

1 multimedia: Multimodal Mediation Analysis 2 of Microbiome Data

3 Hanying Jiang^{1,†}, Xinran Miao^{1,†}, Margaret W. Thairu², Mara Beebe², Dan W. Grupe³,

4 Richard J. Davidson^{3,4,5}, Jo Handelsman^{2,6}, Kris Sankaran^{1,2}

5 ¹Statistics Department, UW-Madison, Madison, WI, USA

6 ²Wisconsin Institute for Discovery, UW-Madison, Madison, WI, USA

7 ³Center for Healthy Minds, UW-Madison, Madison, WI, USA

8 ⁴Psychology Department, UW-Madison, Madison, WI, USA

9 ⁵Psychiatry Department, UW-Madison, Madison, WI, USA

10 ⁶Plant Pathology Department, UW-Madison, Madison, WI, USA

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

12 † Equal contribution.

13 ABSTRACT

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

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

34 **IMPORTANCE**

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

42 **INTRODUCTION**

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

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

53

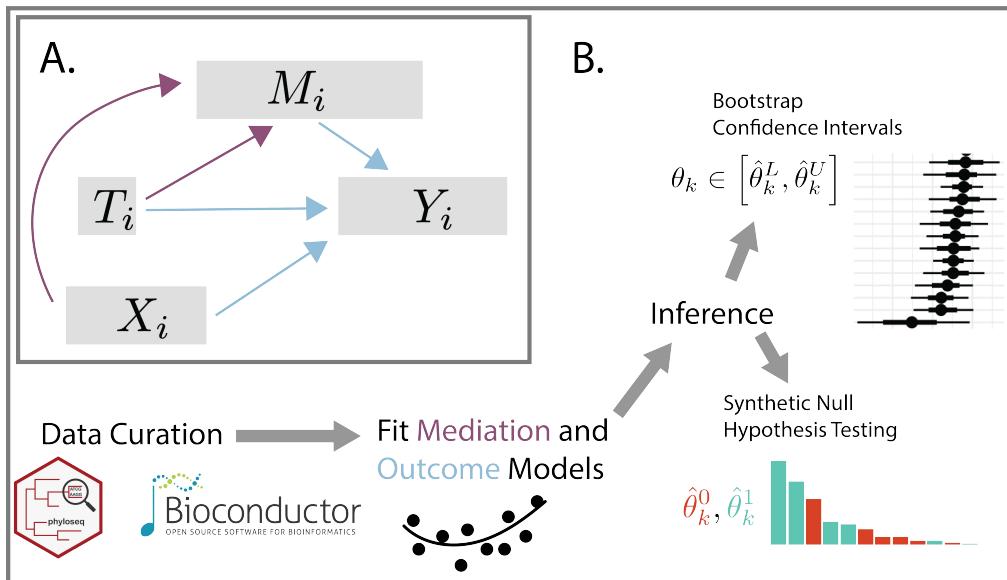


FIG 1 A. The graphical model underlying mediation analysis. Using combined mediation (purple) and outcome (blue) models, mediation analysis makes it possible to distinguish between direct and indirect causal pathways between treatments and outcomes. The conventional mediation analysis typically requires 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.

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

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

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

82 RESULTS

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

90 package functions is given in Table 1. The last two sections demonstrate the package
91 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.

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

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

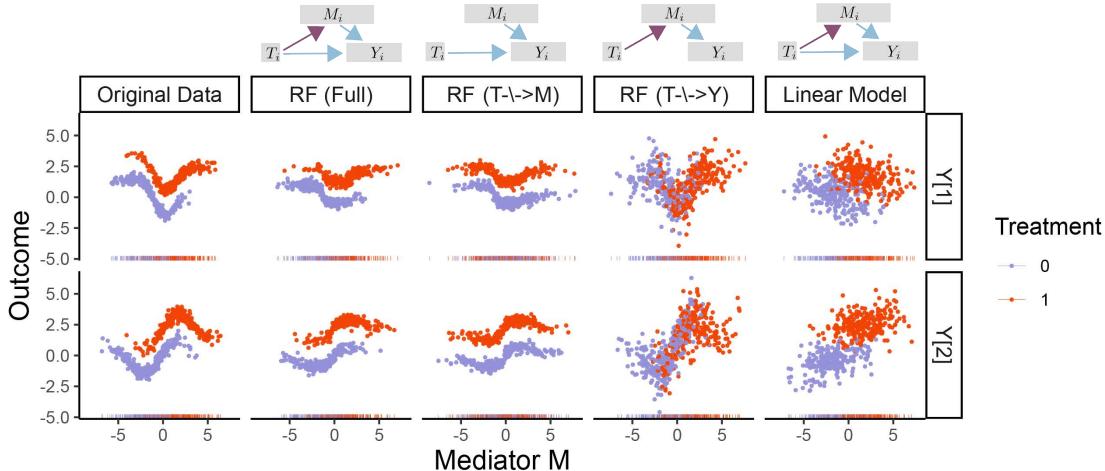
108 Next, we must specify the mediation and outcome models. The package exports
109 wrappers to several regression families, ensuring that, despite their differing underlying
110 methodology, all families can be used interchangeably for estimation, sampling, and
111 prediction in the overall mediation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

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

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

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

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



164

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.

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

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

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

194 **Microbiome-Metabolome Integration** We next illustrate the multimedia workflow
195 with case studies. Our first concerns Inflammatory Bowel Disease (IBD), which is closely
196 tied to gut microbiome community composition [31]. [16] investigated the relationship
197 between the gut microbiome and metabolome between IBD patients and healthy controls,
198 concluding that microbial community members may be partly responsible for the
199 formation of metabolites that lead to inflammation and IBD. By applying clustering and
200 canonical correlation analysis to untargeted mass spectrometry data, they flagged a
201 number of disease-relevant metabolites. We re-analyze the data using model-based
202 mediation analysis, viewing IBD status – Healthy Control, Ulcerative Colitis (UC),
203 Crohn’s Disease (CD) – as treatments T , metabolic profile as the outcome Y , and
204 microbiome community composition as a mediator M . The data are downloaded from the
205 microbiome-metabolome curated data repository [32]. We have further filtered to the top
206 173 and 155 most abundant microbes and metabolites, and we apply centered log-ratio

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

209 We use parallel linear and ℓ^1 -regularized regression for mediation and outcome
210 models, respectively. Note that treatment is the only predictor in the mediation model,
211 which is why no regularization is required. We ran the bootstrap for 1000 iterations, and
212 95% confidence intervals and bootstrap distributions for the features with the strongest
213 direct and overall indirect effects contrasting CD with healthy controls are shown in Fig. 3.
214 Metabolites with strong indirect effects are influenced by IBD-induced changes in
215 microbiome community composition, while those with large direct effects change due to
216 other unknown factors. Fig. 4 contextualizes a small subset of these overall effects by
217 overlaying metabolite abundances onto multidimensional scaling (MDS) plots derived
218 from microbiome community profiles. Though metabolites with strong direct effects have
219 differential abundance across IBD and healthy groups, only metabolites with indirect
220 effects show variation that is also associated with microbiome composition.

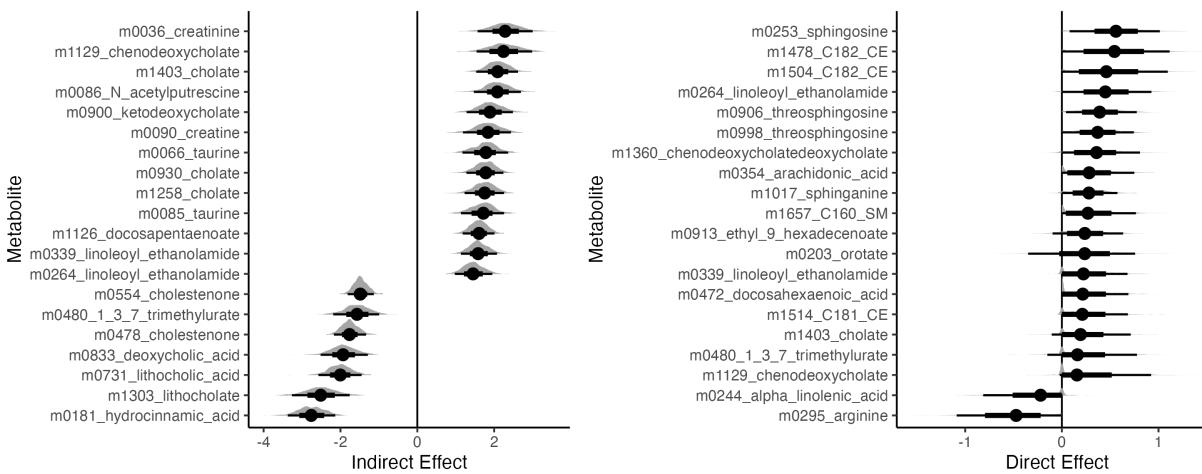
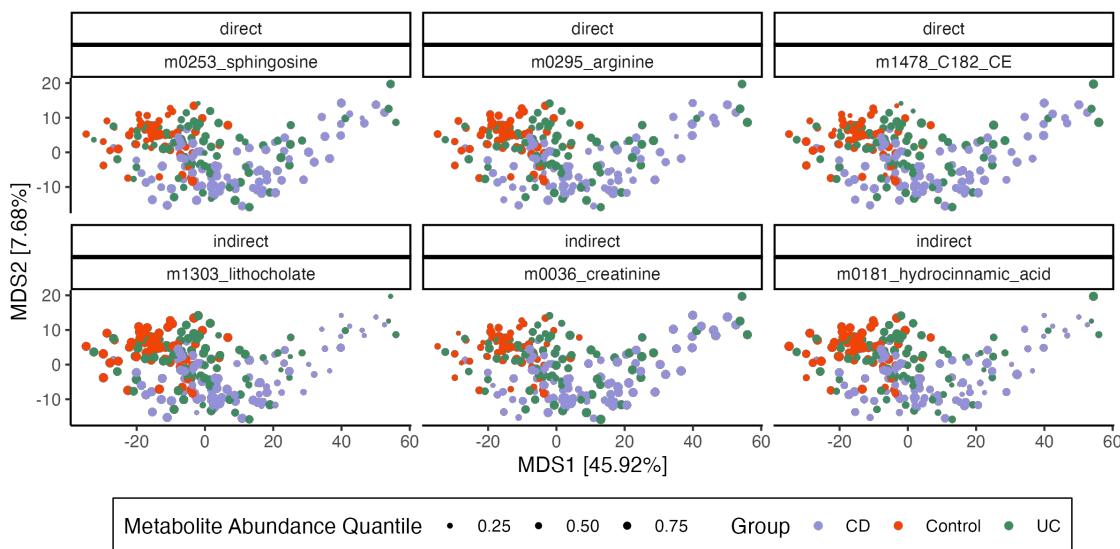


FIG 3 95% Bootstrap confidence intervals for metabolites with the strongest estimated direct and overall indirect effects associated with CD. Effects are sorted according to magnitude, and only the top 15 of each type are shown. Within the interval, the inner rectangle captures 66% of the bootstrap samples.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

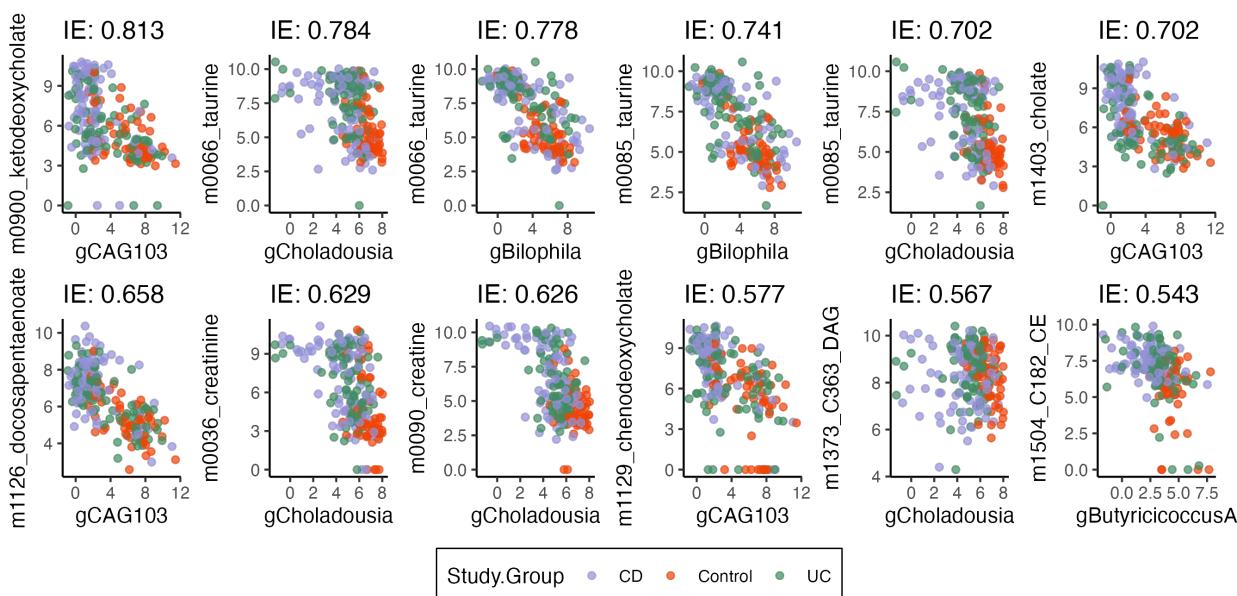


222

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.

223 Moreover, by analyzing pathwise indirect effects, we can uncover genus-level
224 relationships. A subset of the strongest pathwise indirect effects are shown in Fig. 5.
225 Among the microbe-metabolite pairs with the strongest pathwise indirect effects, we find
226 a relationship between metabolites of taurine and *Bilophila* (Fig. 5). High levels of fecal
227 taurine, one of the primary conjugates of primary bile acids [52], has been previously
228 associated with IBD [25, 46]. It has also been found that *Bilophila wadsworthia*, one of the
229 most prominent taurine metabolizers, is often associated with lower levels of taurine [46].
230 Here, our results suggest that higher levels of taurine in IBD patients is mediated in part,
231 by the abundance of *Bilophila*. We also find microbes in the genus *Firmicutes* bacterium
232 CAG:103, are paired with several metabolites: cholate, chenodeoxycholate, and
233 7-ketodeoxycholate (Fig. 5). Cholate and chenodeoxycholate are primary bile acids

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



249

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 ($T \rightarrow M$) and also an association between microbe and metabolite abundance ($M \rightarrow Y$).

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

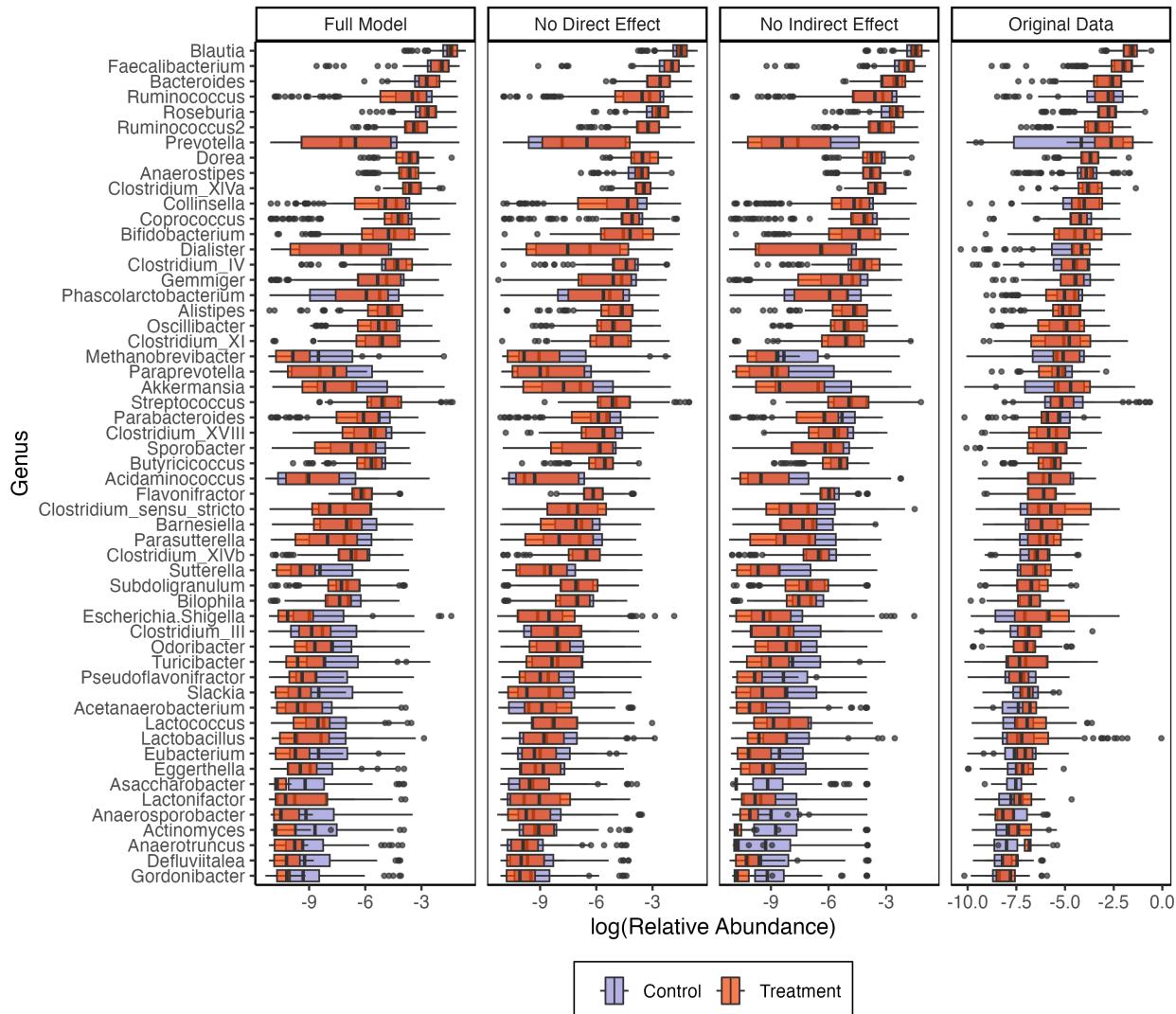
257 **Evaluating a Mindfulness Intervention** Studies of the gut-brain axis have yielded
258 experimental evidence for interactions between the gut microbiome and the brain. For
259 example, germ-free mice colonized with the microbiota from human patients with clinical
260 depression develop depression-like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

275 For mediation and outcome models, we apply ridge and logistic-normal multinomial
276 regressions, respectively [22, 54]. We choose a ridge regression model so that intercepts
277 across the large number of participants are shrunk towards their global mean. We choose
278 logistic-normal multinomial regression to jointly model microbiome composition. We also
279 define altered submodels where all direct and indirect effects have been removed.

280 Simulated genera compositions from all models are shown in Fig. 6. In the newly
281 simulated data, subjects have been randomly re-assigned to the treatment and control
282 groups. These submodels can support synthetic null hypothesis testing, since the
283 synthetic null data appear to capture relevant properties of the real microbiome
284 composition profiles, like the average relative abundances across genera and the range of
285 observed abundances within most genera. Their main limitation is that some genera, like
286 *Methanobrevibacter*, *Paraprevotella*, and *Akkermansia*, have much wider ranges than the
287 synthetic data, and Fig. S1 suggests that this is due to a failure to capture the unusually
288 high zero inflation present in these genera.

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

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



304

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 ($T \rightarrow Y$) and indirect ($M \rightarrow Y$) effects removed.

305

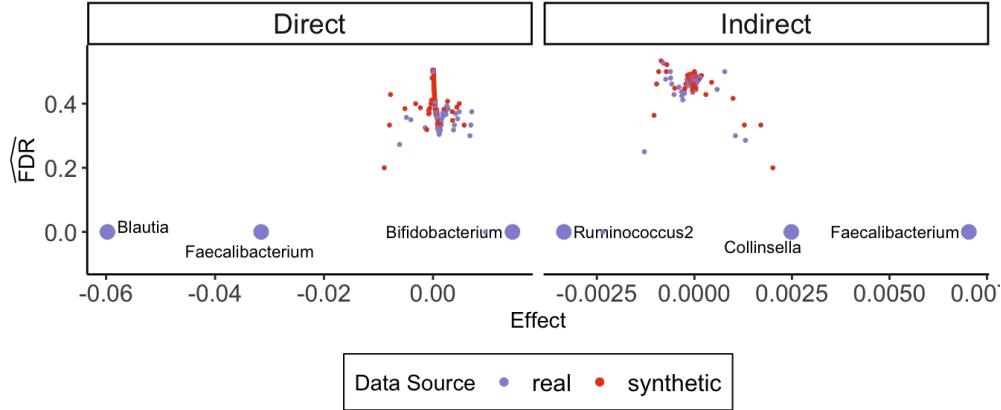


FIG 7 Estimated overall effects and false discovery rates derived from real and synthetic null data. Each point corresponds to one genus in either real (blue) or simulated (orange) data. The genera selected to control the false discovery rate at $q \leq 0.15$ are drawn larger than the rest. Direct effects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

306 DISCUSSION

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

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

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

324 MATERIALS AND METHODS

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

330 To what extent is the effect of the treatment on the outcome modulated by
331 intermediate variables? Mediation analysis answers this question by positing mediators
332 M_i on the causal path from T_i to Y_i . Adopting a counterfactual perspective, we define
333 $M_i(t')$ as the potential outcome of the mediator under $T_i = t'$ and $Y_i(t, m)$ as the potential
334 outcome of the response under $T_i = t$ and $M_i = m$ with $t, t' \in \{0, 1\}$. Therefore, we can
335 express the outcome variable as $Y_i(t, M_i(t'))$. In a randomized experiment, we can only
336 ever observe the case where t and t' are the same, i.e., $Y_i(1, M_i(1))$ in the treated group
337 and $Y_i(0, M_i(0))$ in the control group – but conceptually t and t' can be different. For
338 example, $Y_i(0, M_i(1))$ represents the potential outcome when only mediators are
339 intervened upon and $Y_i(1, M_i(0))$ represents the potential outcome when we make
340 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,

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

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

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

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

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

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

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

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

341 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
342 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,

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

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

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 $M_i = (M_{i1}, \dots, M_{iK})$,

$$\bar{\omega}_k = \frac{1}{2} \sum_{t'=0}^1 \mathbb{E}\{Y_i(t', (M_{i1}(t'), \dots, M_{ik}(1), \dots, M_{iK}(t'))) - Y_i(t', (M_{i1}(t'), \dots, M_{ik}(0), \dots, M_{iK}(t')))\}. \quad (9)$$

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

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

343 **Bootstrap and synthetic null testing** Form a bootstrap resample of the data
 344 $\mathcal{D}^* = (\mathbf{X}^*, \mathbf{M}^*, \mathbf{T}^*, \mathbf{Y}^*)$ by independently resampling the n observations with replacement.
 345 A summary statistic computed on the b^{th} resampled dataset is denoted by $\hat{\theta}^{*b}(\mathcal{D}^*)$. For
 346 brevity, we will omit the data arguments. For example, $\hat{\theta}^{*b}$ could correspond to an
 347 estimator of δ or ζ derived from mediation and outcome models learned from \mathcal{D}^* .
 348 Repeating this process B times, we refit $\hat{M}(t, x)$, $\hat{Y}(t, m, x)$ and the provided summary
 349 statistic $\hat{\theta}$ for each of the B bootstrap datasets. We can obtain the bootstrap distribution
 350 $(\hat{\theta}^{*b})_{b=1}^B$. 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
 351 $[q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}]$ 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 $(\hat{\theta}_d^1)_{d=1}^D$ and $(\hat{\theta}_d^0)_{d=1}^D$ based on the original and the synthetic null data, respectively. For example, $\hat{\theta}_d^1$ could estimate taxon d 's direct effect $\hat{\delta}_d$ 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

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

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

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

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

366 A total of 11,720 taxa and 8,848 metabolites are present in the public data. We applied
367 a centered log-ratio transformation to the microbiome relative abundances profiles:
368 $\text{CLR}(x_1, \dots, x_D) := \left(\log(x_d) - \frac{1}{D} \sum_{d'} \log x_{d'} \right)_{d=1}^D$. We then filtered to taxa whose average
369 transformed abundance was larger than 3, which reduced the number of taxa to 173. We
370 kept only metabolites with confident HMDB assignments, applied a $\log(1 + x)$
371 transformation, and further filtered to those whose average transformed intensity was
372 larger than 6. This resulted in 155 well-annotated and generally abundant metabolites.

373 **Mindfulness study design and processing** The initial Center for Healthy Minds study
374 recruited 114 police officers participants across two cohorts. Microbiome samples were
375 obtained only from participants in the second cohort ($n = 54$), who were randomly
376 assigned to mindfulness training or waitlist control with 27 cases each. We removed four
377 participants due to incomplete responses – three lacked microbiome data, and one had
378 missing mediators. Our analysis considers a mindfulness training treatment group of size
379 $n = 24$ and a waitlist control group of size $n = 26$. Participants in the mindfulness group
380 took part in an 8-week, 18-hour mindfulness training developed specifically for their
381 career and inspired by Mindfulness-Based Stress Reduction and Mindfulness-Based
382 Resilience Training [7]. Weekly two-hour classes (and a four-hour class in week 7)
383 consisted of didactic instruction, embodied mindfulness practices, and individual and
384 group-based inquiry (for full intervention details, see [21]). Microbiota and behavioral
385 survey data were gathered at 2 - 3 timepoints for each participant — samples in the
386 treatment group provided data before, within two weeks following, and, in a subset of
387 cases, four months after the 8-week intervention, resulting in 118 samples total.

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

390 replicates (on average, 2.6) were sequenced for each 16S rRNA gene sample, resulting in
391 307 microbiome composition profiles in total. Amplicon Sequence Variants (ASV) were
392 called using the DADA2 pipeline [4]. The first ten base pairs were removed, and all reads
393 were truncated to a length of 250. Otherwise, we set all pipeline hyperparameters to their
394 defaults. Since the total number of participants is relatively small, we chose to concentrate
395 on the core microbiome [37]. To this end, we assigned taxonomic identity to each ASV
396 using the RDP database and aggregated all counts to the genus level [9]. We removed any
397 genera that did not appear in at least 40% of the samples, thereby generating a core
398 microbiome. On average, this preserved 98.7% of the reads within each sample. After
399 filtering to the core microbiome, sequences for 55 genera remained. To define mediators,
400 we manually selected four variables from the National Cancer Institute Quick Food Scan
401 and self-reported questionnaires on fatigue and sleep disturbance scores based on the
402 Patient-Reported Outcomes Measurement Information System subscale [6]. We
403 concentrated on these questions because changes in both diet and sleep have previously
404 been associated with mindfulness interventions and the microbiome [19, 12, 47].

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

412 **ACKNOWLEDGMENTS**

413 **DATA AVAILABILITY STATEMENT**

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

416 FUNDING

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

419 CONFLICTS OF INTEREST

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

421 References

- 422 [1] Thomaz FS Bastiaanssen, Sofia Cussotto, Marcus J Claesson, Gerard Clarke,
423 Timothy G Dinan, and John F Cryan. Gutted! unraveling the role of the microbiome
424 in major depressive disorder. *Harvard Review of Psychiatry*, 28(1):26, 2020.
- 425 [2] Jos A Bosch, Max Nieuwdorp, Aeilko H Zwinderman, Mélanie Deschaseaux, Djawad
426 Radjabzadeh, Robert Kraaij, Mark Davids, Susanne R de Rooij, and Anja Lok. The
427 gut microbiota and depressive symptoms across ethnic groups. *Nature
428 Communications*, 13(1):7129, 2022.
- 429 [3] Paul-Christian Bürkner. brms: An R package for bayesian multilevel models using
430 stan. *Journal of Statistical Software*, 080:1–28, 2017.
- 431 [4] Benjamin J Callahan, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy Jo A
432 Johnson, and Susan P Holmes. Dada2: High-resolution sample inference from
433 illumina amplicon data. *Nature Methods*, 13(7):581–583, 2016.
- 434 [5] Kyle M. Carter, Meng Lu, Hongmei Jiang, and Lingling An. An information-based
435 approach for mediation analysis on high-dimensional metagenomic data. *Frontiers in
436 Genetics*, 11, 2020.
- 437 [6] David Celli, William Riley, Arthur Stone, Nan Rothrock, Bryce Reeve, Susan Yount,
438 Dagmar Amtmann, Rita Bode, Daniel Buysse, Seung Choi, et al. The patient-reported
439 outcomes measurement information system (promis) developed and tested its first

- 440 wave of adult self-reported health outcome item banks: 2005–2008. *Journal of clinical*
441 *epidemiology*, 63(11):1179–1194, 2010.
- 442 [7] Michael S Christopher, Richard J Goerling, Brant S Rogers, Matthew Hunsinger, Greg
443 Baron, Aaron L Bergman, and David T Zava. A pilot study evaluating the
444 effectiveness of a mindfulness-based intervention on cortisol awakening response
445 and health outcomes among law enforcement officers. *Journal of police and criminal*
446 *psychology*, 31:15–28, 2016.
- 447 [8] Dylan Clark-Boucher, Xiang Zhou, Jiacong Du, Yongmei Liu, Belinda L Needham,
448 Jennifer A. Smith, and Bhramar Mukherjee. Methods for mediation analysis with
449 high-dimensional dna methylation data: Possible choices and comparisons. *PLOS*
450 *Genetics*, 19, 2023.
- 451 [9] James R Cole, Qiong Wang, Jordan A Fish, Benli Chai, Donna M McGarrell, Yanni
452 Sun, C Titus Brown, Andrea Porras-Alfaro, Cheryl R Kuske, and James M Tiedje.
453 Ribosomal database project: data and tools for high throughput rrna analysis. *Nucleic*
454 *acids research*, 42(D1):D633–D642, 2014.
- 455 [10] Sean P. Colgan, Ruth X. Wang, Caroline H.T. Hall, Geetha Bhagavatula, and J. Scott
456 Lee. Revisiting the “starved gut” hypothesis in inflammatory bowel disease.
457 *Immunometabolism*, 5(1):e0016, January 2023.
- 458 [11] John F Cryan and Sarkis K Mazmanian. Microbiota–brain axis: Context and causality.
459 *Science*, 376(6596):938–939, 2022.
- 460 [12] Lawrence A David, Corinne F Maurice, Rachel N Carmody, David B Gootenberg,
461 Julie E Button, Benjamin E Wolfe, Alisha V Ling, A Sloan Devlin, Yug Varma,
462 Michael A Fischbach, et al. Diet rapidly and reproducibly alters the human gut
463 microbiome. *Nature*, 505(7484):559–563, 2014.
- 464 [13] Bradley Efron. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*,
465 7:1–26, 1979.
- 466 [14] Bradley Efron. The jackknife, the bootstrap, and other resampling plans. *SIAM*, 38,
467 1987.

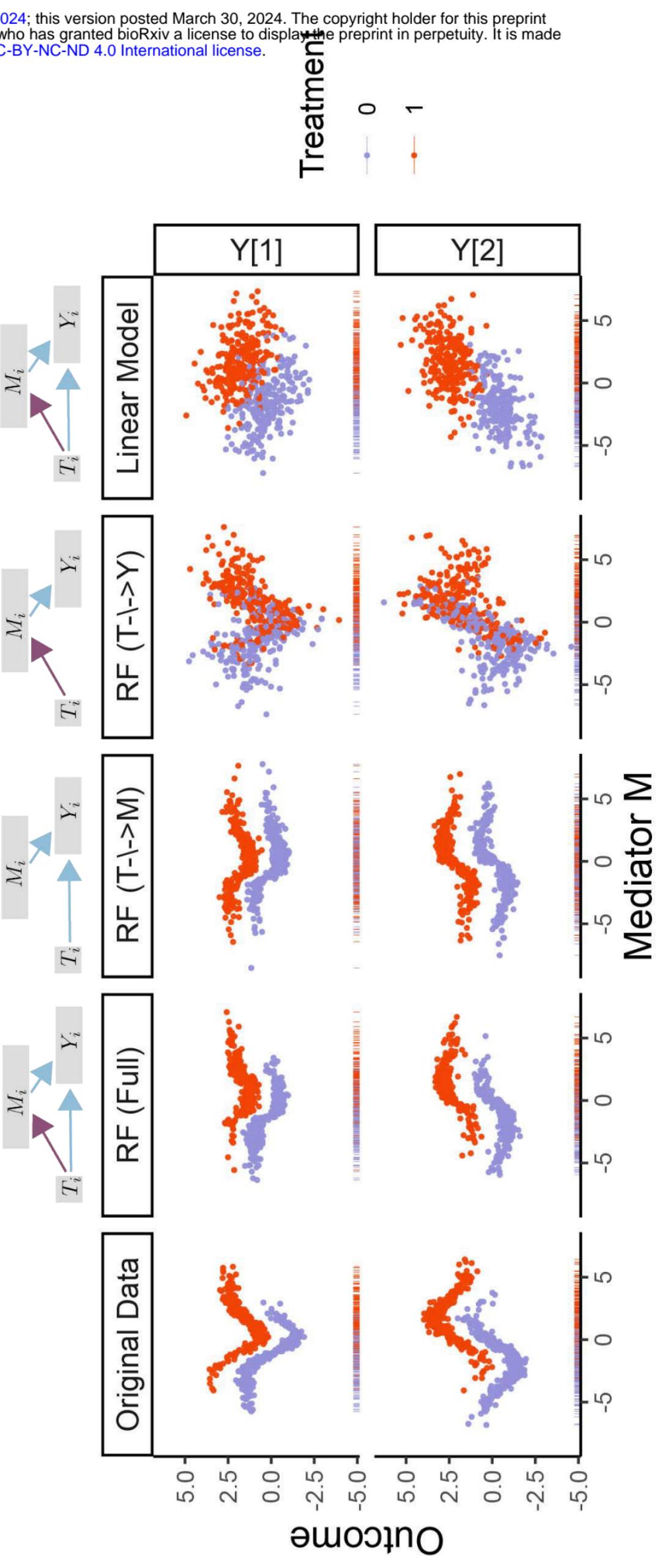
- 468 [15] Bradley Efron and Robert Tibshirani. *An Introduction to the Bootstrap*. Chapman &
469 Hall/CRC Monographs on Statistics and Applied Probability. Chapman &
470 Hall/CRC, Philadelphia, PA, May 1994.
- 471 [16] Eric A. Franzosa, Alexandra Sirota-Madi, Julián Ávila-Pacheco, Nadine Fornelos,
472 Henry J. Haiser, Stefan Reinker, Tommi Vatanen, A. Brantley Hall, FASA
473 Himel Mallick, PhD, Lauren J. McIver, Jenny S. Sauk, Robin G. Wilson, Betsy W.
474 Stevens, Justin Scott, Kerry A. Pierce, Amy Anderson Deik, Kevin Bullock, Floris
475 Imhann, Jeffrey A. Porter, Alexandra Zhernakova, Jingyuan Fu, Rinse K. Weersma,
476 Cisca Wijmenga, Clary B. Clish, Hera Vlamakis, Curtis Huttenhower, and Ramnik J.
477 Xavier. Gut microbiome structure and metabolic activity in inflammatory bowel
478 disease. *Nature microbiology*, 4:293 – 305, 2018.
- 479 [17] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Regularization paths for
480 generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1),
481 2010.
- 482 [18] Robert Gentleman, Vincent J. Carey, Douglas M. Bates, Benjamin M. Bolstad, Marcel
483 Dettling, Sandrine Dudoit, Byron Ellis, Laurent Gautier, Yongchao Ge, Jeff Gentry,
484 Kurt Hornik, Torsten Hothorn, Wolfgang Huber, S. Iacus, Rafael A. Irizarry, Friedrich
485 Leisch, Cheng Li, Martin Maechler, Anthony J. Rossini, Günther Sawitzki, Colin
486 Smith, Gordon Smyth, Luke Tierney, Jean YH Yang, and Jianhua Zhang.
487 Bioconductor: open software development for computational biology and
488 bioinformatics. *Genome Biology*, 5:R80 – R80, 2004.
- 489 [19] Desleigh Gilbert and Jennifer Waltz. Mindfulness and health behaviors. *Mindfulness*,
490 1(4):227–234, 2010.
- 491 [20] Daniel W. Grupe, Jonah L. Stoller, Carmen Alonso, C. R. Mcgehee, Christion Smith,
492 Jeanette A. Mumford, Melissa A. Rosenkranz, and Richard J. Davidson. The impact
493 of mindfulness training on police officer stress, mental health, and salivary cortisol
494 levels. *Frontiers in Psychology*, 12, 2021.
- 495 [21] Daniel W Grupe, Jonah L Stoller, Carmen Alonso, Chad McGehee, Chris Smith,
496 Jeanette A Mumford, Melissa A Rosenkranz, and Richard J Davidson. The impact of

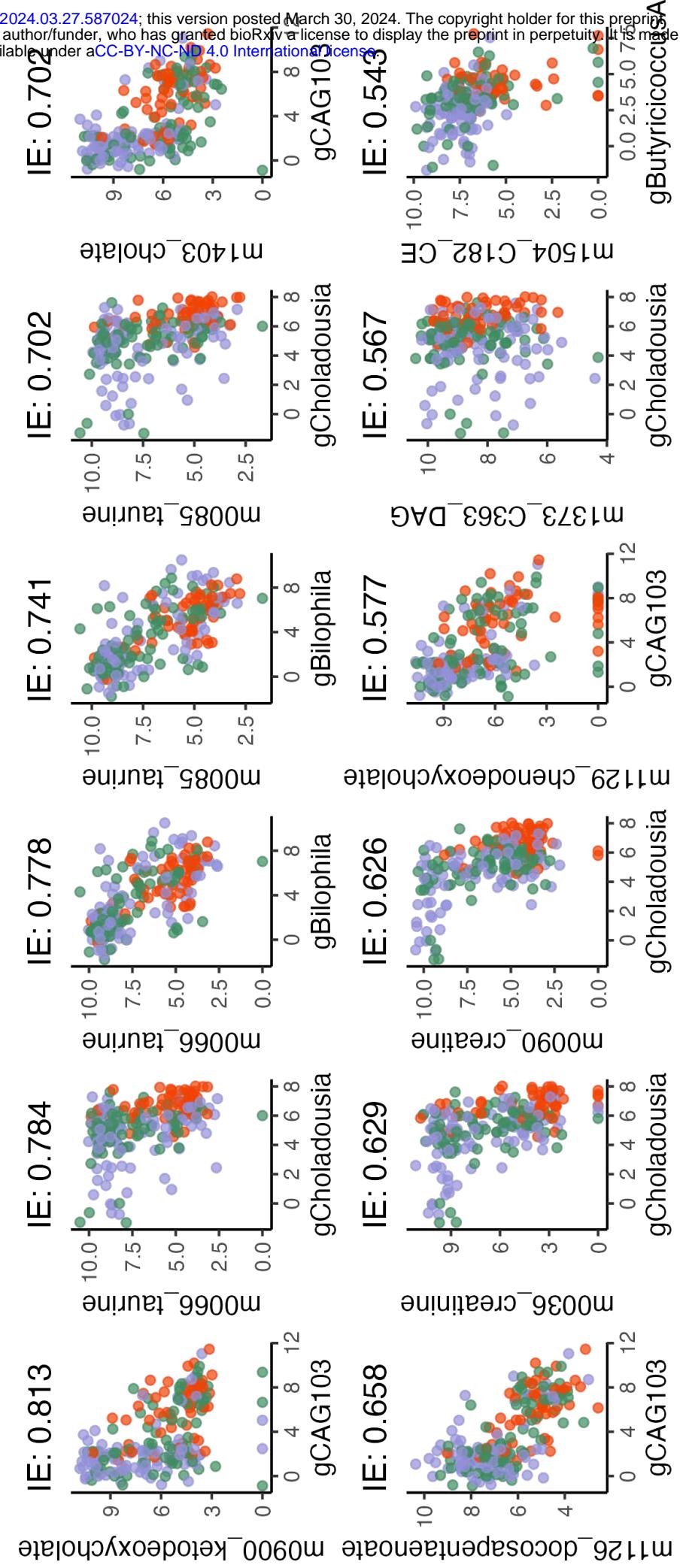
- 497 mindfulness training on police officer stress, mental health, and salivary cortisol
498 levels. *Frontiers in Psychology*, 12:720753, 2021.
- 499 [22] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical
500 Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer
501 New York, New York, NY, 2009.
- 502 [23] Qilin Hong, Guanhua Chen, and Zheng-Zheng Tang. A phylogeny-based test of
503 mediation effect in microbiome. 2021.
- 504 [24] Kosuke Imai, Luke J. Keele, and Dustin Tingley. A general approach to causal
505 mediation analysis. *Political Methods: Quantitative Methods eJournal*, 2010.
- 506 [25] Jasmijn Z Jagt, Eduard A Struys, Ibrahim Ayada, Abdellatif Bakkali, Erwin E W
507 Jansen, Jürgen Claesen, Johan E van Limbergen, Marc A Benninga, Nanne K H
508 de Boer, and Tim G J de Meij. Fecal amino acid analysis in newly diagnosed pediatric
509 inflammatory bowel disease: A multicenter case-control study. *Inflammatory Bowel
510 Diseases*, 28(5):755–763, November 2021.
- 511 [26] Leo Lahti, Felix G. M. Ernst, Sudarshan A. Shetty, Tuomas Borman, Ruizhu Huang,
512 Domenick J. Braccia, and Hector Corrado Bravo. Microbiome data science in the
513 summarizedexperiment universe. *F1000Research*, 10, 2021.
- 514 [27] Wei Vivian Li and Jingyi Jessica Li. A statistical simulator scdesign for rational
515 scrna-seq experimental design. *Bioinformatics*, 35:i41 – i50, 2018.
- 516 [28] Yuanyuan Luo, Benhua Zeng, Li Zeng, Xiangyu Du, Bo Li, Ran Huo, Lanxiang Liu,
517 Haiyang Wang, Meixue Dong, Junxi Pan, et al. Gut microbiota regulates mouse
518 behaviors through glucocorticoid receptor pathway genes in the hippocampus.
519 *Translational psychiatry*, 8(1):1–10, 2018.
- 520 [29] David P. Mackinnon, Amanda J. Fairchild, and Matthew S. Fritz. Mediation analysis.
521 *Annual review of psychology*, 58:593–614, 2019.
- 522 [30] Paul J. McMurdie and Susan P. Holmes. phyloseq: An r package for reproducible
523 interactive analysis and graphics of microbiome census data. *PLoS ONE*, 8, 2013.

- 524 [31] Indrani Mukhopadhyay, Richard Hansen, Emad M El-Omar, and Georgina L Hold.
525 Ibd—what role do proteobacteria play? *Nature reviews Gastroenterology & hepatology*,
526 9(4):219–230, 2012.
- 527 [32] Efrat Muller, Yadic M Algavi, and Elhanan Borenstein. The gut
528 microbiome-metabolome dataset collection: a curated resource for integrative
529 meta-analysis. *NPJ Biofilms and Microbiomes*, 8, 2022.
- 530 [33] Daniela Parada Venegas, Marjorie K. De la Fuente, Glauben Landskron, María Julieta
531 González, Rodrigo Quera, Gerard Dijkstra, Hermie J. M. Harmsen, Klaas Nico Faber,
532 and Marcela A. Hermoso. Short chain fatty acids (scfas)-mediated gut epithelial and
533 immune regulation and its relevance for inflammatory bowel diseases. *Frontiers in*
534 *Immunology*, 10, March 2019.
- 535 [34] Edoardo Pasolli, Lucas Schiffer, Paolo Manghi, Audrey Renson, Valerie Obenchain,
536 Duy Tin Truong, Francesco Beghini, Fa Malik, Marcel Ramos, Jennifer Beam Dowd,
537 Curtis Huttenhower, Martin T. Morgan, N. Segata, and Levi Waldron. Accessible,
538 curated metagenomic data through experimenthub. *Nature Methods*, 14:1023–1024,
539 2017.
- 540 [35] Djawad Radjabzadeh, Jos A Bosch, André G Uitterlinden, Aeilko H Zwinderman,
541 M Arfan Ikram, Joyce BJ van Meurs, Annemarie I Luik, Max Nieuwdorp, Anja Lok,
542 Cornelia M van Duijn, et al. Gut microbiome-wide association study of depressive
543 symptoms. *Nature Communications*, 13(1):7128, 2022.
- 544 [36] Jason M. Ridlon, Steven L. Daniel, and H. Rex Gaskins. The hylemon-björkhem
545 pathway of bile acid 7-dehydroxylation: history, biochemistry, and microbiology.
546 *Journal of Lipid Research*, 64(8):100392, August 2023.
- 547 [37] Ashley Shade and Jo Handelsman. Beyond the venn diagram: the hunt for a core
548 microbiome. *Environmental microbiology*, 14(1):4–12, 2012.
- 549 [38] Michael B. Sohn and Hongzhe Li. Compositional mediation analysis for microbiome
550 studies. *The Annals of Applied Statistics*, 13(1), March 2019.

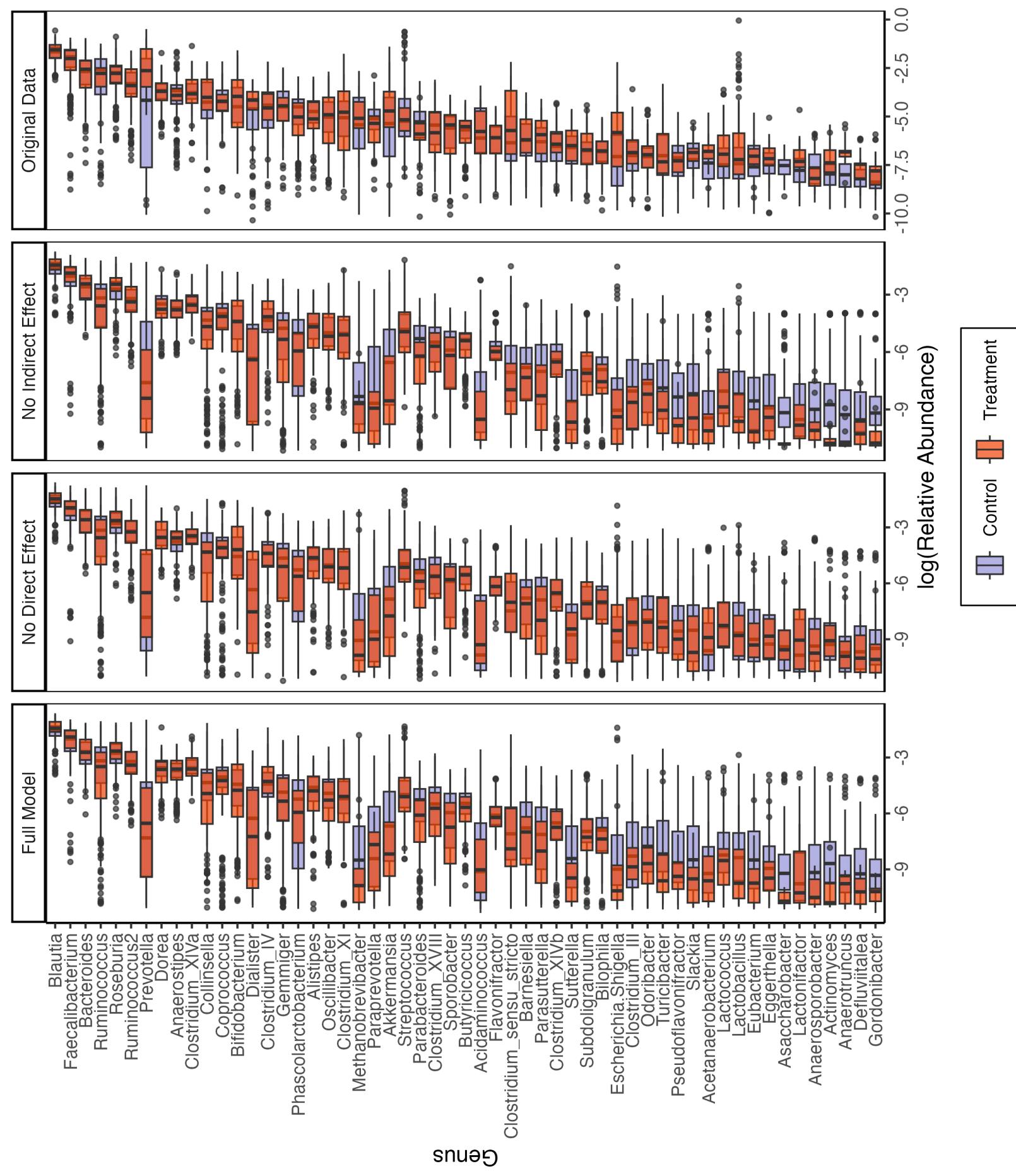
- 551 [39] Dongyuan Song, Qingyang Wang, Guanao Yan, Tianyang Liu, Tianyi Sun, and
552 Jingyi Jessica Li. scdesign3 generates realistic in silico data for multimodal single-cell
553 and spatial omics. *Nature Biotechnology*, 42:247–252, 2023.
- 554 [40] Tianyi Sun, Dongyuan Song, Wei Vivian Li, and Jingyi Jessica Li. scdesign2: a
555 transparent simulator that generates high-fidelity single-cell gene expression count
556 data with gene correlations captured. *Genome Biology*, 22, 2021.
- 557 [41] Ying Taur and Eric G. Pamer. Microbiome mediation of infections in the cancer
558 setting. *Genome Medicine*, 8, 2016.
- 559 [42] John P. Thomas, Dezso Modos, Simon M. Rushbrook, Nick Powell, and Tamas
560 Korcsmaros. The emerging role of bile acids in the pathogenesis of inflammatory
561 bowel disease. *Frontiers in Immunology*, 13, February 2022.
- 562 [43] Robert Tibshirani, Jacob Bien, Jerome Friedman, Trevor Hastie, Noah Simon,
563 Jonathan Taylor, and Ryan J. Tibshirani. Strong rules for discarding predictors in
564 lasso-type problems. *Journal of the Royal Statistical Society Series B: Statistical
565 Methodology*, 74(2):245–266, November 2011.
- 566 [44] Dustin Tingley, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai.
567 mediation: R package for causal mediation analysis. *Journal of Statistical Software*,
568 59:1–38, 2014.
- 569 [45] Arnau Vich Vila, Shixian Hu, Sergio Andreu-Sánchez, Valerie Collij, Bernadien H
570 Jansen, Hannah E Augustijn, Laura A Bolte, Renate A A A Ruigrok, Galeb Abu-Ali,
571 Cosmas Giallourakis, Jessica Schneider, John Parkinson, Amal Al-Garawi, Alexandra
572 Zhernakova, Ranko Gacesa, Jingyuan Fu, and Rinse K Weersma. Faecal metabolome
573 and its determinants in inflammatory bowel disease. *Gut*, 72(8):1472–1485, March
574 2023.
- 575 [46] Marius Vital, Tatjana Rud, Silke Rath, Dietmar H. Pieper, and Dirk Schlüter. Diversity
576 of bacteria exhibiting bile acid-inducible 7-dehydroxylation genes in the human gut.
577 *Computational and Structural Biotechnology Journal*, 17:1016–1019, 2019.

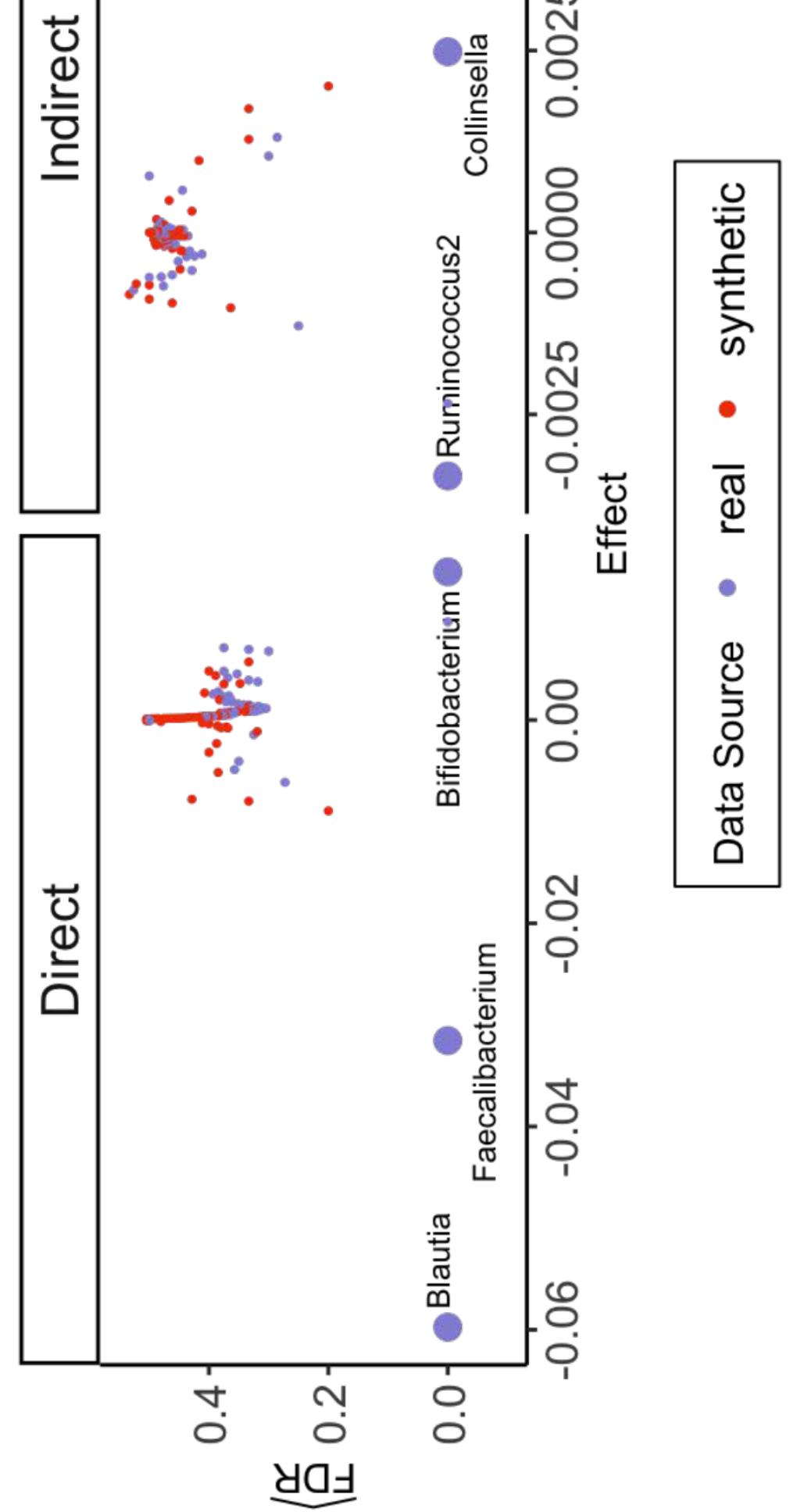
- 578 [47] Jolana Wagner-Skacel, Nina Dalkner, Sabrina Moerkl, Kathrin Kreuzer, Aitak Farzi,
579 Sonja Lackner, Annamaria Painold, Eva Z Reininghaus, Mary I Butler, and Susanne
580 Bengesser. Sleep and microbiome in psychiatric diseases. *Nutrients*, 12(8):2198, 2020.
- 581 [48] Chan Wang, Jiyuan Hu, Martin J Blaser, and Huilin Li. Estimating and testing the
582 microbial causal mediation effect with high-dimensional and compositional
583 microbiome data. *Bioinformatics*, 36(2):347–355, July 2019.
- 584 [49] Hadley Wickham. *ggplot2 - elegant graphics for data analysis*. In *Use R!*, 2009.
- 585 [50] Hadley Wickham. *Advanced R*. Chapman and Hall/CRC, September 2014.
- 586 [51] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan,
587 Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max
588 Kuhn, Thomas Pedersen, Evan Miller, Stephanie Bache, Kirill Müller, Jeroen Ooms,
589 David G. Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis
590 Vaughan, Claus Wilke, Kara H. Woo, and Hiroaki Yutani. Welcome to the tidyverse. *J.*
591 *Open Source Softw.*, 4:1686, 2019.
- 592 [52] Jenessa A. Winston and Casey M. Theriot. Diversification of host bile acids by
593 members of the gut microbiota. *Gut Microbes*, 11(2):158–171, October 2019.
- 594 [53] Marvin N. Wright and Andreas Ziegler. ranger: A fast implementation of random
595 forests for high dimensional data in c++ and r. *Journal of Statistical Software*, 077:1–17,
596 2015.
- 597 [54] Fan Xia, Jun Chen, Wing Kam Fung, and Hongzhe Li. A logistic normal multinomial
598 regression model for microbiome compositional data analysis. *Biometrics*, 69, 2013.
- 599 [55] Dongyang Yang and Wei Xu. Estimation of mediation effect on zero-inflated
600 microbiome mediators. *Mathematics*, 11(13):2830, June 2023.
- 601 [56] Haixiang Zhang, Jun Chen, Zhigang Li, and Lei Liu. Testing for mediation effect
602 with application to human microbiome data. *Statistics in Biosciences*, 13(2):313–328,
603 July 2019.

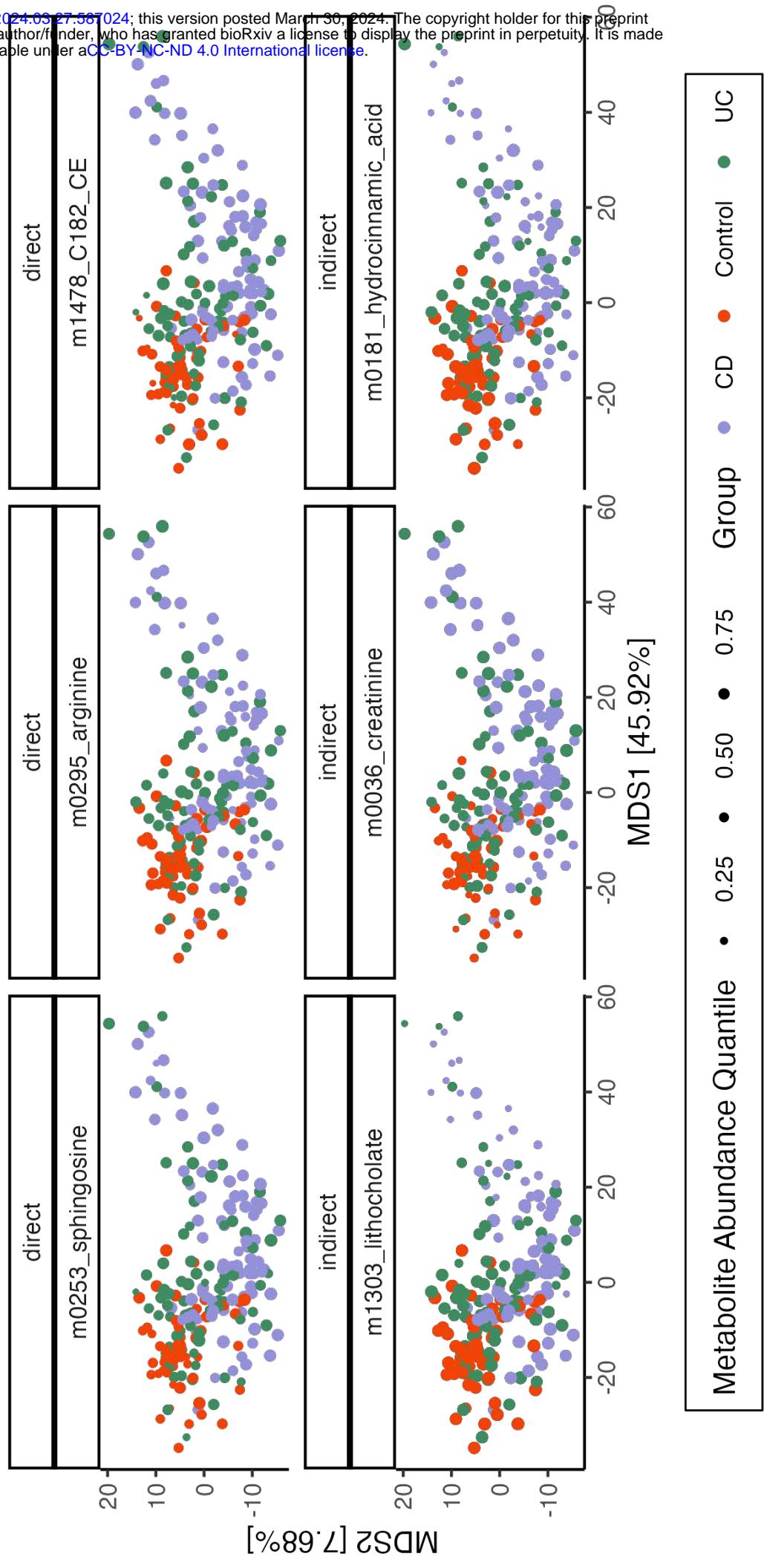


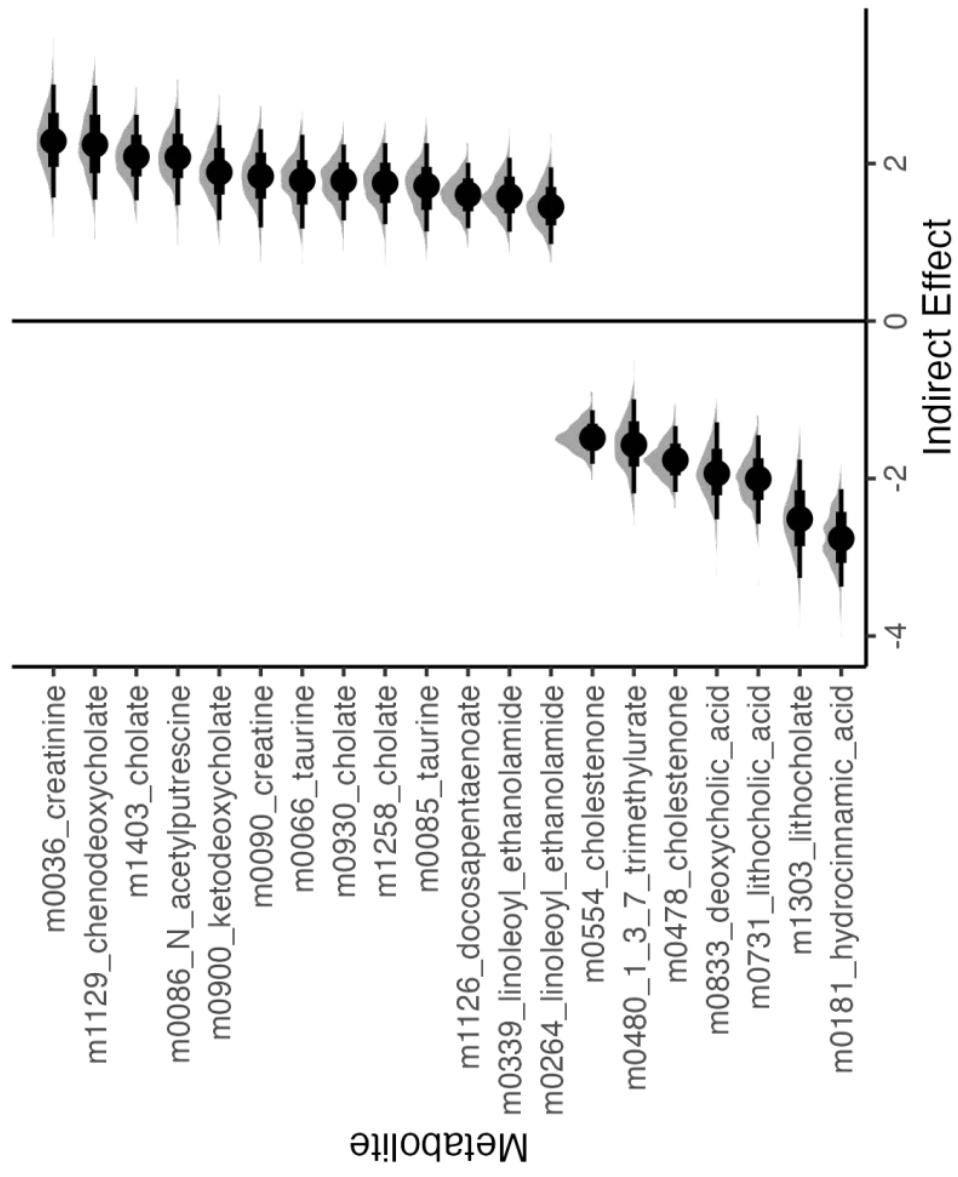
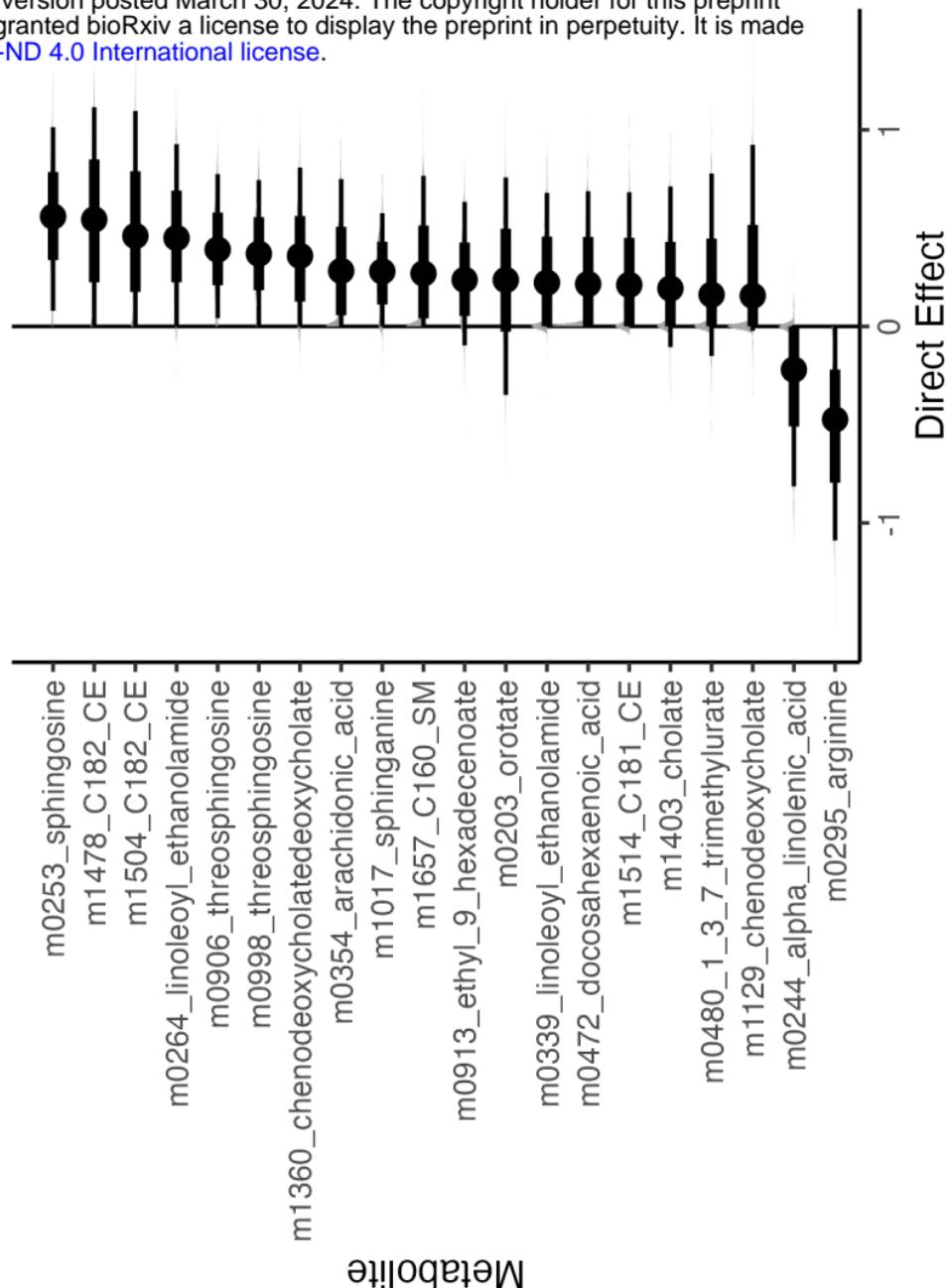


Study.Group ● Control ● CD ● UC

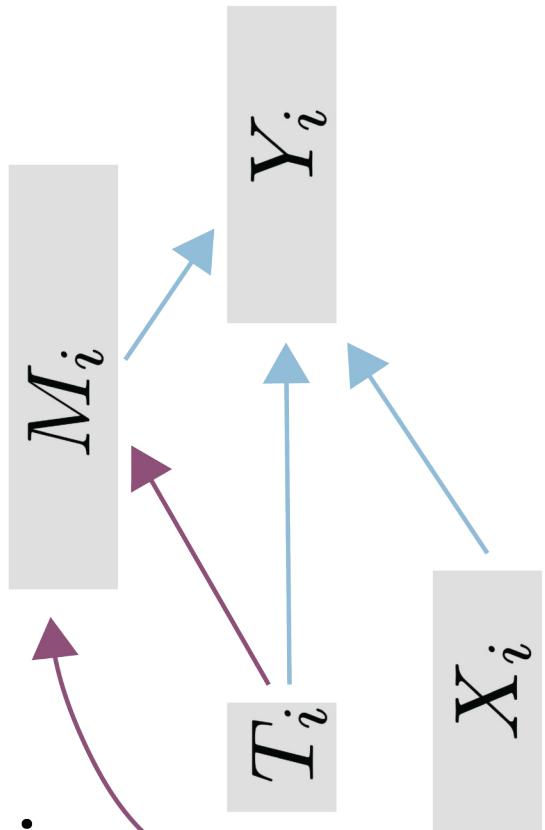




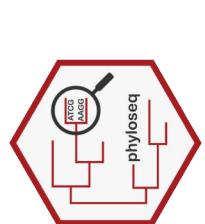




A.



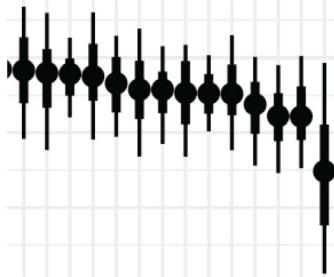
Data Curation



B.

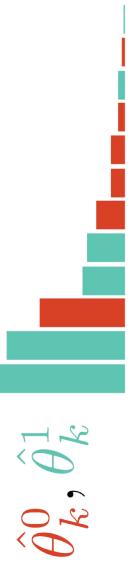
Bootstrap
Confidence Intervals

$$\theta_k \in [\hat{\theta}_k^L, \hat{\theta}_k^U]$$

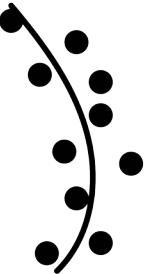


Inference

Synthetic Null
Hypothesis Testing



Fit Mediation and
Outcome Models



Fit Mediation and
Outcome Models