multimedia: Multimodal Mediation Analysis of Microbiome Data

Hanying Jiang[1,†], Xinran Miao[1,†], Margaret W. Thairu[2], Mara Beebe[2], Dan W. Grupe[3], Richard J. Davidson[3,4,5], Jo Handelsman[2,6], Kris Sankaran[1,2]

[1]Statistics Department, UW-Madison, Madison, WI, USA

[2]Wisconsin Institute for Discovery, UW-Madison, Madison, WI, USA

[3]Center for Healthy Minds, UW-Madison, Madison, WI, USA

[4]Psychology Department, UW-Madison, Madison, WI, USA

[5]Psychiatry Department, UW-Madison, Madison, WI, USA

[6]Plant Pathology Department, UW-Madison, Madison, WI, USA

[*]Address correspondence to Kris Sankaran, ksankaran@wisc.edu.

† Equal contribution.

ABSTRACT

Mediation analysis has emerged as a versatile tool for answering mechanistic questions in microbiome research because it provides a statistical framework for attributing treatment effects to alternative causal pathways. Using a series of linked regression models, this analysis quantifies how complementary data modalities relate to one another and respond to treatments. Despite these advances, the rigid modeling assumptions of existing software often results in users viewing mediation analysis as a black box, not something that can be inspected, critiqued, and refined. We designed the multimedia R package to make advanced mediation analysis techniques accessible to a wide audience,

ensuring that all statistical components are easily interpretable and adaptable to specific problem contexts. The package provides a uniform interface to direct and indirect effect estimation, synthetic null hypothesis testing, and bootstrap confidence interval construction. We illustrate the package through two case studies. The first re-analyzes a study of the microbiome and metabolome of Inflammatory Bowel Disease patients, uncovering potential mechanistic interactions between the microbiome and disease-associated metabolites, not found in the original study. The second analyzes new data about the influence of mindfulness practice on the microbiome. The mediation analysis identifies a direct effect between a randomized mindfulness intervention and microbiome composition, highlighting shifts in taxa previously associated with depression that cannot be explained by diet or sleep behaviors alone. A gallery of examples and further documentation can be found at https://go.wisc.edu/830110.

IMPORTANCE

Microbiome studies routinely gather complementary data to capture different aspects of a microbiome's response to a change, such as the introduction of a therapeutic. Mediation analysis clarifies the extent to which responses occur sequentially via mediators, thereby supporting causal, rather than purely descriptive, interpretation. multimedia is a modular R package with close ties to the wider microbiome software ecosystem that makes statistically rigorous, flexible mediation analysis easily accessible, setting the stage for precise and causally informed microbiome engineering.

INTRODUCTION

Treatments often cause change indirectly, triggering a chain of effects that eventually influences outcomes of interest. A standard approach to disentangling these pathways is to distinguish between indirect paths through candidate mediators and direct paths from treatment to outcome. Fig. 1A represents this graphically, with separate paths for treatment $T \to$ mediator $M \to$ outcome Y and treatment $T \to$ outcome Y. In the causal inference literature, this exercise is called mediation analysis, and various techniques have

emerged to support it [29, 8]. Several adaptations have been proposed for the microbiome setting, where mediators, outcomes, and controls may be high-dimensional [38, 48, 5, 23]. These efforts have already uncovered clinically relevant relationships, like the existence of microbial taxa that mediate the success of chemotherapy treatments [41].

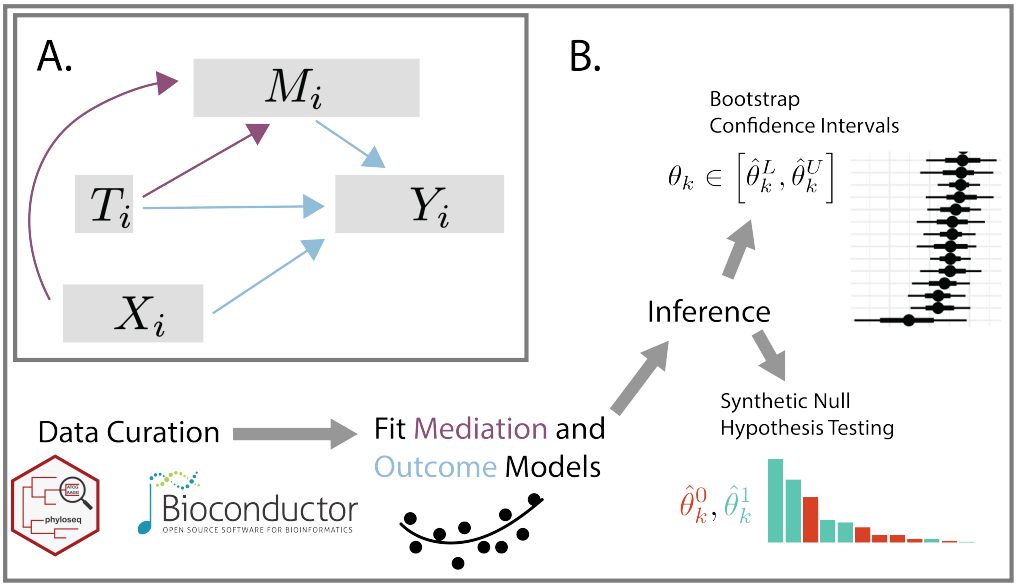


FIG 1 A. The graphical model underlying mediation analysis. Using combined mediation (purple) and outcome (blue) models, mediation analysis makes it possible to distinguish between direct and indirect causal pathways between treatments and outcomes. The conventional mediation analysis typically requires all nodes except for the covariates X to be univariate, whereas our package operates without such constraints. B. The overall multimedia workflow. Multimedia defines a modular interface to mediation analysis with utilities for summarizing and evaluating uncertainty in estimated effects.

Despite these successes, existing methodology places strong requirements on the distribution of the mediators or outcome variables and the functional form of their relationships. For example, [38, 56, 48, 55] assume that mediators are compositional and that outcomes are univariate, focusing on how microbiome relative abundance profiles mediate treatment effects on downstream host phenotypes, like the relationship between fat intake and body mass index [38]. This precludes analysis where outcomes are multidimensional, like metabolic profiles, or where mediators are clinical measurements. Further, with the exception of the mediation package [44], existing implementations are

not modular, fixing the estimator used in both the mediation and outcome regressions. This rigidity limits the range of settings in which mediation analysis can be applied. Moreover, it discourages critical evaluation or interactive model building, since model components are difficult (or impossible) to interchange. Unfortunately, even the adaptable mediation package is limited to one-dimensional mediator and outcome variables.

To enable more flexible and transparent mediation analysis of microbiome data, we extend the methodology of [24, 44] to high-dimensional mediator and outcome variables. This makes it possible to include sparse regression, logistic-normal multinomial, random forest, and hierarchical bayesian mediation and outcome models within a uniform package interface. Moreover, we have documented the process of inserting custom models into the overall workflow. These models can all be specified using R's formula notation, and components can be easily interchanged according to context. We include operations for summarization, alteration, and uncertainty quantification for the resulting models, encouraging interactive and critical microbiome mediation analysis. We ensure strong ties to the wider microbiome software ecosystem by including methods to convert to and from phyloseq [30] and SummarizedExperiment [18, 26] data structures. Altogether, the multimedia package unlocks the potential for mediation analysis for microbiome studies with complex experimental designs, enabling model-based integration of diverse data types, including microbial community composition, high-throughput molecular profiles, and host health surveys.

RESULTS

Mediation analysis with our package is a three-step process. First, users specify the hypothesized causal relationships between variables with a concise syntax that represents diverse modeling choices (**Model Setup**). Next, they estimate the model parameters and the associated causal effects (**Counterfactual Analysis**). Finally, they can compare synthetic data from alternative models and calibrate inferences using either bootstrap confidence intervals or hypothesis tests (**Evaluating Uncertainty**). This overall workflow is illustrated in Fig. **1**B and detailed in the first three sections below. A summary of key

package functions is given in Table 1. The last two sections demonstrate the package workflow with case studies on metabolomic data integration and the gut-brain axis.

Stage	Function	Description
Model Setup	mediation_data	Convert phyloseq, SummarizedExperiment, or data.frame objects into S4 classes representing all components of a mediation analysis study.
	multimedia	Define the form of the mediation and outcome models for estimation and effect calculations.
Counterfactual Analysis	direct_effect	Given fitted models, estimate direct effects for each outcome using Equation 7.
	indirect_overall	Given fitted models, estimate aggregate indirect effects for each outcome using Equation 8.
	indirect_pathwise	Given fitted models, estimate indirect effects for each mediator-outcome pair using Equation 9.
Statistical Inference	bootstrap	Re-estimate models and effects on bootstrap resampled versions of the experiment.
	nullify	Define a version of an existing model with a subset of edges removed from either the mediation or outcome model.
	fdr_summary	Calibrate a false discovery rate controlling selection rule using synthetic null data and Equation 11.

TABLE 1 Core functions for problem specification, effect estimation, and uncertainty quantification available through the multimedia package.

Model Setup To estimate a mediation model, it is necessary to fully specify the nodes and edges in Fig. 1A. The nodes are used to divide data sources into categories according to their role in the causal model. Edges correspond to mediation and outcome models. Rather than requiring specification of all mediation analysis components at once in a single function, multimedia allows users to define separate components and then glue them together to define an overall analysis. The package exports a mediation_data data structure for storing the samples used in model fitting. We use an R's S4 system [50] to define separate slots for each node in Fig. 1A. This data structure can be created by applying the accompanying mediation_data function to accompanying R data.frame,

phyloseq, and SummarizedExperiment objects. We support tidyverse-style syntax [51], meaning that many variables can be assigned to a node using concise queries. For example, mediation = starts_with("diet") will search the input data for any features starting with the string "diet" and will tag them as mediators in the downstream analysis. This efficient matching simplifies data manipulation in high-dimensional settings, where the user may need to work with hundreds of mediators or outcomes.

Next, we must specify the mediation and outcome models. The package exports wrappers to several regression families, ensuring that, despite their differing underlying methodology, all families can be used interchangeably for estimation, sampling, and prediction in the overall mediation analysis workflow. Specifically, multimedia includes (1) linear regression, which ensures that the package generalizes the earlier mediation package, (2) ℓ^1 and ℓ^2-regularized linear regression [17, 43], which can be more stable and interpretable in the presence of numerous predictors, (3) random forests [53], which supports detection of nonlinear relationships between variables, and (4) hierarchical Bayesian regression [3], which can be useful for sharing information across related groups.

Counterfactual Analysis After using the estimate function to fit models to the observed data, we can reason about potential outcomes under different treatment regimes. This allows us to clarify the relative importance of direct and indirect pathways. For example, to estimate a direct effect ($T \to Y$), we can block effects that travel along the indirect path ($T \to M \to Y$) and measuring the changes to the response that persist. Formally, in the counterfactual language of the Materials and Methods, direct and indirect effects are estimated using predicted mediators $\hat{M}(t_i)$ and outcomes $\hat{Y}\left(t_i', \hat{M}(t_i)\right)$, where t_i and t_i' correspond to mediator and outcome-specific treatment assignments. To this end, multimedia defines a data structure for storing $\left(t_i, t_i'\right)$ within two data.frames whose rows are samples and columns are treatment settings. The predict and sample methods allow users to compute expected values and draw samples according to arbitrary treatment profiles $\left(t_i, t_i'\right)$. Note that, in addition to the standard treatment vs. control setup, multimedia supports treatment profiles with multiple concurrent treatments and multilevel or continuous treatment.

Given a fitted model, multimedia outputs estimated direct and indirect effects. We formally define these effects in Equations 8 - 10. Here, we offer an overview of their motivation and interpretation. Direct effects are the changes we would observe in the outcome if we changed the treatment node in Fig. **1**A but held all the mediators fixed. This is the effect that travels along the edge $T \to Y$, and it measures the extent to which the treatment can influence the outcome while bypassing the mediators. We estimate different direct effects for each outcome. For example, in the mindfulness case study below, direct effects can be interpreted as microbiome shifts (changes in Y) following the mindfulness training (treatment T) that are not a consequence of changes in participant sleep or diet behaviors (mediators M). Next, we support estimation of two types of indirect effects. Overall indirect effects measure the changes in the outcome when setting all mediators to their predicted values when the treatment is present, keeping the contribution of the direct path $T \to Y$ fixed. This aggregates the effect across the full collection of indirect paths. In contrast, pathwise indirect effects measure the changes in outcome when comparing counterfactuals that are equal except at a single mediator. This isolates the indirect effect along a single indirect path. In this case, an indirect effect is reported for each outcome-mediator pair, rather than only for each outcome.

To increase modeling transparency, multimedia includes functions for interacting with and altering fitted models. Direct and indirect effects can be visualized within context of the original data. This can serve as a sanity check and guide further model refinements. Outputs are created with ggplot2 [49], which allows users to customize plot appearance. The case studies include outputs from these helper visualization functions. Further, given a fitted model, we allow users to refit new versions with sets of edges removed. Fig. **2** illustrates the main idea with a toy dataset. In the second column, the mediator takes on a larger value under the red treatment, while in the third, the mediators have identical distributions under the two treatments. Similarly, in the fourth, the relationship between the mediator and outcome no longer depends on treatment status. We can also alter the overall model structure, like the switch to a linear outcome model in the last column. If the model quality deteriorates significantly in a altered submodel, then those edges play a critical role. This heuristic is formalized in the synthetic null hypothesis testing strategy

discussed below. Finally, we have built the package with extensibility in mind. If functions can be written for estimation and prediction from a new model type, then it can be passed in to multimedia as a custom mediation or outcome model.

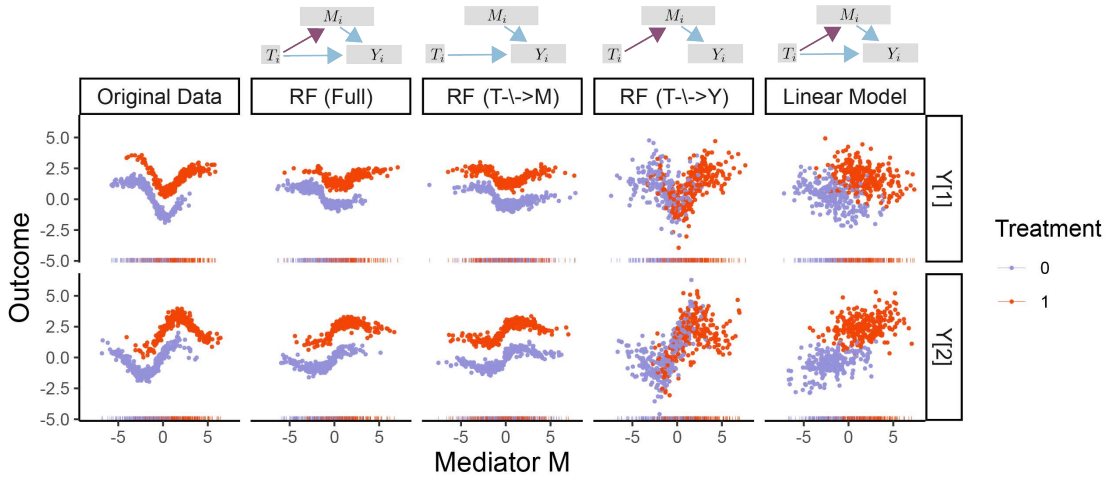


FIG 2 Samples from altered versions of a mediation analysis model fitted to the toy data at the far left. Each row describes a different outcome variable, and colors represent different treatments. The first column gives the original data, and the remaining columns give simulated data from alternative models specified by the DAGs on the top and column titles.

Statistical Inference The multimedia package offers bootstrap [13, 14, 15] and synthetic null hypothesis testing [27, 40, 39] approaches for quantifying uncertainty in estimates of mediation effects. To bootstrap in the mediation analysis context, we refit the mediation and outcome models to bootstrap resampled versions of the data and compute summary statistics (e.g., direct effect estimates) on each bootstrap sample. The percentiles of the resulting summary statistic distribution defines the bootstrap confidence interval. Importantly, the bootstrap is model agnostic and can apply to any instantiation of the counterfactual mediation analysis framework. The primary assumption made by the bootstrap is that its test statistics vary smoothly to small perturbations of the data. For this reason, it is worthwhile to check that the histogram associated with the full bootstrap distribution is well-behaved before computing confidence intervals. Like the `boot` function in base R, multimedia's `bootstrap` uses a functional implementation – any function that transforms an experiment and fitted model into a summary statistic can be

used as input. For example, it can accept a list of direct and indirect effect estimators, and these will be computed on bootstrap resample.

An alternative approach to inference in high-dimensions is based on synthetic null hypothesis testing. In this approach, rather than resampling the original data, the modeler simulates synthetic data from an assumed null distribution. Effect estimates are computed using both the original and the synthetic null data, and the fraction of synthetic null "negative controls" among the strongest observed effects can be used to calibrate a selection rule with false discovery rate control. The alteration functions above can be used to define synthetic nulls; e.g., after zeroing out the edges from either $T \to M$ or $M \to Y$, any estimated indirect effects can be treated as negative controls. Two advantages of the synthetic null approach are that (1) it only requires the mediation and outcome models be estimated twice and (2) multiple hypothesis testing is accounted for via the false discovery rate. The key disadvantage of this approach, relative to the bootstrap, is that it requires a realistic synthetic null data generating mechanism. For example, if the synthetic null data are generated from a linear model, but real effects are nonlinear, then the resulting selection sets will not provide valid false discovery rate control.

Microbiome-Metabolome Integration We next illustrate the multimedia workflow with case studies. Our first concerns Inflammatory Bowel Disease (IBD), which is closely tied to gut microbiome community composition [31]. [16] investigated the relationship between the gut microbiome and metabolome between IBD patients and healthy controls, concluding that microbial community members may be partly responsible for the formation of metabolites that lead to inflammation and IBD. By applying clustering and canonical correlation analysis to untargeted mass spectrometry data, they flagged a number of disease-relevant metabolites. We re-analyze the data using model-based mediation analysis, viewing IBD status – Healthy Control, Ulcerative Colitis (UC), Crohn's Disease (CD) – as treatments T, metabolic profile as the outcome Y, and microbiome community composition as a mediator M. The data are downloaded from the microbiome-metabolome curated data repository [32]. We have further filtered to the top 173 and 155 most abundant microbes and metabolites, and we apply centered log-ratio

(CLR) and $\log(1+x)$ to each source, respectively. Further details about the experimental cohort and data preparation are available in the Materials and Methods.

We use parallel linear and ℓ^1-regularized regression for mediation and outcome models, respectively. Note that treatment is the only predictor in the mediation model, which is why no regularization is required. We ran the bootstrap for 1000 iterations, and 95% confidence intervals and bootstrap distributions for the features with the strongest direct and overall indirect effects contrasting CD with healthy controls are shown in Fig. 3. Metabolites with strong indirect effects are influenced by IBD-induced changes in microbiome community composition, while those with large direct effects change due to other unknown factors. Fig. 4 contextualizes a small subset of these overall effects by overlaying metabolite abundances onto multidimensional scaling (MDS) plots derived from microbiome community profiles. Though metabolites with strong direct effects have differential abundance across IBD and healthy groups, only metabolites with indirect effects show variation that is also associated with microbiome composition.

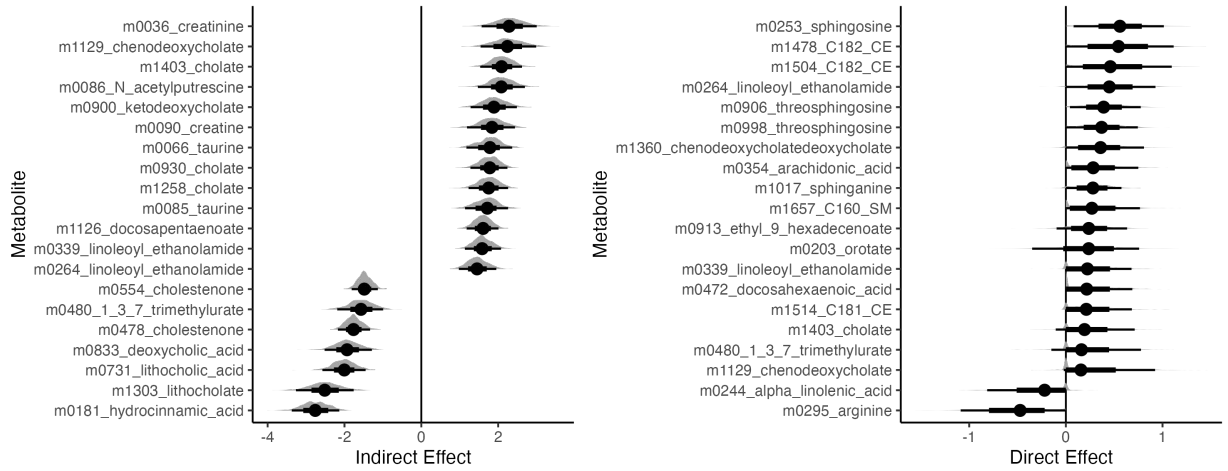


FIG 3 95% Bootstrap confidence intervals for metabolites with the strongest estimated direct and overall indirect effects associated with CD. Effects are sorted according to magnitude, and only the top 15 of each type are shown. Within the interval, the inner rectangle captures 66% of the bootstrap samples. In this data, indirect effects are stronger than direct effects.

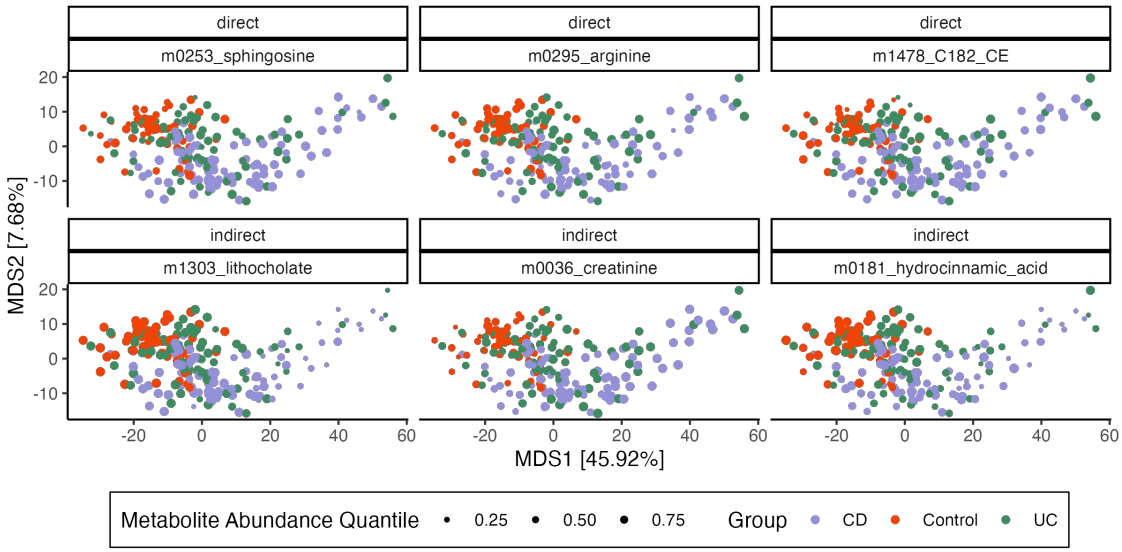


FIG 4 Microbiome composition and metabolite abundance for three metabolites with the strongest direct (top row) and indirect (bottom row) effects. Samples (points) are arranged according to an MDS on CLR transformed microbiome profiles with Euclidean Distance. Axis titles give $\frac{\lambda_k}{\sum_{k'} \lambda_{k'}}$ from the associated eigenvalues. Each panel corresponds to a metabolite, and point size encodes metabolite abundance, normalized to panel-specific quantiles. Metabolites with strong indirect effects vary more systematically with microbiome composition – for example, samples with low abundance of lithocholate are localized to the right of the MDS plot.

Moreover, by analyzing pathwise indirect effects, we can uncover genus-level relationships. A subset of the strongest pathwise indirect effects are shown in Fig. 5. Among the microbe-metabolite pairs with the strongest pathwise indirect effects, we find a relationship between metabolites of taurine and *Bilophila* (Fig. 5). High levels of fecal taurine, one of the primary conjugates of primary bile acids [52], has been previously associated with IBD [25, 46]. It has also been found that *Bilophila wadsworthia*, one of the most prominent taurine metabolizers, is often associated with lower levels of taurine [46]. Here, our results suggest that higher levels of taurine in IBD patients is mediated in part, by the abundance of *Bilophila*. We also find microbes in the genus *Firmicutes* bacterium CAG:103, are paired with several metabolites: cholate, chenodeoxycholate, and 7-ketodeoycholate (Fig. 5). Cholate and chenodeoxycholate are primary bile acids

produced by the host, which are the metabolized by gut bacteria to form secondary bile acids. 7α-dehydroxylation, is one of the pathways that bacteria metabolize primary bile acids, an intermediate of which is 7-ketodeoycholate [36]. Recent work has found that bacteria closely related to *Firmicutes* bacterium CAG:103 contain the majority of predicted genes associated with the 7α-dehydroxylation pathway within metagenomic samples [45]. Our results suggest that the increasing abundance of *Firmicutes* bacterium CAG:103, may be driving to the decrease in these primary bile acid metabolites and intermediates, which is associated more with the non-IBD controls [42]. Host deficiency in creatine uptake has been associated with poor mucosal health in IBD patients [10]. In our results we find that there is a strong microbe-metabolite pair between microbes in the genus *Choladousia* (family: *Lachnospiraceae*) and creatine/creatinine levels. *Lachnospiraceae*, (which is often at lower levels in IBD patients), are known to produce short chain fatty acids, that have been shown to help with mucosal health [33] (Fig. **5**). Overall, these results suggest that *Choladousia* may utilize creatine/creatinine as a nitrogen source, thus explaining its higher abundance in the controls.

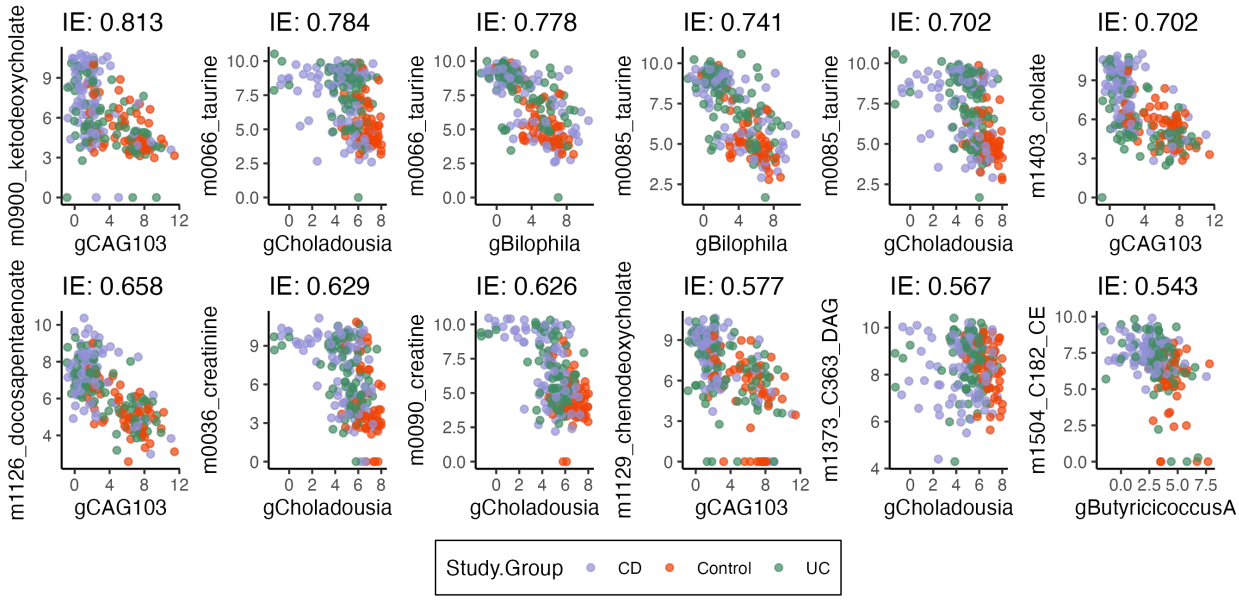


FIG 5 Microbe-metabolite pairs with the strongest pathwise indirect effects from IBD status. Each panel corresponds to one pair, CLR-transformed genus abundance is given on the x-axis, and $\log(1+x)$-transformed metabolite abundance is given on the y-axis. Effects are sorted from most negative (top left)

to most positive (bottom right). For a pathwise indirect effect to be strong, there must be both a shift in microbe abundance due to IBD state ($T \to M$) and also an association between microbe and metabolite abundance ($M \to Y$).

Note that, since this mediation model is built from a regularized linear regression outcome model, it is more sensitive to linear associations between microbe and metabolite abundances. The official package documentation includes an alternative bayesian hurdle outcome model, which exhibits higher sensitivity to outcomes with changes in metabolite presence-absence probability. The easy interchangeability of mediation analysis components makes this contrasting analysis simple to implement — it only requires change in a single line of code — and reflects multimedia's modular design.

Evaluating a Mindfulness Intervention Studies of the gut-brain axis have yielded experimental evidence for interactions between the gut microbiome and the brain. For example, germ-free mice colonized with the microbiota from human patients with clinical depression develop depression-like symptoms [28, 11], and observational studies have linked particular bacterial taxa to depression [2, 35]. Given this growing body of evidence, a team from the UW-Madison Center for Healthy Minds and the Wisconsin Institute for Discovery profiled microbiome composition, surveyed psychological symptoms, and tracked behavior change among 54 subjects before and after participation in a two-month mindfulness training [7, 20] – see the Methods and Materials for details of the study design and data processing. This study aimed to determine the nature of the mindfulness-microbiome relationship and to identify potential causal pathways. Such understanding could lead to novel interventions that influence mood through the microbiome. As a first step, we use mediation analysis to understand the mechanisms linking mindfulness and the microbiome in this randomized controlled trial. Our intervention T is the mindfulness training program, the outcome of interest is microbiome composition Y, and mediators M are survey responses related to diet and sleep that are hypothesized to influence the microbiome. To control for subject-to-subject level variation, participant ID is used as a pretreatment variable X.

For mediation and outcome models, we apply ridge and logistic-normal multinomial regressions, respectively [22, 54]. We choose a ridge regression model so that intercepts across the large number of participants are shrunk towards their global mean. We choose logistic-normal multinomial regression to jointly model microbiome composition. We also define altered submodels where all direct and indirect effects have been removed. Simulated genera compositions from all models are shown in Fig. 6. In the newly simulated data, subjects have been randomly re-assigned to the treatment and control groups. These submodels can support synthetic null hypothesis testing, since the synthetic null data appear to capture relevant properties of the real microbiome composition profiles, like the average relative abundances across genera and the range of observed abundances within most genera. Their main limitation is that some genera, like *Methanobrevibacter*, *Paraprevotella*, and *Akkermansia*, have much wider ranges than the synthetic data, and Fig. S1 suggests that this is due to a failure to capture the unusually high zero inflation present in these genera.

For synthetic null hypothesis testing, models without $T \to Y$ and $M \to Y$ associations are used to generate negative controls for direct and overall indirect effect estimates, respectively. Fig. 7 shows the estimated effects from real and synthetic data, together with the estimated false discovery rates. At a level $q = 0.15$, five genera are selected as having either significant direct or indirect effects. Fig. S2 provides the analog of Fig. 5 for this case study. Indirect effects are an order of magnitude weaker than direct effects, suggesting that changes in microbiome composition following the mindfulness intervention cannot simply be attributed to changes in diet or sleep alone.

We cannot externally validate these findings, since there is no consensus on the relationship between specific taxonomic groups and common psychiatric disorders (for a description of current sources of controversy, see [1]). However, our findings are broadly consistent with those from a recent large-scale human cohort, which found that most genera belonging to the families *Ruminococcaceae* were depleted in people with more symptoms of depression and that *Bifidobacterium* was an important predictor of depressive symptoms in a random forest classifier [2].

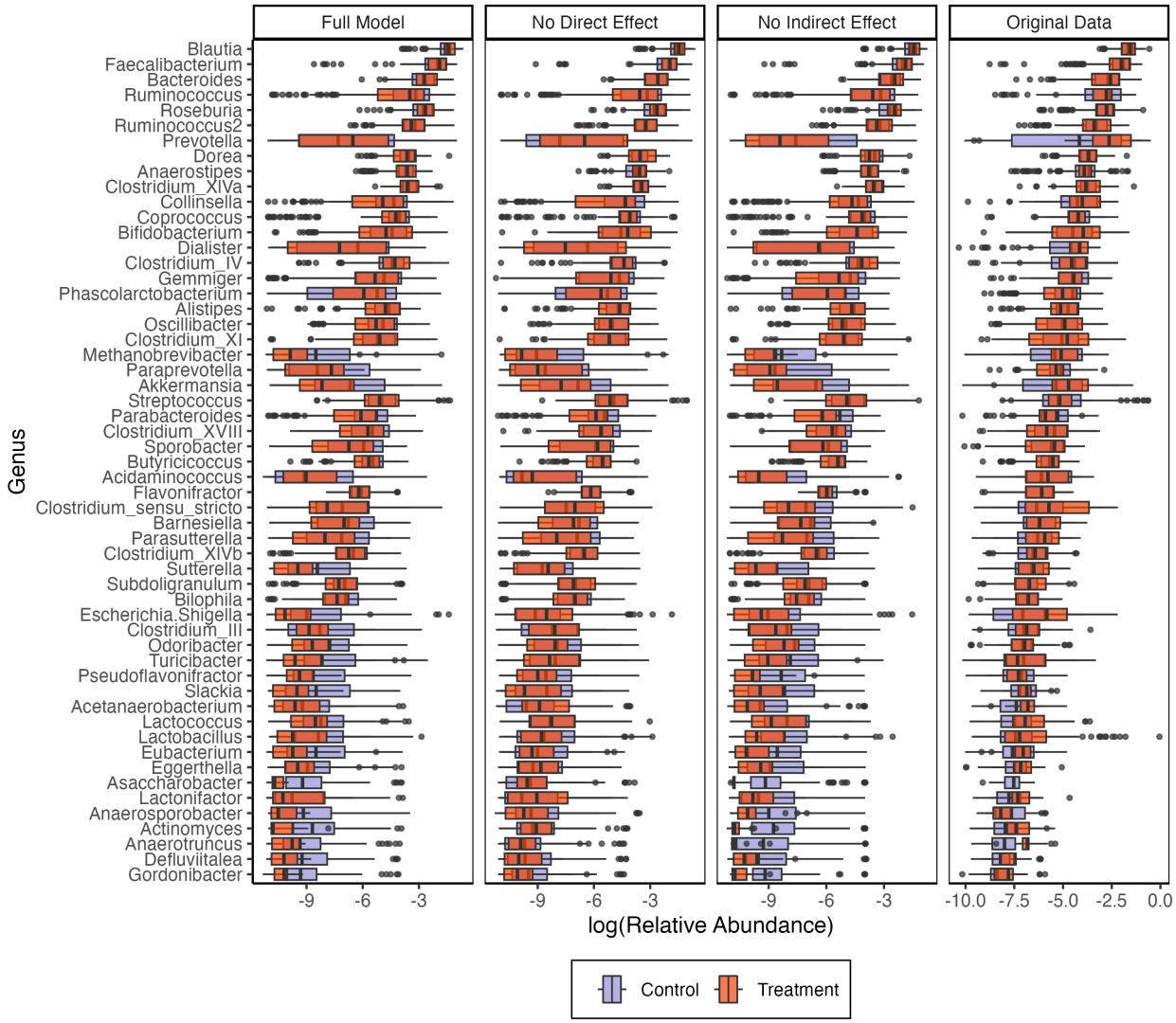


FIG 6 Real and synthetic null relative abundances across a subset of genera at different overall relative abundances. Color distinguishes whether the participant belonged to the treatment (mindfulness training) or control groups. The full model (left panel) captures the overall abundances and trajectories present in the real data, though it tends to underestimate the heaviness of the tails. The second and third panels show the analogous models with direct ($T \to Y$) and indirect ($M \to Y$) effects removed.

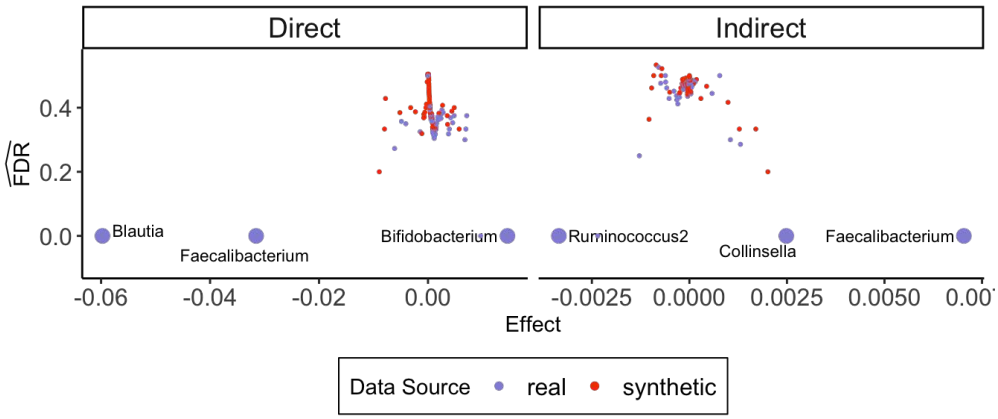


FIG 7 Estimated overall effects and false discovery rates derived from real and synthetic null data. Each point corresponds to one genus in either real (blue) or simulated (orange) data. The genera selected to control the false discovery rate at $q \leq 0.15$ are drawn larger than the rest. Direct effects are both larger in magnitude and easier to distinguish than their indirect counterparts.

DISCUSSION

Mediation analysis makes it possible to study causal pathways in multimodal microbiome data, and it is an essential tool for discovery of subtle relationships that span multiple host measurements and high-throughput assays. Statistical techniques in this space are needed to support interrogation of varied causal relationships, not simply studies where microbiome profiles serve as mediators and outcomes are one-dimensional, as has been the historical focus of the field.

Our case studies illustrate the flexibility and analytical depth supported by multimedia. Unlike traditional microbiome mediation analysis software, the package allows specification of diverse regression components, and the interface simplifies interpretation of effect types and model criticism. In this way, multimedia encourages interactive, rigorous mediation analysis for microbiome data. It is written to interface closely with the existing microbiome software ecosystem, and since analysis are carried out in reproducible code notebooks, it supports scientific transparency.

We have created a gallery of example notebooks that use the multimedia package. These include alternative analyses of the IBD and mindfulness data explored here. We invite users to contribute further examples, and we plan to structure further developments according to community needs.

MATERIALS AND METHODS

Counterfactual framework Consider random samples indexed by i. Let $T_i \in \mathbb{R}$ be the experimental treatment of interest, $Y_i \in \mathbb{R}$ be the outcome, and $X_i \in \mathbb{R}^P$ be the pretreatment covariates. For simplicity, we assume T_i is a binary indicator of either treatment ($T_i = 1$) or control ($T_i = 0$), though multimedia supports categorical, continuous, and multi-treatment cases.

To what extent is the effect of the treatment on the outcome modulated by intermediate variables? Mediation analysis answers this question by positing mediators M_i on the causal path from T_i to Y_i. Adopting a counterfactual perspective, we define $M_i(t')$ as the potential outcome of the mediator under $T_i = t'$ and $Y_i(t, m)$ as the potential outcome of the response under $T_i = t$ and $M_i = m$ with $t, t' \in \{0, 1\}$. Therefore, we can express the outcome variable as $Y_i(t, M_i(t'))$. In a randomized experiment, we can only ever observe the case where t and t' are the same, i.e., $Y_i(1, M_i(1))$ in the treated group and $Y_i(0, M_i(0))$ in the control group – but conceptually t and t' can be different. For example, $Y_i(0, M_i(1))$ represents the potential outcome when only mediators are intervened upon and $Y_i(1, M_i(0))$ represents the potential outcome when we make interventions while keeping mediators at their values under the control.

Analogous to the traditional average treatment effect, [24] defines the indirect effect to be the treatment effect obtained through mediators,

$$\delta(t) = \mathbb{E}\{Y_i(t, M_i(1)) - Y_i(t, M_i(0))\} \tag{1}$$

and the direct effect to be the effect of treatment through other mechanisms,

$$\zeta(t') = \mathbb{E}\{Y_i(1, M_i(t')) - Y_i(0, M_i(t'))\} \tag{2}$$

for $t, t' \in \{0, 1\}$. It has been shown that both effects are identifiable under assumptions (3) to (6):

$$\{Y_i(t', m), M_i(t)\} \perp\!\!\!\perp T_i \mid X_i = x \tag{3}$$

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) \mid T_i = t, X_i = x \tag{4}$$

$$\mathbb{P}\left(T_i = t \mid X_i = x\right) > 0 \tag{5}$$

$$p_{M_i(t)}(m \mid T_i = t, X_i = x) > 0 \tag{6}$$

for $t, t' \in \{0, 1\}$ and all $m \in \mathcal{M}$ and $x \in \mathcal{X}$, where $\mathcal{M}$ and $\mathcal{X}$ represent the supports of M_i and X_i, respectively.

In addition, under a no interaction assumption, $\delta(t) = \delta(t')$ and $\zeta(t) = \zeta(t')$ for any $t \neq t'$. We define the overall indirect effect and direct effect as follows,

$$\bar{\delta} = \frac{1}{2} \sum_{t=0}^{1} \mathbb{E}\{Y_i(t, M_i(1)) - Y_i(t, M_i(0))\} \tag{7}$$

$$\bar{\zeta} = \frac{1}{2} \sum_{t'=0}^{1} \mathbb{E}\{Y_i(1, M_i(t')) - Y_i(0, M_i(t'))\} \tag{8}$$

Large magnitudes of $\bar{\delta}$ and $\bar{\zeta}$ suggest strong effects of the treatment on the outcome via mediators and mechanisms other than mediators, respectively. To define pathwise indirect effects, we apply different treatment assignments across coordinates of the mediators $M_i = (M_{i1}, \ldots, M_{iK})$,

$$\begin{aligned} \bar{\omega}_k = \frac{1}{2} \sum_{t'=0}^{1} \mathbb{E}\{ & Y_i\left(t', \left(M_{i1}\left(t'\right), \ldots, M_{ik}\left(1\right), \ldots, M_{iK}\left(t'\right)\right)\right) - \\ & Y_i\left(t', \left(M_{i1}\left(t'\right), \ldots, M_{ik}\left(0\right), \ldots, M_{iK}\left(t'\right)\right)\right)\}. \end{aligned} \tag{9}$$

In practice, the population quantities $\bar{\delta}, \bar{\zeta}$, and $\bar{\omega}$ are unknown, and the expectations are replaced with fitted values from the mediation and outcome models $\hat{M}$ and $\hat{Y}$, respectively. For example, the direct effect is estimated using

$$\hat{\bar{\zeta}} = \frac{1}{2} \sum_{t'=0}^{1} \sum_{i=1}^{n} \hat{Y}_i(1, \hat{M}_i(t')) - \hat{Y}_i(0, \hat{M}_i(t')). \tag{10}$$

Bootstrap and synthetic null testing Form a bootstrap resample of the data $\mathcal{D}^* = (\mathbf{X}^*, \mathbf{M}^*, \mathbf{T}^*, \mathbf{Y}^*)$ by independently resampling the n observations with replacement. A summary statistic computed on the bth resampled dataset is denoted by $\hat{\theta}^{*b}(\mathcal{D}^*)$. For brevity, we will omit the data arguments. For example, $\hat{\theta}^{*b}$ could correspond to an estimator of $\bar{\delta}$ or $\bar{\zeta}$ derived from mediation and outcome models learned from $\mathcal{D}^*$. Repeating this process B times, we refit $\hat{M}(t,x)$, $\hat{Y}(t,m,x)$ and the provided summary statistic $\hat{\theta}$ for each of the B bootstrap the datasets. We can obtain the bootstrap distribution $\left(\hat{\theta}^{*b}\right)_{b=1}^{B}$. Let $q_{\frac{\alpha}{2}}$ and $q_{1-\frac{\alpha}{2}}$ represent the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ quantiles of this collection. Then $\left[q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}\right]$ forms an α-level bootstrap confidence interval associated with the statistic $\hat{\theta}$.

For synthetic null hypothesis testing, estimate mediation and outcome models $\hat{M}_{\text{sub}}(t,x)$, $\hat{Y}_{\text{sub}}(t,m,x)$ using only a subset of edges within the DAG. This defines the null data generating mechanism. Using the same pretreatment and treatment profiles X_i, T_i from the original experiment, simulate synthetic null data $\mathbf{M}^{*0}, \mathbf{Y}^{*0}$ from the submodel. For D taxa of interest, compute summary statistics $\left(\hat{\theta}_d^1\right)_{d=1}^{D}$ and $\left(\hat{\theta}_d^0\right)_{d=1}^{D}$ based on the original and the synthetic null data, respectively. For example, $\hat{\theta}_d^1$ could estimate taxon d's direct effect $\hat{\bar{\delta}}_d$ using the original data, and $\hat{\theta}_d^0$ could be the corresponding estimate derived from synthetic null data. Next, for any threshold t, we estimate the false discovery rate using

$$\widehat{\text{FDR}}(t) := \frac{\#\left\{d : \left|\hat{\theta}_d^0\right| > t\right\}}{\#\left\{d : \left|\hat{\theta}_d^0\right| > t\right\} + \#\left\{d : \left|\hat{\theta}_d^1\right| > t\right\}}. \tag{11}$$

The numerator counts the number of estimates from the synthetic null data that are larger than t, and the denominator counts the number of discoveries across either simulated or real data at that threshold. Given a desired FDR level q, the selection rule is defined by selecting $t^* = \min\left\{t : \widehat{\text{FDR}}(t) \leq q\right\}$ and selecting all features d such that $\left|\hat{\theta}_d^1\right| > t^*$. Under the null samples generated by $\hat{M}_{\text{sub}}(t,x)$, $\hat{Y}_{\text{sub}}(t,m,x)$, this rule controls the false discovery rate below level q, regardless of the choice of estimator $\hat{\theta}_d$, though better estimators lead to improved power.

Microbiome-metabolome data processing We obtained the data from the microbiome-metagenome curated database. Details of the library preparation and

bioinformatics can be found in [34]. Briefly, metagenomic sequencing was done on an Illumina HiSeq 2500, and metabolites were profiled using LC-MS in non-targeted mode. For metagenomics, fastp was applied to raw reads for quality filtering, adapter trimming, and deduplication. bowtie2 was used to remove human reads by aligning to the hg38. kraken2.1.1 and braken 2.8 were used to estimate taxonomic relative abundances.

A total of 11,720 taxa and 8,848 metabolites are present in the public data. We applied a centered log-ratio transformation to the microbiome relative abundances profiles: $\mathrm{CLR}(x_1, \ldots, x_D) := \left(\log(x_d) - \frac{1}{D}\sum_{d'} \log x_{d'}\right)_{d=1}^{D}$. We then filtered to taxa whose average transformed abundance was larger than 3, which reduced the number of taxa to 173. We kept only metabolites with confident HMDB assignments, applied a $\log(1+x)$ transformation, and further filtered to those whose average transformed intensity was larger than 6. This resulted in 155 well-annotated and generally abundant metabolites.

Mindfulness study design and processing The initial Center for Healthy Minds study recruited 114 police officers participants across two cohorts. Microbiome samples were obtained only from participants in the second cohort ($n = 54$), who were randomly assigned to mindfulness training or waitlist control with 27 cases each. We removed four participants due to incomplete responses – three lacked microbiome data, and one had missing mediators. Our analysis considers a mindfulness training treatment group of size $n = 24$ and a waitlist control group of size $n = 26$. Participants in the mindfulness group took part in an 8-week, 18-hour mindfulness training developed specifically for their career and inspired by Mindfulness-Based Stress Reduction and Mindfulness-Based Resilience Training [7]. Weekly two-hour classes (and a four-hour class in week 7) consisted of didactic instruction, embodied mindfulness practices, and individual and group-based inquiry (for full intervention details, see [21]). Microbiota and behavioral survey data were gathered at 2 - 3 timepoints for each participant — samples in the treatment group provided data before, within two weeks following, and, in a subset of cases, four months after the 8-week intervention, resulting in 118 samples total.

Gut microbiome composition was assessed using 16S rRNA gene sequencing, and participants completed surveys, as reported previously [21]. One to four technical

replicates (on average, 2.6) were sequenced for each 16S rRNA gene sample, resulting in 307 microbiome composition profiles in total. Amplicon Sequence Variants (ASV) were called using the DADA2 pipeline [4]. The first ten base pairs were removed, and all reads were truncated to a length of 250. Otherwise, we set all pipeline hyperparameters to their defaults. Since the total number of participants is relatively small, we chose to concentrate on the core microbiome [37]. To this end, we assigned taxonomic identity to each ASV using the RDP database and aggregated all counts to the genus level [9]. We removed any genera that did not appear in at least 40% of the samples, thereby generating a core microbiome. On average, this preserved 98.7% of the reads within each sample. After filtering to the core microbiome, sequences for 55 genera remained. To define mediators, we manually selected four variables from the National Cancer Institute Quick Food Scan and self-reported questionnaires on fatigue and sleep disturbance scores based on the Patient-Reported Outcomes Measurement Information System subscale [6]. We concentrated on these questions because changes in both diet and sleep have previously been associated with mindfulness interventions and the microbiome [19, 12, 47].

In detail, we consider four mediators – two diet mediators from the National Cancer Institute Quick Food Scan and two stress variables from the Patient-Reported Outcomes Measurement Information System (43-item inventory; version 2.0) following [6]. They are all calculated from questionnaires. The two diet variables indicate the frequency that participants eat cold cereal and fruit (not juices), respectively, in the past 12 months (Supplementary Table 1). The two stress variables, fatigue and sleep disturbance, profile the stress of a participant in the past 7 days (Supplementary Table 2).

ACKNOWLEDGMENTS

DATA AVAILABILITY STATEMENT

The multimedia package is available at https://go.wisc.edu/830110. Notebooks to reproduce the case studies are available at https://go.wisc.edu/787g25.

FUNDING

K.S. and J.H. were funded by award R01GM152744 from the National Institute of General Medical Sciences of the National Institutes of Health.

CONFLICTS OF INTEREST

State all conflicts of interest here or say, "The authors declare no conflict of interest."

References

[1] Thomaz FS Bastiaanssen, Sofia Cussotto, Marcus J Claesson, Gerard Clarke, Timothy G Dinan, and John F Cryan. Gutted! unraveling the role of the microbiome in major depressive disorder. *Harvard Review of Psychiatry*, 28(1):26, 2020.

[2] Jos A Bosch, Max Nieuwdorp, Aeilko H Zwinderman, Mélanie Deschasaux, Djawad Radjabzadeh, Robert Kraaij, Mark Davids, Susanne R de Rooij, and Anja Lok. The gut microbiota and depressive symptoms across ethnic groups. *Nature Communications*, 13(1):7129, 2022.

[3] Paul-Christian Bürkner. brms: An R package for bayesian multilevel models using stan. *Journal of Statistical Software*, 080:1–28, 2017.

[4] Benjamin J Callahan, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy Jo A Johnson, and Susan P Holmes. Dada2: High-resolution sample inference from illumina amplicon data. *Nature Methods*, 13(7):581–583, 2016.

[5] Kyle M. Carter, Meng Lu, Hongmei Jiang, and Lingling An. An information-based approach for mediation analysis on high-dimensional metagenomic data. *Frontiers in Genetics*, 11, 2020.

[6] David Cella, William Riley, Arthur Stone, Nan Rothrock, Bryce Reeve, Susan Yount, Dagmar Amtmann, Rita Bode, Daniel Buysse, Seung Choi, et al. The patient-reported outcomes measurement information system (promis) developed and tested its first

wave of adult self-reported health outcome item banks: 2005–2008. *Journal of clinical epidemiology*, 63(11):1179–1194, 2010.

[7] Michael S Christopher, Richard J Goerling, Brant S Rogers, Matthew Hunsinger, Greg Baron, Aaron L Bergman, and David T Zava. A pilot study evaluating the effectiveness of a mindfulness-based intervention on cortisol awakening response and health outcomes among law enforcement officers. *Journal of police and criminal psychology*, 31:15–28, 2016.

[8] Dylan Clark-Boucher, Xiang Zhou, Jiacong Du, Yongmei Liu, Belinda L Needham, Jennifer A. Smith, and Bhramar Mukherjee. Methods for mediation analysis with high-dimensional dna methylation data: Possible choices and comparisons. *PLOS Genetics*, 19, 2023.

[9] James R Cole, Qiong Wang, Jordan A Fish, Benli Chai, Donna M McGarrell, Yanni Sun, C Titus Brown, Andrea Porras-Alfaro, Cheryl R Kuske, and James M Tiedje. Ribosomal database project: data and tools for high throughput rrna analysis. *Nucleic acids research*, 42(D1):D633–D642, 2014.

[10] Sean P. Colgan, Ruth X. Wang, Caroline H.T. Hall, Geetha Bhagavatula, and J. Scott Lee. Revisiting the "starved gut" hypothesis in inflammatory bowel disease. *Immunometabolism*, 5(1):e0016, January 2023.

[11] John F Cryan and Sarkis K Mazmanian. Microbiota–brain axis: Context and causality. *Science*, 376(6596):938–939, 2022.

[12] Lawrence A David, Corinne F Maurice, Rachel N Carmody, David B Gootenberg, Julie E Button, Benjamin E Wolfe, Alisha V Ling, A Sloan Devlin, Yug Varma, Michael A Fischbach, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, 505(7484):559–563, 2014.

[13] Bradley Efron. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7:1–26, 1979.

[14] Bradley Efron. The jackknife, the bootstrap, and other resampling plans. *SIAM*, 38, 1987.

[15] Bradley Efron and Robert Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, Philadelphia, PA, May 1994.

[16] Eric A. Franzosa, Alexandra Sirota-Madi, Julián Ávila-Pacheco, Nadine Fornelos, Henry J. Haiser, Stefan Reinker, Tommi Vatanen, A. Brantley Hall, FASA Himel Mallick, PhD, Lauren J. McIver, Jenny S. Sauk, Robin G. Wilson, Betsy W. Stevens, Justin Scott, Kerry A. Pierce, Amy Anderson Deik, Kevin Bullock, Floris Imhann, Jeffrey A. Porter, Alexandra Zhernakova, Jingyuan Fu, Rinse K. Weersma, Cisca Wijmenga, Clary B. Clish, Hera Vlamakis, Curtis Huttenhower, and Ramnik J. Xavier. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature microbiology*, 4:293 – 305, 2018.

[17] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 2010.

[18] Robert Gentleman, Vincent J. Carey, Douglas M. Bates, Benjamin M. Bolstad, Marcel Dettling, Sandrine Dudoit, Byron Ellis, Laurent Gautier, Yongchao Ge, Jeff Gentry, Kurt Hornik, Torsten Hothorn, Wolfgang Huber, S. Iacus, Rafael A. Irizarry, Friedrich Leisch, Cheng Li, Martin Maechler, Anthony J. Rossini, Günther Sawitzki, Colin Smith, Gordon Smyth, Luke Tierney, Jean YH Yang, and Jianhua Zhang. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, 5:R80 – R80, 2004.

[19] Desleigh Gilbert and Jennifer Waltz. Mindfulness and health behaviors. *Mindfulness*, 1(4):227–234, 2010.

[20] Daniel W. Grupe, Jonah L. Stoller, Carmen Alonso, C. R. Mcgehee, Christion Smith, Jeanette A. Mumford, Melissa A. Rosenkranz, and Richard J. Davidson. The impact of mindfulness training on police officer stress, mental health, and salivary cortisol levels. *Frontiers in Psychology*, 12, 2021.

[21] Daniel W Grupe, Jonah L Stoller, Carmen Alonso, Chad McGehee, Chris Smith, Jeanette A Mumford, Melissa A Rosenkranz, and Richard J Davidson. The impact of

mindfulness training on police officer stress, mental health, and salivary cortisol levels. *Frontiers in Psychology*, 12:720753, 2021.

[22] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer New York, New York, NY, 2009.

[23] Qilin Hong, Guanhua Chen, and Zheng-Zheng Tang. A phylogeny-based test of mediation effect in microbiome. 2021.

[24] Kosuke Imai, Luke J. Keele, and Dustin Tingley. A general approach to causal mediation analysis. *Political Methods: Quantitative Methods eJournal*, 2010.

[25] Jasmijn Z Jagt, Eduard A Struys, Ibrahim Ayada, Abdellatif Bakkali, Erwin E W Jansen, Jürgen Claesen, Johan E van Limbergen, Marc A Benninga, Nanne K H de Boer, and Tim G J de Meij. Fecal amino acid analysis in newly diagnosed pediatric inflammatory bowel disease: A multicenter case-control study. *Inflammatory Bowel Diseases*, 28(5):755–763, November 2021.

[26] Leo Lahti, Felix G. M. Ernst, Sudarshan A. Shetty, Tuomas Borman, Ruizhu Huang, Domenick J. Braccia, and Hector Corrado Bravo. Microbiome data science in the summarizedexperiment universe. *F1000Research*, 10, 2021.

[27] Wei Vivian Li and Jingyi Jessica Li. A statistical simulator scdesign for rational scrna-seq experimental design. *Bioinformatics*, 35:i41 – i50, 2018.

[28] Yuanyuan Luo, Benhua Zeng, Li Zeng, Xiangyu Du, Bo Li, Ran Huo, Lanxiang Liu, Haiyang Wang, Meixue Dong, Junxi Pan, et al. Gut microbiota regulates mouse behaviors through glucocorticoid receptor pathway genes in the hippocampus. *Translational psychiatry*, 8(1):1–10, 2018.

[29] David P. Mackinnon, Amanda J. Fairchild, and Matthew S. Fritz. Mediation analysis. *Annual review of psychology*, 58:593–614, 2019.

[30] Paul J. McMurdie and Susan P. Holmes. phyloseq: An r package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, 8, 2013.

[31] Indrani Mukhopadhya, Richard Hansen, Emad M El-Omar, and Georgina L Hold. Ibd—what role do proteobacteria play? *Nature reviews Gastroenterology & hepatology*, 9(4):219–230, 2012.

[32] Efrat Muller, Yadid M Algavi, and Elhanan Borenstein. The gut microbiome-metabolome dataset collection: a curated resource for integrative meta-analysis. *NPJ Biofilms and Microbiomes*, 8, 2022.

[33] Daniela Parada Venegas, Marjorie K. De la Fuente, Glauben Landskron, María Julieta González, Rodrigo Quera, Gerard Dijkstra, Hermie J. M. Harmsen, Klaas Nico Faber, and Marcela A. Hermoso. Short chain fatty acids (scfas)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. *Frontiers in Immunology*, 10, March 2019.

[34] Edoardo Pasolli, Lucas Schiffer, Paolo Manghi, Audrey Renson, Valerie Obenchain, Duy Tin Truong, Francesco Beghini, Fa Malik, Marcel Ramos, Jennifer Beam Dowd, Curtis Huttenhower, Martin T. Morgan, N. Segata, and Levi Waldron. Accessible, curated metagenomic data through experimenthub. *Nature Methods*, 14:1023–1024, 2017.

[35] Djawad Radjabzadeh, Jos A Bosch, André G Uitterlinden, Aeilko H Zwinderman, M Arfan Ikram, Joyce BJ van Meurs, Annemarie I Luik, Max Nieuwdorp, Anja Lok, Cornelia M van Duijn, et al. Gut microbiome-wide association study of depressive symptoms. *Nature Communications*, 13(1):7128, 2022.

[36] Jason M. Ridlon, Steven L. Daniel, and H. Rex Gaskins. The hylemon-björkhem pathway of bile acid 7-dehydroxylation: history, biochemistry, and microbiology. *Journal of Lipid Research*, 64(8):100392, August 2023.

[37] Ashley Shade and Jo Handelsman. Beyond the venn diagram: the hunt for a core microbiome. *Environmental microbiology*, 14(1):4–12, 2012.

[38] Michael B. Sohn and Hongzhe Li. Compositional mediation analysis for microbiome studies. *The Annals of Applied Statistics*, 13(1), March 2019.

[39] Dongyuan Song, Qingyang Wang, Guanao Yan, Tianyang Liu, Tianyi Sun, and Jingyi Jessica Li. scdesign3 generates realistic in silico data for multimodal single-cell and spatial omics. *Nature Biotechnology*, 42:247–252, 2023.

[40] Tianyi Sun, Dongyuan Song, Wei Vivian Li, and Jingyi Jessica Li. scdesign2: a transparent simulator that generates high-fidelity single-cell gene expression count data with gene correlations captured. *Genome Biology*, 22, 2021.

[41] Ying Taur and Eric G. Pamer. Microbiome mediation of infections in the cancer setting. *Genome Medicine*, 8, 2016.

[42] John P. Thomas, Dezso Modos, Simon M. Rushbrook, Nick Powell, and Tamas Korcsmaros. The emerging role of bile acids in the pathogenesis of inflammatory bowel disease. *Frontiers in Immunology*, 13, February 2022.

[43] Robert Tibshirani, Jacob Bien, Jerome Friedman, Trevor Hastie, Noah Simon, Jonathan Taylor, and Ryan J. Tibshirani. Strong rules for discarding predictors in lasso-type problems. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 74(2):245–266, November 2011.

[44] Dustin Tingley, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai. mediation: R package for causal mediation analysis. *Journal of Statistical Software*, 59:1–38, 2014.

[45] Arnau Vich Vila, Shixian Hu, Sergio Andreu-Sánchez, Valerie Collij, Bernadien H Jansen, Hannah E Augustijn, Laura A Bolte, Renate A A A Ruigrok, Galeb Abu-Ali, Cosmas Giallourakis, Jessica Schneider, John Parkinson, Amal Al-Garawi, Alexandra Zhernakova, Ranko Gacesa, Jingyuan Fu, and Rinse K Weersma. Faecal metabolome and its determinants in inflammatory bowel disease. *Gut*, 72(8):1472–1485, March 2023.

[46] Marius Vital, Tatjana Rud, Silke Rath, Dietmar H. Pieper, and Dirk Schlüter. Diversity of bacteria exhibiting bile acid-inducible 7-dehydroxylation genes in the human gut. *Computational and Structural Biotechnology Journal*, 17:1016–1019, 2019.

[47] Jolana Wagner-Skacel, Nina Dalkner, Sabrina Moerkl, Kathrin Kreuzer, Aitak Farzi, Sonja Lackner, Annamaria Painold, Eva Z Reininghaus, Mary I Butler, and Susanne Bengesser. Sleep and microbiome in psychiatric diseases. *Nutrients*, 12(8):2198, 2020.

[48] Chan Wang, Jiyuan Hu, Martin J Blaser, and Huilin Li. Estimating and testing the microbial causal mediation effect with high-dimensional and compositional microbiome data. *Bioinformatics*, 36(2):347–355, July 2019.

[49] Hadley Wickham. ggplot2 - elegant graphics for data analysis. In *Use R!*, 2009.

[50] Hadley Wickham. *Advanced R*. Chapman and Hall/CRC, September 2014.

[51] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Pedersen, Evan Miller, Stephanie Bache, Kirill Müller, Jeroen Ooms, David G. Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara H. Woo, and Hiroaki Yutani. Welcome to the tidyverse. *J. Open Source Softw.*, 4:1686, 2019.

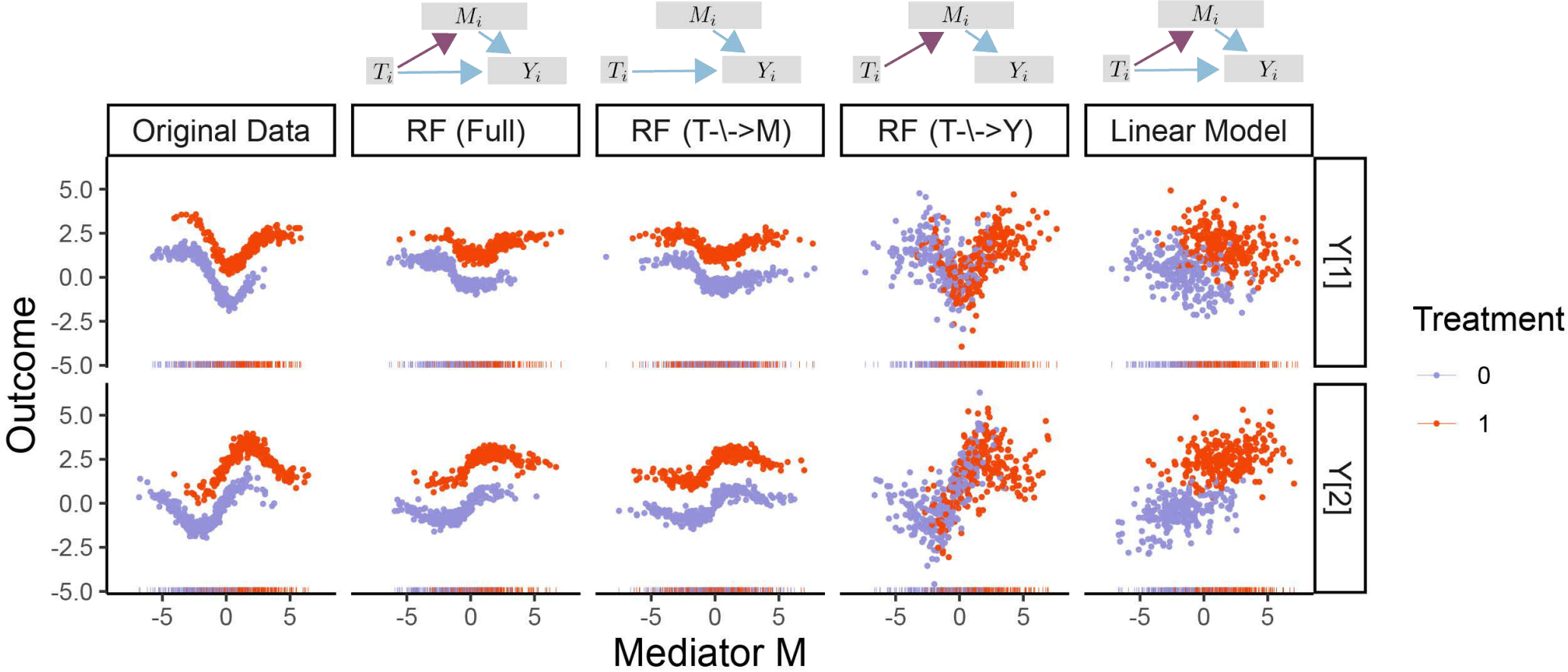
[52] Jenessa A. Winston and Casey M. Theriot. Diversification of host bile acids by members of the gut microbiota. *Gut Microbes*, 11(2):158–171, October 2019.

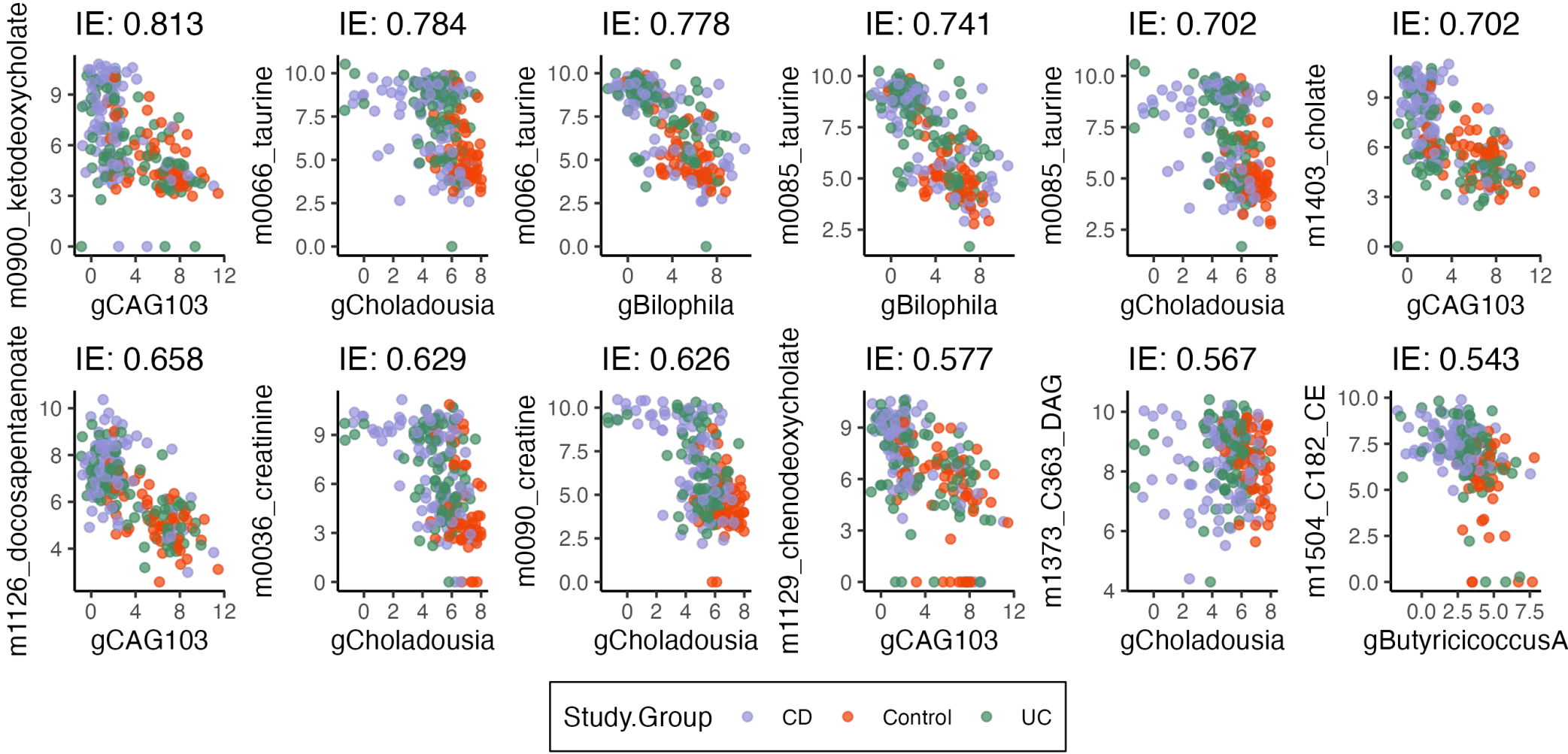
[53] Marvin N. Wright and Andreas Ziegler. ranger: A fast implementation of random forests for high dimensional data in c++ and r. *Journal of Statistical Software*, 077:1–17, 2015.

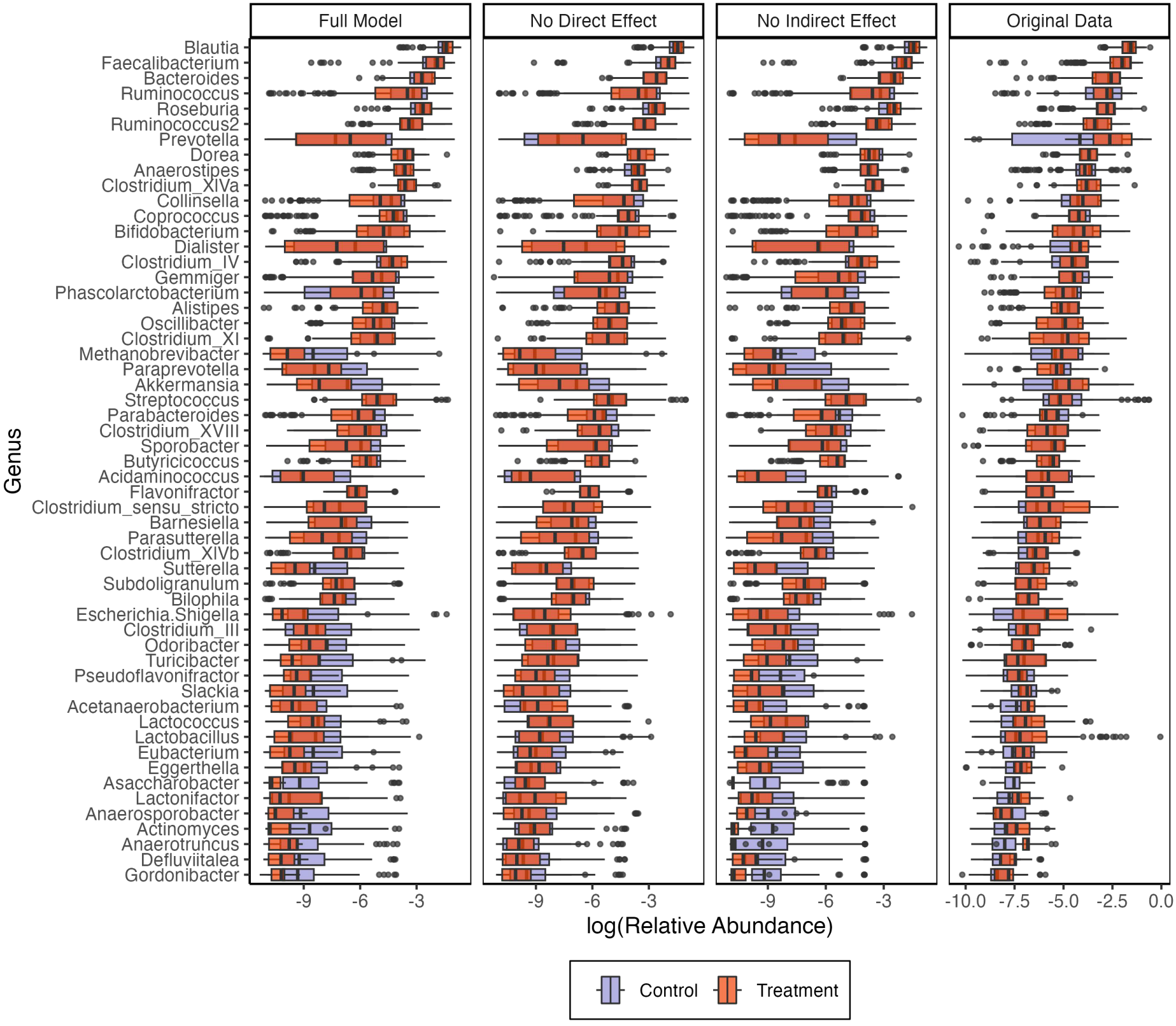
[54] Fan Xia, Jun Chen, Wing Kam Fung, and Hongzhe Li. A logistic normal multinomial regression model for microbiome compositional data analysis. *Biometrics*, 69, 2013.

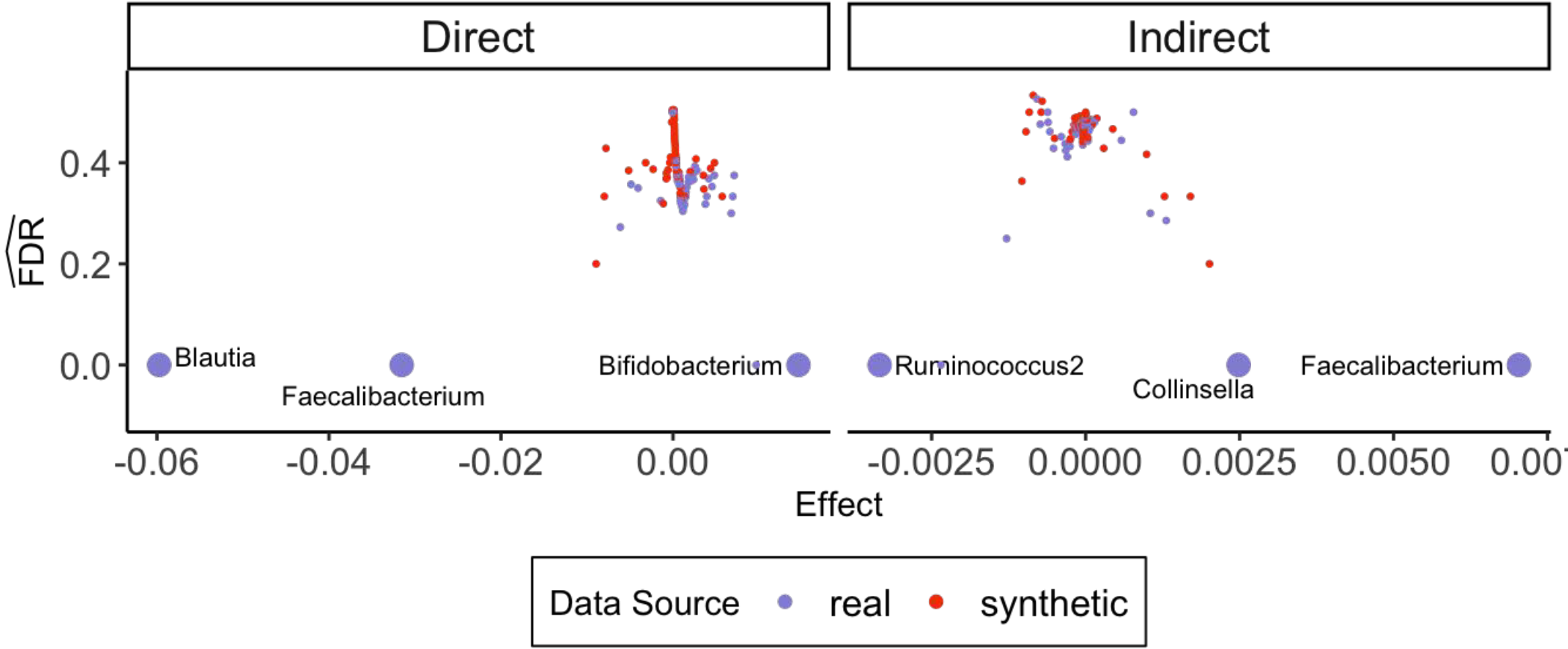
[55] Dongyang Yang and Wei Xu. Estimation of mediation effect on zero-inflated microbiome mediators. *Mathematics*, 11(13):2830, June 2023.

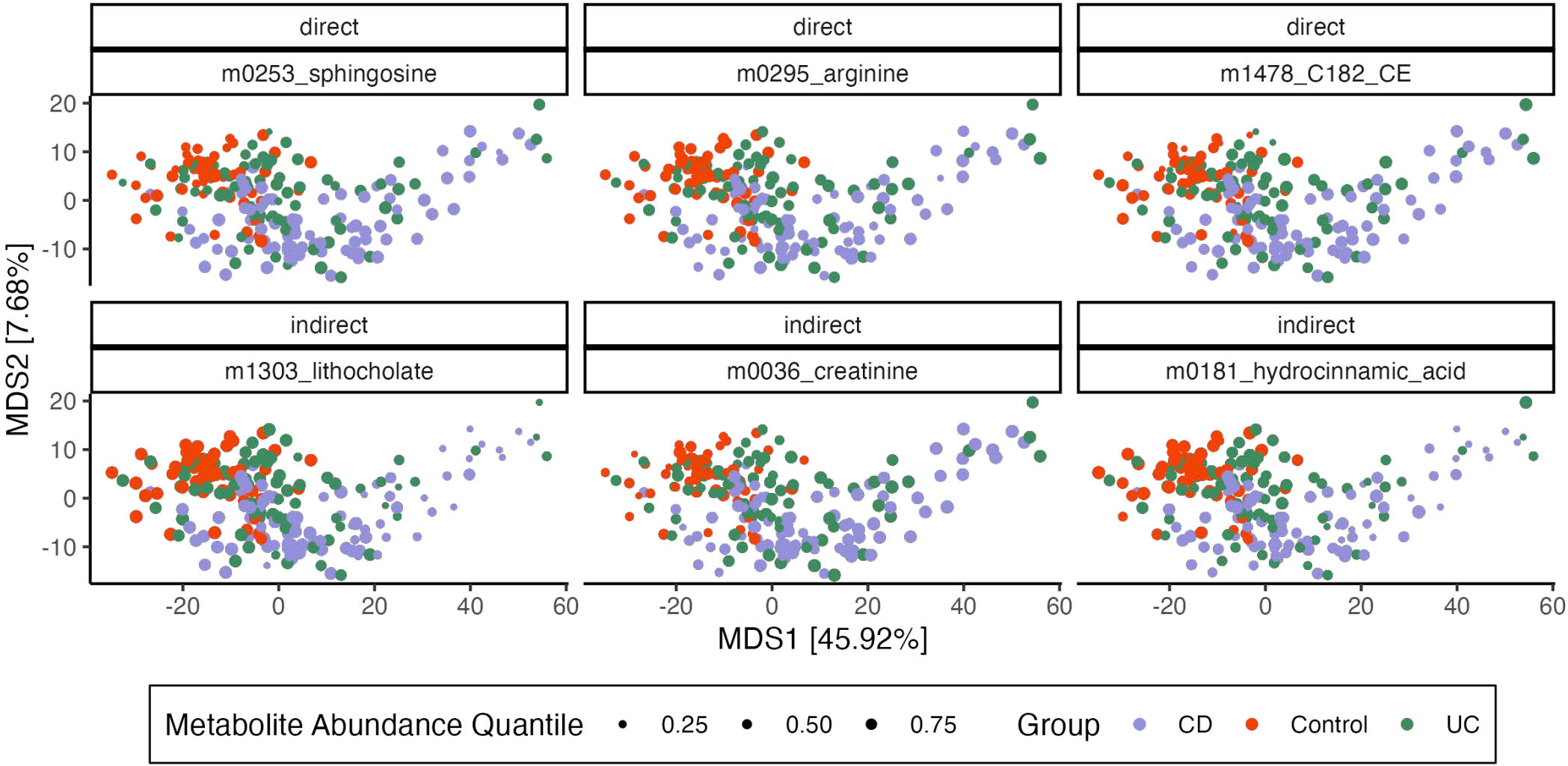
[56] Haixiang Zhang, Jun Chen, Zhigang Li, and Lei Liu. Testing for mediation effect with application to human microbiome data. *Statistics in Biosciences*, 13(2):313–328, July 2019.

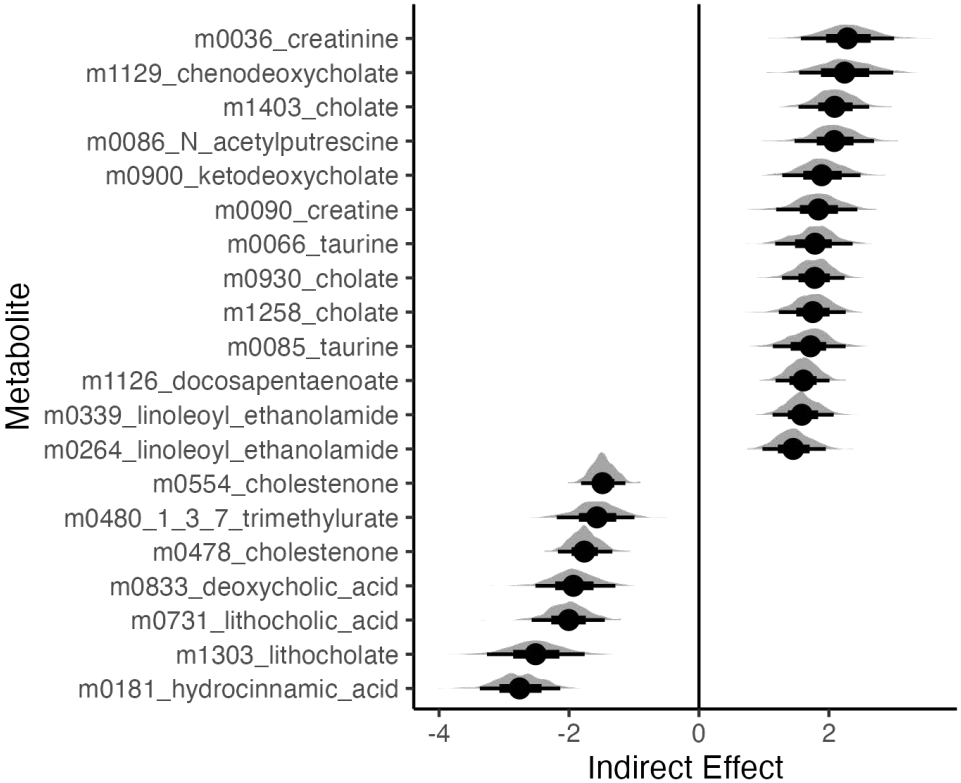

M_i
T_i
Y_i
Original Data
RF (Full)
RF (T-\->M)
RF (T-\->Y)
Linear Model
Y[1]
Y[2]
Outcome
5.0
2.5
0.0
-2.5
-5.0
-5
0
5
Mediator M
Treatment
0
1

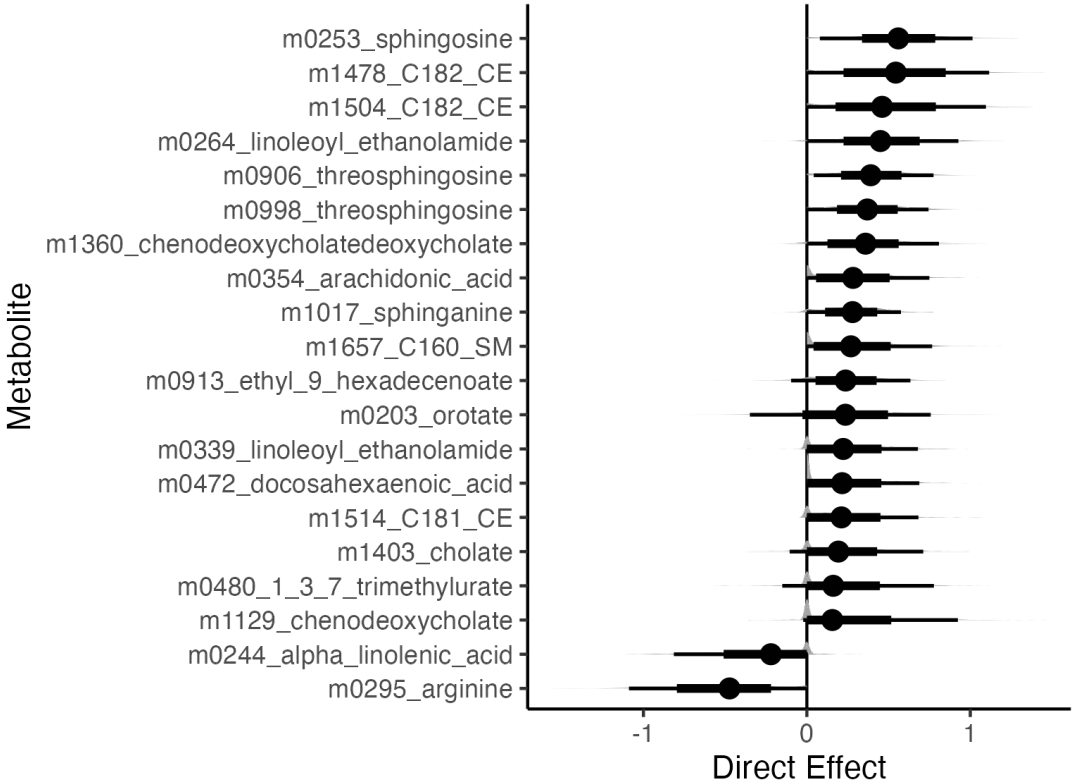

IE: 0.813
m0900_ketodeoxycholate
gCAG103
IE: 0.784
m0066_taurine
gCholadousia
IE: 0.778
m0066_taurine
gBilophila
IE: 0.741
m0085_taurine
gBilophila
IE: 0.702
m0085_taurine
gCholadousia
IE: 0.702
m1403_cholate
gCAG103
IE: 0.658
m1126_docosapentaenoate
gCAG103
IE: 0.629
m0036_creatinine
gCholadousia
IE: 0.626
m0090_creatine
gCholadousia
IE: 0.577
m1129_chenodeoxycholate
gCAG103
IE: 0.567
m1373_C363_DAG
gCholadousia
IE: 0.543
m1504_C182_CE
gButyricicoccusA
Study.Group
CD
Control
UC

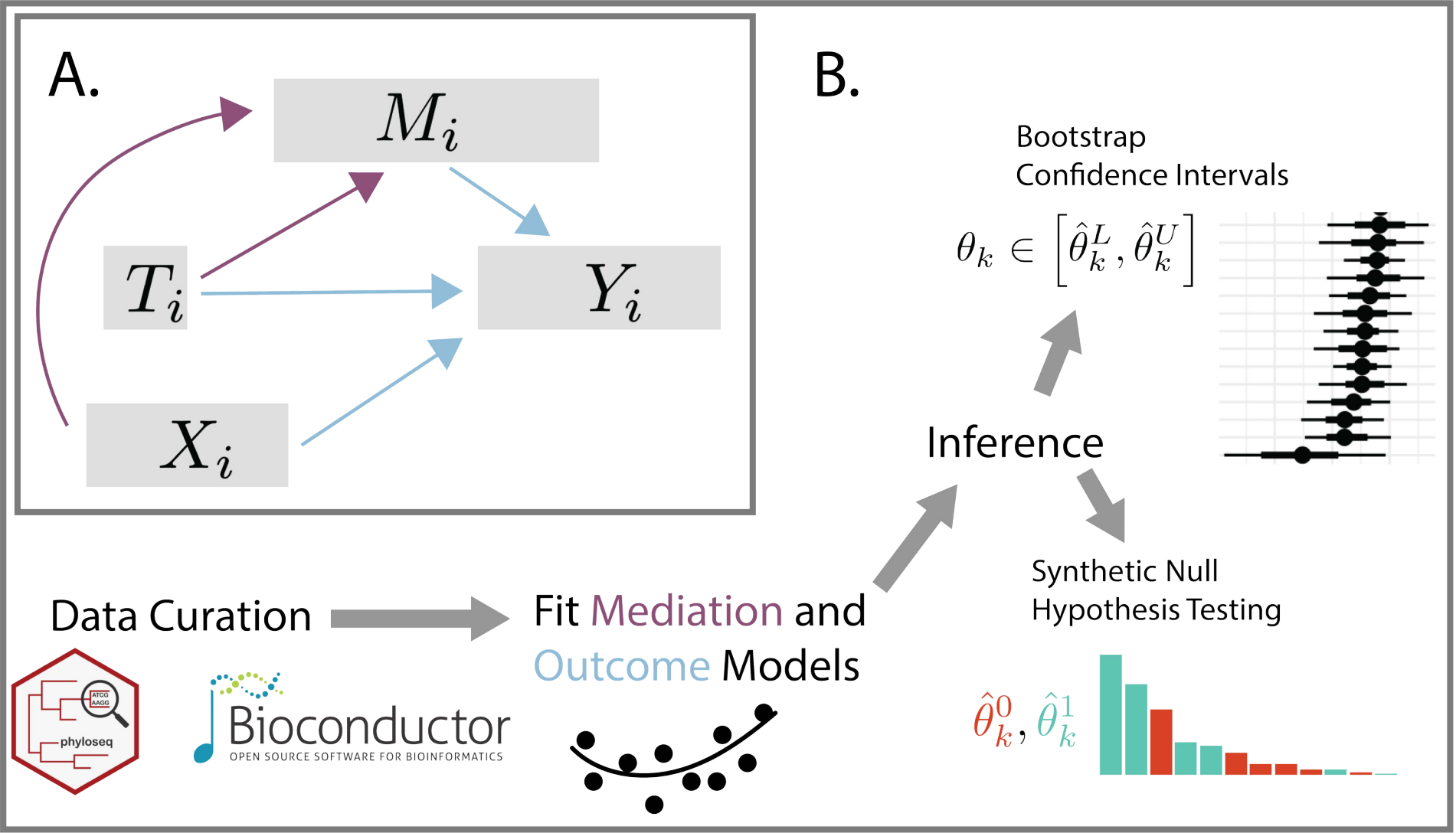

Full Model
No Direct Effect
No Indirect Effect
Original Data
Genus
Blautia
Faecalibacterium
Bacteroides
Ruminococcus
Roseburia
Ruminococcus2
Prevotella
Dorea
Anaerostipes
Clostridium_XIVa
Collinsella
Coprococcus
Bifidobacterium
Dialister
Clostridium_IV
Gemmiger
Phascolarctobacterium
Alistipes
Oscillibacter
Clostridium_XI
Methanobrevibacter
Paraprevotella
Akkermansia
Streptococcus
Parabacteroides
Clostridium_XVIII
Sporobacter
Butyricicoccus
Acidaminococcus
Flavonifractor
Clostridium_sensu_stricto
Barnesiella
Parasutterella
Clostridium_XIVb
Sutterella
Subdoligranulum
Bilophila
Escherichia.Shigella
Clostridium_III
Odoribacter
Turicibacter
Pseudoflavonifractor
Slackia
Acetanaerobacterium
Lactococcus
Lactobacillus
Eubacterium
Eggerthella
Asaccharobacter
Lactonifactor
Anaerosporobacter
Actinomyces
Anaerotruncus
Defluviitalea
Gordonibacter
-9
-6
-3
-10.0
-7.5
-5.0
-2.5
0.0
log(Relative Abundance)
Control
Treatment


Direct
Indirect
FDR
0.4
0.2
0.0
Blautia
Faecalibacterium
Bifidobacterium
Ruminococcus2
Collinsella
Faecalibacterium
-0.06
-0.04
-0.02
0.00
-0.0025
0.0000
0.0025
0.0050
Effect
Data Source
real
synthetic


direct
m0253_sphingosine
direct
m0295_arginine
direct
m1478_C182_CE
indirect
m1303_lithocholate
indirect
m0036_creatinine
indirect
m0181_hydrocinnamic_acid
MDS2 [7.68%]
MDS1 [45.92%]
20
10
0
-10
-20
0
20
40
60
Metabolite Abundance Quantile
0.25
0.50
0.75
Group
CD
Control
UC


m0036_creatinine
m1129_chenodeoxycholate
m1403_cholate
m0086_N_acetylputrescine
m0900_ketodeoxycholate
m0090_creatine
m0066_taurine
m0930_cholate
m1258_cholate
m0085_taurine
m1126_docosapentaenoate
m0339_linoleoyl_ethanolamide
m0264_linoleoyl_ethanolamide
m0554_cholestenone
m0480_1_3_7_trimethylurate
m0478_cholestenone
m0833_deoxycholic_acid
m0731_lithocholic_acid
m1303_lithocholate
m0181_hydrocinnamic_acid
Metabolite
-4
-2
0
2
Indirect Effect


m0253_sphingosine
m1478_C182_CE
m1504_C182_CE
m0264_linoleoyl_ethanolamide
m0906_threosphingosine
m0998_threosphingosine
m1360_chenodeoxycholatedeoxycholate
m0354_arachidonic_acid
m1017_sphinganine
m1657_C160_SM
m0913_ethyl_9_hexadecenoate
m0203_orotate
m0339_linoleoyl_ethanolamide
m0472_docosahexaenoic_acid
m1514_C181_CE
m1403_cholate
m0480_1_3_7_trimethylurate
m1129_chenodeoxycholate
m0244_alpha_linolenic_acid
m0295_arginine
Metabolite
-1
0
1
Direct Effect


A.
M_i
T_i
Y_i
X_i
B.
Bootstrap
Confidence Intervals
$\theta_k \in \left[\hat{\theta}_k^L, \hat{\theta}_k^U\right]$
Inference
Synthetic Null
Hypothesis Testing
$\hat{\theta}_k^0, \hat{\theta}_k^1$
Data Curation
Fit Mediation and
Outcome Models
phyloseq
Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS