bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

Authors: Nick Bohmann[1,2], Tomasz Wilmanski[1], Lisa Levy[3], Johanna W. Lampe[3], Thomas Gurry[4,5], Noa Rappaport[1], Christian Diener[1,*], Sean M. Gibbons[1,2,6,7,*]

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

Abstract

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

Keywords

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

The human gut microbiota maintains intestinal barrier function, regulates peripheral and systemic inflammation, and breaks down indigestible dietary components and host substrates into a wide range of bioactive compounds [1,2]. One of the primary mechanisms by which the gut microbiota impacts human health is through the production of small molecules that enter the circulation and are absorbed and transformed by host tissues [3–5]. Approximately half of the metabolites detected in human blood are known to be significantly associated with cross-sectional variation in gut microbiome composition [6].

Short-chain-fatty-acids (SCFAs) are among the most abundant metabolic byproducts produced by the gut microbiota, largely through the fermentation of indigestible dietary fibers and resistant starches, with acetate, propionate and butyrate being the most abundant SCFAs [7–9]. Deficits in SCFA production have been repeatedly associated with disease [10,11]. Therefore, SCFA production is a crucial ecosystem service that the gut microbiota provides to its host, with far-reaching impacts on health [1,11–13]. However, different human gut microbiota provided with the same exact dietary substrate can show variable SCFA production profiles [14,15], and predicting this heterogeneity remains a fundamental challenge to the microbiome field. Measuring SCFA abundances in blood or feces is rarely informative of *in situ* production rates, due to the volatility of SCFAs, cross-feeding among microbes, and the rapid consumption and transformation of these metabolites by the colonic epithelium [10,16,17] . Furthermore, SCFA production fluxes (i.e., the amount of a metabolite produced over a given period of time) within an individual can vary longitudinally, depending upon dietary inputs and the availability of host substrates [18]. In order to account for this inter- and intra-individual heterogeneity, we propose the use of microbial community-scale metabolic models (MCMMs), which mechanistically account for metabolic interactions between gut microbes, host substrates, and dietary inputs, to estimate personalized, context-specific SCFA production profiles.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

Results

MCMMs capture SCFA production rates in vitro

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

We next investigated whether MCMM-predicted SCFA production rates could be leveraged to help explain inter-individual differences in phenotypic response following a dietary intervention. Specifically, we looked at data from 18 individuals who were placed on a high-fiber diet for ten weeks [26]. These individuals fell into three distinct immunological response groups: one in which high inflammation was observed over the course of the intervention (high-inflammation group), and two other distinct response groups that both exhibited lower levels of inflammation (low-inflammation groups I and II; **Fig. 4A**). We hypothesized that these immune response groups could be explained, in part, by differences in MCMM-predicted SCFA production profiles. Using 16S amplicon sequencing data from seven time points collected from each of these 18 individuals over the 10-week intervention, we built MCMMs for each study participant at each time point in the study. Growth was then simulated for each model using a standardized high-fiber diet, rich in resistant starch (see Material and Methods). Throughout the study, individuals in the high-inflammation group showed significantly lower predicted butyrate plus propionate production on average (i.e., the two SCFAs with the strongest anti-inflammatory effects), compared to the individuals in each of the low-inflammation groups (High vs. Low I: 284.6 ± 7.7 vs 327.5 ± 3.8 mmol/(gDW h) on average, Mann-Whitney p = 1.9e-5. High vs. Low II: 284.6 ± 7.7 vs 337.8 ± 6.4 mmol/(gDW h), Mann-Whitney p = 7.2e-6) (**Fig. 4B**). Predicted levels of butyrate plus propionate production in the high-inflammation group decreased throughout the

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

MCMM-predicted SCFA profiles are associated with a wide range of blood-based clinical markers

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

Leveraging MCMMs to design precision dietary, prebiotic, and probiotic interventions

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

Discussion

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

Model predictions of butyrate production in synthetically constructed *in vitro* co-cultures showed significant agreement between measured and predicted butyrate fluxes (**Fig. 2**). Due to the phylogenetic resolution of 16S data and the lack of strain-level GEMs that match the organisms present in some samples, we built genus-level MCMMs for all analyses. The decreasing accuracy of butyrate predictions as community richness declined may reflect a limitation of building models at the genus-level, as reconstructions contain a summarized aggregation of the metabolic capability of the genus as a whole, without species- or strain-level resolution. Consequently, pathways included in the metabolic model may be absent in a low-richness experimental system, reflecting a mismatch between the modeling framework and reality. In high richness models, predictions became more accurate, indicating this mismatch is less impactful as more taxa are included in the system. Real-world microbiomes are often more species-rich than synthetic *in vitro* communities. Fortunately, we are likely not operating in a community regime where missingness in individual models has a large influence on genus-level MCMM SCFA predictions in the human gut. However, future work should focus on increasing the availability of diverse strain-level GEMs. The recent release of AGORA2[28], containing GEMs for 7,302 microorganisms, may help to overcome this limitation and aid in the construction of MCMMs at finer levels of taxonomic resolution.

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

MCMMs built using data from a 10-week high-fiber dietary intervention allowed us to assess our predictions *in vivo* in the context of immunological responses to diet. Predictions of combined butyrate and propionate (i.e., SCFAs with the strongest anti-inflammatory effects on the host [31]) production rates over the course of the high-fiber intervention showed distinct differences between pre-defined immune response groups (**Fig. 4**). The low-inflammation groups showed stable butyrate and propionate production fluxes over time, while the high inflammation group showed lower average butyrate and propionate production and a decreasing production rate over time following the initiation of the high-fiber diet (**Fig. 4B-C**). Given the strong anti-inflammatory effects of butyrate and propionate, we expected to see lower production of these molecules in the context of higher inflammation. Prior work has shown that higher doses of inulin can actually induce an inflammatory response, which may explain, in part, the inflammatory immune response in these individuals [32]. Overall, our results indicate that

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

Several biomarkers of metabolic health, inflammation, liver function, and cardiovascular health were associated with MCMM-predicted SCFA production profiles in a large, generally-healthy cohort, assuming an average European diet (**Fig. 5**). CRP, a marker of systemic inflammation [33], showed a significant negative association with butyrate production predictions. Markers of cardiovascular health, such as LDL and triglyceride levels, were also negatively associated with butyrate production rates, supporting the role of butyrate as protective against cardiovascular disease [34]. Many significant associations were inverted when looking across butyrate and acetate flux predictions (**Fig. 5**). For instance, CRP, HOMA-IR, glucose, insulin, LDL cholesterol and uric acid all showed significant negative associations with butyrate production and significant positive associations with acetate production. This result may be related to the apparent tradeoff between acetate production and the production of both butyrate and propionate (**Fig. 5C-E**). While the overall production of SCFAs has been implicated in lowering inflammation[35], the potency of butyrate in driving down inflammation and improving overall metabolic health is greater than that of acetate [36], which, given this apparent tradeoff in the production of these different SCFAs, could help to explain these inverted associations.

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 MCMMs 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 interventions [37].
322 MCMMs 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 2021 [25]. 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., 2021[14].
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,

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 FASTP [40]. 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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 before [43]. 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 v0.32.3 [17]. 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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

between analyses (see *Media Construction).* Steady state growth rates and fluxes for all samples were then inferred using cooperative tradeoff flux balance analysis (ctFBA). In brief, this is a two-step optimization scheme, where the first step finds the largest possible biomass production rate for the full microbial community and the second step infers taxon-specific growth rates and fluxes, while maintaining community growth within a fraction of the theoretical maximum (i.e., the tradeoff parameter), thus balancing individual growth rates and the community-wide growth rate [17]. For all models in the manuscript we used a tradeoff parameter of 0.7. This parameter value was chosen through cooperative tradeoff analysis in MICOM. Multiple parameters were tested, and the highest parameter value (i.e., the value closest to the maximal community growth rate at 1.0) that allowed most (>90%) of taxa to grow was chosen (i.e., 0.7). Predicted growth rates from the simulation were analyzed to validate correct behavior of the models. All models were found to grow with minimum community growth rate of 0.3 h^{-1}. Predicted values for export fluxes of SCFAs were collected from each MCMM using the production_rates() function, which calculates the overall production from the community that would be accessible to the colonic epithelium.

Media Construction

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

To mimic the medium used in *ex vivo* cultures of fecally-derived microbial communities, a diluted, carbon-stripped version of a standard European diet was used. First, a standard European diet was collected from the Virtual Metabolic Human database (www.vmh.life/#nutrition) [44]. Components in the medium which could be imported by the host, as defined by an existing uptake reaction in the Recon3D model [45], were diluted to 20% of their original flux, to adjust for absorption in the small intestine[45]. Additionally, host-supplied metabolites such as mucins and bile acids were added to the medium. As most carbon sources are consumed in the body and are likely not present in high concentrations in stool, this diet was then algorithmically stripped of carbon sources by removing metabolites with greater than six carbons and no nitrogen, to avoid removing nitrogen sources. Additionally, the remaining metabolites in the medium were diluted to 10% of their original flux, mimicking the nutrient-depleted fecal homogenate. This medium was also augmented using the fix_medium() function in MICOM. To simulate fiber supplementation, single fiber additions were made to the medium, either pectin (0.75 mmol/gDW*h) or inulin (10.5 mmol/gDW*h). Bounds for fiber supplementation were chosen to balance the carbon content of each, as represented in the model (pectin: 2535 carbons, inulin: 180 carbons).

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

Probiotic Intervention

To model a probiotic intervention, 10% relative abundance of the genus *Faecalibacterium*, a known butyrate-producing taxon [47], was added to the MCMMs by adding a pan-genus model of the taxon derived from the AGORA database version 1.03. Measured taxonomic abundances were scaled to 90% of their initial values, after which *Faecalibacterium* was artificially added to the model.

External Data Collection

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

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

Statistical analysis

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 **Citations**
569

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

https://doi.org/10.3389/fendo.2020.00025 (2020).

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

intervention pilot trial. *BMC Med.* **20**, 56 (2022).

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

32. Arifuzzaman, M. *et al.* Inulin fibre promotes microbiota-derived bile acids and type 2 inflammation. *Nature* **611**, 578–584 (2022).

33. Sproston, N. R. & Ashworth, J. J. Role of C-Reactive Protein at Sites of Inflammation and Infection. *Front. Immunol.* **9**, 754 (2018).

34. Amiri, P. *et al.* Role of Butyrate, a Gut Microbiota Derived Metabolite, in Cardiovascular Diseases: A comprehensive narrative review. *Front. Pharmacol.* **12**, 837509 (2021).

35. Vinolo, M. A. R., Rodrigues, H. G., Nachbar, R. T. & Curi, R. Regulation of inflammation by short chain fatty acids. *Nutrients* **3**, 858–876 (2011).

36. Tedelind, S., Westberg, F., Kjerrulf, M. & Vidal, A. Anti-inflammatory properties of the short-chain fatty acids acetate and propionate: a study with relevance to inflammatory bowel disease. *World J. Gastroenterol.* **13**, 2826–2832 (2007).

37. Gurry, T., Nguyen, L. T. T., Yu, X. & Alm, E. J. Functional heterogeneity in the fermentation capabilities of the healthy human gut microbiota. *PLoS One* **16**, e0254004 (2021).

38. Gasaly, N., de Vos, P. & Hermoso, M. A. Impact of Bacterial Metabolites on Gut Barrier Function and Host Immunity: A Focus on Bacterial Metabolism and Its Relevance for Intestinal Inflammation. *Front. Immunol.* **12**, 658354 (2021).

39. Agus, A., Clément, K. & Sokol, H. Gut microbiota-derived metabolites as central regulators in metabolic disorders. *Gut* **70**, 1174–1182 (2021).

40. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

41. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).

42. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104 (2017).

43. Gauglitz, J. M. *et al.* Enhancing untargeted metabolomics using metadata-based source

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

annotation. *Nat. Biotechnol.* **40**, 1774–1779 (2022).

44. Elmadfa, I. Österreichischer Ernährungsbericht 2012. **1**, (2012).

45. Brunk, E. *et al.* Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* **36**, 272–281 (2018).

46. Waldmann, A., Koschizke, J. W., Leitzmann, C. & Hahn, A. Dietary intakes and lifestyle factors of a vegan population in Germany: results from the German Vegan Study. *Eur. J. Clin. Nutr.* **57**, 947–955 (2003).

47. Zhou, L. *et al.* Faecalibacterium prausnitzii Produces Butyrate to Maintain Th17/Treg Balance and to Ameliorate Colorectal Colitis by Inhibiting Histone Deacetylase 1. *Inflamm. Bowel Dis.* **24**, 1926–1940 (2018).

48. Coppola, S., Avagliano, C., Calignano, A. & Berni Canani, R. The Protective Role of Butyrate against Obesity and Obesity-Related Diseases. *Molecules* **26**, (2021).

49. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure Captions

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

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

Figure 3. Human stool *ex vivo* assays show quantitative agreement between measured and predicted SCFA production fluxes. SCFA production flux predictions from MCMMs are shown on the y-axes and measured values are shown on the x-axes, along with R^2 and p-values from a Pearson's correlation. Marginal rug plots show separation in SCFA production between treatment groups. Error bars show standard error as calculated from measured and predicted values for each sample in triplicate. **(A-C),** Results from a two-donor *ex vivo* study (Study A) showed significant agreement between measured and prediction rates for all three SCFAs following inulin treatment. **(D-F)** Results from Study B[14], which included pectin treatments to stool homogenates from 10 individuals [14]. Samples showed significant association between predicted and measured production rates for all three SCFAs. **(G-I)** Results from Study C, which included both pectin and inulin interventions across stool homogenates from 9 individuals. Samples treated with inulin and pectin showed significant associations between predicted and measured fluxes for both propionate and butyrate, but not for acetate.

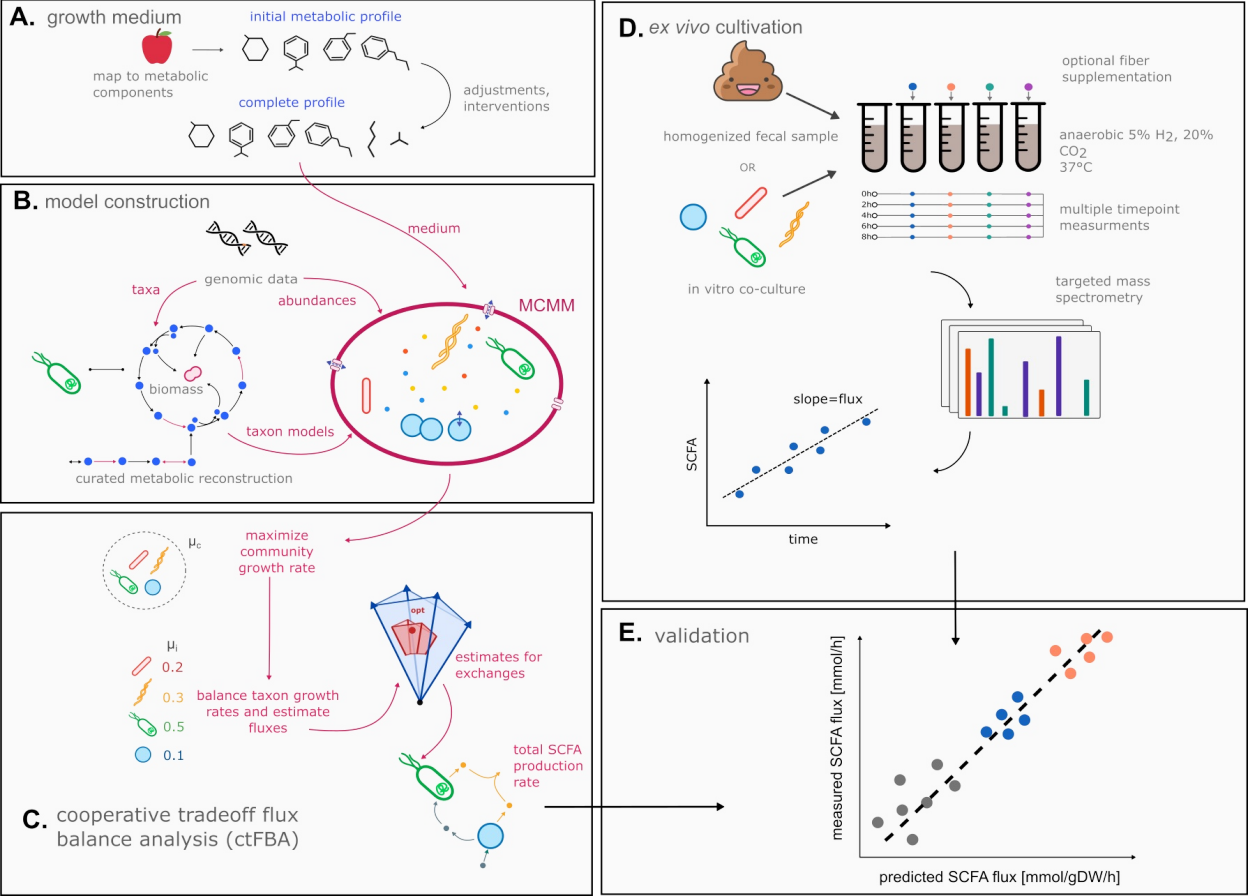
Figure 4. Predicted SCFA production profiles were associated with variable immune response groups following a high-fiber dietary intervention. (A) Summary of the study from Wastyk et al.[26], where a cohort of 18 individuals participated in a 10-week high-fiber dietary intervention. Immune profiling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

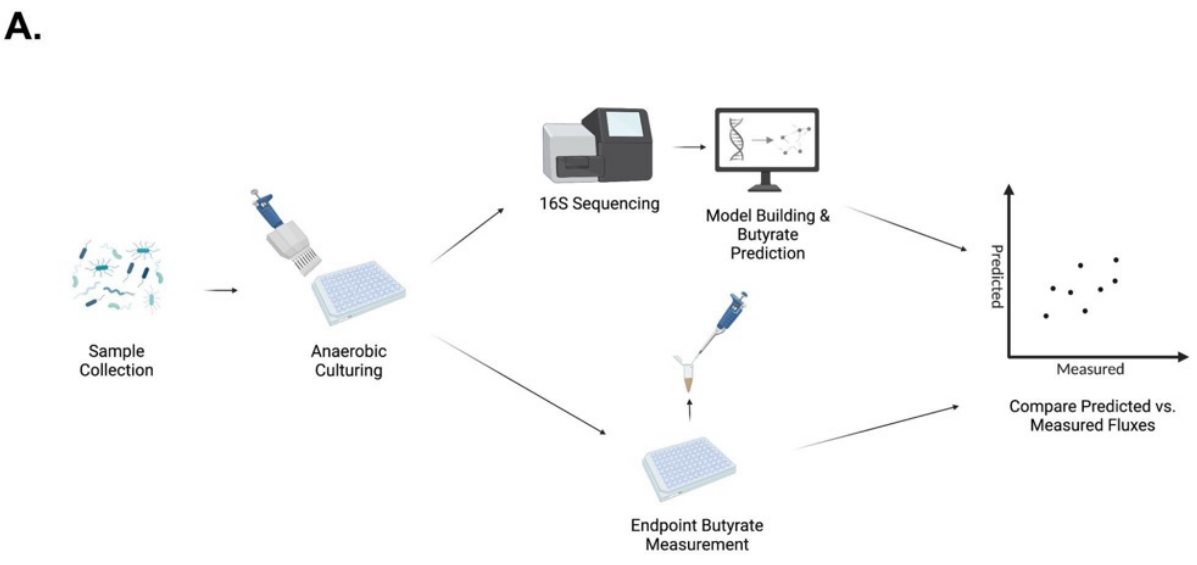
bioRxiv preprint doi: https://doi.org/10.1101/2023.02.28.530516; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

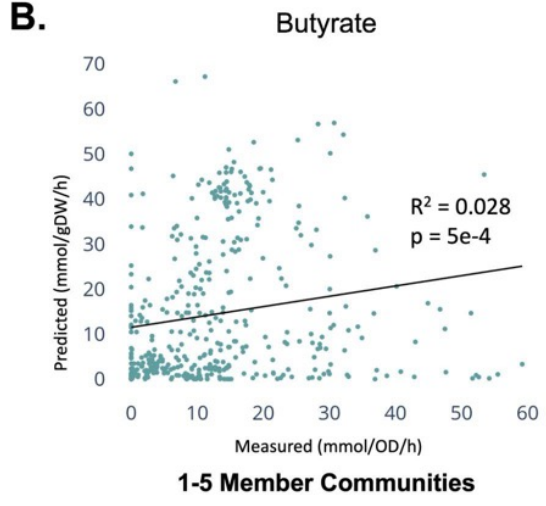
response groups identified in the original study. **(C)** Predicted total butyrate and propionate production rates across the duration of the intervention, stratified by immune response group **(D)** Average predicted total acetate production across the three immune response groups. **(E)** Predicted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

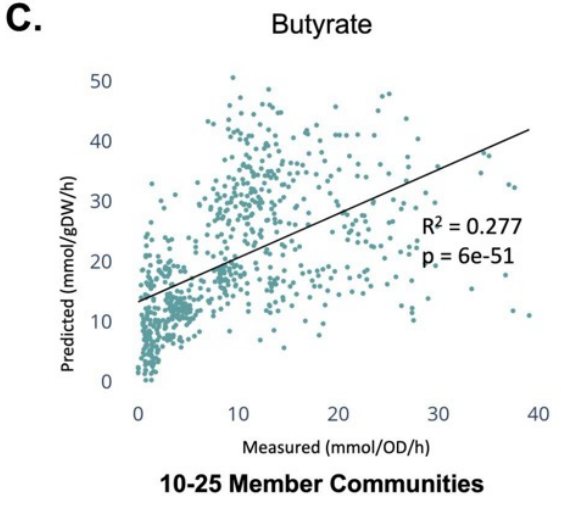
Figure 5. SCFA flux predictions are significantly associated with blood-derived clinical markers. (A) MCMMs were constructed for 2,687 Arivale participants, assuming an average European diet, to predict SCFA production profiles. SCFA predictions were regressed against a set of 128 blood-based clinical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

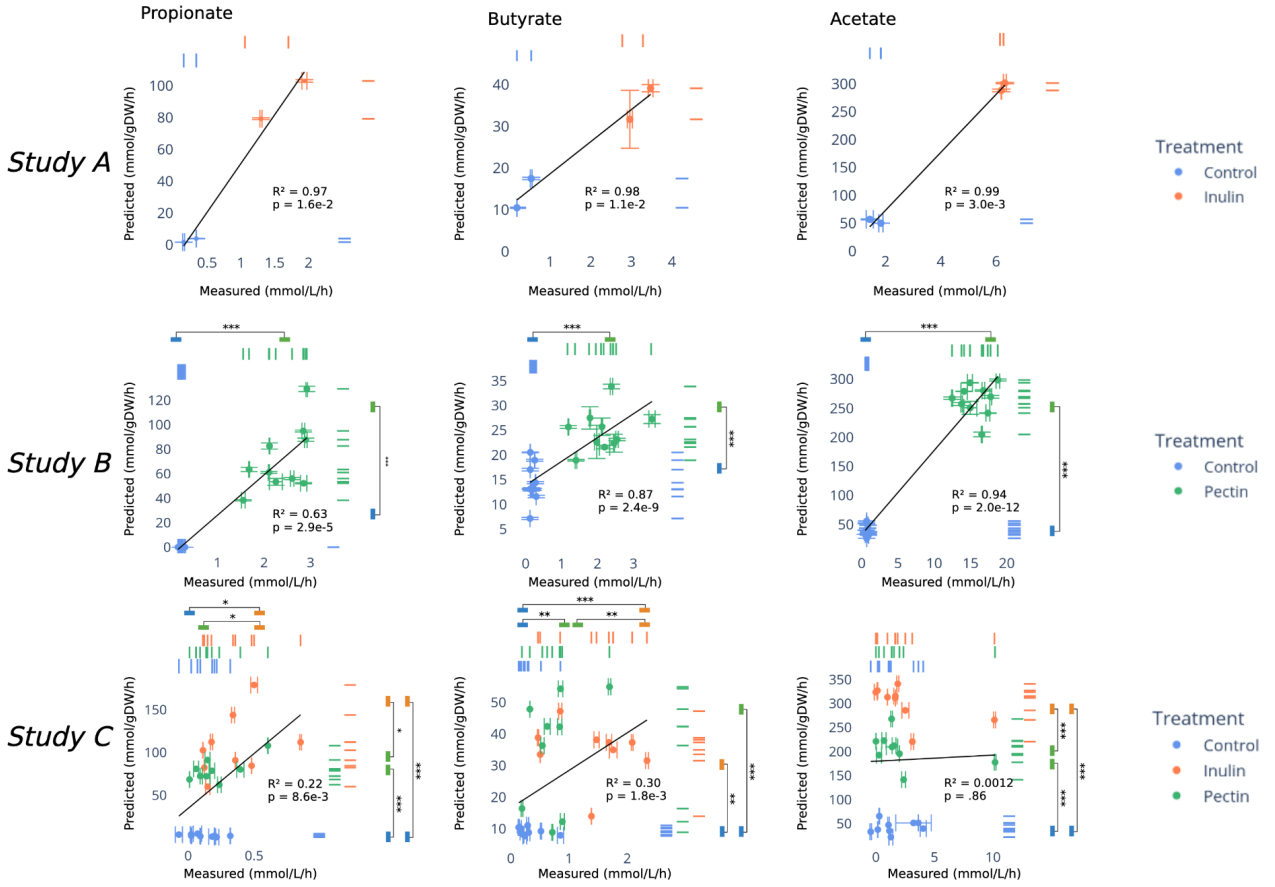
Figure 6. Microbial MCMMs can be used to design, build, and test personalized prebiotic, probiotic, and dietary interventions aimed at optimizing SCFA production profiles. (A) MCMMs built from the Arivale cohort (N = 2,687) were used to test personalized responses to dietary interventions. Personalized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

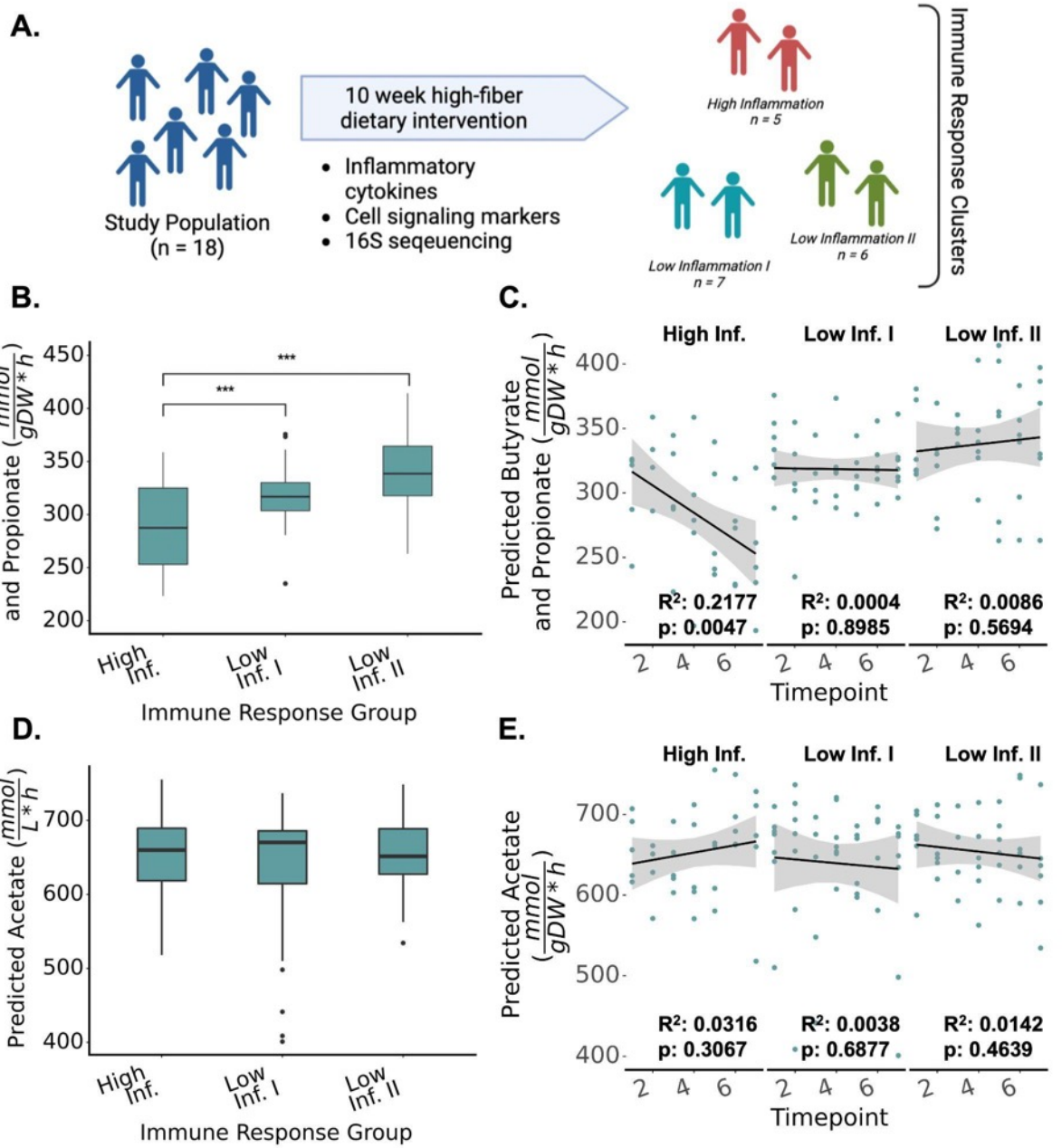

A. growth medium
initial metabolic profile
map to metabolic components
adjustments, interventions
complete profile
B. model construction
medium
genomic data
taxa
abundances
MCMM
biomass
taxon models
curated metabolic reconstruction
C. cooperative tradeoff flux balance analysis (ctFBA)
maximize community growth rate
balance taxon growth rates and estimate fluxes
estimates for exchanges
total SCFA production rate
D. ex vivo cultivation
homogenized fecal sample
OR
in vitro co-culture
optional fiber supplementation
anaerobic 5% H2, 20% CO2 37°C
multiple timepoint measurments
targeted mass spectrometry
slope=flux
SCFA
time
E. validation
measured SCFA flux [mmol/h]
predicted SCFA flux [mmol/gDW/h]

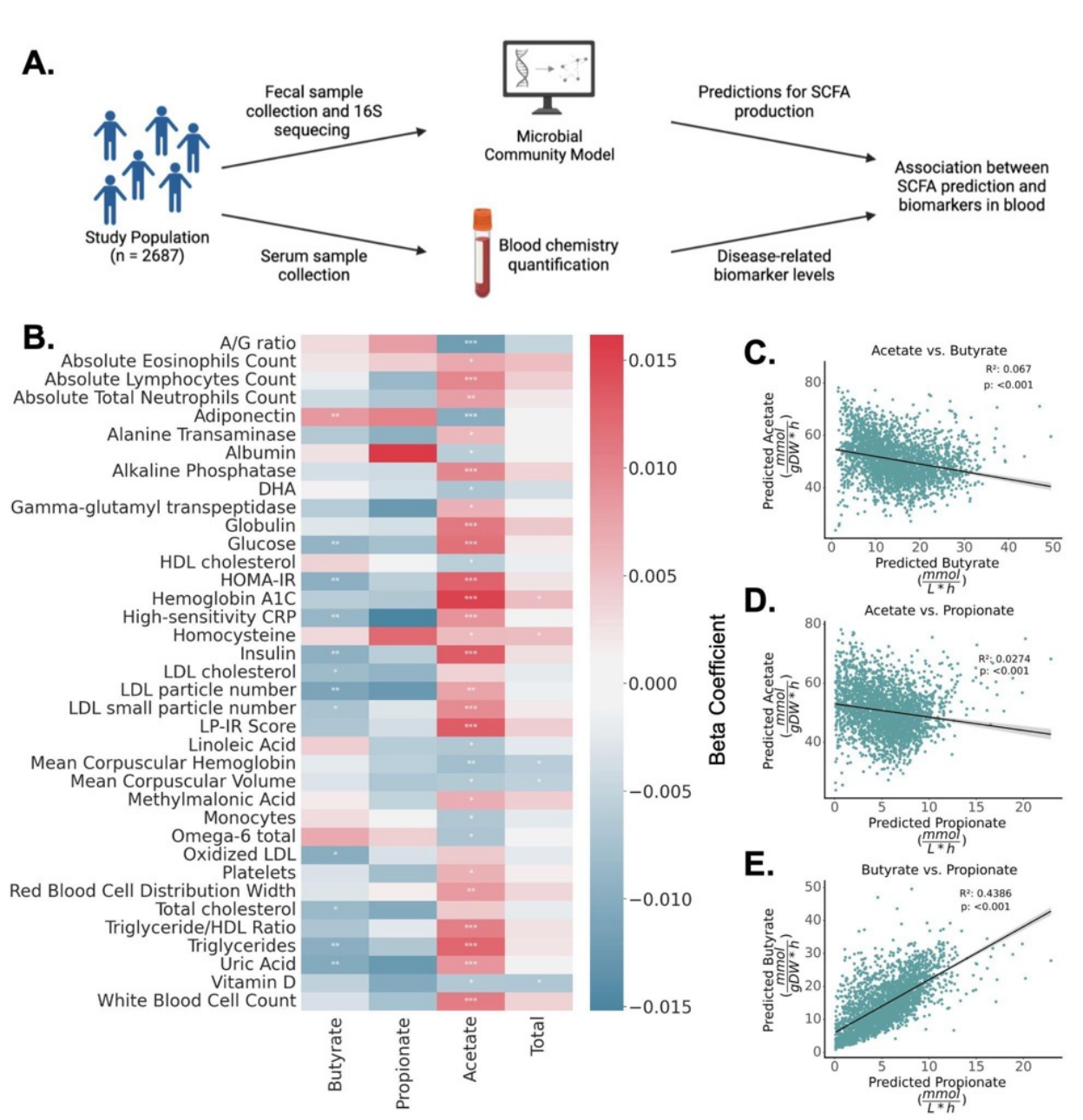

A.
16S Sequencing
Model Building & Butyrate Prediction
Predicted
Measured
Sample Collection
Anaerobic Culturing
Compare Predicted vs. Measured Fluxes
Endpoint Butyrate Measurement

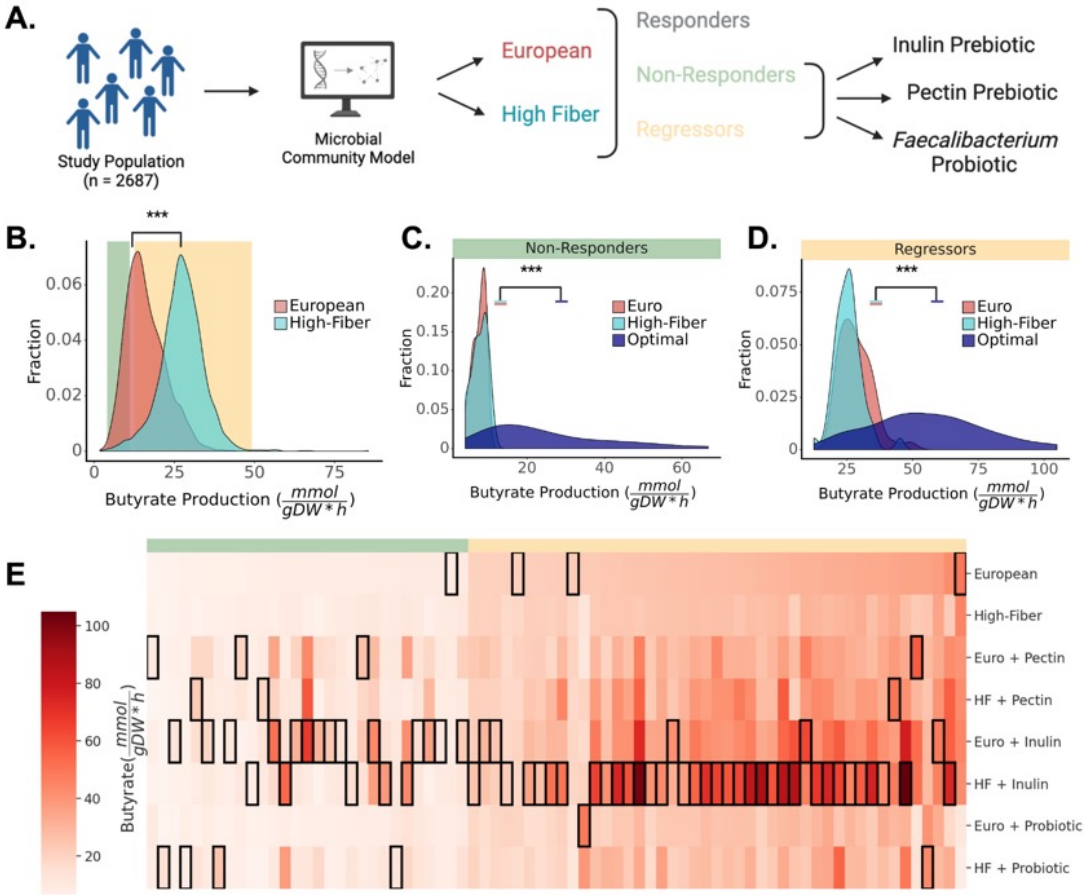

B.
Butyrate
Predicted (mmol/gDW/h)
Measured (mmol/OD/h)
R² = 0.028
p = 5e-4
1-5 Member Communities


C.
Butyrate
Predicted (mmol/gDW/h)
Measured (mmol/OD/h)
R² = 0.277
p = 6e-51
10-25 Member Communities


Propionate
Butyrate
Acetate
Study A
Study B
Study C
Predicted (mmol/gDW/h)
Measured (mmol/L/h)
R² = 0.97
p = 1.6e-2
R² = 0.98
p = 1.1e-2
R² = 0.99
p = 3.0e-3
R² = 0.63
p = 2.9e-5
R² = 0.87
p = 2.4e-9
R² = 0.94
p = 2.0e-12
R² = 0.22
p = 8.6e-3
R² = 0.30
p = 1.8e-3
R² = 0.0012
p = .86
Treatment
Control
Inulin
Pectin


A.
Study Population
(n = 18)
10 week high-fiber
dietary intervention
Inflammatory cytokines
Cell signaling markers
16S seqeuencing
High Inflammation
n = 5
Low Inflammation I
n = 7
Low Inflammation II
n = 6
Immune Response Clusters
B.
Predicted Butyrate and Propionate (mmol/gDW*h)
***
***
High Inf.
Low Inf. I
Low Inf. II
Immune Response Group
C.
High Inf.
Low Inf. I
Low Inf. II
R²: 0.2177
p: 0.0047
R²: 0.0004
p: 0.8985
R²: 0.0086
p: 0.5694
Timepoint
D.
Predicted Acetate (mmol/L*h)
Immune Response Group
E.
R²: 0.0316
p: 0.3067
R²: 0.0038
p: 0.6877
R²: 0.0142
p: 0.4639
Timepoint


A.
Fecal sample collection and 16S sequecing
Microbial Community Model
Predictions for SCFA production
Association between SCFA prediction and biomarkers in blood
Study Population (n = 2687)
Serum sample collection
Blood chemistry quantification
Disease-related biomarker levels
B.
A/G ratio
Absolute Eosinophils Count
Absolute Lymphocytes Count
Absolute Total Neutrophils Count
Adiponectin
Alanine Transaminase
Albumin
Alkaline Phosphatase
DHA
Gamma-glutamyl transpeptidase
Globulin
Glucose
HDL cholesterol
HOMA-IR
Hemoglobin A1C
High-sensitivity CRP
Homocysteine
Insulin
LDL cholesterol
LDL particle number
LDL small particle number
LP-IR Score
Linoleic Acid
Mean Corpuscular Hemoglobin
Mean Corpuscular Volume
Methylmalonic Acid
Monocytes
Omega-6 total
Oxidized LDL
Platelets
Red Blood Cell Distribution Width
Total cholesterol
Triglyceride/HDL Ratio
Triglycerides
Uric Acid
Vitamin D
White Blood Cell Count
Butyrate
Propionate
Acetate
Total
Beta Coefficient
0.015
0.010
0.005
0.000
−0.005
−0.010
−0.015
C.
Acetate vs. Butyrate
R²: 0.067
p: <0.001
Predicted Acetate (mmol/gDW*h)
Predicted Butyrate (mmol/L*h)
D.
Acetate vs. Propionate
R²: 0.0274
p: <0.001
Predicted Acetate (mmol/gDW*h)
Predicted Propionate (mmol/L*h)
E.
Butyrate vs. Propionate
R²: 0.4386
p: <0.001
Predicted Butyrate (mmol/gDW*h)
Predicted Propionate (mmol/L*h)


A.
Study Population
(n = 2687)
Microbial
Community Model
European
High Fiber
Responders
Non-Responders
Regressors
Inulin Prebiotic
Pectin Prebiotic
Faecalibacterium
Probiotic
B.
***
Fraction
European
High-Fiber
Butyrate Production (mmol/gDW*h)
C.
Non-Responders
***
Euro
High-Fiber
Optimal
Fraction
Butyrate Production (mmol/gDW*h)
D.
Regressors
***
Euro
High-Fiber
Optimal
Fraction
Butyrate Production (mmol/gDW*h)
E
Butyrate(mmol/gDW*h)
European
High-Fiber
Euro + Pectin
HF + Pectin
Euro + Inulin
HF + Inulin
Euro + Probiotic
HF + Probiotic