

# Bugs as features (part 2): a perspective on enriching microbiome–gut–brain axis analyses

Received: 22 May 2023

Accepted: 22 September 2023

Published online: 5 December 2023

 Check for updates

Thomaz F. S. Bastiaanssen<sup>1,2</sup>✉, Thomas P. Quinn<sup>3,5</sup> & Amy Loughman<sup>4,5</sup>

The microbiome–gut–brain axis field is multidisciplinary, benefiting from the expertise of microbiology, ecology, psychiatry, computational biology, and epidemiology among other disciplines. As the field matures and moves beyond a basic demonstration of its relevance, it is critical that study design and analyses are robust and foster reproducibility. In this companion piece to Bugs as features (part 1), we present techniques from adjacent and disparate fields to enrich and inform the analysis of microbiome–gut–brain axis data. Emerging techniques built specifically for the microbiome–gut–brain axis are also demonstrated. All of these methods are contextualized to inform several common challenges: how do we establish causality; how can we integrate data from multiple ‘omics techniques; how might we account for the dynamicism of host–microbiome interactions? This perspective is offered to experienced and emerging microbiome scientists alike to assist with these questions and others at the study conception, design, analysis, and interpretation stages of research.

The microbiome–gut–brain axis is informed by biological and epistemological knowledge from many disciplines, spanning microbiology, ecology, psychiatry, and others. Similarly, in its analysis, it is strengthened by methods from across the scientific landscape, as well as some truly interdisciplinary approaches developed specifically for the microbiome–gut–brain axis field (Fig. 1).

In part 1, we introduced core concepts and foundations of compositional data analysis of the microbiome–gut–brain axis<sup>1</sup>, ranging from study design and pre-registration of analysis, to selecting the most suitable diversity metrics, and the options for functional inference. In part 2, we provide a perspective on how to leverage techniques from other disciplines, and provide future directions for the microbiome–gut–brain axis field. We hope that this mapping of the broader landscape will provide useful navigation from which the reader may explore original sources as per their needs and interests.

One aim of this piece is to provide context for the methods borrowed, adapted, and developed from both adjacent and far-flung fields and to aid the reader in appraising their respective strengths and weaknesses for microbiome analysis.

As a guiding principle, we believe that the microbiome–gut–brain axis field has an imperative to become a more reproducible science and to operate from a place of deeper statistical and biological understanding. The techniques described in the following have been carefully examined and selected to ensure that they are fit to drive the field toward this goal.

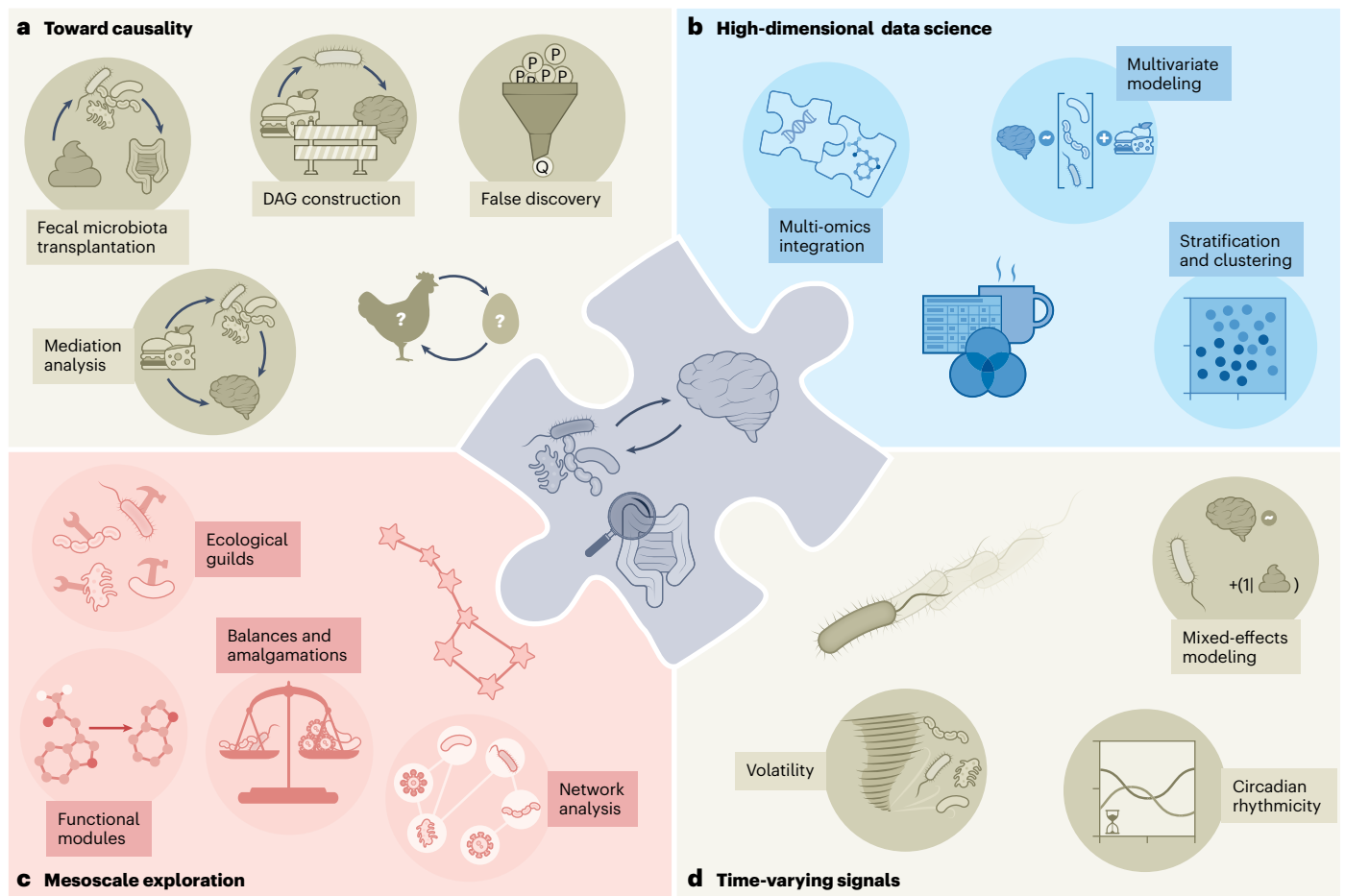
## Causality, uncertainty and the microbiome

There has been a growing call for experiments that can establish causality in the microbiome–gut–brain axis field<sup>2,3</sup>. Causality is a philosophically and statistically contentious term. Granger causality can be thought of as a pragmatic approach to estimating causality between

<sup>1</sup>APC Microbiome Ireland, University College Cork, Cork, Ireland. <sup>2</sup>Department of Anatomy and Neuroscience, University College Cork, Cork, Ireland.

<sup>3</sup>Independent Scientist, Geelong, Victoria, Australia. <sup>4</sup>IMPACT (the Institute for Mental and Physical Health and Clinical Translation), Food and Mood Centre, Deakin University, Geelong, Victoria, Australia. <sup>5</sup>These authors jointly supervised this work: Thomas P. Quinn, Amy Loughman.

✉ e-mail: [thomazbastiaanssen@gmail.com](mailto:thomazbastiaanssen@gmail.com)



**Fig. 1 | Multidisciplinary techniques that enrich the microbiome-gut-brain axis field that are discussed in this Perspective.** **a**, Constructing a DAG facilitates statistical techniques such as causal inference, and experimental procedures such as FMT can be used to interrogate causality and directionality. **b**, Techniques such as multivariate modeling and ordination can be used to analyze and interpret large-omics datasets. **c**, Higher-order patterns within

microbiome data such as interaction networks, functional modules, ecological guilds, and amalgamations, called mesoscale features, are used to ask and answer ecologically relevant questions. **d**, Microbiome time series can be analyzed using mixed-effect models. Special cases of time-series microbiome data, such as microbial volatility, as well as circadian rhythms can be used to ask and answer targeted questions.

occurrences A and B. In a nutshell, if knowledge of the occurrence A helps predict the occurrence of B, A is said to ‘Granger-cause’ B. However, in the case of complex systems such as the microbiome, where nonlinear dynamics are ubiquitous, Granger causality may not be appropriate<sup>4</sup>. Historically compelling sets of criteria to establish causality between a microorganism and a disease exist, including Koch’s postulates<sup>5</sup> and the Bradford Hill criteria<sup>6</sup>. Since these are less applicable to the ecosystem approach required for the microbiome-gut-brain axis, we will not elaborate on these criteria (but see Box 1, which leverages experimental design to interrogate causality). Rather, we provide an overview of causality concepts from epidemiology and econometrics that have been applied to the microbiome-gut-brain axis, as well as some important pitfalls.

**Causal inference analysis**

Causal inference is commonly the underlying motivation for microbiome-gut-brain axis studies, even when it is not being explicitly tested. As outlined by ref. 7, being explicit about the causal motivations of (even) an observational analysis “reduces ambiguity in the scientific question, errors in the data analysis, and excesses in the interpretation of the results”. Rather than avoiding causal language because, as the oft-repeated cautionary tale goes, correlation does not mean causation, Hernán<sup>7</sup> suggests that we instead ask clearer causal questions

and improve use of causal inference methods such as adjusting for confounding. There are many occasions where randomized controlled trials that would provide stronger evidence of causality are not feasible, biologically plausible, or indeed ethical. The complexity of the gastrointestinal and microbial environments is certainly difficult to replicate completely as interventions in clinical and even preclinical trials.

A directed acyclic graph (DAG<sup>8</sup>), or causal diagram, is a useful first step in making explicit the causal hypotheses and underlying assumptions about variables in a study. Creating a DAG serves as a prompt to consider, discuss with colleagues, and design analyses. It is best done at the conception phase of a study so that it may inform aspects of study design, from the timing of data collection to the list of potentially confounding variables about which to collect data. We stress here that while DAGs are a helpful tool to ask causal questions, they do not necessarily allow the user to quantify causality from cross-sectional data. Specialized mechanistic follow-up studies remain the gold standard in this regard. In brief, hypothesized relationships between variables are represented by arrows between them, pointing from cause to effect. By convention, causal diagrams point left to right, with exposure variables on the left and outcome variable on the right. Then add any variables that causally impact the main exposure of interest, or the outcome, using arrows between variables to depict the direction of causality. DAGs must be acyclic; that is, variables must not contain feedback

## BOX 1

## Fecal microbiota transplantation

Fecal microbiota transplantation (FMT) involves transferring the microbiome from a donor to a recipient host, often after the recipient microbiome has been washed out using antibiotics or by inducing diarrhea. The idea behind the procedure is to transfer a microbial ecosystem, and potentially the influence said ecosystem has on its host along with it. FMT has shown promise as a therapy for a wide array of disorders<sup>86</sup>.

FMT is also used as an experimental procedure to investigate causality in a preclinical setting. Often in preclinical studies, murine models are used as the recipient of human microbiomes. FMT can be a useful tool to establish that a phenotype can be transferred by the microbiome, implying that the microbiome is a causal factor in the development of a phenotype. FMT experiments have been used in studies that provide evidence for the microbiome–gut–brain axis, including in depression and aging<sup>87,88</sup>. While FMT can be a powerful experimental tool, experimental designs involving FMT are non-trivial and have been the subject of valid criticism<sup>2</sup>. One such criticism is that observational and experimental units are often conflated in FMT studies. Observational units are the number of recipients, whereas experimental units are the number of donors. Essentially, the criticism is that in a study where a behavioral phenotype is transferred from a patient donor to 20 mouse recipients, the *N* number is not 20 but rather 1, as there was only 1 donor microbiome that transferred the phenotype.

Pooling donor fecal microbiome samples comes with drawbacks and should not be considered the default option for several reasons. First, pooling masks the inter-donor variance of the microbiome, which makes it difficult to trace back and investigate what features

of the donor microbiome may have caused a phenotype to be transferred. Second, pooling produces microbiome compositions never found in nature. It is well known that numerous taxa display competitive exclusion; that is, they never stably appear in the same ecosystem. Pooling can therefore create unstable microbial ecosystems, which may end up in distinct compositional equilibria in the recipients. The recommendation is therefore to power the study on the basis of the number of donors rather than number of recipients<sup>89</sup>. Currently, inter-recipient variance in microbiome composition is hard to estimate due to the differences in methodology between FMT studies, and we recommend taking several recipients per donor to estimate the inter-recipient variance in colonization, which may also depend on the donor.

In terms of statistics, we suggest generalized linear mixed models to account for inter-donor variance, using donor identity as a random effect. These types of models can, for example, be found in the highly cited R library lme4<sup>70</sup>.

Strain transmission and engraftment analysis is a relatively new field of study with applicability to FMT studies as well as to non-FMT horizontal microorganism transmission studies<sup>90–93</sup>. Confirming whether a strain has been transferred from a donor sample—rather than simply being very similar but unrelated—requires shotgun metagenomics-level resolution, as the 16S rRNA gene alone does not allow for this level of precision (but compare recent developments in full-length 16S sequencing analysis<sup>94</sup>). The most recent versions of MetaPhlan<sup>95</sup> and StrainPhlan<sup>95</sup> come with a specialized script to estimate microbial strain transmission and engraftment alike.

loops; relationships between variables must be depicted as unidirectional. This differs from infographics of gut–brain interactions, which are frequently bidirectional as per biological reality. An example DAG can be found in Fig. 2a, and we expand on DAG creation in Box 2. This DAG was created using the dagitty R library and reflects variables relevant to the previously published schizophrenia dataset<sup>9</sup> used in the accompanying Rmarkdown script.

**Mediation analysis.** Mediation analysis is used to investigate whether a variable transmits its effect on the outcome through another mediator variable<sup>10</sup>. For example, an effect of diet on host behavior is well documented, as are effects of diet on the microbiome<sup>11</sup>. Similarly, the gut microbiome is also known to affect host behavior<sup>12</sup>. If we were to test whether diet could affect host behavior via its effects on the microbiome, that would require a mediation analysis. Where mediation explains a relationship, there are two main possibilities:

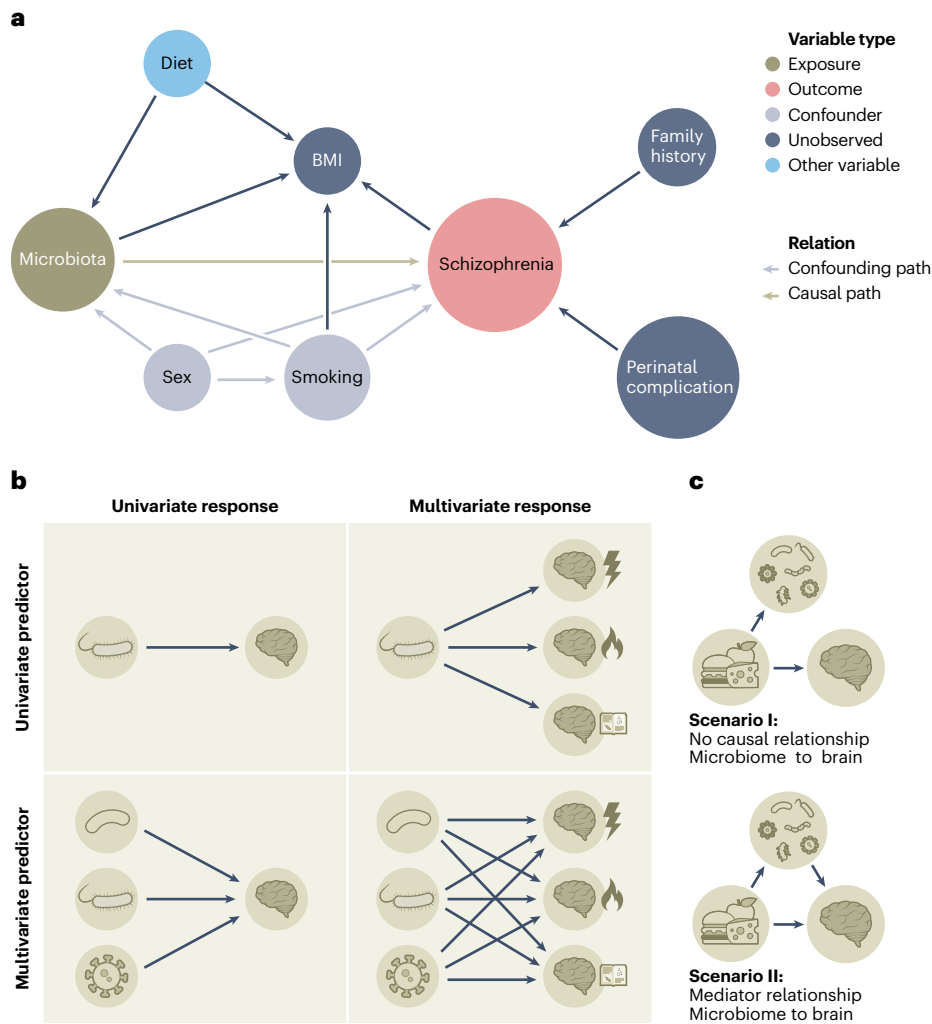
- Partial mediation refers to the scenario where there is both a direct effect and an indirect (mediation) effect; for example, if diet were to both directly affect behavior and indirectly affect behavior by modulating the microbiome—which in turn affects behavior.
- Complete mediation refers to the scenario where—using the preceding example again—diet affects only the microbiome, which in turn affects behavior, but diet on its own does not directly affect behavior.

One recent example of how mediation analysis can be used in the microbiome–gut–brain axis field can be found in the context of autism, diet, and the microbiome<sup>13</sup>. The authors convincingly

showed that alterations in the microbiomes of autistic children can be explained by a restricted diet, a common trait in autistic children. They concluded that since diet can explain the altered microbiome, that altered microbiome does not play a causal role in the occurrence of autism. In a letter to Yap et al.<sup>13</sup>, Morton et al.<sup>14</sup> argued that their model implicitly assumed the absence of a relationship between diet and the microbiome (for example, independence), which is known to be untrue. Morton et al.<sup>14</sup> argued that a more appropriate model would be one where diet affects (1) host phenotype directly and (2) the microbiome, which in turn affects phenotype. Essentially, they argue that the microbiome acts as a partial mediator in this autism example. See Fig. 2c for two miniature DAGs illustrating these two scenarios.

Several excellent tools exist to perform mediation analysis. The mediation package in R takes standard generalized linear model fits as input<sup>15</sup>. Also see the primer on how to perform a mediation analysis in R in the Supplementary Information.

We note that mediation is accompanied by inherently longitudinal assumptions. One presumes that due to the occurrence of some exposure at time 1, a mediating variable is affected at time 2, and the outcome shift as a result is observed at some point in the future (time 3). The use of mediation analysis in cross-sectional observational data, although common, is not considered best practice<sup>16</sup>. The reason for this is that it presumes that the causal chain being tested is correct and precludes an examination of a potential alternative temporal order of the variables. This is particularly relevant for variables that are dynamic, such as diet, the microbiome, and mental states. So mediation analysis in cross-sectional observational studies is correlational and needs to be validated in targeted follow-up studies. Some alternative options



**Fig. 2 | An example DAG followed by graphical representations of multivariate and mediation analyses using miniature DAGs. a,** A DAG describes a hypothetical causal pathway, where an arrow from A to B ( $A \rightarrow B$ ) suggests a causal relationship between A and B. Ideally, the framework of constructs and variables is considered before (and thereby informing) data collection. Some examples of relevant unobserved variables are shown here in gray. Under the assumptions of this DAG and with only three available covariates (sex, smoking status, and body mass index), sex and smoking status comprise the minimal adjustment set required to estimate the total effect of gut microbiota on schizophrenia. **b,** Table illustrating the relationship between models where predictors and responses can be either univariate or multivariate. Predictor

variables are on the left of each diagram, whereas response variables are on the right. Univariate predictor variable to multivariate response variables was included for the sake of completion. Practically, a univariate predictor variable to multivariate response variables can be approached in the same fashion as multivariate predictor variable to a univariate response variable. **c,** Two DAGs illustrating the scenarios from the section on mediation analysis. Diet affects the microbiome, and diet affects the brain. Mediation analysis helps us ask and answer whether measured associations between the microbiome and the brain are spurious (scenario I) or whether the microbiome affects the brain through diet (scenario II). BMI, body mass index.

with fewer data requirements have been trialed through data simulation,<sup>17</sup> demonstrating that sequential mediation—when data for the exposure, mediator, and outcome are collected only once each, but at least longitudinally and in a meaningful temporal sequence—can provide adequate sensitivity to identifying the presence of mediation. The gold standard is the resource-intensive multilevel longitudinal mediation, where variables that represent enduring exposures (such as diet) are collected repeatedly, and path coefficients between variables are allowed to vary across individuals. This may also be ideal for contexts in which a large degree of inter-individual variability might be expected (such as in host–microbiome studies).

Notably, estimating an indirect effect through mediation analysis requires substantially more power than estimating direct effects in traditional analysis. For example, one popular method to estimate the effect size of a mediation analysis is to multiply the two coefficients (exposure to mediator and mediator to outcome), which will always

yield a smaller absolute coefficient compared to its two component coefficients<sup>18</sup>. Also see Box 3 on power calculations.

### Mendelian randomization and the microbiome

In contrast to causal inference, Mendelian randomization is a statistical method from the field of epidemiology, often used to estimate the causal effects of genetic factors on a phenotype in large cohorts.<sup>19–21</sup> In a nutshell, Mendelian randomization leverages the fact that genotype is fixed at conception and therefore takes place before the manifestation of a phenotype. This, along with other assumptions, allows the researcher to assess causality and directionality of the exposure (genotype) on the outcome (phenotype).

Recently, Mendelian randomization has been applied to microbiome data in the sense that genotype is replaced with microbiome metagenomic content. Particular care is therefore necessary. Unlike genotype, the microbiome is not fixed at conception but remains in

**BOX 2**

## Constructing a directed acyclic graph

In an illustrative example of the power of interdisciplinary expertise, ref. 96 builds on the hybrid field of molecular epidemiology to demonstrate the application of causal inference analysis in 'omics, with the following phases:

**A: Ask specific, detailed research questions.** Build a DAG. A DAG may be used to identify the hypothesized relationships among the exposure (microbiome), outcome (for example, schizophrenia), and potentially confounding variables of interest on the basis of prior knowledge<sup>97</sup>. By convention, causal diagrams point left to right, with exposure variables on the left and outcome variable on the right. Then add any variables that causally impact the main exposure of interest, or the outcome, using arrows between variables to depict the direction of causality. DAGs must be acyclic; that is, variables must not contain feedback loops; relationships between variables must be depicted as unidirectional. For DAGs with bidirectional relationships, assumptions need to be made as to the dominant direction of action in a given model.

**B: Test the exposure–outcome association.** Run unadjusted association analyses between exposure and outcome. This will require an operational definition of each. For example, which microbial diversity metric will be used? How is schizophrenia assessed? Is the exposure–disease association linear or nonlinear in form?

**C: Consider other variables.** Using the DAG, identify potentially confounding variables. This could include any common causes of exposure and outcome (for example, cigarette smoking, which may affect both the gut microbiome and risk of schizophrenia) and include any proxy measures of unmeasured common causes of both exposure and outcome (for example, family history of schizophrenia as a proxy for unmeasured genetic factors that could impact both the gut microbiome and risk of schizophrenia)<sup>98</sup>. In addition, consider technical or processing variables that might affect measurement precision, for example, microbiome sequencing batch effects<sup>99,100</sup>.

Understanding how each of these causal and non-causal potentially confounding variables associate with the exposure and outcome will provide information to help assess whether the putative factor is a mediator, an antecedent, an instrumental variable (antecedent of exposure), or a disease consequence. If so, the putative factor is not a confounder and should not be adjusted for. As well as adjusting for confounding, one may include disease determinants that are independent of outcome<sup>101</sup>. Too many variables will negatively impact power, so is dimension reduction a possibility? Is there collinearity or redundancy?

**D: Build multivariable models.** Consider refining the a priori DAG on the basis of the data, adding or removing variables as required, and reporting models that are adjusted on the basis of the original DAG, updated DAG, and with any additional processing and precision-enhancing variables that reduce measurement error. While it is desirable to investigate all associations manually, DAGitty software<sup>102</sup> does provide identification of 'minimal adjustment sets,' which can be used to block all non-causal paths, to estimate the total or direct effects between exposure and outcome. Non-causal confounders such as batch effects do not fit strictly within this tool of causal inference but comprise an unwanted source of variance that should nonetheless feature in adjusted models<sup>103</sup>.

**E: Evaluate non-causal and causal explanations.** Interpret the findings of both unadjusted and adjusted models. Consider possible biases, such as measurement and selection bias, and other explanations of effects such as reverse causality.

We stress here that DAGs are a tool to formulate causal questions; specialized mechanistic follow-up experiments remain the gold standard to establish causality in the microbiome. For a definitive guide on constructing a DAG, we refer the reader to an excellent free online course<sup>104</sup> informed by the authors of the corresponding text<sup>105</sup>. There is also a useful free online tool<sup>102</sup>, although in reality DAGs can be drawn as a proverbial back-of-napkin sketch almost as effectively.

constant flux throughout life (Time-varying signals). While it makes sense to assess the causal effect of host genotype on the microbiome, for example, in the case of a host metabolic disorder altering the host gut and hence the microbiome<sup>22</sup>, it seems much less clear whether taking the microbial metagenome as a fixed exposure is appropriate.

### High-dimensional data science

Microbiome–gut–brain axis experiments tend to yield complex, high-dimensional datasets. Here, we will discuss techniques to handle these types of data and strategies to integrate multiple high-dimensional data from the same experiment.

### Stratifying and clustering samples

In some cases, it is necessary to stratify data into clusters, distinct subgroups based on microbiome signature. Stratification is a common method of defining enterotypes, which are large subgroups based on microbial taxonomic composition. The precise number of true enterotypes, as well as the best way to define them, is still up for debate (although 3–4 enterotypes are often cited<sup>23,24</sup>). Initial efforts involved calculating a Jensen–Shannon dissimilarity matrix and performing cluster analysis (clusters here corresponding to enterotypes) using

the partition around medoids approach<sup>25</sup>. More recently, studies have employed the Dirichlet multinomial mixtures approach, a promising technique to estimate enterotypes from the Bayesian school<sup>26</sup>. In brief, the method involves estimating a probability vector for each sample and then estimating whether these vectors came from the same source (metacommunity, enterotype) or from separate enterotypes. Enterotypes appear to be important constructs because they are related to factors such as host health, diet, and exercise, despite some known limitations<sup>27</sup>. Notably, bacterial load is not easy to estimate using metagenomic techniques such as 16S and shotgun (although compare ref. 28) but can rather be assessed by pan-bacterial quantitative PCR (qPCR) or, most accurately, using flow cytometry. Bacterial load is associated with enterotype identity and may bias results<sup>23,24,27</sup>.

Stratifying samples on the basis of feature abundance is a defensible approach under some circumstances, for example, when pursuing functional groups of microorganisms that might exhibit competitive exclusion. However, it is rarely advisable to stratify samples into subgroups while in the middle of an analysis or when working with datasets comprising only 10s–100s of samples because there are too few samples for validation. It is especially important to validate data-driven stratification, either in a new cohort or in a subsection of withheld data

## BOX 3

## Power calculations and adjustment for multiple testing

Power calculations can be daunting but are an important part of figuring out how many samples per group are required to register an effect of a given magnitude. Using power calculations can help avoid two undesirable scenarios. First, it saves us from going through the trouble of running an experiment that would not be able to find any effects that may be present (that is, an underpowered set-up). Second, it saves us from collecting more samples than necessary to test a hypothesis (that is, an overpowered set-up). This second scenario is especially important as part of our ethical commitment to avoid excess animal suffering (that is, 'to Reduce' as per the 3Rs of Animal Welfare), as well as participant burden in human studies. Both underpowered and overpowered studies waste precious resources.

In the case of feature-wise analyses such as differential abundance analysis, microbiome power calculations differ from other power calculations for one big reason: microbiome data are highly dimensional. Usually, the number of features is larger than the sample size. This necessitates an adjustment for multiple testing (for example, by Bonferroni's correction, the Benjamini–Hochberg procedure, or Storey's  $q$  value<sup>106</sup>), which must be accounted for as part of the power calculation. This is because without a false discovery rate (FDR) correction and instead using just a traditional  $P < 0.05$  threshold, approximately 1 in 20 tests would give a false positive result, making the  $P < 0.05$  threshold too lenient of an evidence threshold for most applications. Unless there is a compelling reason not to do so, researchers should always correct for multiple testing in microbiome and other high-dimensional datasets. Also see ref. 67 for further discussion on differential abundance testing. For some adjustments, notably the Benjamini–Hochberg procedure and Storey's  $q$  value, the adjustment depends on the distribution of  $P$  values, which itself depends on the number of true positives. Thus, to account for adjustment, one must estimate how many features will be differentially abundant between treatment groups. This can be difficult to estimate, making it hard to choose the effective FDR-adjusted  $\alpha$  threshold for the power calculations.

So how to overcome this challenge? At one extreme, we could consider what sample size is needed to detect a single differentially abundant feature if all other features are unaltered. Let us say only 1 microorganism out of 100 tested features is actually different. (Incidentally, in the case that only one microorganism increases in abundance, the counted proportion of all the others in the sample would decrease. See the discussion on compositionality in the part 1 companion piece to this Perspective<sup>1</sup>.) Then the adjusted  $P$  value needed to reject the null hypothesis would be 100 times lower than the initial  $\alpha$ . Thus, to obtain a conservative sample size estimate, one should perform a power calculation where the significance threshold  $\alpha$  is divided by the number of features expected to be tested (not the number anticipated to be differentially abundant).

At the other extreme, we could consider what sample size is needed to detect a single differentially abundant feature if all features differ between groups. In the case that all individual  $P$  values are lower than the alpha threshold, the Benjamini–Hochberg correction will not adjust any of the  $P$  values to the point that they no longer are under  $\alpha$ . Thus, to obtain a liberal sample size estimate, one should perform a power calculation where the significance threshold  $\alpha$  is not adjusted. From these two extreme scenarios, we can formulate bounds: for  $D$  features, the adjusted  $\alpha$  used for power calculations should fall between  $\frac{\alpha}{D}$  and  $\alpha$ , depending on the number of features we expect to be differentially abundant.

In the latter case, we recommend a heuristic:  $\alpha_{\text{adjusted}} = \alpha \times \frac{M}{D}$  where  $D$  is the total number of features tested, and  $M$  is the total number of features expected to differ significantly between the groups. This adjusted alpha can then be used for power calculations as done under ordinary circumstances. Once a study is appropriately powered for differential abundance analysis, it seems reasonable to assume that, as a rule of thumb, the study is also appropriately powered for testing differences in alpha and beta diversity. Although, to be safe, we recommend that analysts add the total number of planned alpha and beta diversity analyses to  $D$  in the preceding formula.

that can be used as a validation set. Spurious strata can frequently arise from technical or biological artifacts, leading enthusiastic researchers on long and fruitless tangents. Clustering algorithms, by design, will cluster and can even find seemingly impressive clusters among random noise.

### Multi-omics integration

The microbiome refers to the collection of microbial genes in a sample. While the present work focuses on this type of data, other 'omics also exist<sup>29–32</sup>. Microbial genetic data provide evidence of microorganism presence as well as their functional potential.<sup>33</sup> Besides metagenomics, the three most common types of 'omics data in microbiome–gut–brain axis studies are the following:

- **Metabolomics:** the metabolites and small molecules in a sample. Mass spectrometry or nuclear magnetic resonance spectroscopy are the most common techniques to measure the metabolome. Metabolomics can shed light on the functional consequences of a given microbiome.
- **Metatranscriptomics:** the sequencing of RNA in a sample. In practice, metatranscriptomics can be thought of as RNAseq on

a microbial community rather than a single organism. Metatranscriptomics can tell us about the transcriptional activity of a microbial community. A microorganism may be present and have a certain gene, but it may not be transcribing that gene<sup>34</sup>.

- **Metaproteomics:** the proteins in a sample. Typically, metaproteomics relies on specialized mass spectrometry techniques to identify proteins and derive their sequences. Metaproteomics goes further than metatranscriptomics and tells us whether the transcribed genes are translated to proteins.

There are three broad approaches to data integration, treating datasets as either univariate or multivariate. The suffixes -variable and -variate are often used interchangeably, but they refer to subtly but meaningfully distinct concepts<sup>35</sup>. In short, -variate refers to the structural nature of the data, whereas -variable refers to the structure and number of variables in the statistical model (also Fig. 2b):

- **Univariate-univariate:** with two separate multivariate datasets, one can perform an acceptable analysis using 'simple' univariate methods by correlating each microbiome feature (for example, taxa or gene) with individual features in the other dataset, one at

a time. Metrics such as Pearson's and Spearman's Rank correlation coefficients are commonly used for this purpose. Both of these metrics can be thought of as special cases of a linear model, which we particularly recommend as it allows for the inclusion of covariates.

- **Univariate-multivariate:** treat one feature from one dataset as a dependent variable, and use all features from the other dataset as the predictors. By repeating this for each feature, all associations between the datasets are described.
- **Multivariate–multivariate:** multivariate regression, such as a canonical correlation analysis or redundancy analysis<sup>36</sup>, to obtain a single model that associates all features from one dataset with all features from the other dataset. The mixOmics package provides a user-friendly implementation of multivariate methods for microbiome research<sup>37,38</sup>. Similarly, neural networks or other machine learning can be used<sup>39–41</sup>.

For multivariate–multivariate analysis, one compelling method, DIABLO<sup>42</sup>, extends this approach by comparing association networks between phenotypes, focusing on the interactions between two 'omics data tests rather than the values within the two individual datasets. This permits the discovery of patterns not necessarily visible in either of the individual datasets. Note also that it is possible to extend any of these approaches to incorporate external information about known relationships within the individual dataset or across the two datasets (for example, via a gene ontology database). For example, joint pathway analysis takes advantage of existing biological knowledge structures by mapping two 'omics datasets to the same metabolic pathways and then assessing the joint coverage as a readout of pathway enrichment<sup>43,44</sup>. Such knowledge structures could also be leveraged to constrain an analysis to include only feature pairs that are canonically able to interact according to the database, thus potentially preserving power by avoiding unnecessary hypothesis testing, as exemplified by the anansi framework.<sup>45</sup> Whatever approach one uses, analysts should take care to normalize or transform their data appropriately, especially since correlations can yield spurious results when measured for compositional data<sup>46,47</sup>. As with differential abundance analysis, multiple multivariate tests should always be accompanied by an FDR adjustment. When null hypothesis testing is not straightforward in multivariate methods, permutations or algorithmic validation (for example, cross-validation) may be used there instead.

## Exploring the mesoscale

Mesoscale features of the microbiome contain information about patterns within parts of a microbiome that can be seen across samples—not necessarily its smallest parts (the microscale) or about the whole system (the macroscale). Mesoscale analysis focuses on identifying community-level patterns that define the ecosystem(s). This is useful because phenomena in a microbiome may be more readily explained by aggregated patterns in the data rather than by any individual feature. The mesoscale is an important object of study in theoretical ecology<sup>48</sup>. These emerging techniques derive microbiome mesoscale features. The first three make use of external knowledge. The final two are purely data driven:

- **Ecological guilds:** ecological guilds are taxonomically unrelated but functionally related clusters of microorganisms that have a shared role in the microbiome (for example, occupy a common niche). For example, microbial communities across a wide span of environments, including soil, the ocean, and the human gut, could be assigned to trophic groups on the basis of how they feed on certain substrates and subsequently pass on metabolites to another trophic group<sup>49</sup>. While ecological guilds are a promising concept in microbiome science, to our knowledge there are currently no standardized pipelines or databases that can be used to detect and compare ecological guilds across cohorts and experiments

(but compare ref. 50). Such tools would be welcome additions to the field<sup>51,52</sup>.

- **Functional modules:** functional modules are a list of curated metabolic pathways encoding for processes that are related to a specific aspect of the microbiome. We will consider two classes of functional modules. Gut–brain modules cover pathways that are related to gut–brain communication, such as serotonin degradation and histamine synthesis. The complete list of gut–brain modules can be accessed as a table in the supplementary files of ref. 53. Gut–metabolic modules, from the same group, encompass metabolic processes in the microbiome. Changes in gut–metabolic modules can indicate a shift in the microbial metabolic environment and thereby in the fitness landscape, thus allowing for microorganisms with different metabolic features to thrive. The complete list of gut–metabolic modules can be accessed as a table in the supplementary files of the paper that introduced them<sup>54</sup>. Functional modules are especially interpretable and help develop hypotheses for future experiments. We note that functional module analysis depends on the availability of a functional abundance table (see the section on functions in the companion piece of this Perspective<sup>1</sup>).
- **Enrichment analysis:** differential abundance (DA) analysis is first performed on taxa or genes or other 'omics features, and then a functional database is used to summarize the DA results. In the simplest case, the DA results can be dichotomized into significant or non-significant, and functional status can be dichotomized as present or absent. For each function, one could perform a Fisher exact test (or similar) to test over-enrichment among the significant taxa, genes, or features<sup>43,55</sup>. Gene set enrichment analysis is a popular generalization of this concept and is commonplace in gene expression analysis<sup>55</sup>.
- **Network analysis:** network analysis is most often applied to study or visualize associations between microbiome features such as taxa or genes. This requires some measure of association. The Pearson's correlation is the most popular; however, correlations have been shown to yield spurious results when applied to compositional data<sup>56</sup>. For this reason, several alternatives have been designed specifically for microbiome data<sup>46,57,58</sup> (also see the discussion on compositionality in the companion piece to this Perspective<sup>1</sup>). These metrics build on a log-ratio transformation that makes them more robust to the biases introduced by compositionality<sup>59</sup>, although they can still be prone to false positives<sup>60</sup>. A recent benchmark of 213 single-cell datasets has shown that proportionality has excellent performance for sparse high-dimensional data such as those encountered in microbiome research<sup>61</sup>.
- **Balance selection and summed log-ratios:** balance selection and data-driven amalgamation are two new approaches to learning mesoscale features directly from the data. In both cases, the motivation is to find mesoscale features that serve as a biomarker to predict another variable of interest. These mesoscale features are unique in that they are defined explicitly as a ratio between groups of taxa, similar to the Firmicutes-to-Bacteroidetes ratio<sup>62</sup>. By using a ratio of taxa, any normalization factors would cancel, thus making the method normalization-free. When the groups of taxa are summarized by a geometric mean, the resultant mesoscale feature is called a balance. When they are summarized by a sum, the resultant mesoscale feature is called a summed log-ratio. Software tools such as selbal<sup>63</sup>, balance<sup>64</sup>, amalgam<sup>65</sup>, and CoDaCoRe<sup>66</sup> enable analysts to learn mesoscale features in a few lines of code. It is customary to validate the reliability of these features by measuring predictive performance in a withheld test set<sup>67</sup>.

## Time-varying signals

While 16S and shotgun sequencing allow only for snapshot measurements of the microbiome, in reality, microbiomes are dynamic ecosystems in constant flux. To account for this, it has become more common

for studies to include multiple (repeated) measures of the microbiome. However, time-series analysis necessitates special considerations<sup>68,69</sup>.

### Statistical considerations with time-varying data analysis

Time-series data, where the same microbiomes are sampled repeatedly, intrinsically break the assumption of independence between samples that many statistical tests rely on. Mixed-effects models are well equipped to handle this type of data, using the resampled microbiome as a random effect. The well-documented and widely used lme4 package in R provides an excellent framework for this<sup>70</sup>. More-specialized microbiome tools such as MaAsLin2<sup>71</sup> are also available.

A recent study on the temporal variation of the microbiome estimated that inter-individual variation is smaller than intra-individual variation<sup>72</sup>. Taking several microbiome measurements over time may therefore be necessary to increase power to detect group differences. Another approach to deal with this high intra-individual variation is to include microbial variance in the model<sup>73</sup>. This allows investigation of whether microbial variability itself is associated with the phenotype of interest. The idea that microbial variance rather than abundance can be informative for a phenotype is core to the idea of volatility.

### Volatility

The microbiome is a dynamic ecosystem that undergoes constant change. The degree of change in the microbiome over time is called volatility, which is inversely related to stability. The term was first coined during the early days of the Human Microbiome Project in the context of instability<sup>74</sup> and was soon thereafter used to describe the degree of change in the microbiome between two time points<sup>75</sup>. It can be helpful to think of volatility as a change in sample diversity (alpha or beta) over time. In a neutral setting, without intervention, a higher volatility is generally considered to be associated with negative health outcomes<sup>76</sup>. One way to calculate volatility is to measure the beta diversity between two or more time points corresponding to the same host. When measuring volatility in this fashion, it is especially useful to choose a beta diversity metric that is also a distance (that is, follows triangle inequality, such as PhiLR or Aitchison distance) so that any comparisons are standardized for all time points. Volatility has recently been shown to differ between enterotypes, indicating that microbiome composition at least partially explains microbiome volatility<sup>72</sup>. Because sampling depth is known to affect beta diversity indices, it may be worth subsampling before volatility analysis<sup>77</sup>.

### Circadian rhythms

Circadian rhythms, or 24-hour biological cycles, are key in maintaining physical and mental health<sup>78</sup>. The microbiome is an example of a biological system that displays such a 24-hour cycle<sup>79,80</sup>. Typically, models that assess rhythmicity will make use of a sinusoidal model rather than a conventional linear model. Circadian rhythms are a special case of time-varying data as there is an implicit assumption that microbial taxa will oscillate around a set mean (mesor). Due to the 24-hour period of a circadian rhythm, time of sampling becomes an important source of variance and thus a relevant covariate even when the researcher is not interested in investigating circadian rhythms per se. We recently developed the kronos package in R to analyze circadian rhythms in the microbiome<sup>81</sup>.

### Consolidating and looking forward

As the microbiome–gut–brain axis field continues its maturation, we shift our priorities away from a basic demonstration of relevance and toward formulating and addressing more mechanistic questions. In the final section of this Perspective, we briefly look forward to efforts to consolidate findings in the field.

### Meta-analyses

In a nutshell, meta-analyses incorporate outcomes from numerous studies on the same subject to estimate a ‘true effect’ on the basis

of a weighted summary of the component studies. When planning a meta-analysis of microbiome–gut–brain axis studies, it is particularly important to consider which features to analyze. For example, it may be preferable to investigate the role of microbial functions in a disorder rather than taxonomy-level data. In addition, due to large inter-study heterogeneity in methodology, it may not be appropriate to compare reported outcomes from studies at all, and a ‘meta-re-analysis’ from raw data may be warranted. This again underlines the importance of making microbiome data publicly available. We note and applaud the burgeoning development of meta-analysis methods for microbiome studies such as MMUPHin by ref. 82, which account for the heterogeneity in pre-processing that precludes standard meta-analysis tools and techniques.<sup>83–85</sup> A groundswell of attempts at reproducing previous findings and quantitative synthesis of the literature to date will improve the robustness of the field, as it has done in others.

### Toward enriching microbiome–gut–brain axis research

In this part 2 of our Bugs as features Perspective, we have taken you on a tour of both adjacent and far-flung topics to enrich contemporary microbiome–gut–brain axis research, from Mendelian randomization and mediation analysis to numerous ways to explore microbiome patterns of the mesoscale. In combination with the concepts and foundations detailed in part 1, and the corresponding supplementary code tutorial, we have described the key considerations for microbiome–gut–brain axis analysis<sup>1</sup>. In our opinion, establishing causality, integrating multi-omics data, and accounting for the dynamic nature of the microbiome are key. We hope that this Perspective has assisted with confident navigation of the microbial landscape. We trust that the increased use of biologically and statistically sound methods such as those described here will improve our understanding of the complex phenomenon known as the microbiome–gut–brain axis.

### References

- Bastiaanssen, T. F. S., Quinn, T. P. & Loughman, A. Bugs as features (part 1): concepts and foundations for the compositional data analysis of the microbiome–gut–brain axis. *Nat. Ment. Health* <https://doi.org/10.1038/s44220-023-00148-3> (2023).
- Walter, J., Armet, A. M., Finlay, B. B. & Shanahan, F. Establishing or exaggerating causality for the gut microbiome: lessons from human microbiota-associated rodents. *Cell* **180**, 221–232 (2020).
- Bastiaanssen, T. F. S. & Cryan, J. F. The microbiota–gut–brain axis in mental health and medication response: parsing directionality and causality. *Int. J. Neuropsychopharmacol.* **24**, 216–220 (2021).
- Sugihara, G. et al. Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).
- Koch, R. Untersuchungen über bakterien v. die aetiologie der milzbrand-krankheit, begründer auf die entwicklungsgeschichte bacillus anthracis. *Beiträge zur biologie der pflanzen* **2**, 277–310 (1877).
- Hill, A. B. The environment and disease: association or causation? *Proc. R. Soc. Med.* **58**, 295–300 (1965).
- Hernán, M. A. The c-word: scientific euphemisms do not improve causal inference from observational data. *Am. J. Public Health* **108**, 616–619 (2018).
- VanderWeele, T. J. & Robins, J. M. Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. *Am. J. Epidemiol.* **166**, 1096–1104 (2007).
- Zhu, F. et al. Metagenome-wide association of gut microbiome features for schizophrenia. *Nat. Commun.* **11**, 1612 (2020).
- MacKinnon, D. P., Fairchild, A. J. & Fritz, M. S. Mediation analysis. *Annu. Rev. Psychol.* **58**, 593–614 (2007).
- Logan, A. C. & Jacka, F. N. Nutritional psychiatry research: an emerging discipline and its intersection with global urbanization, environmental challenges and the evolutionary mismatch. *J. Physiol. Anthropol.* <https://doi.org/10.1186/1880-6805-33-22> (2014).



12. Cryan, J. F. et al. The microbiota–gut–brain axis. *Physiol. Rev.* **99**, 1877–2013 (2019).
13. Yap, C. X. et al. Autism-related dietary preferences mediate autism–gut microbiome associations. *Cell* **184**, 5916–5931 (2021).
14. Morton, J. T., Donovan, S. M. & Taroncher-Oldenburg, G. Decoupling diet from microbiome dynamics results in model mis-specification that implicitly annuls potential associations between the microbiome and disease phenotypes—ruling out any role of the microbiome in autism (Yap et al. 2021) likely a premature conclusion. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.02.25.482051> (2022).
15. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. mediation: R package for causal mediation analysis. *J. Stat. Softw.* <https://doi.org/10.18637/jss.v059.i05> (2014).
16. Fairchild, A. J. & McDaniel, H. L. Best (but oft-forgotten) practices: mediation analysis. *Am. J. Clin. Nutr.* **105**, 1259–1271 (2017).
17. Cain, M. K., Zhang, Z. & Bergeman, C. Time and other considerations in mediation design. *Educ. Psychol. Meas.* **78**, 952–972 (2018).
18. Baron, R. M. & Kenny, D. A. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* **51**, 1173 (1986).
19. Smith, G. D. & Ebrahim, S. ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
20. Gagliano Taliun, S. A. & Evans, D. M. Ten simple rules for conducting a Mendelian randomization study. *PLoS Comput. Biol.* **17**, e1009238 (2021).
21. Sanderson, E. et al. Mendelian randomization. *Nat. Rev. Methods Primers* **2**, 6 (2022).
22. Sanna, S. et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nat. Genet.* **51**, 600–605 (2019).
23. Costea, P. I. et al. Enterotypes in the landscape of gut microbial community composition. *Nat. Microbiol.* **3**, 8–16 (2018).
24. Vandeputte, D. et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature* **551**, 507–511 (2017).
25. Arumugam, M. et al. Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
26. Holmes, I., Harris, K. & Quince, C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS ONE* **7**, e30126 (2012).
27. Knights, D. et al. Rethinking ‘enterotypes’. *Cell Host Microbe* **16**, 433–437 (2014).
28. Cruz, G. N. F., Christoff, A. P. & de Oliveira, L. F. V. Equivolumetric protocol generates library sizes proportional to total microbial load in 16S amplicon sequencing. *Front. Microbiol.* **12**, 425 (2021).
29. Lloyd-Price, J. et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).
30. Smolinska, A. et al. Volatile metabolites in breath strongly correlate with gut microbiome in CD patients. *Anal. Chim. Acta* **1025**, 1–11 (2018).
31. Tang, Z.-Z. et al. Multi-omic analysis of the microbiome and metabolome in healthy subjects reveals microbiome-dependent relationships between diet and metabolites. *Front. Genet.* **10**, 00454 (2019).
32. Yachida, S. et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat. Med.* **25**, 968–976 (2019).
33. Aguiar-Pulido, V. et al. Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis: supplementary issue: bioinformatics methods and applications for big metagenomics data. *Evol. Bioinform.* **12**, EBO–S36436 (2016).
34. Abu-Ali, G. S. et al. Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat. Microbiol.* **3**, 356–366 (2018).
35. Mallick, H. et al. Experimental design and quantitative analysis of microbial community multiomics. *Genome Biol.* **18**, 228 (2017).
36. Meng, C. et al. Dimension reduction techniques for the integrative analysis of multi-omics data. *Brief. Bioinform.* **17**, 628–641 (2016).
37. Lê Cao, K.-A., Rossouw, D., Robert-Granié, C. & Besse, P. A sparse PLS for variable selection when integrating omics data. *Stat. Appl. Genet. Mol. Biol.* **7** 35 (2008).
38. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixomics: an r package for ‘omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, e1005752 (2017).
39. Le, V., Quinn, T. P., Tran, T. & Venkatesh, S. Deep in the bowel: highly interpretable neural encoder–decoder networks predict gut metabolites from gut microbiome. *BMC Genom.* **21**, 256 (2020).
40. Morton, J. T. et al. Learning representations of microbe–metabolite interactions. *Nat. Methods* **16**, 1306–1314 (2019).
41. Reiman, D., Layden, B. T. & Dai, Y. Mimenet: exploring microbiome–metabolome relationships using neural networks. *PLoS Comput. Biol.* **17**, e1009021 (2021).
42. Singh, A. et al. Diablob: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* **35**, 3055–3062 (2019).
43. Chong, J. et al. Metaboanalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **46**, W486–W494 (2018).
44. Pang, Z. et al. Metaboanalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, W388–W396 (2021).
45. Bastiaanssen, T. F. S., Quinn, T. P. & Cryan, J. F. Knowledge-based integration of multi-omic datasets with anansi: annotation-based analysis of specific interactions. Preprint at <https://arxiv.org/abs/2305.10832> (2023).
46. Quinn, T. P., Richardson, M. F., Lovell, D. & Crowley, T. M. propr: an r-package for identifying proportionally abundant features using compositional data analysis. *Sci. Rep.* **7**, 16252 (2017).
47. Quinn, T. P. & Erb, I. Examining microbe–metabolite correlations by linear methods. *Nat. Methods* **18**, 37–39 (2021).
48. Hogeweg, P. in *Simulating Complex Systems by Cellular Automata* (eds Kroc, J. et al.) 19–28 (Springer, 2010).
49. Gralka, M., Szabo, R., Stocker, R. & Cordero, O. X. Trophic interactions and the drivers of microbial community assembly. *Curr. Biol.* **30**, R1176–R1188 (2020).
50. Frioux, C. et al. Enterosignatures define common bacterial guilds in the human gut microbiome. *Cell Host Microbe* **31**, 1111–1125 (2023).
51. Lam, Y. Y., Zhang, C. & Zhao, L. Causality in dietary interventions—building a case for gut microbiota. *Genome Med.* **10**, 62 (2018).
52. Zhao, L. et al. Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. *Science* **359**, 1151–1156 (2018).
53. Valles-Colomer, M. et al. The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat. Microbiol.* **4**, 623–632 (2019).
54. Vieira-Silva, S. et al. Species–function relationships shape ecological properties of the human gut microbiome. *Nat. Microbiol.* **1**, 16088 (2016).
55. Irizarry, R. A., Wang, C., Zhou, Y. & Speed, T. P. Gene set enrichment analysis made simple. *Stat. Methods Med. Res.* **18**, 565–575 (2009).
56. Lovell, D., Pawlowsky-Glahn, V., Egozcue, J. J., Marguerat, S. & Bähler, J. Proportionality: a valid alternative to correlation for relative data. *PLoS Comput. Biol.* **11**, e1004075 (2015).

57. Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8**, e1002687 (2012).
58. Kurtz, Z. D. et al. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.* **11**, e1004226 (2015).
59. Quinn, T. P., Erb, I., Richardson, M. F. & Crowley, T. M. Understanding sequencing data as compositions: an outlook and review. *Bioinformatics* **34**, 2870–2878 (2018).
60. Erb, I. & Notredame, C. How should we measure proportionality on relative gene expression data? *Theory Biosci.* **135**, 21–36 (2016).
61. Skinnider, M. A., Squair, J. W. & Foster, L. J. Evaluating measures of association for single-cell transcriptomics. *Nat. Methods* **16**, 381–386 (2019).
62. Mariat, D. et al. The firmicutes/bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiol.* **9**, 123 (2009).
63. Rivera-Pinto, J. et al. Balances: a new perspective for microbiome analysis. *mSystems* **3**, e00053-18 (2018).
64. Quinn, T. P. Visualizing balances of compositional data: a new alternative to balance dendrograms. *F1000Res.* **7**, 1278 (2018).
65. Quinn, T. P. & Erb, I. Amalgams: data-driven amalgamation for the dimensionality reduction of compositional data. *NAR Genom. Bioinform.* **2**, lqaa076 (2020).
66. Gordon-Rodriguez, E., Quinn, T. P. & Cunningham, J. P. Learning sparse log-ratios for high-throughput sequencing data. *Bioinformatics* **38**, 157–163 (2021).
67. Quinn, T. P., Gordon-Rodriguez, E. & Erb, I. A critique of differential abundance analysis, and advocacy for an alternative. Preprint at <https://arxiv.org/abs/2104.07266> (2021).
68. Kodikara, S., Ellul, S. & Lê Cao, K.-A. Statistical challenges in longitudinal microbiome data analysis. *Brief. Bioinform.* **23**, bbac273 (2022).
69. Bokulich, N. A. et al. q2-longitudinal: longitudinal and paired-sample analyses of microbiome data. *mSystems* **3**, e00219-18 (2018).
70. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* <https://doi.org/10.18637/jss.v067.i01> (2015).
71. Mallick, H. et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput. Biol.* **17**, e1009442 (2021).
72. Vandeputte, D. et al. Temporal variability in quantitative human gut microbiome profiles and implications for clinical research. *Nat. Commun.* **12**, 6740 (2021).
73. Martino, C. et al. Context-aware dimensionality reduction deconvolutes gut microbial community dynamics. *Nat. Biotechnol.* **39**, 165–168 (2021).
74. Weinstock, G. M. The volatile microbiome. *Genome Biol.* **12**, 114 (2011).
75. Goodrich, J. K. et al. Conducting a microbiome study. *Cell* **158**, 250–262 (2014).
76. Bastiaanssen, T. F. S. et al. Volatility as a concept to understand the impact of stress on the microbiome. *Psychoneuroendocrinology* **124**, 105047 (2021).
77. Park, D. J. & Plantinga, A. M. Impact of data and study characteristics on microbiome volatility estimates. *Genes* **14**, 218 (2023).
78. Caliyurt, O. Role of chronobiology as a transdisciplinary field of research: its applications in treating mood disorders. *Balkan Med. J.* **34**, 514–521 (2017).
79. Thaïss, C. A. et al. Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. *Cell* **159**, 514–529 (2014).
80. Liang, X., Bushman, F. D. & FitzGerald, G. A. Rhythmicity of the intestinal microbiota is regulated by gender and the host circadian clock. *Proc. Natl Acad. Sci. USA* **112**, 10479–10484 (2015).
81. Bastiaanssen, T. F. S. et al. Kronos: a computational tool to facilitate biological rhythmicity analysis. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.04.21.537503> (2023).
82. Ma, S. et al. Population structure discovery in meta-analyzed microbial communities and inflammatory bowel disease using mmuphin. *Genome Biol.* **23**, 208 (2022).
83. Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A. & Alm, E. J. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat. Commun.* **8**, 1784 (2017).
84. Chong, J., Liu, P., Zhou, G. & Xia, J. Using microbiomeanalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nat. Protoc.* **15**, 799–821 (2020).
85. Morton, J. T. et al. Multi-level analysis of the gut–brain axis shows autism spectrum disorder-associated molecular and microbial profiles. *Nat. Neurosci.* **26**, 1208–1217 (2023).
86. Wortelboer, K., Nieuwdorp, M. & Herrema, H. Fecal microbiota transplantation beyond clostridioides difficile infections. *EBioMedicine* **44**, 716–729 (2019).
87. Kelly, J. R. et al. Transferring the blues: depression-associated gut microbiota induces neurobehavioural changes in the rat. *J. Psychiatr. Res.* **82**, 109–118 (2016).
88. Boehme, M. et al. Microbiota from young mice counteracts selective age-associated behavioral deficits. *Nat. Aging* **1**, 666–676 (2021).
89. Gheorghie, C. E. et al. Investigating causality with fecal microbiota transplantation in rodents: applications, recommendations and pitfalls. *Gut Microbes* **13**, 1941711 (2021).
90. Ferretti, P. et al. Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. *Cell Host Microbe* **24**, 133–145 (2018).
91. Podlesny, D. et al. Metagenomic strain detection with samestr: identification of a persisting core gut microbiota transferable by fecal transplantation. *Microbiome* **10**, 53 (2022).
92. Valles-Colomer, M. et al. Variation and transmission of the human gut microbiota across multiple familial generations. *Nat. Microbiol.* **7**, 87–96 (2022).
93. Valles-Colomer, M. et al. The person-to-person transmission landscape of the gut and oral microbiomes. *Nature* **614**, 125–135 (2023).
94. Johnson, J. S. et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **10**, 5029 (2019).
95. Blanco-Míguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlan 4. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-023-01688-w> (2023).
96. Ponsonby, A.-L. Reflection on modern methods: building causal evidence within high-dimensional molecular epidemiological studies of moderate size. *Int. J. Epidemiol.* **50**, 1016–1029 (2021).
97. VanderWeele, T. J., Hernán, M. A. & Robins, J. M. Causal directed acyclic graphs and the direction of unmeasured confounding bias. *Epidemiology* **19**, 720–728 (2008).
98. VanderWeele, T. J. Principles of confounder selection. *Eur. J. Epidemiol.* **34**, 211–219 (2019).
99. McLaren, M. R., Willis, A. D. & Callahan, B. J. Consistent and correctable bias in metagenomic sequencing experiments. *Elife* **8**, e46923 (2019).
100. Wang, Y. & LêCao, K.-A. Managing batch effects in microbiome data. *Brief Bioinform.* **21**, 1954–1970 (2020).
101. Schisterman, E. F., Cole, S. R. & Platt, R. W. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* **20**, 488–495 (2009).

102. Textor, J., van der Zander, B., Gilthorpe, M. S., Liškiewicz, M. & Ellison, G. T. Robust causal inference using directed acyclic graphs: the r package 'dagitty'. *Int. J. Epidemiol.* **45**, 1887–1894 (2016).
103. Bross, I. D. Spurious effects from an extraneous variable. *J. Chronic Dis.* **19**, 637–647 (1966).
104. Hernán, M. Causal diagrams: draw your assumptions before your conclusions; <https://pll.harvard.edu/course/causal-diagrams-draw-your-assumptions-your-conclusions> (Harvard Univ., 2023).
105. Hernán, M. A. & Robins, J. A. *Causal Inference: What If* (Chapman & Hall/CRC, 2020).
106. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).

## Acknowledgments

We thank A.-L. Ponsonby for her expert comments on DAGs, D. L. Dahly for his insights on statistical analysis, and J. F. Cryan for his excellent advice. We are grateful for their help and support. APC Microbiome Ireland is a research center funded by Science Foundation Ireland (SFI), through the Irish Governments' national development plan (grant no. 12/RC/2273\_P2).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s44220-023-00149-2>.

**Correspondence** should be addressed to Thomaz F. S. Bastiaanssen.

**Peer review information** *Nature Mental Health* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature America, Inc. 2023