can have considerable effects on the microbiome. Pigs represent another useful model system to study microbiome heritability. A previous study compared the faecal and caecal microbiome of full siblings, half-siblings and unrelated individuals from two pig populations that were raised under the same conditions and fed a similar diet. The analysis identified heritable microbial taxa (81 in faeces and 67 in caecum) that had a heritability value greater than 0.15, with 31 of these taxa showing consistent heritability in both sample types, suggesting that host genetic effects on gut microbiota are relatively stable across different gut locations[36]. Similarly, a study using hybrid mouse lines to estimate the heritability of the mucosa-associated microbiome found that 21 taxa show significant heritability, with heritability values as high as 0.83 (ref. 37). Lastly, research in a wild baboon population shows that intensive sampling across multiple time points may help identify heritable variation in the face of environmental heterogeneity, such as seasonal changes in food availability and fluctuating social dynamics[31]. More than 95% of gut microbial taxa were significantly heritable, but with modest effect sizes (median <0.1, maximum slightly greater than 0.2), illustrating how the statistical power gained from repeated longitudinal sampling enables the detection of subtle genetic effects on the microbiome that may otherwise be missed with single time-point studies.

Taken together, human and animal studies highlight several important considerations in microbiome heritability analysis: the inclusion of longitudinal data can have a substantial impact on heritability estimates, as this allows controlling for day-to-day variation in gut microbiome composition; the ability to account for environmental effects, especially diet, is critical in accurately estimating

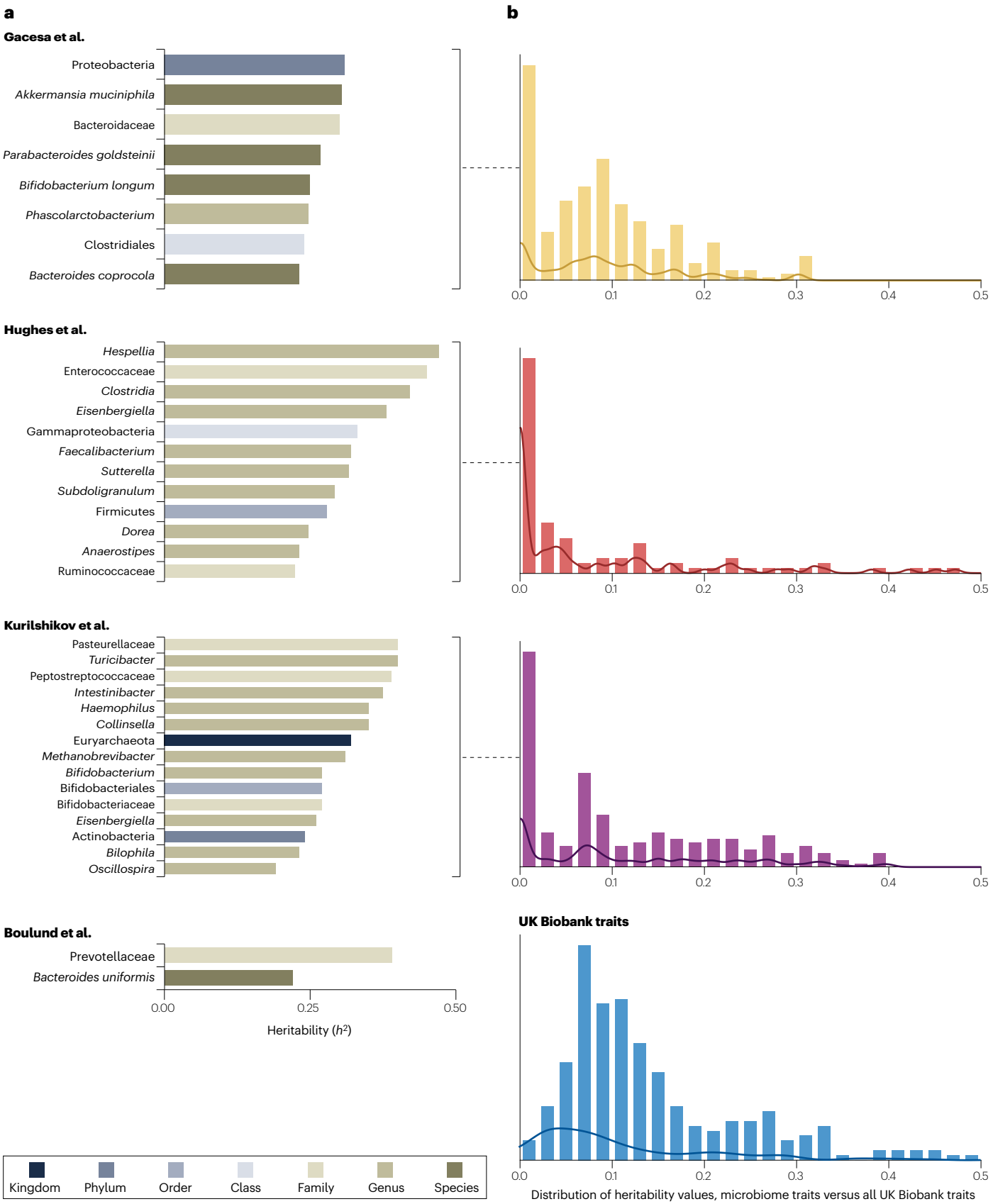

a
b
Gacesa et al.
Proteobacteria
Akkermansia muciniphila
Bacteroidaceae
Parabacteroides goldsteinii
Bifidobacterium longum
Phascolarctobacterium
Clostridiales
Bacteroides coprocola
Hughes et al.
Hespellia
Enterococcaceae
Clostridia
Eisenbergiella
Gammaproteobacteria
Faecalibacterium
Sutterella
Subdoligranulum
Firmicutes
Dorea
Anaerostipes
Ruminococcaceae
Kurilshikov et al.
Pasteurellaceae
Turicibacter
Peptostreptococcaceae
Intestinibacter
Haemophilus
Collinsella
Euryarchaeota
Methanobrevibacter
Bifidobacterium
Bifidobacteriales
Bifidobacteriaceae
Eisenbergiella
Actinobacteria
Bilophila
Oscillospira
Boulund et al.
Prevotellaceae
Bacteroides uniformis
0.00
0.25
0.50
Heritability (h^2)
0.0
0.1
0.2
0.3
0.4
0.5
UK Biobank traits
Kingdom
Phylum
Order
Class
Family
Genus
Species
Distribution of heritability values, microbiome traits versus all UK Biobank traits

Fig. 2 | Overview of microbiome heritability results. a, Heritability estimates of bacterial taxa across genome-wide association studies (GWAS), focusing on the four largest human GWAS with publicly accessible heritability data[23,25–27]. Taxa are organized by study and coloured according to taxonomic level (kingdom, phylum, class, family, genus and species). Despite some variability in the specific taxa identified as heritable, most studies report similar ranges of heritability values, with estimates typically between 0.1 and 0.4. **b**, Distribution of the heritability estimates of bacterial taxa across microbiome studies compared with heritability estimates of non-microbiome traits in the UK Biobank (as of April 2019). The overall distribution of heritability values is similar between microbiome traits and other complex human traits, suggesting that the genetic architecture of microbiome features follows patterns observed in other heritable phenotypes. Notably, most microbiome traits show low to moderate heritability, with a small number of taxa exhibiting higher heritability values, consistent with a complex trait influenced by both genetic and environmental factors.

microbiome heritability; and the heritability of microbiome traits is context-dependent, and can vary considerably between host species, populations, seasons and environments.

Microbiome-associated host genetic loci

Studies that identified quantitative trait loci associated with abundances of microbial taxa in the mouse gut demonstrated that genome-wide approaches could be used to identify host genetic loci associated with microbiome composition[14,38,39]. The first human microbiome GWAS in 2015 used relatively small sample sizes of fewer than 100 individuals[18,30]. Subsequent microbiome GWAS have used larger cohorts, increased the diversity of participants beyond those of European ancestry and shifted from 16S rRNA gene amplicon to shotgun metagenomic sequencing, enabling genetic association testing of species-level taxa and microbial genetic pathway abundances (Fig. 1b,c). Although the majority of microbiome GWAS have focused on the gut microbiome[17,23,25,26,40–51], numerous studies have applied this methodology to additional body sites, including skin[52], nasal[53] and milk[54] microbiota, providing evidence for host genetic influences across the human body. As microbiome GWAS of other body sites tend to be small and have not yet been replicated by multiple studies, we focus on insights from gut microbiome GWAS with a sample size of at least 1,000 participants (Table 1 and Supplementary Table 1). Taken together, these 14 studies have identified 34 genetic loci that have genome-wide significant associations with gut microbiome traits in at least 2 studies ($P < 5 \times 10^{-8}$) (Fig. 3a and Supplementary Table 2). Here, we discuss potential mechanisms controlling host gene–microorganism interaction at two of these loci that are associated with the same microbial traits across four studies — the lactase gene (*LCT*) locus and the *ABO* locus (Fig. 3a).

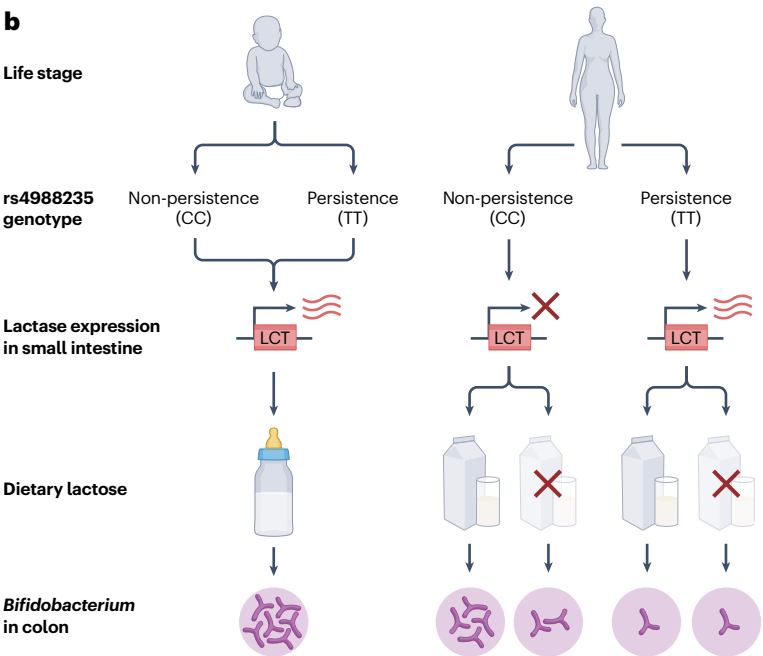
The *LCT* locus, diet and gut *Bifidobacterium*

An association between genetic variants near the *LCT* locus and *Bifidobacterium* spp. in the gut was first reported in one of the original small microbiome GWAS[30] and has since been replicated at genome-wide significance in multiple European ancestry cohorts[17,25,47,48]. The genetic variants associated with *Bifidobacterium* abundance include the functional variant (rs4988235) that confers lactase persistence (that is, the continued expression of lactase in the small intestine into adulthood) and nearby linked variants in European

Table 1 | Summary of microbiome GWAS with more than 1,000 participants

Study	Cohort size	Method	Country	Genome-wide significant loci ($P<5\times10^{-8}$)	Ref.
Goodrich et al. (2016)	1,126 twin pairs	16S rRNA gene sequencing	UK	22 taxa	139
Turpin et al. (2016)	1,561	16S rRNA gene sequencing	Canada, USA, Israel	55 taxa	40
Bonder et al. (2016)	1,514	Shotgun metagenomic sequencing	The Netherlands	9 taxa, 21 pathway, 12 gene ontology	41
Wang et al. (2016)	1,812	16S rRNA gene sequencing	Germany	40 taxa, 42 beta diversity	42
Hughes et al. (2020)	2,223+950+717	16S rRNA gene sequencing	Belgium, Germany	13 taxa	23
Xu et al. (2020)	1,475	16S rRNA gene sequencing	China	10 taxa, 1 beta diversity	150
Ishida et al. (2020)	1,068	16S rRNA gene sequencing	Japan	0	43
Rühlemann et al. (2021)	8,956	16S rRNA gene sequencing	Germany	34 taxa, 4 beta diversity	44
Kurilshikov et al. (2021)	18,340	16S rRNA gene sequencing	Many	31 taxa	25
Liu et al. (2021)	1,295 (632+663 replication)	Shotgun metagenomic sequencing	China	36 taxa, 8 pathway, 4 beta diversity	45
Liu et al. (2022)	3,432 (1,539+1,430 replication)	Shotgun metagenomic sequencing	China	230 taxa, 312 pathway, 6 taxa and pathway	46
Lopera-Maya et al. (2022)	7,738	Shotgun metagenomic sequencing	The Netherlands	6 taxa, 14 pathway	47
Qin et al. (2022)	5,959	Shotgun metagenomic sequencing	Finland	422 taxa	48
Boulund et al. (2022)	4,117	16S rRNA gene sequencing	The Netherlands	51 taxa	26

GWAS, genome-wide association studies.



c

FUT2 secretor status	Non-secretor		Secretor	
ABO blood type	O	A/AB/B	O	A/AB/B
Collinsella abundance				
Lactose and galactose degradation pathway abundance	+	+	+	++
ABO blood type	O/B	A/AB	O/B	A/AB
Faecalibacterium prauznitzii SV 577–579 frequency	+	+	+	++

Fig. 3 | Overview of genetic loci identified in gut microbiome GWAS. **a**, Summary of genetic loci with genome-wide significant associations ($P < 5 \times 10^{-8}$) in at least two gut microbiome genome-wide association studies (GWAS). Summary statistics from gut microbiome GWAS including at least 1,000 participants were included (Table 1). Loci were defined by clustering genetic associations within 100,000 base pairs of each other and are labelled with the name of the gene nearest to the variant with the smallest *P* value, except at the *ABO* and lactase gene (*LCT*) loci where the causal gene is known and labelled. Data for this figure are presented in Supplementary Tables 1 and 2. **b**, Flow chart illustrating gene-by-diet and gene-by-age interactions for the *LCT* association with gut *Bifidobacterium*. In infancy and early childhood, lactase is highly expressed in the small intestine regardless of the lactase persistence genotype, and *Bifidobacterium* spp. are the most abundant gut microorganisms. In lactase non-persistent adults, high dairy consumption is associated with higher levels of gut *Bifidobacterium* spp., whereas lower dairy consumption is associated with lower levels of *Bifidobacterium*. In lactase-persistent adults, there is no association between dairy consumption and *Bifidobacterium* levels. **c**, Gene–gene interactions between FUT2 secretor status and ABO blood group are associated with several microbial features related to antigen utilization: *Collinsella* abundance, the microbial lactose and galactose degradation pathway, and a structural variant in *Faecalibacterium prausnitzii* that contains an *N*-acetylgalactosamine (GalNAc) degradation gene cluster.

populations[55]. Initial comparisons of individuals with and without this persistence-associated genotype (that is, lactase persisters and non-persisters, respectively) showed that non-persisters had higher levels of *Bifidobacterium*[17,30]. In addition, multiple studies have found evidence for an interaction between host genotype and dairy consumption in the association with *Bifidobacterium*: in individuals with the non-persister genotype, *Bifidobacterium* abundance is positively correlated with dairy intake, whereas no relationship between diet and *Bifidobacterium* is observed in individuals homozygous for the lactase persistence allele[41,47,48] (Fig. 3b). These observations suggest that, in lactase persisters, dietary lactose is metabolized by lactase in the small intestine, thus providing no additional benefit to lactose-utilizing bacteria such as Bifidobacteria in the colon[48]. However, in adults expressing little lactase in their small intestine, dietary lactose makes its way to the colon, where it can be utilized by *Bifidobacterium*. This model is supported by an additional interaction with age, whereby the association between the lactase persistence genotype and Bifidobacteria is attenuated in younger individuals[25] owing to the age dependence of the lactase persistence phenotype[56] (Fig. 3b). Although a GWAS of gut microbiome composition specifically in infants is lacking, *Bifidobacterium* spp. are the most abundant microorganisms in infants, with breastfed infants tending to have higher *Bifidobacterium* levels than formula-fed infants[57] (Fig. 3b). The association between *LCT* and *Bifidobacterium* has not been found in cohorts of non-European populations, likely because of differences in lactase persistence alleles between populations[26,43,45,46]. A dataset from a cohort that included individuals of Ghanaian and African Surinamese ancestry included only one genetic variant associated with lactase persistence in sub-Saharan African groups[26], limiting the ability to test for associations with *Bifidobacterium* in the *LCT* region. The lack of *Bifidobacterium*–*LCT* associations in cohorts of East Asian ancestry is also consistent with the observation that the known lactase persistence alleles are essentially absent from populations of East Asian ancestry[58].

ABO blood group and gut microbiome

Polymorphisms in the *ABO* gene determine an individual's ABO blood group: that is, their ability to produce A-antigen, which contains a terminal *N*-acetylgalactosamine (GalNAc); B-antigen which contains a terminal galactose; and O-antigen, which lacks the addition of this terminal monosaccharide[59]. Associations between ABO blood group and gut microbiome traits were first observed in a candidate gene study[60]. This association was later identified in a microbiome GWAS in a German cohort[48] and subsequently replicated in European and Chinese cohorts[44,46–48]. Microbiome GWAS generally assume that genetic associations at this highly polymorphic locus are tagging the linked variants that determine an individuals' ABO alleles. By inferring each participant's ABO blood group from their genotype at the multiple variants known to tag this phenotype, studies then directly test for associations between ABO blood group and microbial phenotypes[44].

The microbial taxa and pathways associated with the *ABO* locus are more heterogeneous than the *Bifidobacterium* association at the *LCT* locus, but mounting evidence suggests a common mechanism of mucosal antigen utilization among the associated microbial traits. Microbial taxa and genetic pathways associated with the *ABO* locus in at least two studies include *Collinsella*[47,48], the lactose and galactose degradation pathway[46,47] and *Faecalibacterium*[28,44]. Importantly, microbiome associations at this locus are dependent on an interaction with the individual's FUT2 secretor genotype (Fig. 3c). *FUT2* encodes a fucosyltransferase that is required for mucosal secretion of ABO antigens, with non-secretors lacking mucosal ABO antigen[59]. Associations between the gut microbiome and ABO blood group are only observed in secretors[47,48], suggesting that, when present, mucosal ABO antigens promote the growth of different mucosal antigen-utilizing bacteria. FUT2 secretor status has also been previously associated with the composition of the gut microbiome in candidate gene studies in mice[61] and humans[62–65].

Large-scale shotgun metagenomic sequencing enables association analyses between host genetic variation and microbial genetic variation, an approach that recently shed additional light on the *ABO* association with the gut microbiome. The first GWAS of human host genotypes and microbial structural variants, performed in four Dutch cohorts, found a higher frequency of structural variants in *Faecalibacterium prausnitzii* (Fig. 3c) in individuals with A or AB blood types, which was replicated in a Tanzanian cohort[28]. The associated *F. prausnitzii* structural variants contain a cluster of genes in the GalNAc utilization pathway, the terminal monosaccharide in A-antigen. This study further showed that only *F. prausnitzii* strains with this GalNAc utilization gene cluster grew on media with GalNAc as the sole carbon source and that the GalNAc utilization pathway is present across other taxa previously associated with the ABO locus, such as *Collinsella aerofaciens*[47]. Moreover, the abundance of this microbial gene cluster was found to be highly correlated with the individuals' A-antigen secretor status. The study of the gut microbiome in a large mosaic pig population discussed above (see 'Microbiome heritability in animals' section) similarly identified an association between a deletion in the porcine *N*-acetyl-galactosaminyltransferase gene underlying the ABO blood group in humans and species of Erysipelotrichaceae[66]. This study showed that pigs with the deletion have a null allele resulting in reduced GalNAc in their intestine and lower levels of Erysipelotrichaceae. Moreover, this study found that the genome of the associated species encodes genetic pathways to utilize GalNAc but are unable to regulate their expression in response to GalNAc levels, suggesting that these species are at a disadvantage in a low GalNAc environment such as in

pigs homozygous for the null allele. Furthermore, the authors present evidence that balancing selection has occurred at this gene across porcine species, as has previously been shown in primates[67]. Although the associated bacterial species differ between humans and pigs, the evidence supports a similar mechanism across species whereby a common polymorphism under balancing selection causes variation in intestinal GalNAc levels leading to differential abundances of GalNAc-utilizing gut bacteria.

The *LCT* and *ABO* examples illustrate two mechanisms by which host genetic variation may shape microbial fitness and, subsequently, the frequency of specific microbial taxa. As larger and more diverse cohorts with host and microbial genomic sequence data become available, more powerful studies with additional types of microbial genetic variation (for example, single-nucleotide variants) will become possible.

Mendelian randomization and the microbiome

Observational associations have been reported between gut microbiota and a wide range of health outcomes, but whether these associations are causal has been difficult to determine. Mendelian randomization is a statistical approach to test for putative causal relationships between an exposure and an outcome by testing whether genetic variants associated with the exposure also correlate with the outcome of interest[68–70]. Mendelian randomization has recently been widely applied to test for relationships between abundances of gut microbial taxa or genetic pathways (exposures) and health outcomes, by utilizing the output of microbiome GWAS.

Several studies have implemented Mendelian randomization to identify potential causal effects of microbial abundances and validated their results in independent cohorts. Mendelian randomization analysis of the LifeLines DEEP cohort of 952 individuals from The Netherlands, which includes host genotype data, gut shotgun metagenomic sequencing data and faecal short-chain fatty acid (SCFA) profiles, identified putative causal relationships between microorganisms, SCFAs and insulin secretion. In particular, the study found evidence that the microbial GABA degradation pathway was causally associated with insulin secretion during an oral glucose challenge test, with higher pathway abundance associated with improved insulin response[71]. The GABA degradation pathway produces butyrate, an SCFA with potential anti-diabetic effects[72]. Another study used two Chinese cohorts to perform Mendelian randomization testing for causal impacts of microbiome traits on blood metabolites measured in the same cohorts[46]. The authors identified 58 putative causal relationships in their primary cohort and were able to replicate 43 of these associations in their second cohort. The strongest signal represented an increase in *Oscillibacter* linked to decreased blood triglycerides and alanine, as well as lower body mass index and waist to hip ratio. Both of these studies applied bidirectional Mendelian randomization, testing for the effect of the microbiome trait on the health outcome and vice versa, to elucidate the direction of effect. These studies highlight the potential for Mendelian randomization to link microorganisms with health.

The examples above represent one-sample Mendelian randomization studies, in which genotypes, exposure and outcome have been measured in a single study. The wide availability of GWAS summary statistics from microbiome and other trait GWAS has enabled an explosion in two-sample Mendelian randomization studies, which test for potential causal relationships between gut microbiota and health outcomes by comparing summary statistics from separate microbiome and outcome GWAS. In any Mendelian randomization analysis, genetic associations with the exposure (that is, the microbiome trait) should ideally be strongly and replicably associated with microorganism abundance and modify the outcome only through microorganism abundance and not through a parallel pathway involving an intermediate trait associated with the genetic variant. However, the relatively small sample size and limited number of microbiome GWAS means that there is a paucity of high-confidence genetic associations with microbial traits to utilize in the Mendelian randomization framework, making it challenging to address microbiome traits with this method[73,74] and leading many existing studies to use a lenient statistical threshold for inclusion of microbiome-associated genetic variants. Moving forward, larger sample sizes should enable future studies to employ standard thresholds and additional evidence to support Mendelian randomization results with microbiome traits.

Host gene regulation and the microbiome

In addition to identifying associations between host genetics and the microbiome, studies have also explored the relationship between host gene expression and the microbiome. To understand host transcription–microbiome associations, researchers have used a range of techniques, including bulk RNA sequencing (RNA-seq), single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics, to profile host gene expression, and jointly analysed these data with microbiome features to uncover key insights into host–microbiota interactions (Fig. 4 and Box 1). These studies have revealed associations between microbiome composition and host gene expression, including host genes whose expression levels correlate with the presence, absence or changes in abundance of specific bacterial taxa or taxonomic groups. Whereas microbiome GWAS suggest that host genetics, likely through host gene regulation, shape the composition of the microbiome, studies linking host gene expression to the microbiome highlight a bidirectional cross-talk, in which microbial communities both influence and respond to host transcriptional activity through diverse biological pathways. Complementary to microbiome GWAS, host gene expression–microbiome interaction data capture dynamic tissue-specific and condition-specific responses, providing insights into active biological processes and functional mechanisms linking microorganisms to host biology.

Host transcript–microbiome interactions

To determine functional interactions between the microbiome and host genome, recent studies have characterized associations between host gene expression and the microbiome across body sites and diseases (Fig. 5 and Supplementary Table 3). Studies investigating these relationships in the gastrointestinal tract have uncovered connections between the gut microbiome and host transcripts that potentially contribute to the pathogenesis of human gastrointestinal disorders[75–82]. An early study that examined how host gene expression and the mucosal microbiome interact in patients who underwent ileal pouch–anal anastomosis for ulcerative colitis found that host transcripts that were most strongly associated with microbial abundances were enriched for genes related to the complement cascade and IL-12 pathways[75]. These host pathways were inversely correlated with the abundance of beneficial microorganisms such as *Sutterella*, *Akkermansia* and *Bifidobacterium*, and positively correlated with the abundance of *Escherichia*, a frequently over-represented bacteria in inflammatory bowel disease (IBD)[83]. A longitudinal analysis of gut microbiome dynamics in individuals with IBD found that the expression of mucosal chemokine genes with antimicrobial properties, such

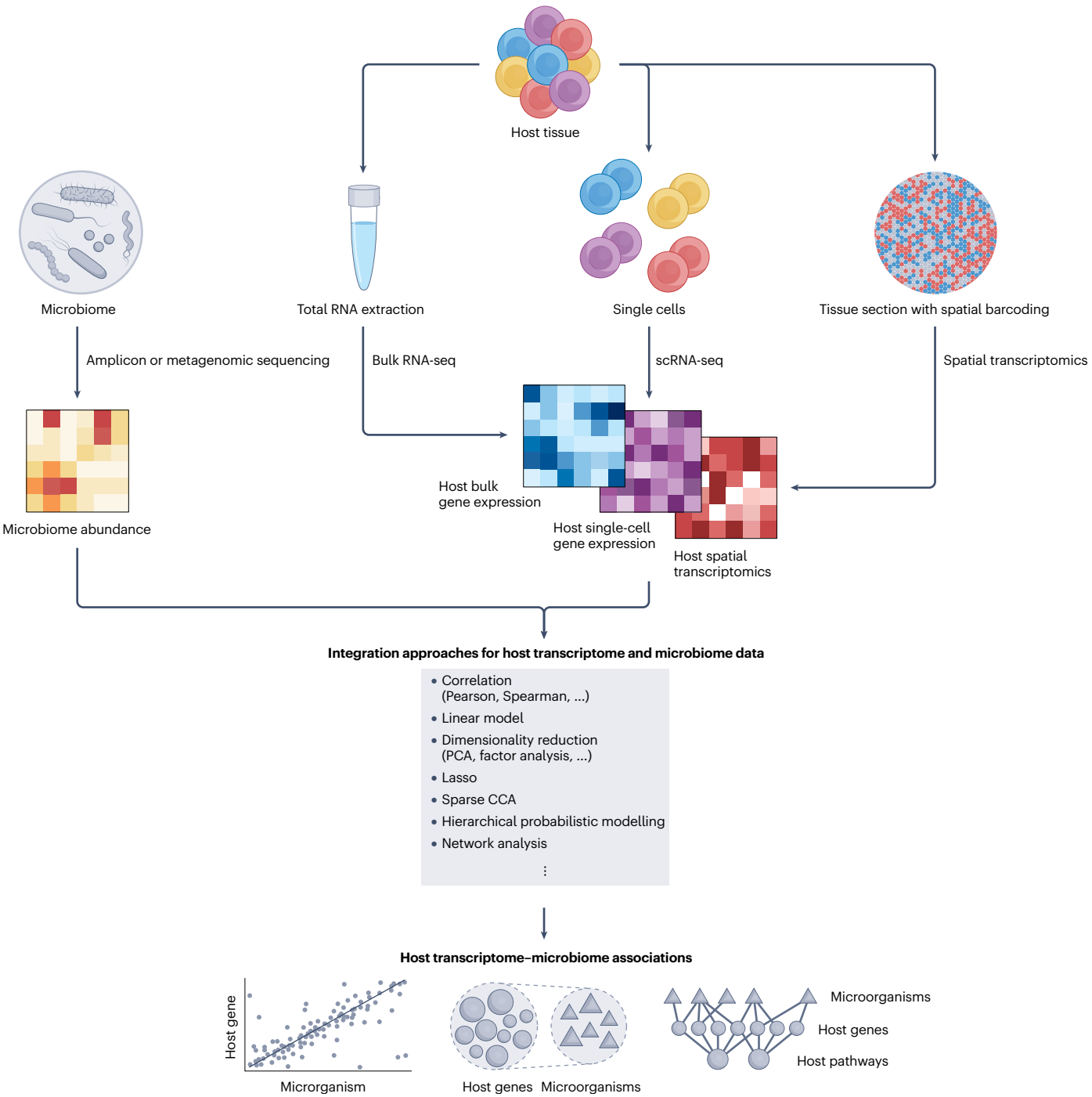
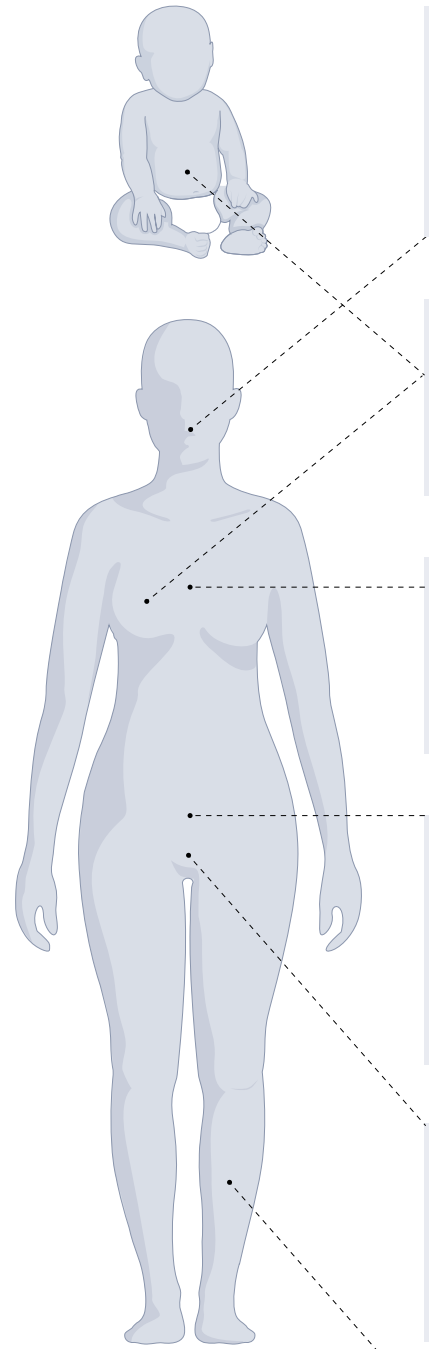


Fig. 4 | Conceptual framework for characterizing associations between host gene expression and microbiome. Host tissues are characterized using various techniques to profile host transcriptional activity at different levels. Total RNA is extracted from host tissues and subjected to bulk RNA sequencing (RNA-seq) to obtain host bulk gene expression data; host tissues are dissociated into individual cells and subjected to single-cell RNA sequencing (scRNA-seq) to generate host single-cell gene expression profiles; or spatial barcoding is applied to host tissue sections, followed by spatial transcriptomics, to generate host spatial transcriptomic data. On the microbiome side, 16S rRNA gene amplicon or metagenomic sequencing is performed to generate microbiome abundance data. Integrative approaches are used to jointly analyse the microbiome and host transcriptional datasets to identify associations between host gene expression and the microbiome. CCA, canonical correlation analysis; PCA, principal component analysis.



Oral cavity

Oral squamous cell carcinoma

Fusobacterium nucleatum | *Treponema* spp. | *Catonella morbi*

CXCL10, CXCL11, MMP9, MMP3 | *COLGALT2* | *PRKN* | *CBX7*

Interferon response, JAK–STAT signalling pathway | Tumour suppressor

Breast milk and infant gut

Bifidobacterium spp. | Amino acid degradation | *Escherichia* spp. | *Bifidobacterium infantis*

SOCS3

Fatty acid metabolism | Lysosome | JAK–STAT pathway

Lung and airway

Chronic obstructive pulmunary disease

Haemophilus | *Moraxella* | Tyrosine degradation | L-Glutamate production | Glycerophospholipid metabolism

IRAK1 pathway, T cell differentiation | Interferon signalling | Neutrophil chemotaxis | Eosinophil chemotaxis, CCL3 signalling

Colon and rectum

Colorectal cancer | **Irritable bowel syndrome** | **Inflammatory bowel disease**

Bacteroidales | *Streptococcus* | *Odoribacter* | *Bacteroides massiliensis* | *Clostridium sensu stricto 1* | *Streptococcus*

CCR2, FPR1 | *RIPK4* | *PLA2G4A* | *CXCL6, CCL20*

IL-10 signalling | RAC1 pathway | Chemokine

Cervicovaginal region

Atopobium | *Gardnerella* | *Sneathia* | *Prevotella* | *Lactobacillus* spp. | *Pepto-streptococcus* | *Mycoplasma*

Arachidonic and linoleic acid metabolism pathways | IL-1β regulation, stress response and keratinocyte differentiation | Innate immune response, antigen processing and presentation

Skin

Atopic dermatitis | **Psoriasis**

Staphylococcus aureus | *Corynebacterium*

DEFB4, S100A9 | *KYNU, TDO2*

Skin barrier, antimicrobial factors | Tryptophan metabolism | Interferon signalling, T_H1/T_H2 cell activation and cell cycle

Microbe or microbial pathway

Host-expressed gene

Host pathway or function

Infant gut microorganism or microbial pathway

Shared gene or microorganism across diseases

Fig. 5 | Overview of associations between the microbiome and host gene expression and pathways in humans. Known associations between gut microorganisms and microbial pathways with host gene expression and pathways across various body sites — including the oral cavity, lungs and airways, skin, gut, cervicovaginal region and breast milk — and in different health conditions. These associations vary by body site and disease state, suggesting that niche-specific microorganisms may both influence and respond to host gene expression through diverse biological pathways. The section 'Breast milk and infant gut' refers to associations between gene expression in milk and microbial abundance in the infant gut. An overview of the body sites and diseases studied in the studies whose findings were used to generate this network figure[85–90,92–94,151] is provided in Supplementary Table 3. T_H1 cell, T helper 1 cell.

as *CXCL6* and *CCL20*, were negatively correlated with the abundance of *Streptococcus*[79]. Similarly, the expression levels of *DUOX2* (which produces reactive oxygen species) and its maturation factor *DUOXA2* were negatively associated with the abundance of *Ruminococcaceae UCG 005* (ref. 79). A study that used a Mendelian randomization approach for causal inference identified oxidative stress-related genes, including *MUC1* and *PRKAB1*, as putative candidates for Crohn's disease[84]. The expression of *MUC1* was associated with *Bacillus aciditolerans*, a species capable of degrading myo-inositol, a process that contributes to SCFA production and may modulate mucosal immunity and inflammation in Crohn's disease[84]. By contrast, *PRKAB1* expression was associated with the opportunistic pathogen *Escherichia coli*[84]. Furthermore, studies on host transcript–microbiome associations in colorectal cancer have identified links between pathogenic mucosal bacteria and the expression of host genes involved in gastrointestinal inflammation and tumorigenesis[77,80].

Most studies examining host transcript–microbiome associations analyse interactions between a subset of host genes and gut microorganisms (for example, differentially expressed genes and/or differentially abundant microorganisms), and typically focus on a single disease at a time. This targeted approach can limit the ability to characterize association patterns across multiple diseases. To address these limitations, a study developed a machine learning-based multi-omic integration framework to characterize associations between the gut microbiome and host transcriptome across gastrointestinal disorders, including IBD, colorectal cancer and irritable bowel syndrome[85]. This approach revealed several disease-specific and shared associations. For example, the study found that a common set of host genes and pathways that regulate energy metabolism and intestinal mucosal repair (for example, the RAC1 pathway) are associated with disease-specific gut microorganisms in colorectal cancer (*Streptococcus*), IBD (*Clostridium sensu stricto 1*) and irritable bowel syndrome (*Odoribacter*). This finding suggests that distinct gut microorganisms can sometimes modulate commonly dysregulated host genes and pathways across different gut pathologies.

In addition to the gut, host transcript–microbiome associations have been characterized in other body sites, including the oral cavity, lungs/airways, skin, cervicovaginal region and breast milk[86–92] (Fig. 5). A few recent studies have explored the associations between the oral microbiota and host gene expression in individuals with oral squamous cell carcinoma[93,94]. One study found that tumour-enriched bacteria, such as *Catonella morbi* and *Treponema medium*, exhibit significant associations with the expression of host genes involved in oncogenic pathways[94]. For example, *C. morbi* is negatively associated with the tumour suppressor genes *PRKN* and *CBX7*, whereas *T. medium* shows a negative association with *COLGALT2*, which is implicated in cancer suppression. Another study investigated interactions between intratumoural bacteria and host transcription at spatial and single-cell resolution in oral squamous cell carcinoma[93]. It found that *Fusobacterium* and *Treponema* spp. are associated with significant upregulation of genes enriched for the interferon response and JAK–STAT signalling pathways, including increased expression of chemokines such as *CXCL10* and *CXCL11* and metalloproteinases such as *MMP9* and *MMP3*. Studies exploring airway host–microbiome interactions in chronic obstructive pulmonary disease found that specific microorganisms and microbial pathways are associated with host immune response and inflammatory pathways[88,89]. One study found *Haemophilus* to be linked to host interleukin-1 receptor-associated kinase 1 (IRAK1) and T cell differentiation pathways, whereas *Moraxella* was predominantly associated with interferon signalling pathways during chronic obstructive pulmonary disease exacerbations[88]. Another study identified tyrosine degradation and glycerophospholipid metabolism as key microbial pathways linked to distinct chronic obstructive pulmonary disease inflammatory phenotypes, including neutrophilic and eosinophilic inflammation, respectively[89]. When investigating host–microorganism interplay in the skin, it was observed that, in atopic dermatitis, colonization by *Staphylococcus aureus* triggers the upregulation of antimicrobial genes, such as *DEFB4* and *S100A9*, and tryptophan metabolism genes, including *KYNU* and *TDO2*, leading to inflammation and skin barrier disruption[86]. Published[87] and preprint[92] studies examining host gene expression–microbiome interactions in the cervicovaginal region have identified bacteria linked to bacterial vaginosis (including *Gardnerella*, *Atopobium*, *Sneathia* and certain *Prevotella* spp.) as being associated with host arachidonic and linoleic acid metabolism pathways. These microorganisms are also positively correlated with the expression of host genes enriched for stress responses and IL-1β production[87]. By contrast, most *Lactobacillus* spp. exhibit a protective role in the cervicovaginal region, showing negative associations with stress response and cytokine production, thus contributing to epithelial barrier stability and lower inflammation[87,92]. In a recent study, data from exclusively breastfeeding mother–infant pairs was used to examine how maternal genetics and milk gene expression influence the infant gut microbiome[90]. It was observed that the expression of genes involved in fatty acid metabolism in milk is positively correlated with the abundance of *Bifidobacterium* spp. in the infant gut at 1 month postpartum. By contrast, expression levels of lysosome genes in milk are negatively associated with microbial amino acid degradation pathways in the infant gut at 6 months postpartum. Furthermore, the abundance and growth of *Bifidobacterium infantis* in the infant gut are negatively correlated with the expression in milk of genes in the JAK–STAT pathway and are also linked to milk IL-6 and glucose levels. Together, these observations suggest that milk gene expression shapes the infant gut microbiome[90].

Overall, the findings of these studies show that the interplay between host gene expression and the microbiome varies by body site and disease state, and suggest that niche-specific microbiota could potentially influence, and be influenced by, host gene expression via various biological pathways.

Mechanisms of host transcript–microbiome cross-talk

Most studies on host transcript–microbiome association discussed so far are based on correlational analysis, where it is difficult to infer the

causal role and directionality of the associations. Given the challenges associated with disentangling causal mechanisms in human studies, researchers have used model organisms (such as mice and zebrafish) and in vitro human cell culture experiments to investigate the mechanisms underlying host gene expression–microbiome cross-talk[95–104].

One study found that gut microbiota in zebrafish and mice can suppress the activity of the transcription factor HNF4α, which is crucial for activating genes involved in lipid metabolism and inflammatory responses in the intestinal epithelium[98]. The conservation of this effect across species suggests that microbiome-mediated suppression of HNF4α may disrupt gene expression related to intestinal homeostasis, potentially contributing to the development of IBD in humans[98]. Another study in mice revealed that skin microbiota modulate immune responses and epidermal differentiation via regulation of key transcription factors such as KLF4, AP-1 and SP-1, which are critical for epidermal barrier function, and are implicated in conditions such as atopic dermatitis and psoriasis in humans[105].

Microbiota can also modulate epigenetic modifications, including DNA methylation, chromatin accessibility and histone acetylation, which can substantially alter host gene expression programmes[97,99,106–110]. Commensal bacteria have been shown to induce TET2/3-mediated DNA demethylation of regulatory elements in mouse intestinal epithelial cells, leading to alterations in gene expression programmes associated with colitis and colon cancer[106]. Another cell line-based approach used human intestinal epithelial cells treated with live microbial communities to assess host gene regulatory responses[99]. This study showed that the gut microbiota both broadly and specifically alters host gene expression, affecting more than 5,000 genes, including many linked to complex traits such as obesity, colorectal cancer and lipid metabolism, and identified specific microorganisms (such as *Collinsella*) that modulate host gene expression through changes in chromatin accessibility[99]. This approach was used to measure the effect of live microbial communities from different primate species on host gene regulation, identifying conserved transcriptional programmes as well as species-specific responses of host pathways[111]. A study investigating the influence of an obesogenic diet on gut microbiome and host gene expression in a mouse model found that the obesity-associated gut microbiome reprogrammes the intestinal epigenome by altering histone modifications at enhancer regions, leading to changes in gene expression linked to metabolic and cancer pathways[112].

These studies highlight diverse mechanisms through which microbiota can influence host gene expression, from modulating transcription factor activity to inducing epigenetic changes, offering insights into how these interactions can drive both homeostasis and disease across various tissues. Although host gene expression and the microbiome potentially engage in a bidirectional interaction, mechanistic evidence demonstrating that host gene expression can shape the microbiome remains limited, highlighting the need for more targeted experimental studies to establish causal mechanisms in this direction.

Emerging approaches for host–microbiome analysis

Although studies in humans and model organisms have revealed key insights into host gene expression–microbiome cross-talk, there remain many areas that need to be explored to obtain a more comprehensive understanding of host–microbiome interactions. Most existing studies associate bacterial abundances with host gene expression to identify specific taxonomic shifts, but this approach does not fully capture the functional potential of the microbiome. Although 16S rRNA gene amplicon and shotgun metagenomic data provide valuable insights, integrating microbial metatranscriptomic, metabolomic and metaproteomic data with host transcriptomic data could provide deeper insight into how microbial functions influence host gene expression. A few studies have taken important steps towards integrating multi-omic layers across the host and microbiome[79,81,89], but such efforts remain rare owing to the experimental and computational challenges of obtaining and aligning complex, heterogeneous datasets.

Microorganisms affect host gene regulation in a cell type-specific manner, and studies have used new techniques, such as scRNA-seq and spatial transcriptomics approaches, to identify the spatial, cellular and molecular interactions between microorganisms and host cells in mice, zebrafish and the human tumour microenvironment[93,113–116]. One study used 10× Visium spatial transcriptomics, GeoMx digital spatial profiling and the INVADEseq scRNA-seq method to map the spatial distribution, cellular interactions and transcriptional effects of intratumoural microbiota[93]. The study discovered that bacteria localize within specific intratumoural microniches characterized by the upregulation of immunosuppressive pathways, and that specific microorganisms, including *Fusobacterium* and *Treponema*, are predominantly associated with epithelial and macrophage cell types, driving transcriptional changes linked to metastasis and inflammation. Another study introduced spatial host–microbiome sequencing, a novel technique that combines spatial transcriptomics and 16S rRNA gene amplicon sequencing to simultaneously profile host gene expression and microbial composition in tissues with spatial resolution[114]. Using spatial host–microbiome sequencing, the authors identified distinct spatial niches in the mouse gut, where bacterial genera such as *Pseudobutyrivibrio* and *Oscillibacter* influenced expression of host genes such as *Muc2* and *Ceacam20*, which are involved in mechanisms critical for maintaining gut barrier integrity and immune signalling.

In the future, scRNA-seq of both the microbiome and host tissue could help reveal how the microbiome interacts with different cell types and influences cell-specific host gene expression. In addition, organoid systems co-cultured with microorganisms can be used as near-physiological in vitro models to study host–microorganism interactions, enabling researchers to conduct high-resolution studies of epithelial–microorganism dynamics[117,118]. Finally, novel computational models will be needed to integrate complex, high-dimensional and heterogeneous microbiome and host genomics datasets to characterize complex associations between the gut microbiome and host gene expression.

Challenges and future directions

Although initial microbiome GWAS and heritability studies have successfully identified associations at loci such as *LCT* and *ABO*, the field faces both biological and technical barriers to establishing additional robust and replicated genetic associations. These include limited power due to relatively small sample sizes, confounding factors such as diet, limited genetic ancestry diversity and a lack of standardized approaches for sample processing, sequencing and data analysis. Furthermore, the scalability of host–microbiome genomics analysis remains a challenge. Despite these limitations, microbiome heritability studies and GWAS have enriched our understanding of host–microbiome interactions, and novel technologies and analytical approaches — such as the integration of microbiome structural variation, host transcriptomics, single-cell data, functional genomics techniques, novel in vitro and ex vivo experimental models, and deeper longitudinal sampling — are being used to address these gaps and advance the field.

Limited sample size

The limited sample size of current microbiome GWAS and heritability studies represents one of the most pressing limitations in the field. The number of individuals included in microbiome GWAS has increased substantially in the past decade, from an average of 208 individuals in 2015 to 9,022 in 2022 (Table 1 and Fig. 1b). However, a detailed power analysis revealed for common taxa (prevalence of 50–90%) that a sample size between 25,000 and 50,000 is needed to detect associations with an effect of 0.4% variance explained, which represents approximately half of the effect of variants in the *LCT* locus on Bifidobacteria[119]. Larger sample sizes would be required to detect smaller effect sizes and/or less prevalent taxa: for example, an effect of 0.1% variance explained would require between 50,000 and 100,000 samples, and the same effect (0.1%) would require more than 100,000 samples to be detected for taxa with low prevalence (10–50%)[119]. Such large cohorts are challenging to assemble for single research groups, even considering the relatively low cost of 16S rRNA gene amplicon sequencing. Meta-analyses that combine publicly available cohorts could potentially reach the lower boundary of an estimated 30,000 individuals if data and metadata are harmonized, highlighting the importance of making both data and metadata of microbiome heritability and GWAS available to the research community. To date, microbiome GWAS have only explored interactions between bacteria and the host genome, but these considerations also apply to microbiome GWAS that focus on other components of the microbiome, such as fungi, archaea and viruses.

Multiple test correction and statistical analysis

The complex nature of microbiome data complicates the statistical analysis and reproducibility of microbiome GWAS and heritability studies. Unlike traditional GWAS, microbiome data include many possible readouts, including beta diversity, presence of specific taxa, relative abundances, metabolic pathways and microbial genetic variants. As each host genetic variant must be tested against each microbial trait, the number of hypotheses being tested is substantially larger than for standard GWAS, ranging from dozens or hundreds when examining bacterial genera to millions when studying microbial genetic variants. This extensive multiple testing reduces statistical power, leaving few significant associations after multiple test correction, ultimately affecting the reproducibility of results. Several statistical methods have been proposed to deal with the complexity of gut microbiome data, including parametric and non-parametric models, and zero-inflated models[120,121]. In addition, approaches that analyse the microbiome globally (that is, consider all taxa jointly) can help reduce the multiple testing burden and provide information on the genetic architecture of subtle combined effects[122]. However, these approaches provide limited insight into specific microbial traits that could be targets for clinical applications. Another strategy is to select contextually relevant microbial traits – for example, studying specific bacteria known to be important in IBD[123]. Methods based on machine-learning approaches have been especially promising[124]; for example, Lasso linear regression has been used for identifying associations between host genetic variation and microbiome composition[125]. Similar approaches have also been successful in identifying associations between the microbiome and multi-omic data[81,85,90,126]. Moving forward, the field could benefit from the development and application of techniques borrowed from linear algebra, graph theory and machine learning to help manage the complex nature of microbiome data while maintaining statistical power to detect meaningful associations.

Technical factors and other confounders

The choice of sequencing and analysis methods can also hinder the reproducibility of microbiome GWAS and heritability analyses[122]. For example, 16S rRNA gene amplicon sequencing provides lower sensitivity and taxonomic resolution than shotgun metagenomic sequencing[127], which with sufficient sequencing depth can provide reliable strain-level taxonomic classification, even for taxa present at low abundances[128]. Shotgun metagenomic sequencing can also provide information on bacterial SNPs that are potentially relevant for cancer susceptibility, inflammation and host DNA damage[129,130]. In addition, different taxonomic profiling tools might provide different taxonomic classification[131]. This variation can result from differences in detection and classification methods or the reference databases used. In 16S rRNA gene amplicon sequencing studies, the choice of the targeted hypervariable region, DNA lysis and extraction methods, and 16S primer selection can also affect observed microbial composition[132–134]. In addition, microbial composition can be affected by factors such as the sampling time[135] and microbial load[136]. The choice of profiling tool can also affect the taxonomic classification of the non-bacterial components of the microbiome (viruses, fungi and archaea).

Among the many confounding factors in microbiome GWAS, diet is of particular interest. A prominent example of a gene-by-diet association with the microbiome is the *LCT* gene (see 'The *LCT* locus, diet and gut *Bifidobacterium*' section). As diet has an important role in shaping and modulating gut microbiome composition and function[12], accurate dietary information is critical for analysing and interpreting microbiome GWAS and heritability results. In addition, dietary preference itself is a heritable trait[137,138], further complicating the interpretability of microbiome GWAS and emphasizing the need to account for dietary composition in these analyses. The dietary information associated with current microbiome GWAS is based on the use of high-level questionnaires, which might be affected by personal biases and usually do not provide fine-grained information on the intake of key molecules that can shape gut microbiome composition and functionality, such as lactose, butyrate and inulins[139,140]. The ongoing development of methods to infer precise and quantitative dietary metadata from stool shotgun metagenomic data may help overcome this limitation[141–143].

Future microbiome GWAS and heritability meta-analyses aiming to identify associations with microbial taxa should ideally reanalyse the raw microbiome data consistently to address differences introduced by various taxonomic profiling tools. In addition, researchers should favour shotgun metagenomics over 16S rRNA gene amplicon-based sequencing, as it provides richer data on microbial pathways, gene content and genetic variation, improving the usability of the data in future meta-analyses. Detailed metadata such as the sampling time, Bristol stool scale, host diet and medication history will allow more biologically meaningful stratification of the data and, potentially, more robust associations. Finally, functional validation studies are needed to advance the field from correlation to causation studies.

Low population heterogeneity

An additional important limitation is the low diversity in the genetic ancestry of current microbiome GWAS. To date, 51% of the participants included in microbiome GWAS were recruited either in Germany or in The Netherlands (Fig. 1c), whereas only a few studies have included individuals from non-westernized countries[26,43,45], mimicking a less pronounced but similar trend found in microbiome studies[144]. Although individuals from 13 countries have been included in microbiome GWAS so far, these countries represent less than 30% of the world population

Glossary

16S rRNA gene amplicon sequencing
A targeted sequencing approach that amplifies and sequences regions of the 16S ribosomal RNA gene, which is present in the genomes of all bacteria. This method enables taxonomic profiling of microbial communities without sequencing entire genomes, allowing for a cost-effective way to characterize the taxonomic composition of microbiome samples.

Alpha diversity
A measure of microbial community complexity within a single sample, quantifying the number of different species and their relative abundances. Alpha-diversity metrics can be used to assess how factors such as genetics, diet and disease affect the ecological structure of microbial communities.

Beta diversity
A measure of the difference in microbial community composition between two or more samples. Beta-diversity metrics quantify how samples differ from each other in terms of which microorganisms are present and their relative abundances, allowing the quantification of microbiome variation across conditions, environments and host factors.

Heritability
The proportion of phenotypic variation in a population that can be attributed to genetic differences; heritability is expressed as a value between 0 and 1, where 0 indicates that none of the observed variation is due to genetic factors and 1 indicates that all variation is genetic.

Quantitative trait locus
A specific region of DNA that contributes to variation in a measurable characteristic or quantitative trait, such as height, blood pressure or disease susceptibility. In microbiome research, quantitative trait locus mapping identifies host genetic regions that influence microbial community features, revealing how host genetic variation shapes microbiome composition.

Shotgun metagenomic sequencing
A comprehensive sequencing technique that sequences all DNA found in a microbiome sample, which enables the capture of genomic information from bacteria, archaea, viruses and fungi found in a microbial community. Unlike 16S rRNA amplicon sequencing, shotgun metagenomic sequencing provides information on not only the taxonomic composition but also the functional potential, gene composition and genetic variation within microbial communities.

(Fig. 1c). For example, although Germany is home to less than 2% of the world population, it has contributed 28.5% of the samples collected so far in microbiome GWAS (Fig. 1c). Limited population heterogeneity can introduce false positive or false negative associations[119] when considering the relationship between population heterogeneity and dietary and environmental factors, both known to significantly shape the gut microbiome composition. Moreover, microbiome GWAS results from one population may not translate to another, as is the case for GWAS of other traits[145]. Populations from non-urbanized contexts have also been under-represented in current studies. Thus, the lack of global representation in microbiome heritability studies and GWAS limits the applicability and global relevance of results from these studies.

Lessons from GWAS

With fewer than 20 studies conducted in the past decade, the field of microbiome GWAS is still in its infancy. Many of the challenges and limitations described above were also encountered in the early stages of traditional GWAS. Similarly to current microbiome GWAS, early GWAS of non-microbiome traits were hindered by small sample sizes, reproducibility issues and limited statistical power. The lessons learned from those studies can, at least partially, be used to improve microbiome GWAS and develop standards for future studies[146]. For example, the adoption of more stringent statistical requirements for testing and reporting associations, common data analysis workflows, consistent guidelines for reporting metadata and reproducibility requirements will substantially improve the status quo.

In addition, when using data from cohorts that include individuals from different ancestries, it is important to account for population substructure in the analysis. Such population substructure in microbiome GWAS can lead to spurious associations when genetic differences between populations are confounded with differences in the composition of the microbiome. This issue can be partially addressed through methods such as principal component analysis to control for ancestry[147] and mixed linear models incorporating genetic relatedness matrices[148].

Lastly, traditional GWAS clearly showed that the genetic architecture of common traits is more polygenic than initially thought, with many traits associated with hundreds or thousands of genetic variants with small effect sizes[149]. Given the relatively low heritability estimates and small GWAS effect sizes, it is reasonable to assume that a similar genetic architecture underlies microbiome traits, which further highlights the need for larger sample sizes and more robust and consistent data analysis approaches.

Conclusions

The past decade has seen remarkable progress in our understanding of the relationship between host genetics and the microbiome. Numerous studies have quantified the heritability of the microbiome, identified host genetic variants associated with the microbiome composition and shed light on the interaction between the microbiome and host gene regulation. The field is also making progress by developing and applying novel genomic and computational approaches that allow interactions between host genes and the microbiota to be profiled across study systems and health conditions. Although current research focuses primarily on bacteria, expanding microbiome heritability and GWAS to include non-bacterial components such as fungi and viruses represents an important future direction. Several challenges reviewed here, including low replicability across studies, will likely require the concerted effort of consortia that will enable leveraging larger study cohorts, collection of detailed metadata and development of standards for data generation and processing. Lastly, functional validation of host gene–microorganism associations, as well as experimental assessment of their impact on host health, are critical next steps. Understanding the interactions between human genetics and the microbiome will be crucial for developing targeted therapeutic approaches and advancing precision health interventions, highlighting the immense potential of this rapidly evolving field.

Published online: 04 June 2025

References

1. Valdes, A. M., Walter, J., Segal, E. & Spector, T. D. Role of the gut microbiota in nutrition and health. *BMJ* **361**, k2179 (2018).
2. Zheng, D., Liwinski, T. & Elinav, E. Interaction between microbiota and immunity in health and disease. *Cell Res.* **30**, 492–506 (2020).
3. Tarracchini, C. et al. Exploring the vitamin biosynthesis landscape of the human gut microbiota. *mSystems* **9**, e0092924 (2024).
4. de Vos, W. M., Tilg, H., Van Hul, M. & Cani, P. D. Gut microbiome and health: mechanistic insights. *Gut* **71**, 1020–1032 (2022).
5. Geng, J., Ni, Q., Sun, W., Li, L. & Feng, X. The links between gut microbiota and obesity and obesity related diseases. *Biomed. Pharmacother.* **147**, 112678 (2022).

6. Van Hul, M. & Cani, P. D. The gut microbiota in obesity and weight management: microbes as friends or foe? *Nat. Rev. Endocrinol.* **19**, 258–271 (2023).
7. Kandalai, S., Li, H., Zhang, N., Peng, H. & Zheng, Q. The human microbiome and cancer: a diagnostic and therapeutic perspective. *Cancer Biol. Ther.* **24**, 2240084 (2023).
8. Vivarelli, S. et al. Gut microbiota and cancer: from pathogenesis to therapy. *Cancers* **11**, 38 (2019).
9. Vich Vila, A. et al. Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. *Nat. Commun.* **11**, 362 (2020).
10. Amato, K. R. et al. The human gut microbiome and health inequities. *Proc. Natl Acad. Sci. USA* **118**, e2017947118 (2021).
11. Leeming, E. R., Johnson, A. J., Spector, T. D. & Le Roy, C. I. Effect of diet on the gut microbiota: rethinking intervention duration. *Nutrients* **11**, 2862 (2019).
12. Schmidt, T. S. B., Raes, J. & Bork, P. The human gut microbiome: from association to modulation. *Cell* **172**, 1198–1215 (2018).
13. Turnbaugh, P. J. et al. A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
14. Benson, A. K. et al. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc. Natl Acad. Sci. USA* **107**, 18933–18938 (2010).
15. Goodrich, J. K. et al. Human genetics shape the gut microbiome. *Cell* **159**, 789–799 (2014).
This important early study uses data from twins to estimate gut microbiome heritability.
16. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era—concepts and misconceptions. *Nat. Rev. Genet.* **9**, 255–266 (2008).
17. Goodrich, J. K. et al. Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* **19**, 731–743 (2016).
18. Davenport, E. R. et al. Genome-wide association studies of the human gut microbiota. *PLoS One* **10**, e0140301 (2015).
19. O'Callaghan, A. & van Sinderen, D. Bifidobacteria and their role as members of the human gut microbiota. *Front. Microbiol.* **7**, 925 (2016).
20. Samuel, B. S. et al. Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc. Natl Acad. Sci. USA* **104**, 10643–10648 (2007).
21. Ignatyeva, O. et al. *Christensenella minuta*, a new candidate next-generation probiotic: current evidence and future trajectories. *Front. Microbiol.* **14**, 1241259 (2023).
22. Rothschild D et al. Environment dominates over host genetics in shaping human gut microbiota. *Yearb. Pediatr. Endocrinol.* https://doi.org/10.1530/ey.15.14.5 (2018).
This key study argues that microbiome heritability is low and that environment is stronger than genetics in shaping the gut microbiome.
23. Hughes, D. A. et al. Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. *Nat. Microbiol.* **5**, 1079–1087 (2020).
24. Wang, J. et al. Meta-analysis of human genome-microbiome association studies: the MiBioGen consortium initiative. *Microbiome* **6**, 101 (2018).
25. Kurilshikov, A. et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat. Genet.* **53**, 156–165 (2021).
26. Boulund, U. et al. Gut microbiome associations with host genotype vary across ethnicities and potentially influence cardiometabolic traits. *Cell. Host. Microbe* **30**, 1464–1480.e6 (2022).
27. Gacesa, R. et al. Environmental factors shaping the gut microbiome in a Dutch population. *Nature* **604**, 732–739 (2022).
28. Zhernakova, D. V. et al. Host genetic regulation of human gut microbial structural variation. *Nature* **625**, 813–821 (2024).
This study investigates the heritability of structural variation in the gut microbiome.
29. Beaumont, M. et al. Heritable components of the human fecal microbiome are associated with visceral fat. *Genome Biol.* **17**, 189 (2016).
30. Blekhman, R. et al. Host genetic variation impacts microbiome composition across human body sites. *Genome Biol.* **16**, 191 (2015).
This important early study demonstrates the feasibility of testing for genome-wide host genetic associations with gut microbiome measurements as quantitative traits.
31. Grieneisen, L. et al. Gut microbiome heritability is nearly universal but environmentally contingent. *Science* **373**, 181–186 (2021).
This article describes a large study of microbiome heritability in wild primates.
32. Zhernakova, A. et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* **352**, 565–569 (2016).
33. Hejazi, J., Amiri, R., Nozarian, S., Tavasolian, R. & Rahimlou, M. Genetic determinants of food preferences: a systematic review of observational studies. *BMC Nutr.* **10**, 24 (2024).
34. Vandeputte, D. et al. Temporal variability in quantitative human gut microbiome profiles and implications for clinical research. *Nat. Commun.* **12**, 6740 (2021).
35. Org, E. et al. Genetic and environmental control of host–gut microbiota interactions. *Genome Res.* **25**, 1558–1569 (2015).
36. Chen, C. et al. Contribution of host genetics to the variation of microbial composition of cecum lumen and feces in pigs. *Front. Microbiol.* **9**, 2626 (2018).
37. Doms, S. et al. Key features of the genetic architecture and evolution of host–microbe interactions revealed by high-resolution genetic mapping of the mucosa-associated gut microbiome in hybrid mice. *eLife* **11**, e75419 (2022).
38. McKnite, A. M. et al. Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. *PLoS One* **7**, e39191 (2012).
39. Leamy, L. J. et al. Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. *Genome Biol.* **15**, 552 (2014).
40. Turpin, W. et al. Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat. Genet.* **48**, 1413–1417 (2016).
41. Bonder, M. J. et al. The effect of host genetics on the gut microbiome. *Nat. Genet.* **48**, 1407–1412 (2016).
This paper describes an early study on the effects of host genetics on the gut microbiome.
42. Wang, J. et al. Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat. Genet.* **48**, 1396–1406 (2016).
This early microbiome GWAS finds associations with the gene encoding for the vitamin D receptor.
43. Ishida, S. et al. Genome-wide association studies and heritability analysis reveal the involvement of host genetics in the Japanese gut microbiota. *Commun. Biol.* **3**, 686 (2020).
44. Rühlemann, M. C. et al. Genome-wide association study in 8956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. *Nat. Genet.* **53**, 147–155 (2021).
This large GWAS identifies effects of the ABO blood group on the human gut microbiome.
45. Liu, X. et al. A genome-wide association study for gut metagenome in Chinese adults illuminates complex diseases. *Cell Discov.* **7**, 9 (2021).
46. Liu, X. et al. Mendelian randomization analyses support causal relationships between blood metabolites and the gut microbiome. *Nat. Genet.* **54**, 52–61 (2022).
47. Lopera-Maya, E. A. et al. Effect of host genetics on the gut microbiome in 7738 participants of the Dutch Microbiome Project. *Nat. Genet.* **54**, 143–151 (2022).
This large GWAS of the human gut microbiome validates the association of variants at the *LCT* locus and finds that the association is modulated by lactose intake.
48. Qin, Y. et al. Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. *Nat. Genet.* **54**, 134–142 (2022).
49. Kolde, R. et al. Host genetic variation and its microbiome interactions within the human microbiome project. *Genome Med.* **10**, 6 (2018).
50. Scepanovic, P. et al. A comprehensive assessment of demographic, environmental, and host genetic associations with gut microbiome diversity in healthy individuals. *Microbiome* **7**, 130 (2019).
51. Tomofuji, Y. et al. Analysis of gut microbiome, host genetics, and plasma metabolites reveals gut microbiome–host interactions in the Japanese population. *Cell Rep.* **42**, 113324 (2023).
52. Moitinho-Silva, L. et al. Host genetic factors related to innate immunity, environmental sensing and cellular functions are associated with human skin microbiota. *Nat. Commun.* **13**, 6204 (2022).
53. Liu, X. et al. A genome-wide association study reveals the relationship between human genetic variation and the nasal microbiome. *Commun. Biol.* **7**, 139 (2024).
54. Fang, Z. Y. et al. Networks of human milk microbiota are associated with host genomics, childhood asthma, and allergic sensitization. *Cell Host Microbe* **32**, 1838–1852.e5 (2024).
55. Enattah, N. S. et al. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30**, 233–237 (2002).
56. Rasinperä, H. et al. Transcriptional downregulation of the lactase (*LCT*) gene during childhood. *Gut* **54**, 1660–1661 (2005).
57. Garrido, D., Barile, D. & Mills, D. A. A molecular basis for bifidobacterial enrichment in the infant gastrointestinal tract. *Adv. Nutr.* **3**, 415S–421S (2012).
58. Itan, Y., Jones, B. L., Ingram, C. J. E., Swallow, D. M. & Thomas, M. G. A worldwide correlation of lactase persistence phenotype and genotypes. *BMC Evol. Biol.* **10**, 36 (2010).
59. Jajosky, R. P. et al. ABO blood group antigens and differential glycan expression: perspective on the evolution of common human enzyme deficiencies. *iScience* **26**, 105798 (2023).
60. Mäkivuokko, H. et al. Association between the ABO blood group and the human intestinal microbiota composition. *BMC Microbiol.* **12**, 94 (2012).
61. Rausch, P. et al. Multigenerational influences of the *Fut2* gene on the dynamics of the gut microbiota in mice. *Front. Microbiol.* **8**, 991 (2017).
62. Folseraas, T. et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. *J. Hepatol.* **57**, 366–375 (2012).
63. Rausch, P. et al. Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. *Proc. Natl Acad. Sci. USA* **108**, 19030–19035 (2011).
64. Wacklin, P. et al. Secretor genotype (FUT2 gene) is strongly associated with the composition of Bifidobacteria in the human intestine. *PLoS One* **6**, e20113 (2011).
65. Tong, M. et al. Reprograming of gut microbiome energy metabolism by the FUT2 Crohn's disease risk polymorphism. *ISME J.* **8**, 2193–2206 (2014).
66. Yang, H. et al. ABO genotype alters the gut microbiota by regulating GalNAc levels in pigs. *Nature* **606**, 358–367 (2022).
67. Ségurel, L. et al. The ABO blood group is a trans-species polymorphism in primates. *Proc. Natl Acad. Sci. USA* **109**, 18493–18498 (2012).
68. Burgess, S. et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open. Res.* **4**, 186 (2019).
69. Davey Smith, G. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
70. Sanderson, E. et al. Mendelian randomization. *Nat. Rev. Methods Prim.* **2**, 6 (2022).
71. Sanna, S. et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nat. Genet.* **51**, 600–605 (2019).
Together with Liu et al. (2022), this work exemplifies the use of Mendelian randomization to identify potential causal effects of the microbiome on human health.

72. Arora, T. & Tremaroli, V. Therapeutic potential of butyrate for treatment of type 2 diabetes. *Front. Endocrinol.* **12**, 761834 (2021).
73. Stender, S., Gellert-Kristensen, H. & Smith, G. D. Reclaiming Mendelian randomization from the deluge of papers and misleading findings. *Lipids Health Dis.* **23**, 286 (2024).
74. Wade, K. H. & Hall, L. J. Improving causality in microbiome research: can human genetic epidemiology help? *Wellcome Open. Res.* **4**, 199 (2019).
75. Morgan, X. C. et al. Associations between host gene expression, the mucosal microbiome, and clinical outcome in the pelvic pouch of patients with inflammatory bowel disease. *Genome Biol.* **16**, 67 (2015).
76. Häsler, R. et al. Uncoupling of mucosal gene regulation, mRNA splicing and adherent microbiota signatures in inflammatory bowel disease. *Gut* **66**, 2087–2097 (2017).
77. Omar Al-Hassi, H., Ng, O. & Brookes, M. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut* **67**, 395 (2018).
78. Bennet, S. M. P. et al. Altered intestinal antibacterial gene expression response profile in irritable bowel syndrome is linked to bacterial composition and immune activation: XXXX. *Neurogastroenterol. Motil.* **30**, e13468 (2018).
79. Lloyd-Price, J. et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).
Together with Morgan et al. (2015), this key paper describes associations between the gut microbiome and host gene regulation in IBD.
80. Dayama, G., Priya, S., Niccum, D. E., Khoruts, A. & Blekhman, R. Interactions between the gut microbiome and host gene regulation in cystic fibrosis. *Genome Med.* **12**, 12 (2020).
81. Mars, R. A. T. et al. Longitudinal multi-omics reveals subset-specific mechanisms underlying irritable bowel syndrome. *Cell* **182**, 1460–1473.e17 (2020).
82. Chang, H. et al. Unveiling the links between microbial alteration and host gene disarray in Crohn's disease via TAHMC. *Adv. Biol.* **8**, e2400064 (2024).
83. Mirsepasi-Lauridsen, H. C., Vallance, B. A., Krogfelt, K. A. & Petersen, A. M. *Escherichia coli* pathobionts associated with inflammatory bowel disease. *Clin. Microbiol. Rev.* **32**, e00060-18 (2019).
84. Xu, S. et al. Oxidative stress gene expression, DNA methylation, and gut microbiota interaction trigger Crohn's disease: a multi-omics Mendelian randomization study. *BMC Med.* **21**, 179 (2023).
85. Priya, S. et al. Identification of shared and disease-specific host gene–microbiome associations across human diseases using multi-omic integration. *Nat. Microbiol.* **7**, 780–795 (2022).
This important study elucidates patterns of host gene–microbiome interactions across diseases.
86. Fyhrquist, N. et al. Microbe–host interplay in atopic dermatitis and psoriasis. *Nat. Commun.* **10**, 4703 (2019).
87. Edfeldt, G. et al. Distinct cervical tissue-adherent and luminal microbiome communities correlate with mucosal host gene expression and protein levels in Kenyan sex workers. *Microbiome* **11**, 67 (2023).
88. Wang, Z. et al. Airway host–microbiome interactions in chronic obstructive pulmonary disease. *Respir. Res.* **20**, 113 (2019).
89. Yan, Z. et al. Multi-omics analyses of airway host–microbe interactions in chronic obstructive pulmonary disease identify potential therapeutic interventions. *Nat. Microbiol.* **7**, 1361–1375 (2022).
90. Johnson, K. E. et al. Human milk variation is shaped by maternal genetics and impacts the infant gut microbiome. *Cell Genom.* **4**, 100638 (2024).
91. Yang, Y. et al. Exploratory multi-omics analysis reveals host–microbe interactions associated with disease severity in psoriatic skin. *EBioMedicine* **105**, 105222 (2024).
92. Zhu, B. et al. Characteristics of vaginal microbes and classification of the vaginal microbiome. Preprint at *bioRxiv* https://doi.org/10.1101/2023.08.16.553525 (2023).
93. Galeano Niño, J. L. et al. Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. *Nature* **611**, 810–817 (2022).
94. Cai, L. et al. Integrative analysis reveals associations between oral microbiota dysbiosis and host genetic and epigenetic aberrations in oral cavity squamous cell carcinoma. *NPJ Biofilms Microbiomes* **10**, 39 (2024).
95. Camp, J. G. et al. Microbiota modulate transcription in the intestinal epithelium without remodeling the accessible chromatin landscape. *Genome Res.* **24**, 1504–1516 (2014).
96. Sommer, F., Nookaew, I., Sommer, N., Fogelstrand, P. & Bäckhed, F. Site-specific programming of the host epithelial transcriptome by the gut microbiota. *Genome Biol.* **16**, 62 (2015).
97. Pan, W.-H. et al. Exposure to the gut microbiota drives distinct methylome and transcriptome changes in intestinal epithelial cells during postnatal development. *Genome Med.* **10**, 27 (2018).
98. Davison, J. M. et al. Microbiota regulate intestinal epithelial gene expression by suppressing the transcription factor hepatocyte nuclear factor 4α. *Genome Res.* **27**, 1195–1206 (2017).
99. Richards, A. L. et al. Gut microbiota has a widespread and modifiable effect on host gene regulation. *mSystems* **4**, e00323–18 (2019).
Together with Davison et al. (2017), this key paper investigates the mechanisms of host gene–microbiome cross-talk.
100. Semenkovich, N. P. et al. Impact of the gut microbiota on enhancer accessibility in gut intraepithelial lymphocytes. *Proc. Natl Acad. Sci. USA* **113**, 14805–14810 (2016).
101. Dirksen, P. et al. CeMbio—the *Caenorhabditis* elegans microbiome resource. *G3* **10**, 3025–3039 (2020).
102. Yang, W. et al. The inducible response of the nematode *Caenorhabditis elegans* to members of its natural microbiota across development and adult life. *Front. Microbiol.* **10**, 1793 (2019).
103. Loomis, K. H. et al. A mixed community of skin microbiome representatives influences cutaneous processes more than individual members. *Microbiome* **9**, 22 (2021).
104. Brusilovsky, M. et al. Host–microbiota interactions in the esophagus during homeostasis and allergic inflammation. *Gastroenterology* **162**, 521–534.e8 (2022).
105. Meisel, J. S. et al. Commensal microbiota modulate gene expression in the skin. *Microbiome* **6**, 20 (2018).
106. Ansari, I. et al. The microbiota programs DNA methylation to control intestinal homeostasis and inflammation. *Nat. Microbiol.* **5**, 610–619 (2020).
107. Krautkramer, K. A. et al. Diet–microbiota interactions mediate global epigenetic programming in multiple host tissues. *Mol. Cell* **64**, 982–992 (2016).
108. Kuang, Z. et al. The intestinal microbiota programs diurnal rhythms in host metabolism through histone deacetylase 3. *Science* **365**, 1428–1434 (2019).
109. Ryan, F. J. et al. Colonic microbiota is associated with inflammation and host epigenomic alterations in inflammatory bowel disease. *Nat. Commun.* **11**, 1512 (2020).
110. Nshanian, M. et al. Short-chain fatty acid metabolites propionate and butyrate are unique epigenetic regulatory elements linking diet, metabolism and gene expression. *Nat. Metab.* **7**, 196–211 (2025).
111. Muehlbauer, A. L. et al. Interspecies variation in hominid gut microbiota controls host gene regulation. *Cell Rep.* **37**, 110057 (2021).
112. Qin, Y. et al. An obesity-associated gut microbiome reprograms the intestinal epigenome and leads to altered colonic gene expression. *Genome Biol.* **19**, 7 (2018).
113. Drayman, N., Patel, P., Vistain, L. & Tay, S. HSV-1 single-cell analysis reveals the activation of anti-viral and developmental programs in distinct sub-populations. *eLife* **8**, e46339 (2019).
114. Lötstedt, B., Stražar, M., Xavier, R., Regev, A. & Vickovic, S. Spatial host–microbiome sequencing reveals niches in the mouse gut. *Nat. Biotechnol.* **42**, 1394–1403 (2024).
Together with Galeano Niño et al. (2022), this notable paper describes spatial profiling of host–microbiome interactions.
115. Massaquoi, M. S. et al. Cell-type-specific responses to the microbiota across all tissues of the larval zebrafish. *Cell Rep.* **42**, 112095 (2023).
116. Willms, R. J., Jones, L. O., Hocking, J. C. & Foley, E. A cell atlas of microbe-responsive processes in the zebrafish intestine. *Cell Rep.* **38**, 110311 (2022).
117. Puschhof, J. et al. Intestinal organoid cocultures with microbes. *Nat. Protoc.* **16**, 4633–4649 (2021).
118. Williamson, I. A. et al. A high-throughput organoid microinjection platform to study gastrointestinal microbiota and luminal physiology. *Cell. Mol. Gastroenterol. Hepatol.* **6**, 301–319 (2018).
119. Sanna, S., Kurilshikov, A., van der Graaf, A., Fu, J. & Zhernakova, A. Challenges and future directions for studying effects of host genetics on the gut microbiome. *Nat. Genet.* **54**, 100–106 (2022).
120. Xu, L., Paterson, A. D., Turpin, W. & Xu, W. Assessment and selection of competing models for zero-inflated microbiome data. *PLoS One* **10**, e0129606 (2015).
121. Fisher, C. K. & Mehta, P. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS One* **9**, e102451 (2014).
122. Weissbrod, O., Rothschild, D., Barkan, E. & Segal, E. Host genetics and microbiome associations through the lens of genome wide association studies. *Curr. Opin. Microbiol.* **44**, 9–19 (2018).
123. Sham, P. C. & Purcell, S. M. Statistical power and significance testing in large-scale genetic studies. *Nat. Rev. Genet.* **15**, 335–346 (2014).
124. Chetty, A. & Blekhman, R. Multi-omic approaches for host–microbiome data integration. *Gut Microbes* **16**, 2297860 (2024).
125. Lynch, J. et al. HOMINID: a framework for identifying associations between host genetic variation and microbiome composition. *Gigascience* **6**, 1–7 (2017).
126. Mills, R. H. et al. Multi-omics analyses of the ulcerative colitis gut microbiome link *Bacteroides vulgatus* proteases with disease severity. *Nat. Microbiol.* **7**, 262–276 (2022).
127. Poretsky, R., Rodriguez-R, L. M., Luo, C., Tsementzi, D. & Konstantinidis, K. T. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One* **9**, e93827 (2014).
128. Durazzi, F. et al. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Sci. Rep.* **11**, 3030 (2021).
129. Sharafutdinov, I. et al. A single-nucleotide polymorphism in *Helicobacter pylori* promotes gastric cancer development. *Cell Host Microbe* **31**, 1345–1358.e6 (2023).
130. Berthenet, E. et al. A GWAS on *Helicobacter pylori* strains points to genetic variants associated with gastric cancer risk. *BMC Biol.* **16**, 84 (2018).
131. Sczyrba, A. et al. Critical assessment of metagenome interpretation—a benchmark of metagenomics software. *Nat. Methods* **14**, 1063–1071 (2017).
132. Meisel, J. S. et al. Skin microbiome surveys are strongly influenced by experimental design. *J. Invest. Dermatol.* **136**, 947–956 (2016).
133. Deissová, T. et al. 16S rRNA gene primer choice impacts off-target amplification in human gastrointestinal tract biopsies and microbiome profiling. *Sci. Rep.* **13**, 12577 (2023).
134. Elie, C. et al. Comparison of DNA extraction methods for 16S rRNA gene sequencing in the analysis of the human gut microbiome. *Sci. Rep.* **13**, 10279 (2023).
135. Allaband, C. et al. Time of sample collection is critical for the replicability of microbiome analyses. *Nat. Metab.* **6**, 1282–1293 (2024).
136. Nishijima, S. et al. Fecal microbial load is a major determinant of gut microbiome variation and a confounder for disease associations. *Cell* **188**, 222–236.e15 (2025).
137. Pallister, T. et al. Food preference patterns in a UK twin cohort. *Twin Res. Hum. Genet.* **18**, 793–805 (2015).
138. Fildes, A. et al. Nature and nurture in children's food preferences. *Am. J. Clin. Nutr.* **99**, 911–917 (2014).

139. Goodrich, J. K., Davenport, E. R., Waters, J. L., Clark, A. G. & Ley, R. E. Cross-species comparisons of host genetic associations with the microbiome. *Science* **352**, 532–535 (2016).
140. Corrêa, R. O. et al. Inulin diet uncovers complex diet–microbiota–immune cell interactions remodeling the gut epithelium. *Microbiome* **11**, 90 (2023).
141. Shinn, L. M. et al. Fecal metagenomics to identify biomarkers of food intake in healthy adults: findings from randomized, controlled, nutrition trials. *J. Nutr.* **154**, 271–283 (2024).
142. Diener, C. et al. Metagenomic estimation of dietary intake from human stool. *Nat. Metab.* **7**, 617–630 (2025).
143. Zuppinger, C. et al. Performance of the digital dietary assessment tool MyFoodRepo. *Nutrients* **14**, 635 (2022).
144. Abdill, R. J., Adamowicz, E. M. & Blekhman, R. Public human microbiome data are dominated by highly developed countries. *PLoS Biol.* **20**, e3001536 (2022).
145. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The missing diversity in human genetic studies. *Cell* **177**, 1080 (2019).
146. McCarthy, M. I. et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9**, 356–369 (2008).
147. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
148. Yang, J. et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519–525 (2011).
149. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
150. Xu, F. et al. The interplay between host genetics and the gut microbiome reveals common and distinct microbiome features for complex human diseases. *Microbiome* **8**, 145 (2020).
151. Lloyd-Price, J. et al. Strains, functions and dynamics in the expanded human microbiome project. *Nature* **550**, 61–66 (2017).
152. Weiss, S. et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27 (2017).
153. Luca, F., Kupfer, S. S., Knights, D., Khoruts, A. & Blekhman, R. Functional genomics of host–microbiome interactions in humans. *Trends Genet.* **34**, 30–40 (2018).
154. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220–230 (2012).
155. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome datasets are compositional: and this is not optional. *Front. Microbiol.* **8**, 2224 (2017).
156. Franzosa, E. A. et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* **15**, 962–968 (2018).

Acknowledgements
The authors thank T. Vilgalys, R. Abdill and members of the Blekhman laboratory for their valuable feedback. R.B. acknowledges funding from National Institutes of Health (NIH) grants R35-GM128716 and R01-HD109830. K.J. acknowledges funding from NIH grants K99-HD113834 and R01-HD109830. S.P. acknowledges funding from the Duchossois Family Institute Fellowship (Chicago Fellows programme) at the University of Chicago.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no conflict of interest.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41576-025-00849-8.

Peer review information *Nature Reviews Genetics* thanks Andre Franke, John Rawls and Rafael Valdes-Mas for their contribution to the peer review of this work.