

Microbial Community-Scale Metabolic Modeling: Personalized Predictions for Short-Chain Fatty Acid Production from the Human Gut Microbiome

Nicklas Quinn-Bohmann

A dissertation

submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Sean Gibbons, Chair

Nathan Price

Amy Willis

Program Authorized to Offer Degree:

Molecular Engineering and Sciences

©Copyright 2024

Nicklas Quinn-Bohmann

University of Washington

Abstract

Microbial Community-Scale Metabolic Modeling: Personalized Predictions for Short-Chain Fatty Acid Production from the Human Gut Microbiome

Nicklas Quinn-Bohmann

Chair of the Supervisory Committee:

Sean Gibbons

Institute for Systems Biology

Small-molecule production in response to dietary input is a primary mechanism in which the gut microbiome can affect the health of the host. 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, individuals exposed to identical diets often show significant variation in the level of produced SCFAs, which remains poorly understood at a systems scale. Addressing this gap, we propose employing microbial community-scale metabolic models (MCMMs) to predict individual-specific SCFA production profiles. To validate this approach, we leverage *in vitro* and *ex vivo* fluxomic data across several independent datasets. Furthermore, the clinical relevance of MCMMs is illustrated by associating predicted SCFA profiles with physiological markers of inflammation and cardiometabolic health, and by demonstrating the use of this approach to design and test precision microbiome-targeted interventions. Finally, work towards integration of MCMMs with a metabolic model of the human colonic epithelium allows for a comprehensive exploration of metabolic exchanges

influencing health and disease. Ultimately MCMMs offer a promising avenue for the future of precision nutrition, aimed at modulating metabolic production from the gut microbiome to benefit human health.

Table of Contents

<i>List of Figures</i>	<i>iii</i>
<i>List of Tables</i>	<i>iv</i>
<i>List of Abbreviations</i>	<i>v</i>
<i>Acknowledgements</i>	<i>vi</i>
<i>Chapter 1. Introduction</i>	<i>1</i>
1.1 Health-Associated Functions of the Gut Microbiome.....	1
1.2 Small Molecule Production by the Gut Microbiome	2
1.3 Computational Modeling of the Gut Microbiome.....	6
1.4 Thesis Objectives and Overview of Chapters	14
<i>Chapter 2. Validation of MCMM-Based Predictions with in vitro and ex vivo Data</i>	<i>18</i>
2.1 Introduction	18
2.2 Methodology	20
2.3 Results	23
2.4 Discussion.....	26
<i>Chapter 3. Analyzing the Clinical Relevance of MCMM-Based Predictions.....</i>	<i>41</i>
<i>with in vivo Data</i>	<i>41</i>
3.1 Introduction	41
3.2 Methodology	43
3.3 Results	45
3.4 Discussion.....	46
<i>Chapter 4. Using MCMMs to Design and Test Personalized Interventions</i>	<i>56</i>

4.1 Introduction	56
4.2 Methodology	58
4.3 Results	60
4.4 Discussion.....	61
<i>Chapter 5. Toward Integration of MCMMs with a Metabolic Reconstruction of the Colonic Epithelium</i>	68
5.1 Introduction	68
5.2 Methodology	70
5.3 Results	72
5.4 Discussion.....	73
<i>Chapter 6. Study Limitations, Future Directions, and Significance</i>	82
6.1 Study Limitations	82
6.2 Future Directions.....	83
6.3 Significance	85
<i>References</i>	87

List of Figures

Figure 1.1: Visual representation of the flux balance analysis methodology

Figure 1.2: Cooperative tradeoff strategy in MICOM

Figure 2.1. Relationship between predicted and measured butyrate production rates in *in vitro* co-cultures.

Figure 2.2 Relationship between predicted and measured butyrate production rates in fiber-treated *in vitro* co-cultures

Figure 2.3 Relationship between predicted and measured butyrate production rates in longitudinally-sampled *in vitro* co-cultures

Figure 2.4 Stool *ex vivo* assays show strong associations between measured and predicted SCFA production fluxes

Figure 2.5 16S amplicon vs shotgun metagenomics for MCMM construction

Figure 2.6 MCMM prediction accuracy in genus- vs. species-level models

Figure 2.7 Alpha diversity in *ex vivo*-derived MCMMs

Figure 2.8 Bacterial biomass on measured SCFA production

Figure 3.1 Predicted SCFA production profiles were associated with variable immune response groups following a high-fiber dietary intervention

Fig 3.2 Alpha diversity in longitudinal high fiber study

Figure 3.3 SCFA flux predictions are significantly associated with blood-derived clinical markers

Figure 3.4 Association of MCMM-predicted butyrate with blood-measured butyrate

Figure 4.1. Microbial MCMMs can be used to design and select personalized prebiotic, probiotic, and dietary interventions aimed at optimizing SCFA production profiles

Figure 4.2. Predicted butyrate production associates with physiological response to HAMSAB diet

Figure 5.1 Compartmentalization of integrated host-microbiome multi-system model

Fig 5.2 Objective function fluxes from human colonocyte model

Figure 5.3 Associations between predicted colonic imports and blood-based clinical labs

List of Tables

Table 2.1 List of *in vitro* and *ex vivo* datasets leveraged for MCMM validation

Table 2.2 Within treatment group correlations from MCMMs for *ex vivo* cultures

Table 3.1 List of *in vivo* datasets leveraged for analysis of clinical relevance

Table 3.2 Clinical measures showing significant association with MCMM-predicted butyrate production

Table 4.1 List of datasets leveraged for dietary design and response prediction

Table 5.1 Metabolic tasks for colonocyte model

Table 5.2 Data inclusion for colonocyte reconstruction

Table 5.3 Colonocyte model objective function

List of Abbreviations

Abbreviation	Definition
AGORA	Assembly of gut organisms through reconstruction and analysis
AGORA2	Assembly of gut organisms through reconstruction and analysis, version 2
BMI	Body mass index
CRC	Colorectal cancer
CRP	C-reactive protein
FBA	Flux balance analysis
FDR	False detection rate
FFAR2	Free fatty acid receptor 2
FFAR3	Free fatty acid receptor 3
FMO3	Flavin monooxygenase 3
FOS	Fructooligosaccharides
GABA	γ -aminobutyric acid
gDW	Gram dry weight
GEM	Genome-scale metabolic model
GLP-1	Glucagon-like peptide-1
GPR41	G-protein coupled receptor 41
GPR43	G-protein coupled receptor 43
HDAC	Histone deacetylase
HDL	High-density lipoprotein
IL-6	Interleukin 6
LDL	Low-density lipoprotein
LP-IR	Lipoprotein-insulin resistance score
MCOMM	Microbial community-scale metabolic model
ML	Machine learning
NLRP3	NOD-like receptor protein 3
NRP-1	Neuropilin 1
OD	Optical density
RECON3D	Human metabolic reconstruction, version 3
SCFA	Short-chain fatty acid
TLR2	Toll-like receptor 2
TMA	Trimethylamine
TMAO	Trimethylamine N-Oxide
Treg	Regulatory T-cells
VMH	Virtual Metabolic Human Database

Acknowledgements

I will begin by thanking my advisor, Dr. Sean Gibbons. Since joining the lab during the peak of the COVID-19 pandemic in 2020, Sean has consistently provided steadfast support and thoughtful guidance through every step of my doctoral work. His dedicated guidance and unwavering belief in me helped me grow as a scientist and find success in ways I could never have accomplished alone. I am deeply thankful for the opportunity to have worked with Sean for the past four years, and I know his mentorship will continue to shape my growth even after I depart from the Gibbons Lab.

I would like to thank the entirety of my advisory committee, Dr. Nathan Price, Dr. Amy Willis, and Dr. Lauren Rajakovich. They have all provided invaluable insight to my work throughout the research process and encouraged me to approach problems from new angles. Their thoughtful advice and continued support through every milestone helped me grow as a researcher and develop a project I am proud of.

I am very grateful for Dr. Christian Diener's guidance during my tenure at ISB. He not only expanded my understanding of microbiome science and metabolic modeling beyond what I imagined possible, but also did so with constant patience and contagious enthusiasm. Any successes I've experienced thus far are undoubtedly influenced by his mentorship.

I would also like to extend my gratitude to the rest of the Gibbons Lab, as well as the entire ISB community, for cultivating an atmosphere of collaboration and growth that played a pivotal role in shaping me as a scientist. Their support and camaraderie have been invaluable throughout my journey, and the friendships will last long after my departure.

I am truly thankful for the support of all my family and friends, from Seattle to Rochester to Virginia and everywhere in between. You have all been there along the way to celebrate with me through the ups and support me through the downs, and I could not be here today without each of you. Special thanks to Dr. Ovitt and Drs. Dirk and Theo Bohmann for the lifelong inspiration to get to this point.

Finally, thank you to my partner. Throughout everything, you've kept me grounded and constantly excited for what lies ahead. I am grateful for your constant support and thoughtfulness, and for being the project manager of our lives. Thank you, Tay.

Chapter 1

Introduction

1.1 Health-Associated Functions of the Gut Microbiome

The human gut microbiome, consisting of a rich diversity of bacteria, archaea, fungi, protists, and viruses, plays an integral role in regulating various physiological functions critical for the maintenance of host health ¹⁻³. Housed within the colonic lumen, the microbiome provides several functions critical to human health. Development of the immune system is a primary function of the microbiota ⁴. During development, presentation of foreign stimuli originating from the microbiome leads to modulation of immune cell response and the expansion of adaptive immunity in the host ⁵. These stimuli can take the form of secondary metabolites produced by the microbiome, foreign molecular patterns or foreign antigens ⁵. In germ-free mice lacking immune education by the microbiome, a significant decrease in immune response to infection and concordant increase in mortality was observed when compared with conventionally colonized mice ⁶. Essentially, the microbiome trains and modulates the immune system to discriminate pathogenic microbes from commensals and respond appropriately.

Another contribution of the gut microbiome to host health centers on maintaining the structure and function of the gut barrier ⁷. Primarily, this occurs through the maintenance of a robust mucus layer and healthy tight junctions between cells in the colonic epithelium. Certain bacterial species such as *Bacteroides thetaiotaomicron* and *Lactobacillus rhamnosus* aid in maintaining tight junction integrity through the induction of specific proteins such as sprr2A, p40 and p75 ⁸. Additionally, the transport of microbially-derived metabolites from the gut microbiota to the host contributes to the maintenance of homeostasis of cells in the colonic epithelium, maintaining to the proper function of the gut barrier ⁹.

A process fundamental to the effect of the gut microbiome on human health is the production and subsequent transport of small molecules, also called microbially-derived metabolites, to the host through the colonic epithelium ^{10,11}. This exchange of metabolites at the microbiome-host interface is mediated by dietary inputs, taxonomic composition and spatial structure of the microbiota, cross-feeding between taxa, as well as host physiology. Microbially-produced metabolites are implicated in an abundance of host-related functions, from immunomodulation through histone deacetylase (HDAC) inhibition to neurotransmission by γ -aminobutyric acid (GABA) ^{12,13}. While directly measuring small molecule production and consumption rates from the microbiome is difficult due to interference from the host, computationally predicting these rates is an increasingly attractive research avenue.

1.2 Small Molecule Production by the Gut Microbiome

1.2.1 Short-Chain Fatty Acids

Many microbially-produced metabolites transported across the luminal wall are beneficial, or even critical, for the health of the host. Short-chain fatty acids (SCFAs) are one

such class, with several salient functions in the body. These metabolites are primarily produced by the microbiome through the fermentation of indigestible dietary fibers and resistant starches¹⁴. The most commonly produced SCFAs are acetate, butyrate, and propionate, accounting for upwards of 95% of fatty acid production in microbiome, generally in a ratio of 60:20:20, respectively¹⁵. Microbial production of butyrate is of particular interest, as it provides several well-established health benefits in humans^{16,17}. Butyrate acts as the primary energy source for the colonic enterocytes (i.e. colonocytes) that line the colon, thereby aiding in sustaining healthy gut barrier function through homeostatic maintenance and the development of tight junctions, as alluded to previously⁹. It is also important in energy balance¹⁸, regulating systemic lipid and glucose metabolism, has been shown to ameliorate insulin resistance and can be preventative of diabetes^{19,20}. Increased butyrate production by the gut microbiome has even been shown to benefit the reduction of obesity²⁰. Critically, butyrate production by the gut microbiome is a primary method through which the immune system is regulated by secondary microbially-produced metabolites, as described above. Butyrate acts as a systemic inhibitor of inflammation via immune activation of regulatory T-cells (Tregs)²¹ through epigenetic regulation via histone deacetylase (HDAC) inhibition¹². Inhibition of HDAC results in chromatin remodeling at the FoxP3 locus in CD4⁺ T cells, thereby activating the immunosuppressive function of these immune cells. Butyrate also acts as a ligand for GPR41 and GPR43 (also known as FFAR2 & FFAR3) on the surface of colonic enterocytes²². Upon binding to these receptors, butyrate triggers intracellular signaling pathways that lead to the inhibition of anti-inflammatory responses²². This includes the production of anti-inflammatory cytokines, further promotion of regulatory T cell differentiation, and enhancement of intestinal barrier function via maintenance of tight junctions⁹. Overall, the interaction between butyrate and GPR41 and GPR43 plays a

crucial role in maintaining gut homeostasis and modulating immune responses in the colon. Butyrate even exhibits anticarcinogenic effects, specifically toward colorectal cancer (CRC) cells: it decreases NRP-1 expression to suppress angiogenesis and metastasis, inhibits CRC cell proliferation and induces CRC cell apoptosis via activation of the Wnt pathway and HDAC inhibition ^{23,24}.

Other SCFAs produced by the gut microbiome show similar, but distinct properties. Both acetate and propionate show some HDAC-inhibiting activity, but neither is as potent as butyrate ²⁵. Propionate downregulates hepatic lipogenesis and increases insulin sensitivity in obesity, and has been shown to reduce serum cholesterol, protecting against cardiovascular disease ^{26,27}. As with butyrate, propionate has been shown to inhibit cytokine production and inflammation via binding of GPR41 and GPR43 ²⁸. Acetate controls the release of satiety-stimulating hormones from the gut including GLP-1 and peptide YY, helping with weight control and metabolic regulation, although it is less potently anti-inflammatory than either butyrate or propionate ²⁹. Microbial production of SCFAs also lowers the pH of the colonic environment, which has been associated with increased absorption of vitamins, electrolytes and iron, while minimizing absorption of ammonia and other amines ³⁰. Low pH has also been found to combat the invasion of pathogenic taxa like *Clostridioides difficile* ³⁰

1.2.2 Heterogenous SCFA production between individuals

Certainly, methods for augmenting SCFA production from the gut microbiome can be beneficial to the host by increasing the availability of these necessary metabolites. It is well established that diets high in fiber and omega-3 fatty acids increase SCFA production from the microbiota ³¹. However, functional capacity for SCFA production when given the same dietary

input varies significantly between individuals, with some individuals showing much larger increases in SCFA production than others upon fiber supplementation ³². In some cases, switching to a high-fiber diet can have negative effects, increasing bile acid production from the microbiome and leading to inflammation ³³. For instance, communities lacking the capacity to degrade the dietary fiber inulin exhibit elevated inflammation through activation of the NLRP3 and TLR2 pathways by unfermented β -fructans, and indirectly via changes to the bile acid pool ^{33,34}. Given the unique composition and the heterogeneous metabolic outputs of every individual's microbiome, methods for computationally predicting response to dietary interventions are an attractive research avenue in precision nutrition.

1.2.3 Selected examples of other microbially-derived molecules

Not all molecules produced by the gut microbiome are beneficial to the host. One such metabolite, trimethylamine (TMA), is produced by the microbiome from the dietary precursors choline and L-carnitine, most commonly found in red meat and eggs ^{35,36}. Upon transport into the host, TMA is oxidized to trimethylamine N-oxide (TMAO) in the liver by flavin monooxygenase (FMO3) ³⁷. Circulating TMAO has been identified as a significant risk factor for cardiovascular disease, metabolic syndrome and overall mortality ^{38,39}. As with SCFA production, due to differences in the abundance of TMA-producing species, some individuals show higher levels of TMA production than others given the same dietary input ³⁶. As opposed to SCFA production, a targeted intervention aimed at reducing TMA production from the gut microbiome would be beneficial to at-risk patients.

Secondary bile acids are produced by the gut microbiome through the deconjugation and dehydroxylation of primary bile acids, primarily in the colon ⁴⁰. These secondary bile acids, such

as deoxycholic acid and lithocholic acid, play a crucial role in various physiological processes, including fat digestion and absorption ⁴¹. However, they can also exert cytotoxic effects on colonic epithelial cells, potentially contributing to colorectal cancer development and inflammatory bowel diseases, highlighting the complex interplay between the gut microbiota and human health ⁴⁰.

Indoles are metabolites produced by the gut microbiota from dietary tryptophan ⁴². These compounds have been shown to exert diverse effects on human health, including modulation of the immune system, regulation of intestinal barrier function, and influence on neurotransmitter pathways ^{42,43}. Indoles have also been associated with anti-inflammatory and anti-carcinogenic properties, suggesting their potential role in protecting against certain diseases such as inflammatory bowel disease and colorectal cancer ^{42,44}. However, the precise mechanisms underlying these effects and their overall impact on human health require further investigation.

1.3 Computational Modeling of the Gut Microbiome

1.3.1 Benefits of Mechanistic Modeling

A growing interest in the gut microbiome as a clinical target has driven advances in computationally modeling aimed at predicting the growth behavior and metabolic capacity of these microbial communities ⁴⁵. Statistical modeling and machine-learning (ML) approaches for predicting metabolic output from the microbiome have shown some promise results in recent years. For example, postprandial blood glucose responses can be predicted by machine-learning algorithms trained on large human cohorts ^{46,47}. Other ML-directed studies have successfully predicted human-bacteria protein-protein interactions, and even developed a classifier of overall health status based on microbial composition of the microbiome ^{48,49}. Despite this progress, ML

methods are limited by the measurements, heterogeneity, and interventions represented within the collected training data, making it difficult to model the output of several combinatorial interventions or to make predictions in human populations that are underrepresented in research cohorts. Considering the massive combinatorial space of possible taxonomic composition in the gut microbiome, to which a nearly infinite number of possible diets or interventions could be applied, this is a difficult problem. The amount of data necessary to train and validate such a comprehensive ML model would be too prohibitively large, given current technologies.

Mechanistic models of microbial metabolism, on the other hand, do not rely on training data. Instead, these models are constructed using deep mechanistic understanding of all the reaction pathways present in a given system. This makes it possible to extrapolate to arbitrary combinations of communities or interventions. While limited by our knowledge bases, these approaches often show more consistent performance across populations and contexts ⁴⁵. The capability of mechanistic models to capture high dimensional microbial abundance data and apply any of a set of boundless possible dietary inputs represents a significant advantage over ML methods. Additionally, due to the mechanistic natures of these models, the organisms and pathways responsible for producing a given metabolite are accounted for. This enables the exploration of causal relationships between variables, which is not generally possible in mechanistically-opaque statistical models ⁴⁵.

1.3.2 Flux Balance Analysis

Flux balance analysis (FBA), a methodology for solving metabolic models, has been increasingly employed as a method to understand microbial systems and interrogate their metabolic capability ^{50–53}. Classical FBA makes two primary assumptions. The first is that the

system being modeled is at steady state; that is, the rate of production and the rate of consumption of metabolites within the system remain constant over time ⁵³. The steady state assumption differentiates FBA modeling from dynamic models, as all rates are fixed. This assumption is appropriate when modeling bacterial growth in the log phase, as growth rate is essentially constant in this regime. The second assumption is that the system is optimized for a biological goal, which is most commonly biomass production in simple bacterial models ^{54,55}. This goal is subsequently used to define an objective function, a reaction or set of reactions to be maximized in the optimal solution. Flux balance analysis works by first determining a reaction network, representing all the metabolic processes taking place in a system (**Fig. 1.1A**). This can be done through a combination of biochemical knowledge, genome annotation, and computational gap filling. This reaction network is then represented as a stoichiometrically consistent matrix, denoted **S** (**Fig. 1.1B**) ⁵⁶. In this matrix, each reaction is represented as a row, and each metabolite as a column. The stoichiometric coefficient of a metabolite in a given reaction is found in the corresponding cell. Applying the steady state assumption, one can constrain the set of possible fluxes (**v**) through matrix multiplication via the following equation.

$$S \cdot v = 0$$

This yields a limited solution set of possible fluxes. Additionally, constraints applied to individual fluxes further limit the possible solutions, which can be represented with iteration of the following equation, wherein a_i represents the lower flux bound and b_i represents the upper flux bound.

$$a_i < v_i < b_i$$

Individual fluxes can be constrained based on thermodynamic limitations, metabolite availability, or other knowledge of the system. Directionality of reactions can also be applied via individual flux constraints: for instance, a lower bound (a_i) of zero allows only the forward reaction to occur, and vice versa. Application of the steady state assumption and individual bounds results in a minimized subset of solutions for the flux vector $\mathbf{v}$ which satisfies all constraints (**Fig. 1.1D**). Finally, a single solution within the minimized solution space is selected through linear programming which maximizes the following expression, wherein c^T is the transpose of the coefficient vector, containing the coefficients of the objective function, and v is the vector of fluxes to be determined.

$$c^T v$$

Through maximization of this expression, FBA finds the solution of fluxes in vector $\mathbf{v}$ which optimizes the value of the objective function (**Fig. 1.1E**).

FBA is well validated for predicting the metabolic capacity of an individual genome-scale metabolic model, and is commonly used in biomanufacturing⁵⁴. Metabolic modeling of individual commensal taxa has been used to predict plasma concentrations of microbially derived metabolites, but special considerations are required to extend these methods into diverse, real-world microbiomes^{57,58}.

1.3.3 Microbial Community-Scale Metabolic Modeling

While FBA is a robust and useful approach for genome-scale metabolic modeling, direct application to microbial community-scale metabolic models (MCMs), such as those representing the gut microbiome, requires additional considerations. For this purpose, we utilize MICOM - a tool specifically built for flux balance analysis of microbial communities⁵⁸.

Community growth rate in such a system can be described by the sum of the growth rates of individual taxa, scaled by their relative abundance in the community, illustrated by the following equation, where μ_c is community growth rate, b_i is relative abundance of a given taxon, and μ_i is its individual growth rate.

$$\mu_c = \sum_i b_i \mu_i$$

In a community with several members, however, no unique solution exists for a set of fluxes that maximizes community growth rate - rather, there is an infinite set of solutions that satisfies this criterion (**Fig. 1.2A**). Additionally, when introducing community biomass production as an objective, many of these solutions optimize the growth of fast-growing taxa while sacrificing the growth of slower-growing taxa. When sampling the gut microbiome, however, we expect that taxa identified in a sample should exhibit a non-zero growth rate, to avoid dilution-to-extinction. To overcome this, MICOM utilizes an L2 regularization strategy, finding an optimal solution wherein the sum of the squares of the individual growth rates is minimized by the expression shown below.

$$\text{minimize } \sum_i \mu_i^2$$

$$\text{subject to } \mu_c = \max \mu_c$$

This constraint finds the shortest path from a region of no growth to a region that reflects the steady state relative biomass optimum of all the taxa in the system, as measured by 16S or metagenomic sequencing, returning one unique solution where growth rates are distributed more evenly across the community (**Fig. 1.2B**). In many situations, the optimal community growth rate as calculated by FBA is only achievable with the exclusion of most taxa. To address this,

MICOM also includes a cooperative tradeoff parameter, wherein L2 regularization finds a unique solution at a predefined suboptimal fraction of the maximal community growth rate where most of the taxa present can grow, called a tradeoff parameter (**Fig. 1.2C**). This is shown in the expression below, where α is the tradeoff parameter.

$$\begin{aligned} & \text{minimize } \sum_i \mu_i^2 \\ & \text{subject to } \mu_c \geq \alpha * \max \mu_c \end{aligned}$$

This method has been validated to return biologically relevant results for community and individual growth rates, not achieved by using FBA alone ⁵⁸.

As mentioned previously, flux bounds placed on individual reactions constrain the solution space when finding an optimal solution. When constructing MCMMs to represent the gut microbiome, specific bounds are placed on the exchange reactions that control the import of metabolites into the environment ⁵⁹. These bounds are defined by the *in silico* growth medium, representing the dietary components available to the microbiota. Tuning the composition of this medium allows context specific solutions to be predicted; for instance, the predicted metabolic behavior of a community on a healthy diet can be compared to that of an unhealthy diet. Importantly, due to the mechanistic nature of MCMMs, any combination of prebiotic, probiotic or dietary interventions can be tested.

MCMMs are constructed using existing knowledge bases, including curated genome-scale metabolic models (GEMs) of individual taxa. MCMMs are limited by the availability of well-curated GEMs for abundant taxa present within every individual in a population and by information on individual-specific dietary variation. These limitations are further exacerbated in human populations that are generally underrepresented in microbiome research, where our

databases are also less representative. However, as our knowledge bases grow, so too will the power and scope of MCMMs. Using MICOM as a platform, personalized community-scale metabolic models of the gut microbiome can be constructed, and simulated within any dietary context. Overall, MCMMs have the potential to serve as powerful, transparent, knowledge-driven tools for predicting community-specific responses to a wide array of interventions or perturbations of a dietary, prebiotic or probiotic nature.

1.3.4 Tissue-Specific Metabolic Modeling

In addition to community-scale metabolic models of the gut microbiome community, mechanistic modeling can also be applied to human tissues and cells ^{60,61}. Tissue-specific metabolic modeling is an established methodology for predicting the metabolic behavior of a specific cell, organ or system in the body ^{61,62}. Using cell- or tissue-specific gene expression data, reaction networks can be constructed from a generic human model, such as RECON3D ⁶³. Generally, expression-based and connectivity-based evidence is used to determine the likelihood of reactions taking place in a given tissue, and a set of core reactions critical to tissue function is identified ⁶². Non-core reactions are iteratively pruned from the generic human model like RECON3D, assuming removal does not affect flux through core reactions. One is then left with a reaction set specific to the tissue or cell of interest, based on gene-expression evidence. This reaction set is then used to construct a stoichiometric matrix as described previously. Using classical FBA, a unique solution can be found for the reactions in this model. An important consideration in the construction of such a model is the objective function used in model optimization. For bacterial genome-scale metabolic models, biomass optimization makes sense, since the primary biological goal of a microbe is to grow. However, such an optimum is not physiologically relevant for a model of a human tissue, as they do not exhibit constant growth.

Instead, an objective focused on the primary metabolic functions of a given tissue is required. For instance, an applicable objective function for a model of the colonic epithelium would be β -oxidation of fatty acids, as SCFAs from the microbiome provide most of the cellular energy for these tissues ⁶⁴. This process is described in more detail in Chapter 5.

Several robust tools for constructing tissue-specific models exist, including mCADRE, tINIT, CORDA and iMAT ^{60, 62,65,66}. Tissue-specific modeling can give insight into the context-specific metabolic function of a given tissue, and further understanding of how it may interact with other tissues or body systems ⁶⁰.

1.3.5 Multi-System Metabolic Models

To interrogate the interface between metabolic systems within the body, multiple metabolic reconstructions can be combined to construct a multi-compartment metabolic reconstruction ^{67,68}. In this scenario, reactions corresponding to each tissue-specific reconstruction are compartmentalized ⁶⁷. For instance, an integrated host-microbiome model could be constructed with two primary compartments: one containing a community-scale metabolic model of the microbiota, and another containing a tissue-specific reconstruction of the colonic epithelium. Technically speaking, each of these compartments contains additional sub-compartments, segregating reactions as biologically relevant. However, an interesting point of analysis in these models is the metabolic exchange between the primary compartments, in this case the microbiota and the colonic epithelium. Exchange reactions at this interface are limited to those accounting for gene-associated transporter reactions and free diffusion, allowing metabolic transport between compartments ⁶⁷.

Using the resulting model, the complex interplay between the microbiome and the surrounding host tissue can be investigated. This analysis allows for the identification of metabolites consumed by the host, as well as the subsequent production and export of metabolites from the colonic epithelium into the host circulation. This approach of multi-tissue metabolic modeling has proven successful in modeling various organ systems and has even been applied to construct comprehensive whole-body models ^{67–69}. However, to date, no work thus far has shown full predictive capability by a community-scale model of the gut microbiome.

1.4 Thesis Objectives and Overview of Chapters

The objective of this thesis is to demonstrate the practicality of microbial community-scale metabolic modeling in predicting short-chain fatty acid production from the gut microbiome and to showcase its potential impact on human health and wellness. Each chapter contributes to this overarching objective as follows:

Chapter 2 presents a thorough validation of model predictions across various independent datasets, leveraging *in vitro* culturing data and *ex vivo* fluxomic data to validate model outcomes.

Chapter 3 demonstrates the clinical relevance of model-derived predictions concerning human cardiometabolic and inflammatory health using extensively phenotyped human observational and interventional cohorts, emphasizing the significance of these predictions.

Chapter 4 illustrates the application of these models in designing and evaluating dietary, prebiotic, and probiotic interventions, offering a glimpse into the translational potential of this methodology.

Chapter 5 details the integration of existing community-scale models with a model of the human colonic epithelium, indicating the future trajectory of the project and how predictions can be directly linked to human physiology and health.

Chapter 6 discusses the limitations, potential future directions, and overall significance of the research presented, underscoring the progress made in predictive modeling of the gut microbiome and its potential impact on precision nutrition.

Taken together, the content of this dissertation reflects significant advancements in the field of predictive modeling of the gut microbiome, highlighting its potential benefits for precision nutrition.

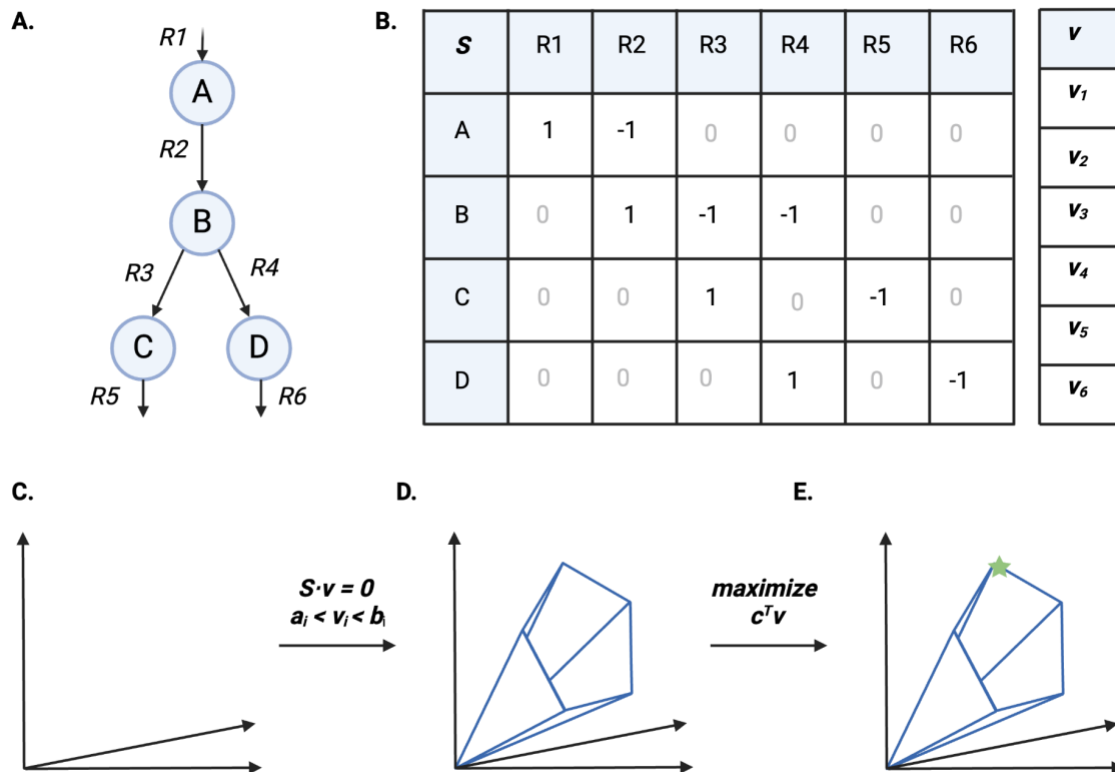


Figure 1.1 Visual representation of the flux balance analysis methodology. Flux balance analysis (FBA) leverages a mechanistic understanding of a system to predict the most likely set of fluxes in the community being modeled. **(A)** To begin, a metabolic network consisting of metabolites and reactions must be identified. **(B)** Next, the reaction network is represented as a stoichiometrically-consistent matrix, with rows representing metabolites and columns representing reactions. Values within the matrix represent the stoichiometric coefficients of a metabolite in a given reaction. Ultimately, FBA seeks to determine the most likely set of fluxes in vector v , representing flux through each reaction. **(C)** Prior to applying constraints the possible solution space for the unknown vector v is infinite. **(D)** The application of the steady-state assumption combined with the imposition of individual flux bounds constrains the feasible solutions within the flux vector space. **(E)** Maximization of the objective function leads to a unique solution of fluxes.

(This figure is adapted from a similar figure found in "What is Flux Balance Analysis?" by Orth et al. (2010)⁵³)

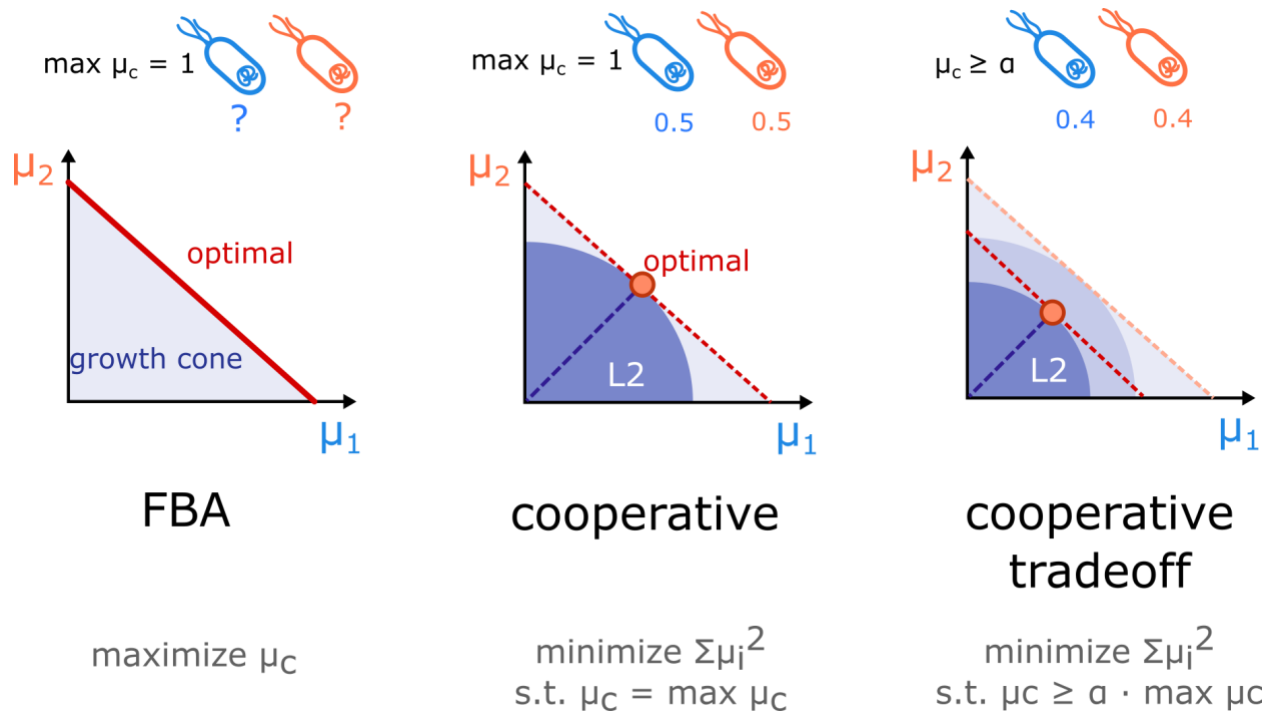


Figure 1.2 Cooperative tradeoff strategy in MICOM. MICOM, a tool for building and simulating MCMs, incorporates a novel strategy for simulating growth of microbial communities. **(A)** In a microbial community, there is an infinite set of combinations that maximize the community growth rate (μ_c), represented by any point along the red line. **(B)** To find a unique solution that distributes growth evenly across community members, an L2 parameter is introduced that minimizes the sum of the squares of the individual growth rates of community members, while maintaining the maximum community growth rate. **(C)** In some cases, maximal community growth can only be achieved at the expense of some slow-growing community members. Here, MICOM introduces a tradeoff parameter that minimizes the L2 parameter while maintaining community growth rate at a predefined fraction of the maximal community growth rate.

(This figure provided courtesy of the Gibbons Lab)

Chapter 2

Construction and Validation of MCMMs with *in vitro* and *ex vivo* Consortia

Portions of this chapter were published as a manuscript in Nature Microbiology ⁷⁰.

2.1 Introduction

Prior to implementing any computational model for practical use, it is essential to thoroughly validate the predicted outcomes to ensure their reliability and accuracy. Accordingly, predictions should be compared against controlled and quantifiable measurements from a system representative of that being modeled. This methodology should be repeated, to ensure not only model accuracy and reliability, but also to identify potential sources of error or uncertainty. Iterative validation allows for refinement of the computational model, leading to improved predictive capabilities and increased confidence in its applicability to real-world scenarios. Additionally, ongoing validation provides insight into the robustness of the model across varying conditions and helps uncover any biases or limitations that may exist.

Microbial production of SCFAs from the gut microbiome is technically challenging to assess *in vivo*, as these metabolites are quickly and preferentially imported by the colonic epithelium, where they enter the beta-oxidation pathway for energy harvesting¹⁷. For this reason, we turned to *in vitro* and *ex vivo* cultures of microbial communities to validate MCMM-predicted SCFA production. In these communities, lack of host consumption of these metabolites allows direct, quantifiable measurements of SCFA production to be compared with model predictions.

In the first case, *in vitro* cultures consisting of microbial communities synthetically constructed from common gut microbial taxa were modeled. These cultures offer a simplified but controlled environment for both constructing and validating models, as all parameters associated with MCMM construction, including composition, abundance, and metabolic availability, are detailed and tunable. Direct measurement of SCFAs at the end of the culturing period can be used to calculate experimental production rates, which are subsequently compared to model-predicted fluxes. While these communities do not fully reflect the complexity of the gut microbiome, they act as a perfect test case for validation. Using data from 1,579 synthetic *in vitro* communities, we investigated the ability of MCMMs to predict butyrate production when grown using a limited, defined medium. Additionally, we validated predictions in a second *in vitro* data set consisting of 1,249 individual cultures, grown using a defined medium including the addition of soluble dietary fibers, specifically inulin or laminarin. Finally, we analyzed the results of a time series analysis of 95 additional *in vitro* cultures, predicting butyrate production rates from cultures sampled across multiple timepoints. Taken together, each of these results underscore the ability of MCMMs to predict SCFA production from microbial communities.

In the second scenario, we utilized data from *ex vivo* cultures derived from fecal samples. These cultures, collected directly from study participants, consist of the unique microbial

communities found within individual-specific microbiomes. This served as a more direct validation, as these communities more thoroughly captured the complexity of the microbiome. Additionally, acetate and propionate production were assessed in these communities, allowing assessment for the prediction of these metabolites, as well. To ensure the validation was robust, we utilized four independent datasets, two from in-house experiments, one from a collaborator, and one publicly available dataset. As these data are directly taken from the patient-derived samples, they provide a proving ground for modeling production of SCFAs from the microbiome itself.

A list of all datasets included in this analysis is available at the end of the chapter (**Table 2.1**).

2.2 Methodology

2.2.1 MCMM construction for in vitro cultures

Models constructed to simulate *in vitro* cultures were built using preprocessed data collected by members of the Venturelli lab at the University of Wisconsin, Madison. In each experiment, cultures were constructed from lab strains and cultured in defined medium under anaerobic conditions for 24 hours, or 48 hours in the case of longitudinally-sampled cultures. Following growth, cultures were analyzed for taxonomic composition, optical density and butyrate production. Endpoint relative abundances included in the collected data, as determined by 16S rRNA amplicon sequencing, were mapped to existing species-level GEMs from the AGORA database (v1.03) to construct MCMMs in MICOM (v0.32.5) with an abundance threshold of 0.001, discarding taxa with less than 0.1% relative abundance in each sample^{58,71}. Sample growth was simulated using an *in silico* medium consisting of each component of the

defined DM38 medium used in the *in vitro* experiments, mapped to metabolites listed in the Virtual Metabolic Human database (VMH, www.vmh.life)⁷². For all models, a cooperative tradeoff parameter of 0.7 was used for simulating growth. Subsequently, predicted production rates were collected for butyrate in each community, to be compared with *in vitro* data. GC-MS measurements of endpoint butyrate were scaled by optical density of each culture, to account for differences in biomass between samples. Endpoint butyrate concentration was divided by culturing time to yield units of mmol/OD/h, given there was no butyrate in the medium at the beginning of the culturing time. These production rates were then compared with model-predicted production given in units of mmol/gDW/h. Associations between measured and predicted butyrate production rates were assessed using Pearson correlation in Scipy (v1.9.1) and significance determined with the associated p-value⁷³. For modeling *in vitro* samples given a similar defined medium with the addition of dietary fibers, the same analysis was performed. However, in this case, the *in silico* medium included the addition of dietary fiber, either inulin or laminarin. For *in vitro* samples assessed at multiple time points, the same analysis was performed. MCMMs for each sample were constructed using sequencing data from each timepoint. The same defined medium (without fiber addition) was used to simulate growth, using identical growth parameters.

2.2.2 MCMM construction for *ex vivo* cultures

In *ex vivo* experiments, models were built using data collected from four independent studies. In each study, fecal samples were collected and homogenized to form a fecal slurry. Homogenate was then split into two groups: experimental samples supplemented with dietary fiber (inulin, pectin, or mixed fructooligosaccharides (FOS)), and control samples that received only a sterile PBS buffer. The final dilution of samples in each study were distinct (Study A 1:5,

Study B 1:19, Study C 1:5, Study D 1:3). Samples were then cultured anaerobically for between 4 and 24 hours, depending on the specific study. 16S (study B) or metagenomic (studies A, C and D) sequencing, as well as mass-spectroscopy-based measures of SCFA production, were collected at the start and end of culturing. Pre-processed baseline sequencing data from these communities was collected and used to construct MCMs for each sample. In each, taxonomic relative abundances were mapped to the AGORA database of curated GEMs. For studies A, C and D, where metagenomic sequencing was available, species level models were used - for study B, where only 16S sequencing was available, models were constructed at the genus level. Models were constructed in MICOM as described previously, with a relative abundance threshold of 0.001. As cultures were grown in fecal homogenate with the addition of dietary fiber, an *in silico* medium was developed to approximate the metabolic availability of this environment. Using an average European diet, we removed easily digestible carbon sources and diluted the remaining metabolites to 10% of their initial concentration. Host-supplied metabolites such as mucins and bile acids were then added to the medium. Finally, the medium was augmented with a minimal set of metabolites to allow community growth, using a built-in function in MICOM. As mentioned previously, cultures in each study were diluted to a different ratio. To account for this, residual fiber included in the medium (resistant starch, dextrin and cellulose) was scaled to match the study conditions. For instance, Study A was diluted 1:5, so the residual fiber in the medium used to simulate growth for these samples was diluted by a factor of 5. Growth of models was simulated using a cooperative tradeoff parameter of 0.7. Finally, model predicted SCFA production fluxes were compared with calculated SCFA production rates from the *ex vivo* experiments, calculated from the endpoint concentrations less the starting concentration divided by the culturing time. For these samples, turbidity of the culture prevented

optical density from being measured, so scaling measurements to account for differences in biomass was not possible. Nonetheless, predicted flux in units of mmol/gDW/h could be compared with measured production rates in units of mmol/L/h. To compare across datasets, measured and predicted SCFA production was z-scored using Scipy. Again, the relationship between measured and predicted rates was assessed using Pearson correlation in Scipy.

2.3 Results

2.3.1 Predicting SCFA production from *in vitro* cultures

To initially validate predictions of SCFA production from microbial community-scale metabolic models, existing experimental data from synthetically constructed *in vitro* cultures were leveraged ⁷⁴. This dataset included measurements of relative microbial abundances, SCFA production levels, and the overall optical density for each of 1,579 independent co-cultures. Each co-culture contained between 1 and 25 strains of common gut microbes. Results showed that all models were predicted to grow at a community growth rate of 0.3/h. Comparisons were stratified between low richness (1-5 taxa) and high richness communities (10-25 taxa). In both low and high richness communities, predicted butyrate production was significantly associated with butyrate production fluxes measured *in vitro* (**Fig. 2.1**; Pearson correlation; Low Richness: $r = 0.15$, $p = .0014$; High Richness: $r = 0.53$, $p = 8.7e-52$). Notably, predictions for butyrate production were far more accurate in higher richness communities (**Fig. 2.1A**), while low richness communities showed more over or underestimation (**Fig. 2.1B**). Interestingly, low richness communities showed several examples where little to no butyrate production was predicted, an outcome not observed in higher richness communities. This indicates a dependence on community cross feeding for butyrate production.

2.3.2 Predicting SCFA production from *in vitro* cultures treated with dietary fiber

Furthering this analysis, growth of synthetic *in vitro* communities with dietary fiber addition was simulated. Similar to the previous analysis, communities consisted of pairwise combinations of common microbial species, grown on a defined medium. In this scenario, the dietary fibers inulin or laminarin were added to cultures. This was represented in the *in silico* media applied to the growth simulation. After model construction and growth simulation, predictions of butyrate production were once again compared with measured butyrate production rates. As previously, strong association was observed between measured and predicted rates (**Fig. 2.2**; Pearson correlation; inulin-treated: $r = 0.24$, $p = 1.7e-11$; laminarin-treated: $r = 0.64$, $p = 4.1e-55$). It is important to note that these cultures were not stratified for richness, and spanned from 1-25 taxa per sample. Predictions for laminarin-treated samples (**Fig. 2.2B**) were slightly stronger than those for inulin-treated samples (**Fig. 2.2A**), but both showed a correlation coefficient similar to that of the communities in the previous analysis. In each case, butyrate production was noticeably higher than in the samples with no dietary fiber addition.

2.3.3 Predicting SCFA production from longitudinally sampled *in vitro* cultures

Additional validation was conducted in another *in vitro* dataset, wherein 95 synthetic communities were cultured for 48 hours and sampled thrice throughout the growth period. Again, a strong significant correlation between measured and predicted values for butyrate production was observed (**Fig. 2.3**; Pearson correlation; 16h timepoint: $r = 0.49$, $p = 8.6e-7$; 32h timepoint: $r = 0.43$, $p = 2.2e-5$; 48h timepoint: $r = 0.47$, $p = 1.9e-6$). As measured production was

scaled by culturing time, measured production rates for butyrate did not notably increase. Similarly, predictions for butyrate production remained similar across timepoints, reflected by similar results from correlation analysis. Ultimately, no single sampling timepoint showed significantly improved results over another with respect to predictions for butyrate production.

2.3.4 Predicting SCFA production from fecally-derived *ex vivo* cultures

For further validation of SCFA predictions from community-scale metabolic models, we turned to fecal homogenates cultured *ex vivo*, as a higher-complexity system. Four independent *ex vivo* datasets with nearly identical experimental set up were used, totaling 30 samples. To compare results from across studies, predicted and measured SCFA production fluxes were Z-scored. Among these results, significant agreement was observed between measured and predicted values for butyrate and propionate production (**Fig. 2.4A&B**, Butyrate: Pearson $r = 0.62$, p -value $2.1e-10$; Propionate: Pearson $r = 0.63$, p -value $1.6e-10$). Interestingly, acetate was not well predicted (**Fig. 2.4C**, Pearson $r = 0.14$, p -value $= 0.18$). Significant agreement was observed between measured and predicted production fluxes of butyrate and propionate within each individual data set (**Fig. 2.4D-O**, $r = 0.40$ - 0.97 , Pearson test, $p < 0.05$) apart from samples in Study A, which had a very limited sample size ($N = 2$), and propionate in Study C, which was just above the threshold for significance. Notably, the correlation coefficient (Pearson r) for these associations was similar to that seen in the high-richness *in vitro* cultures. As mentioned, the prediction of acetate was worse, most notably in studies A, C and D, where no significant prediction was observed. In study B, acetate production was more readily predicted, likely due to a strong treatment-effect. Within treatment groups, similar correlations were observed, though statistical power was severely limited by the smaller sample sizes (**Table 2.2**). For Study C, both 16S amplicon and metagenomic shotgun sequencing was available. Models for this dataset were

constructed from both kinds of sequencing data at the species and genus levels and showed highly consistent results (**Fig. 2.5**). Predictions from models built with shotgun metagenomic sequencing data from Studies A, C and D showed slightly higher concordance with measured values when constructed at the species level, as compared to building at the genus level (**Fig. 2.6**). To test whether SCFA production was impacted by sample diversity, we compared measured butyrate and propionate against Shannon index for each sample in each study (**Fig. 2.7**). A weak significant signal was seen in only one of the four studies (Study D). Additionally, bacterial biomass of the initial was shown to not make a significant impact on measured SCFAs, determined using the relative abundance of metagenomic reads mapping to the human genome (**Fig. 2.8**). In summary, we observed agreement between MCMM predicted and measured *in vitro* production rates of butyrate and propionate in the presence or absence of fiber supplementation, with better agreement in more diverse communities.

2.4 Discussion

Utilizing both *in vitro* and *ex vivo* communities facilitated a thorough evaluation and validation of predictions for SCFA production from the gut microbiome based on MCMMs. This approach revealed both advantages and constraints of the modeling scheme. A primary finding was the models' capacity to consistently predict butyrate and propionate production across different individuals, a conclusion corroborated by multiple independent datasets.

An interesting result from *in vitro* modeling showed that predictions based on higher richness communities showed stronger concordance with measured butyrate production rates. We hypothesize that this is due to a summary effect, ultimately benefiting the construction of models of the human gut microbiome. Genus-level reconstructions used to construct the models

in this portion of the analysis contain the summarized metabolic capacity of all species within a given genus. As a result, the functional capacity of low-diversity models may exceed that of the community *in vitro*, resulting in incorrect prediction of butyrate production. In high-richness communities, this effect is of lesser consequence, as it is summarized across more genera. Overall, the observed increase in accuracy with community diversity benefits modeling of real-world microbiomes, which are generally more species-rich than synthetic *in vitro* communities. Additionally, increasing prevalence of shotgun metagenomic sequencing for analyzing the gut microbiome results in increased granularity of taxonomic assignment, meaning construction of models at the species level is increasingly possible, minimizing the summary effect seen in genus-level models. Improvements in the development of GEMs will also serve to minimize the incongruencies between metabolic pathways included in models and those actually found in these taxa.

Modeling the addition of dietary fiber to the *in vitro* cultures served as additional validation of the modeling framework and underscored an important consideration. As one would expect, adding more dietary fiber to the medium used to simulate growth did lead to increased SCFA production when compared to the first set of *in vitro* cultures, as there was more fermentative substrate available to these communities. Cultures treated with either dietary fiber tested showed significant agreement between measured and predicted rates of butyrate production. Since these samples were not stratified between high and low richness, low-richness communities included in this analysis contribute to a slightly diminished accuracy of predictions. Interestingly, laminarin-treated samples showed much stronger agreement than those treated with inulin, a counter-intuitive result given that laminarin is a less commonly consumed dietary fiber. However, examining the individual compartments comprising the community-scale models

reveals that fifteen different taxa are capable of importing and fermenting inulin, while only two taxa are capable of importing and fermenting laminarin. Due to this effect, there is a smaller set of solutions for samples treated with laminarin which maximize the objective of community growth, wherein a maximal amount of the available substrate is consumed. This, in turn, leads to increased certainty regarding the final flux solution and total butyrate production. In future analyses, this can be considered when designing dietary interventions, since the limited solution space may lead to better individual-specific predictions for the effectiveness of dietary interventions.

In an effort to add additional validation for MCMM-based predictions, a third *in vitro* dataset was utilized. Here, predicted rates for butyrate production not only agreed with measured butyrate production rates, but did so across multiple timepoints. Essentially, calculated production rates for butyrate were accurately captured by the metabolic models of each sample. This result, beyond validating even further the accuracy and reliability of these models, shows that the results are robust throughout the growth of microbial cultures. Despite the assumptions of the state steady condition, predictions were still accurate after 48 hours of culturing.

Ex vivo cultures emerged as the optimal environment for modeling SCFA production from the gut microbiome. While individual datasets had limited sample sizes, the combined results yielded promising outcomes for both butyrate and propionate. However, the prediction accuracy for acetate was notably lower in this analysis. This discrepancy is likely attributable to acetate's function as an overflow metabolite, where a substantial portion of available carbon substrate is metabolized into acetate and expelled into the environment ⁷⁵. Since this process isn't directly linked to growth, there exists a wide spectrum of potential production fluxes for acetate relative to a given biomass optimum. Consequently, the FBA optimization method may not

consistently provide accurate predictions for acetate production. On the other hand, butyrate and propionate exhibit a narrower range of production fluxes, suggesting a closer association with biomass production. As such, these metabolites were well predicted across the four datasets utilized in this analysis. MCMMs showed strong predictive capability whether constructed from 16S amplicon or shotgun metagenomic sequencing data, although species level models constructed from shotgun metagenomic data do seem to provide the highest level of reliability. Importantly, we also determined that measured differences in butyrate or propionate production were not due to extrinsic factors, including community diversity or bacterial biomass. This underscores the necessity of a deeper mechanistic understanding of these communities, as models based on these measures alone would perform worse than MCMMs. These findings underscore the reliability of utilizing MCMMs for predicting the production of these metabolites.

One important consideration brought to light by these validation schemata is the importance of accurately representing the metabolic availability of the environment in which communities are cultured. This turned out to be the most laborious task for these analyses, specifically for the *ex vivo* data, wherein no defined medium was used and the fecal microenvironment had to be estimated. The dilution ratio of *ex vivo* cultures played an important role in determining the amount of substrate available for bacterial fermentation. Samples from studies with higher dilution showed less SCFA production on average than in studies with low dilution, and this could be captured in the metabolic models by scaling the available dietary fiber in the *in silico* media to match this dilution. When conducting similar analyses with *in vivo* data, it is vital to try to match the correct diet to each individual, as inaccuracies in metabolic availability can potentially lead to large deviations between model predictions and true values.

Leveraging these datasets from *in vitro* and *ex vivo* data for the comprehensive evaluation of SCFA production has provided valuable insights into the predictive capacity and limitations of this modeling framework. The consistency in predicting butyrate and propionate production across individuals underscores the robustness of MCMMs as a modeling platform. Notably, higher richness in *in vitro* communities enhanced the accuracy of butyrate predictions, suggesting a pivotal role of community diversity in modeling real-world microbiomes. The validation across multiple timepoints and environments reaffirms the reliability of MCMMs in capturing microbial processes. Furthermore, the differential response to dietary fiber treatments elucidates the intricate interplay between microbial community composition and substrate utilization, offering avenues for personalized dietary interventions. However, challenges persist, particularly in accurately representing the metabolic environment, and emphasizing the importance of refining model parameters such as metabolic availability through medium application to match experimental conditions. Overall, these findings advance our understanding of gut microbiome metabolism and highlight the potential of MCMMs in guiding future studies and interventions targeting SCFA production and gut health.

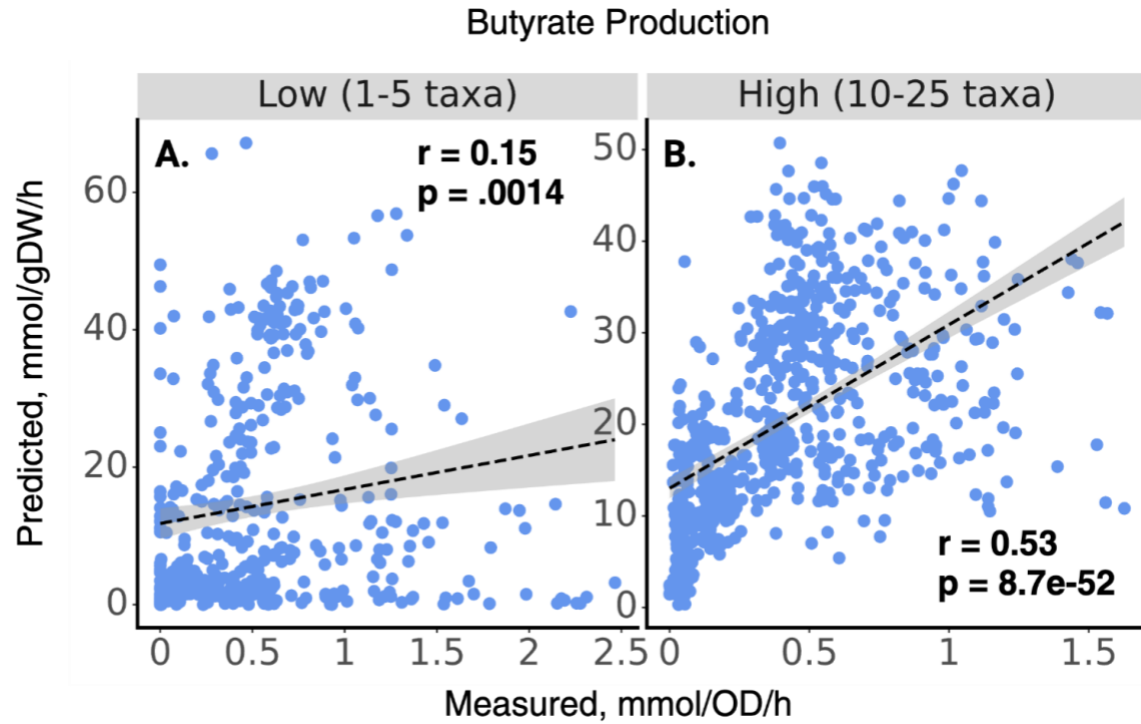


Figure 2.1. Relationship between predicted and measured butyrate production rates in *in vitro* co-cultures. MCOMM-based predictions of butyrate production in high- and low-richness *in vitro* communities are shown on the y-axes and OD-normalized measured production rates are shown on the x-axes, along with Pearson r and p -values from a Pearson correlation. The dashed line represents a linear regression of the data, and the shaded gray region a 95% confidence interval. **(A)** Predicted and measured butyrate fluxes in models of low richness (1-5 taxa) synthetic communities (Pearson $r = 0.15$, $p = .0014$, $N = 882$). **(B)** Predicted and measured butyrate fluxes in models of high richness (10-25 taxa) synthetic communities (Pearson $r = 0.53$, $p = 8.7e-52$, $N = 697$).

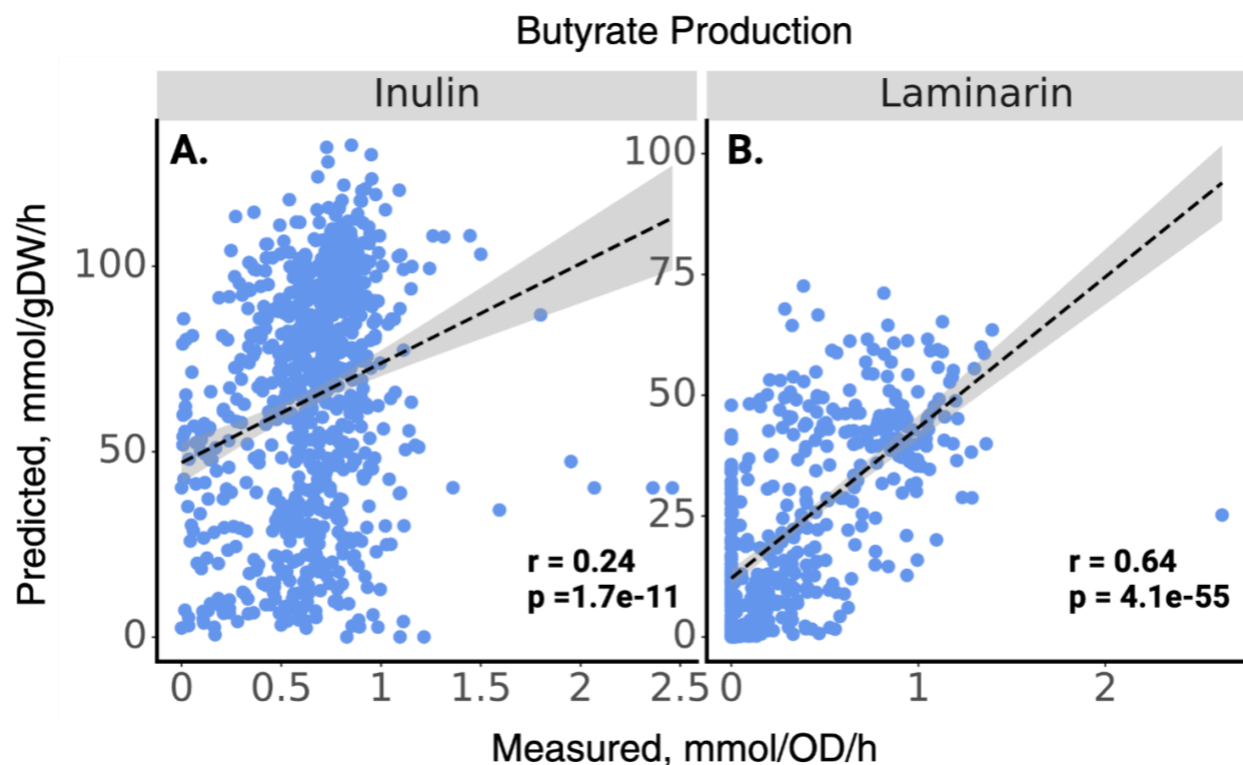


Figure 2.2 Relationship between predicted and measured butyrate production rates in fiber-treated *in vitro* co-cultures. MCMC-based predictions of butyrate production from fiber-treated *in vitro* cultures are shown on the y-axes and OD-normalized measured production rates are shown on the x-axes, along with Pearson r and p -values from a Pearson correlation. The dashed line represents a linear regression of the data, and the shaded gray region a 95% confidence interval. **(A)** Predicted and measured butyrate fluxes in models of inulin-treated synthetic communities (Pearson $r = 0.24$, $p = 1.7e-11$, $N = 774$). **(B)** Predicted and measured butyrate fluxes in models of laminarin-treated synthetic communities (Pearson $r = 0.64$, $p = 4.1e-55$, $N = 475$).

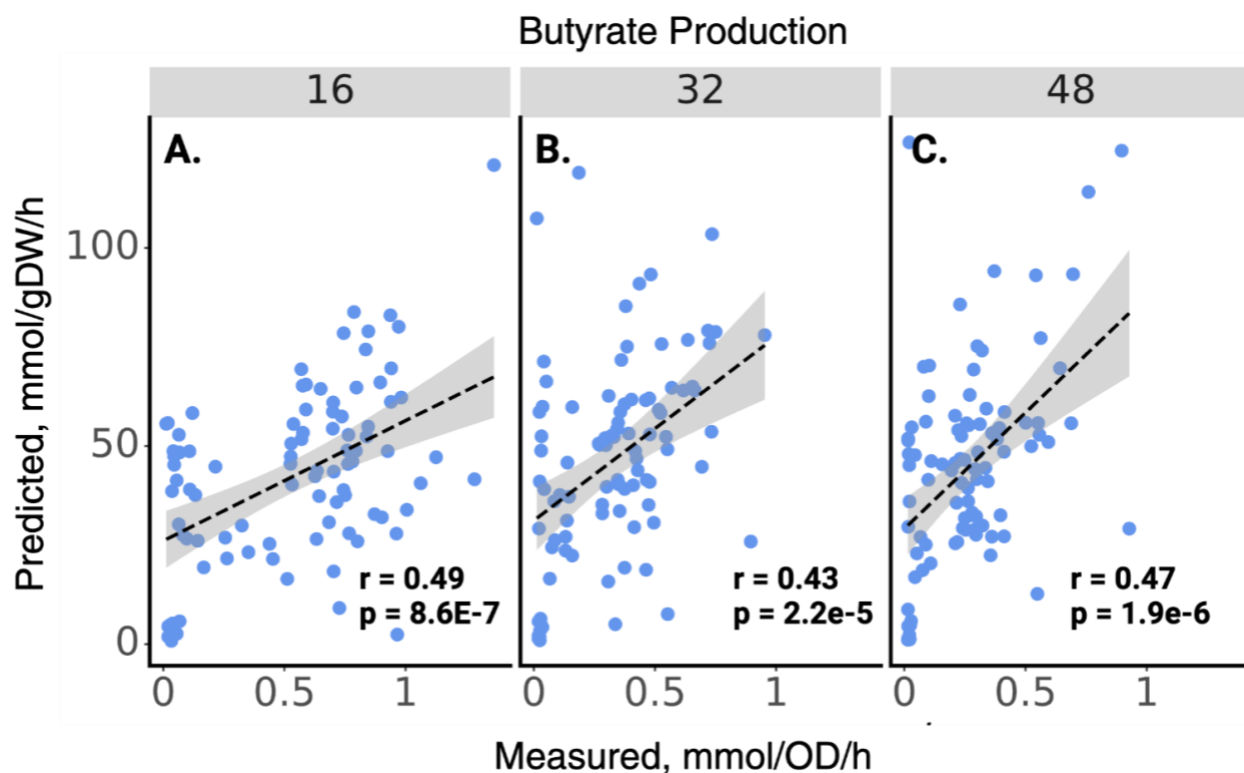


Figure 2.3 Relationship between predicted and measured butyrate production rates in longitudinally-sampled in vitro co-cultures. MCMC-based predictions of butyrate production from longitudinally-sampled in vitro cultures are shown on the y-axes and OD-normalized measured production rates are shown on the x-axes, along with Pearson r and p -values from a Pearson correlation. The dashed line represents a linear regression of the data, and the shaded gray region a 95% confidence interval. **(A)** Predicted and measured butyrate fluxes in models of synthetic communities sampled at 16 hours into culturing ($N = 95$). **(B)** Predicted and measured butyrate fluxes in models of synthetic communities sampled at 32 hours into culturing ($N = 95$). **(C)** Predicted and measured butyrate fluxes in models of synthetic communities sampled at 48 hours into culturing ($N = 95$)

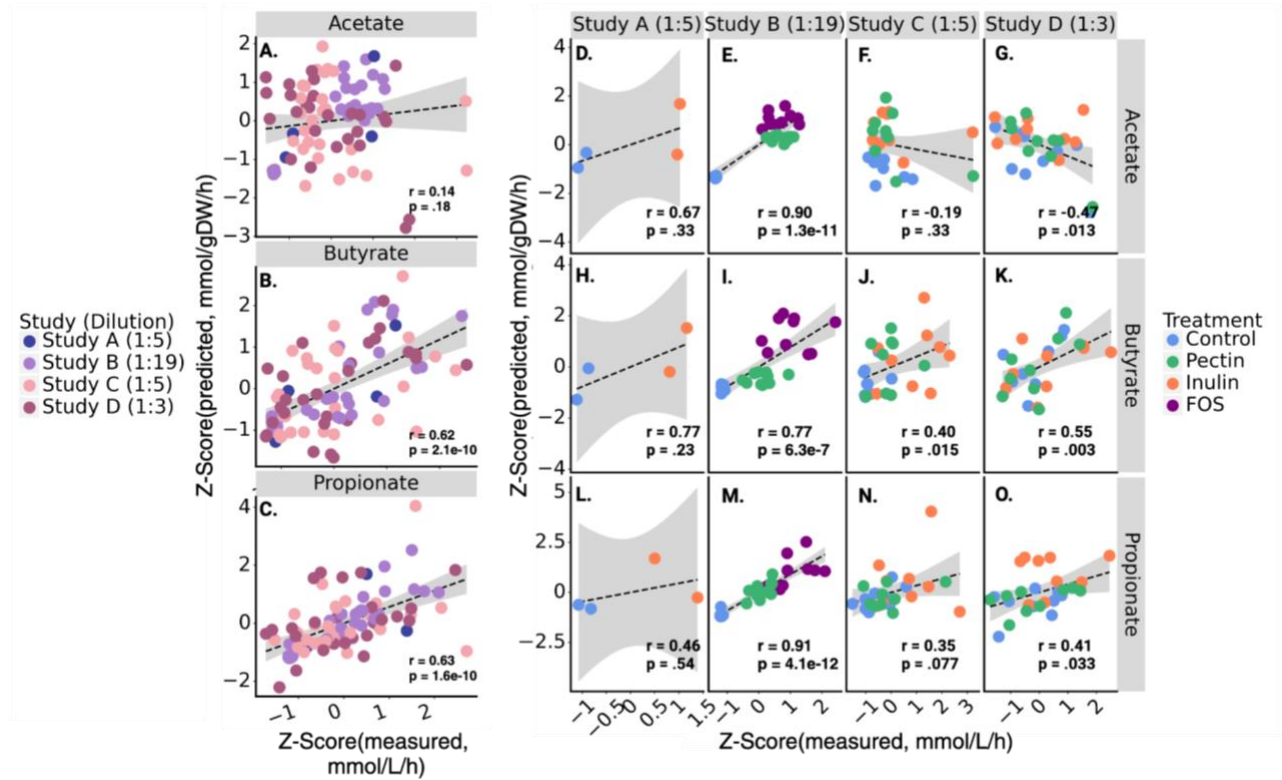


Figure 2.4 Stool ex vivo assays show strong associations between measured and predicted SCFA production fluxes. Z-scored SCFA production flux predictions from MCMs are shown on the y-axes and Z-scored measured values are shown on the x-axes. Pearson r and associated p-value are calculated for all points in each plot. The dashed line denotes a linear regression line, and the gray area denotes the 95% confidence interval of the regression. (A-C) Z-scored predictions from all four independent datasets are shown together for acetate, butyrate, and propionate (Acetate: Pearson r = 0.14, p-value = 0.18; Butyrate: Pearson r = 0.61, p-value 2.1e-10; Propionate: Pearson r = 0.63, p-value 1.6e-10). Residual fiber in the media used to simulate growth of each study was scaled according to the dilution factor, shown next to the study name in the legend. Z-scored predictions versus z-scored measurements within each study are shown for acetate (D-G), butyrate (H-K), and propionate (L-O). Specific dilution of each study is shown in the column name.

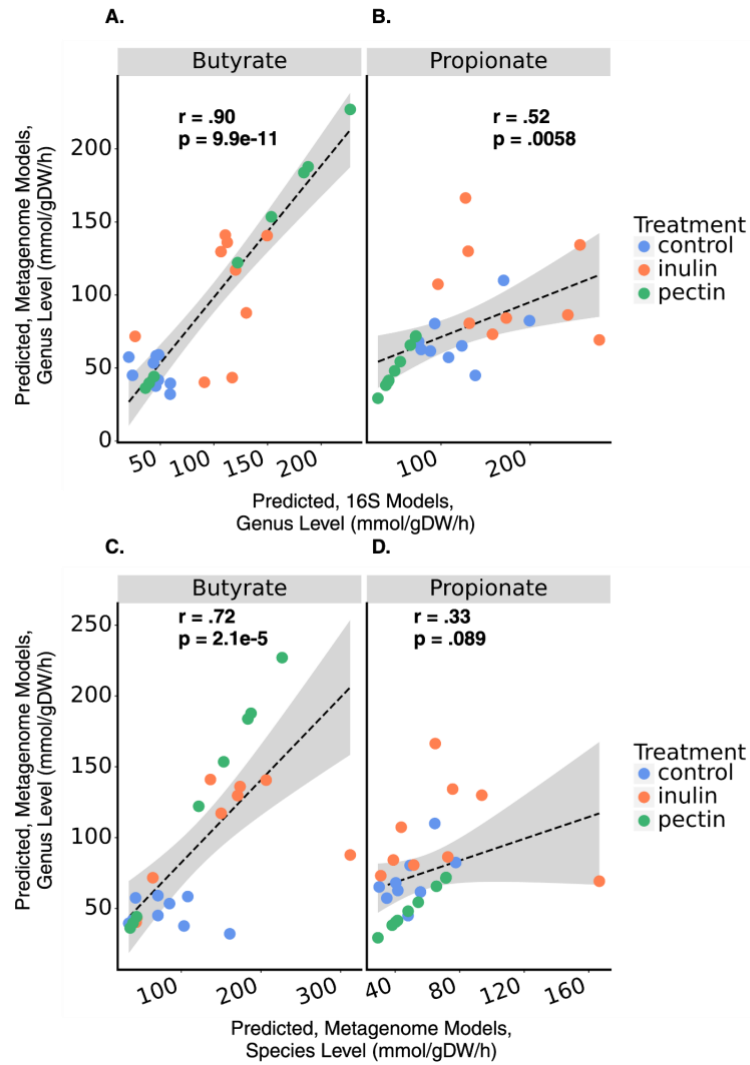


Figure 2.5 16S amplicon vs shotgun metagenomics for MCMM construction. Data from Study C included 16S amplicon sequencing as well as shotgun metagenomic sequencing. Models were built using both types of data, as well as across taxonomic ranks for metagenomic data (**A-B**) Predictions for butyrate and propionate between models summarized to the genus level from 16S amplicon sequencing data and shotgun metagenome data (butyrate: Pearson Correlation $r = 0.90$, $p = 9.9e-11$; propionate: Pearson Correlation $r = 0.52$, $p = .0058$). (**C-D**) Predictions for butyrate and propionate from models built using shotgun metagenome data at the genus level and species level (butyrate: Pearson Correlation $r = 0.72$, $p = 2.1e-5$; propionate: Pearson Correlation $r = 0.33$, $p = .089$). Pearson correlation r value and associated two-tailed p -value were calculated across all points in each subplot. The dashed black line denotes a linear regression line and the gray area denotes the 95% confidence interval of the regression line. Color encoding indicates the specific fiber treatment for each point.

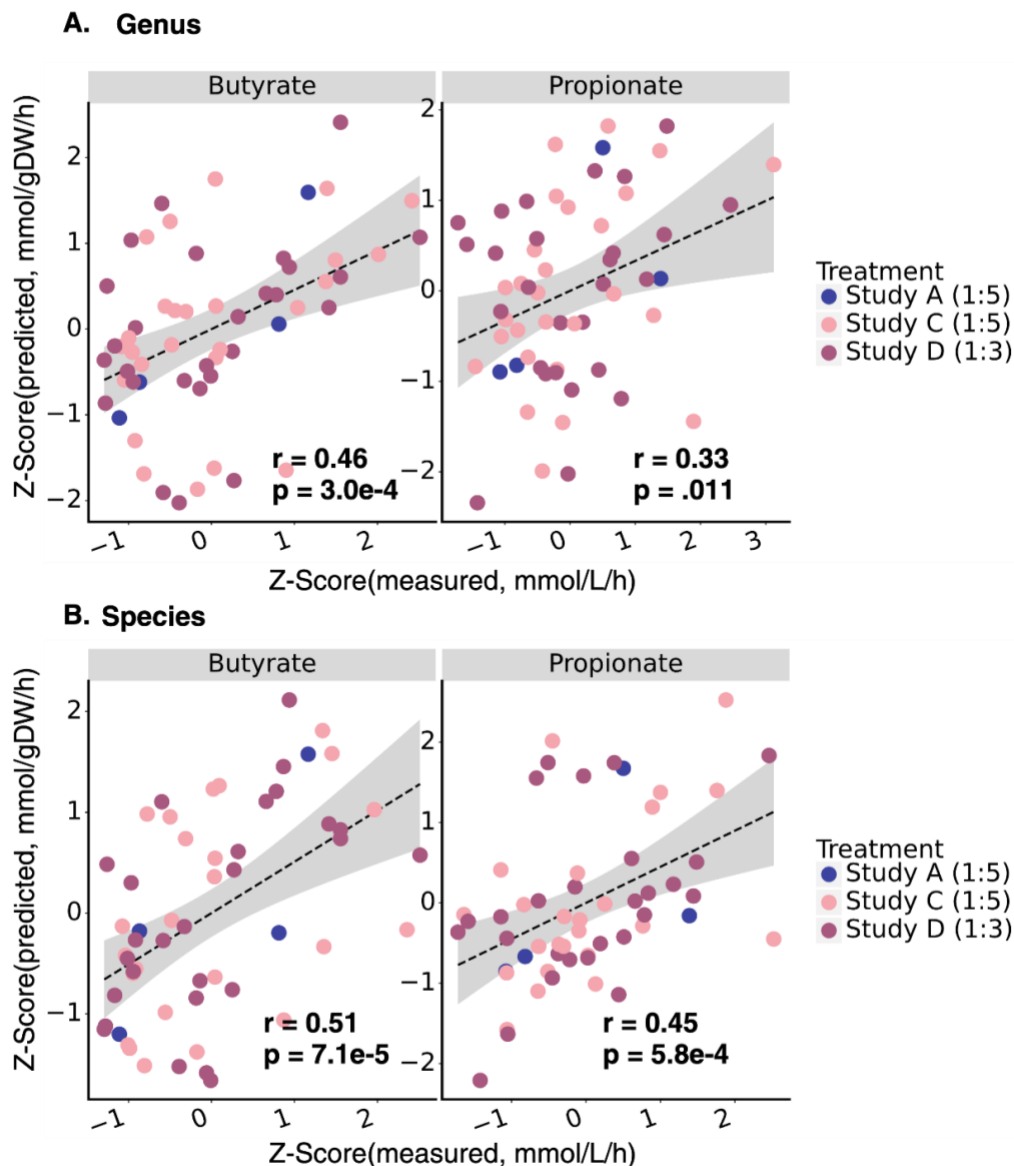


Figure 2.6 MCMM prediction accuracy in genus- vs. species-level models. MCMMs from ex vivo studies A, C and D were constructed at the **(A)** genus (butyrate: Pearson Correlation $r = 0.46$, $p = 3.0 \times 10^{-4}$; propionate: Pearson Correlation $r = 0.33$, $p = .011$) and **(B)** species level (butyrate: Pearson Correlation $r = 0.51$, $p = 7.1 \times 10^{-5}$; propionate: Pearson Correlation $r = 0.45$, $p = 5.8 \times 10^{-4}$). Prediction production rate of butyrate and propionate more closely matched measured production rate in the species level model as compared to the genus level model. Pearson correlation r value and associated two-tailed p -value were calculated across all points in each subplot. The dashed black line denotes a linear regression line and the gray area denotes the 95% confidence interval of the regression line. Color encoding indicates the specific study from which each point originates.

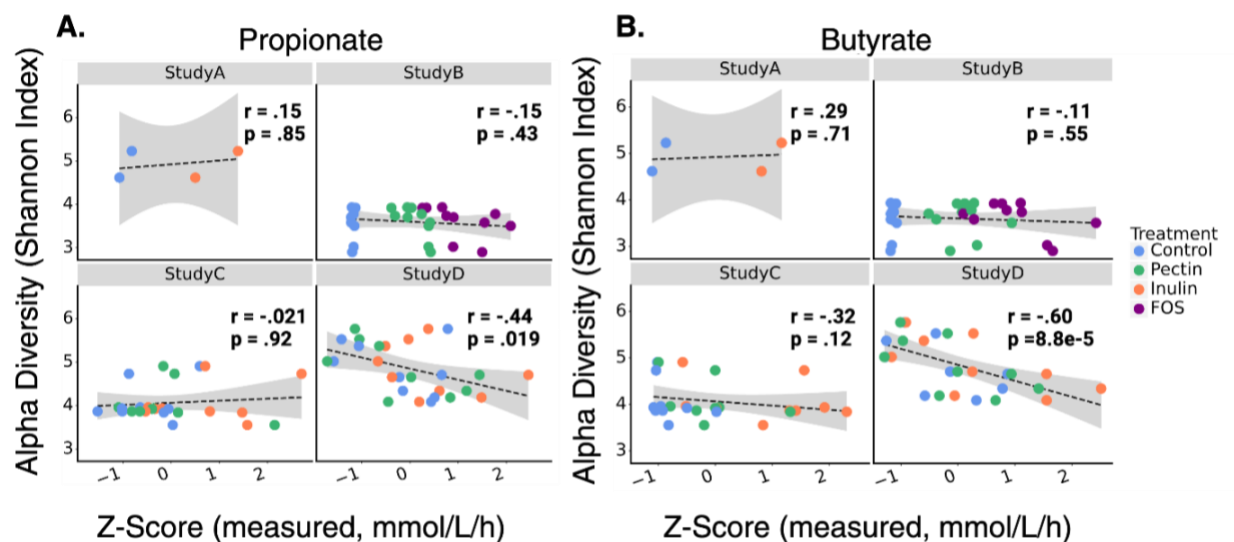


Figure 2.7 Alpha diversity in ex vivo-derived MCMs. We compared Shannon index, a measure of alpha diversity, against SCFA production in ex vivo communities **(A)** Propionate production in four ex vivo datasets was not consistently explained by alpha diversity (Study A: Pearson Correlation $r = 0.15$, $p = .85$; Study B, Pearson Correlation $r = -0.15$, $p = .43$; Study C, Pearson Correlation $r = -0.021$, $p = .92$; Study D, Pearson Correlation $r = -0.44$, $p = .019$). **(B)** Butyrate production also showed no consistent correlation with alpha diversity, although a significant difference was again observed within Study D (Study A: Pearson Correlation $r = 0.29$, $p = .71$; Study B, Pearson Correlation $r = -0.11$, $p = .55$; Study C, Pearson Correlation $r = -0.32$, $p = .12$; Study D, Pearson Correlation $r = -0.60$, $p = 8.8e-5$). Pearson correlation r value and associated two-tailed p -value were calculated across all points in each subplot. The dashed black line denotes a linear regression line, and the gray area denotes the 95% confidence interval of the regression line. Color encoding indicates the specific fiber treatment for each point.

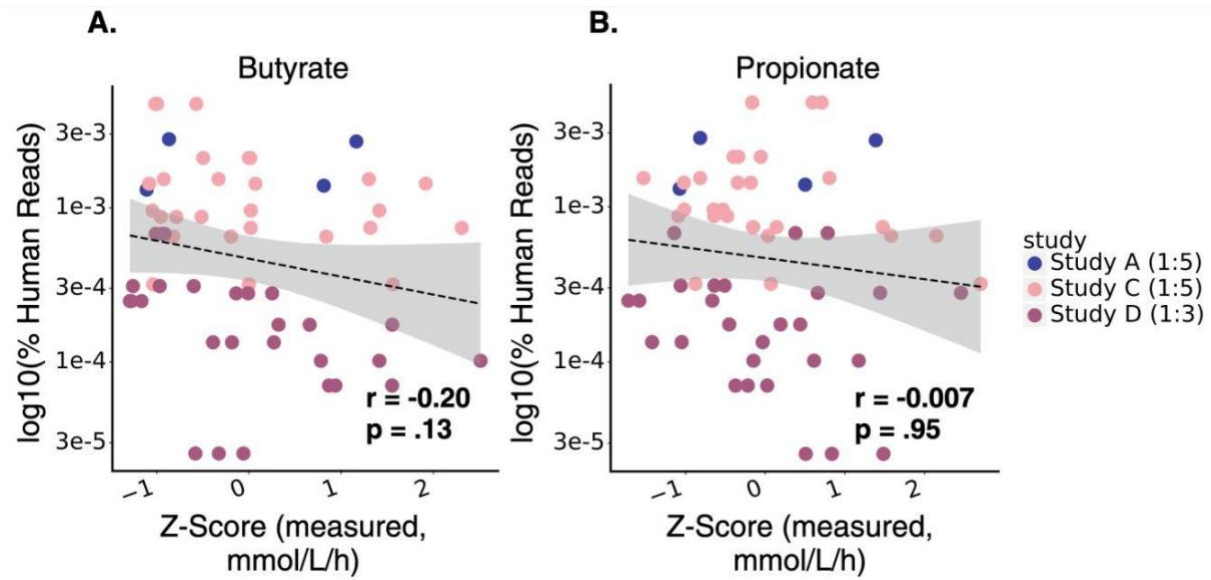


Figure 2.8 Bacterial biomass on measured SCFA production. Bacterial biomass as estimated by shotgun metagenomic reads showed no association with measured **(A)** butyrate or **(B)** propionate production from ex vivo cultures (butyrate: Pearson Correlation $r = -0.20$, $p = .13$, propionate: Pearson Correlation $r = -0.007$, $p = .95$). Pearson correlation r value and associated two-tailed p -value were calculated across all points. The dashed black line denotes a linear regression line and the gray area denotes the 95% confidence interval of the regression line. Color encoding indicates the specific study from which each point originates.

<i>Study</i>	<i>Source</i>	<i>Publication</i>	<i>Sample Size</i>	<i>Total taxa per Sample</i>	<i>Total taxa in model per sample</i>
In Vitro (Low Richness)	Venturelli Lab, University of Wisconsin	Clark et al. (2021) ⁷⁴	882	2.93 ± 0.07	2.93 ± 0.07
In vitro (High Richness)	Venturelli Lab, University of Wisconsin	Clark et al. (2021) ⁷⁴	697	13.08 ± 0.16	13.08 ± 0.16
In vitro (Inulin-Treated)	Venturelli Lab, University of Wisconsin	Unpublished	774	7.65 ± 0.08	7.65 ± 0.08
In vitro (Laminarin-Treated)	Venturelli Lab, University of Wisconsin	Unpublished	475	6.04 ± 0.09	6.04 ± 0.09
In vitro (longitudinally sampled)	Venturelli Lab, University of Wisconsin	Unpublished	95	11.32 ± 0.11	11.32 ± 0.11
Ex Vivo Study A	Gibbons Lab, Institute for Systems Biology	current study (SRA accession: PRJNA1033794)	2	88.7 ± 1.9	51.5 ± 1.0
Ex Vivo Study B	Hamaker Lab, Purdue University	Cantu Jungles et al. (2021) ⁷⁶	10	20.7 ± 0.4	16.4 ± 0.3
Ex Vivo Study C	Gurry Lab, University of Geneva	current study (SRA accession: PRJNA939256)	9	68.4 ± 3.9	28.0 ± 1.8
Ex Vivo Study D	Gibbons Lab, Institute for Systems Biology	current study (SRA accession: PRJNA1033794)	9	58.7 ± 2.4	44.6 ± 1.8

Table 2.1 List of *in vitro* and *ex vivo* datasets leveraged for MCMM validation

	Acetate			Butyrate			Propionate		
Study A	Control	Inulin		Control	Inulin		Control	Inulin	
r	n/a	n/a		n/a	n/a		n/a	n/a	
p	n/a	n/a		n/a	n/a		n/a	n/a	
Study B	Control	Pectin	FOS	Control	Pectin	FOS	Control	Pectin	FOS
r	-0.062	-0.077	0.16	-0.028	0.24	0.060	-0.94	0.64	0.52
p	0.87	0.83	0.66	0.94	0.51	0.87	0.80	0.046	0.12
Study C	Control	Inulin	Pectin	Control	Inulin	Pectin	Control	Inulin	Pectin
r	-0.54	-0.11	-0.54	0.260	0.42	0.19	0.62	0.52	0.48
p	0.17	0.800	0.16	0.53	0.30	0.65	0.100	0.19	0.23
Study D	Control	Inulin	Pectin	Control	Inulin	Pectin	Control	Inulin	Pectin
r	-0.57	-0.08	-0.850	0.67	0.47	0.65	0.39	0.14	0.52
p	0.110	0.84	0.00	0.0510	0.200	0.059	0.29	0.72	0.15

Table 2.2 Within treatment group correlations from MCMs for *ex vivo* cultures

Chapter 3

Analyzing the Clinical Relevance of MCMM-Based Predictions with *in vivo* Data

Portions of this chapter were published as a manuscript in Nature Microbiology ⁷⁰.

3.1 Introduction

Following the successful validation of MCMM-based predictions for butyrate and propionate production, a natural progression was to extend the analysis to include human observational and interventional data. Building upon the demonstrated reliability and accuracy of the modeling approach, this transition allows for a deeper exploration of the complexities inherent in the gut microbiome environment. By venturing into *in vivo* studies, we aim to further elucidate the interplay between microbial metabolism and host physiology, thus providing valuable insights into the intricate mechanisms in which microbially-derived SCFAs are involved.

As alluded to previously, direct measurement of SCFA production *in vivo* is prohibitively challenging and costly, making direct validation of predictions challenging. Enterocytes in the colonic epithelium depend on microbially-produced SCFAs as a primary energy source, and quickly consume these metabolites to enter the mitochondrial beta-oxidation pathway ¹⁷. However, the physiological effects of these metabolites on the host can be observed and assayed through clinical labs and measurements, specifically regarding systemic inflammation and metabolic health. Here, we detail analyses focused on correlating MCMM-predicted SCFA production from the gut microbiome to measured outcomes in the host using *in vivo* data.

In the first analysis, we investigated the effect of a 10-week high fiber dietary intervention on the SCFA production capacity of the microbiome. In this study, 18 participants were fed a diet rich in resistant starch for 10 weeks ⁷⁷. These individuals fell into three distinct immunological response groups, as assessed by the authors of the study: 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) ⁷⁷. Notably, the original paper did not offer a definitive mechanism to account for the variability in immune response across these groups. We hypothesized that these immune response groups could be explained, in part, by differences in MCMM-predicted production rates of anti-inflammatory SCFAs. The longitudinal nature of these data made it straightforward to assess differences in predicted SCFA production across the duration of the intervention.

Next, we shifted our focus to a considerably larger cohort comprising generally healthy individuals with comprehensive phenotypic data, sourced from a former precision wellness company based in Seattle, WA (Arivale). Leveraging blood-based clinical laboratory data

obtained from each participant, we aimed to reevaluate the association between predicted butyrate and propionate production from the gut microbiome and the overall health status of the host. The findings gleaned from modeling *in vivo* communities instilled greater confidence in the reliability of these models, and furthermore, in the clinical significance of the predictions they yield. Beyond merely predicting production levels, understanding the downstream effects underscores the potential of MCMMs as potent clinical tools for microbiome-directed therapies.

A list of all datasets included in this analysis is available at the end of the chapter (**Table 3.1**).

3.2 Methodology

*3.2.1 MCMM construction for *in vivo* data*

Fecally-derived samples from study participants were analyzed using 16S rRNA amplicon sequencing to determine taxonomic composition of the microbiome. As previously, taxonomic abundance data inferred from 16S rRNA amplicon sequencing were mapped to existing GEMs from the AGORA database to construct MCMMs in MICOM with an abundance threshold of 0.001. As only 16S amplicon sequencing data was available, models were constructed at the genus level, which was the lowest taxonomic rank for which the data was reliable. Growth of the models was then simulated after applying an *in silico* medium, using a cooperative tradeoff parameter of 0.7. Two media for these analyses: a high-fiber diet containing high levels of resistant starch for the longitudinal fiber study, and a standard European diet for the Arivale cohort^{78,79}. Both diets were collected from the Virtual Metabolic Human database. 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 medium was augmented with a minimal set of metabolites to allow community growth.

3.2.2 Association of MCMM predictions with immune response groups

MCMMs were constructed for all 18 participants in a longitudinal high fiber study at each of 7 timepoints using timepoint-specific data using the method described above. Model growth was simulated using a high-fiber medium rich in resistant starch representative of the diet used in the intervention. Predicted production of butyrate and propionate was calculated for each sample at each time point, and subsequently compared between immune response groups. Analysis between groups was performed using a non-parametric Mann-Whitney U test, and between timepoints within a group using Pearson correlation. All statistical tests were performed using Scipy.

3.2.3 Association of MCMM predictions with blood-based clinical measures

MCMMs were constructed for 3,129 samples from the Arivale cohort who had paired fecal samples and blood draws within 30 days of each other, using the method described above. Model growth was simulated using an average European diet, in the absence of in-depth knowledge of the dietary intake of participants. Predicted butyrate and propionate rates were calculated for each participant, and regressed against a panel of 128 blood-based clinical labs as well as blood pressures using a set of common covariates including sex and age using statsmodels (v0.14.0) ⁸⁰. Metabolomic measures of butyrate in the blood were also regressed against clinical labs and blood pressure. All measures were z-scored prior to regression to allow for direct comparisons between results.

3.3 Results

3.3.1 Predicted SCFA production in high-fiber dietary intervention

Across the duration of the high-fiber dietary intervention, a trend of decreasing propionate production was predicted in high-inflammation individuals ($r = -0.39$, Pearson test, $p = 0.019$), showing less production as the intervention went on, despite the high fiber content of the diets consumed by participants (**Fig. 3.1A**). Individuals in the high-inflammation group showed significantly lower predicted propionate production, on average, compared to the individuals in each of the low-inflammation groups (High vs. Low I: 131.9 ± 5.8 vs 158.1 ± 5.7 mmol/(gDW h), Mann-Whitney $p = 0.0053$; High vs. Low II: 131.9 ± 5.8 vs $163.08.3 \pm 6.5$ mmol/(gDW h), Mann-Whitney $p = 0.0017$; **Fig. 3.1B**). Butyrate showed no significant differences across immune response groups (**Fig. 3.1C, 3.1D**). To investigate whether sample alpha-diversity was sufficient to explain the differences between the immune response groups, we calculated the alpha diversity for each sample at each timepoint during the study. Across all seven time points tested, only one significant difference in alpha diversity was seen, between the two low inflammation groups at time point 2 (Mann-Whitney U-test, $p < 0.05$), leading us to determine that differences in SCFA production throughout the intervention were not the result of differences in diversity (**Fig. 3.2**).

3.3.2 Predicted SCFA associations with blood-based clinical measures

MCMM-predicted SCFA fluxes for participants in the Arivale cohort were regressed against a panel of 128 clinical chemistries and health metrics collected from each individual, adjusting for a standard set of covariates (i.e., age and sex), as well as blood pressure. Butyrate

levels in blood plasma, as measured by untargeted metabolomics, were also regressed against clinical markers. After FDR correction, 20 markers were significantly associated with the predicted rate of butyrate production (**Fig. 3.3A**). Predicted butyrate production showed significant positive associations with 3 markers, including the health-associated hormone adiponectin, and significant negative associations with 17 markers linked to disease, including C-reactive protein (CRP), low-density lipoprotein (LDL), and mean arterial blood pressure ($p < 0.05$, FDR-corrected t-test). Measured butyrate levels in the bloodstream, however, showed no significant associations with any of the clinical chemistries, indicating that blood-based butyrate measures are poor proxies for *in situ* production (**Fig. 3.3A**). There was, however, a weak, but significant, positive association between measured butyrate levels in blood and MCMM-predicted butyrate production (**Fig. 3.4**). Predicted propionate production rates showed no significant associations with clinical chemistries after covariate adjustment and multiple comparison correction (**Fig. 3.3A**). Predicted butyrate production was significantly negatively associated with BMI (multiple regression, $\beta = -0.10$, t-test, $p < 0.001$), while measured blood butyrate and predicted propionate production were not (**Fig. 3.3C-D**). Results of the regression for significantly associated blood-based measures are found in **Table 3.2**.

3.4 Discussion

Findings from MCMM-based modeling of a longitudinal high-fiber dietary intervention show that production of propionate, a strong inhibitor of inflammation through activation of FFAR2 and FFAR3^{22,81}, was predicted to be significantly lower in individuals who displayed the high inflammation immune response. Given its known anti-inflammatory function, this is an intuitive result. Interestingly, butyrate did not show a similar association - no significant trend was observed across the duration of the study. However, the limited sample size of the study

could limit the power of observing a strong anti-inflammatory relationship. Another mechanism that may contribute to lowering inflammation in these individuals is the more complete utilization of dietary fiber by the gut microbiota, leading to increased production of SCFAs such as propionate. Conversely, lower utilization of dietary fiber, leading to less SCFA production, may be causative of increased inflammation in some individuals. This is consistent with previous reports finding that unfermented β -fructans, such as the dietary fiber inulin, can contribute to inflammation through the NLRP3 and TLR2 pathways, and indirectly via change to the bile acid pool ^{33,34}. Thus, higher utilization of dietary fiber by the microbiota could dampen this effect. While it is impossible to say whether or not propionate flux predictions are underlying these dietary response phenotypes, the observed immune response groups and propionate production fluxes could not be explained by differences in alpha-diversity alone between groups. Generally, this result points to the capacity of metabolic modeling to elucidate the mechanisms underlying microbiome-related conditions, such as inflammation, and the overall relevance of the model-based predictions.

In comparing MCMM-based predictions for butyrate and propionate against blood-based clinical measurements, we sought to relate the production of these metabolites from the gut microbiome to health outcomes in the host. Upon regressing predicted values against a panel of blood-based clinical labs, 20 clinical markers were significantly associated with predicted butyrate production, but no significant associations were observed with MCMM-predicted rates of propionate production.

Among the associations with butyrate were several markers for systemic inflammation and metabolic health. For instance, C-reactive protein (CRP), a common biomarker for systemic inflammation, showed significant negative association with predicted butyrate production. CRP

is commonly elevated autoimmune disease, bacterial infection, and inflammation-related conditions such as atherosclerosis ⁸². Similarly, low-density lipoprotein (LDL) was also negatively associated with predicted butyrate production. LDL is commonly associated with atherosclerosis and the onset of cardiovascular diseases ⁸³. Mean arterial blood pressure was negatively associated with butyrate production, reflecting an effect that was previously reported in patients with hypertension ⁸⁴. Other significant negative associations with predicted butyrate production included inflammation-associated markers like alanine transaminase (increased with hepatic inflammation), arachidonic acid (induces pro-inflammatory eicosanoid production), blood glucose and LP-IR score (elevated with insulin resistance), triglycerides (increases risk of metabolic syndrome and cardiovascular disease) and uric acid (associated with inflammatory markers) ^{85–88}. The analysis also revealed positive correlations between predicted butyrate production and three biomarkers: HDL cholesterol, recognized for its role in reducing the risk of inflammation-associated cardiovascular disease ^{89,90}; adiponectin, a homeostatic adipokine with established functions in glucose regulation and anti-inflammatory processes ⁹¹; and chloride levels, known to be increased in cases of high butyrate transport into the colonic epithelium due to ionic exchange from active transport ⁹². Butyrate showed a significant negative association with BMI, as well. This association is also well-supported as previous results have shown that butyrate supplementation can decrease obesity in some individuals. Each significant association is consistent with what is known about how butyrate is protective against inflammation, cardiovascular disease, obesity, and metabolic disorders, and provide us with further confidence in the predictive power of our MCMM-based approach.

Interestingly, none of the clinical labs or measures were associated with direct blood metabolomic measurements of circulating butyrate *in vivo*. This is unsurprising, given that the

rapid consumption and processing of SCFAs by the host epithelium means that the therapeutic benefit of butyrate production can be observed without high levels of circulating butyrate. However, there was a weak but significant association between blood metabolomic measurements of circulating butyrate and MCMM-predicted butyrate. Importantly, this result further underscores the potential benefit of predicting *in situ* production of these metabolites, as the physiological signals expected to be seen from increased microbial butyrate production can be predicted effectively. Dietary interventions have long been known to elicit variable responses, but a mechanistic framework for predicting this microbiome-mediated heterogeneity has not been available until now.

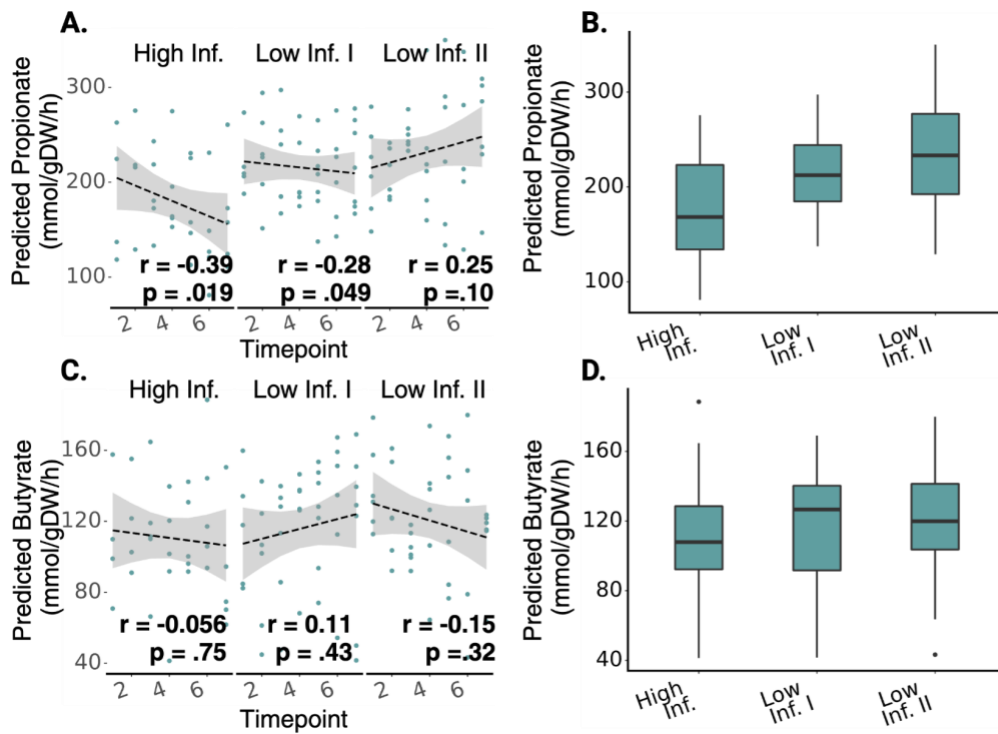


Figure 3.1 Predicted SCFA production profiles were associated with variable immune response groups following a high-fiber dietary intervention. (A) Total predicted propionate production at each timepoint across the three immune-response groups identified in the original study (High Inflammation: Pearson's Correlation $r = -0.39$, $p = .019$; Low Inflammation I: Pearson's Correlation $r = -0.28$, $p = .049$; Low Inflammation II: Pearson's Correlation $r = 0.25$, $p = .10$). Pearson's correlation r value and associated two-tailed p -value are calculated across the timepoints. (B) Average predicted propionate production rates, stratified by immune response group, $N = 18$ individual study participants (High vs. Low I, Mann-Whitney $U = 447$, $p = .0025$; High vs. Low II Mann-Whitney $U = 353$, $p = 2.6e-4$; Low I vs. Low II, Mann-Whitney $U = 744$, $p = 0.17$). The central line of the boxplot denotes the median value and the box contains the 25th to 75th percentile of data. The whiskers extend to the extreme points not more than $1.5 \times$ interquartile range from the median. (C) Total predicted butyrate production at each timepoint across the three immune-response groups identified in the original study (High Inflammation: Pearson's Correlation $r = -0.056$, $p = .75$; Low Inflammation I: Pearson's Correlation $r = 0.11$, $p = .43$; Low Inflammation II: Pearson's Correlation $r = -0.15$, $p = .32$). Pearson's correlation r value and associated two-tailed p -value are calculated across the timepoints. (D) Average predicted butyrate production rates, stratified by immune response group, $N = 18$ individual study participants (High vs. Low I, Mann-Whitney $U = 626$, $p = .22$; High vs. Low II Mann-Whitney $U = 546$, $p = .11$; Low I vs. Low II, Mann-Whitney $U = 899$, $p = 0.98$). In C & E stars denote significance under a Mann-Whitney U-test based on associated two-tailed p -value * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$. In B & D, The dashed black line denotes a linear regression line and the gray area denotes the 95% confidence interval of the regression line.

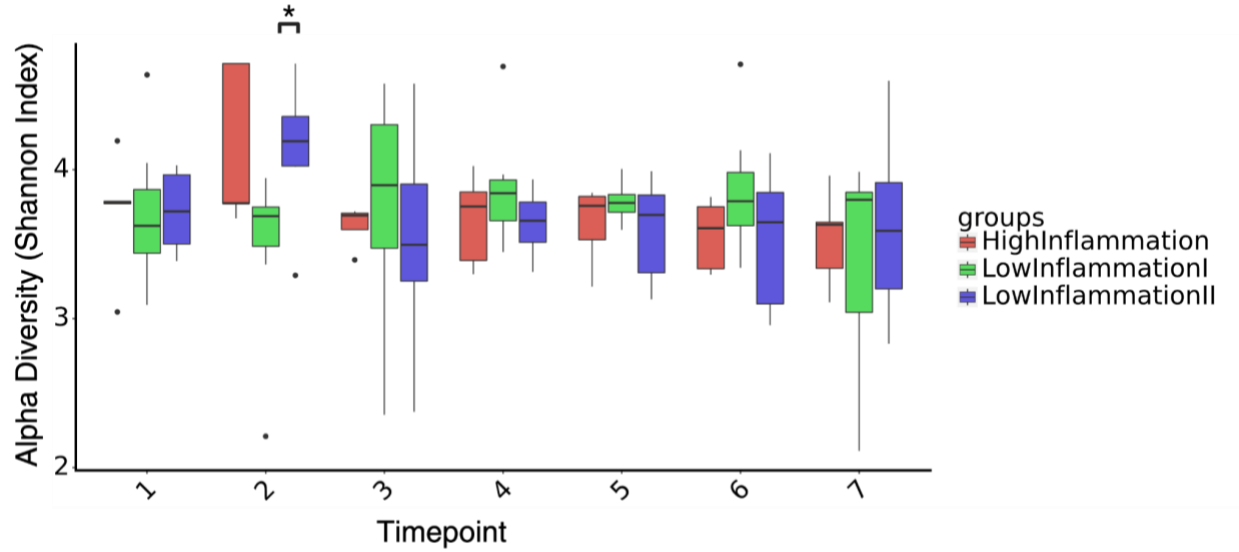


Fig 3.2 Alpha diversity in longitudinal high fiber study. We compared Shannon index, a measure of alpha diversity, between immune response groups in a longitudinal high fiber study. No consistent pattern emerged regarding alpha diversity between immune response groups throughout the course of the high fiber dietary intervention, as determined by Mann Whitney U test for significance. The central line of the boxplot denotes the median value and the box contains the 25th to 75th percentile of data. The whiskers extend to the extreme points not more than 1.5*interquartile range from the median. N = 18 individual study participants, * = $p < 0.05$.

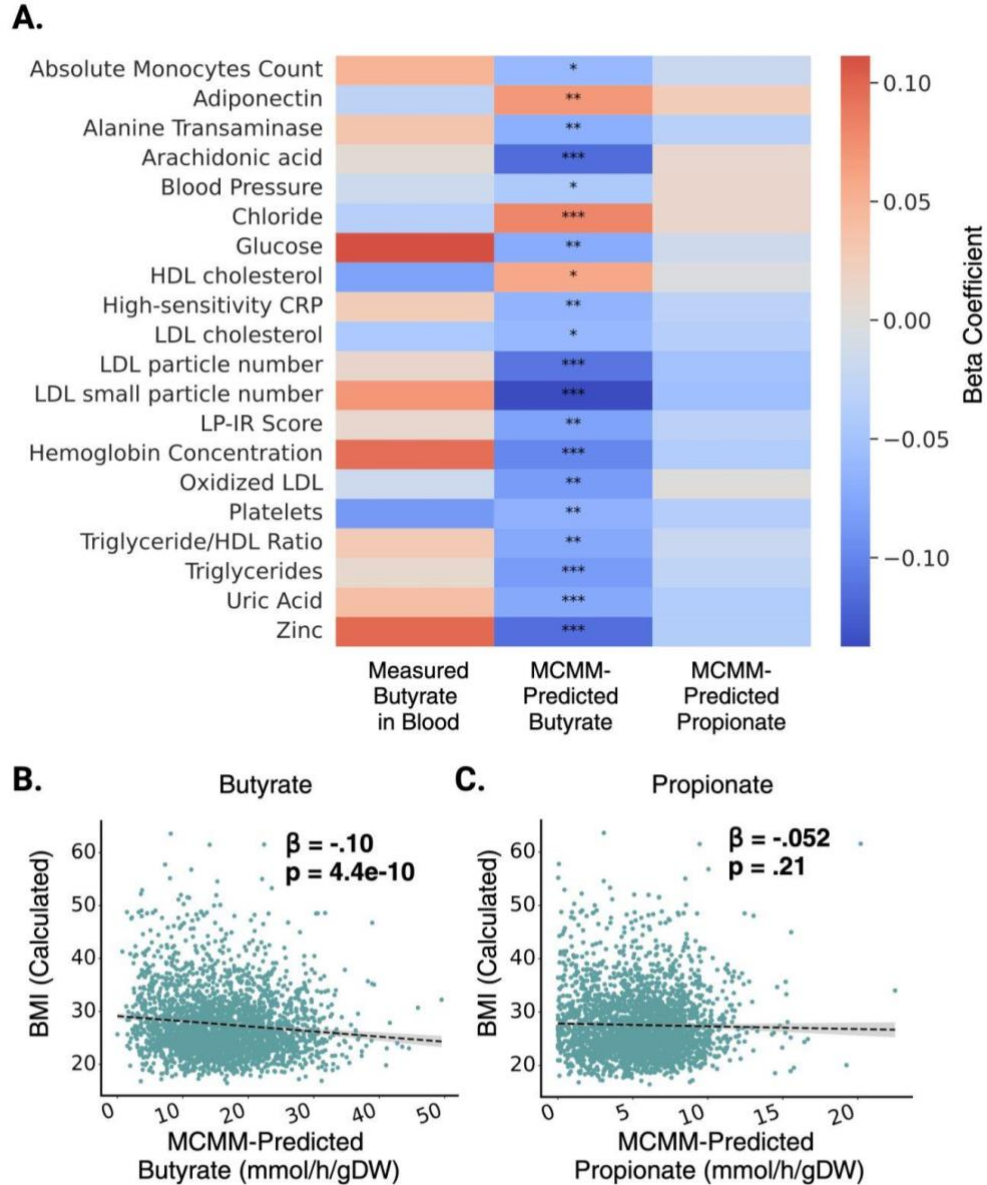


Figure 3.3 SCFA flux predictions are significantly associated with blood-derived clinical markers. (A) Heatmap showing the 20 significant associations between measured blood markers and predicted SCFA production rates, as well as measured butyrate in blood for individuals in the Arivale cohort (FDR-corrected p-value, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, $N = 3,129$). Color encoding indicates β -coefficient for individual variables in linear regression correcting for age and sex **(B-C)** Predictions for butyrate were significantly correlated with reported BMI measurements for respective participants (two-sided t-test $p = 4.4e-10$), but not for propionate (two-sided t-test $p = .21$), according to linear regression. Each dot denotes an individual model reconstructed for a single sample in the Arivale study ($N = 3,129$). The dashed black line denotes a linear regression line and the gray area denotes the 95% confidence interval of the regression line.

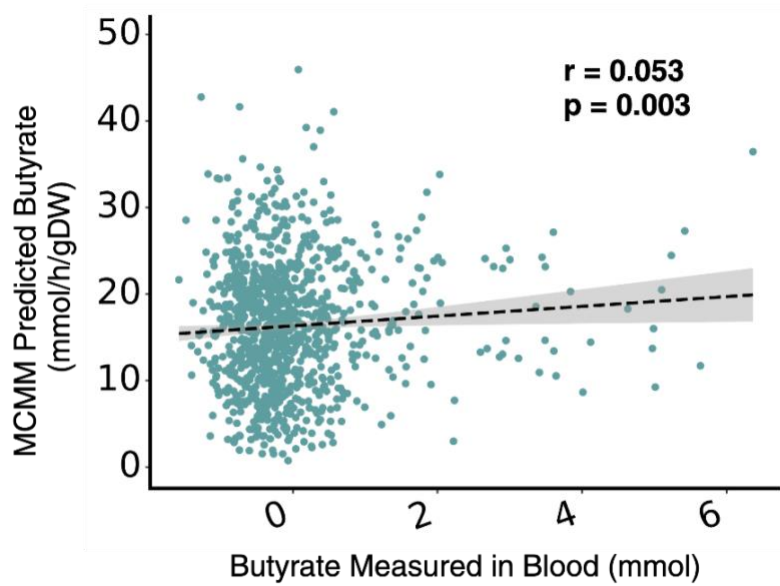


Figure 3.4 Association of MCMM-predicted butyrate with blood-measured butyrate. MCMM predicted values of butyrate production fluxes from the gut microbiome show weak, but significant association with blood metabolomic measures of circulating butyrate in the Arivale cohort (Pearson Correlation $r = 0.053$, $p = .003$). Each point represents one individual. Pearson correlation r value and associated two-tailed p -value were calculated across all points. The dashed black line denotes a linear regression line and the gray area denotes the 95% confidence interval of the regression line.

<i>Study</i>	<i>Source</i>	<i>Publication</i>	<i>Sample Size</i>	<i>Total taxa per Sample</i>	<i>Total taxa in model per sample</i>
Fiber Dietary Intervention	Sonnenburg Lab, Stanford University	Wastyk et al. 2021 ⁷⁷	18	16.7 ± 0.23	16.7 ± 0.23
Arivale	Arivale, Inc. Seattle WA	Manor et al. 2020 ⁹³	3,129	43.2 ± 0.14	29.5 ± 0.08

Table 3.1 List of *in vivo* datasets leveraged for analysis of clinical relevance

Beta Coefficients				p-values			
	Measured Butyrate in Blood	MCMM-Predicted Butyrate	MCMM-Predicted Propionate	Name	Measured Butyrate in Blood	MCMM-Predicted Butyrate	MCMM-Predicted Propionate
Absolute Monocytes Count	0.0484	-0.0595	-0.0183	Absolute Monocytes Count	1.0E+00	3.8E-02	1.0E+00
Adiponectin	-0.0295	0.0687	0.0260	Adiponectin	1.0E+00	4.7E-03	1.0E+00
Alanine Transaminase	0.0328	-0.0689	-0.0339	Alanine Transaminase	1.0E+00	4.7E-03	1.0E+00
Arachidonic acid	0.0064	-0.1168	0.0103	Arachidonic acid	1.0E+00	2.2E-06	1.0E+00
Chloride	-0.0348	0.0809	0.0136	Chloride	1.0E+00	7.5E-04	1.0E+00
Glucose	0.1114	-0.0710	-0.0150	Glucose	8.1E-01	4.7E-03	1.0E+00
HDL cholesterol	-0.0788	0.0581	-0.0030	HDL cholesterol	8.5E-01	1.8E-02	1.0E+00
High-sensitivity CRP	0.0263	-0.0651	-0.0310	High-sensitivity CRP	1.0E+00	4.7E-03	1.0E+00
LDL cholesterol	-0.0433	-0.0605	-0.0363	LDL cholesterol	1.0E+00	3.6E-02	1.0E+00
LDL particle number	0.0138	-0.1107	-0.0504	LDL particle number	1.0E+00	9.5E-08	9.3E-01
LDL small particle number	0.0705	-0.1375	-0.0548	LDL small particle number	1.0E+00	1.2E-11	9.3E-01
LP-IR Score	0.0112	-0.0788	-0.0322	LP-IR Score	1.0E+00	6.1E-03	1.0E+00
Hemoglobin Concentration	0.0949	-0.0991	-0.0391	Hemoglobin Concentration	8.3E-01	2.5E-05	1.0E+00
Oxidized LDL	-0.0158	-0.0846	0.0013	Oxidized LDL	1.0E+00	7.0E-03	1.0E+00
Platelets	-0.0867	-0.0673	-0.0368	Platelets	8.1E-01	6.3E-03	1.0E+00
Triglyceride/HDL Ratio	0.0283	-0.0726	-0.0197	Triglyceride/HDL Ratio	1.0E+00	4.7E-03	1.0E+00
Triglycerides	0.0100	-0.0835	-0.0273	Triglycerides	1.0E+00	5.0E-04	1.0E+00
Uric Acid	0.0400	-0.0735	-0.0385	Uric Acid	1.0E+00	2.2E-04	9.9E-01
Zinc	0.0983	-0.1145	-0.0378	Zinc	8.1E-01	2.2E-04	1.0E+00

Table 3.2 Clinical measures showing significant association with MCMM-predicted butyrate production

Chapter 4

Using MCMMs to Design and Test Personalized Interventions

Portions of this chapter were published as a manuscript in Nature Microbiology ⁷⁰.

4.1 Introduction

Community-scale metabolic models of the gut microbiome have been shown to make accurate predictions of SCFA production in response to fiber supplementation, and these predictions accurately reflect immunological responses to dietary interventions and systemic inflammation levels. This ultimately raises the question of whether this approach can be used to design and test personalized dietary and probiotic interventions *in silico*. As described, the metabolic response of the gut microbiome to the same dietary input can vary between individuals based on taxonomic composition and abundance of a community ³². This can alter the efficacy of interventions, making a computational method for testing them a useful tool. In effect, metabolic modeling could be used by clinicians to assess the suitability of one intervention over another for a given individual. As alluded to previously, incomplete fiber degradation by the microbiome

can contribute to increased inflammation in some cases ^{33,34}. Consequently, administering a high-fiber diet to individuals at risk of systemic inflammation with the aim of mitigating this issue through SCFA production may not always be beneficial, as it can potentially exacerbate rather than alleviate the problem.

Importantly, this methodology could indicate the necessary severity of mediation in a given diet. For instance, in the case of improved SCFA production from the microbiome, the model could point to one of three possible cases. First, that a full dietary shift is required to sufficiently increase production; second, that augmentation of SCFA production may only require the addition of a small set of metabolic precursors or prebiotics; or finally, that the metabolic capacity of the given microbiome is inadequate, and compositional shifts in the form of probiotics are necessary. Given the heterogeneous response of different microbial communities to the same dietary input, having this information *a priori* could be a major benefit to the design and implementation of dietary interventions, and oppose the outdated one-size-fits-all approach to nutrition-based interventions ⁹⁴.

To investigate an MCMM-based approach to this avenue, we again leveraged the community-scale metabolic models constructed from 3,129 participants in the Arivale cohort. First, all models were grown in two dietary contexts: a standard European diet, and a standardized high-fiber diet, with the expectation that higher levels of fiber would serve as increased substrate for bacterial fermentation into SCFAs. Based on response to these dietary interventions, we then designed and tested a number of combinatorial interventions aimed at increasing butyrate production. Interventions included single fiber additions of inulin or pectin, as well as a simulated probiotic intervention of *Faecalibacterium*. The results informed the most optimal intervention for each individual.

Similarly, we sought to analyze the capability of metabolic modeling of short-chain fatty acid production to capture differences between responders and non-responders to a HAMSAB (acetylated or butyrylated high-amylose maize starch) diet. In this study, participants were split into two study arms, and fed either a placebo or HAMSAB-rich diet. Responders to the diet showed a decrease in systolic blood pressure, while non-responders did not ⁸⁴. Given the blood pressure lowering properties of butyrate in the body, we hypothesize that metabolic predictions of butyrate production by the gut microbiome would be associated with physiological response to the diet. Using individual-specific MCMMs from baseline sequencing data, we could identify responders and nonresponders to the HAMSAB diet by examining predicted butyrate production.

A list of all datasets included in this analysis is available at the end of the chapter (**Table 4.1**).

4.2 Methodology

4.2.1 MCMM construction for designing personalized interventions

MCMMs constructed from data for individuals in the Arivale cohort were reused from Chapter 3. Growth of each model was simulated using both the average European diet and high fiber diet described in Chapter 3, using a tradeoff parameter of 0.7. From these simulations, predicted community butyrate production was calculated for each sample for both dietary contexts. From these results, samples were categorized into three groups. Non-responders, who produced <15 mmol/gDW/h of butyrate on the European diet and showed an increase of <20% on the high-fiber diet; regressors, who showed a decrease in butyrate production on the high-fiber diet when compared to the European diet; and responders, who did not fit into these categories. For both the non-responders and regressors, additional interventions were tested.

Using the European and high-fiber diets as a base, single fiber interventions were created. First, inulin was added to each, with an upper import bound of 14 mmol/gDW/h. Second, pectin was added, with an upper import bound of 1 mmol/gDW/h. Next, a simulated probiotic intervention was included by scaling the relative abundance of existing models to 95% of their original, and adding 5% relative abundance of *Faecalibacterium*, simulating engraftment. Growth for each of the non-responders and regressors was simulated on each of the diets and simulated interventions, for a total of 8 simulations per sample. The optimal intervention that resulted in the largest value for predicted butyrate was selected for each individual.

4.2.2 MCMM construction for HAMSAB dietary study

For participants in the HAMSAB diet study, shotgun metagenomic data from baseline-collected fecal samples for individuals in the treatment arm of the study were collected. Preprocessing of raw sequencing reads from Study A and D was performed using FASTP⁹⁵. The first 5bp on the 5' end of each read were trimmed, and the 3' end was trimmed using a sliding window quality filter that would trim the read as soon as the average window quality fell below 20. Reads containing ambiguous base calls or with a length of less than 15bp after trimming were removed from the analysis. Bacterial species abundances were quantified using Kraken2 (v2.0.8) and Bracken (v2.2) using the Kraken2 default database which was based on Refseq release 94, retaining only those species with at least 10 assigned reads^{96,97}. MCMMs were constructed by mapping relative abundance data for each individual to species-level GEMs from the AGORA database (v1.03) in MICOM, with an abundance threshold of 0.001. A carbon-depleted average European medium containing high levels of amylose fiber was used to simulate the growth of samples, and predicted butyrate production was calculated. Using dietary survey data, the exact amount of fiber added was scaled to reflect the reported fiber consumption of

each individual. To account for butyrylation, amylose consumption was calculated and added to predicted butyrate production to account for the butyrate directly liberated from starch degradation (independent of microbial butyrate production pathways). Resulting predictions for butyrate production were compared between responders and non-responders using a two-tailed Student's t-test.

4.3 Results

4.3.1 MCMM-based prediction of individual response to dietary interventions

MCMMs were built using two different dietary contexts: an average European diet, and a vegan, high-fiber diet rich in resistant starch. As expected, models grown on a high-fiber diet showed higher average predicted butyrate production (**Fig. 4.1B**, 87.78 ± 0.67 mmol/(gDW h) vs 16.29 ± 0.13 mmol/(gDW h), two-tailed Student's t-test, $t = 104.3$, $p < 0.001$). 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' (**Fig. 4.1C**, $N = 10$) and a set of 'regressors' (**Fig. 4.1D**, $N = 6$) who showed only small increases or decreases in butyrate production on the high-fiber diet when compared to the European diet. 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. Predictions for butyrate production were then observed for each of the samples between the interventions.

For 15/16 individuals in the regressor or non-responder groups, supplementation of the background diet with a specific prebiotic or probiotic increased the butyrate production rate (**Fig. 4.1C-E**). For both regressors and non-responders, the optimal intervention showed substantial increases over the standard European diet ($+290 \pm 80\%$ for non-responders; $+239 \pm 102\%$ for regressors). The exact intervention that yielded the highest butyrate production for any given individual across both populations varied widely (**Fig. 4.1E**). For example, the probiotic intervention was more successful in raising predictions for butyrate production in non-responders than it was in regressors (**Fig. 4.1E**). Overall, no single combinatorial intervention was ideal for every individual in the population.

4.3.2 HAMSAB diet shows differential SCFA production between groups

When analyzing the model-predicted butyrate production from individuals fed a HAMSAB-rich diet, a marked disparity emerged between individuals categorized as non-responders and those classified as responders (responders: $N = 5$ individual participants, mean predicted production = 17.6 ± 4.9 ; non-responders: $N = 7$, mean predicted production = 6.5 ± 2.0). Specifically, the responder group exhibited a significant increase in butyrate production compared to their non-responder counterparts, despite the very limited sample size of this study (**Fig 4.2**, Student's t-test, t statistic = -2.34 , $p = 0.041$).

4.4 Discussion

This chapter builds upon previous validation and health-related associations of microbial community-scale metabolic modeling and details its clinical promise, with respect to the design of, and predicted response to, dietary interventions. Specifically, we sought to determine the most effective solutions for each individual with regard to elevating microbial butyrate

production. It is well demonstrated that butyrate is strongly anti-inflammatory, regulates blood pressure and is associated with several clinical markers of health. Having depressed levels of butyrate production, whether due to improper diet, lack of fermentation capacity by the microbiota, or both, is associated with increased inflammation and elevated risk for the onset of clinical pathologies like cardiovascular disease, metabolic syndrome and diabetes. Engineering metabolic responses in these individuals to improve fermentation capacity of the microbiome and increase SCFA production can contribute to disease prevention and improved health outcomes.

By identifying individuals who did not respond as expected to a standard high-fiber dietary intervention with regard to predicted butyrate production, we can demonstrate that the one-size-fits all approach to nutrition is antiquated, and that expectations of population-scale interventions do not necessarily match individual results. For instance, the group of non-responders showed low baseline butyrate production and did not show a putatively expected increase on the high-fiber diet, and required extra consideration. Importantly, in many individuals, predictions showed that a limited response to a full dietary switch could be recovered with the addition of a single fiber not present in the original diet. It has been suggested that making small adjustments in habits, such as incorporating single food groups or utilizing prebiotic supplements, may foster greater adherence among participants compared to implementing large-scale changes, such as completely altering one's diet ⁹⁸. Having a computational framework for predicting responses is additionally beneficial versus the alternative guess-and-check approach.

It is important to consider, too, that the interventions tested here represent only a small fraction of possibilities; again, the mechanistic nature of these models means that an infinite set of combinatorial interventions can be tested, combining dietary components in any number of

formulations and ratios. It is possible, and entirely likely, that the most optimal intervention for each individual was not shown in this proof-of-concept study. Thus, increasing the number and complexity of tested dietary additions would increase the set of effective options for treating individuals exhibiting low microbial butyrate production. For instance, alternative fibers such as mixed fructooligosaccharides or seaweed-derived laminarin could be included in the computational assessment.

An interesting observation of this analysis is that in some individuals, varying dietary substrates can show little to no increase in butyrate production. Rather, a shift in taxonomic composition is required to show any substantial improvements at all. This is rather intuitive, in that butyrate-producing species must necessarily be present for butyrate production to occur. Probiotic interventions can accomplish this by introducing microbes to the community with high capacity for fermentation. While we demonstrate the effect of introducing just one taxon, *Faecalibacterium*, mechanistic models such as these provide the capability of testing any number and combination of potential probiotic interventions. Having the capacity to identify individuals that fall into this category, and subsequently to determine the effectiveness of any combination of probiotic treatments, could provide a potential benefit for precision microbiome-targeted therapies, especially in the case of microbial dysbiosis.

Beyond designing *de novo* personalized diets, MCMMs also have the promise of predicting individual response to a specific given diet. Tailored diets aimed at improving patient health are central to precision nutrition, but their efficacy can be difficult to predict prior to their application. In the example of the HAMSAB diet, we hoped to show that predicted microbial response to a particular high-fiber diet could be indicative of the physiological response to said diet. In this case, the dietary intervention, which was aimed at utilizing butyrylated starch to

ameliorate systolic blood pressure, showed a beneficial response in some individuals but not in others. Our results show that there is a distinct signature with regard to production of butyrate from the microbiome in response to application of the HAMSAB diet. Further work is required to fully validate these results. In this case, a carbon-depleted average European diet was used, in order to isolate the individual effects of high-amylose starch. Given the availability of individual dietary surveys, patient-specific diets will be constructed, increasing the context specificity of the modeling scheme. However, given the very small sample size of the cohort used in this analysis, the significance of the results presented here display an encouraging capability of this modeling platform to predict physiological responses to HAMSAB-rich diets.

It is well established that the microbiome can impact trajectories of human health ^{2,4}. This chapter aimed to show how community-scale metabolic modeling can effectively assist in designing and testing microbiome-directed therapies. By taking a mechanistic approach to predict outcomes, we broaden the range of interventions that can be explored beyond what traditional machine learning methods can achieve.

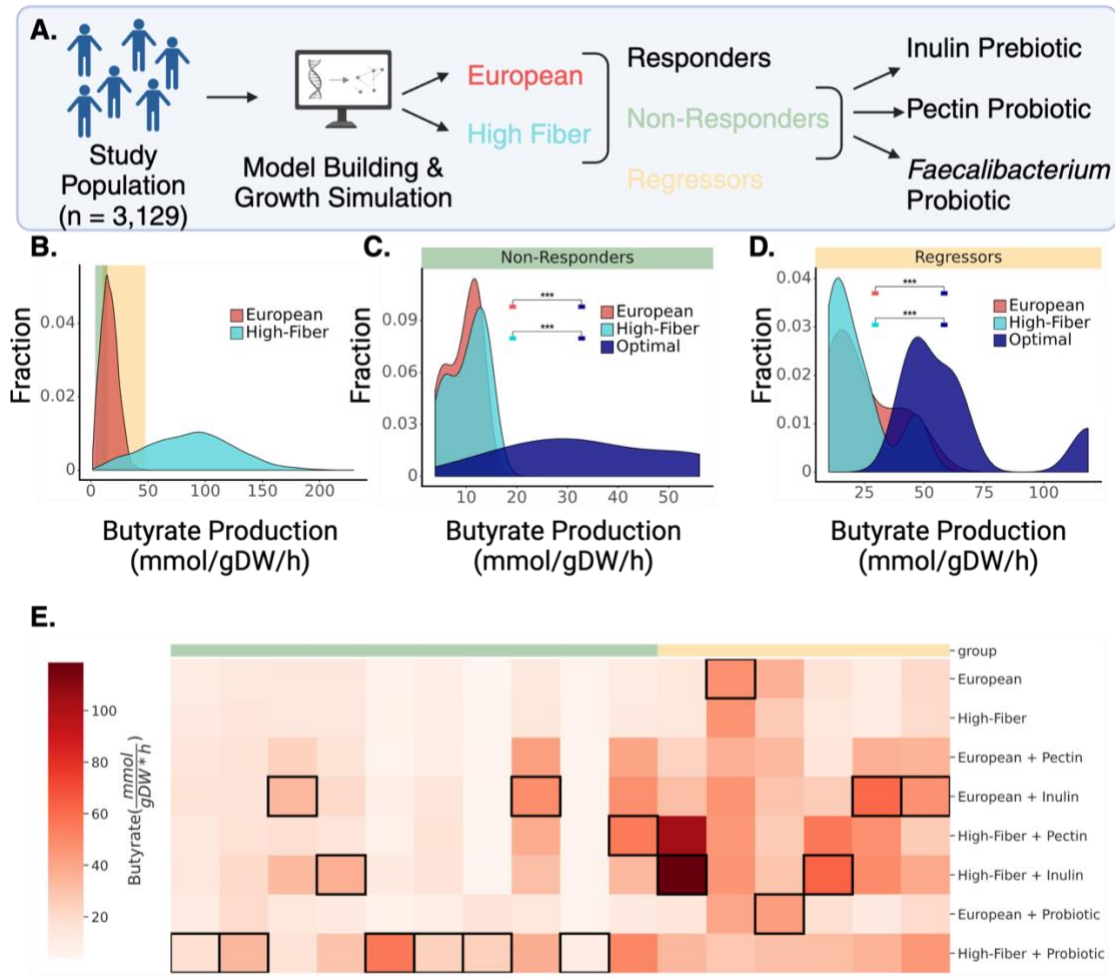


Figure 4.1. Microbial MCMs can be used to design and select personalized prebiotic, probiotic, and dietary interventions aimed at optimizing SCFA production profiles. (A) MCMs built from the Arivale cohort (N = 3,129 individual participants) were used to test personalized responses to dietary interventions. Personalized models divided into responders, non-responders, and regressors, based on the changes in predicted butyrate production in response to increasing dietary fiber. (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 = 10) and regressors (N = 6) are highlighted in green and yellow shaded areas, respectively. (C) Distributions of butyrate production rates for the non-responder group (N = 10). The intervention resulting in the highest butyrate production is shown in blue. *** = p < 0.001, two-sided t-test. (D) Butyrate production rates for the regressor group (N = 6). The optimal intervention that resulted in the highest butyrate production is shown in blue. *** = p < 0.001, two-sided t-test. (E) Heatmap of butyrate production rates across simulated interventions for the individuals in the non-responder and regressor groups. Rows denotes specific interventions, columns denote individuals in the response groups (N = 16). Cell shading (white-to-red) denotes predicted butyrate production rate. Added interventions tested on both non-responders and regressors included prebiotic supplementation (inulin or pectin) as well as probiotic supplementation (5% relative abundance *Faecalibacterium*). The most successful intervention for each individual is denoted by a black border around that cell in the corresponding column.

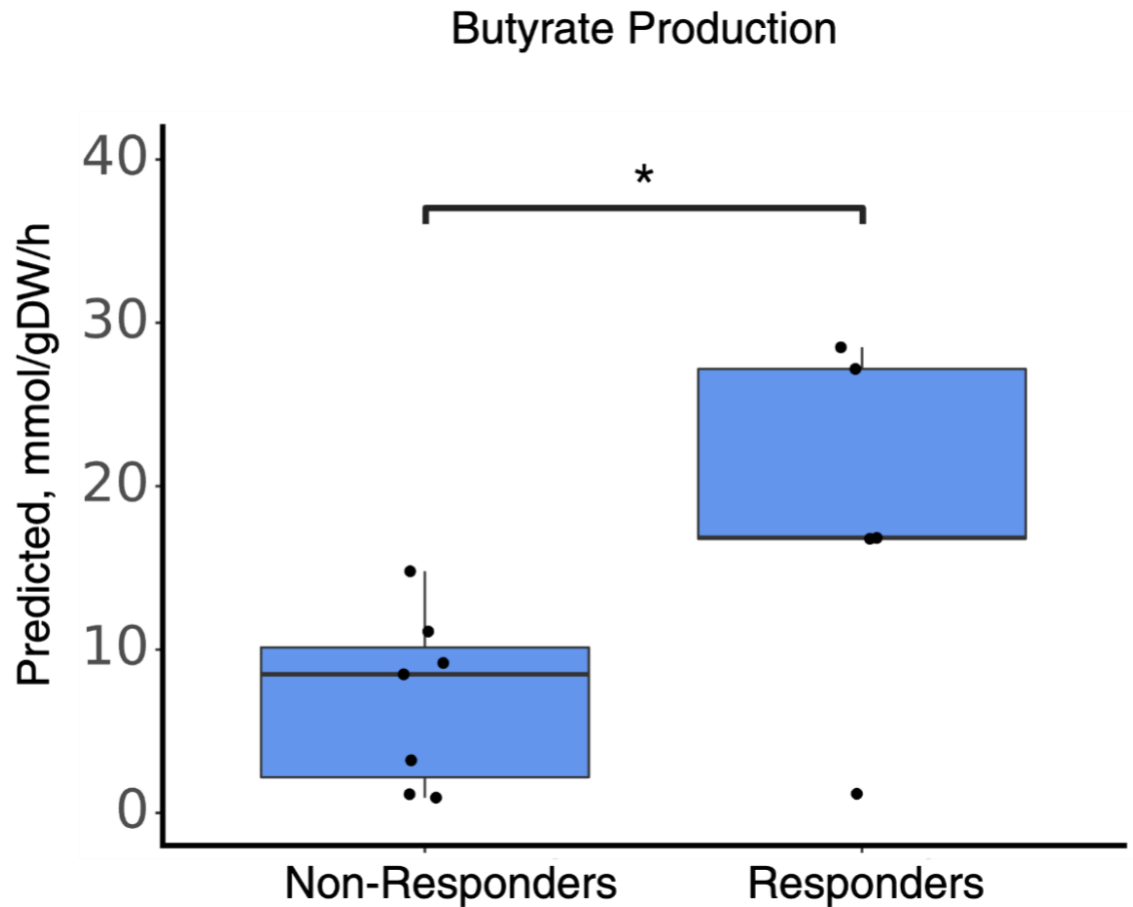


Figure 4.2. Predicted butyrate production associated with physiological response to HAMSAB diet. MCMC-based predictions for HAMSAB-treated participants shown in the boxplots (responders: N = 5 individual participants, mean production = 17.60 ± 4.91 ; non-responders: N = 7, mean production = 6.52 ± 2.01). Significance determined by two-tailed t-test (t statistic = -2.34, p = 0.041). The central line of the boxplot denotes the median value and the box contains the 25th to 75th percentile of data. The whiskers extend to the extreme points not more than 1.5*interquartile range from the median.

Arivale	Arivale, Inc. Seattle WA	Manor et al 2020 ⁹³	3,129	43.2 ± 0.14	29.5 ± 0.08
HAMSAB Diet Trial	Marques Lab, Monash University	Unpublished	13	53.8 ± 3.6	30.4 ± 1.8

Table 4.1 List of datasets leveraged for dietary design and response prediction

Chapter 5

Toward integration of MCMs with a Metabolic Reconstruction of the Colonic Epithelium

5.1 Introduction

Predictive modeling of the gut microbiome holds promise for clinical applications, given the established links between microbiome-derived metabolic production and human health. Understanding the mechanisms underlying metabolite production and consumption rates and their variation across individuals paves the way for targeted therapeutic interventions in precision nutrition. However, the impact of these microbially-derived metabolites is primarily driven by transport into the host through the colonic epithelium ¹¹. For example, TMA produced by the microbiome shows an increased association with cardiovascular diseases, but only following its transport to the liver and subsequent conversion to TMAO ³⁸. To comprehensively capture the metabolic exchange of small molecules with the host and the subsequent phenotypic effects, we aimed to develop a functional multi-compartment, host-microbial metabolic model. This model would integrate microbial community-scale metabolic models of the gut microbiome with tissue-specific models of the colonic epithelium. Such an approach would not only enable the

prediction of metabolism at the microbial community level but also facilitate the examination of metabolite exchange with the host.

The construction of a tissue-specific metabolic model of the colonic epithelium was carried out in collaboration with the Baloni Laboratory at the School of Health Sciences, Purdue University. Utilizing single-cell RNA sequencing data, four methods were employed for reductive model construction (tINIT, iMAT, CORDA, and pymCADRE), utilizing the comprehensive human reconstruction model RECON3D as a universal framework. Gene expression measures were binarized - present calls were given a score of 1, and absent calls given a score of 0. This was repeated across all samples present in the gene expression data. After binarization, ubiquity scoring of genes calculated how often a gene was measured as present across all samples. Essentially, an average of the binarization score across all samples for a gene was calculated, ranging from 0 to 1. Ubiquity scores were used by each of the model construction methodologies to calculate an expression-based evidence score, calculating how likely a reaction is to take place given the expression data. Additionally, connectivity-based evidence score is calculated from expression-based evidence along with network topology, calculating how likely a reaction is to be present based upon its relationship with other reactions in the model ⁶². Using these evidence scores, the non-core reactions in the generic human model are ranked. Working iteratively through the model, each non-core reaction was pruned starting from the lowest ranked reaction. If removing a reaction does not affect flux through the set of core reactions or the production of key metabolites, it is removed from the model. Following this process across four methodologies left a consensus draft reconstruction. Model specificity of the reconstruction was assessed across 47 metabolic functions known to be essential for the colonocyte function (**Table 5.1**). An important consideration in the construction of the model

was the application of a relevant metabolic objective function. In this case, an objective function was selected that optimized flux through three reactions: a biomass maintenance reaction, and two reactions included in the β -oxidation of fatty acids (**Table 5.2**). As butyrate is the primary oxidative substrate for colonocytes, this inclusion made sense to properly capture the metabolic objectives of the cells being modeled. A list of data included in construction of the colonocyte model can be found in **Table 5.3**.

Upon completion of a cell-type specific metabolic reconstruction of the colonic epithelium, the model had to be integrated with community-scale metabolic models of the gut microbiome. A fully functional, multi-tissue metabolic model would enable future investigations into the downstream effects of microbiome directed therapies and interventions. However, due to the complexity of the model and number of distributed compartments, special consideration must be given to the segregation of reactions. Here, we present the results of an initial approach at this integration, and the subsequent results obtained from of this host-microbiome metabolic model. By elucidating the steps involved in model construction, we lay the foundation for our overarching objective of developing a comprehensive multi-compartment metabolic model capable of capturing the complex interplay between the gut microbiome and host metabolism. This paves the way for deep exploration of the mechanisms underlying the connection between diet, the microbiome and human health.

5.2 Methodology

The fully constructed and validated metabolic model of the colonic epithelium was obtained from collaborators in the Baloni Lab at the College of Health and Human Sciences at Purdue University, and analyzed using the COBRA Toolbox (v0.26.0) ⁵⁶. Reactions within the

model were segregated into ten distinct compartments, including eight intracellular compartments and two extracellular, one representing the basolateral (host-facing) side of the cell and one representing the apical (luminal) side. Without manipulating internal reactions, biologically relevant exchange reactions into and out of the model were added, such that microbiome-supplied metabolites imported into the apical side and host-related metabolites (i.e., present in the bloodstream) transported through the basolateral side. These added transport reactions facilitated integration with MICOM models.

To examine the behavior of the colonic model, MCMMs built from the Arivale cohort were utilized. Growth of these models was first simulated using an average European diet, collected from VMH and amended as described previously. Bounds on exchange reactions into the colonic model from the apical side were then individually restricted to match predicted metabolic production from these MCMMs, as well as unconsumed medium metabolites. Additionally, basolateral exchanges were limited to just those included in a previously published description of the blood metabolome, effectively limiting metabolic availability to the model to those metabolites provided by the microbiome or through circulation ⁶⁹. Using classical flux balance analysis, growth of the model under this initial integration scheme was analyzed, and predictions collected for metabolic exchange into and out of the colonocyte model. To assess clinical relevance of predicted fluxes, transports into the colonocyte model were regressed against blood-based clinical labs collected from each participant using statsmodels, correcting for age and sex. Standardized beta-coefficients for each association were reported, as well as FDR-corrected p-values for each association.

5.3 Results

5.3.1 *Integration of the colonocyte model with MCMMs*

A diagram of the compartmentalization of the integrated model displays how reactions were segregated, keeping biologically distinct functions separate (**Fig. 5.1**). Under the nutrient-limited conditions described above, all components included in the objective function were able to carry flux, indicating the metabolic precursors available were sufficient to optimize the central function of the model (**Fig. 5.2**). Flux through the first step of the β -oxidation pathway varied between individuals, dependent on the predicted butyrate production from the microbiome, which bounded butyrate import into the colonocyte model.

5.3.2 *Association of predicted colonic imports with blood-based clinical measures*

Regression analysis of predicted metabolic imports into the colonic epithelium against a panel of blood-based clinical labs showed several salient associations (**Fig. 5.3**). Butyrate, as shown previously, showed the most associations, mainly with the same markers for cardiometabolic health and inflammation discussed in Chapter 3. Other notable associations uncovered in this analysis were negative associations with C-Reactive Protein for formate, a positive association with total antioxidant capacity for riboflavin, and a positive association between thiamine (vitamin B1) and ergocalciferol (vitamin D2).

5.4 Discussion

Developing a metabolic model of the colonic epithelium, coupled with community-scale models of the gut microbiota, holds great potential. Although the impacts of common microbially-derived metabolites such as SCFAs on physiology are widely recognized, such a model could facilitate the understanding of connections between less-explored molecules and health outcomes in the host. Here, we present an initial exploration into this potential research direction.

By leveraging a functional reconstruction depicting the metabolic capabilities of a human colonocyte, we successfully merged this model with pre-existing individual specific MCMs. Employing solely physiologically pertinent metabolites accessible to the model, we successfully simulated metabolic intake and output patterns of the colonocyte, with feasible maximization of the objective value. This achievement is promising for model viability, yet warrants further validation, as reported fluxes may not always align with biological relevance.

Due to the infeasibility of direct measurement of *in vivo* exchanges, we instead turned to associations found between the transport of metabolites from the colonic lumen into the colonic epithelium, and a panel of blood-based clinical labs. Here, several metabolic transports showed interesting and well-supported associations. Butyrate, as discussed in Chapter 3, was strongly associated with several markers for cardiometabolic health and health. Interestingly, other metabolic transports showed similar associations. Riboflavin (vitamin B2) showed a strong positive association with total antioxidant capacity. Considering riboflavin's primary biological

function as an antioxidant that reduces reactive oxygen species, this result is well-supported ⁹⁹. However, not all associations were well supported by evidence in literature. For instance, formate showed a negative association with C-reactive protein, though general consensus is that it is largely a proinflammatory molecule and is commonly associated with colitis ¹⁰⁰. That said, formate production is also commonly involved in microbial overflow metabolism, meaning that MCMM-predicted rates of formate prediction may not accurately reflect true values, as discussed previously. This may lead to an inaccurate prediction of colonic import of the metabolite, explaining this association. An interesting relationship was the positive association between thiamine (vitamin B1) and ergocalciferol (vitamin D2). These vitamins may be co-enriched in specific foods, leading to their co-occurrence.

The current construction of the integrated model raises a critical consideration regarding the utilization of β -oxidation as the primary objective function. This choice aligns with our understanding of colonocyte function, where the majority of cellular energy is derived from microbially-produced butyrate ¹⁸. However, this poses a technical challenge concerning the predicted transport of butyrate. By maximizing flux through β -oxidation pathways, the model simultaneously maximizes the import of butyrate, resulting in 100% utilization of microbially-produced butyrate across all models. While this discrepancy is not substantial in absolute terms, given that approximately 95% of microbially-produced butyrate is indeed imported by colonocytes, it enforces high rates of butyrate import, rather than allowing it to be an emergent property of the model ⁶⁴. Although valuable insights can still be gleaned from the transport of other metabolites, this consideration must be addressed in future endeavors aimed at modeling the microbiome-host interface.

Along with the promise of metabolic modeling of the gut microbiome for precision nutrition, we anticipate that this ongoing work toward the inclusion of a human colonocyte model will allow for deeper understanding of some of the mechanisms underlying human health and disease. While significant efforts are still required for the full integration of the colonocyte model, especially accounting for resource competition between the host and microbiota in a single optimization step, the work presented here is an encouraging step towards a comprehensive understanding of how diet and the microbiome shape health or disease trajectories in humans.

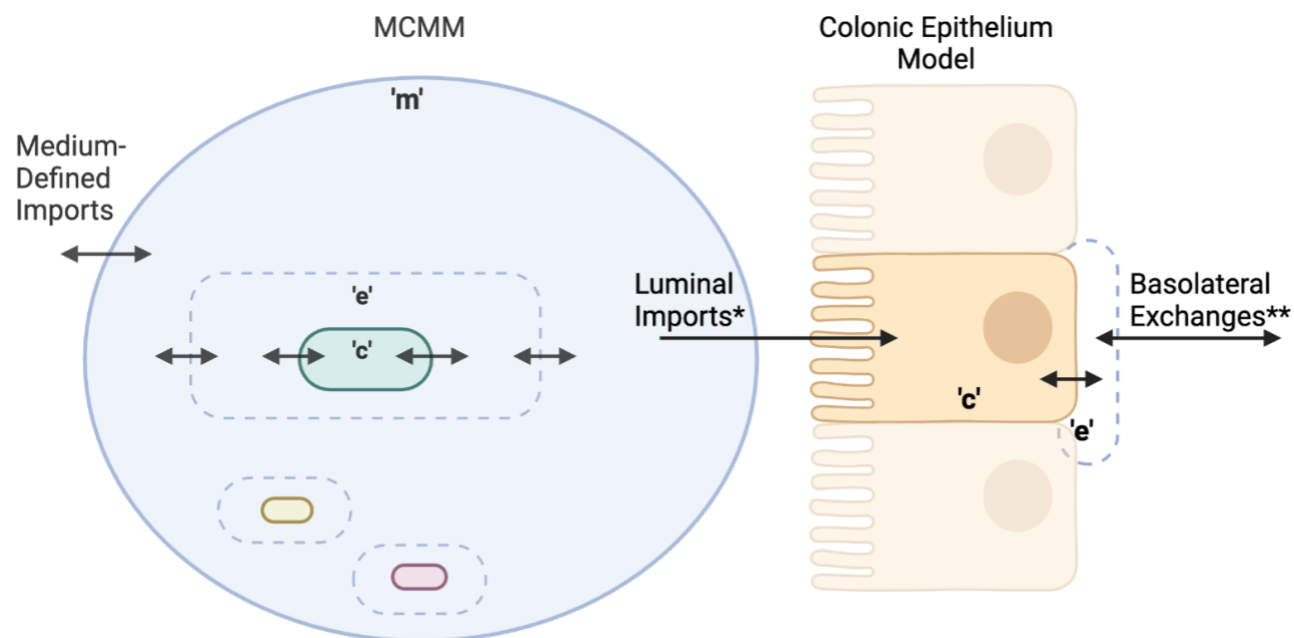


Figure 5.1 Compartmentalization of integrated host-microbiome multi-system model. A functional metabolic model of the colonic epithelium was integrated with existing MCMs of the gut microbiome. Following optimization of the microbial community, luminal imports were limited to MCM-predicted production rates and unconsumed medium components. Incoming basolateral exchanges were limited to metabolites found in a published description of the blood metabolome.

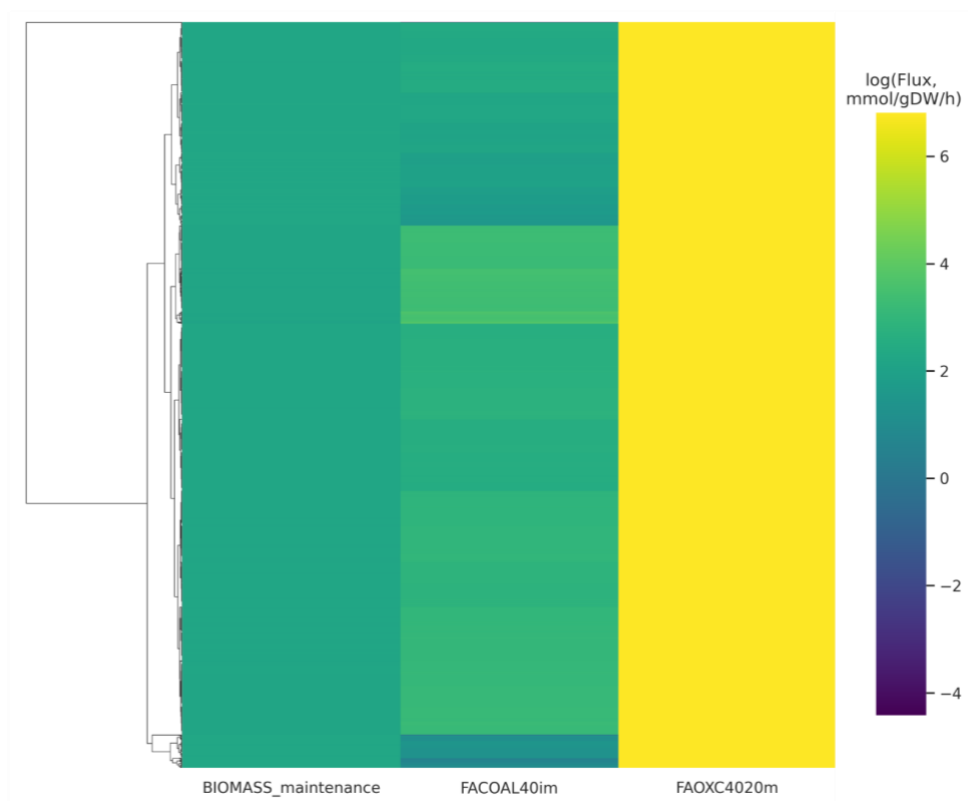


Fig 5.2 Objective function fluxes from human colonocyte model. Flux through the components of the objective function was observed for all samples. The biomass maintenance reaction, representing metabolic precursors required for growth, was required for proper functioning of models. Flux through the first step in the β -oxidation pathway, represented by the middle column of the heatmap, was dependent on predicted levels of microbially-produced butyrate.

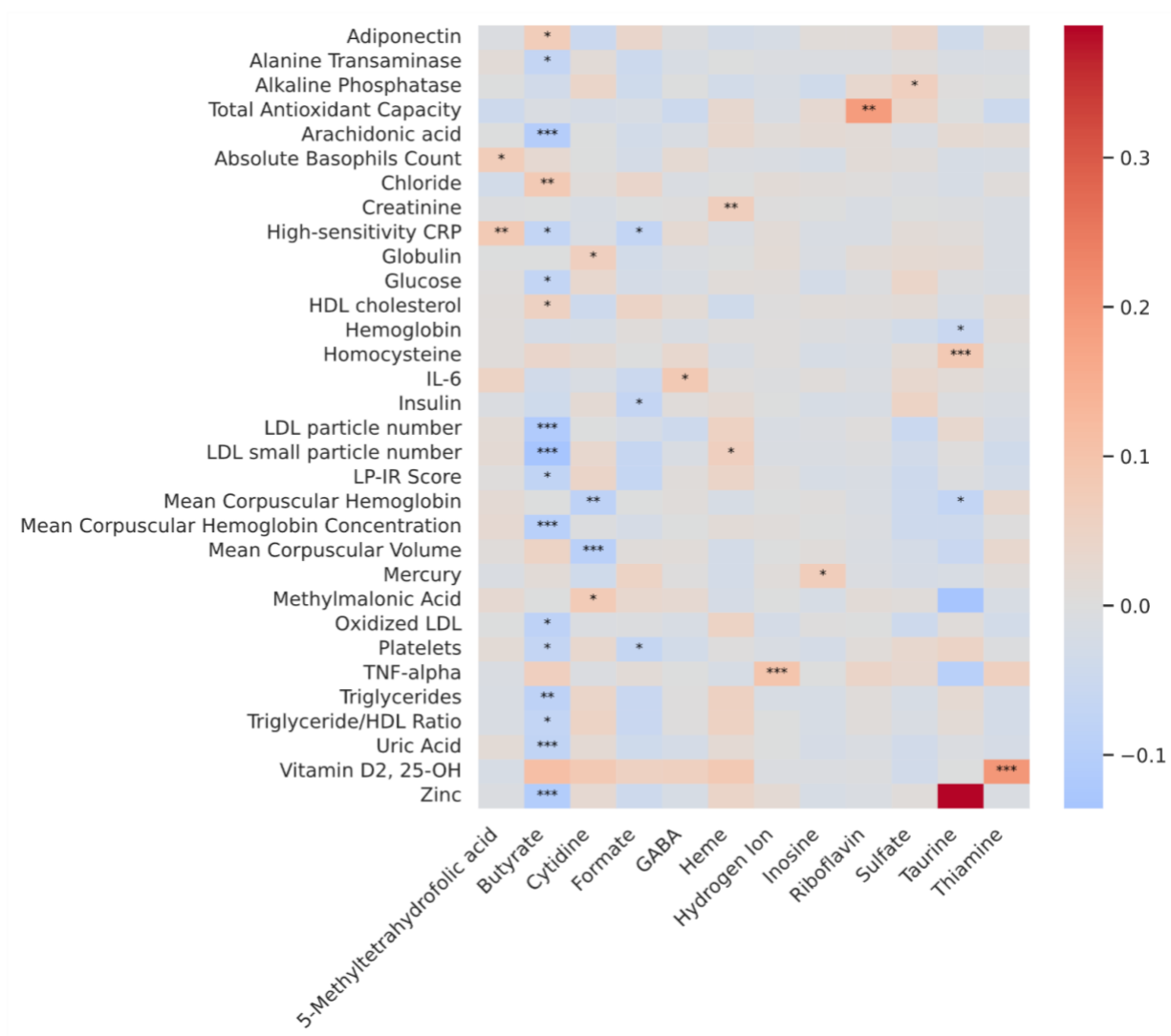


Figure 5.3 Associations between predicted colonic imports and blood-based clinical labs. Heatmap showing the 32 blood-based clinical labs showing significant association with predicted metabolic transports into the colonic epithelium, using MCMs for individuals in the Arivale cohort integrated with the human colonocyte metabolic model. Associations are based on a linear regression accounting for sex and age as covariates (FDR-corrected p-value, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, $N = 3,129$). Color encoding indicates standardized β -coefficient for individual variables in linear regression correcting for age and sex.

HMR_0156: but_c -> btcoa_c	Butyrate	Butanoyl-CoA
ACGPID: acgpail_hs_c -> ac_c	N-Acetyl-D-glucosaminylphosphatidylinositol	Acetate
ACLYSHYc: aclys_c -> ac_c	Acetyl-L-Lysine	Acetate
ACOA: accoa_c -> ac_c	Acetyl-CoA	Acetate
RE2649C: ppcoa_c -> ppa_c	Propionyl-CoA	Propionate
DCASULT: dchac_c -> dca3s_c	Deoxycholate	Deoxycholic acid 3-sulfate
LCASULT: HC02191_c -> lca3s_c	Lithocholate	Lithocholic acid 3-sulfate
UDCASULT: HC02194_c -> udca3s_c	Ursodeoxycholate	Ursodeoxycholic acid 3-sulfate
ADD: ade_c -> hxa_c	Adenine	Hypoxanthine
PUNP6: din_c -> hxa_c	Deoxyinosine	Hypoxanthine
PUNP5: ins_c -> hxa_c	Inosine	Hypoxanthine
SPMS: ametam_c -> spmd_c	S-Adenosylmethioninamine	Spermidine
SPRMS: spmd_c -> sprm_c	Spermidine	Spermine
HISDC: his_L_c -> hista_c	L-Histidine	Histamine extracellular
LTDC: trp_L_c -> trypta_c	L-Tryptophan	Tryptamine
r0403: trp_L_c -> 5htrp_c	L-Tryptophan	5-Hydroxy-L-tryptophan
5HLTDL : 5htrp_c -> srtn_c	5-Hydroxy-L-tryptophan	Serotonin
ENO: 2pg_c -> pep_c	D-Glycerate 2-phosphate	Phosphoenolpyruvate
FBA: fdp_c -> g3p_c	D-Fructose 1,6-bisphosphate	Glyceraldehyde 3-phosphate
GAPD: g3p_c -> 13dpg_c	Glyceraldehyde 3-phosphate	3-Phospho-D-glyceroyl phosphate
PGI: g6p_c -> f6p_c	D-Glucose 6-phosphate	D-Fructose 6-phosphate
PGK: 3pg_c -> 13dpg_c	3-Phospho-D-glycerate	3-Phospho-D-glyceroyl phosphate
PGM: 2pg_c -> 3pg_c	D-Glycerate 2-phosphate	3-Phospho-D-glycerate
PYK: pep_c -> pyr_c	Phosphoenolpyruvate	Pyruvate
TPI: dhap_c -> g3p_c	Dihydroxyacetone phosphate	Glyceraldehyde 3-phosphate
PYK2: pep_c -> pyr_c	Phosphoenolpyruvate	Pyruvate
r0354: glc_D_c -> g6p_c	D-Glucose	D-Glucose 6-phosphate
AGPOP: pep_c -> pyr_c	Phosphoenolpyruvate	Pyruvate
PYK6: pep_c -> pyr_c	Phosphoenolpyruvate	Pyruvate
ACYC: 13dpg_c -> 3pg_c	13dpg_c	3pg_c
r0173: lac_L_x -> pyr_x	L-Lactate	Pyruvate
HEX1: glc_D_c -> g6p_c	D-Glucose	D-Glucose 6-phosphate
LDH_L: lac_L_c -> pyr_c	L-Lactate	Pyruvate
ACONTm: cit_m -> icit_m	Citrate	Isocitrate
AKGDm: akc_m -> succoa_m	2-Oxoglutarate	Succinyl-CoA
CSm: oaa_m -> cit_m	Oxaloacetate	Citrate
FUMm: fum_m -> mal_L_m	Fumarate	L-Malate
ICDHxm: icit_m -> akc_m	Isocitrate	2-Oxoglutarate
MDHm: mal_L_m -> oaa_m	L-Malate	Oxaloacetate
SUCD1m: succ_m -> fum_m	Succinate	Fumarate
SUCOAS1m: succ_m -> succoa_m	Succinate	Succinyl-CoA
r0317: cit_m -> HC00342_m	Citrate	Cis-aconitate(3-); (1Z)-prop-1-ene-1,2,3-tricarboxylat
r0423: icit_m -> HC01434_m	Isocitrate	Oxalosuccinate(3-)
r0426: icit_m -> HC00342_m	Isocitrate	Cis-aconitate(3-); (1Z)-prop-1-ene-1,2,3-tricarboxylat
FADH2ETC: q10_m -> q10h2_m	Ubiquinone-10	Ubiquinol-10
G3PDm: glyc3p_m -> dhap_m	Glycerol 3-phosphate	Dihydroxyacetone phosphate
SALMCOM: amet_c -> ahcys_c	S-Adenosyl-L-methionine	S-Adenosyl-L-homocysteine
SALMCOM2: amet_c -> ahcys_c	S-Adenosyl-L-methionine	S-Adenosyl-L-homocysteine
RE2235R: estrone_r -> C05300_r	Estrone cytosol	16alpha-Hydroxyestrone
UGT1A3r: andrstn_r -> andrstnglc_r	Andrstn c	Androsterone glucuronide
O16G1e: glygn4_e -> glygn5_e	Glycogen, structure 4	Glycogen, structure 5
MEVK1: mev_R_c -> 5pmev_c	R Mevalonate	R 5 Phosphomevalonate
HISTASE: hista_c -> im4act_c	Histamine extracellular	Imidazole-4-acetaldehyde

Table 5.1 Metabolic tasks for colonocyte model

BioProject	Series Accession	Year	Number of health control colon epithelium samples	Disease type	Sample type	Experiment type
PRJNA173220	GSE40220	2013	3	Health	biopsy	Microarray
PRJNA509201	GSE123553	2018	8	Health	organoid	Microarray
PRJNA849958	GSE206285	2022	18	UC	biopsy	Microarray
PRJNA273501	GSE64392	2015	15	CRC	organoid	Microarray
PRJNA151571	GSE35279	2013	5	CRC	biopsy	Microarray
PRJNA291533	GSE71571	2015	44	Health	biopsy	Mircoarray
PRJNA788129	GSE190705	2022	3	Health	organoid	RNAseq
PRJNA787886	GSE190634	2022	4	Health	organoid	RNAseq
PRJNA494560	GSE120795	2018	6	Health	biopsy	RNAseq
PRJNA742380	GSE179130	2021	3	Health	organoid	RNAseq
PRJNA742364	GSE179128	2022	6	CD, UC	biopsy	RNAseq
PRJNA723260	GSE172404	2021	3	UC	organoid	RNAseq
PRJNA899901	GSE217663	2022	3	Health	organoid	RNAseq
PRJNA691157	GSE164541	2021	5	CRC	biopsy	RNAseq
PRJNA762103	GSE183853	2022	3	UC	biopsy	RNAseq
PRJNA525840	GSE127938	2019	3	Health	organoid	RNAseq
PRJNA558821	GSE135428	2021	3	Health	organoid	RNAseq
PRJNA310005	GSE77312	2016	20	Colon Cancer	biopsy	RNAseq
PRJNA565188	GSE137327	2019	9	Colon cancers	biopsy	RNAseq
PRJNA613086	GSE147133	2020	6	Health	organoid	RNAseq
PRJNA218851	GSE50760	2014	18	CRC	biopsy	RNAseq

Table 5.2 Data inclusion for colonocyte reconstruction

Reaction ID	Formula
BIOMASS_maintenance	0.50563 ala_L[c] + 0.35926 arg_L[c] + 0.27942 asn_L[c] + 0.35261 asp_L[c] + 20.7045 atp[c] + 0.020401 chsterol[c] + 0.011658 clpn_hs[c] + 0.039036 ctp[c] + 0.046571 cys_L[c] + 0.27519 g6p[c] + 0.326 gln_L[c] + 0.38587 glu_L[c] + 0.53889 gly[c] + 0.036117 gtp[c] + 20.6508 h2o[c] + 0.12641 his_L[c] + 0.28608 ile_L[c] + 0.54554 leu_L[c] + 0.59211 lys_L[c] + 0.15302 met_L[c] + 0.023315 pail_hs[c] + 0.15446 pchol_hs[c] + 0.055374 pe_hs[c] + 0.002914 pglyc_hs[c] + 0.25947 phe_L[c] + 0.41248 pro_L[c] + 0.005829 ps_hs[c] + 0.39253 ser_L[c] + 0.017486 sphmyln_hs[c] + 0.31269 thr_L[c] + 0.013306 trp_L[c] + 0.15967 tyr_L[c] + 0.053446 utp[c] + 0.35261 val_L[c] --> 20.6508 adp[c] + 20.6508 h[c] + 20.6508 pi[c]
FACOAL40im	atp[m] + but[m] + coa[m] --> amp[m] + btcoa[m] + ppi[m]
FAOXC4020m	btcoa[m] + coa[m] + fad[m] + h2o[m] + nad[m] --> 2.0 accoa[m] + fadh2[m] + h[m] + nadh[m]

Table 5.3 Colonocyte model objective function

Chapter 6

Study Limitations, Future Directions, and Significance

6.1 Study Limitations

The work presented here has several limitations. First, we faced constraints due to the limited availability of high-quality fluxomic datasets for model validation, particularly evident in the small sample sizes in the fecal *ex vivo* studies. While existing fluxomic datasets are difficult to find, the Gibbons Lab is currently developing a very large dataset of *ex vivo* fluxomic data, which will enable further validation in future work. Second, our predictions relied on the availability of GEMs, sourced mainly from the publicly available AGORA model database ⁷¹. While AGORA models undergo some manual curation, many are not fully validated, and they are known to contain inaccuracies ¹⁰¹. However, GEM databases are rapidly expanding and undergoing further curation and validation. As ongoing GEM refinement leads to ever-improving understanding of microbial metabolism, predictions from MCMM-based methods will only continue to improve. A major limitation of the outlined project is the lack of individual-specific dietary information for most individuals, which reduces prediction accuracy. This is a notable

limitation, as the dietary input to the model is one of the primary variables impacting the final predictions. Specific dietary data can be very difficult to gather, as adherence to specific diets is difficult and dietary recall data is often limited. However, applying an incorrect medium to the modeling framework will lead to drastic deviations from true values of metabolic production. Fortunately, new work from Diener et al. (2024) aims to combat this shortcoming, utilizing fecal metagenomic sequencing data to estimate dietary intake, allowing direct determination of diet¹⁰². While this methodology is still being validated, it represents a significant step toward the improvement of MCMM-based modeling. Despite the limitations outlined here, MCMMs still explained an impressive percentage of the variance in microbial butyrate and propionate production (i.e. ~35-40%). The accuracy of these MCMMs should only increase with time, as databases and modeling methods improve.

6.2 Future Directions

The successful implementation of this modeling framework has indeed provided valuable insights into the metabolic capacity of the gut microbiome. Nevertheless, our exploration of this subject is far from complete. As alluded to in the Study Limitations section above, ongoing efforts in the Gibbons Lab involve the collection of a large volume of fluxomic data, employing methodologies similar to those utilized for the *ex vivo* validation of MCMMs. This expanded dataset holds the potential to facilitate more comprehensive validation of the modeling framework. Importantly, the inclusion of additional interventions and metabolomic measures from these datasets will enable us to further investigate the predictive capacity of MCMMs in estimating the production of a broader range of metabolites. For instance, trimethylamine (TMA) is a clinically relevant microbially-derived metabolite, prediction of which has yet to be fully

validated from metabolic modeling-based tools. Using these data, microbial production of this potentially harmful metabolite may be better understood and predicted.

One attractive avenue of research is the accurate representation of non-growth associated metabolic pathways within MCMMs. Acetate production, for instance, is not well predicted due to its role as an overflow metabolite ⁷⁵. However, work by Versluis et al. (2022) demonstrated that adding a fixed enzymatic constraint to FBA models of the gut microbiome can help to overcome this limitation and more accurately assess overflow metabolism ¹⁰³. In the future, applying this methodology to MCMM-based efforts could improve the prediction of acetate, and many other non-growth-associated metabolites.

Despite the successful construction and integration of a model of the human colonocyte model with community models of the gut microbiome (**Chapter 5**), more work is required to fully validate the efficacy of this model, beyond the scope of what is presented here. This will be a challenging task, as direct measurement of fluxes into and out of the colonic epithelium is an invasive and costly endeavor. However, *in vitro* co-cultures of microbes with differentiated colonocytes or organoid-based models may offer an easier methodology for collecting these data. Complete validation of a fully functional integrated model will ultimately prompt more work toward translational applications, including the design of dietary, prebiotic and probiotic interventions.

Another future direction for this modeling framework would be to leverage the newly published AGORA2 database of microbial GEMs to predict xenobiotic metabolism, specifically of pharmaceuticals ¹⁰⁴. This new database included metabolic degradation pathways for several drugs, including statins, one of the most prescribed medications worldwide ¹⁰⁵. Work in recent

years has uncovered that the gut microbiome can significantly modulate the efficacy of these drugs ¹⁰⁶. Individual-specific prediction of this effect through metabolic modeling could theoretically lead to advances in precision application of statins, or other pharmaceuticals. Unfortunately, metabolic models as currently published in AGORA2 were found to be unreliable (i.e. unreliable), and drug metabolism pathways were largely ineffective at predicting xenobiotic degradation. However, as these microbial GEMs continue to improve, this objective remains a top goal for future research, especially given integration with a human colonocyte model.

6.3 Significance

As research in the microbiome field becomes increasingly mainstream, the translation of key findings into clinical interventions has become a focus ¹⁰⁷. The paradigm of using food as medicine has gained increasing popularity in recent years to combat diet-related diseases, increasing interest in the field of precision nutrition ^{108,109}. Generally, precision nutrition aims to provide an answer to a difficult question: What does a healthy diet look like for a particular individual ¹⁰⁸? Nutrition is a fundamental pillar of health, and what we eat is one of the most prevalent lifestyle factors affecting overall health ^{108,110}. Indeed, it has been shown that a causal link exists between diet and the incidence of non-communicable disease, such as cardiovascular disease and cancer ¹¹¹. Individuals with diets low in fruits and vegetables, but high in sodium and red meat show an increased risk of mortality from heart disease, stroke and type 2 diabetes ¹¹². Despite the well-established link between unhealthy diet and transition into disease, strategies for improving public health through dietary intervention have had limited success ⁹⁴. While this effect is partially due to lack of adherence ^{113,114}, is it also a function of the one-size-fits-all paradigm through which wellness has often been viewed. Given the countless variables involved, including genetics, the microbiome, health status, physical activity and food availability, there is

no easy answer to this question that applies widely to a heterogeneous population, opening the door to novel precision nutrition and healthcare solutions.

Applying metabolic modeling to the gut microbiome helps to resolve one of the factors with which precision medicine is concerned; metabolic response of the microbiome to different dietary contexts. By applying a collection of diets to an individual model *in silico*, clinicians could examine personalized metabolic responses, and use these results to modify and select a diet with the most beneficial metabolic output. For instance, given an individual with high levels of systemic inflammation at risk for transition into cardiovascular disease, one could develop a personalized diet increasing production of anti-inflammatory SCFAs. Probiotic and prebiotic interventions, easily implemented into the modeling framework, could be tested in the same manner. It is important to note that the number of interventions that could be tested is limited only by the number of annotated metabolic pathways present in the model. When grounding predictions in a comprehensive mechanistic understanding of the systems, extensive sets of training data aren't necessary, thereby broadening the scope of potential interventions.

A commonly referenced aphorism in mathematical modeling states that “all models are wrong, but some are useful”. The models presented here fit into this paradigm. Each model contains inherent inaccuracies. Nevertheless, despite these imperfections, the results presented above demonstrate the utility of this methodology. When applied to precision nutrition, having a methodological approach that can predict outcomes, even with some margin of error, offers significant benefits. This is particularly evident when compared to alternative methods such as implementing one-size-fits-all diets with the hope that an individual responds or collecting massive amounts of training data for machine learning predictions. Ultimately, the integration of microbial-scale metabolic models with host metabolic models allows for robust, tractable, and

extensible predictions of how the microbiome responds to various prebiotic, probiotic, and dietary interventions.

References

1. Shreiner, A. B., Kao, J. Y. & Young, V. B. The gut microbiome in health and in disease. *Curr. Opin. Gastroenterol.* **31**, 69–75 (2015).
2. Rackerby, B., Van De Grift, D., Kim, J. H. & Park, S. H. Effects of Diet on Human Gut Microbiome and Subsequent Influence on Host Physiology and Metabolism. *Gut Microbiome and Its Impact on Health and Diseases* 63–84 Preprint at https://doi.org/10.1007/978-3-030-47384-6_3 (2020).
3. Oliphant, K. & Allen-Vercoe, E. Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. *Microbiome* **7**, 91 (2019).
4. Ogunrinola, G. A., Oyewale, J. O., Oshamika, O. O. & Olasehinde, G. I. The Human Microbiome and Its Impacts on Health. *Int. J. Microbiol.* **2020**, 8045646 (2020).
5. Graham, D. B. & Xavier, R. J. Conditioning of the immune system by the microbiome. *Trends Immunol.* **44**, 499–511 (2023).
6. Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* **9**, 313–323 (2009).
7. Viggiano, D. *et al.* Gut barrier in health and disease: focus on childhood. *Eur. Rev. Med. Pharmacol. Sci.* **19**, 1077–1085 (2015).
8. Jandhyala, S. M. *et al.* Role of the normal gut microbiota. *World J. Gastroenterol.* **21**, 8787–8803 (2015).
9. Peng, L., Li, Z.-R., Green, R. S., Holzman, I. R. & Lin, J. Butyrate enhances the intestinal barrier by facilitating tight junction assembly via activation of AMP-activated protein

- kinase in Caco-2 cell monolayers. *J. Nutr.* **139**, 1619–1625 (2009).
10. Donia, M. S. & Fischbach, M. A. HUMAN MICROBIOTA. Small molecules from the human microbiota. *Science* **349**, 1254766 (2015).
 11. Glotfelty, L. G., Wong, A. C. & Levy, M. Small molecules, big effects: microbial metabolites in intestinal immunity. *Am. J. Physiol. Gastrointest. Liver Physiol.* **318**, G907–G911 (2020).
 12. Steliou, K., Boosalis, M. S., Perrine, S. P., Sangerman, J. & Faller, D. V. Butyrate histone deacetylase inhibitors. *Biores. Open Access* **1**, 192–198 (2012).
 13. Bravo, J. A. *et al.* Ingestion of *Lactobacillus* strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16050–16055 (2011).
 14. Ríos-Covián, D. *et al.* Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. *Front. Microbiol.* **7**, 185 (2016).
 15. Otten, B. M. J., Sthijns, M. M. J. P. E. & Troost, F. J. A Combination of Acetate, Propionate, and Butyrate Increases Glucose Uptake in C2C12 Myotubes. *Nutrients* **15**, (2023).
 16. Hamer, H. M. *et al.* Review article: the role of butyrate on colonic function. *Aliment. Pharmacol. Ther.* **27**, 104–119 (2008).
 17. Hodgkinson, K. *et al.* Butyrate's role in human health and the current progress towards its clinical application to treat gastrointestinal disease. *Clin. Nutr.* **42**, 61–75 (2023).
 18. Donohoe, D. R. *et al.* The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell Metab.* **13**, 517–526 (2011).
 19. Arora, T. & Tremaroli, V. Therapeutic Potential of Butyrate for Treatment of Type 2

- Diabetes. *Front. Endocrinol.* **12**, 761834 (2021).
20. Peng, K. *et al.* Butyrate and obesity: Current research status and future prospect. *Front. Endocrinol.* **14**, 1098881 (2023).
 21. Kespohl, M. *et al.* The Microbial Metabolite Butyrate Induces Expression of Th1-Associated Factors in CD4⁺ T Cells. *Front. Immunol.* **8**, 1036 (2017).
 22. Ulven, T. Short-chain free fatty acid receptors FFA2/GPR43 and FFA3/GPR41 as new potential therapeutic targets. *Front. Endocrinol.* **3**, 111 (2012).
 23. Wu, X. *et al.* Effects of the intestinal microbial metabolite butyrate on the development of colorectal cancer. *J. Cancer* **9**, 2510–2517 (2018).
 24. Lazarova, D. L., Chiaro, C. & Bordonaro, M. Butyrate induced changes in Wnt-signaling specific gene expression in colorectal cancer cells. *BMC Res. Notes* **7**, 226 (2014).
 25. Zou, F. *et al.* Effects of short-chain fatty acids in inhibiting HDAC and activating p38 MAPK are critical for promoting B10 cell generation and function. *Cell Death Dis.* **12**, 582 (2021).
 26. Hosseini, E., Grootaert, C., Verstraete, W. & Van de Wiele, T. Propionate as a health-promoting microbial metabolite in the human gut. *Nutr. Rev.* **69**, 245–258 (2011).
 27. Weitkunat, K. *et al.* Importance of propionate for the repression of hepatic lipogenesis and improvement of insulin sensitivity in high-fat diet-induced obesity. *Mol. Nutr. Food Res.* **60**, 2611–2621 (2016).
 28. 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).
 29. 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).
30. Yamamura, R., Inoue, K. Y., Nishino, K. & Yamasaki, S. Intestinal and fecal pH in human health. *Frontiers in Microbiomes* **2**, (2023).
 31. Nogal, A., Valdes, A. M. & Menni, C. The role of short-chain fatty acids in the interplay between gut microbiota and diet in cardio-metabolic health. *Gut Microbes* **13**, 1–24 (2021).
 32. 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).
 33. Arifuzzaman, M. *et al.* Inulin fibre promotes microbiota-derived bile acids and type 2 inflammation. *Nature* **611**, 578–584 (2022).
 34. Armstrong, H. K. *et al.* Unfermented β -fructan Fibers Fuel Inflammation in Select Inflammatory Bowel Disease Patients. *Gastroenterology* **164**, 228–240 (2023).
 35. Rajakovich, L. J., Fu, B., Bollenbach, M. & Balskus, E. P. Elucidation of an anaerobic pathway for metabolism of l-carnitine-derived γ -butyrobetaine to trimethylamine in human gut bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
 36. Rath, S., Heidrich, B., Pieper, D. H. & Vital, M. Uncovering the trimethylamine-producing bacteria of the human gut microbiota. *Microbiome* **5**, 54 (2017).
 37. Gatarek, P. & Kaluzna-Czaplinska, J. Trimethylamine N-oxide (TMAO) in human health. *EXCLI J.* **20**, 301–319 (2021).
 38. Lee, Y. *et al.* Longitudinal Plasma Measures of Trimethylamine N-Oxide and Risk of Atherosclerotic Cardiovascular Disease Events in Community-Based Older Adults. *J. Am. Heart Assoc.* **10**, e020646 (2021).
 39. Fretts, A. M. *et al.* Association of Trimethylamine N-Oxide and Metabolites With Mortality

- in Older Adults. *JAMA Netw Open* **5**, e2213242 (2022).
40. Zeng, H., Umar, S., Rust, B., Lazaro, D. & Bordonaro, M. Secondary Bile Acids and Short Chain Fatty Acids in the Colon: A Focus on Colonic Microbiome, Cell Proliferation, Inflammation, and Cancer. *International Journal of Molecular Sciences* vol. 20 1214 Preprint at <https://doi.org/10.3390/ijms20051214> (2019).
 41. Larabi, A. B., Masson, H. L. P. & Bäuml, A. J. Bile acids as modulators of gut microbiota composition and function. *Gut Microbes* **15**, 2172671 (2023).
 42. Ye, X. *et al.* Dual Role of Indoles Derived From Intestinal Microbiota on Human Health. *Front. Immunol.* **13**, 903526 (2022).
 43. Jaglin, M. *et al.* Indole, a Signaling Molecule Produced by the Gut Microbiota, Negatively Impacts Emotional Behaviors in Rats. *Front. Neurosci.* **12**, 216 (2018).
 44. Wan, Y., Li, Y., Yan, C., Yan, M. & Tang, Z. Indole: A privileged scaffold for the design of anti-cancer agents. *Eur. J. Med. Chem.* **183**, 111691 (2019).
 45. 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).
 46. Berry, S. E. *et al.* Human postprandial responses to food and potential for precision nutrition. *Nat. Med.* **26**, 964–973 (2020).
 47. Zeevi, D. *et al.* Personalized Nutrition by Prediction of Glycemic Responses. *Cell* **163**, 1079–1094 (2015).
 48. Gupta, V. K. *et al.* A predictive index for health status using species-level gut microbiome profiling. *Nat. Commun.* **11**, 4635 (2020).
 49. Lian, X., Yang, S., Li, H., Fu, C. & Zhang, Z. Machine-Learning-Based Predictor of

- Human-Bacteria Protein-Protein Interactions by Incorporating Comprehensive Host-Network Properties. *J. Proteome Res.* **18**, 2195–2205 (2019).
50. Price, N. D., Reed, J. L. & Palsson, B. Ø. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* **2**, 886–897 (2004).
 51. Jamshidi, N. & Palsson, B. O. Using in silico models to simulate dual perturbation experiments: procedure development and interpretation of outcomes. *BMC Syst. Biol.* **3**, 44 (2009).
 52. Feist, A. M. & Palsson, B. Ø. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat. Biotechnol.* **26**, 659–667 (2008).
 53. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
 54. Raman, K. & Chandra, N. Flux balance analysis of biological systems: applications and challenges. *Brief. Bioinform.* **10**, 435–449 (2009).
 55. Segrè, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 15112–15117 (2002).
 56. Becker, S. A. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat. Protoc.* **2**, 727–738 (2007).
 57. Shoaie, S. *et al.* Quantifying Diet-Induced Metabolic Changes of the Human Gut Microbiome. *Cell Metab.* **22**, 320–331 (2015).
 58. Diener, C., Gibbons, S. M. & Resendis-Antonio, O. MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. *mSystems* **5**, (2020).
 59. Marinos, G., Kaleta, C. & Waschina, S. Defining the nutritional input for genome-scale metabolic models: A roadmap. *PLoS One* **15**, e0236890 (2020).

60. Schultz, A. & Qutub, A. A. Reconstruction of Tissue-Specific Metabolic Networks Using CORDA. *PLoS Comput. Biol.* **12**, e1004808 (2016).
61. Jerby, L., Shlomi, T. & Ruppin, E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol. Syst. Biol.* **6**, 401 (2010).
62. Wang, Y., Eddy, J. A. & Price, N. D. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst. Biol.* **6**, 153 (2012).
63. Brunk, E. *et al.* Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* **36**, 272–281 (2018).
64. Salvi, P. S. & Cowles, R. A. Butyrate and the Intestinal Epithelium: Modulation of Proliferation and Inflammation in Homeostasis and Disease. *Cells* **10**, (2021).
65. Agren, R. *et al.* Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol. Syst. Biol.* **10**, 721 (2014).
66. Zur, H., Ruppin, E. & Shlomi, T. iMAT: an integrative metabolic analysis tool. *Bioinformatics* **26**, 3140–3142 (2010).
67. Bordbar, A. *et al.* A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. *BMC Syst. Biol.* **5**, 180 (2011).
68. Martins Conde, P., Pfau, T., Pires Pacheco, M. & Sauter, T. A dynamic multi-tissue model to study human metabolism. *NPJ Syst Biol Appl* **7**, 5 (2021).
69. Thiele, I. *et al.* Personalized whole-body models integrate metabolism, physiology, and the gut microbiome. *Mol. Syst. Biol.* **16**, e8982 (2020).
70. Quinn-Bohmann, N. *et al.* Microbial community-scale metabolic modeling predicts personalized short chain fatty acid production profiles in the human gut. *bioRxiv* (2023) doi:10.1101/2023.02.28.530516.

71. Magnúsdóttir, S. *et al.* Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).
72. Noronha, A. *et al.* The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res.* **47**, D614–D624 (2019).
73. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
74. Clark, R. L. *et al.* Design of synthetic human gut microbiome assembly and butyrate production. *Nat. Commun.* **12**, 3254 (2021).
75. Shimada, T. *et al.* Acetate overflow metabolism regulates a major metabolic shift after glucose depletion in *Escherichia coli*. *FEBS Lett.* **595**, 2047–2056 (2021).
76. 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).
77. Wastyk, H. C. *et al.* Gut-microbiota-targeted diets modulate human immune status. *Cell* **184**, 4137–4153.e14 (2021).
78. 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).
79. Elmadfa, I. Österreichischer Ernährungsbericht 2012. **1**, (2012).
80. Seabold, S. & Perktold, J. Statsmodels: Econometric and statistical modeling with python. in *Proceedings of the 9th Python in Science Conference (SciPy, 2010)*. doi:10.25080/majora-92bf1922-011.
81. Tough, I. R., Forbes, S. & Cox, H. M. Signaling of free fatty acid receptors 2 and 3 differs

- in colonic mucosa following selective agonism or coagonism by luminal propionate. *Neurogastroenterol. Motil.* **30**, e13454 (2018).
82. Nehring, S. M., Goyal, A. & Patel, B. C. *C Reactive Protein*. (StatPearls Publishing, 2023).
 83. Venugopal, S. K., Anoruo, M. & Jialal, I. *Biochemistry, Low Density Lipoprotein*. (StatPearls Publishing, 2023).
 84. Jama, H. A. *et al.* Prebiotic intervention with HAMSAB in untreated essential hypertensive patients assessed in a phase II randomized trial. *Nature Cardiovascular Research* **2**, 35–43 (2023).
 85. Feingold, K. R. & Grunfeld, C. *The Effect of Inflammation and Infection on Lipids and Lipoproteins*. (MDText.com, Inc., 2022).
 86. Li, Q., Zhou, Y., Huang, C., Li, W. & Chen, L. A novel diagnostic algorithm to predict significant liver inflammation in chronic hepatitis B virus infection patients with detectable HBV DNA and persistently normal alanine transaminase. *Sci. Rep.* **8**, 15449 (2018).
 87. Kimura, Y., Tsukui, D. & Kono, H. Uric Acid in Inflammation and the Pathogenesis of Atherosclerosis. *Int. J. Mol. Sci.* **22**, (2021).
 88. Higgins, A. J. & Lees, P. The acute inflammatory process, arachidonic acid metabolism and the mode of action of anti-inflammatory drugs. *Equine Vet. J.* **16**, 163–175 (1984).
 89. Castelli, W. P. Cholesterol and lipids in the risk of coronary artery disease--the Framingham Heart Study. *Can. J. Cardiol.* **4 Suppl A**, 5A–10A (1988).
 90. Barter Philip *et al.* HDL Cholesterol, Very Low Levels of LDL Cholesterol, and Cardiovascular Events. *N. Engl. J. Med.* **357**, 1301–1310.
 91. Nguyen, T. M. D. Adiponectin: Role in Physiology and Pathophysiology. *Int. J. Prev. Med.* **11**, 136 (2020).

92. Resta-Lenert, S., Truong, F., Barrett, K. E. & Eckmann, L. Inhibition of epithelial chloride secretion by butyrate: role of reduced adenylyl cyclase expression and activity. *Am. J. Physiol. Cell Physiol.* **281**, C1837–49 (2001).
93. Manor, O. *et al.* Health and disease markers correlate with gut microbiome composition across thousands of people. *Nat. Commun.* **11**, 5206 (2020).
94. Ordovas, J. M., Ferguson, L. R., Tai, E. S. & Mathers, J. C. Personalised nutrition and health. *BMJ* **361**, bmj.k2173 (2018).
95. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
96. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
97. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104 (2017).
98. Hills, A. P., Byrne, N. M., Lindstrom, R. & Hill, J. O. ‘Small changes’ to diet and physical activity behaviors for weight management. *Obes. Facts* **6**, 228–238 (2013).
99. Ashoori, M. & Saedisomeolia, A. Riboflavin (vitamin B₂) and oxidative stress: a review. *Br. J. Nutr.* **111**, 1985–1991 (2014).
100. Hughes, E. R. *et al.* Microbial Respiration and Formate Oxidation as Metabolic Signatures of Inflammation-Associated Dysbiosis. *Cell Host Microbe* **21**, 208–219 (2017).
101. Babaei, P., Shoaie, S., Ji, B. & Nielsen, J. Challenges in modeling the human gut microbiome. *Nat. Biotechnol.* **36**, 682–686 (2018).
102. Diener, C. & Gibbons, S. M. Metagenomic estimation of dietary intake from human stool. *bioRxiv* 2024.02.02.578701 (2024) doi:10.1101/2024.02.02.578701.

103. Versluis, D. M. *et al.* A Multiscale Spatiotemporal Model Including a Switch from Aerobic to Anaerobic Metabolism Reproduces Succession in the Early Infant Gut Microbiota. *mSystems* **7**, e0044622 (2022).
104. Heinken, A. *et al.* AGORA2: Large scale reconstruction of the microbiome highlights widespread drug-metabolising capacities. *bioRxiv* 2020.11.09.375451 (2020)
doi:10.1101/2020.11.09.375451.
105. Salami, J. A. *et al.* National Trends in Statin Use and Expenditures in the US Adult Population From 2002 to 2013: Insights From the Medical Expenditure Panel Survey. *JAMA Cardiol* **2**, 56–65 (2017).
106. Wilmanski, T. *et al.* Heterogeneity in statin responses explained by variation in the human gut microbiome. *Med (N Y)* **3**, 388–405.e6 (2022).
107. Lynch, S. V., Ng, S. C., Shanahan, F. & Tilg, H. Translating the gut microbiome: ready for the clinic? *Nat. Rev. Gastroenterol. Hepatol.* **16**, 656–661 (2019).
108. Berciano, S. *et al.* Precision nutrition: Maintaining scientific integrity while realizing market potential. *Front Nutr* **9**, 979665 (2022).
109. Downer, S., Berkowitz, S. A., Harlan, T. S., Olstad, D. L. & Mozaffarian, D. Food is medicine: actions to integrate food and nutrition into healthcare. *BMJ* **369**, m2482 (2020).
110. Ordovas, J. M. & Corella, D. Nutritional genomics. *Annu. Rev. Genomics Hum. Genet.* **5**, 71–118 (2004).
111. Iriti, M., Varoni, E. M. & Vitalini, S. Healthy Diets and Modifiable Risk Factors for Non-Communicable Diseases-The European Perspective. *Foods* **9**, (2020).
112. Micha, R. *et al.* Association Between Dietary Factors and Mortality From Heart Disease, Stroke, and Type 2 Diabetes in the United States. *JAMA* **317**, 912–924 (2017).

113. Martin, K. A., Bowen, D. J., Dunbar-Jacob, J. & Perri, M. G. Who will adhere? Key issues in the study and prediction of adherence in randomized controlled trials. *Control. Clin. Trials* **21**, 195S–9S (2000).
114. Leme, A. C. B., Hou, S., Fisberg, R. M., Fisberg, M. & Haines, J. Adherence to Food-Based Dietary Guidelines: A Systemic Review of High-Income and Low- and Middle-Income Countries. *Nutrients* **13**, (2021).