

1 **Title:** Microbial community-scale metabolic modeling predicts personalized short-chain-fatty-
2 acid production profiles in the human gut.

3
4 **Authors:** Nick Bohmann^{1,2}, Tomasz Wilmanski¹, Lisa Levy³, Johanna W. Lampe³, Thomas
5 Gurry^{4,5}, Noa Rappaport¹, Christian Diener^{1,*}, Sean M. Gibbons^{1,2,6,7,*}

6
7 **Affiliations:** ¹ Institute for Systems Biology, Seattle, WA 98109, USA; ² Molecular Engineering
8 Graduate Program, University of Washington, Seattle, WA 98195, USA; ³ Fred Hutchinson
9 Cancer Center, Seattle, WA 98109, USA; ⁴ Pharmaceutical Biochemistry Group, School of
10 Pharmaceutical Sciences, University of Geneva, Switzerland; ⁵ Myota GmbH, Berlin, Germany; ⁶
11 Departments of Bioengineering and Genome Sciences, University of Washington, Seattle, WA
12 98195, USA; ⁷ eScience Institute, University of Washington, Seattle, WA 98195, USA; *
13 correspondence can be addressed to cdiener@isbscience.org and sgibbons@isbscience.org

14 15 **Abstract**

16 Microbially-derived short-chain fatty acids (SCFAs) in the human gut are tightly coupled to host
17 metabolism, immune regulation, and integrity of the intestinal epithelium. However, the
18 production of SCFAs can vary widely between individuals consuming the same diet, with lower
19 levels often associated with disease. A mechanistic understanding of this heterogeneity is
20 lacking. We present a microbial community-scale metabolic modeling (MCMM) approach to
21 predict individual-specific SCFA production profiles. We assess the quantitative accuracy of our
22 MCMMs using *in vitro*, *ex vivo*, and *in vivo* data. Next, we identify associations between MCMM
23 SCFA predictions and a panel of blood-based clinical chemistries in a large human cohort.
24 Finally, we demonstrate how MCMMs can be leveraged to design personalized dietary,
25 prebiotic, and probiotic interventions that optimize SCFA production in the gut. Our results
26 represent an important advance in engineering gut microbiome functional outputs for precision
27 health and nutrition.

28 29 **Keywords**

30 gut microbiome, short chain fatty acids, flux balance analysis, metabolic model, precision
31 nutrition

32
33

34

35

36

37

38

39

40

41

42 Introduction

43 The human gut microbiota maintains intestinal barrier function, regulates peripheral and
44 systemic inflammation, and breaks down indigestible dietary components and host substrates
45 into a wide range of bioactive compounds^{1,2}. One of the primary mechanisms by which the gut
46 microbiota impacts human health is through the production of small molecules that enter the
47 circulation and are absorbed and transformed by host tissues³⁻⁵. Approximately half of the
48 metabolites detected in human blood are known to be significantly associated with cross-
49 sectional variation in gut microbiome composition⁶.

50 Short-chain-fatty-acids (SCFAs) are among the most abundant metabolic byproducts
51 produced by the gut microbiota, largely through the fermentation of indigestible dietary fibers
52 and resistant starches, with acetate, propionate and butyrate being the most abundant SCFAs⁷⁻
53 ⁹. Deficits in SCFA production have been repeatedly associated with disease^{10,11}. Therefore,
54 SCFA production is a crucial ecosystem service that the gut microbiota provides to its host, with
55 far-reaching impacts on health^{1,11-13}. However, different human gut microbiota provided with the
56 same exact dietary substrate can show variable SCFA production profiles^{14,15}, and predicting
57 this heterogeneity remains a fundamental challenge to the microbiome field. Measuring SCFA
58 abundances in blood or feces is rarely informative of *in situ* production rates, due to the volatility
59 of SCFAs, cross-feeding among microbes, and the rapid consumption and transformation of
60 these metabolites by the colonic epithelium^{10,16,17}. Furthermore, SCFA production fluxes (i.e.,
61 the amount of a metabolite produced over a given period of time) within an individual can vary
62 longitudinally, depending upon dietary inputs and the availability of host substrates¹⁸. In order to
63 account for this inter- and intra-individual heterogeneity, we propose the use of microbial
64 community-scale metabolic models (MCMs), which mechanistically account for metabolic
65 interactions between gut microbes, host substrates, and dietary inputs, to estimate
66 personalized, context-specific SCFA production profiles.

67 Statistical modeling and machine-learning approaches for predicting metabolic output
68 from the microbiome have shown promising results in recent years. For example, postprandial
69 blood glucose responses can be predicted by machine-learning algorithms trained on large
70 human cohorts ^{19,20}. However, machine-learning methods are limited by the measurements and
71 interventions represented within the training data ²¹. Mechanistic models like MCMMs, on the
72 other hand, do not rely on training data and provide causal insights ¹⁷. MCMMs are constructed
73 using existing knowledge bases, including curated genome-scale metabolic models (GEMs) of
74 individual taxa ²². MCMMs can be limited by the inability to find well-curated GEMs for abundant
75 taxa present in certain samples, and this underrepresentation in GEMs tends to be worse in
76 human populations that are generally underrepresented in microbiome research ²³. Despite
77 this, MCMMs can be powerful, transparent, knowledge-driven tools for predicting community-
78 specific responses to a wide array of interventions or perturbations.

79 Here, we demonstrate the utility of MCMMs for the prediction of personalized SCFA
80 production profiles in the context of different dietary, prebiotic, and probiotic inputs. We first
81 validate our modeling platform using synthetic *in vitro* gut microbial communities (N=1,387) and
82 *ex vivo* stool incubation assays (N=21). Next, we investigate the relevance of this modeling
83 strategy *in vivo* using data from a 10-week high-fiber dietary intervention cohort (N=18), where
84 individuals showed a variety of immune responses. We assess the clinical significance of these
85 precision SCFA predictions by looking at associations between predicted SCFA production on
86 an average European diet and a panel of blood-based clinical lab tests in a large human cohort
87 (N=2,687). Finally, we demonstrate the power of MCMMs in designing personalized prebiotic,
88 probiotic, and dietary interventions that optimize individual-specific butyrate production rates.

89

90 **Results**

91 *MCMMs capture SCFA production rates in vitro*

92 We sought to investigate whether MCMs can predict production rates of the major SCFAs
93 (i.e., acetate, propionate, and butyrate) under controlled experimental conditions (**Fig. 1**). We
94 assembled microbial community models for *in vitro* data sets spanning 4 independent studies
95 with varying levels of complexity. Models were constructed by combining manually-curated
96 GEMs from the AGORA database ²⁴, constraining taxon abundances using 16S amplicon
97 sequencing relative abundance estimates, and applying an appropriate growth medium.
98 Sample-specific metabolic models were then solved using cooperative tradeoff flux balance
99 analysis (ctFBA), a previously-reported two-step quadratic optimization strategy that yields
100 empirically-validated estimates of the steady state growth rates and metabolic uptake and
101 secretion fluxes for each taxon in the model ¹⁷ (see Materials and Methods). Models were
102 summarized at the genus level, which was the finest level of phylogenetic resolution that the
103 16S data allowed for.

104 First, we looked at data from synthetically constructed *in vitro* cultures of human gut
105 microbial communities obtained from a recent publication ²⁵. This data set included
106 measurements of relative microbial abundances, butyrate production levels, and the overall
107 optical density for each of 1,387 independent co-cultures (**Fig. 2A**). Cultures varied in richness
108 from 1-25 strains. MCMs were constructed for each co-culture as described above, simulating
109 growth of each of the models using a defined, componentized medium, matching the
110 composition of the medium used in the *in vitro* experiments (see Material and Methods). Model-
111 predicted fluxes of butyrate were compared with measured butyrate production rates normalized
112 to the OD600, stratifying results into low richness (1-5 genera) and high richness (10-25 genera)
113 communities. Model predictions for butyrate production fluxes were significantly associated with
114 measured butyrate production fluxes (Pearson's correlation; Low Richness: $R^2=0.028$, $p=5e-4$;
115 High Richness: $R^2=0.277$, $p=6e-51$), but predictions were more accurate in the higher richness
116 communities (**Fig. 2B-C**).

117 Next, we compared MCMM predictions to anaerobic *ex vivo* incubations of human stool
118 samples from a small number of individuals (N=21) cultured after supplementation with sterile
119 PBS buffer or with different dietary fibers across three independent studies. Study A contained
120 samples from two donors cultured for 7 hours, Study B ¹⁴ contained samples from 10 donors
121 cultured for 24 hours, and Study C contained samples from 9 donors cultured for 4 hours. Fecal
122 *ex vivo* assays allow for the direct measurement of bacterial SCFA production fluxes without
123 interference from the host. For all three studies, *ex vivo* incubations were performed by
124 homogenizing fecal material in sterile buffer under anaerobic conditions, adding control or fiber
125 interventions to replicate fecal slurries, and measuring the resulting SCFA production rates *in*
126 *vitro* at 37°C (see Materials and Methods). Metagenomic or 16S amplicon sequencing data from
127 these *ex vivo* cultures were used to construct MCMMs, using relative abundances obtained from
128 sequencing data as a proxy for relative biomass for each bacterial genus (see Materials and
129 Methods). MCMMs were simulated using a diluted standardized European diet (i.e., to
130 approximate residual dietary substrates still present in the stool slurry), with or without specific
131 fiber amendments, matching the experimental treatments (see Material and Methods). The
132 resulting SCFA flux predictions were then compared to the measured fluxes. We observed an
133 agreement between MCMM-predicted and measured SCFA production fluxes across all three
134 *ex vivo* data sets (**Fig. 3**). Models predicted significantly higher SCFA production fluxes for fiber-
135 treated samples across all studies (Independent Student's t-test, $p < 0.05$; Study A was omitted
136 from this analysis due to low sample size, although the separation between controls and fiber-
137 treated samples is visually apparent). The same held true for measured SCFA fluxes, with the
138 exception of acetate in Study C (i.e., the study with the shortest incubation time), where there
139 was not always significant separation in measured SCFA production between control and fiber
140 treatments (**Fig. 3G-I**). With one exception (**Fig. 3J**), a significant positive correlation was
141 observed between predicted and measured SCFA fluxes across treatment groups for all three

142 studies ($R^2=0.22-0.99$, Pearson test, $p<0.05$). In summary, we observed agreement between
143 MCMM-predicted and measured *in vitro* SCFA production rates in the presence or absence of
144 fiber supplementation, with better agreement in more diverse communities and over longer
145 experimental incubation times (**Fig. 2-3**).

146

147 *MCMM predictions correspond with variable immunological responses to a 10-week high-fiber*
148 *dietary intervention*

149 We next investigated whether MCMM-predicted SCFA production rates could be leveraged to
150 help explain inter-individual differences in phenotypic response following a dietary intervention.
151 Specifically, we looked at data from 18 individuals who were placed on a high-fiber diet for ten
152 weeks²⁶. These individuals fell into three distinct immunological response groups: one in which
153 high inflammation was observed over the course of the intervention (high-inflammation group),
154 and two other distinct response groups that both exhibited lower levels of inflammation (low-
155 inflammation groups I and II; **Fig. 4A**). We hypothesized that these immune response groups
156 could be explained, in part, by differences in MCMM-predicted SCFA production profiles. Using
157 16S amplicon sequencing data from seven time points collected from each of these 18
158 individuals over the 10-week intervention, we built MCMMs for each study participant at each
159 time point in the study. Growth was then simulated for each model using a standardized high-
160 fiber diet, rich in resistant starch (see Material and Methods). Throughout the study, individuals
161 in the high-inflammation group showed significantly lower predicted butyrate plus propionate
162 production on average (i.e., the two SCFAs with the strongest anti-inflammatory effects),
163 compared to the individuals in each of the low-inflammation groups (High vs. Low I: 284.6 ± 7.7
164 vs 327.5 ± 3.8 mmol/(gDW h) on average, Mann-Whitney $p = 1.9e-5$. High vs. Low II: $284.6 \pm$
165 7.7 vs 337.8 ± 6.4 mmol/(gDW h), Mann-Whitney $p = 7.2e-6$) (**Fig. 4B**). Predicted levels of
166 butyrate plus propionate production in the high-inflammation group decreased throughout the

167 duration of the high-fiber intervention (Pearson $r = -0.47$, Pearson test, $p = 4.7e-3$) (**Fig. 4C**),
168 while predicted levels of butyrate and propionate production in the low-inflammation groups
169 were constant over time (Low I: Pearson $r = -0.020$, Pearson test $p = 0.90$, Low II: Pearson $r =$
170 0.093 , Pearson test, $p = 0.57$) (**Fig. 4C**). Acetate production did not appear to differ across
171 groups (High vs. Low I: 652.7 ± 52.4 vs 639.8 ± 79.3 mmol/(gDW h) on average, Mann-Whitney
172 $p = .99$. High vs. Low II: 652.7 ± 52.4 vs 653.8 ± 50.0 mmol/(gDW h), Mann-Whitney $p = .97$.
173 Low I vs. Low II: 652.7 ± 52.4 vs 639.8 ± 79.3 mmol/(gDW h), Mann-Whitney $p = .80$) (**Fig.**
174 **4D**), although there was a slight trend towards increasing acetate production over time in the
175 high inflammation group (Pearson $r = 0.18$, $p = 0.31$) (**Fig. 4E**).

176

177 *MCMC-predicted SCFA profiles are associated with a wide range of blood-based clinical*
178 *markers*

179 To further evaluate the clinical relevance of personalized MCMCs, we generated SCFA
180 production rate predictions from stool 16S amplicon sequencing data for 2,687 individuals in a
181 deeply phenotyped, generally-healthy cohort from the West Coast of the United States (i.e., the
182 Arivale cohort)²⁷. Baseline MCMCs were built for each individual assuming the same dietary
183 input (i.e., an average European diet) in order to compare SCFA production rate differences,
184 independent of background dietary variation. MCMC-predicted SCFA fluxes were then
185 regressed against a panel of 128 clinical chemistries and health metrics collected from each
186 individual, adjusting for a standard set of common covariates (i.e., age, sex, and microbiome
187 sequencing vendor) (**Fig. 5A**). After FDR correction, 37 markers were significantly associated
188 with the predicted production rate of at least one SCFA (**Fig. 5B**). Predicted butyrate production
189 showed significant positive association with the health-associated hormone adiponectin, and
190 significant inverse association with 11 metabolites associated with poorer health, including C-
191 reactive protein (CRP), HOMA-IR, and low-density lipoprotein (LDL; $P < 0.05$, FDR-corrected

192 Wald test). Acetate showed significant positive associations with 23 blood metabolites and
193 inverse associations with 11 metabolites (**Fig. 5B**), which tended to be in the opposite direction
194 as the butyrate associations. Propionate showed no significant associations, while overall SCFA
195 production showed 2 positive and 3 negative associations (**Fig. 5B**). Butyrate and propionate
196 production tended to be positively correlated within an individual, while higher acetate
197 production was inversely associated with both butyrate and propionate production (**Fig. 5C-D**).
198 This inverse association may be responsible, in part, for the flipped associations with clinical
199 chemistries between butyrate and acetate.

200

201

202 *Leveraging MCMs to design precision dietary, prebiotic, and probiotic interventions*

203

204 As a proof-of-concept for *in silico* engineering of the metabolic outputs of the human gut
205 microbiome, we screened a set of potential interventions designed to increase SCFA production
206 for individuals from the Arivale cohort (**Fig. 6A**). MCMs were built using two different dietary
207 contexts: an average European diet, and a vegan, high-fiber diet (see Material and Methods).
208 Predicted butyrate production rates were then compared across the two diets. As expected,
209 models grown on a high-fiber diet showed higher average predicted butyrate production: 27.35
210 ± 6.77 mmol/(gDW h) vs 16.17 ± 6.22 mmol/(gDW h), paired t-test, $t = 92.74$, $p < .001$ (**Fig. 6B**).
211 However, this increase in butyrate production between the European and high-fiber diets was
212 not uniform across individuals. On the high-fiber diet, some individual gut microbiota
213 compositions showed very large increases in butyrate production, some showed little-to-no
214 change, and a small subset of samples actually showed a decrease in butyrate production,
215 relative to the European diet. We identified a set of ‘non-responders’ ($n = 29$) who produced less
216 than $10 \frac{\square\square\square\square}{\square\square\square*\square}$ of butyrate on the European diet and showed an increase in butyrate production

217 of less than 20% on the high-fiber diet (**Fig. 6C**). We also identified a set of ‘regressors’ (n = 45)
218 who produced more than 20 $\frac{\square\square\square\square}{\square\square\square*\square}$ of butyrate on the European diet and showed lower butyrate
219 production on the high-fiber diet (**Fig. 6D**). We then simulated a handful of simple prebiotic and
220 probiotic interventions across these individuals, to identify optimal intervention combinations for
221 each individual (**Fig. 6C-E**). MCMMs for each subset of individuals were simulated with prebiotic
222 and probiotic interventions in the context of either the European or the high-fiber diet.
223 Specifically, diets were supplemented with the dietary fiber inulin, with the dietary fiber pectin, or
224 with a simulated probiotic intervention that consisted of introducing 10% relative abundance of
225 the butyrate-producing genus *Faecalibacterium* to the MCMM. In general, optimal combination
226 interventions significantly increased the population-level butyrate production above either
227 dietary intervention alone (**Fig. 6C-D**).

228 For 70/74 individuals, supplementation of the background diet with a specific pre- or
229 probiotic increased the butyrate production rate (**Fig. 6C-E**). Neither response group had an
230 intervention that was optimal for all individuals. In general, the most successful intervention for
231 non-responders was the addition of inulin to the European diet (156.6% \pm 183.3% increase vs
232 standard European diet), and for regressors it was the addition of inulin to the high-fiber diet
233 (88.2% \pm 75.7% increase). However, the exact intervention that yielded the highest butyrate
234 production for any given individual across both populations varied widely (**Fig. 6E**). For
235 example, the probiotic intervention was more successful in raising predictions for butyrate
236 production in non-responders than it was in regressors (**Fig. 6E**). The optimal intervention
237 combination in the non-responder subpopulation was more heterogeneous than in the regressor
238 subpopulation. Overall, no single intervention combination was optimal for every individual in the
239 population.

240

241 **Discussion**

242 Here we present an approach to the rational engineering of SCFA production rates from the
243 human gut microbiome through prebiotic, probiotic, and dietary interventions, validated using *in*
244 *vitro* and *in vivo* experimental data. We demonstrated the MCMMs can be used to formulate
245 personalized interventions designed to optimize SCFA production profiles.

246 Model predictions of butyrate production in synthetically constructed *in vitro* co-cultures
247 showed significant agreement between measured and predicted butyrate fluxes (**Fig. 2**). Due to
248 the phylogenetic resolution of 16S data and the lack of strain-level GEMs that match the
249 organisms present in some samples, we built genus-level MCMMs for all analyses. The
250 decreasing accuracy of butyrate predictions as community richness declined may reflect a
251 limitation of building models at the genus-level, as reconstructions contain a summarized
252 aggregation of the metabolic capability of the genus as a whole, without species- or strain-level
253 resolution. Consequently, pathways included in the metabolic model may be absent in a low-
254 richness experimental system, reflecting a mismatch between the modeling framework and
255 reality. In high richness models, predictions became more accurate, indicating this mismatch is
256 less impactful as more taxa are included in the system. Real-world microbiomes are often more
257 species-rich than synthetic *in vitro* communities. Fortunately, we are likely not operating in a
258 community regime where missingness in individual models has a large influence on genus-level
259 MCMM SCFA predictions in the human gut. However, future work should focus on increasing
260 the availability of diverse strain-level GEMs. The recent release of AGORA2²⁸, containing GEMs
261 for 7,302 microorganisms, may help to overcome this limitation and aid in the construction of
262 MCMMs at finer levels of taxonomic resolution.

263 Data from *ex vivo* anaerobic fecal incubations showed agreement between SCFA flux
264 predictions and measurements. Fiber-treated samples showed significant increases in both
265 predicted and measured SCFA production, compared to the controls (**Fig. 3**). Additionally,
266 measured and predicted production rates of butyrate and propionate showed quantitative

267 agreement across three independent studies. Acetate production rates were accurately
268 predicted in all but one study. Acetate is known to act as an overflow metabolite^{29,30}, with a
269 wide range of possible fluxes for a given biomass optimum, so it is perhaps not surprising that
270 the predictions for this metabolite tended to be less accurate. Finally, predictions were generally
271 the weakest in Study C (**Fig. 3G-I**). One possible reason for poorer predictions in this study is
272 that the incubation time was shorter than for the other two studies (4 hours in Study C, vs 7
273 hours and 24 hours in Studies A and B, respectively), resulting in less divergence in
274 accumulated SCFA concentrations between controls and treated samples. Thus, it is likely that
275 SCFA levels did not build up to high enough levels in this study to accurately reflect *in situ*
276 production. Overall, the observed correspondence between our SCFA production profile
277 predictions and *in vitro* data provided us with some confidence in our MCMC modeling strategy
278 and prompted us to explore how these predictions might be applied to an *in vivo* setting.

279 MCMCs built using data from a 10-week high-fiber dietary intervention allowed us to
280 assess our predictions *in vivo* in the context of immunological responses to diet. Predictions of
281 combined butyrate and propionate (i.e., SCFAs with the strongest anti-inflammatory effects on
282 the host³¹) production rates over the course of the high-fiber intervention showed distinct
283 differences between pre-defined immune response groups (**Fig. 4**). The low-inflammation
284 groups showed stable butyrate and propionate production fluxes over time, while the high
285 inflammation group showed lower average butyrate and propionate production and a decreasing
286 production rate over time following the initiation of the high-fiber diet (**Fig. 4B-C**). Given the
287 strong anti-inflammatory effects of butyrate and propionate, we expected to see lower
288 production of these molecules in the context of higher inflammation. Prior work has shown that
289 higher doses of inulin can actually induce an inflammatory response, which may explain, in part,
290 the inflammatory immune response in these individuals³². Overall, our results indicate that

291 different immunological responses to a high fiber diet may be explained, in part, by the observed
292 heterogeneity in MCMM-predicted butyrate and propionate production rates (**Fig. 4**).

293 Several biomarkers of metabolic health, inflammation, liver function, and cardiovascular
294 health were associated with MCMM-predicted SCFA production profiles in a large, generally-
295 healthy cohort, assuming an average European diet (**Fig. 5**). CRP, a marker of systemic
296 inflammation³³, showed a significant negative association with butyrate production predictions.
297 Markers of cardiovascular health, such as LDL and triglyceride levels, were also negatively
298 associated with butyrate production rates, supporting the role of butyrate as protective against
299 cardiovascular disease³⁴. Many significant associations were inverted when looking across
300 butyrate and acetate flux predictions (**Fig. 5**). For instance, CRP, HOMA-IR, glucose, insulin,
301 LDL cholesterol and uric acid all showed significant negative associations with butyrate
302 production and significant positive associations with acetate production. This result may be
303 related to the apparent tradeoff between acetate production and the production of both butyrate
304 and propionate (**Fig. 5C-E**). While the overall production of SCFAs has been implicated in
305 lowering inflammation³⁵, the potency of butyrate in driving down inflammation and improving
306 overall metabolic health is greater than that of acetate³⁶, which, given this apparent tradeoff in
307 the production of these different SCFAs, could help to explain these inverted associations.

308 Given this set of promising associations between SCFA predictions and host phenotypic
309 variation, we next wanted to demonstrate the potential of MCMMs for designing precision
310 prebiotic, probiotic, and dietary interventions that optimize SCFA production profiles. Using the
311 Arivale cohort, we identified two classes of individuals that responded differently to an *in silico*
312 high-fiber dietary intervention: non-responders and regressors (**Fig. 6**). We found significant
313 heterogeneity in the optimal intervention across individuals from each of these response groups,
314 but most notably in the non-responders (**Fig. 6E**). Given that the non-responders had low
315 baseline levels of butyrate production and did not respond to a high-fiber diet, this underscores

316 the importance of personalized predictions for those who tend not to respond well to population-
317 scale interventions.

318 Personalized prediction of SCFA production profiles from human gut MCMs represents
319 an important technological step forward in leveraging artificial intelligence for precision nutrition.
320 Mechanistic modeling allowed us to translate the ecological composition of the gut microbiome
321 into concrete, individual-specific metabolic outputs, in response to specific interventions³⁷.
322 MCMs are transparent models that do not require training data, with clear causal and
323 mechanistic explanations behind each prediction. Microbially-produced metabolites have an
324 substantial impact on host physiology and health^{38,39}, and a rational framework for engineering
325 the production or consumption rates of these metabolites has broad potential applications in
326 precision nutrition and personalized healthcare.

327

328

329 **Materials and Methods**

330 *In vitro culturing*

331 Culturing of the synthetically assembled gut microbial communities is described in Clark et al.,
332 2021²⁵. Culturing of *ex vivo* samples in Study A was done using the methodology described
333 below. Culturing of *ex vivo* samples in Study B is described in Cantu-Jungles et al., 2021¹⁴.
334 Culturing of *ex vivo* samples in Study C was conducted by co-author Dr. Thomas Gurry, using
335 the methodology described below.

336

337 *In vitro culturing of fecal-derived microbial communities (Study A)*

338 Fecal samples were collected in 1200 mL 2-piece specimen collectors (Medline, USA) in the
339 Public Health Science Division of the Fred Hutchinson Cancer Center (IRB Protocol number
340 5722) and transferred into an large vinyl anaerobic chamber (Coy, USA, 37°C, 5% hydrogen,

341 20% carbon dioxide, balanced with nitrogen) at the Institute for Systems Biology within 20
342 minutes of defecation. All further processing and sampling was then run inside the anaerobic
343 chamber. 50 g of fecal material was transferred into sterile 50 oz Filter Whirl-Paks (Nasco, USA)
344 with sterile PBS + 0.1% L-cysteine at a 1:2.5 w/v ratio and homogenized with a Stomacher
345 Biomaster (Seward, USA) for 15 minutes. After homogenization, each sample was transferred
346 into three sterile 250 mL serum bottles and another 2.5 parts of PBS + 0.1% L-cysteine was
347 added to bring the final dilution to 1:5 in PBS. 87 ug/mL inulin or an equal volume of sterile PBS
348 buffer were added to treatment or control bottles, respectively. Samples were immediately
349 pipetted onto sterile round-bottom 2 mL 96-well plates in triplicates. Baseline samples were
350 aliquoted into sterile 1.5 mL Eppendorf tubes and the plates were covered with Breathe-Easy
351 films (USA Scientific Inc., USA). Plates were incubated for 7 h at 37°C and gently vortexed
352 every hour within the chamber. Final samples were aliquoted into 1.5 mL Eppendorf tubes at the
353 end of incubation. Baseline and 7 h samples were kept on ice and immediately processed after
354 sampling. 500 uL of each sample were aliquoted for metagenomics and kept frozen at -80°C
355 before and during transfer to the commercial sequencing service (Diversigen, Inc). The
356 remaining sample was transferred to a table-top centrifuge (Fisher Scientific accuSpin, USA)
357 and spun at 1,500 rpm for 10 minutes. The supernatant was then transferred to collection tubes
358 kept on dry ice and transferred to the commercial metabolomics provider Metabolon, USA, for
359 targeted SCFA quantification.

360

361 *In vitro culturing of fecal-derived microbial communities (Study C)*

362 Homogenized fecal samples in this study again underwent anaerobic culturing at 37°C, as
363 described above, but with a shorter culturing time of 4 hours. The slurry was diluted 2.5x in
364 0.1% L-cysteine PBS buffer solution. Cultures were supplemented with the dietary fibers pectin
365 or inulin to a final concentration of 10g/L, or a sterile PBS buffer control treatment. Aliquots were

366 taken at 0h and 4h and further processed for measurement of SCFA concentrations, which were
367 used to estimate experimental production flux (concentration[4h] - concentration[0h]/4h). SCFA
368 concentrations were measured using GC-FID. Briefly, the pH of the aliquots was adjusted to 2-3
369 with 1% aqueous sulfuric acid solution, after which they were vortexed for 10 minutes and
370 centrifuged for 10 minutes at 10,000 rpm. 200 uL aliquots of clear supernatant were transferred
371 to vials containing 200 uL of MeCN and 100 uL of a 0.1% v/v 2-methyl pentanoic acid solution.
372 The resulting solutions were analyzed by GC-FID on a Perkin Elmer Clarus 500 equipped with a
373 DB-FFAP column (30m, 0.250mm diameter, 0.25um film) and a flame ionization detector.

374

375 *Metagenomic sequencing and analysis*

376 For Study A, shallow metagenomic sequencing was performed by the sequencing vendor
377 Diversigen, USA (i.e., their BoosterShot service). In brief, DNA was extracted from the fecal
378 slurries with the DNeasy PowerSoil Pro Kit on a QiaCube HT (Qiagen, Germany) and quantified
379 using the Qiant-iT Picogreen dsDNA Assay (Invitrogen, USA). Library preparation was
380 performed with a proprietary protocol based on the Nextera Library Prep kit (Illumina, USA) and
381 the generated libraries were sequenced on a NovaSeq (Illumina, USA) with a single-end 100bp
382 protocol. Demultiplexing was performed using Illumina BaseSpace to generate the final FASTQ
383 files used during analysis.

384 Preprocessing of raw sequencing reads was performed using FASTP⁴⁰. The first 5bp on
385 the 5' end of each read were trimmed, and the 3' end was trimmed using a sliding window
386 quality filter that would trim the read as soon as the average window quality fell below 20. Reads
387 containing ambiguous base calls or with a length of less than 15bp after trimming were removed
388 from the analysis.

389 Bacterial species abundances were quantified using Kraken2 v2.0.8 and Bracken v2.2
390 using the Kraken2 default database which was based on Refseq release 94, retaining only

391 those species with at least 10 assigned reads^{41,42}. The analysis pipeline can be found at
392 https://github.com/Gibbons-Lab/pipelines/tree/master/shallow_shotgun.

393

394 *Metabolomics*

395 Targeted metabolomics were performed using Metabolon's high-performance liquid
396 chromatography (HPLC)–mass spectrometry (MS) platform, as described before⁴³. In brief,
397 fecal supernatants were thawed on ice, proteins were removed using aqueous methanol
398 extraction, and organic solvents were removed with a TurboVap (Zymark, USA). Mass
399 spectroscopy was performed using a Waters ACQUITY ultra-performance liquid
400 chromatography (UPLC) and Thermo Scientific Q-Exactive high resolution/accuracy mass
401 spectrometer interfaced with a heated electrospray ionization (HESI-II) source and an Orbitrap
402 mass analyzer operated at 35,000 mass resolution. For targeted metabolomics ultra-pure
403 standards of the desired short-chain fatty acids were used for absolute quantification. Fluxes for
404 individual metabolites were estimated as the rate of change of individual metabolites during the
405 incubation period (concentration[7h] - concentration[0h]/7h).

406

407 *Model Construction*

408 Taxonomic abundance data summarized to the genus level, inferred from 16S amplicon
409 sequencing or shotgun metagenomic sequencing, were used to construct all MCMMs in this
410 analysis using the community-scale metabolic modeling platform MICOM v0.32.3¹⁷. Models
411 were built using the MICOM build() function with a relative abundance threshold of 0.001,
412 omitting taxa that made up less than 0.1% relative abundance. The AGORA database (v1.03) of
413 taxonomic reconstructions summarized to the genus level was used to collect genome-scale
414 metabolic models for taxa present in each model. *In silico* media were applied to the grow()
415 function, defining the bounds for metabolic imports by the MCMM. Medium composition varied

416 between analyses (see *Media Construction*). Steady state growth rates and fluxes for all
417 samples were then inferred using cooperative tradeoff flux balance analysis (ctFBA). In brief,
418 this is a two-step optimization scheme, where the first step finds the largest possible biomass
419 production rate for the full microbial community and the second step infers taxon-specific growth
420 rates and fluxes, while maintaining community growth within a fraction of the theoretical
421 maximum (i.e., the tradeoff parameter), thus balancing individual growth rates and the
422 community-wide growth rate¹⁷. For all models in the manuscript we used a tradeoff parameter
423 of 0.7. This parameter value was chosen through cooperative tradeoff analysis in MICOM.
424 Multiple parameters were tested, and the highest parameter value (i.e., the value closest to the
425 maximal community growth rate at 1.0) that allowed most (>90%) of taxa to grow was chosen
426 (i.e., 0.7). Predicted growth rates from the simulation were analyzed to validate correct behavior
427 of the models. All models were found to grow with minimum community growth rate of 0.3 h⁻¹.
428 Predicted values for export fluxes of SCFAs were collected from each MCMM using the
429 `production_rates()` function, which calculates the overall production from the community that
430 would be accessible to the colonic epithelium.

431

432 *Media Construction*

433 Individual media were constructed based on the context of each individual analysis. For the
434 synthetic *in vitro* cultures conducted by Clark et al. (2021), a defined medium (DM38) was used
435 that supported growth of all taxa used in the experiments, excluding *Faecalibacterium*
436 *prausnitzii*. Manually mapping each component to the Virtual Metabolic Human database, we
437 constructed an *in silico* medium with flux bounds scaled to component concentration. All
438 metabolites were found in the database. Using the MICOM `fix_medium()` function, a minimal set
439 of metabolites necessary for all models to grow to a minimum community growth rate of 0.3 h⁻¹

440 was added to the medium - here, only iron(III) was added (*in silico* medium available here:
441 https://github.com/Gibbons-Lab/scfa_predictions/tree/main/media).

442 To mimic the medium used in *ex vivo* cultures of fecally-derived microbial communities,
443 a diluted, carbon-stripped version of a standard European diet was used. First, a standard
444 European diet was collected from the Virtual Metabolic Human database
445 (www.vmh.life/#nutrition)⁴⁴. Components in the medium which could be imported by the host,
446 as defined by an existing uptake reaction in the Recon3D model⁴⁵, were diluted to 20% of their
447 original flux, to adjust for absorption in the small intestine⁴⁵. Additionally, host-supplied
448 metabolites such as mucins and bile acids were added to the medium. As most carbon sources
449 are consumed in the body and are likely not present in high concentrations in stool, this diet was
450 then algorithmically stripped of carbon sources by removing metabolites with greater than six
451 carbons and no nitrogen, to avoid removing nitrogen sources. Additionally, the remaining
452 metabolites in the medium were diluted to 10% of their original flux, mimicking the nutrient-
453 depleted fecal homogenate. This medium was also augmented using the `fix_medium()` function
454 in MICOM. To simulate fiber supplementation, single fiber additions were made to the medium,
455 either pectin (0.75 mmol/gDW*h) or inulin (10.5 mmol/gDW*h). Bounds for fiber
456 supplementation were chosen to balance the carbon content of each, as represented in the
457 model (pectin: 2535 carbons, inulin: 180 carbons).

458 For *in vivo* modeling, two diets were used: a high-fiber diet containing high levels of
459 resistant starch, and a standard European diet^{44,46}. Again, both diets were collected from the
460 Virtual Metabolic Human database (www.vmh.life/#nutrition). Each medium was subsequently
461 adjusted to account for absorption in the small intestine by diluting metabolite flux as described
462 previously. Additionally, host-supplied metabolites such as mucins and bile acids were added to
463 the medium, to match the composition of the medium *in vivo*. Finally, the `complete_medium()`
464 function was again used to augment the medium, as described above.

465 Prebiotic interventions were designed by supplementing the high-fiber or average
466 European diet with single fiber additions, either pectin or inulin. As before, bounds for fiber
467 addition were set as 0.75 mmol/gDW*h for pectin and 10.5 mmol/gDW*h for inulin.

468

469 *Probiotic Intervention*

470 To model a probiotic intervention, 10% relative abundance of the genus *Faecalibacterium*, a
471 known butyrate-producing taxon⁴⁷, was added to the MCMs by adding a pan-genus model of
472 the taxon derived from the AGORA database version 1.03. Measured taxonomic abundances
473 were scaled to 90% of their initial values, after which *Faecalibacterium* was artificially added to
474 the model.

475

476 *External Data Collection*

477 Data containing taxonomic abundance, optical density, and endpoint butyrate concentration for
478 synthetically-constructed *in vitro* microbial cultures were collected from Clark et al. (2021)²⁵.
479 Endpoint taxonomic abundance data, calculated from fractional read counts collected via 16S
480 amplicon sequencing, was used to construct individual MCMs for each co-culture (see *Model*
481 *Construction*). Resulting models ranged in taxonomic richness from 1 to 25 taxa.

482 From a second study by Cantu-Jungles et al. (2021)¹⁴ (*ex vivo* Study B), preprocessed
483 taxonomic abundance and SCFA metabolomics data was collected. Homogenized fecal
484 samples in this study underwent a similar culturing process, with a culturing time of 24 hours.
485 Cultures were supplemented with the dietary fiber pectin, or a PBS control. Initial and endpoint
486 metabolomic SCFA measurements were used to estimate experimental production flux
487 (concentration[24h] - concentration[0h]/24h). Taxonomic abundance data was used to construct
488 MCMs for each individual (see *Model Construction*).

489 Data from a third (Study C) was collected from the Pharmaceutical Biochemistry Group
490 at the University of Geneva, Switzerland, under study protocol 2019-00632, containing
491 sequencing data in FASTQ format and targeted metabolomics SCFA measurements.

492 Data was collected from Wastyk, et al 2021 ²⁶, which provided 16S amplicon sequencing
493 data at 9 timepoints spanning 14 weeks, along with immunological phenotyping, for 18
494 participants undergoing a high-fiber dietary intervention. Only 7 timepoints spanning 10 weeks
495 were included in subsequent analysis, as the last 2 timepoints were taken after the conclusion
496 of the dietary intervention. MCMs were constructed for each participant at each timepoint at
497 the genus level (see *Model Construction*). Mean total butyrate and propionate production, as
498 well as acetate production, were compared between immune response groups.

499 De-identified data was obtained from a former scientific wellness program run by Arivale,
500 Inc. (Seattle, WA) ²⁷. Arivale closed its operations in 2019. Taxonomic abundances, inferred
501 from 16S amplicon sequencing data, for 2,687 research-consenting individuals were collected
502 and used to construct MCMs. 128 paired blood-based clinical chemistries taken within 30
503 days of fecal sampling were also collected and used to find associations between MCM SCFA
504 predictions on a standard European diet and clinical markers.

505

506 *Statistical analysis*

507 Statistical analysis was performed using SciPy (v1.9.1) and statsmodels (v0.14.0) in Python
508 (v3.8.13). Pearson correlation coefficients and p-values were calculated between measured and
509 predicted SCFA production fluxes in *in vitro* cultures, as well as for predicted SCFA production
510 fluxes across timepoints for an *in vivo* high-fiber intervention. Significance in overall SCFA
511 production between immune response groups in the high-fiber intervention was determined by
512 pairwise Mann-Whitney U test for butyrate+propionate production and for acetate production.
513 Association of MCM-predicted SCFA production flux with paired blood-based clinical labs was

514 tested using OLS regression, adjusting for age, sex, microbiome sequencing vendor, and
515 clinical lab vendor, and tested for significance by two-sided Wald test. BMI was not included as
516 a confounder in the analysis because it was itself negatively correlated with butyrate production
517 ⁴⁸. Multiple comparison correction for p-values was done using the Benjamini–Hochberg method
518 for adjusting the False Discovery Rate (FDR) ⁴⁹. Comparison of butyrate production between
519 dietary interventions was tested using paired Student's t-tests. In all analyses, significance was
520 considered at the $p < 0.05$ threshold.

521

522 *Data, Software, and Code Availability*

523 Code used to run analysis and create figures for this manuscript can be found at

524 https://github.com/Gibbons-Lab/scfa_predictions.

525

526 Processed data for synthetically constructed cultures can be found at

527 <https://github.com/RyanLincolnClark/DesignSyntheticGutMicrobiomeAssemblyFunction>. Raw

528 sequencing data can be found at <https://doi.org/10.5281/zenodo.4642238>.

529

530 Raw sequencing data for Study A can be found in the NCBI SRA under accession number

531 PRJNA937304.

532

533 Processed data for *ex vivo* Study B can be found at

534 https://github.com/ThaisaJungles/fiber_specificity. Raw sequencing data can be found in the

535 NCBI SRA under accession number PRJNA640404.

536

537 Raw sequencing data for *ex vivo* Study C can be found in the NCBI SRA under accession

538 number PRJNA939256.

539

540 Qualified researchers can access the full Arivale deidentified dataset supporting the findings in
541 this study for research purposes through signing a Data Use Agreement (DUA). Inquiries to
542 access the data can be made at data-access@isbscience.org and will be responded to within 7
543 business days.

544

545 *Acknowledgements*

546 We thank members of the Gibbons Lab for helpful discussions and suggestions regarding this
547 work. Thanks to Nathan Price, Amy Willis, Lauren Rajakovich, and Ophelia Venturelli for helpful
548 comments and suggestions.

549

550 *Funding*

551 This research was funded by Washington Research Foundation Distinguished Investigator
552 Award and by startup funds from the Institute for Systems Biology (to SMG). Fecal sample
553 collection at Fred Hutchinson Cancer Center was supported by P30 CA015704. Research
554 reported in this publication was supported by the National Institute of Diabetes and Digestive
555 and Kidney Diseases of the National Institutes of Health (NIH) under award no. R01DK133468
556 (to SMG), as well as the National Institute on Aging of the National Institutes of Health (NIH)
557 under award no. U19AG023122 (to NR).

558

559 *Author contributions*

560 N.B., S.M.G. and C.D. conceptualized the study. N.B. ran the analyses, interpreted results and
561 authored the first draft of the manuscript. S.M.G. and C.D. provided funding, materials and
562 resources for the work, and supervised the work. J.W.L., L.L., and T.G. contributed data and

563 resources. T.W. and N.R. provided support with analyses and statistical interpretation. All
564 authors reviewed and edited the manuscript.

565
566
567
568
569

Citations

- 570 1. Oliphant, K. & Allen-Vercoe, E. Macronutrient metabolism by the human gut microbiome:
571 major fermentation by-products and their impact on host health. *Microbiome* **7**, 91 (2019).
- 572 2. Rackerby, B., Van De Grift, D., Kim, J. H. & Park, S. H. Effects of Diet on Human Gut
573 Microbiome and Subsequent Influence on Host Physiology and Metabolism. *Gut*
574 *Microbiome and Its Impact on Health and Diseases* 63–84 Preprint at
575 https://doi.org/10.1007/978-3-030-47384-6_3 (2020).
- 576 3. Tomasova, L., Grman, M., Ondrias, K. & Ufnal, M. The impact of gut microbiota metabolites
577 on cellular bioenergetics and cardiometabolic health. *Nutr. Metab.* **18**, 72 (2021).
- 578 4. Glotfelty, L. G., Wong, A. C. & Levy, M. Small molecules, big effects: microbial metabolites
579 in intestinal immunity. *Am. J. Physiol. Gastrointest. Liver Physiol.* **318**, G907–G911 (2020).
- 580 5. Donia, M. S. & Fischbach, M. A. HUMAN MICROBIOTA. Small molecules from the human
581 microbiota. *Science* **349**, 1254766 (2015).
- 582 6. Diener, C. *et al.* Genome-microbiome interplay provides insight into the determinants of the
583 human blood metabolome. *Nat Metab* **4**, 1560–1572 (2022).
- 584 7. Ríos-Covián, D. *et al.* Intestinal Short Chain Fatty Acids and their Link with Diet and Human
585 Health. *Front. Microbiol.* **7**, 185 (2016).
- 586 8. Nogal, A., Valdes, A. M. & Menni, C. The role of short-chain fatty acids in the interplay
587 between gut microbiota and diet in cardio-metabolic health. *Gut Microbes* **13**, 1–24 (2021).
- 588 9. Silva, Y. P., Bernardi, A. & Frozza, R. L. The Role of Short-Chain Fatty Acids From Gut
589 Microbiota in Gut-Brain Communication. *Frontiers in Endocrinology* vol. 11 Preprint at

- 590 <https://doi.org/10.3389/fendo.2020.00025> (2020).
- 591 10. Morrison, D. J. & Preston, T. Formation of short chain fatty acids by the gut microbiota and
592 their impact on human metabolism. *Gut Microbes* **7**, 189–200 (2016).
- 593 11. Cong, J., Zhou, P. & Zhang, R. Intestinal Microbiota-Derived Short Chain Fatty Acids in
594 Host Health and Disease. *Nutrients* **14**, (2022).
- 595 12. Tan, J. *et al.* The role of short-chain fatty acids in health and disease. *Adv. Immunol.* **121**,
596 91–119 (2014).
- 597 13. Mortensen, P. B. & Clausen, M. R. Short-chain fatty acids in the human colon: relation to
598 gastrointestinal health and disease. *Scand. J. Gastroenterol. Suppl.* **216**, 132–148 (1996).
- 599 14. Cantu-Jungles, T. M. *et al.* Dietary Fiber Hierarchical Specificity: the Missing Link for
600 Predictable and Strong Shifts in Gut Bacterial Communities. *MBio* **12**, e0102821 (2021).
- 601 15. Healey, G. R., Murphy, R., Brough, L., Butts, C. A. & Coad, J. Interindividual variability in
602 gut microbiota and host response to dietary interventions. *Nutr. Rev.* **75**, 1059–1080
603 (2017).
- 604 16. Boets, E. *et al.* Quantification of in Vivo Colonic Short Chain Fatty Acid Production from
605 Inulin. *Nutrients* **7**, 8916–8929 (2015).
- 606 17. Diener, C., Gibbons, S. M. & Resendis-Antonio, O. MICOM: Metagenome-Scale Modeling
607 To Infer Metabolic Interactions in the Gut Microbiota. *mSystems* **5**, (2020).
- 608 18. van Deuren, T., Blaak, E. E. & Canfora, E. E. Butyrate to combat obesity and obesity-
609 associated metabolic disorders: Current status and future implications for therapeutic use.
610 *Obes. Rev.* **23**, e13498 (2022).
- 611 19. Zeevi, D. *et al.* Personalized Nutrition by Prediction of Glycemic Responses. *Cell* **163**,
612 1079–1094 (2015).
- 613 20. Rein, M. *et al.* Effects of personalized diets by prediction of glycemic responses on
614 glycemic control and metabolic health in newly diagnosed T2DM: a randomized dietary

- 615 intervention pilot trial. *BMC Med.* **20**, 56 (2022).
- 616 21. Gibbons, S. M. *et al.* Perspective: Leveraging the Gut Microbiota to Predict Personalized
617 Responses to Dietary, Prebiotic, and Probiotic Interventions. *Adv. Nutr.* **13**, 1450–1461
618 (2022).
- 619 22. Heinken, A. *et al.* Genome-scale metabolic reconstruction of 7,302 human microorganisms
620 for personalized medicine. *Nat. Biotechnol.* (2023) doi:10.1038/s41587-022-01628-0.
- 621 23. Abdill, R. J., Adamowicz, E. M. & Blekhman, R. Public human microbiome data are
622 dominated by highly developed countries. *PLoS Biol.* **20**, e3001536 (2022).
- 623 24. Magnúsdóttir, S. *et al.* Generation of genome-scale metabolic reconstructions for 773
624 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).
- 625 25. Clark, R. L. *et al.* Design of synthetic human gut microbiome assembly and butyrate
626 production. *Nat. Commun.* **12**, 3254 (2021).
- 627 26. Wastyk, H. C. *et al.* Gut-microbiota-targeted diets modulate human immune status. *Cell*
628 **184**, 4137–4153.e14 (2021).
- 629 27. Manor, O. *et al.* Health and disease markers correlate with gut microbiome composition
630 across thousands of people. *Nat. Commun.* **11**, 5206 (2020).
- 631 28. Heinken, A. *et al.* AGORA2: Large scale reconstruction of the microbiome highlights wide-
632 spread drug-metabolising capacities. *bioRxiv* 2020.11.09.375451 (2020)
633 doi:10.1101/2020.11.09.375451.
- 634 29. Valgepea, K. *et al.* Systems biology approach reveals that overflow metabolism of acetate
635 in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase.
636 *BMC Syst. Biol.* **4**, 166 (2010).
- 637 30. Wolfe, A. J. The acetate switch. *Microbiol. Mol. Biol. Rev.* **69**, 12–50 (2005).
- 638 31. Li, M. *et al.* Pro- and anti-inflammatory effects of short chain fatty acids on immune and
639 endothelial cells. *Eur. J. Pharmacol.* **831**, 52–59 (2018).

- 640 32. Arifuzzaman, M. *et al.* Inulin fibre promotes microbiota-derived bile acids and type 2
641 inflammation. *Nature* **611**, 578–584 (2022).
- 642 33. Sproston, N. R. & Ashworth, J. J. Role of C-Reactive Protein at Sites of Inflammation and
643 Infection. *Front. Immunol.* **9**, 754 (2018).
- 644 34. Amiri, P. *et al.* Role of Butyrate, a Gut Microbiota Derived Metabolite, in Cardiovascular
645 Diseases: A comprehensive narrative review. *Front. Pharmacol.* **12**, 837509 (2021).
- 646 35. Vinolo, M. A. R., Rodrigues, H. G., Nachbar, R. T. & Curi, R. Regulation of inflammation by
647 short chain fatty acids. *Nutrients* **3**, 858–876 (2011).
- 648 36. Tedelind, S., Westberg, F., Kjerrulf, M. & Vidal, A. Anti-inflammatory properties of the short-
649 chain fatty acids acetate and propionate: a study with relevance to inflammatory bowel
650 disease. *World J. Gastroenterol.* **13**, 2826–2832 (2007).
- 651 37. Gurry, T., Nguyen, L. T. T., Yu, X. & Alm, E. J. Functional heterogeneity in the fermentation
652 capabilities of the healthy human gut microbiota. *PLoS One* **16**, e0254004 (2021).
- 653 38. Gasaly, N., de Vos, P. & Hermoso, M. A. Impact of Bacterial Metabolites on Gut Barrier
654 Function and Host Immunity: A Focus on Bacterial Metabolism and Its Relevance for
655 Intestinal Inflammation. *Front. Immunol.* **12**, 658354 (2021).
- 656 39. Agus, A., Clément, K. & Sokol, H. Gut microbiota-derived metabolites as central regulators
657 in metabolic disorders. *Gut* **70**, 1174–1182 (2021).
- 658 40. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
659 *Bioinformatics* **34**, i884–i890 (2018).
- 660 41. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2.
661 *Genome Biol.* **20**, 257 (2019).
- 662 42. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species
663 abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104 (2017).
- 664 43. Gauglitz, J. M. *et al.* Enhancing untargeted metabolomics using metadata-based source

- 665 annotation. *Nat. Biotechnol.* **40**, 1774–1779 (2022).
- 666 44. Elmadfa, I. Österreichischer Ernährungsbericht 2012. **1**, (2012).
- 667 45. Brunk, E. *et al.* Recon3D enables a three-dimensional view of gene variation in human
668 metabolism. *Nat. Biotechnol.* **36**, 272–281 (2018).
- 669 46. Waldmann, A., Koschizke, J. W., Leitzmann, C. & Hahn, A. Dietary intakes and lifestyle
670 factors of a vegan population in Germany: results from the German Vegan Study. *Eur. J.*
671 *Clin. Nutr.* **57**, 947–955 (2003).
- 672 47. Zhou, L. *et al.* Faecalibacterium prausnitzii Produces Butyrate to Maintain Th17/Treg
673 Balance and to Ameliorate Colorectal Colitis by Inhibiting Histone Deacetylase 1. *Inflamm.*
674 *Bowel Dis.* **24**, 1926–1940 (2018).
- 675 48. Coppola, S., Avagliano, C., Calignano, A. & Berni Canani, R. The Protective Role of
676 Butyrate against Obesity and Obesity-Related Diseases. *Molecules* **26**, (2021).
- 677 49. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful
678 approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).

679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696

697 Figure Captions

698

699 **Figure 1. Microbial community-scale metabolic models (MCMMs) predict personalized SCFA**
700 **production profiles.** Schematic of our workflow for validating MCMM-based personalized predictions
701 for SCFA production. **(A)** Prior to modeling, an *in silico* medium is constructed, containing a matched
702 diet mapped to its constituent metabolic components. The medium is depleted in compounds absorbed
703 by the host in the small intestine and augmented with other host-supplied compounds, in addition to
704 adding a minimal set of metabolites required for growth. **(B)** MCMMs are constructed, combining
705 abundance and taxonomic data with pre-curated GEMs into a community model. **(C)** Growth in the
706 MCMM is simulated through cooperative tradeoff flux balance analysis (ctFBA), yielding predicted
707 growth rates and SCFA production fluxes. **(D)** To validate predicted levels of SCFA production fluxes,
708 measured values of production fluxes are collected from fecal samples cultured anaerobically *ex vivo* at
709 37°C over time. **(E)** Predicted and measured SCFA production fluxes are compared to assess the
710 accuracy of the model.
711

712 **Figure 2. Relationship between predicted and measured butyrate production rates in *in***
713 ***vitro* co-cultures.** Each point denotes one of 1,387 anaerobic co-culture assays. Butyrate
714 production flux predictions from MCMMs are shown on the y-axes and measured values are
715 shown on the x-axes, along with R^2 and p-values from a Pearson's correlation **(A)** Synthetically
716 constructed communities were cultured anaerobically in a defined medium. Endpoint butyrate
717 concentration was measured and compared with MCMM-predicted flux. **(B)** Predicted and
718 measured butyrate fluxes in models of low richness synthetic communities (1-5 genera per
719 model, N = 882). **(C)** Predicted and measured butyrate fluxes in models of high richness
720 synthetic communities (10-25 genera, N = 697). In (B-C) the dashed line denotes a linear model
721 fit to the data.
722

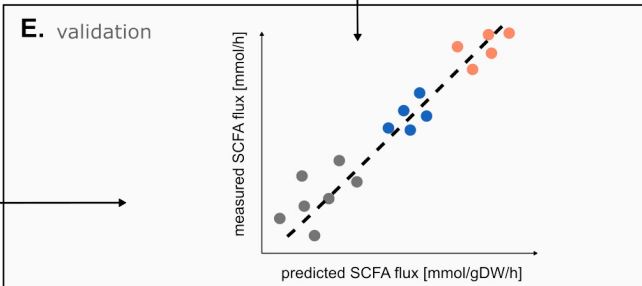
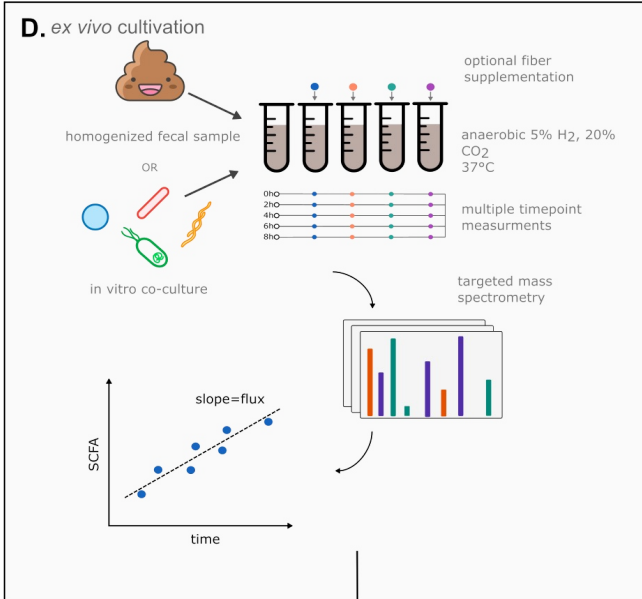
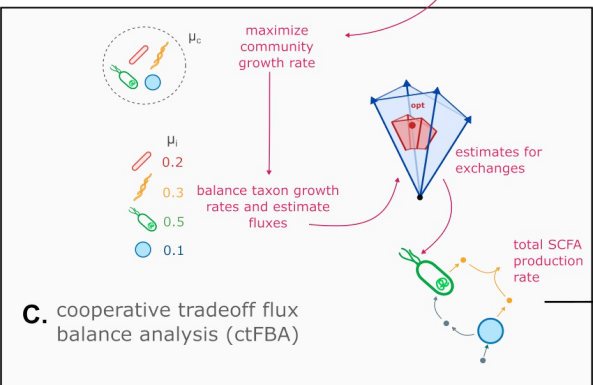
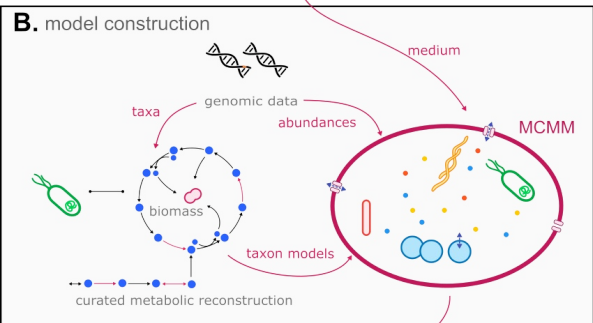
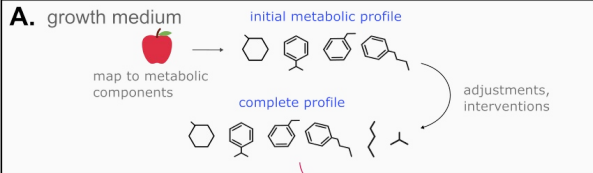
723 **Figure 3. Human stool *ex vivo* assays show quantitative agreement between measured**
724 **and predicted SCFA production fluxes.** SCFA production flux predictions from MCMMs are
725 shown on the y-axes and measured values are shown on the x-axes, along with R^2 and p-
726 values from a Pearson's correlation. Marginal rug plots show separation in SCFA production
727 between treatment groups. Error bars show standard error as calculated from measured and
728 predicted values for each sample in triplicate. **(A-C)**, Results from a two-donor *ex vivo* study
729 (Study A) showed significant agreement between measured and prediction rates for all three
730 SCFAs following inulin treatment. **(D-F)** Results from Study B¹⁴, which included pectin
731 treatments to stool homogenates from 10 individuals¹⁴. Samples showed significant association
732 between predicted and measured production rates for all three SCFAs. **(G-I)** Results from Study
733 C, which included both pectin and inulin interventions across stool homogenates from 9
734 individuals. Samples treated with inulin and pectin showed significant associations between
735 predicted and measured fluxes for both propionate and butyrate, but not for acetate.
736

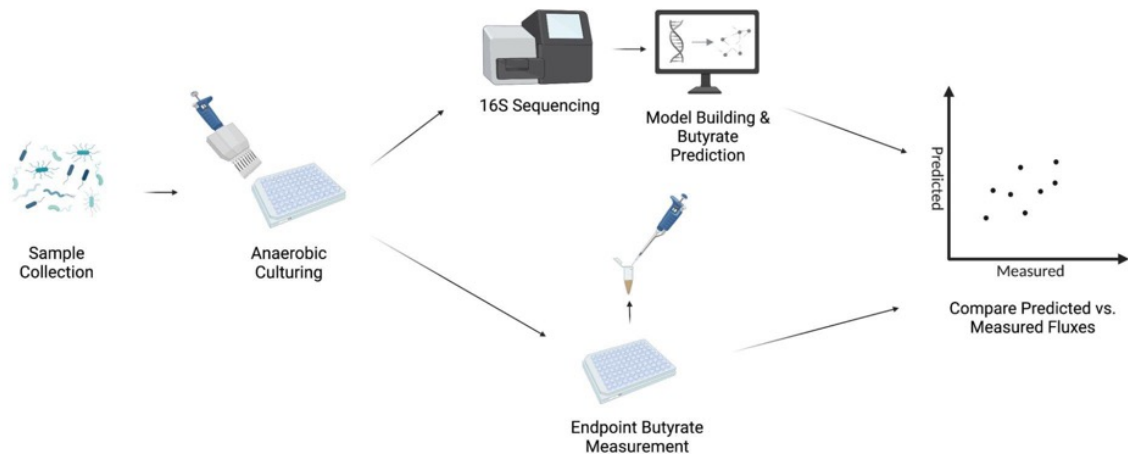
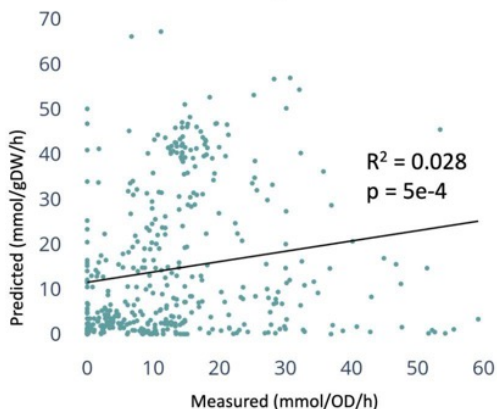
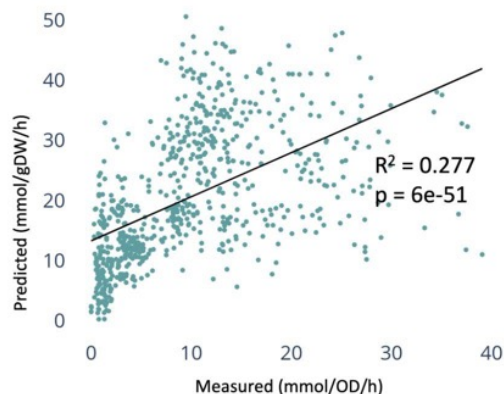
737 **Figure 4. Predicted SCFA production profiles were associated with variable immune**
738 **response groups following a high-fiber dietary intervention.** **(A)** Summary of the study from
739 Wastyk et al.²⁶, where a cohort of 18 individuals participated in a 10-week high-fiber dietary
740 intervention. Immune profiling based on circulating inflammatory cytokines and immune cells
741 clustered individuals into three groups: two low-inflammation groups and one high-inflammation
742 group. **(B)** Average predicted total butyrate and propionate production across the three immune-

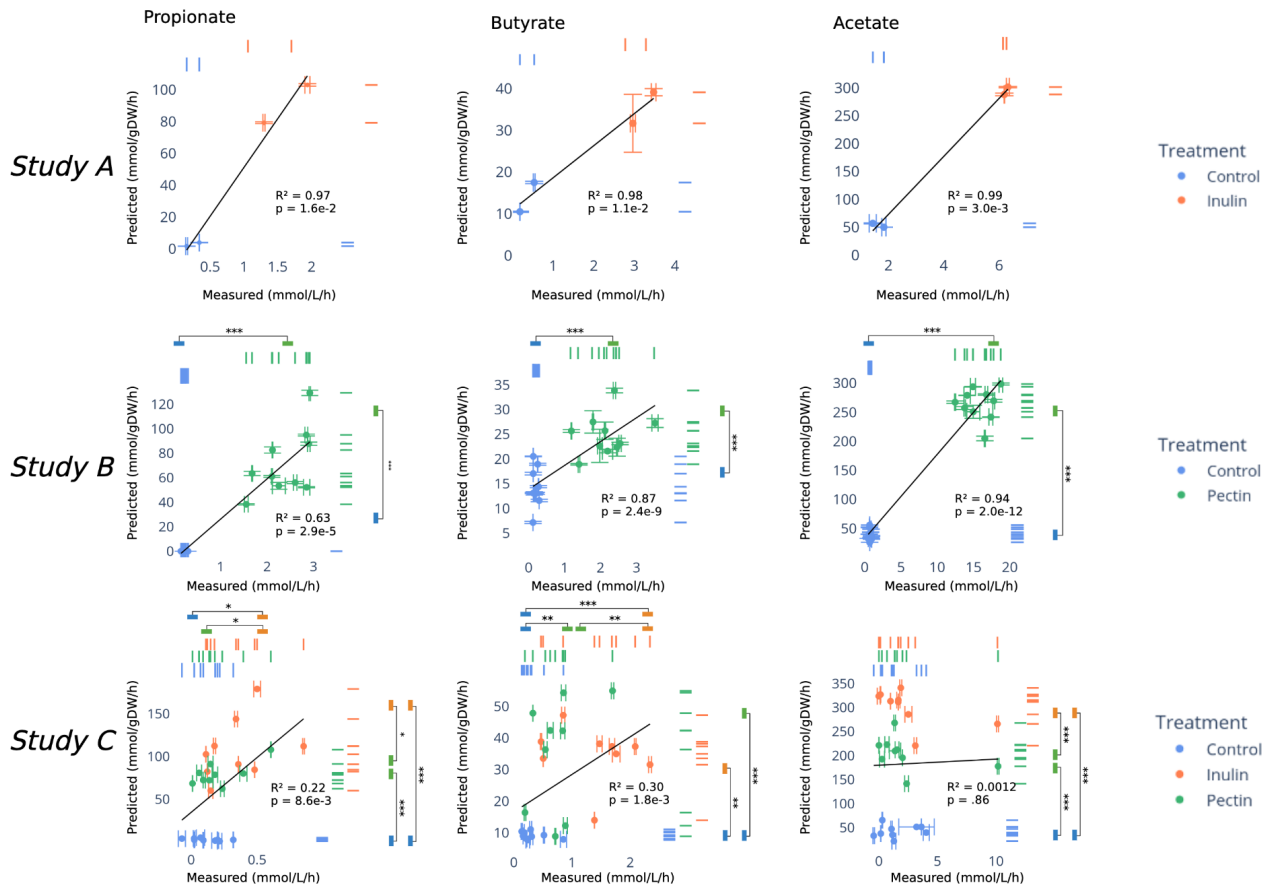
743 response groups identified in the original study. **(C)** Predicted total butyrate and propionate
744 production rates across the duration of the intervention, stratified by immune response group
745 **(D)** Average predicted total acetate production across the three immune response groups. **(E)**
746 Predicted acetate production rates across the intervention, stratified by immune response
747 group. In (A-E) stars denote significance under an Independent Student's t-test, *** = $p < 0.001$.
748

749 **Figure 5. SCFA flux predictions are significantly associated with blood-derived clinical**
750 **markers. (A)** MCMMs were constructed for 2,687 Arivale participants, assuming an average
751 European diet, to predict SCFA production profiles. SCFA predictions were regressed against a
752 set of 128 blood-based clinical labs and health markers, with sex, age, and sequencing vendor
753 as covariates in the regressions. **(B)** Heatmap showing the 37 significant associations (FDR-
754 corrected Wald test $p < 0.05$) between measured blood markers and predicted SCFA production
755 rates. **(C-E)** Relationship between pairs of predicted SCFA production rates. Each dot denotes
756 an individual model reconstructed for a single sample in the Arivale study ($n=2,687$). The black
757 line denotes a linear regression line and the gray area denotes the 95% confidence interval of
758 the regression. R^2 and p -values from Pearson's correlations.
759

760 **Figure 6. Microbial MCMMs can be used to design, build, and test personalized prebiotic,**
761 **probiotic, and dietary interventions aimed at optimizing SCFA production profiles. (A)**
762 MCMMs built from the Arivale cohort ($N = 2,687$) were used to test personalized responses to
763 dietary interventions. Personalized models were simulated on an average European (Euro) diet,
764 as well as on a high-fiber diet, and divided into responders, non-responders, and regressors,
765 based on the changes in predicted butyrate production in response to increasing dietary fiber.
766 Non-responders were defined as individuals who produced less than $10 \frac{\square\square\square\square}{\square\square\square*\square}$ of butyrate on the
767 European diet and showed an increase of less than 20% in butyrate production on the high-fiber
768 diet. Regressors were defined as individuals who produced at least $20 \frac{\square\square\square\square}{\square\square\square*\square}$ butyrate on the
769 European diet and showed a drop in butyrate production on the high-fiber diet. Single-fiber and
770 probiotic interventions were applied to non-responders and regressors. **(B)** Distribution of
771 butyrate production rates on two different diets simulated for all participants in the study.
772 Butyrate production ranges that contain non-responders ($N=29$) and regressors ($N=45$) are
773 highlighted in green and yellow shaded areas, respectively. **(C)** Distributions of butyrate
774 production rates for the non-responder group ($N=29$). The optimal intervention resulting in the
775 highest butyrate production is shown in blue. **(D)** Butyrate production rates for the regressor
776 group ($N=45$). The optimal intervention that resulted in the highest butyrate production is shown
777 in blue. **(E)** Heatmap of butyrate production rates across simulated interventions for the
778 individuals in the non-responder and regressor groups. Rows denotes specific interventions
779 (Euro - average European diet, HF - high fiber diet). Columns denote individuals in the response
780 groups ($N=74$). Cell shading (white-to-red) denotes butyrate production rate. Added
781 interventions tested on both non-responders and regressors included probiotic supplementation
782 (inulin or pectin) as well as prebiotic supplementation (10% relative abundance
783 *Faecalibacterium*). The most successful intervention for each individual is denoted by a black
784 border around that cell in the corresponding column.



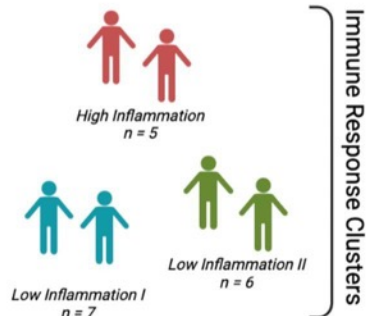
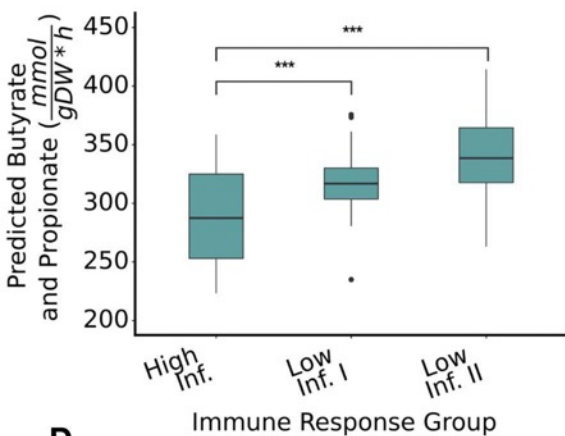
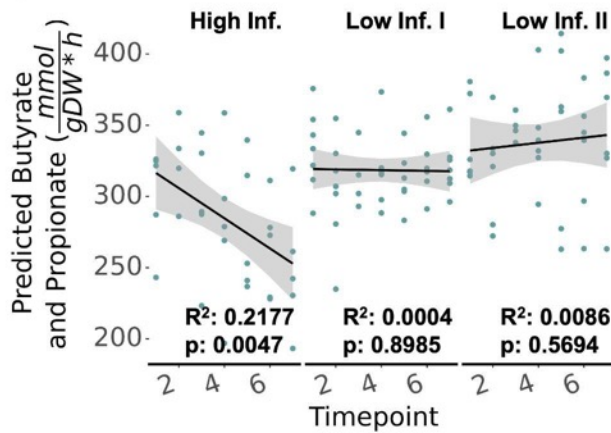
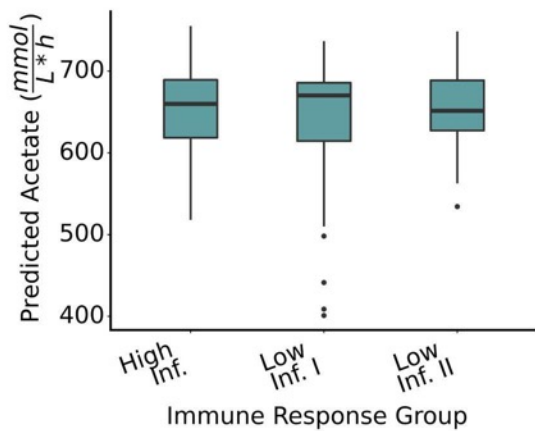
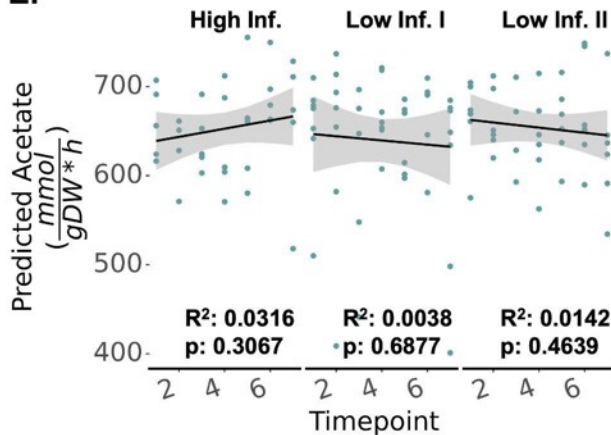
A.**B.****Butyrate****1-5 Member Communities****C.****Butyrate****10-25 Member Communities**

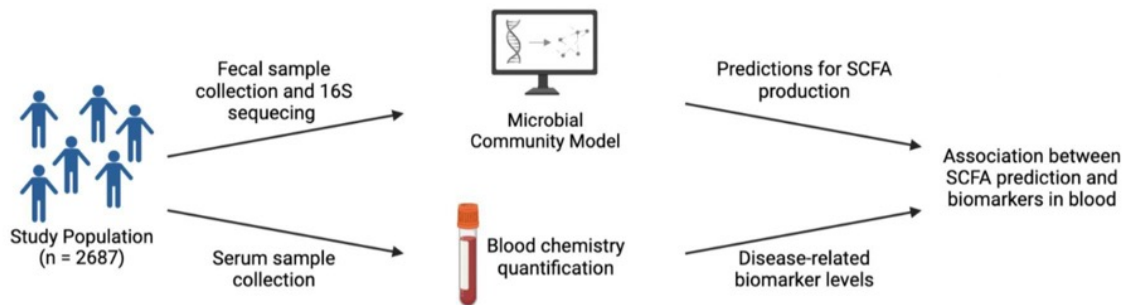
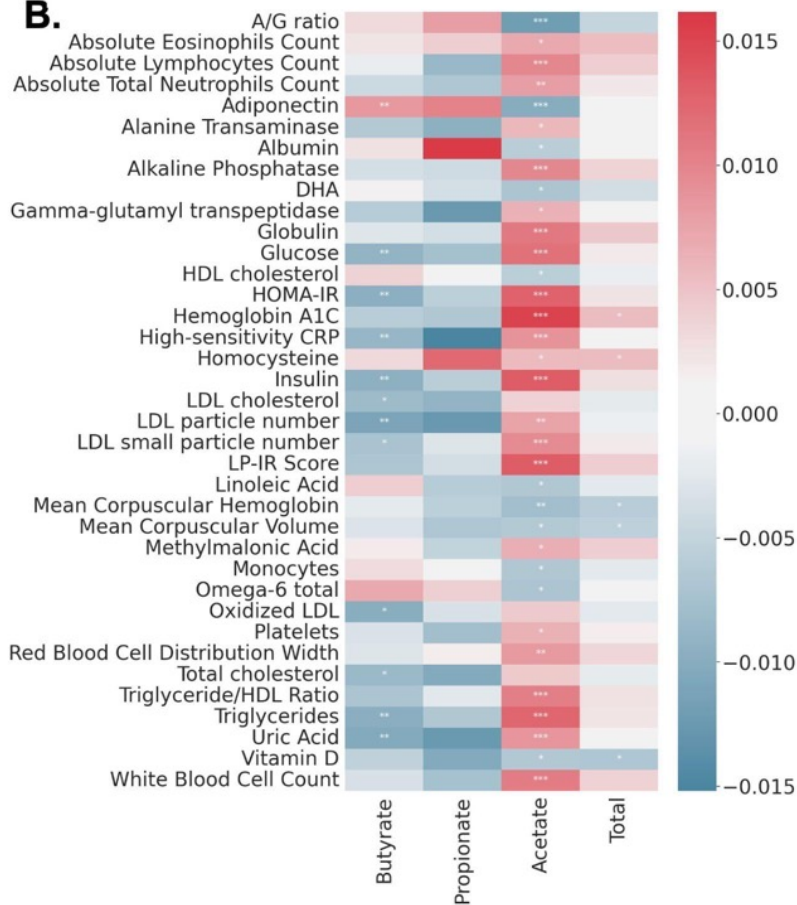
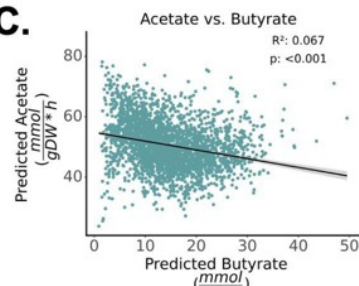
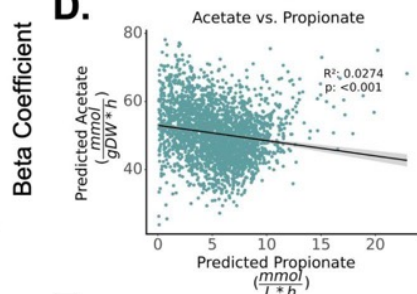


A.

10 week high-fiber
dietary intervention

- Inflammatory cytokines
- Cell signaling markers
- 16S sequencing

**B.****C.****D.****E.**

A.**B.****C.****D.****E.**