RESEARCH ARTICLE

Functional heterogeneity in the fermentation capabilities of the healthy human gut microbiota

Thomas Gurry[1,2,3,4☉], **Le Thanh Tu Nguyen**[1,2,3☉], **Xiaoqian Yu**[1,5], **Eric J. Alm**[1,2,3]*

1 Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, United States of America, **2** Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology, Cambridge, MA, United States of America, **3** The Broad Institute of MIT and Harvard, Cambridge, MA, United States of America, **4** Institute for Pharmaceutical Sciences of Western Switzerland, University of Geneva, Geneva, Switzerland, **5** Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, United States of America

☉ These authors contributed equally to this work.
* ejalm@mit.edu

Abstract

The human gut microbiota is known for its highly heterogeneous composition across different individuals. However, relatively little is known about functional differences in its ability to ferment complex polysaccharides. Through *ex vivo* measurements from healthy human donors, we show that individuals vary markedly in their microbial metabolic phenotypes (MMPs), mirroring differences in their microbiota composition, and resulting in the production of different quantities and proportions of Short Chain Fatty Acids (SCFAs) from the same inputs. We also show that aspects of these MMPs can be predicted from composition using 16S rRNA sequencing. From experiments performed using the same dietary fibers *in vivo*, we demonstrate that an ingested bolus of fiber is almost entirely consumed by the microbiota upon passage. We leverage our *ex vivo* data to construct a model of SCFA production and absorption *in vivo*, and argue that inter-individual differences in quantities of absorbed SCFA are directly related to differences in production. Though *in vivo* studies are required to confirm these data in the context of the gut, in addition to *in vivo* read outs of SCFAs produced in response to specific fiber spike-ins, these data suggest that optimizing SCFA production in a given individual through targeted fiber supplementation requires quantitative understanding of their MMP.

Introduction

The symbiotic relationship between host and gut microbiota is intimately related to host diet. For example, ruminants derive the majority of their caloric intake from the microbial fermentation of indigestible polysaccharides in their diet. In humans, only a fraction of total caloric intake is derived from fermentation of dietary fibers and other Microbiota Accessible Carbohydrates (MACs), but the resulting metabolites also play other important physiological roles [1–3]. In particular, the Short Chain Fatty Acids (SCFAs) acetate, propionate and butyrate,

which are the major by-products of the microbial fermentation of dietary fibers in the gut, exert a number of forces on the host's physiology. While it has long been known that butyrate serves as the dominant energy source for colonocytes [4, 5], and is therefore critically important to maintaining a healthy gut, it has been shown more recently that imbalances in SCFA production can be associated with disease. In the case of microbial acetate, increased production and turnover have been shown to activate glucose-stimulated insulin secretion via a gut-brain axis mediated process, which can result in insulin resistance and subsequent obesity, in addition to increasing a patient's risk of developing Type 2 diabetes [6]. In contrast, propionate and butyrate activate gluconeogenesis, with beneficial effects to host metabolism and the regulation of plasma glucose levels [7–9]. Indeed, increased butyrate production has been causally linked to improved insulin response after an oral glucose-tolerance test, whereas abnormalities in propionate production were associated with an increased risk of Type-2 Diabetes [10].

Many of these effects are mediated by short chain fatty acid receptors in the gut epithelium. However, a critical property of colonic SCFAs is their activity as histone deacetylase (HDAC) inhibitors [1]. This ability to regulate gene expression in host cells has associated SCFAs with a growing list of clinical indications. Butyrate, in particular, has been proposed to exert anti-inflammatory pressure on the host's immune system through several mechanisms, which include differentiation of regulatory and IL-10-producing T cells, down regulation of IL-6 production, pro-inflammatory T cell apoptosis, and suppression of IFN-γ-mediated inflammation in the colonic epithelium [11]. Consistent with these findings, when Clostridia strains were computationally ranked by regulatory T cell induction capability, the highest ranked strains were predicted to produce significantly higher butyrate than the lower ranked strains [12]. These data complement measured associations between Inflammatory Bowel Disease (IBD) and depletions in butyrate-producing organisms (e.g. *Faecalibacterium prausnitzii*) [13] to indicate that SCFAs likely play important roles in the disease etiology. Moreover, the fact that SCFAs produced by the microbiota in the colon can be absorbed into the blood stream suggests that their effects on gene expression may transcend the gut and affect distal tissues in ways that currently poorly understood. It is therefore of significant clinical interest to improve our quantitative understanding of SCFA production in the gut, in order that it may be modulated towards a particular clinical outcome.

Several different biochemical fermentation pathways in the gut metagenome result in SCFAs as a final product. These pathways begin with the hydrolysis of complex dietary polysaccharides to their constituent oligo- and monosaccharides by members of the microbiota encoding the appropriate Polysaccharide Utilization Loci (PULs) and Glycoside Hydrolases (GHs), which differ widely between different bacterial species [14] and even strains within a species [15]. These monosaccharides can then be fermented by a number of pathways to ultimately result in acetate, propionate or butyrate. Importantly, there is cross-talk between different pathways: for example, the most prevalent butyrate-producing pathway in the human gut involves the enzyme butyryl-CoA: acetate-CoA transferase [16], which exchanges a butyrate moiety for an acetate moiety on the CoA molecule and releases free butyrate. A similar enzyme exists for propionate in one of its three main production pathways [17]. Thus, the pool of available acetate affects the production of both propionate and butyrate. Other products of bacterial fermentation can also serve as intermediates in SCFA production: lactate, in particular, can act as a substrate for further fermentation into acetate, propionate and butyrate. Moreover, the SCFA-producing organisms may not be able to directly ferment a specific polysaccharide themselves; instead, they rely on the preliminary degradation of these dietary fibers into hexoses and pentoses by other bacteria encoding these carbohydrate active enzymes. This gives rise to a host of cooperative microbial networks which act in concert to produce the overall SCFA profile present in an individual's colon. For example, lactic acid-utilizing, butyrate-

producing bacteria (e.g. *Eubacterium hallii* and *Anaerostipes caccae*) depend on the presence of lactic acid-producing bacteria such as *Bifidobacterium adolescentis* to produce butyrate [18]. Thus, the combination of dietary inputs (which specific dietary fibers, in which quantities) and the composition of an individual's microbiota together dictate what ratios and absolute quantities of SCFAs are produced in their gut, with potentially important effects on physiology. In this study, we sought to gain a better understanding of the variation in fermentation capabilities in the microbiota of different individuals within the healthy human population. We present an *ex vivo* framework for measuring the production of SCFAs of an individual's stool microbiota in response to challenge with specific dietary fibers. Performing these experiments *ex vivo* allows us to quantify SCFA accumulation accurately and circumvents the technical difficulties associated with measurements of SCFA production *in vivo* (invasiveness) or SCFA measurements from raw stool upon passage (unknown extent of SCFA absorption by the gut epithelium).

Results

Measuring a microbial metabolic phenotype *ex vivo*

We first sought to obtain *ex vivo* measurements of SCFA production in response to different dietary polysaccharides, in order to quantify the degree of heterogeneity in the fermentation capabilities of the healthy population's gut microbiota. To do this, stool from 40 healthy human participants was homogenized into a slurry in anaerobic conditions and spiked with inulin, pectin or cellulose (cf. Methods). The slurry was then incubated over time and samples obtained at regular intervals in order to quantify SCFA concentration at each time point (Fig 1A). In order to determine the appropriate sampling frequency, we performed pilot experiments in which we analyzed the trajectory of each SCFA concentration over a 24h period. We found that only a subset of participants appeared to converge to a final SCFA concentration prior to the 24h timepoint, but that all participants exhibited a linear production rate in the 0-4h time window (Fig 1B). These data were in good agreement with concentrations of inulin measured from the stool over time using an inulin-specific ELISA assay: after 4h, a significant fraction of inulin substrate remained, but this was almost entirely consumed by 24h in five of the six participants tested (S3A Fig). Since we sought to perform the experiments in an environment that best mimicked conditions in the colon, we analyzed 16S rRNA data and determined that community structure or diversity were not significantly altered between timepoints 0 and 4h, with minor changes being observed between 4h and 24h (S4 and S5A Figs). Participants' slurry did not resemble each other more at 4h than they did at 0h, but did increase in similarity by the 24h timepoint (S6 Fig). Samples noticeably clustered by participant rather than by timepoint or condition (S5B Fig). Moreover, changes in pH due to an accumulation of acidic SCFAs was limited to a drop from approximately neutral to 5.5 during the first four hours (S7 Fig). Together, these data informed our decision to use the linear production regime observed between 0-4h as the most appropriate period in which to measure SCFA production *ex vivo* as a proxy for SCFA production *in vivo*. As a result, we chose to measure two biological replicates for the timepoints 0, 2h and 4h, and compute production rates between timepoints 2h and 4h. There was generally good agreement between biological replicates, with an overall coefficient of variation of 23.5% and no discernible bias in any specific SCFA or at any specific timepoint (S1 Fig). Linear regressions between the sample's Bristol stool scale and total SCFA concentration produced, as well as between the C_T values from the qPCR amplification of 16S rRNA from the stool, did not suggest any apparent artifact induced by stool consistency or density, arguing that a correction factor for stool consistency or microbial load was not required and samples of various consistencies could be compared to one another (S8 Fig).
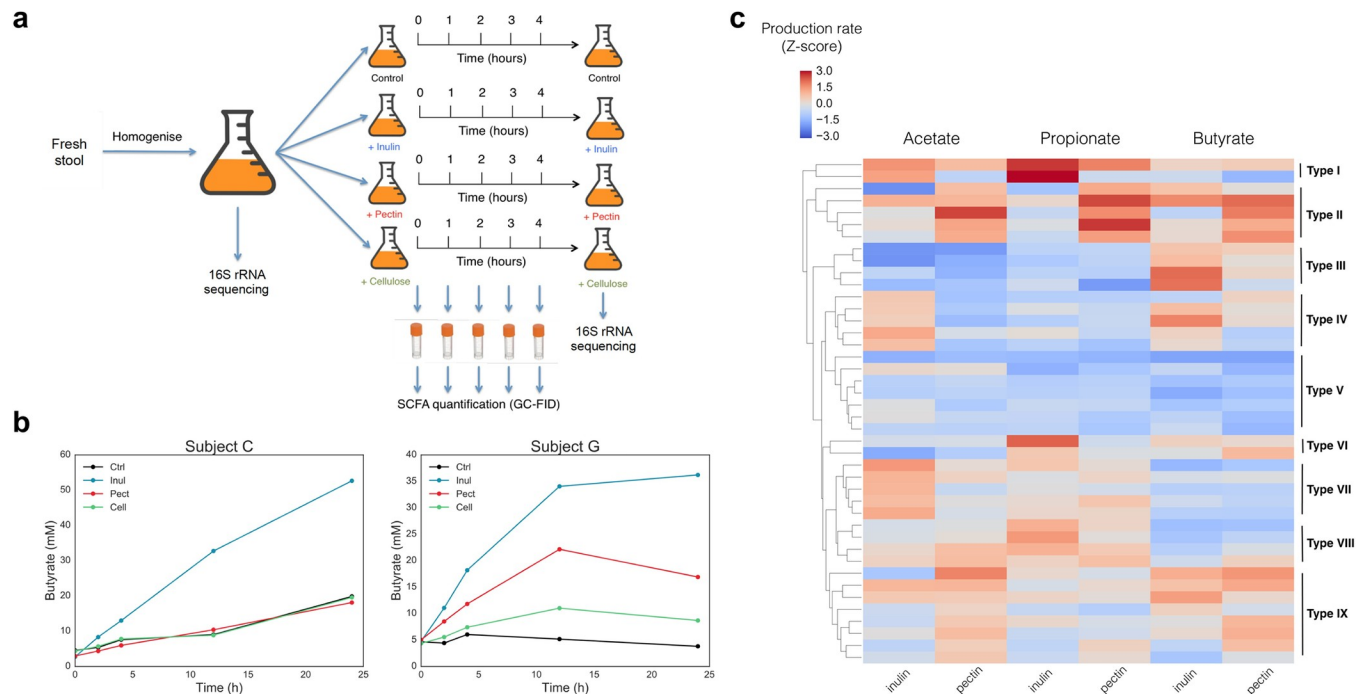
Fig 1. Microbial metabolic phenotype varies significantly in the healthy population. (a) Schematic of assay setup and sampling frequency. (b) 24h time traces of butyrate concentration over time in response to inulin, pectin, cellulose and control, in two different participants. Participant C only produces butyrate from inulin, while participant G produces it from pectin as well. Both participants have linear production regimes in the 0-4h window used to calculate production rate. (c) Acetate, propionate and butyrate production rates from inulin and pectin in different participants, presented as Z-scores computed across all participants. Each row represents measurements from a single sample. SCFA production rates were measured *ex vivo* in mM/h for each participant in response to each condition. Production rates were computed between timepoints 2h and 4h for each condition and production rates from the control condition (no spike-in) were subtracted. Cellulose timepoints were indistinguishable from control and therefore were not presented.

We found that participants differed greatly in their SCFA production profiles (Fig 1C, here Z-scores were chosen to facilitate visualization of the difference between participants with different fibers and different SCFAs, since each of these are on separate scales). From the same dietary fiber input, the different microbiotas produced significantly different quantities and ratios of SCFAs. We define the resulting SCFA production rates in response to different dietary fibers, quantified as standardized scores compared to the other participants in the dataset, as an individual's Microbial Metabolic Phenotype (MMP). Hierarchical clustering of MMPs (using UPGMA algorithm) indicated discernible groups: for example, individuals with MMP Type I were strong producers of propionate from inulin; in contrast, participants with MMP Type II were strong producers of propionate from pectin (Fig 1C). Thus, improving an individual's production of a given SCFA will not necessarily rely on the same polysaccharide to reach the same effect; put differently, the same polysaccharide will have different effects in different individuals depending on their MMP. These data argue that there is significant heterogeneity in the healthy human population when it comes to functional degradation of fibers in the gut and the SCFAs produced, but that MMPs cluster into discernible types, which can be used to guide future dietary interventions.

Predicting microbial metabolic phenotype from community composition

We then asked the question whether we could predict a participant's MMP from community composition alone, defined here as the relative abundances of 97% *de novo* OTUs obtained from 16S rRNA sequencing of the stool prior to incubation with the different fibers. We
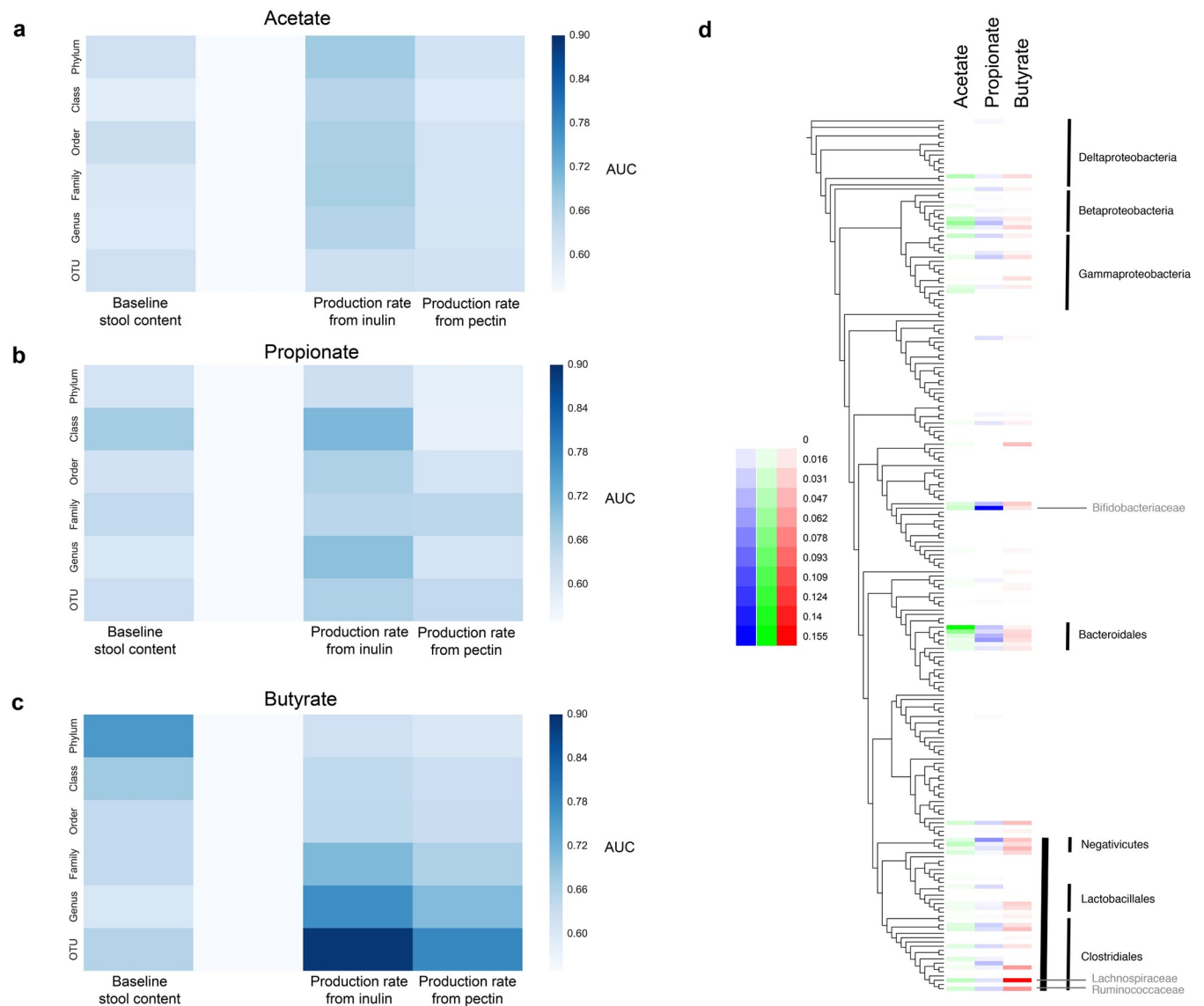
Fig 2. SCFA production capacity can be predicted from individual bacterial OTUs. (a) AUC values for different RFCs trained either to predict high or low SCFA content in stool at baseline, or high or low SCFA production rate *ex vivo* in response to specific dietary fibers. High and low production is defined according to the z-score across all participants in the study. (b) Relationship between propionate and butyrate production rates, and the relative abundance of an unassigned OTU of the Lachnospiraceae family, showing a relationship between its relative abundance and butyrate production in response to inulin, specifically. (c) Similar relationship but specific to the relative abundance of a *Prevotella copri* OTU and propionate production in response to inulin. (d) Feature importances from RFCs trained to predict high or low SCFA production in response to inulin from bacterial family abundances.

https://doi.org/10.1371/journal.pone.0254004.g002

trained Random Forest Classifiers (RFCs) to predict whether a given microbiota had a high or low production rate of a given SCFA in response to a given fiber, defined by a production rate z-score of greater than or equal to 0, or less than 0, respectively. Performance varied by SCFA, with the highest accuracies obtained in predicting butyrate production in response to inulin (AUC = 0.87) and pectin (AUC = 0.79) (Fig 2A).

We also tested whether straight stool SCFA contents could be predicted from 16S rRNA sequencing and found moderate predictive power for acetate and butyrate (AUC = 0.76 and AUC = 0.73, respectively). Though these data indicated that community

composition is somewhat predictive of the resulting stool SCFA contents, manual inspection of specific OTU features that were highly ranked in terms of importance for the inulin- and pectin-specific RFCs found these to often only be associated with SCFA production in response to that specific fiber. For example, the relative abundance of an unassigned Lachnospiraceae OTU was only associated with butyrate production in response to inulin (Fig 2B). In contrast, a specific *Prevotella copri* OTU only appeared to be associated with propionate production from inulin (Fig 2C). These data are consistent with the fact that members of the *Prevotella* genus are known propionate producers [19] and vary in their polysaccharide degradation capabilities [15]. Thus, a Type I MMP (Fig 1C) is associated with a high relative abundance of the latter *P. copri* OTU. Similarly, known butyrate-producing families (e.g. Lachnospiraceae and Ruminococcaceae) are ranked highly in importance when RFCs for high/low SCFA production in response to inulin are trained on 16S rRNA data collapsed at the family level (Fig 2D). Though more data and whole genome characterizations are required to demonstrate these relationships rigorously, these results suggest that individual OTUs are predictive of SCFA production capability from specific polysaccharides, likely due to their specific polysaccharide degradation machinery and internal fermentation pathways.

Stability of an individual's MMP through time

While it is known that the gut microbiota of individuals can be relatively stable for long periods of time in the absence of large perturbations [20], it is unclear whether an individual's MMP will also be similarly stable through time. We therefore repeated the experiment for eight participants at timepoints separated by at least 6 months (Fig 3A). Though some variability between timepoints was observed, the extrema of each individual's MMP were generally preserved (Fig 3B). A Fisher test on the contingency table resulting from pairwise comparison of each SCFA:fiber pair at the two timepoints for all individuals indicated that this stability was statistically significant ($p = 0.003$; two-tailed Fisher test). These results were consistent with the fact that MMPs are associated with the relative abundance of specific members of the microbiota: an individual with a high relative abundance of the aforementioned *P. copri* OTU is likely to retain a high relative abundance of this OTU during the six month period between timepoints compared to the general population, thus retaining their ability to produce high levels of propionate from inulin through time.

In vivo model of SCFA production

In order to better understand microbial SCFA production *in vivo*, we sought to develop a quantitative model of the process (Fig 4A). In a given participant, when a quantity of dietary fiber, *[F]*, is consumed, it is fermented into acetate, propionate and butyrate with rates that are functions of their specific microbiota. As defined, a person's MMP is the aggregate total response of an individual's fecal microbiota to a challenge with specific dietary fibers. The process of producing SCFAs from fiber inputs requires several steps (S9B Fig). The first step involves breaking the fiber/polysaccharide (*F*) into smaller oligo- and monosaccharides (*O*), usually through a hydrolysis reaction encoded by extracellular enzymes [14]; the second step consists of fermenting *O* into a reduced biochemical species *P* (e.g. lactate, acetate, pyruvate), which can act as a substrate for the third step, the final fermentation reaction that leads to the final product or SCFA in question. A person's MMP is thus the bulk total of all these individual reactions (S9A Fig).

In principle, some or all of the dietary fiber can be excreted in the stool without having undergone any fermentation in the colon. From the *ex vivo* experiment, we observed that at
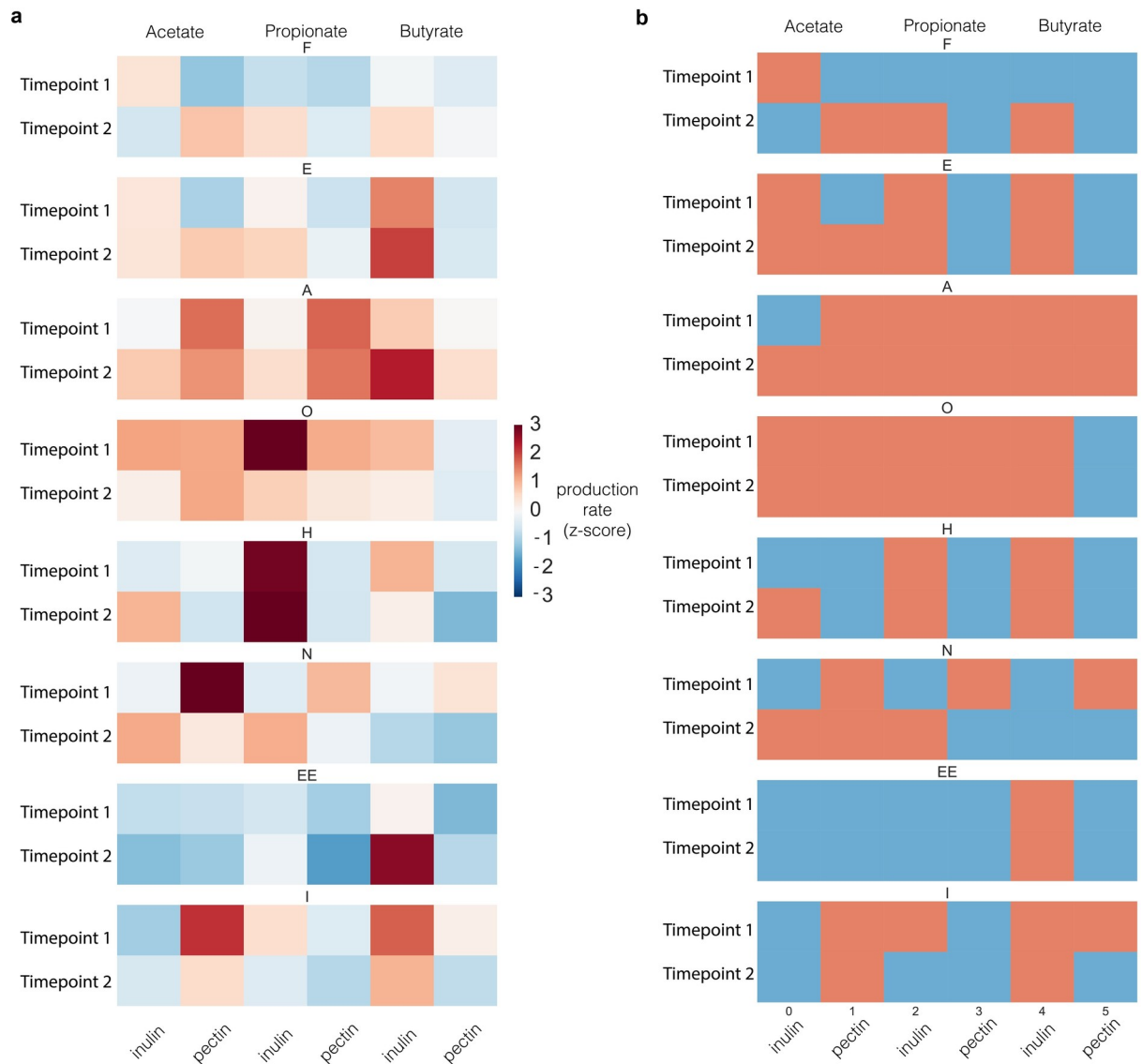
Fig 3. General features of an individual's MMP are stable over time. (a) Continuous production rates of each SCFA in response to inulin and pectin for two timepoints separated by at least six months, expressed as z-scores relative to the population in the dataset. (b) The same data, but collapsed to high or low producers of a given SCFA in response to a given fiber (red = high, blue = low).

least some of the ingested fiber is fermented and give rise to SCFAs production. For our model, in order to evaluate the limitations of SCFAs' absorption, we assume that the fermentable dietary fiber is almost entirely utilized within the gut and its excretion rate is zero (i.e. $[F]_{excreted} = 0$). (We sought to determine what portion of consumed dietary fiber is fermented in the gut by measuring fecal inulin concentration in samples taken from a previous *in vivo* study [21] as well as samples from all timepoints of the *ex vivo* experiment, these results can be found in S3 Fig).

Gut luminal concentrations of microbially-derived SCFAs is a balance between net production in the lumen and absorption by the host epithelium. Unfortunately, quantitative understanding of the rate of absorption of a given SCFA by cells in the gut epithelia is limited. Previous data collected using a dialysis bag technique suggest that SCFA absorption rate is
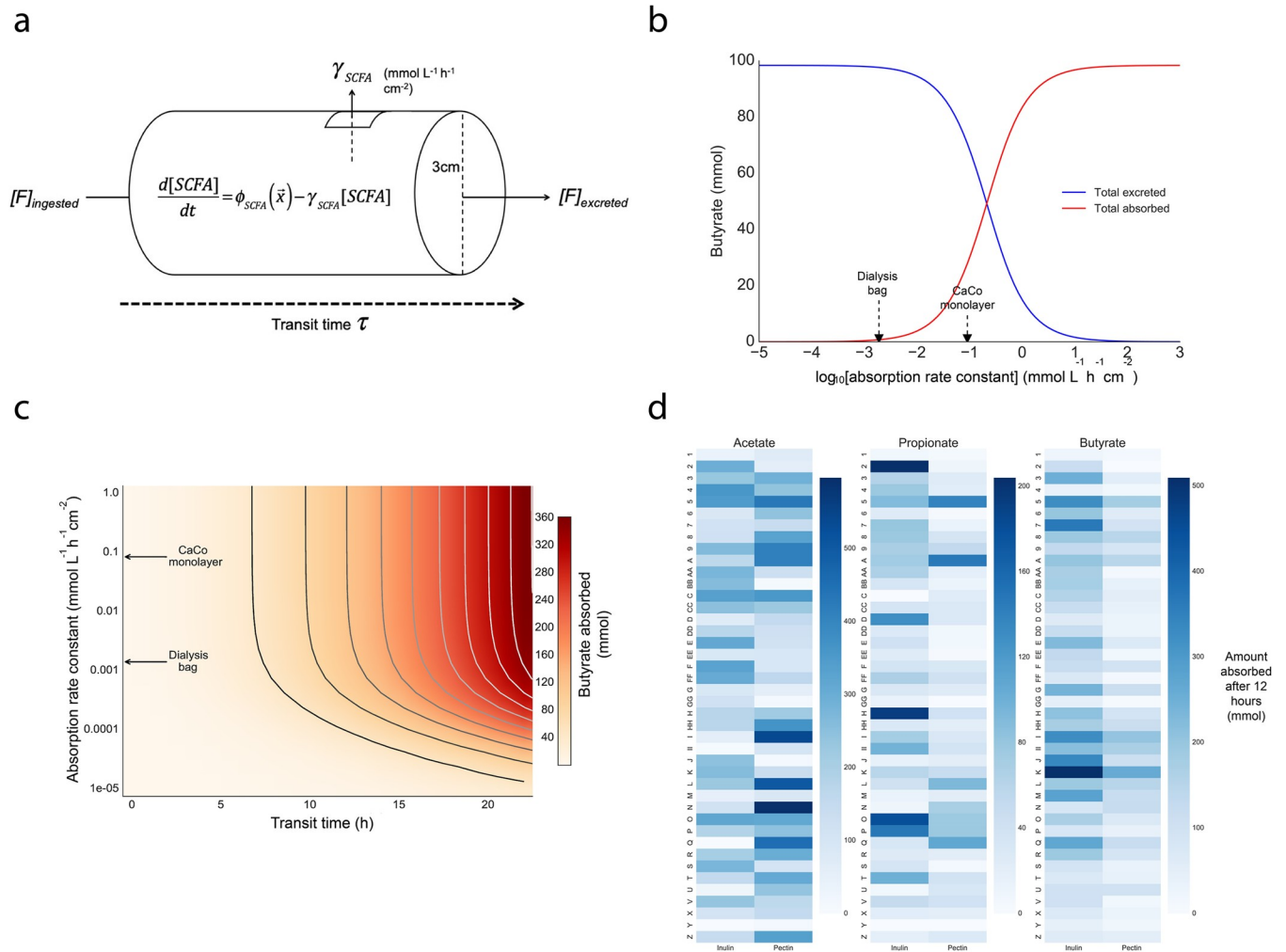
Fig 4. Phenomenological model of *in vivo* SCFA production predicts that inter-individual differences affects quantities absorbed. (a) Schematic of model parameters. (b) Predicted quantities of butyrate absorbed versus excreted in the stool in participant H as a function of the colonic epithelial absorption rate constant, assuming a transit time of 12 hours. Values of the rate constant measured by the dialysis bag and CaCo monolayer approaches discussed in the text are shown explicitly. (c) Phase diagram of predicted quantity of butyrate absorbed as a function of the epithelial absorption rate constant and transit time (participant H). Values of the rate constant measured by the dialysis bag and CaCo monolayer approaches discussed in the text are shown explicitly. (d) Predicted amount of each SCFA absorbed (in mmol) using a transit time of 12 hours for each subject and the dialysis bag rate constant parameters.

linear with concentration at typical physiological concentrations of SCFAs [22]. Moreover, studies have shown that SCFA absorption exhibits an unexpectedly modest pH-dependence [23]. We therefore modeled the absorption rate as proportional to the luminal SCFA concentration. We sought to estimate the order of magnitude of this rate constant, to better understand the fate of SCFAs produced by the microbiota in the colon (i.e. the balance between quantities absorbed versus quantities excreted). For this purpose, we used CaCo cell monolayers grown in a trans-well and measured SCFA concentrations from the media sampled from the apical and basal sides of the monolayer over a duration of 24h (cf. S1 File for details). Assuming the SCFA production rates measured *ex vivo* are representative of *in vivo* rates of production with the same inputs, and explicitly including the absorption rate parameters, we obtain the following system of phenomenological equations describing the time evolution of

each SCFA's concentration in the colonic lumen:

$$\frac{d[A]}{dt} = \phi_A(\overrightarrow{x}) - \gamma_A[A]$$

$$\frac{d[P]}{dt} = \phi_P(\overrightarrow{x}) - \gamma_P[P]$$

$$\frac{d[B]}{dt} = \phi_B(\overrightarrow{x}) - \gamma_B[B]$$

where $\phi_A(\overrightarrow{x})$, $\phi_P(\overrightarrow{x})$ and $\phi_B(\overrightarrow{x})$ are bulk production rates of acetate, propionate and butyrate respectively, produced in response to a given input concentration of fiber *[F]*, defining an individual's MMP.

To better understand the impact of the absorption rate constant on quantities of SCFA absorbed by the gut epithelium, we used our model to calculate the amount of butyrate absorbed in a given participant as a function of this rate constant:

$$B_{absorbed}(\tau) = V_{colon} S_{colon} \int\limits_0^\tau \gamma_B[B](t)dt,$$

where $B_{absorbed}$ is the quantity (in mmol) of butyrate absorbed, V_{colon} and S_{colon} are the volume and surface area of the colon, respectively, $[B]$ is the concentration of butyrate in the colon, and γ_B is the absorption rate of butyrate in mmol L^{-1}h^{-1}cm^{-2}. We used two different estimates of γ_B: the first corresponded to a published rate constant estimated using a dialysis bag technique [22], and the second was obtained from our own measurements using a previously reported CaCo monolayer model system [24]. We found that, assuming a transit time of 12 hours, these two rate constants have significantly different effects on the resulting dynamics: in the case of the dialysis bag constant (0.0019mmol L^{-1}h^{-1}cm^{-2}), the majority of produced butyrate is excreted, while in the case of CaCo monolayer (0.091mmol L^{-1}h^{-1}cm^{-2}), a significant quantity is absorbed (~40%) while the remainder is excreted (Fig 4B). This indicates that, depending on the value of the absorption rate constant, the relationship between excreted stool SCFA quantities and quantities of SCFA produced and absorbed in the gut are not necessarily related.

Plotting the quantity of butyrate absorbed as a function of both absorption rate and transit time indicates that transit time is a significant variable controlling the quantities of SCFA absorbed by the gut in the orders of magnitude of the absorption rates considered (Fig 4C). This relationship holds across different participants with different production rates of butyrate from the same input, related by a scaling factor. However, our model does not account for depletion of the fiber substrate. Naturally, slowing colonic transit time is therefore only an effective way of increasing absorbed SCFAs for as long as there remain SCFAs to absorb and fiber substrate to ferment. Nonetheless, it is clear from our model that, despite assuming equal absorption rate constants across participants and a transit time of 12h, different individuals absorb different quantities of an SCFA from a given fiber (Fig 4D), mirroring their differences in production and overall MMP. Absorption is linearly related to the produced concentration, and changing absorption rate constant parameters only affects the scale of absorbed quantities but the relative ratios between individuals are unchanged.

Discussion

In this study, we have used an *ex vivo* setup to measure the inter-individual differences in gut microbial SCFA production in response to different dietary fiber inputs. We have shown that

there are significant differences in the MMPs of different individuals, i.e. differences in the capacity of their microbiota to ferment a given fiber substrate into a given SCFA. Moreover, we showed that MMP could to a certain extent be predicted from stool microbiota community composition, to a greater extent than can raw stool SCFA content. These data are consistent with a recent study where participants with particular microbiota compositions were more likely to yield higher butyrate concentrations in the stool after consuming resistant starch [25].

In addition, we showed that the dominant features of an individual's MMP are relatively stable through time, consistent with the fact that an individual's MMP is related to their microbiota community composition. Though we did not study such cases, it is likely that treatment with broad-spectrum antibiotics or other extreme perturbations to the microbiota would have significant impact to an individual's MMP, which concomitant implications to their physiology.

In order to explore the relationship between our *ex vivo* results and the implications for *in vivo* SCFA production and absorption, we sought to develop a phenomenological model of the process. We found that quantities of absorbed SCFAs mirrored the quantities produced in the different parameter regimes considered, though the absolute amounts varied significantly as a function of the epithelial absorption rate constants used in our model. This highlights the critical importance of this variable in understanding the relationship between stool SCFA quantities and *in vivo* production and absorption for clinical applications. It is therefore of great importance to the field to obtain accurate measurements of these rate constants in addition to quantitative descriptions of their behavior and the underlying kinetics of absorption at different concentrations of SCFAs and in different regions of the gut.

Our model also showed that increased colonic transit time results in significantly greater quantities of absorbed SCFAs while these remain at appreciable concentrations in the gut, or that there remains substrate to ferment. In our *ex vivo* experiments, the microbiota was given a highly accessible, powdered form of dietary fiber as a substrate. It is likely that a less accessible form of fiber input (e.g. in the form of raw chicory root in the case of inulin) would take longer to ferment, in which case colonic transit time is an important variable in determining the quantities of SCFAs produced. Moreover, dietary fiber supplementation therapeutic strategies aiming to improve quantities of SCFAs produced would benefit from highly accessible forms of these fibers (purified powder), rather than grains, fruits and vegetables high in these quantities, which will likely take longer for the microbiota to ferment and therefore may result in a larger quantity of excreted, unfermented substrate.

Our model has several other limitations, which should be taken into account in the interpretation of our data. First of all, we do not consider fiber input concentration as a variable in the time evolution of our system, since we only considered a single concentration across participants (corresponding to significantly larger doses than those found in an ordinary diet) and assumed a linear production rate through time based on our *ex vivo* measurements (Fig 1B). In addition, we assumed that epithelial absorption rates are identical across all participants. Moreover, it is unclear whether experimental production rates measured *ex vivo* are related to *in vivo* production in the gut.

Despite these limitations, our model suggests a high degree of robustness to the association between differences in MMPs and quantities of absorbed SCFAs across different individuals, despite the fact that varying the absorption rate constant had significant impact on both absorbed and excreted SCFA quantities. Importantly, depending on the actual magnitude of the absorption rate constant, the relationship between stool SCFA concentration and *in vivo* production and absorption can be completely non-informative. Stool SCFA concentration is related to the differential between production and absorption, the colonic transit time, and whether or not the entire fermentable substrate was consumed. For example, in a regime

where production rate significantly exceeds absorption rate (as in both sets of absorption rate constant parameters considered in this study), stool SCFA is a function of the differential between the transit time (τ) and the time to depletion of the fermentable substrate (t_1). If $\tau > t_1$, the stool SCFA concentration will mostly be a function of the time during which the stool was transiting but no further SCFAs were being produced (and luminal SCFAs were merely being absorbed). This is an important consideration that suggests that stool SCFA quantities may not be the relevant quantity of interest in the absence of knowledge of these other variables. Put differently, a lack of association between stool SCFA concentration and a clinical outcome variable is not necessarily indicative of a lack of involvement of microbial SCFAs in the disease process.

Taken together, our data indicate that a quantitative understanding of a patient's MMP can inform personalized dietary supplementation strategies that aim to increase the production and resulting absorption of specific SCFAs in a patient's gut. More broadly, they suggest a framework for modulating SCFA production in a patient through two separate but complimentary means: modification of dietary inputs as a function of their existing MMP, and modification of the underlying community composition of their microbiota towards a given MMP of interest.

Methods

Human participants

Healthy human participants (14 females and 19 males) with ages ranging from 23 to 38 were consented to participate in this study under COUHES protocol number 1510271631.

Human fecal materials used in this study were collected under the protocol "Culturing of Bacterial Strains within Healthy Individuals"; the participants gave their written informed consent to participate in this study. The above-referenced study was reviewed and approved by the Ethical review Board of MIT, the Committee on the Use of Human as Experimental Subjects (COUHES), protocol number 1510271631.

Stool sample processing and *ex vivo* setup

Fresh stool samples were collected and weighed before being transferred to anaerobic conditions, in which they were diluted in reduced PBS containing 0.1% L-Cysteine to a ratio of 1g/5ml. Samples were homogenized into a slurry before being aliquoted into 96-well plates. Samples of the unaltered slurry were taken for a baseline sample, after which inulin (CAS number 9005-80-5, Alfa Aesar), pectin (CAS number 9000-69-5, MP Biomedicals, LLC), and cellulose (NutriCology) were added from stock solutions to final slurry concentrations of 10g/L, 5g/L and 20g/L, respectively. Concentrations of inulin and pectin were determined based on the maximum stock concentration we were able to obtain in which the dietary fiber was fully dissolved. Four conditions were measured: inulin, pectin, cellulose and control (no spike-in). The samples were incubated anaerobically at 37°C, and two biological replicates (two different wells in the 96-well plates) were collected at each time point. Samples were collected after 2 and 4 hours from all participants and sent out for SCFA quantification on a GC-FID. Linear production rate was measured between 2h and 4h because this allowed us maximum accuracy in measuring the time interval without introducing artefacts due to delays between conditions introduced during setup. Results obtained from slurry in large flasks on a shaker were in good agreement with data obtained in a 96-well plate format.

Short-chain Fatty acid measurements

Chromatographic analysis was carried out using an Agilent 7890B system with a flame ionization detector (FID) (Agilent Technologies, Santa Clara, CA). A high resolution gas

chromatography capillary column 30m x 0.25 mm coated with 0.25 µm film thickness was used (DB-FFAP) for the volatile acids (Agilent Technologies) and a high resolution gas chromatography capillary column 30m x 0.25 mm coated with 0.50 µm film thickness was used (DB-FFAP) for the nonvolatile acids. Nitrogen was used as the carrier gas. The oven temperature was 145˚C and the FID and injection port was set to 225˚C. The injected sample volume was 1µL and the runtime for each analysis was 12 minutes. Chromatograms and data integration was carried out using the OpenLab ChemStation software (Agilent Technologies).

For chromatographic analyses, a volatile acid mix containing 10mM of acetic, propionic, isobutyric, butyric, isovaleric, valeric, isocaproic, caproic, and heptanoic acids was used (Supelco CRM46975, Bellefonte, PA). A standard stock solution containing 1% 2-methyl pentanoic acid (Sigma-Aldrich St. Louis, MO) was prepared as an internal standard control for the volatile acid extractions.

Samples were kept frozen at -80˚C until analysis. The samples were removed from the freezer and allowed to thaw. A sample of the raw fecal material was transferred to a 2mL tube, the weight of the fecal material was determined and 1.5 mL of HPLC water was added to each sample. The samples were vortexed for 5 minutes until the material was homogenized. The pH of raw fecal suspension and the thawed fecal slurry samples was adjusted to 2–3 with 50% sulfuric acid. The acidified samples were kept at room temperature and vortexed for 10 minutes. The samples were centrifuged for 10 minutes at 5000g. 1000uL of the clear supernatant was transferred into a glass tube with a PTFE faced rubber lined screw cap for further processing. 50uL of the internal standard (1% 2-methyl pentanoic acid solution) and 1ml of ethyl ether anhydrous were added to the volatile samples. The tubes were mixed end over end for 10 minutes and then centrifuged at 2000g for 1 minute. The upper ether layer was transferred to an Agilent sampling vial for analysis.

For quantification of SCFAs, 1ml of each of the standard mixes were used and processed as described for the samples. The retention times and peak heights of the acids in the standard mix were used as references for the sample unknowns. These acids were identified by their specific retention times and the concentrations determined and expressed as mM concentrations per gram of sample for the raw fecal material and as mM concentrations per mL of fecal slurry.

16S rRNA sequencing

For DNA extraction, the MoBioPowersoil 96 kit (now Qiagen Cat No./Id: 12955–4) was used with minor modifications. All samples were thawed on ice and 250uL of the 5x dilution fecal slurry from the *ex vivo* assay from each sample were transferred to the Mobio High Throughput PowerSoil bead plate (12955–4 BP) for sample loading steps. We then proceeded through the extraction procotol on the same day following the manufacturer's instructions.

Paired-end Illumina sequencing libraries were constructed using a two-step PCR approach targeting 16S rRNA genes V4 region, previously described by Preheim et al. [26]. All paired-end libraries were multiplexed into lanes (at maximum 200 individual samples pooled per lane) and sequenced with paired end 150 bases on each end on an Illumina MiSeq platform.

The sequencing data is uploaded to NCBI Sequence Read Archive as BioProject ID PRJNA675106.

Inulin-specific ELISA

We use BioPAL's inulin immunoassay kit (BioPAL Worcester MA) to measure inulin concentration in filtrate from stool samples and followed the manufacturer's protocol. For stool samples that came from the *ex vivo* assay, they were already diluted 5x and needed to be centrifuged (10'000xg for 2 minutes) before passing the supernatant through a 0.2 µm syringe

filter (Pall Corporation, Port Washington NY). For stool samples that came from the *in vivo* diet study [21], they were thawed on ice and homogenized to a 5x dilution fecal slurry in PBS buffer before being spun down and filtered similarly to the *ex vivo* samples. The fecal filtrates were stored at -80˚C, always thawed on ice, and diluted to an appropriate dilution to be in the detection range of the ELISA Inulin kit. For reference, the *ex vivo* inulin conditions needed to be diluted at least 1'000-fold while the *ex vivo* control or *in vivo* samples only needed to be diluted at 10-fold.

Gut monolayer SCFA absorption measurements

We used a gut monolayer system without the immune component, where the gut monolayer was prepared as described previously by WLK Chen et al. [24, 27]. Briefly Caco-2 or C2BBe1 along with HT29-MTX-E21 cells (Sigma) were seeded onto rat-tail collagen I- (corning 354236) coated Transwell inserts (Corning 3460) in a 9:1 ratio (1 x 105 cells/cm2) in 500 uL seeding medium. The top and bottom compartments of the Transwell Plate were fed with 500uL and 1.5mL of seeding medium. The medium was changed every 2–3 days. After 20 days of gut insert maturation, the cell monolayer was ready to be used.

Barrier integrity was quantified by TEER (TransEpithelial Electrical Resistance) using the EVOM2 and the Enfohm-12 (World Precision Instruments) at 37˚C. We measured cell monolayer integrity at the beginning and the end of the 24h long experiment. At the in-between time points, we surveilled the cell monolayer's (dis)continuity under the microscope.

SCFAs were added to the mammalian cell media in their salt forms: Sodium Butyrate (Sigma, B5887), Sodium propionate (Alfa Aesar, A17440) and Sodium acetate (Sigma, S2889). They were added to the gut cell media on the apical side while the basal side was left unchanged, and the experiment was run for 24 hours. At each of the time points (0, 2, 4, 8 and 24h) we collected 100μL from the apical side and 200μL from the basal side from each gut cell Transwell insert (previous experiments done by WLK Chen et al. [27] have shown that the resulting reduction in volume did not affect the cell culture). Collected media from apical and basal sides were stored at -80˚C until analyzed for their SCFA contents. Each condition was run in triplicates except for the controls.

We ran 5 different concentrations of SCFA combinations. The Butyrate:Propionate:Acetate ratio (1:1:5, respectively) was kept constant across the different conditions. Concentrations tested were 40mM, 20mM, 10mM, 5mM and 2mM of Butyrate, with corresponding amounts of Propionate and Acetate according to the aforementioned ratio. The controls were either single SCFA (5mM Butyrate, 5mM Propionate or 25mM Acetate) or without added SCFA. At 24h, all the cells monolayers were washed, lysed and kept frozen at -80˚C for further analysis if needed.

16S rRNA sequencing analysis

Raw paired-end 16 S rRNA Illumina sequencing reads were merged, demultiplexed, and quality trimmed with a cut-off of Q = 25 using usearch8, before being trimmed to a common length of 226 bases. Dereplicated reads were then clustered into OTUs to 97% identity using UPARSE [28]. OTU centroids were assigned a taxonomy using the RDP classifier using an uncertainty cut-off of 0.5 [29].

Samples containing fewer than 5'000 reads (9 out of 370 samples) were discarded from OTU tables and all downstream OTU analyses. OTU tables were then rarefied to the lowest read-count. Alpha and beta-diversity analyses were conducted using raw filtered read counts, while OTU relative abundance analyses were conducted using tables where sample OTU counts were normalized by the total sample read count.

Machine learning

Random Forest Classifiers (RFCs) were built using the scikit-learn Python package. 5-fold cross-validation was used to construct an average Receiver Operator Characteristic (ROC) from which to compute an AUC.

Statistical analysis

MMP clusters were identified de novo with hierarchical clustering using the UPGMA algorithm from the script.cluster.hierarchy Python package.

Supporting information

S1 Fig. Percentage difference in SCFA quantification between biological replicates. Percentage difference in SCFA quantification by GC-FID between biological replicates, broken down by timepoint and SCFA. The overall distribution, across all participants, timepoints, and SCFAs, is plotted in blue, while the individual distributions for each timepoint and SCFA are shown in red.
(PNG)

S2 Fig. Individual butyrate production over 24h in pilot study. Individual 24h trajectories for Butyrate in five pilot participants. The different conditions are shown in different colors.
(PNG)

S3 Fig. Inulin measurements. (a) Concentrations of inulin measured at t = 0, 2h, 4h, 6h and 24h in six separate donors (donor IDs K, S, Z, EE, 1 and 2), determined using an inulin-specific ELISA assay. (b) Concentration of inulin in participant stool from a previous study [21] where participants were fed 10g of inulin daily on days 4, 5 and 6 against a constant (fiber-impoverished) dietary background. Fecal inulin concentrations ranged from undetectable to a detectable level (maximum of 25 mg/L) on the days following inulin consumption.
(PNG)

S4 Fig. Shannon diversity index. Shannon diversity index of 16S rRNA communities in each condition and each timepoint (0h, 4h, 24h).
(PNG)

S5 Fig. Beta diversity index between timepoints. (a) Jensen-Shannon Distances (left) and Weighted Unifrac (right) beta-diversities between timepoints at the level of 16S rRNA (0-4h, 0-24h, 4-24h). (b) Multidimensional scaling analysis of all 16S rRNA samples, colored by timepoint (left) and by participant (right).
(PNG)

S6 Fig. Weighted Unifracs between participants. Between participant Weighted Unifracs at 0h, 4h and 24h.
(PNG)

S7 Fig. pH measurements over time. pH of the slurry over time, measured across all participants in the study.
(PNG)

S8 Fig. Pairwise linear regressions. Pairwise linear regressions between between Bristol score of the sample, total SCFAs produced, and qPCR amplification cycle (C_T values).
(PNG)

S9 Fig. Microbial SCFA production proceeds through the cooperation of different members of the microbiota. (a) Schematic illustrating the generic steps involved in dietary fiber degradation. A bacterial OTU i of relative abundance x_i hydrolyses the polysaccharide (dietary fiber) F into oligosaccharides O. These are then fermented into a reduced intermediate P by OTU j with relative abundance x_j. Finally, P may be further fermented to an SCFA by OTU k with relative abundance x_k. In addition, the ability of an OTU to carry out a given reaction can itself be inhibited by a separate OTU (e.g. x_i is inhibited by x_o). (b) Bulk measurement of the overall production rate of a given SCFA, $\phi_{SCFA}(\boldsymbol{x})$, which is itself a function of the composition of the stool microbiota, $\boldsymbol{x}$, and corresponds to the quantities measured using our *ex vivo* experiments.
(PNG)

S10 Fig. Growth comparison *in vivo* and *ex vivo*. Comparison in relative abundances of two different OTUs of the genus *Blautia* in the same participants *in vivo* and *ex vivo*.
(PNG)

S1 Table. Raw SCFA dataset. (1) Subset 1 of *ex vivo* SCFA data. Samples ID include Donor ID–Culture condition (Ctrl- Control; Cell- Cellulose; Inul-Inulin; Pect- Pectin)–Timepoint ID (00- 0h; 02 – 2h; 04 – 4h). (2) Subset 2 of *ex vivo* SCFA data. Samples ID include Donor ID–Culture condition (Ctrl- Control; Cell- Cellulose; Inul-Inulin; Pect- Pectin)–Timepoint ID (00- 0h; 01 -2h; 02- 4h). (3) Gut-on-a-chip experiment SCFA data. Samples ID include Concentration ID–Timepoint ID (00- 0h; 01- 2h; 02- 4h; 03- 8h)–Replica ID (a, b or c where applicable). The different concentrations tested here were C1 – 2mM Butyrate, 2mM Propionate, 10mM Acetate; C2 – 5mM Butyrate, 5mM Propionate, 25 Propionate; C3- 10mM Butyrate, 10mM Butyrate, 50mM Acetate; But– 5mM Butyrate; Pro- 5mM Propionate; Ace– 25 mM Acetate.
(XLSX)

S1 File.
(DOCX)

Acknowledgments

We are very grateful for the assistance provided by Prof. Linda Griffith and Jason Velazquez with the colonocyte monolayer experiments. We also wish to thank Mary Delaney from the Harvard Digestive Disease Core for conducting the SCFA measurements, and Vicki Mountain for logistical support.

Author Contributions

Conceptualization: Thomas Gurry, Le Thanh Tu Nguyen, Eric J. Alm.

Data curation: Thomas Gurry, Le Thanh Tu Nguyen, Xiaoqian Yu.

Formal analysis: Thomas Gurry, Le Thanh Tu Nguyen.

Funding acquisition: Eric J. Alm.

Methodology: Thomas Gurry, Le Thanh Tu Nguyen.

Project administration: Thomas Gurry, Le Thanh Tu Nguyen.

Supervision: Eric J. Alm.

Validation: Thomas Gurry, Le Thanh Tu Nguyen.

Visualization: Thomas Gurry, Le Thanh Tu Nguyen.

Writing – original draft: Thomas Gurry, Le Thanh Tu Nguyen.

Writing – review & editing: Thomas Gurry, Le Thanh Tu Nguyen, Eric J. Alm.

References

1. Davie JR. Inhibition of Histone Deacetylase Activity by Butyrate. J Nutrition. 2003; 133: 2485S–2493S. https://doi.org/10.1093/jn/133.7.2485S PMID: 12840228

2. Iraporda C, Errea A, Romanin DE, Cayet D, Pereyra E, Pignataro O, et al. Lactate and short chain fatty acids produced by microbial fermentation downregulate proinflammatory responses in intestinal epithelial cells and myeloid cells. Immunobiology. 2015; 220: 1161–1169. https://doi.org/10.1016/j.imbio.2015.06.004 PMID: 26101138

3. Kuo S-M. The Interplay Between Fiber and the Intestinal Microbiome in the Inflammatory Response. Adv Nutrition Int Rev J. 2013; 4: 16–28. https://doi.org/10.3945/an.112.003046 PMID: 23319119

4. Roediger WE. Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man. Gut. 1980; 21: 793. https://doi.org/10.1136/gut.21.9.793 PMID: 7429343

5. Roediger WEW. Utilization of Nutrients by Isolated Epithelial Cells of the Rat Colon. Gastroenterology. 1982; 83: 424–429. https://doi.org/10.1016/s0016-5085(82)80339-9 PMID: 7084619

6. Perry RJ, Peng L, Barry NA, Cline GW, Zhang D, Cardone RL, et al. Acetate mediates a microbiome–brain–β-cell axis to promote metabolic syndrome. Nature. 2016; 534: 213–217. https://doi.org/10.1038/nature18309 PMID: 27279214

7. De Vadder F, Kovatcheva-Datchary P, Goncalves D, Vinera J, Zitoun C, Duchampt A, et al. Microbiota-Generated Metabolites Promote Metabolic Benefits via Gut-Brain Neural Circuits. Cell. 2014; 156: 84–96. https://doi.org/10.1016/j.cell.2013.12.016 PMID: 24412651

8. Pingitore A, Chambers ES, Hill T, Maldonado IR, Liu B, Bewick G, et al. The diet-derived short chain fatty acid propionate improves beta-cell function in humans and stimulates insulin secretion from human islets in vitro. Diabetes Obes Metabolism. 2016; 19: 257–265. https://doi.org/10.1111/dom.12811 PMID: 27761989

9. Zhao L, Zhang F, Ding X, Wu G, Lam YY, Wang X, et al. Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. Science. 2018; 359: 1151–1156. https://doi.org/10.1126/science.aao5774 PMID: 29590046

10. Sanna S, Zuydam NR van, Mahajan A, Kurilshikov A, Vila AV, Võsa U, et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nat Genet. 2019; 51: 600–605. https://doi.org/10.1038/s41588-019-0350-x PMID: 30778224

11. Zimmerman MA, Singh N, Martin PM, Thangaraju M, Ganapathy V, Waller JL, et al. Butyrate suppresses colonic inflammation through HDAC1-dependent Fas upregulation and Fas-mediated apoptosis of T cells. Am J Physiol-gastr L. 2012; 302: G1405–G1415. https://doi.org/10.1152/ajpgi.00543.2011 PMID: 22517765

12. Stein RR, Tanoue T, Szabady RL, Bhattarai SK, Olle B, Norman JM, et al. Computer-guided design of optimal microbial consortia for immune system modulation. Elife. 2018; 7: e30916. https://doi.org/10.7554/eLife.30916 PMID: 29664397

13. Machiels K, Joossens M, Sabino J, Preter VD, Arijs I, Eeckhaut V, et al. A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. Gut. 2014; 63: 1275. https://doi.org/10.1136/gutjnl-2013-304833 PMID: 24021287

14. Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, et al. Specificity of Polysaccharide Use in Intestinal Bacteroides Species Determines Diet-Induced Microbiota Alterations. Cell. 2010; 141: 1241–1252. https://doi.org/10.1016/j.cell.2010.05.005 PMID: 20603004

15. Filippis FD, Pasolli E, Tett A, Tarallo S, Naccarati A, Angelis MD, et al. Distinct Genetic and Functional Traits of Human Intestinal Prevotella copri Strains Are Associated with Different Habitual Diets. Cell Host Microbe. 2019; 25: 444–453.e3. https://doi.org/10.1016/j.chom.2019.01.004 PMID: 30799264

16. Duncan SH, Barcenilla A, Stewart CS, Pryde SE, Flint HJ. Acetate Utilization and Butyryl Coenzyme A (CoA):Acetate-CoA Transferase in Butyrate-Producing Bacteria from the Human Large Intestine. Appl Environ Microb. 2002; 68: 5186–5190. https://doi.org/10.1128/AEM.68.10.5186-5190.2002 PMID: 12324374

17. Reichardt N, Duncan SH, Young P, Belenguer A, Leitch C, Scott KP, et al. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. The ISME Journal. 2014; 8: 1323–1335. https://doi.org/10.1038/ismej.2014.14 PMID: 24553467

18. Belenguer A, Duncan SH, Calder AG, Holtrop G, Louis P, Lobley GE, et al. Two Routes of Metabolic Cross-Feeding between Bifidobacterium adolescentis and Butyrate-Producing Anaerobes from the Human Gut. Appl Environ Microb. 2006; 72: 3593–3599. https://doi.org/10.1128/AEM.72.5.3593-3599.2006 PMID: 16672507

19. Chen T, Long W, Zhang C, Liu S, Zhao L, Hamaker BR. Fiber-utilizing capacity varies in Prevotella— versus Bacteroides -dominated gut microbiota. Scientific Reports. 2017; 7: 2594. https://doi.org/10.1038/s41598-017-02995-4 PMID: 28572676

20. David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, et al. Host lifestyle affects human microbiota on daily timescales. Genome Biol. 2014; 15: R89. https://doi.org/10.1186/gb-2014-15-7-r89 PMID: 25146375

21. Gurry T, Gibbons SM, Kearney SM, Ananthakrishnan A, Jiang X, Duvallet C, et al. Predictability and persistence of prebiotic dietary supplementation in a healthy human cohort. Scientific reports. 2018; 8: 12699. https://doi.org/10.1038/s41598-018-30783-1 PMID: 30139999

22. McNeil NI, Cummings JH, James WP. Short chain fatty acid absorption by the human large intestine. Gut. 1978; 19: 819. https://doi.org/10.1136/gut.19.9.819 PMID: 30683

23. Occurrence Bugaut M., absorption and metabolism of short chain fatty acids in the digestive tract of mammals. Comp Biochem Physiology Part B Comp Biochem. 1987; 86: 439–472. https://doi.org/10.1016/0305-0491(87)90433-0 PMID: 3297476

24. Trapecar M, Communal C, Velazquez J, Maass CA, Huang Y-J, Schneider K, et al. Gut-Liver Physiomimetics Reveal Paradoxical Modulation of IBD-Related Inflammation by Short-Chain Fatty Acids. Cell Syst. 2020. https://doi.org/10.1016/j.cels.2020.02.008 PMID: 32191873

25. Baxter NT, Schmidt AW, Venkataraman A, Kim KS, Waldron C, Schmidt TM. Dynamics of Human Gut Microbiota and Short-Chain Fatty Acids in Response to Dietary Interventions with Three Fermentable Fibers. Mbio. 2019; 10: e02566–18. https://doi.org/10.1128/mBio.02566-18 PMID: 30696735

26. Preheim SP, Perrotta AR, Martin-Platero AM, Gupta A, Alm EJ. Distribution-Based Clustering: Using Ecology To Refine the Operational Taxonomic Unit. Appl Environ Microb. 2013; 79: 6593–6603. https://doi.org/10.1128/AEM.00342-13 PMID: 23974136

27. Chen WLK, Edington C, Suter E, Yu J, Velazquez JJ, Velazquez JG, et al. Integrated gut/liver microphysiological systems elucidates inflammatory inter-tissue crosstalk. Biotechnol Bioeng. 2017; 114: 2648–2659. https://doi.org/10.1002/bit.26370 PMID: 28667746

28. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods. 2013; 10: 996–998. https://doi.org/10.1038/nmeth.2604 PMID: 23955772

29. Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM, et al. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res. 2005; 33: D294–D296. https://doi.org/10.1093/nar/gki038 PMID: 15608200