# PNAS

## www.pnas.org

**Supplementary Information Appendix for**

CODY enables quantitatively spatiotemporal predictions on in vivo gut microbial variability induced by diet-intervention.

Jun Geng[1], Boyang Ji[1], Gang Li[1], Felipe López-Isunza[2] & Jens Nielsen[1,3,4,5*]

[1]Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

[2]Departamento de Ingeniería de Procesos e Hidráulica, Universidad Autónoma Metropolitana-Iztapalapa, México City, México

[3]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark

[4]Wallenberg Center for Protein Research, Chalmers University of Technology, Gothenburg, Sweden

[5]BioInnovation Institute, Ole Maaløes Vej 3, DK2200 Copenhagen N, Denmark

* Jens Nielsen

**Email:** jni@novo.dk

**This PDF file includes:**

>    Supplementary Methods 1 to 9
>    Figures S1 to S29
>    Tables S1 to S8
>    Legends for Datasets S1 to S11
>    SI References

**Other supplementary materials for this manuscript include the following:**

>    Datasets S1 to S11

## Supplementary Information Appendix

## Supplementary Methods:

## Supplementary Figures:

## Supplementary Tables:

## Legends for Datasets S1 to S11

Dataset S1. Metabolic pathway modules of B. thetaiotaomicron.

Dataset S2. Metabolic pathway modules of B. fragilis.

Dataset S3. Metabolic pathway modules of B. longum.

Dataset S4. Metabolic pathway modules of B. adolescentis.

Dataset S5. Metabolic pathway modules of B. breve.

Dataset S6. Metabolic pathway modules of E. hallii.

Dataset S7. Metabolic pathway modules of F. prausnitzii.

Dataset S8. Metabolic pathway modules of R. intestinalis.

Dataset S9. Metabolic pathway analysis results of eight microbiome members.

Dataset S10. Summary of parameters of bacterial biology for monoculture of microbiome representatives.

Dataset S11. Simulation results of monoculture and coculture of two species consortium.

# Supplementary Methods

**Method 1 | Dynamic models of bacterial biology using enzyme-centric metabolic framework (ECMF)**

Process of ECMF construction

The detailed process of how the multiscale model of CODY is developed is shown in Fig. S1a. In order to investigate how the gut ecosystem evolves and dynamically adapts to environmental changes[1], we used a cybernetic modeling approach ("Wiener's theory of cybernetics", by Norbert Wiener [2]), to construct dynamic mathematical models for individual bacteria to describe their adaptive biology functions. The emergent higher-order interactions and community-level microbiota dynamic is described by using a hierarchical resource allocation framework (HRAF, Fig. S1b, step3). Therefore, the intrinsic dynamics of the gut microbiota ecosystem could be described by integrating ECMF into HRAF. The impact of colon physical forces, diet, and spatial structure is described by the spatial compartmentalized framework (SPCF, Fig. S1b, step 4-5). By integrating ECFM and HRAF with SPCF, CODY is assembled, and this characterizes how the intrinsic dynamics of the gut microbiota is impacted by the structure and physical forces in different regions of the colon. In this section, the construction process of ECMF is introduced. The construction processes of the other two frameworks are introduced in Method 2-3 (*SI Appendix*), and the assemblage process of CODY is introduced in Method 8 (*SI Appendix*). The step-by-step workflow of ECMF construction is performed for representative species (Fig. S1b, step 2) in the gut microbiota ecosystem (Fig. S1b, step 1). This modeling framework not only incorporates the stoichiometric information of representative species, but also integrates regulatory machinery by postulating metabolic regulation as the controlled response of the organism by adapting to its environment [3].

As can be seen from Fig. 1a and *SI Appendix* Fig. S1b, ECMF contains two components, i.e., metabolism and regulation. The bacterial metabolism is characterized by the repertoire of biochemical reactions the bacteria can carry out. In this work, the metabolism for each species is derived by using reduced genome-scale metabolic models (GEMs) from published GEMs that have been reported in the literature [4, 5]. To this end, we used referenced GEMs as the metabolic backbone, and removed reactions belonging to irrelevant pathways associated with the interested metabolites in this work, which make them mainly confined on the central metabolism, according to the central pathways for most of these species. Besides, the signature pathways for individual species which has been reported in the literature are added to associated reduced GEMs [6-8]. More specifically, for the metabolism of Bacteroides (*B. thetaiotaomicron* and *B. fragilis*), a primitive electron transfer chain using fumarate as the terminal electron acceptor [6] was integrated, and the updated metabolic network was checked to involve known reactions from the published metabolic network [9]. Metabolic networks for Bifidobacterium (*B. longum*, *B. breve* and *B. adolescentis*) were compared with published reports [10, 11] and integrated with the F6P-shunt pathway. For *F. prausnitzii* and *R. intestinalis*, the reduced GEMs were compared manually with published metabolic networks [12] with respect to central metabolic pathways and the Bcd/Etf complex of the electron transfer chain was introduced. The reduced GEMs of *E. hallii* was compared to the published metabolic network for *E. rectale* [10] given their closely phylogenetic distance. Finally, in order to avoid potential dead ends and gaps, after performing the reduction

and integration process, we used the Raven toolbox [13] for all the models to ensure the quality of all reduced GEMs of these bacteria. This was done by performing defined metabolic tasks, i.e., ensuring the reduced GEMs have no dead-ends, ensuring the reduced GEMs cannot produce metabolites from nothing, ensuring each metabolic precursor for growth can be synthesized, and so on[13]. Information on the original GEMs that are referred as the backbone for each bacterium is provided in Table S2. All reduced GEMs of the representatives are provided as Dataset S1-8.

Based on the reduced GEMs which represent the metabolic structure for each species, metabolic pathway analysis was implemented to firstly decompose the metabolic network into a unique (i.e., non-decomposable) set of all possible and feasible pathways. Given the large number of the obtained feasible pathways, metabolic yield analysis [14] was further performed which reduced these pathways to an active subset of metabolic pathway modules (MPMs) encompassing minimal sets of connected reactions that function at steady state. To ensure that the metabolic features of each species are conserved with respect to relevant metabolic yields measured *in vitro*, representation of more than 99% of the original solution space was applied as a criterion that must be achieved by convex combination of the MPMs. It is known that the solution space of the mass balance equations, i.e., by letting Eq. 2 equals to 0, is a convex hull (see equation below). To enable the yield analysis, the solution space was mapped to the yield space by normalizing each row of the MPMs with respect to the corresponding substrate uptake flux in the same row. Some bacteria may have more than one substrate, such as *F. prausnitzii*, which consumes both hexose and acetate, the reference substrate uptake flux is considered as hexose flux. Thus, in the yield space, the substrate column is normalized to 1, and other columns would be converted to a ratio compared to the substrate. Consequently, the solution space would be a bounded convex hull in the yield space. In the yield space, pathways that contribute negligibly are first excluded by examining the impact by removing it. Next, an initial subset of pathways can be determined which constitute large enough space in the yield space. Following this, an iterative process by assessing the pathway that contribute the most to increasing the area of the yield space would be added at one time. A threshold of 0.99 was set to determine when to terminate the iterative process. Following this procedure, the number of finally obtained MPMs were largely reduced, and the resulting MPMs were used to represent the metabolic features of the individual species.

For each MPM, an aggregated enzyme is allocated to describe the flux through the MPM, i.e. the uniquely derived fundamental pathway modules represented metabolic subunits upon which we incorporated enzyme-centric regulatory processes. The so called "self-regulated" mechanism, that is believed to be retrieved from extended adaptation to evolution[15], is regarded as a distinct feature of metabolism that endows the microorganism to preferentially promote the synthesis and activity of enzymes that catalyze reactions with maximal returns. More specifically, this will allow the limited amount of intracellular resources to be optimally allocated to MPMs by synthesizing the corresponding enzymes and controlling their activities in a regulated manner such that the bacteria growth rate was maximized. Therefore, the regulatory machinery was incorporated in such a way to ensure that a fixed amount of intracellular resources allocated economically and efficiently in an automatic way to meet the nutrition availability. To this end, two cybernetic variables, *u* and *v*, corresponding to the optimal control of enzyme synthesis and enzyme activity level, were assigned to each MPM. The formulations of these variables are derived based on optimal control theory [15]. Incorporation of self-regulatory mechanisms into our approach endowed

individual consortium members with the ability to specify preferential use of alternative metabolic pathways, and thus facilitate metabolic adaptation within the community. The specific uptake rate is thus characterized as the weighted sum of corresponding time-specific reaction rates of individual MPMs. For each MPM, its associated uptake flux rate is characterized by the regulated kinetic formulation adopting Michaelis-Menten functions, which is dependent on the time-specific substrate concentration. In together, the specific uptake rate could be viewed as regulated resource allocation among different MPMs. Such regulation strategy is dependent on both time-specific substrate concentrations and the cellular demand function that is captured by inbuilt cybernetic variables and metabolic pathways intrinsic properties. Therefore, the specific uptake rate is the outcome of cellular regulated behavior, which endows our ECMF to relieve the necessity to measure itself *a priori*. This feature of ECMF thus made it possible to describe the species fitness function, particularly when encountering environmental changes.

ECMF formulations

The mass balance equations of the ECMF for single species could be described as the following set of equations,

$$\frac{1}{c}\Phi(\dot{\mathbf{x}},t)=\mathbf{S}_{\mathbf{x}}\mathbf{r} \tag{1}$$

$$\Phi(\dot{\mathbf{m}},t)=\mathbf{S}_{\mathbf{m}}\mathbf{r}-\mu\mathbf{m} \tag{2}$$

$$\frac{1}{c}\Phi(\dot{\mathbf{c}},t)=\mu \tag{3}$$

Where $\dot{\Phi}$ is the dynamic changes of state variables. $\mathbf{x}$ denotes the vector of extracellular metabolite concentrations, $c$ is biomass concentration, $\mathbf{m}$ is the vector of intracellular metabolite concentrations, $\mathbf{S}_{\mathbf{x}}$ and $\mathbf{S}_{\mathbf{m}}$ are the matrices that stoichiometrically link the flux vector ($\mathbf{r}$) to exchange and intracellular fluxes, $\mu$ is the specific growth rate. Under quasi-steady-state approximation, the flux vector can be represented by a convex (nonnegative) combination of MPMs, i.e.,

$$\mathbf{r}=\mathbf{Z}\mathbf{r}_{\mathbf{M}} \tag{4}$$

Where $\mathbf{Z}(n_R \times n_Z)$ is the matrix of the metabolic pathway modules, with each column representing the relative stoichiometric ratio of each flux through the metabolic pathway modules. $\mathbf{r}_{\mathbf{M}}(n_Z \times 1)$ could be regarded not only as the weights to the MPMs, but also as the vector of fluxes that go through the MPMs (mmol/gDW/h). Thus, substitution of Equation (1) with the above relationship would give:

$$\Phi(\dot{\mathbf{x}},t)=\mathbf{S}_{\mathbf{x}}\mathbf{Z}\mathbf{r}_{\mathbf{M}}c \tag{5}$$

Where $\mathbf{S}_{\mathbf{x}}\mathbf{Z}$ represent the stoichiometric coefficients between extracellular metabolites and the MPMs. The uptake rate vector of MPMs, $\mathbf{r}_{\mathbf{M}}$ is represented by:

$$\mathbf{r}_{\mathbf{M}}=\mathbf{D}(\mathbf{v})\mathbf{D}(\mathbf{e})\mathbf{r}_{\mathbf{M}}^{\text{kin}}(\mathbf{x}) \tag{6}$$

5

Where $\mathbf{r}_M^{kin}$ denotes the unregulated kinetic form of the uptake rate associated with the $j$-th metabolic pathway module (MPM), following the Michalis-Menten function, which is determined by the kinetic parameters of $k_{max}$, i.e.,

$$r_{M,j}^{kin} = k_{max} \frac{\mathbf{x}_S}{\mathbf{x}_S + \mathrm{K}_M} \tag{7}$$

In Eq. 6, $\mathbf{v}$ is the cybernetic variable vector with respect to controlling enzyme activities, $\mathbf{e}$ is the vector of enzyme levels for MPMs which becomes an additional state variable. $\mathbf{r}_M^{kin}$ is the unregulated kinetic uptake rate vector of MPMs which follow Michaelis-Menten forms, which depends on the available substrate vector $\mathbf{x}$. $\mathbf{D}(\bullet)$ is the operator which converts the input vector into the diagonal matrix such that,

$$\mathbf{D}(\mathbf{v}) = \begin{bmatrix} v_1 & 0 & 0 & 0 \\ 0 & v_j & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & v_{n_Z} \end{bmatrix} \tag{8}$$

$$\mathbf{D}(\mathbf{e}) = \begin{bmatrix} e_1 & 0 & 0 & 0 \\ 0 & e_j & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & e_{n_Z} \end{bmatrix} \tag{9}$$

Where $v_j$ is the cybernetic variable controlling enzyme activity of the $j$-th MPM, and $e_j$ is the associated enzyme levels of the $j$-th MPM. Following this, the explicit form of $\mathbf{r}_M$ associated with the $j$-th MPM should be as follows, i.e.,

$$r_{M,j} = v_j k_{max,j} e_j \frac{\mathbf{x}_S}{\mathbf{x}_S + \mathrm{K}_M} \tag{10}$$

The differential equations for the enzyme state variables can be written as

$$\dot{\mathbf{e}} = \boldsymbol{\alpha} + \mathbf{D}(\mathbf{u})\mathbf{r}_{ME}^{kin} - (\mathbf{D}(\boldsymbol{\beta}) + \mu\mathbf{I})\mathbf{e} \tag{11}$$

Where $\mathbf{u}$ is the cybernetic variable regulating the induction of enzyme synthesis rate, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the constitutive synthesis rate and degradation rate.

The formulations of cybernetic variables $\mathbf{u}$ and $\mathbf{v}$ are derived according to optimal control theory[15]:

$$\mathbf{u} = \frac{\mathbf{p}}{\|\mathbf{p}\|_1} \quad \text{and} \quad \mathbf{v} = \frac{\mathbf{p}}{\|\mathbf{p}\|_\infty} \tag{12}$$

Where $\|\mathbf{r}\|_1$ and $\|\mathbf{r}\|_\infty$ denote the $l_1$ and $l_\infty$ vector norms of uptake fluxes individually. $\mathbf{p}$ is the return-on-investment vector, which is given by the uptake rate vector in current context.

6

## Method 2 | Dynamic models of microbial community using the hierarchical resource allocation framework (HRAF)

The next step is to develop community-level dynamic framework to accommodate multiple species interactions (Fig. S1b, step 3). Microbiome assemblages emerging from interspecies interactions are highly complex, and constantly interactive with the environment. To characterize these features at the community level, a hierarchical framework to describe multi-level resource allocation from microbial community to individual members, and further among individual bacterial alternative MPMs, is developed, which is termed as HRAF. The "hierarchical" here is used to strengthen that the dynamic process of the in vivo microbial-microbial interactions in a system-level perspective, i.e., how the microbial-microbial interactions are impacted by the intrinsic dynamics of individual bacterium, how the resulted dynamic microbial growth further impact the common resources variations, and in turn, how the previous two factors determine the microbial-microbial interactions. During this three-stage process, each step is not isolated with the others, but hierarchically coupled/linked to each other thus collectively characterize how the environmental resources determine the higher-order microbial interactions.

At the first level, microbial-accessible carbohydrates (MACs), such as fiber, resistant starch, and human milk oligosaccharides (HMO), are allocated among microbial ecosystem members through degradation (Fig. S1b, step 3). The amount of resources allocated to each member at this level is dependent on individual bacterial accessibility to MACs components contained in the diet (Fig. 1d), i.e., bacterium with higher accessibility will grow with preferential dominance. In this process, the degradation rate of the complex carbohydrate was taken into account using Contois equation, which was adapted to wide range of organic wastes in anaerobic digestion process and therefore regarded as a general model of hydrolysis for complex carbohydrate ($\xi_z$) degradation [16], that is described by:

$$r_h = \rho_{hyd,z} \frac{\xi_z \varphi_{su,i}}{K_{x,z} \varphi_{su,i} + \xi_z} \tag{13}$$

Where $\xi_z$ denotes the stage variable of site-specific complex carbohydrate abundance. $\varphi_{su,i}$ denotes the site-specific absolute abundance of specific microbial species. $\rho_{hyd,z}$ denotes the degradation capacity constant of specific microbial species, and $K_{x,z}$ denotes the hydrolysis constants, which are detemrined based on values taken from the literature [17].

At the second level, the allocated MACs are degraded to simpler forms and metabolized by individual bacterium. At this level, the dynamic variations of individual bacterium are determined by two impact factors, i.e., the intrinsic dynamics characterized by specific ECMF, and the microbial-microbial interactions characterized by the HRAF concept. Microbial-microbial interactions occurred dynamically through the exchange of environmental resources or byproducts, dependent on the available environmental resource that emerged by collective contributions from ecosystem members. During this process, the substrate that act as the main substrate is the simpler forms of the carbohydrates that are derived during microbial degradation process. The extent to which available resource is allocated to each microbiota members could be derived based the total hydrolysis rate for each species i.e.,

7

$$r_{\xi_{hex}} = Y_{Hex,z} r_h \tag{14}$$

Where $r_{\xi_{hex}}$ is the rate of generated carbohydrate resources in the common environment that is shared by all microbial ecosystem members as given in Eq. 11, with $Y_{Hex,z}$ being the hexose yield from degradation of MACs, i.e. the amount of hexose obtained from the carbohydrate.

At the third level, environmental resources are continuously updated along with microbial ecosystem development and through mass transport in the colon, which would further shape the first and the second level (associated formulations with respect to the colon physiology impacts on environmental resources could be referred in Method 3(*SI Appendix*)).

The common resources shared among all ecosystem members are consumed and thus dynamically varied, which further impact the metabolic behavior of individual bacteria. The system equation for dynamic changes of common resources due to the collective impacts of microbial intrisic dynamics, the microbial-microbial interactions as well as the degradation process could be summarized as follows,

$$\Phi(\dot{\xi},t) = \sum_{J=1}^{N} (\mathbf{S_x Z})^J (\mathbf{r_M})^J c^J + Y_{Hex,z} \sum_{J=1}^{N} (\mathbf{r_h})^J_{Mk} \tag{15}$$

Where the first term denotes the consumption by the microbial ecosytem (with the negative number in corresponding rows indicating the resouce as been consumed), and the second term denotes the hydrolysis derived resource amount.

**Method 3 | Biomimetic model of the colon using the spatial compartmentalized framework (SPCF)**

<u>Colon physiology</u>

The schematic view of colon physiology is shown in Fig. 1e. In the cross-sectional direction, the structure of the colon includes two compartments where microbiota resides, i.e., lumen compartment and mucus compartment. In the longitudinal directions, the colon structure resembles a hollow tube with a very large internal surface area for nutrients and water absorption through the wall-membrane. The large intestine wall is lined with colon epithelium, where the secretion of a mucus gel layer is carried out by goblet cells. The polymer-based matrix characteristic of the mucus layer enables the attachment of microorganisms and their resistance to shear forces. The mucus layer is mainly formed by mucins, which are high molecular weight glycoproteins playing an important role in microbial development [18]. In the longitudinal direction of the colon, due to peristalsis movement engendered by peristaltic muscular contraction and relaxation deeper generated by both the circular and longitudinal muscle layer within the bowel wall [19], physiological forces thus give rise to the back-mixing of the material flow, which then impact the microbiota stability. Water absorption by epithelium cell via diffusion through the colon wall membrane contributes to condensing feces. These features would ensure the major function of colon, which is to absorb water and digest complicated molecules like polysaccharides that cannot be accessed by humans, and release energy metabolites (acetate, propionate, butyrate) into the blood circulatory system (BCS) through the mesenteric vein. The collected energy metabolites are therefore converged into the hepatic portal vein through BCS and finally transported to other organs for energy supply. We constructed a multiscale gut ecosystem, which incorporate not only interspecies interaction, but also colon physiology, mass transport and the information on the spatial distribution of microbiota.

<u>Biomimetic model to mimic *in vivo* colon environment</u>

We integrated the intrinsic dynamics of the gut microbiota into the colon environment to describe the colon environmental impacts on the microbiota development (Fig. S1b, step 4-5). To describe the colon microscopic structure and physiological functions, we constructed a biomimetic model of *in vivo* colon milieu using a tubular bioreactor model, which can be mathematically described by well-mixed tanks-in-series reactor model with recycle, which describes the material-flow from cecum downwards the feces, as well as axial and radial back-mixing effects (hydrodynamics) due to peristaltic movements at the colon wall. The spatial compartmentalized framework of the biomimetic model of the colon is termed SPCF in this text.

To model the inverted U shape of the colon structure in the longitudinal direction (Fig. S1b, step 4),  we used four regions lined in two rows of parallel interconnected ideally well-mixed tanks of constant volumes, and a ninth region represented as a fed-batch reactor that mimic the rectum [20]. Within each region of the lumen/mucus- compartments, we considered the physical forces in the longitudinal direction, including the inflow rate, the dietary intake, the peristaltic movement due to periodic contraction of the circular colon muscles, as well as the defecation process.

To model the multi-tissue layer structure of the colon in the radial direction (Fig. S1b, step 5), the compartmentalized spatial structure of the gut microbiota is considered as residing within each region of the lumen and mucus compartments. The physical forces in the radial direction, including perfusion between lumen and mucus and water absorption is modeled as mass

transport, bacteria detachment from mucus to lumen compartment is modeled to occur at a constant rate. Additionally, diffusion of key metabolites between BCS and the four regions of the mucus compartment is modeled by representing BCS as a single continuous well-mixed tank, which connects the tanks representing the mucus compartment. All relevant parameters were determined relying on literature-based knowledge as introduced in the following section.

Parameters determination to characterize *in vivo* colon physiology

Parameters used to describe the colon physiology were taken as typical values reported in the literature (Table S7). The length of the colon has been reported to vary in different populations [21, 22], but is about 150~160 cm on average for healthy individuals. The diameter of the colon is averagely estimated to be about 4.8~6 cm [21, 23]. The volume of infant colon was assumed to be around one fourth of the reported value for an adult colon. The four well-defined regions of the colon were set equally representing the internal lumen space and mucus phases respectively. The total flow rate is distributed proportionally into lumen and mucus with a fixed ratio dependent on their volume ratio. The diffusion of SCFA from the mucus phase to blood circulation was determined by diffusion coefficients using *in vitro* measurements in rats [24]. In the longitudinal direction, a constant flow was set to enter the colon each day, in agreement with colon receiving about 2L liquid per day [25-27]. Due to continuous mass transfer between the lumen, mucus, and BCS, the flow rate of each region kept changing over time, depending on both inflow flux and radial transport flux from lumen to mucus, which varied over time. The nutrients and metabolites in the mucus compartment are not only determined by transport with lumen compartment, but also determined by communication with BCS. The permeability constants were determined based on the fact that 10% of a semi-solid bolus ultimately reaches the rectum and forming the feces [28, 29].

Spatial compartmentalized framework formulation

The system governing equations is the same for each compartment, and here therefore just shown for one region. Considering compartment k for any the four regions (Fig. 1c and Fig. S7), it comprises two rows of interconnected ideal continuous stirred tanks and connected to a third ideal continuous stirred tank representing the blood compartment. The smaller tank ($v_k$) in the first-row structure represents the mucus phase where residing microbiota is attached to the mucus layer, thus protected against detachment by flow shear stresses. $j_{nk}$ is the mass flux of component $n$ transferred from $v_k$ to $V_B$ by perfusion through the wall membrane. $J_{nk}$ is the net mass flux interchange of component $n$ between lumen ($V_k$) and mucus ($v_k$) due to radial concentration gradients between compartments in lumen and mucus. In the second row at the bottom, each tank ($V_k$) represents the lumen region, and conveys most of the material bulk flow, and it is here where axial back-mixing takes place as well as bacterial wash out. $J_{Wk}$ is the net amount of water flux transferred from $V_k$ to $v_k$, and similarly $j_{Wk}$ is the amount of water absorbed through the wall membrane. $F_{k-1}$ and $F_k$, and $f_{k-1}$ and $f_k$ are to-and-from bulk flows for tanks $V_k$ *and* $v_k$. The back flow due to axial back-mixing in the colon is $q_k$ and has same value for all $k$. Finally, for the single tank representing the blood compartment for BCS, $F_{B0}$ and $F_B$ are flows to-and-from this compartment. Different species of the microbiota have different residence times in compartments $V_k$ *and* $v_k$ where the corresponding reactions occur. We assume that there is no change in the value of the density of the medium in lumen and mucus ($\rho_k = \rho$), for all compartments. Taken together, the four regions of the colon are axially connected and the main

flow goes from the ascending colon towards the sigmoid colon, with equal amounts of back-mixing ($q_k$) of mass from tanks $V_k$ to $V_{k-1}$. The ninth region is a fed-batch reactor storing mass leftover due to absorptions of metabolites and water. It only opens at fixed time intervals when defecation is called for.

More specifically, we take the *k-th* region comprising lumen and mucus compartments, i.e., with tank volumes of [$V_k$, $v_k$]. The two regions have different dynamic features. The mass balances for the lumen region, committed to the conditions of constant volume and same back-mixing extent between contiguous tanks, thus

$$\rho\frac{dV_k}{dt} = \rho(F_{k-1} - F_k) - V_k[J_{Tk} + J_{Wk}] + \rho(q_{k+1} - q_k) = 0 \tag{16}$$

Since,

$$q_{k+1} = q_k = q \tag{17}$$

And for all k,

$$F_k = F_{k-1} - \frac{V_k}{\rho}[J_{Tk} + J_{Wk}] \tag{18}$$

Where $J_{Tk}$ is the mass flow of all species *n* transferred from $V_k$ to $v_k$

$$J_{Tk} = \sum_{n=1}^{NC} k_m (C_{Lk}^n - C_{Mk}^n) M_n \tag{19}$$

In order to calculate the volume of the mass transport, the concentrations with unit of mM is converted to masses based on the molar weight of metabolite denoted by $M_n$. Parameter $k_m$ is a mass transfer coefficient (1/h). And this equation gives the net mass of metabolite *n* to be transferred from lumen towards the mucus.

Water absorption in corresponding region is determined by according to [29], where 10% is estimated as the achieved concentration of outer space, thus:

$$J_{Wk} = k_m (C_{Lk}^w - 0.1 * C_{Lk}^w) M_W \tag{20}$$

Similarly, in the mucus phase for any *k* of compartments, the mass balance equation is as following,

$$\rho\frac{dv_k}{dt} = \rho(f_{k-1} - f_k) + V_k(J_{Tk} + J_{Wk}) - v_k(j_{Tk} + j_{Wk}) = 0 \tag{21}$$

For any k

$$f_k = f_{k-1} + \frac{1}{\rho}[V_k(J_{Tk} + J_{Wk}) - v_k(j_{Tk} + j_{Wk})] \tag{22}$$

Where

$$j_{Tk} = \sum_{n=1}^{NC} \pi_n a_v (C_{Mk}^n - C_B^n) M_n \tag{23}$$

Where $\pi_n$ (cm/h) represents the permeability coefficient for component *n*, and $a_v$ is a surface area per unit volume at the membrane-wall for permeation.

For the mucus compartment, the volume of each mucus tank is assumed as constant,

$$f_k = f_{k-1} - v_k \frac{j_{Tk}}{\rho} - v_k \frac{j_{Wk}}{\rho} + V_k \frac{J_{Tk}}{\rho} + V_k \frac{J_{Wk}}{\rho} \tag{24}$$

Therefore, the outflow in mucus compartment could be written as,

$$f_k = f_{k-1} - \frac{v_k}{\rho} \sum_{n=1}^{NC} \pi_n a_v (C_{Mk}^n - C_B^n) M_n + \left( \frac{V_k}{\rho} \sum_{n=1}^{NC} k_m (C_{Lk}^n - C_{Mk}^n) M_n + J_{Wk} \right) - \frac{v_k}{\rho} j_{Wk} \tag{25}$$

The mass balance equation of microbiota in mucus is affected by detachement and intrinsic dynamics due to emergent properties of microbial ecosystem,

$$\frac{d}{dt} \Phi_{Mk}^J = \sum_{J=1}^{N} Y_{i,J} (\mathbf{r_M})_{Mk}^J \Phi_{Mk}^J - k_d \Phi_{Mk}^J \tag{26}$$

The metabolites in mucus compartment are determined mainly by mass transport. The dynamic changes of the complex carbohydrate could be formulated as following,

$$\frac{dz_{Mk}}{dt} = \frac{1}{v_k} [f_{k-1} z_{M,k-1} - f_k z_{Mk}] + \sum_{J=1}^{N} (\mathbf{S_x Z})^J (\mathbf{r_M})_{Mk}^J \Phi_{Mk}^J - \sum_{J=1}^{N} (\mathbf{r_h})_{Mk}^J$$
$$+ \frac{V_k}{v_k} k_m (z_{Lk} - z_{Mk}) - \pi_n a_v (z_{Mk} - z_B) \tag{27}$$

The mass balance for hexose in the mucus compartment is described by following equation,

$$\frac{d}{dt} \xi_{hex,Mk} = \frac{1}{v_k} [f_{k-1} \xi_{hex,,k-1} - f_k \xi_{hex,Mk}] + \sum_{J=1}^{N} (\mathbf{S_x Z})^J (\mathbf{r_M})_{Mk}^J \Phi_{Mk}^J$$
$$+ Y_{Hex,z} \sum_{J=1}^{N} (\mathbf{r_h})_{Mk}^J + \frac{V_k}{v_k} k_m (\xi_{hex,Lk} - \xi_{hex,Mk}) - \pi_n a_v (\xi_{hex,Mk} - \xi_{hex,B}) \tag{28}$$

The mass balance for all other metabolites in this compartment is described by following equation,

$$\frac{d\xi_{Mk}^i}{dt} = \frac{1}{v_k} \left( f_{k-1} \xi_{Mk-1}^i - f_k \xi_{Mk}^i \right) + \sum_{J=1}^{N} (\mathbf{S_x Z})^J (\mathbf{r_M})_{Mk}^J \Phi_{Mk}^J + \frac{V_k}{v_k} k_m (\xi_{Lk}^i - \xi_{Mk}^i) - \pi_n a_v (\xi_{Mk}^i - \xi_B^i) \tag{29}$$

In the rectum ($V_5$), the feces would be accumulated and stored until defecation occurring, they are not removed, i.e., during a very fixed period of time (three times/day during breastmilk-feeding period, and one time/day during solid food-feeding period). The mass balance equation for the rectum satisfies (recalling that the volume is constant),

$$\frac{dV_{Feces}}{dt} = F_4 + f_4 - F_5 \int_t \psi(t_f - t_i) dt \tag{30}$$

where $F_5$ denotes the outflow of rectum during defecation process as following,

$$\psi(t_f - t_i) = \begin{cases} 1 & t_i < t \le t_f \\ 0 & \text{otherwise} \end{cases} \tag{31}$$

It should be noticed that the volume of rectum ($V_5$) is not the same as the volume of feces $V_{Feces}$ (rectum content), which is the reason that we differentiate two terms in above equation.

As the mass of bacteria and metabolites are accumulated in the rectum, we need to know the concentrations of these state variables, which however cannot be obtained from the mass balance in $V_5$, because it behaves as a fed-batch reactor where microbial reactions still occur.

The concentrations of the state variables in the rectum is deduced from the following mass balance equation:

$$\frac{d}{dt}(V_{Feces}\varphi_5^J) = V_{Feces}\frac{d\varphi_5^J}{dt} + \varphi_5^J\frac{dV_{Feces}}{dt} = F_4\varphi_{L4}^J - F_5\psi(t)\varphi_5^J + V_{Feces}\sum_{J=1}^{N}Y_{i,J}\left(\mathbf{r_M}\right)_5^J\varphi_5^J \tag{32}$$

Solving this equation gives,

$$\frac{d\varphi_5^J}{dt} = \frac{1}{V_{Feces}}[F_4(\varphi_{L4}^J - \varphi_5^J) - f_4\varphi_5^J] + \sum_{J=1}^{N}\left(\mathbf{S_x Z}\right)^J\left(\mathbf{r_M}\right)_5^J\varphi_5^J \tag{33}$$

The mass balance for hexose in the rectum is described by following equation,

$$\frac{d\xi_{hex,5}}{dt} = \frac{1}{V_{Feces}}[F_4(\xi_{hex,L4} - \xi_{hex,5}) + f_4(\xi_{hex,M4} - \xi_{hex,5})] + \sum_{J=1}^{N}\left(\mathbf{S_x Z}\right)^J\left(\mathbf{r_M}\right)_5^J\varphi_{M5}^J \tag{34}$$

The mass balance for all other metabolites in the BCS is described by following equation,

$$\frac{d\xi_5^i}{dt} = \frac{1}{V_{Feces}}[F_4(\xi_{L4} - \xi_5) + f_4(\xi_{M4} - \xi_5)] + \sum_{J=1}^{N}\left(\mathbf{S_x Z}\right)^J\left(\mathbf{r_M}\right)_5^J\varphi_5^J \tag{35}$$

The volume of the blood circulatory system $V_B$ is also assumed to be constant, and metabolites from all mucus compartment ($v_k$) are collected into $V_B$ through wall-membrane permeability and are carried away by the portal vein. For the plasma compartment, we formulated the following equations to describe each metabolite that is collected to portal vein,

$$\frac{d}{dt}\xi_B^i = \frac{1}{V_B}[F_{B0}\xi_{B0}^i - F_B\xi_B^i + \sum_{k=1}^{Nk}\sum_{i=1}^{NP}v_k\pi_i\left(\xi_{Mk}^i - \xi_B^i\right)] \tag{36}$$

**Method 4 | Evaluation of dynamic bacterial models captured by enzyme-centric metabolic framework (ECMF)**

*Smaller-scale gut microbial ecosystem*

CODY is not limited to include only a certain number of species in the gut microbiota ecosystem, but for demonstration, we constructed a smaller-scale gut microbial ecosystem composed of the representative species the human gut microbiota in this study. In order to ensure the reasonability and reliability of the prediction results based on this constructed ecosystem, a standard principle comprised of three aspects are adopted, which can be schematically illustrated in Fig. S1b (step 1) and described in detail as below:

i)  The selected species not only represent the most dominant species, but also represent the three dominant phyla of the human gut microbiota since they resemble the proportion of the three major phyla of original community (Fig. S13). Specifically, the abundance of the smaller-scale ecosystem accounts for 20-50% of the three major phyla in the original microbial community.

ii)  The selected species represent the major metabolic capacity of gut microbiota with regard to producing the key human-accessible health-promoting metabolites, i.e., all three short-chain fatty acids: acetate, propionate, and butyrate, and other microbial associated metabolites, including succinate, formate, lactate, and so on.

iii)  Furthermore, in order to ensure CODY predictions on the microbial ecosystem dynamics comparable to the original community, the effect of other community members should be evaluated as well. This was achieved by scaling the prediction results using the sum proportion resembled by the smaller scale ecosystem to the original community.

Furthermore, *in vitro* fermentation experimental data are available for these representatives. Since monoculture fermentation data of *Eubacterium rectale* were not available in the literature, it was substituted by *Eubacteria hallii* that is phylogenetically closely related and has been demonstrated to have a similar metabolic phenotype as *Eubacteria rectale*. The most abundant twenty species at all three stages were analyzed and compared (Fig. S13). The three dominant phyla were denoted by actinobacteria, bacteroidetes, firmicutes in the infant cohort across all three stages. The extent to which the ecosystem represents the whole microbiota at each stage is summarized as the coverage number, i.e., 19.6%, 35.4%, and 8.3%, respectively. Comparatively, the coverage of the three dominant phyla is summarized in Fig. S13b, i.e., 59.8%, 63.7%, and 55.7% respectively. Moreover, the composition of our ecosystem resembles that of the three dominant phyla in the original community very well, thus indicating that our ecosystem can represent the systematic behavior of the original community (Fig. S13c-d).

*ECMF construction*

ECMF is composed by two major components, metabolism and regulation process. The ECMF for each species is constructed following the pipeline as shown in Fig. 1a and Fig. S1b. The metabolism of each of the representatives was derived in two steps. In the first step, backbone metabolic networks that is taken from published literatures [6, 10-12, 30] of corresponding species was employed and incorporating signature pathways to characterize the repertoire of biochemical reactions. More specifically, the metabolism of Bacteroides (*B. thetaiotaomicron* and *B. fragilis*) relies upon the published metabolic network [9] and integrating a primitive electron transfer chain using fumarate as the terminal electron acceptor[6]. Metabolic networks for Bifidobacterium (*B.*

*longum*, *B. breve* and *B. adolescentis*) were referred to the published report [10, 11], and for the Firmicutes species the metabolic networks of *F. prausnitzii* and *R. intestinalis* were taken from the literature [12] and the metabolic network of *E. hallii* was assumed to be similar to that of *E. rectale* for which there is a published network available [10]. Through applying convex analysis, the metabolic network of each species was decomposed into their metabolic pathway modules (MPMs)[31]. METATOOL v5.1[32] was employed to firstly decompose the biochemical stoichiometric structure into nonnegative combinations of all possible pathways. Instead of searching for an optimal solution by maximizing biomass yield from given uptake flux such as FBA, all feasible solutions in the null space of stoichiometric matrix were regarded as pathway representatives that denote the fundamental structural units of the metabolic network. The obtained pathways are a set of thermodynamic feasible metabolic pathways by which all the possible metabolic states can be completely described and any feasible flux vector at steady state can be represented by their linear combinations.

In the next step, to avoid overparameterization, the large number of pathways was reduced to an active subset of metabolic pathway modules, by which the metabolic features of each species were conservatively characterized. By doing this, the number of kinetic parameters is thus much less than that of *in vitro* experimental data points, which eliminates the potential overfitting. The yields of the key metabolites that are the products of extracellular metabolites in our studies, i.e., acetate, butyrate, propionate, succinate and lactate were taken into account based on *in vitro* experimental data [33-38]. The detailed process is described in Method 1(*SI Appendix*), by using metabolic yield analysis [14]. Accordingly, ECMF for the representatives characterized their signature metabolic capacities. For example, *B. thetaiotaomicron* and *B. fragilis* both from Bacteroidetes phylum are well known to produce propionate from succinate; *B. longum* and *B. adolescentis* possess the common F6P-shunt pathways and generate lactate and acetate as the major products[39]. Through applying the metabolic network associated decomposition and yield analysis techniques, the metabolic capacity of individual bacterium can be furthest reserved and well represented (>99%), which can then ensure its accuracy in characterizing the dynamics of individual bacterium metabolism. By viewing each pathway as one lumped biochemical route that converts the substrates to products, there is no need to explicitly calculate kinetic parameters for the intracellular metabolites. The uniquely derived subsets of MPMs, that incorporates structural information of metabolism, are viewed as subunits that can be activated by cellular regulation of the corresponding enzymes. In ECMF, the dynamic uptake flux of individual bacterium at each time point is viewed as being optimally assigned among MPMs in a regulated pattern. This is achieved by introducing two cybernetic variables, i.e., *u* and *v,* to depict the cellular regulation mechanism. These two variables are formulated based on matching law and proportional law [15], in which the "*u*" is responsible for control on enzyme synthesis and "*v*" is responsible for control on enzyme activity. By doing this, cellular regulation can be achieved by preferentially selecting alternative MPMs dependent on the context-specific substrates. The MPMs for all representatives and their stoichiometric relationship are all shown in Dataset S9. The regulation machinery was integrated and the ECMF models were formulated following the ECMF construction workflow as described in Method 1(*SI Appendix*), following Eq. 1-10. Besides, the sensitivity analysis is performed associated with ECMF framework, for individual bacterium. The corresponding results indicate that ECMF simulation results are relative invariant to small variations of the kinetic parameters ($\pm 5\%$) (Fig. S3). Furthermore, another evidence that there is no overfitting here could be seen from the validation process of HRAF. On the other hand, the exclusion of overfitting issue

can be further demonstrated by the evaluation process of HRAF. During this process, the dynamic and metabolic behavior together with the interspecies interactions of the two-species consortium are purely predicted by their kinetic parameters that were identified during the evaluation process of their ECMF construction for monoculture, which shows well agreement with the coculture *in vitro* fermentation dataset. Furthermore, the overfitting issue can be avoided as evidenced by the consistency between pure prediction results of the integrative CODY framework and the metagenomics measurements, where three component frameworks are integrated in a bottom-up fashion to predict a wholly new dataset. Finally, as an extended demonstration strategy, we successfully deployed our CODY to be a visualized and public online application toolbox. This online toolbox provided a random combination of diet strategy upon user's choice. It enables to perform corresponding simulations in response to user's design of different diet combinations, and provides the spatiotemporal microbial/metabolite profiles as the output.

## *ECMF evaluation*

The evaluation of ECMF for representatives was performed by comparing model simulation results and *in vitro* experimental fermentation data [30, 33-37, 40] (Fig. S2). We obtained data from the literature using the PlotDigitizer software. Applying the experimental data at t=0 hr as the model input, the ECMF model would simulate the sequential profiles for both bacterial growth and metabolite production, based on the system governing equations listed in ECMF formulation section in Method 1(*SI Appendix*). Kinetic parameters could be identified for each microbiota representative, through parameter estimation by comparing *in silico* model prediction to the *in vitro* literature fermentation data. For bacterial growth, the standard unit in the dynamic model was gram DW/Liter. To convert the *in vitro* biomass data for specific bacteria with unit of log CFU/mL to the data with unit of gram DW/Liter, the data was transformed by multiplying the cellular wet weight and the coefficient between dry weight and wet weight. For this we assumed that this coefficient was the same as that of *E. coli* [41]. The *in vitro* experimental data for each species were obtained from literature, including *B. thetaitaomicron*, *B. fragilis*, *B. longum*, *B. adolescentis in vitro* experiments with growth on hexose, as well as *E. hallii* growing under both acetate and lactate as substrates [25-29]. For *B. breve*, *F. prausnitzii* and *R. intestinalis*, the *in vitro* experimental fermentation data were taken from the literature [34]. The parameters related with enzyme balance equations were set to be their typical values for each species. A parameter sensitivity analysis was performed as shown in Fig. S3a. It can be seen that most kinetic parameters of the MPMs are not sensitive by perturbing within the range of $\pm$ 5%. However, for *B. adolescentis* and *B. fragilis*, there are a few parameters which are relatively sensitive, given by the relatively large RMSE changes impacted by variations of these parameters. This is due to that the associated MPMs are the major metabolic modules through which bacteria grow or produce the concerned metabolic products, i.e., acetate, succinate, and the corresponding changes can be seen in Fig. S3b.

## *In vitro fermentation experimental data for individual bacterium*

The evaluation of ECMF for each species was performed by comparing model simulation results and *in vitro* experimental data. For *in vitro* experimental validation of *B. thetaiotaomicron* and *B.fragilis*, the fermentations were performed in a complex medium, allowing good growth of human colon bacteria when supplemented with an appropriate energy source[33], containing (in

grams per liter): bacteriological peptone (Oxoid), 6.5; soy peptone (Oxoid), 5.0; tryptone (Oxoid), 2.5; yeast extract (VWR International, Darmstadt, Germany), 3.0; KCl, 2.0; NaHCO$_3$, 0.2; NaCl, 4.5; MgSO$_4$ · 7H$_2$O, 0.5; CaCl$_2$ · 2H$_2$O, 0.45; MnSO$_4$ · H$_2$O, 0.2; FeSO$_4$ · 7H$_2$O, 0.005; ZnSO$_4$ · 7H$_2$O, 0.005; cysteine-HCl, 0.4; hemin, 0.005; menadion, 0.005. The medium also contained H$_3$PO$_4$, 0.5 ml liter$^{-1}$ and Tween 80, 2 ml liter$^{-1}$. The pH of the medium was adjusted to 5.80 before sterilization (210 kPa, 121°C, 20 min). Fructose was used as the sole energy source, at a concentration of 15 g liter$^{-1}$, sterilized separately and aseptically added to the fermentation medium. Fructose was sterilized in an autoclave (210 kPa, 121°C, 20 min). For *in vitro* experimental validation of *B. longum*, the metabolism of this species was first examined using small-scale fermentations (100 ml) in MCB with glucose as the sole added energy source and without pH control. Initially, the strains were transferred from storage at 80°C to de Man-Rogosa-Sharpe medium and incubated anaerobically at 37°C in a modular atmosphere-controlled system (MG anaerobic work station; Don Withley Scientific, West Yorkshire, United Kingdom) that was continuously sparged with a mixture of 80% N$_2$, 10% CO$_2$, and 10% H$_2$ (Air Liquide, Paris, France). After this, the strains were propagated twice in MCB with glucose as the sole added energy source and finally added (5% [vol/vol]) to 100 ml of MCB with glucose as the sole added energy source. After 24 h of anaerobic incubation at 37°C, colony counts were obtained by plating the strains on MCB agar with glucose, and the final pH of the medium was determined. Samples were also removed for further analysis.

For *in vitro* experimental validation of *B.breve*, the detailed fermentation process could be found in literature data[34]. For the *in vitro* experimental validation of *E. hallii*, the detailed fermentation process could be found in literature data [37]. The detailed processes are: Eubacterium hallii L2-7 was isolated from a 2-year-old healthy infant and grown in YCFA medium containing (per 100 mL) 0.1 g casitone, 0.25 g yeast extract, 0.4 g NaHCO$_3$, 0.1 g cysteine, 0.045 g K$_2$HPO$_4$, 0.045 g KH$_2$PO$_4$, 0.09 g NaCl, 0.009 g MgSO$_4$·7H$_2$O, 0.009 g CaCl$_2$, 0.1 mg resazurin, 1 mg haemin, 1 mg biotin, 1 mg cobalamin, 3 mg paminobenzoic acid, 5 mg folic acid and 15 mg pyridoxamine. Final concentrations of the SCFA in the medium were 46mM acetate, 9mM propionate and 1mM each of isobutyrate, iso-valerate and valerate. The medium was adjusted to pH 6.7 and placed in Hungate tubes that were flushed with CO$_2$ and heat-sterilized. Heat-labile vitamins were added after the medium was autoclaved to give a final concentration of 0.05 $\mu$g mL$^{-1}$ thiamine and 0.05 $\mu$g mL$^{-1}$ riboflavin. The glucose-supplemented medium as a carbon source contained a final concentration of 10mM. Each test bacterial strain was inoculated into triplicate tubes and growth measured spectrophotometrically as A650nm. The medium was supplemented with 35mM DL-lactate (pH 6.7±0.1), of which approximately 16 mM were in the L-lactate form. The *in vitro* experiments were carried out in Hungate tubes under anaerobic conditions. Fermentation was monitored by synchronous measurements of the OD650nm and concentrations of glucose, lactate (DL and L), acetate and butyrate. Concentrations of L-lactate and glucose were determined by enzymatic methods. Concentrations of SCFA and DL-lactate were determined by capillary GC. D-Lactate was estimated as the difference between total lactate and L-lactate concentrations. For *in vitro* experimental validation of *F. prausnitzii* and *R. intestinalis*, the detailed fermentation process could be found in literature [30]. The detailed processes are: the strains studied were originally isolated from anaerobic roll tubes with M2GSC medium containing 30% clarified rumen fluid. All seven strains were, however, able to grow in yeast extract-Casitone-fatty acids (YCFA) medium lacking rumen fluid provided that a mixture of sources of carbon and fatty acids (SCFA) (giving final concentrations of 33 mM acetate, 9 mM propionate, 1.2 mM iso-butyrate, 1.0 mM iso-valerate, and 1.0 mM valerate) was added. Medium YCFAGSC contains the

following ingredients (per 100 ml): 1 g of Casitone; 0.25 g of yeast extract; 0.2 g each of glucose, starch, and cellobiose; 0.4 g of $NaHCO_3$; 0.1 g of cysteine; 0.045 g of $K_2HPO_4$; 0.045 g of $K_2HPO_4$; 0.09 g of NaCl; 0.009 g of $MgSO_4 \cdot 7H_2O$; 0.009 g of $CaCl_2$; 0.1 mg of resazurin; 1 mg of hemin; 1 µg of biotin; 1 µg of cobalamin; 3 µg of p-amino benzoic acid; 5 µg of folic acid; 15 µg of pyridoxamine; 5 µg of thiamine; and 5 µg of riboflavin. Thiamine and riboflavin were added to the medium as filter-sterilized solutions after being autoclaved (121°C, 15 min). Media were prepared and maintained under $O_2$-free $CO_2$.

For the evaluation of both monoculture and coculture for *B. thetaiotaomicron* and *B. adolescentis*, the composition of the base medium per 100 ml, is as follows: 0.25 g yeast extract (BD), 1.0 g casitone (BD), 0.4 g $NaHCO_3$ (Merck), 0.045 g $K_2HPO_4$ (Merck), 0.045 g $KH_2PO_4$ (Merck), 0.09 g NaCl (Merck), 0.09 g $(NH_4)_2SO_4$ (Merck), 0.009 g $MgSO_4 \cdot 7H_2O$ (Merck), 0.009 g $CaCl_2$ (Merck), 0.1 mg resazurin (Sigma-Aldrich), and 1 mg hemin (Sigma-Aldrich). In addition, the final concentrations of following short-chain fatty acids (SCFA) were included (final concentrations): acetate (33 mM); propionate (9 mM); isobutyrate, isovalerate, and valerate (1mM each). Finally, the volume was adjusted with $dH_2O$ in a conical Erlenmeyer flask. The medium was boiled in a microwave to dissolve the complex compounds. After the medium cools down, it was supplemented with cysteine (0.1 g), followed by boiling under oxygen-free $CO_2$ atmosphere for 3±4 minutes. After autoclaving at 120 °C for 15 min, filter sterilized solutions of vitamins (1 µg biotin, 1 µg cobalamin, 3 µg p-aminobenzoic acid, 5 µg folic acid, 15 µg pyridoxamine, 5 µg thiamine and 5 µg riboflavin) per 100 ml of medium. All SCFAs and vitamins were purchased from Sigma-Aldrich. The final pH of medium was maintained to 7.2 ± 0.1. In YCGMS medium, carbon sources such as glucose (G), maltose (M), cellobiose (C), and soluble starch (S) were added to a final concentration of 0.2% (wt/vol) of each. To encourage the growth of the fastidious species, simple sugars were added to the medium. This same media was used for the *B. adolescentis* L2-32 and *B. thetaiotaomicron* ATCC 29148 co-culture experiment. All other components were the same between the YCFA and YCGMS media.

**Method 5 | Evaluation of hierarchical resource allocation framework (HARF) by comparing predictions on community dynamic metabolic behavior and experimental data**

In order to evaluate whether HRAF enables to characterize the dynamic behavior of the gut ecosystem also to investigate how the gut microbiota interact with each other in the dynamic manner, we firstly constructed the ECMF models for the individual bacterium that constituted the two-species consortium, and further developed the HRAF of their coculture for predictions on the dynamic metabolic behavior of their consortia, following the HRAF framework pipeline. All associated parameters, including the kinetic parameters and the MPMs of individual bacterium within the two-species consortium were taken from that identified from their corresponding ECMF models (Table S5). The kinetic parameters in ECMF for each species were estimated by fitting model simulation results of growth and product profiles to the *in vitro* experimental data of each species [40]. The initial condition was set based on the *in vitro* coculture experimental data, including the substrate and products. There were no parameter fittings involved in these simulations, thus, the dynamic metabolic profiles of the consortia were purely predicted for growth on a mixture of three substrates, i.e., glucose, maltose and starch, with each being initially available at a concentration of 2g/L. The model predictions were simulated based on the system governing equations provided in Method 2(*SI Appendix*). The dynamic interactions of the consortia member could be indicated by the growth curve of two species based on the HRAF pure predictions on their growth profiles. We further evaluate the HRAF predictions by comparing to *in vitro* fermentation data of the consortia for validation [40]. The pure predictions of HRAF on both growth profiles and metabolite profiles were compared to *in vitro* fermentation data from the literature [40]. The coefficient of determination, i.e., $R^2$ was calculated to evaluate the fidelity of HRAF predictions.

**Method 6 | Evaluation of HRAF by predicting ecosystem-wide ecological interactions**

*In silico* community-models for all possible pairwise combinations (two species consortium), based on a collection of dynamic metabolic models describing our ecosystem representatives and their identified kinetic parameters, were systematically reconstructed and evaluated using a community-level dynamic model framework. Correspondingly, bacterial growth profiles in monoculture and coculture with other species were predicted for 28 pair-wise combinations. Analogous to the community-level model construction, we assumed context-dependent environmental resources were shared by any two species within each pair. The community medium was defined based on the growth requirements for the two consortium members within each pair. To ensure reliable comparison at the system-level, substrates that were utilized by all species were standardized to the same level for both pure culture and co-culture simulations, among all combinations. The kinetic parameters for each species applied in the community models were the same as those identified based on its individual monoculture. For each species, its growth capacity in monoculture was used as the baseline, upon which its growth capacity within pairwise coculture with other species was compared and used to calculate the interaction index (Eq. 35). Four types of interactions were hereby defined based on the interaction index with similar concept to previous studies [42, 43], i.e., negative, where two species compete for the shared resources (competition, -/-); positive, where the metabolites produced by one species are required by another (cooperation, +/+); unequal, including positive/neutral (commensalism, +/0), negative/neutral (amensalism, -/0), or positive/negative (exploitation, +/-) where coculture produced a synergic co-growth benefit; and neutral, where coculture has no net effect. The pairwise interaction index for the *i-th* combination was defined by fold changes between growth capacity in pairwise combinations and the baseline level for each species, i.e.

$$FC_i = \log\left(\frac{\mu_{\max, pair}}{\mu_{\max, pure}}\right) \tag{37}$$

**Method 7 | Evaluation of the biomimetic model described by the spatial compartmentalized framework (SPCF)**

We evaluated whether SPCF is able to correctly predict the colonic physiology by comparing model simulations with the *in vitro* experimental setup, TNO intestine model (TIM), that has been developed since 1998 and successfully used to simulate many different colon experiments[44]. We evaluated SPCF in two parts. First, we evaluated SPCF's ability to characterize the colon physiology, including transit time and water absorption. In this part, with the same inflow rate and volume settings as TIM, SPCF was used to predict the spatiotemporal profiles under multiple conditions with different nutrition signals as model input (Figs. S8-10). SPCF showed consistent predictions on the outflow rate and transit time (Table S6), as compared to *in vitro* observations of TIM. Besides, our model enables the spatiotemporal profiles at uninterrupted time-scale which cannot be achieved by TIM. Rather, TIM only measures the output at the defined sampling outlet. We further performed assessment of SPCF's SCFA production with input of four complex carbohydrates (0.25g for each), by assuming a pseudo carbohydrate degradation reaction in each compartment of SPCF. The hydrolysis parameters were determined by fitting SCFA yields under each carbohydrate (Table S6). Due to the spatiotemporal nature of SPCF predictions, to make our predictions comparable with the TIM time-course total output profiles, we calculated the time-specific SCFA production by taking total account of *in vivo* site-specific SCFA (including ascending lumen, transverse lumen, descending lumen and sigmoid lumen, ascending mucus, transverse mucus, descending mucus and sigmoid mucus), blood SCFA and fecal SCFA. The resulting time-coursed profiles of both SCFA production and composition under each carbohydrate were compared to the TIM *in vitro* observations (Figs. S11-12).

## Method 8 | Evaluation of CODY

CODY formulations

By integrating the intrinsic dynamic of gut microbiota that is described by ECMF and HRAF, into the extrinsic dynamics of the gut microbiota that is the colon environment described by SPCF, CODY could be assembled and formulated by integrating the impacts of species metabolites from both microbiota and the colon. The mass balance equations for the bacteria and metabolite state variables could be described by the following governing equations:

$$\dot{\boldsymbol{\varphi}} = \delta(\boldsymbol{\varphi}, F, k, t) + \Gamma(\boldsymbol{\varphi}, f, k, t) + \Upsilon(\boldsymbol{\varphi}, F, k, t) + \nabla(\boldsymbol{\Phi}, k, t) \tag{38}$$

$$\dot{\boldsymbol{\xi}} = \Omega(S^0, F, t) + \delta(\boldsymbol{\xi}, F, k, t) + \Gamma(\boldsymbol{\xi}, f, k, t) + \Upsilon(\boldsymbol{\xi}, F, t) + \Lambda(\nabla\boldsymbol{\xi}_{L/M}, k_m) + \Pi(\nabla\boldsymbol{\xi}_{M/B}, \pi_n) \tag{39}$$

where $\varphi$ denotes the microbiota profiles, and $\xi$ denotes the metabolite profiles. The mathematical model of microbiota profiles comprises four components, forward flow ($\Upsilon$), back flow ($\Gamma$), dynamic reaction metabolism($\Upsilon$), and detachement ($\nabla$); while that of metabolite comprises six components, besides the three components that are the same as the microbiota, the other three terms correspond to a source term ($\Omega$), mass transport term from lumen phase to mucus phase ($\Lambda$), and membrane permeation from mucus to blood ($\Pi$). The state variable vector is defined as $\boldsymbol{\varphi} = (\varphi_{bth}, \varphi_{bfr}, \varphi_{blg}, \varphi_{bbv}, \varphi_{bad})^{\mathrm{T}}$ for microbiota biomass of functional community members (for the breastmilk-feeding period microbial community), and $\boldsymbol{\xi} = (z, \xi_{hex}, \xi^{\mathrm{T}})^{\mathrm{T}}$ for all involved metabolites, where $z$ is the concentration of the complex polysaccharide accessible by microbiota, $\xi_{hex}$ is the hexose that was released by complex carbohydrates and further supported microbiota development. $\xi^{\mathrm{T}} = (\xi_{lac}, \xi_{ac}, \xi_{prop}, \xi_{succ}, \xi_{for}, \xi_{eth}, \xi_{but}, \xi_{H_2})^{\mathrm{T}}$ is the vector of concentration of other small molecules that are produced by the microbial ecosystem. Furthermore, according to the presented formulations for ECMF, HRAF and SPCF in Method 1-3(*SI Appendix*), the above system governing equations could be detailed as below for the *J-th* microbe and *i-th* metabolite in the *k-th* colon region by substituting each term, which gives:

$$\frac{\partial \varphi_k^J}{\partial t} = \left(F_{k-1}\varphi_{k-1}^J - F_k\varphi_k^J\right)\Big/V_k + q_{k+1}(\varphi_{k+1}^J - \varphi_k^J)/V_k + \left(\sum_{J=1}^N S_x^J Z^J r_k^J \varphi_k^J\right)_J + \nabla(k, \Phi_k^J) \tag{40}$$

$$\frac{\partial \xi_k^i}{\partial t} = \left(F_{k-1}\xi_{k-1}^i - F_k\xi_k^i\right)\Big/V_k + q_{k+1}(\xi_{k+1}^i - \xi_k^i)/V_k + \left(\sum_{J=1}^N S_x^J Z^J r_k^J \varphi_k^J\right)_i + k_m(\xi_{k,lumen}^i, \xi_{k,mucus}^i) \tag{41}$$

Two Longitudinal clinical studies used for demonstrations of CODY

We employed two cohorts to evaluate CODY's predictive power on gut microbiota reprogramming under diet-intervention. For the infant cohort, we adopted the time series infant gut microbiota consisting of 98 stool samples taken from Swedish healthy full-term breastfed infants. Fecal samples at three stages including newborn, aged 4th month and 12th month were involved in this study. Shotgun-sequenced samples were analyzed at the meta-OTU level. The metagenome sequence data are available at GigaDB (http://gigadb.org/) under the accession code ERP005989. For detailed description of this dataset see the published study [1]. By removing

the infants receiving formula-feeding at newborn and 4th month, we identified 68 breastfed infants that were included in our study. For the adult cohort, we adopted the Helsinki hepatic steatosis obese adult fecal samples before and after isocaloric carbohydrate-restriction diet intervention. Fecal samples available at 0 day at control condition, and 1, 3, 7 and 14 days after diet intervention were used to evaluate CODY predictions for this cohort. The relative abundance data of microbiota representatives were performed based on their shotgun metagenomics data. The fecal SCFA measurement were based on GC-MS detection. Details information could be referred to the reported literature [45].

CODY simulation process

The nutrient inputs were determined in order to mimic the real-life daily meals. Specific diet frequency and composition accessed by the microbiota were described by a pattern-specific multiple segmented step-function. The polysaccharides are the microbial resources from degrading complex carbohydrates, and the composition was determined for each period. For example, in the infant cohort, after birth, HMO, which is the third most abundant solid components in breastmilk, with an estimated amount of 14 g/d served as the main nutrient for the gut microbiota [46, 47]. 1 g/d mucin is estimated to be produced during this period [48]. Breastmilk is set being fed to infants by 6 times/day ($\alpha$ =3.5). Diet input during this period is shown in Eq. 40, whereby $S^0$ denotes for the polysaccharide concentration during this stage.

$$\text{Nutrient}_{\text{in}}(t) = \begin{cases} S^0 & \mod(t, 24) \in \left[ \alpha * n - 0.5, \alpha * n \right] \mid n \in [1, 6] \\ 0 & otherwise \end{cases} \quad (42)$$

After diet switching to solid food, the microbial accessible carbohydrates (MACs) became fiber and resistant starch, the amount of which is adjusted to half of the amount taken by adults [49]. Besides, 1.5 g/d mucin glycans are estimated to be produced during this period [48]. This is equivalent to 80g/L of carbohydrate contained in 0.3L/day of fluid which are allocated among three meals ($\alpha$ =5). The diet input during this period is shown in Eq. 41.

$$\text{Nutrient}_{\text{in}}(t) = \begin{cases} S^0 & \mod(t, 24) \in \left[ \alpha * n + 2, \alpha * n + 3 \right] \mid n \in [1, 3] \\ 0 & otherwise \end{cases} \quad (43)$$

The metabolic degradation capacity of representatives differentiated for specific carbohydrate, based on literature-based mining. *B.fragilis* and *B.thetaoitaomicron* are reported to access bovine mucins [50], *B.longum* is reported to possess certain enzymes degrading mucins[51]; *B.longum* is well-known in degrading HMO [52], *B.fragilis* and *B.adolescentis* shows moderate and weak capacity respectively to utilize certain structures of HMO [53], while *B.breve* hardly degrade HMO[54]. Dietary fibers coming from plant cell walls, with main constituents being hemicellulose, i.e., xylan, mannan, xyloglucan, beta-glucan and pectin, can be degraded by *B.fragilis*, *R.intestinalis*, and *F.prausnitzii*[55, 56]; *B.thetaiotaomicron* is reported to hardly access hemicelluloses and beta-glucans from fiber[57]; small portion of other fiber types including inulin (contained in onions, garlic and banana [56]), fructooligosaccharides could be degraded by *B.adolescentis* and *E.hallii* [36]. *B.fragilis*, *B.adolescentis*, *B.thetaiotaomicon* and *E.hallii*[58] are reported to access certain type of resistant starch [50, 56, 59, 60]. Consequently, the degradation capacity of all involved carbohydrate forms by each consortia member were quantitatively summarized (Table S4). All forms of complex carbohydrates were assumed to be degraded to the same extent of hexose that is accessible to

23

all species that require hexose for growth. The degradation capacity of each species was determined by their collective ability to degrade all forms of nutrients involved within the investigated period.

For the obese adult cohort, the MACs were determined based on what was reported in the study. The amount of the constitutes, i.e., fiber and resistant starch, was based on the total insoluble fiber and resistant starch before and after carbohydrate-restriction [45] (Table S4). The diet input during this period was the same as shown in Eq. 41. All the parameters used for CODY are shown in Table S6.

Defecation process was described by a Dirac function in Eq. 42 to represent that rectum would be emptied periodically, with frequency of three times/day during breastmilk-feeding period ($t_c = 8, 16, 24$), and one time/day during solid food feeding period ($t_c = 12$). Before the defecation process, all wastes containing bacteria, undigested molecules and water would accumulate in the rectum continuously.

$$\psi(t_f - t_i) = \begin{cases} 1 & [\mathrm{mod}(t, 24) - t_c] \in [t_i, t_f] \\ 0 & otherwise \end{cases} \tag{44}$$

CODY enables predictions on both *in vivo* microbiota reprogramming and *in vitro* microbiota reprogramming reflected in the feces, employing the initial microbiota composition as the baseline condition. The newly achieved stable composition was employed as the initial conditions for the subsequent simulation by using the diet-intervention strategy as the diet input. We followed this workflow to predict the spatiotemporal microbial and metabolite profiles for both infant and adult cohort studies.

To make our prediction on the feces comparable to the *in vitro* feces metagenomics measurements, we converted the microbial density profiles at the post-stabilized state to the relative microbiota composition thus enabling comparison between model predictions and metagenomics measurements. The transformation was done according to the following equation,

$$RA^J = \frac{(x^J)_f}{\sum_{J=1}^{N} (x^J)_f} \times (RA)_{total} \tag{45}$$

Where $(x^J)_f$ denotes the concentration of the bacteria in the final time of the dynamic trajectory, and $(RA)_{total}$ denotes the total abundance of the ecosystem members with respect to the original gut microbiota.

**Method 9 | PCA for dynamic microbial landscape development**

In order to evaluate the phenotypic difference in the dynamic process of microbial growth between the lumen phase and the mucus layer we used the HCTSA toolbox to analyze the microbial time-series data [61, 62]. The analysis was performed as suggested by the toolbox. Based on system-level comparison of all microbial temporal trajectories, we firstly identified related features that characterize the time-evolved data of microbial growth profiles in the lumen and mucus. 7749 features were encapsulated based on the model libraries that have been evaluated previously by the HCTSA package. Each feature captured one or more specific properties of the time-series data structure (i.e., autocorrelation, entropy, etc.). A filtering step was used to remove features related with location or length of the data structure, as well as those features that have zero-variances. Each feature was scored based on its ability to classify the three labeled groups, i.e., lumen, mucus and feces samples. This step was done by training a one-point split identification tree and calculating the accuracy score using following equation:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{46}$$

where TP, TN, FN, FP are the number of true positives, true negatives, false positives, false negatives, respectively.

A further non-redundancy step was performed. In this step, all features were firstly ranked based on their accuracy score. Next, the representative feature (with the highest accuracy score) was chosen for one distinct feature cluster, remaining features with high Pearson's correlation coefficient (>0.9) with the representative features grouped together, or otherwise set as a new representative feature of the next cluster. Following this, we obtained all the feature clusters. The top ten representative features with the highest accuracy were used for the PCA analysis in Fig. 4f. The correlation coefficients among the selected features are shown in Fig. S25, illustrating that the obtained features are relatively independent from each other. Low correlation coefficients between each other indicate that the obtained features are relatively independent with each other. The top ten representative features are summarized in Table S8.
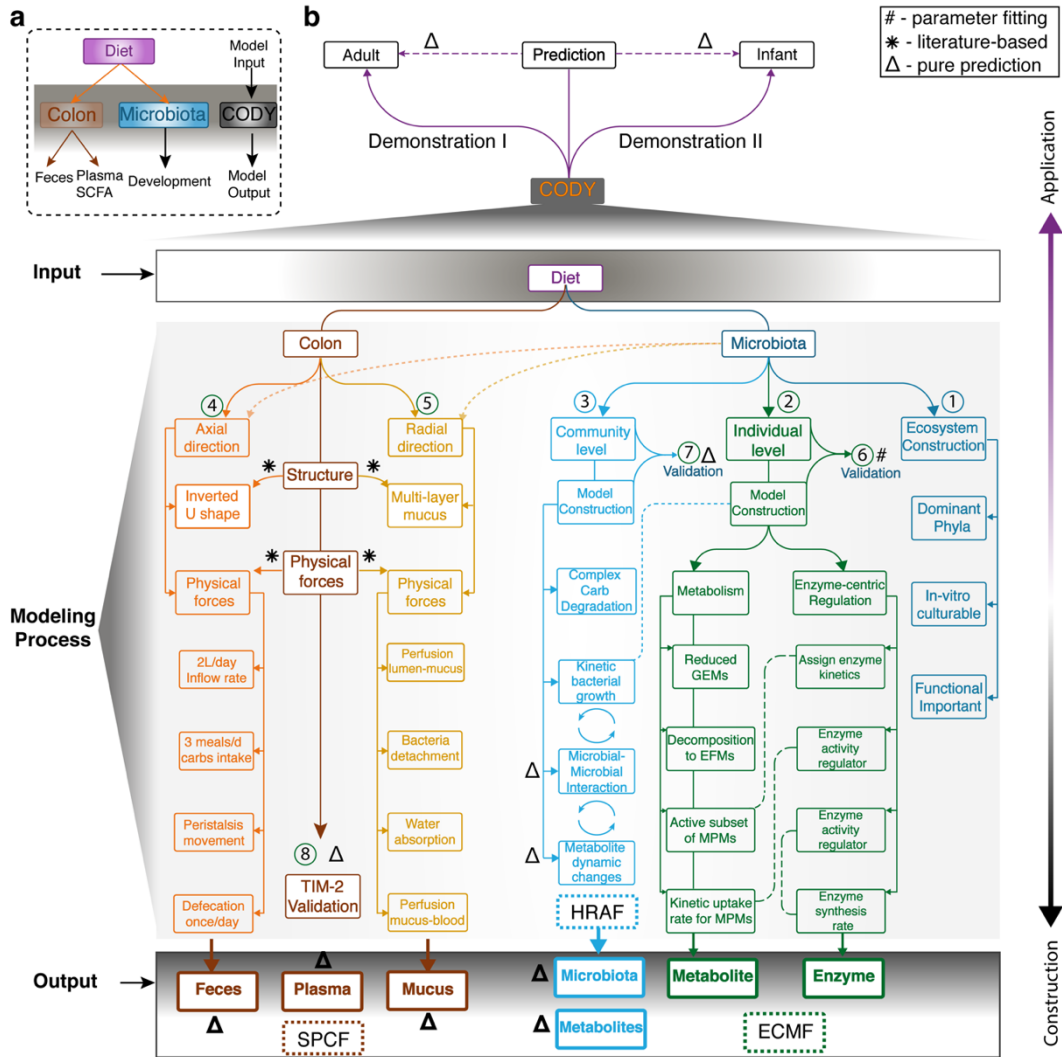
# Supplementary Figures



**Fig. S1 | Step-by-step workflow of the CODY platform and associated applications.** (**a**) Schematic overview of the different components considered in the CODY platform. (**b**) Detailed step-by-step workflow of the CODY construction process. The symbol **"#"** denotes that the associated model was defined using parameter fitting to *in vitro* experimental data; The symbol **"*"** denotes that the model was reconstructed based on literature mining; The symbol **"Δ"** denotes that the model was evaluated through pure model predictions, i.e. without reliance on fitting or training process. 1-ecosystem members determination; 2-enzyme-centric metabolic framework (ECMF) construction process; 3- construction of the community level dynamic microbial model using the hierarchical resource allocation framework (HRAF); 4-spatial compartmentalized framework (SPCF) construction in the axial direction; 5-spatial compartmentalized framework (SPCF) construction in the radial direction. 6- Evaluation of the dynamic mono culture microbial model using enzyme-centric metabolic framework (ECMF); 7- evaluation for community level dynamic models using the hierarchical resource allocation

framework (HRAF); 8-Evaluation of the spatial compartmentalized framework (SPCF) by comparing to TIM model.
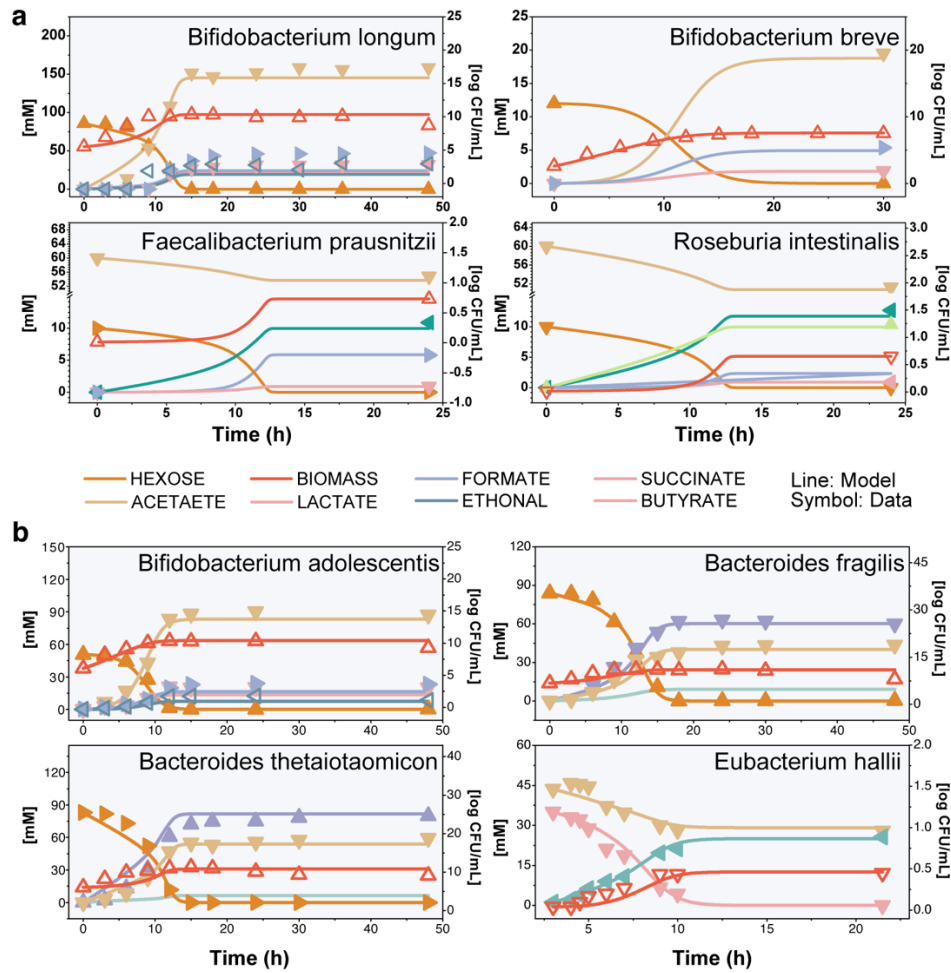
**Fig. S2 | Model validation results of representatives of bacteria-level dynamic models. (a)** Lines show model simulations of each metabolite and growth of microbiota representatives, namely, *Bidifobacterium longum*, *Bifidobacterium breve*, *Faecalibacterium prausnitzii*, and *Roseburia intestinalis.* (**b**) *Bifodobacterium adolescentis*, *Bacteroides fragilis*, *Bacteroides thetaiotaomicron*, and *Eubacterium hallii*. In each figure, the solid lines correspond to model simulation results and the symbols represent *in vitro* experimental data taken from the literature.
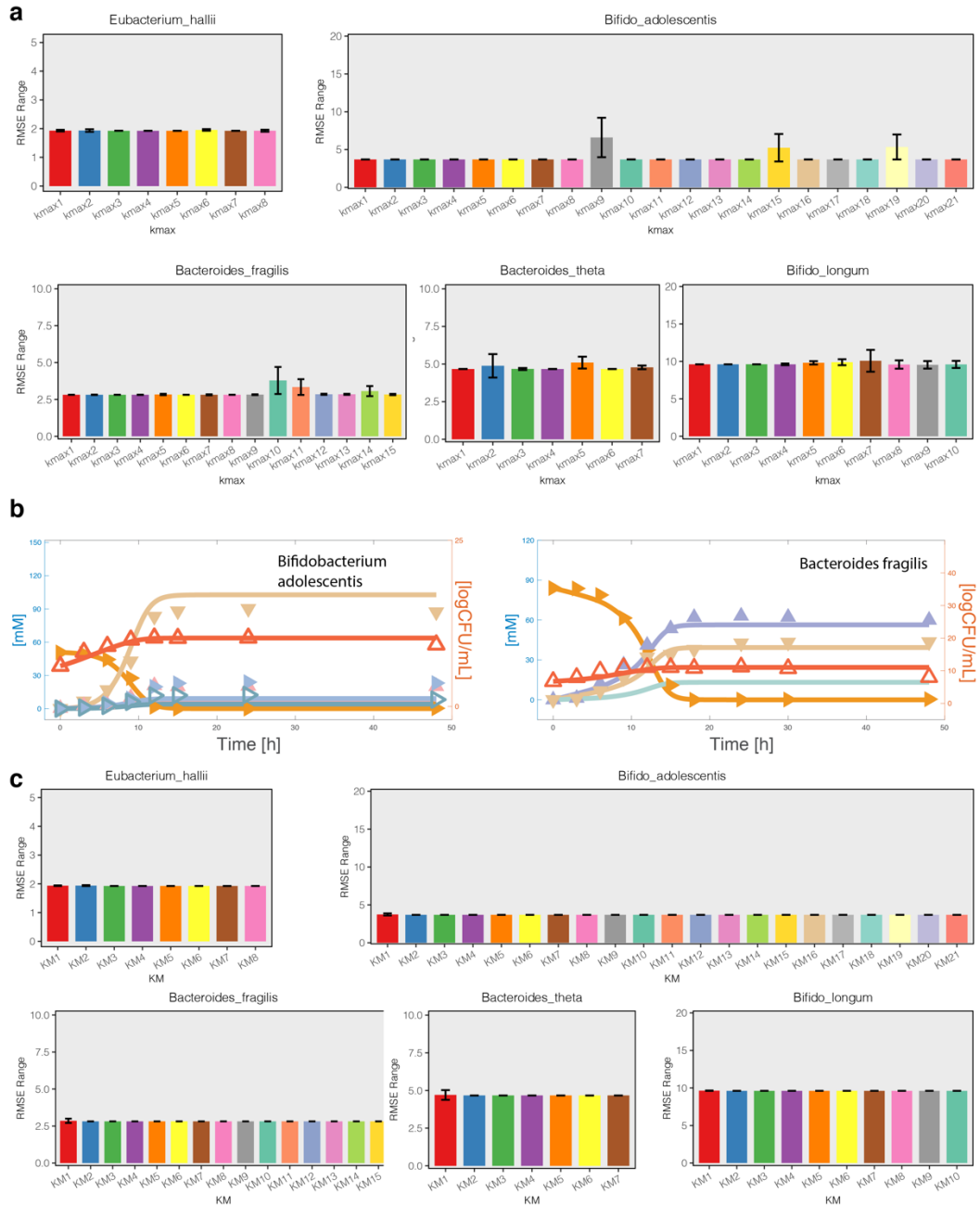
**Fig. S3 | Parameter sensitivity for ECMF model simulations of dynamic metabolism for monoculture bacterial growth.** (**a**) Bar plots show the variations of root mean square errors (RMSE) between model simulation results and *in vitro* experiment data as each kinetic parameter with respect to the $k_{max}$, was individually perturbed within the range from (-5%) to (+5%). **(b)** Sensitivity analysis for individual bacterium with the most significantly impacted RMSE condition for *Bifidobacterium adolescentis* ($k_{max,9}$), and *Bacteroides fragilis* ($k_{max,11}$). Other cases are not shown since the relative impact on RMSE was small. (**c**) Bar plots show the variations of root mean square errors (RMSE) between model simulation results and *in vitro* experiment data as each kinetic parameter (Michalis-Menten constants) was individually perturbed within the range from (-5%) to (+5%) by performing the sensitivity analysis for individual bacterium
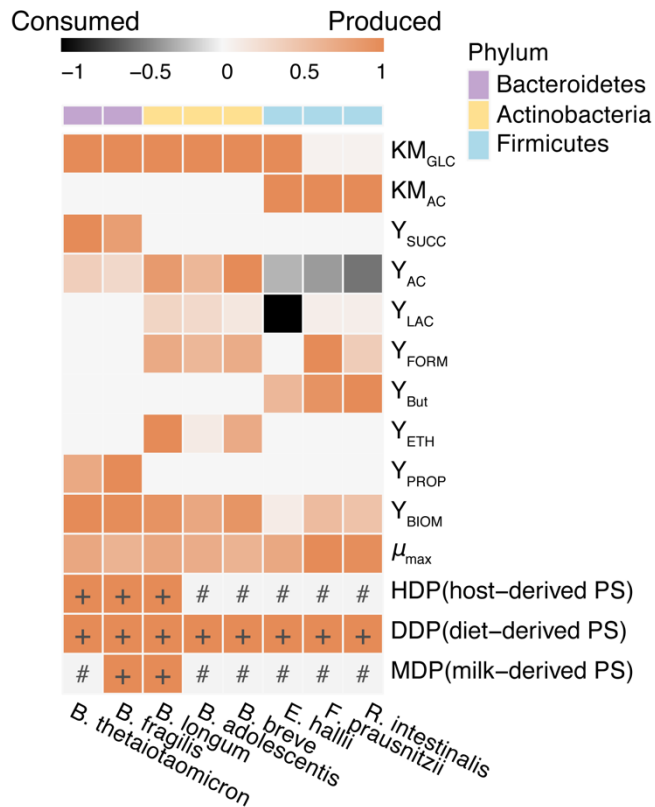
**Fig. S4 | Metabolic capacity comparison among representative microbiota members, represented by metabolite yield of key metabolites, maximal growth rate, and degradation of different complex carbohydrates.** Each column corresponds to one species. KM denotes the Michaelis-Menten constants, Y denote the yield of different metabolites, $\mu$ denotes the specific growth rate under monoculture condition. The subscripts and their corresponding meanings are: GLC-glucose, AC-acetate, SUCC-succinate, LAC-lactate, FORM-formate, But-butyrate, ETH-ethanol, PROP-propionate, BIOM-biomass. The symbol "+" denotes capable to utilize, the symbol "#" denotes cannot utilize. The representatives were annotated with different colors with respect to their phylum info. Positive value denotes production while negative value denotes consumption. Measures used to capture the metabolic capacity properties included yields of specific products (succinate, acetate, lactate, propionate, biomass, ethanol), and growth capacity, as well as their accessibility to three microbial-accessible-carbohydrates (MACs). The color bar was annotated based on the normalized values by dividing the maximal absolute values among all representative species. The metabolic capacity is shown for monoculture conditions.
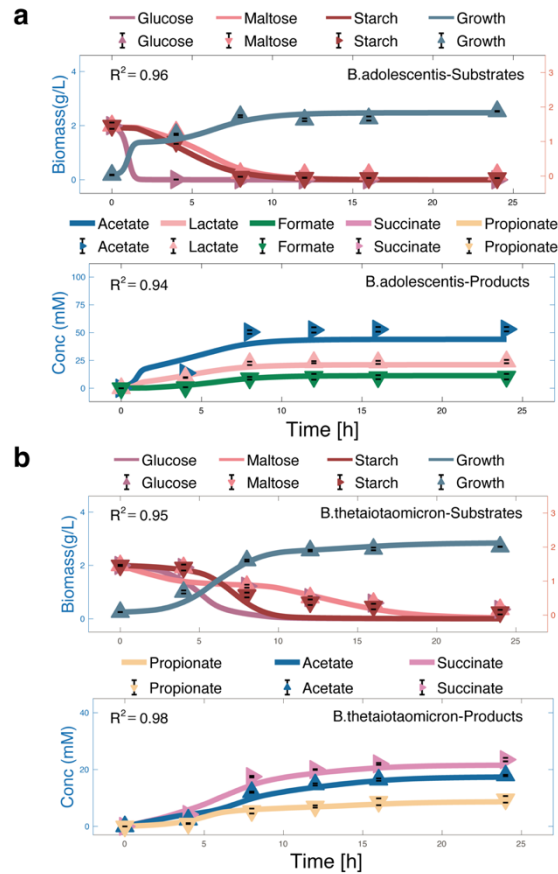
**Fig. S5 | Evaluation of coculture using *in vitro* experimental data of two consortia members.** Comparison of model simulation results on the dynamic metabolic profiles of species-specific utilization of multiple substrates, production of key metabolites, as well as individual bacterial growth to *in vitro* experimental data. (**a**) Model validation result of *Bifidobacterium adolescentis* under multiple substrate condition, namely, glucose, maltose and cellobiose. The coefficient of determination ($R^2$) ranges 0.94~0.96, for the model goodness evaluation. (**b**) Model validation result of *Bacteroides thetaoitaomicron* under the same multiple substrate condition. The coefficient of determination ($R^2$) ranges between 0.95~0.98, for the model goodness evaluation. Lines represent model simulation results; the symbols represent *in vitro* experimental data.
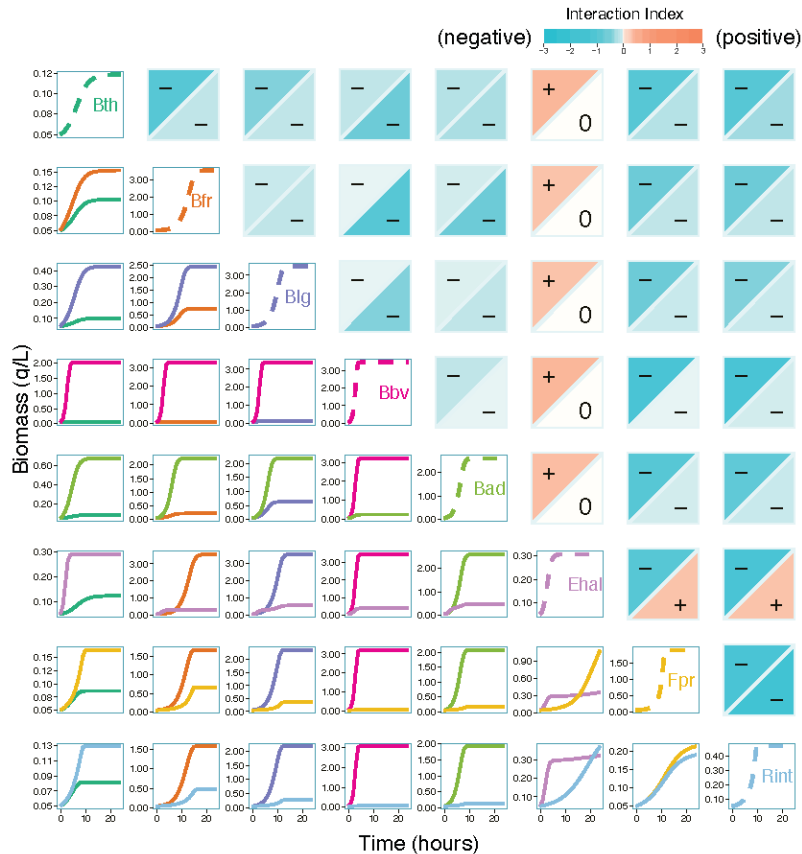
**Fig. S6 | HRAF prediction capacity of microbial-microbial interaction calculated based on dynamic growth profiles for all possible combinations of two-species synthetic coculture (Method 6(*SI Appendix*)).** Predictions on the dynamic growth profiles of monoculture for each microbial coculture member are shown in diagonal plots with colored dashed lines. HRAF predictions for coculture of all consortia of two-species combinations, are presented as the counters of the below triangle plot ruled with eight vertical and eight horizontal lines. The curves inside each counter plot denoted prediction results on the dynamic growth profiles of synthetic microbial ecosystem members within their consortia. Different species were indicated by corresponding line colors. The above triangle presented the calculation results of the microbial-microbial interaction index corresponding to each two-species coculture of the below triangle. Similarly, each counter of the above triangle plot ruled with the same eight vertical and horizontal lines denoted the interaction strength, with species lined in the row corresponding to the lower part of the interaction index and species lined in the column of each pair corresponding to the upper part of the interaction index. Three interaction types are shown based on model predictions of interaction indexes, with competition (-/-) as the dominant type (21 out of 28), others are nonneutral (five commensalism (+/0), two exploitation (+/-)). Interaction strength is shown as the interaction index color bar, with blue color indicating negative and orange color indicating positive. Same colors are shown for denoting the species in the diagonal and triangle below. Abbreviations of all species are listed as below. *Bth: Bacteroides thetaiotaomicron; Bfr: Bacteroides fragilis; Blg: Bifidobacterium longum; Bbv: Bifidobacterium breve; Bad: Bifidobacterium adolescentis; Ehal: Eubacterium hallii; Fpr: Faecalibacterium prausnitzii; and Rint: Roseburia intestinalis.*
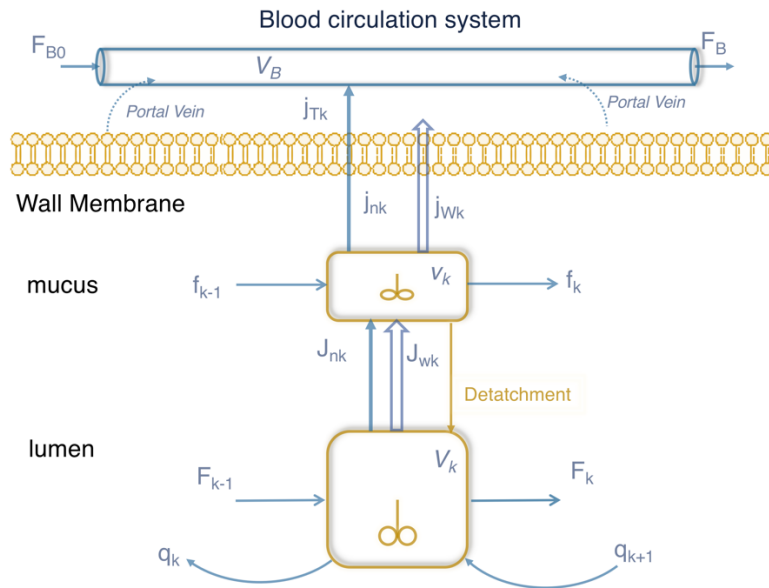
**Fig. S7 | Paradigm of spatial compartmentalized framework that describes colon physiology, shown as one region of each compartment.** Biomimetic colon model structure shown for one region, which is composed by three compartments, i.e., small tank of $v_k$ of mucus layer in the bottom, tank $V_k$ of lumen compartment in the top, and blood circulation system tank outside the membrane. The flow rate of lumen tanks is denoted by $F_k$, the flow rate of mucus tank is denoted by $f_k$, mass transport between lumen and mucus tank is denoted by $J_{nk}$ back-mixing in lumen compartment is denoted by $q_k$, mass transport from mucus to blood circulation system is denoted by $j_{nk}$. Flow rate in the blood circulation system is denoted by $F_B$.
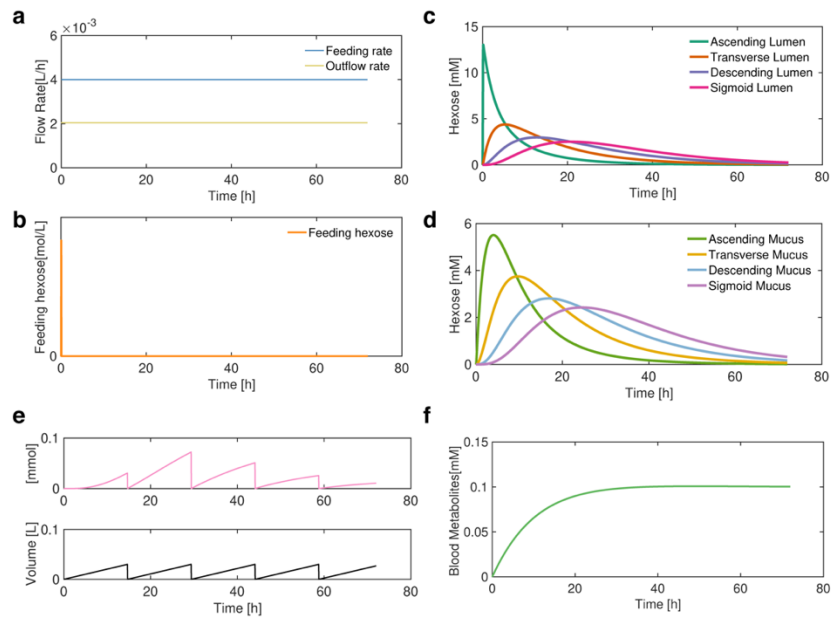
**Fig. S8 | Evaluation of SPCF by using one-time impulse signal of substrate as input.** (**a**) In- and Out- flow rate of the system. Blue line denotes the inflow rate of the model, yellow line denotes the outflow rate of the model. (**b**) Impulse one-time input signal to trace the system. (**c**) The dynamic profiles of how the input signal developed in each lumen- associated compartment. (**d**) The dynamic profiles of how the input signal changes in each mucus- associated compartment. (**e**) The output of the tracer in the feces compartment of fecal metabolites (above), and feces volume (below). (**f**) The dynamic profile of the nutrient signal in SPCF's blood compartment.
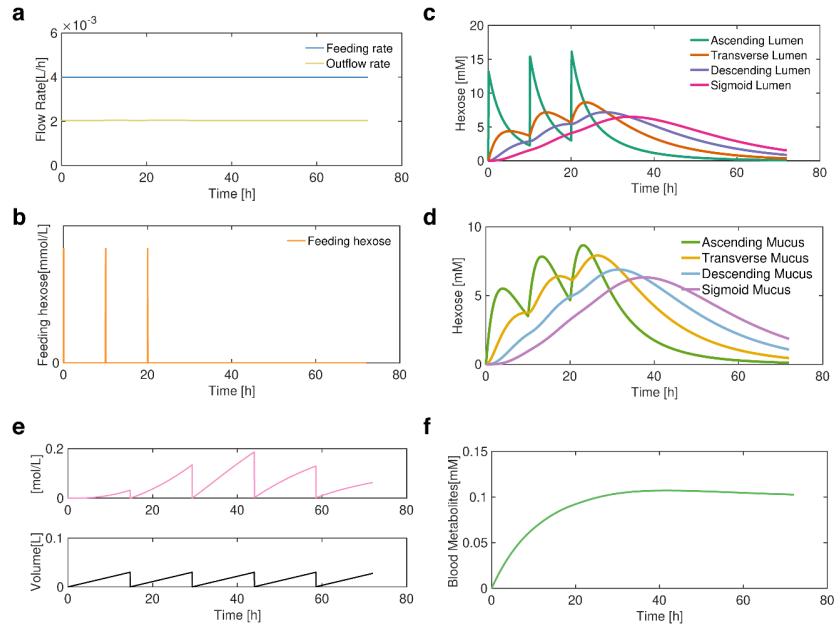
34

**Fig. S9 | Evaluation of SPCF by using three-times impulse signal of substrate as input.** (**a**) In- and Out- flow rate of the system. Blue line denotes the inflow rate of the model, yellow line denotes the outflow rate of the model**.** (**b**) Impulse three-times input signal to trace the system. (**c**) The dynamic profiles of how the input signal developed in each lumen- associated compartment. (**d**) The dynamic profiles of how the input signal developed in each mucus- associated compartment. (**e**) The output of the tracer in the feces compartment of fecal metabolites (above), and feces volume (below). (**f**) The dynamic profile of the signal nutrient in SPCF's blood compartment.
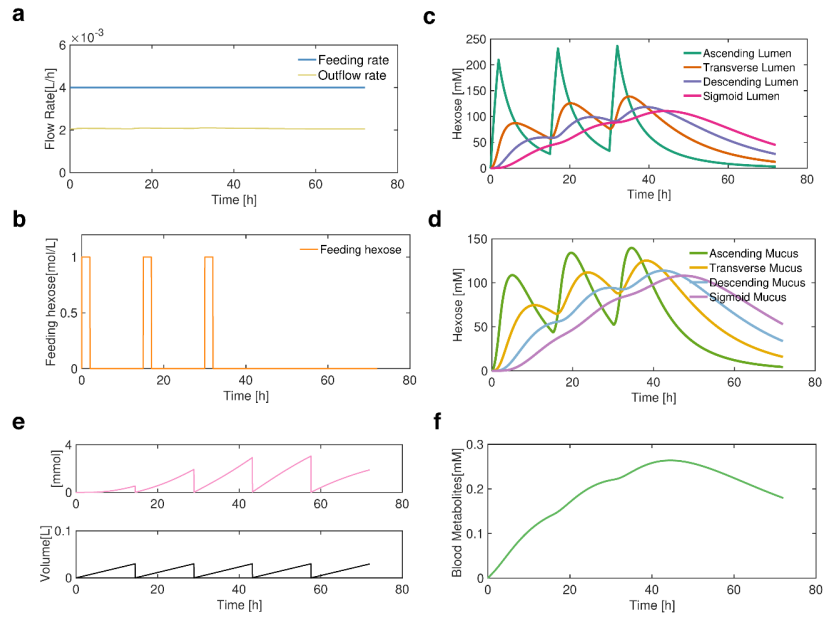
**Fig. S10 | Evaluation of SPCF by using three-step signal of substrate as input.** (**a**) In- and Out- flow rate of the system. Blue line denotes the inflow rate of the model, yellow line denotes the outflow rate of the model. (**b**) Impulse three-step input signal to trace the system. (**c**) The dynamic profiles of how the input signal developed in each lumen- associated compartment. (**d**) The dynamic profiles of how the input signal developed in each mucus- associated compartment. (**e**) The output of the tracer in the feces compartment of fecal metabolites (above), and feces volume (below). (**f**) The dynamic profile of the signal nutrient in SPCF's blood compartment.
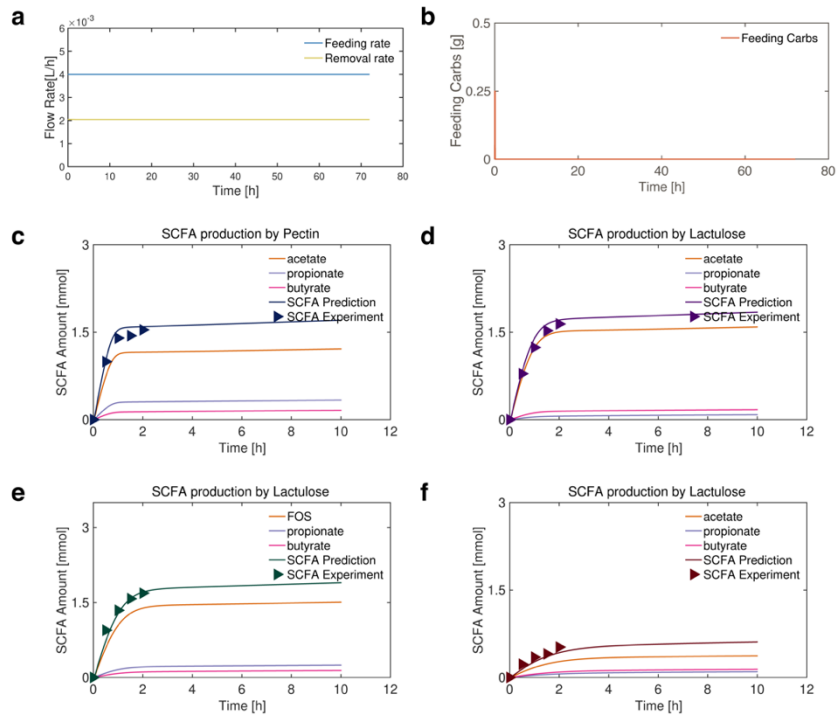
**Fig. S11 | Evaluation of SPCF predictions of SCFA production by using four complex carbohydrate at 0.25g for each.** (**a**) In- and Out- flow rate of the system. Blue line denotes the inflow rate of the model, yellow line denotes the outflow rate of the model. (**b**) Total carbohydrate input to the system, with 0.25g at the t=0 time-point. (**c**) Comparison of model predictions on SCFA total production and composition under Pectin as carbohydrate. Lines denote model predictions for acetate, butyrate and propionate, and their sum (purple lines). Purple dots denote total SCFA production obtained in TIM experimental data under 0.25g pectin. (**d**) Comparison of model predictions on total SCFA production and composition under Lactulose as carbohydrate. Lines denote model predictions for acetate, butyrate and propionate, and their sum (purple red lines). Purple red dots denote total SCFA production observed in TIM experimental data under 0.25g lactulose. (**e**) Comparison of model predictions on SCFA total production and composition under FOS as carbohydrate. Lines denote model predictions for acetate, butyrate and propionate, and their sum (green lines). Green dots denote total SCFA production observed in TIM experimental data under 0.25g FOS. (**f**) Comparison of model predictions on total SCFA production and composition under Lactitol as carbohydrate. Lines denote model predictions for acetate, butyrate and propionate, and their sum (red lines). Red dots denote total SCFA production observed in TIM experimental data under 0.25g lactitol.
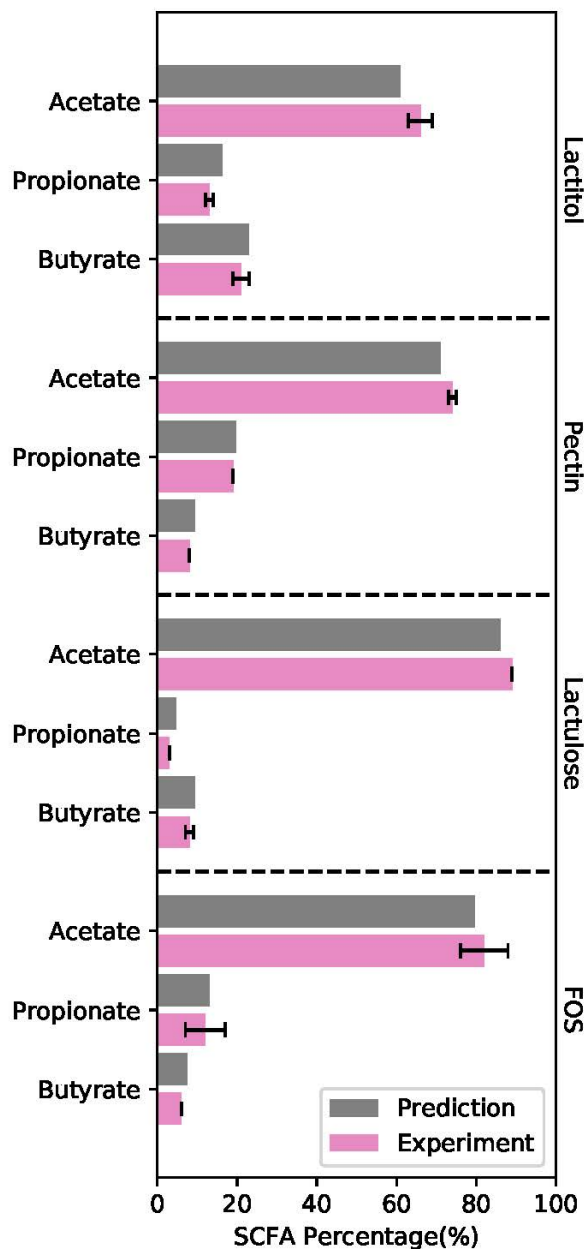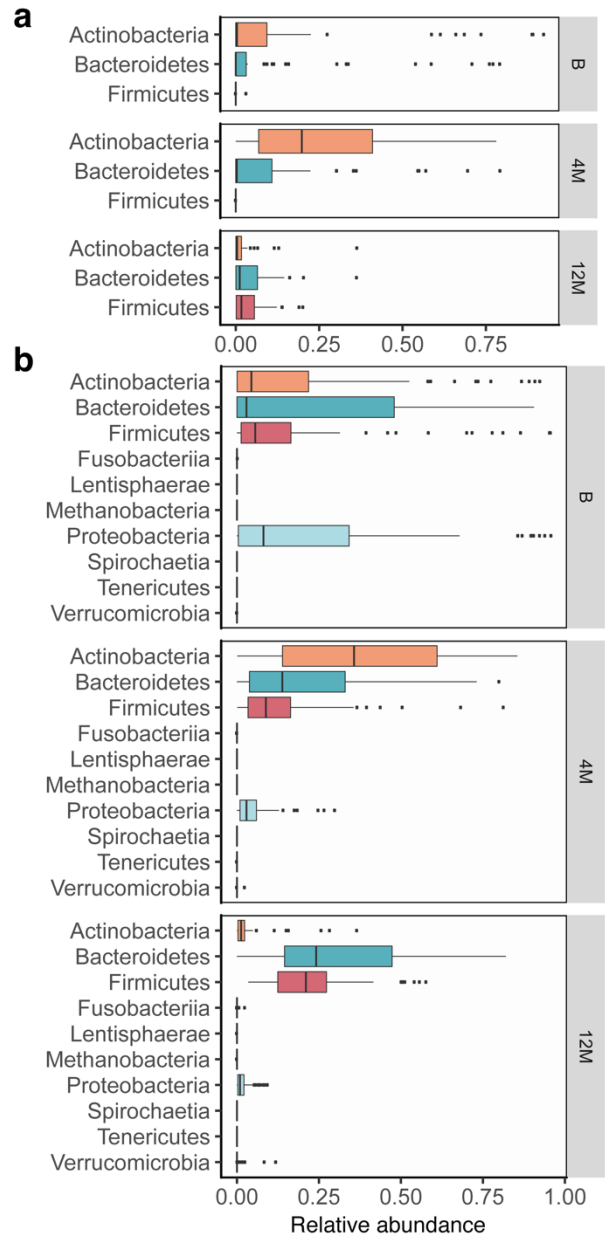
**Fig. S12 | Evaluation of SPCF predictions on SCFA composition profiles (denoted as percentage) produced by four complex carbohydrate (at 0.25g for each) through comparing SPCF predictions to TIM *in vitro* experimental data.** Grey bars denote SPCF predictions, pink bars with error bars denote TIM *in vitro* experimental data.
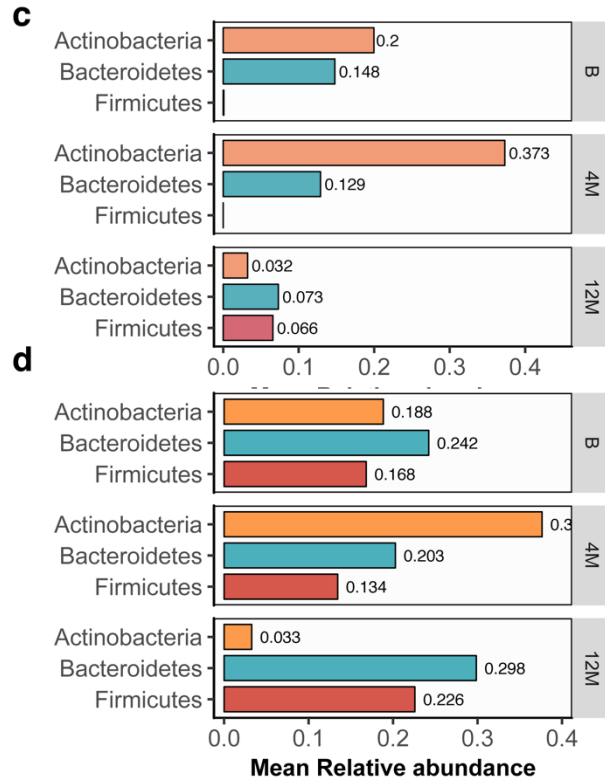
**Fig. S13 | Comparison between microbial ecosystem model encompassing representatives and original community of the infant gut microbiota.** (**a-b**) Boxplots show how microbial ecosystem model resemble the three most prominent phyla in the infant gut microbiota. (**c-d**) Mean relative abundance plots show how microbial ecosystem resemble the three most prominent phyla during each time period.
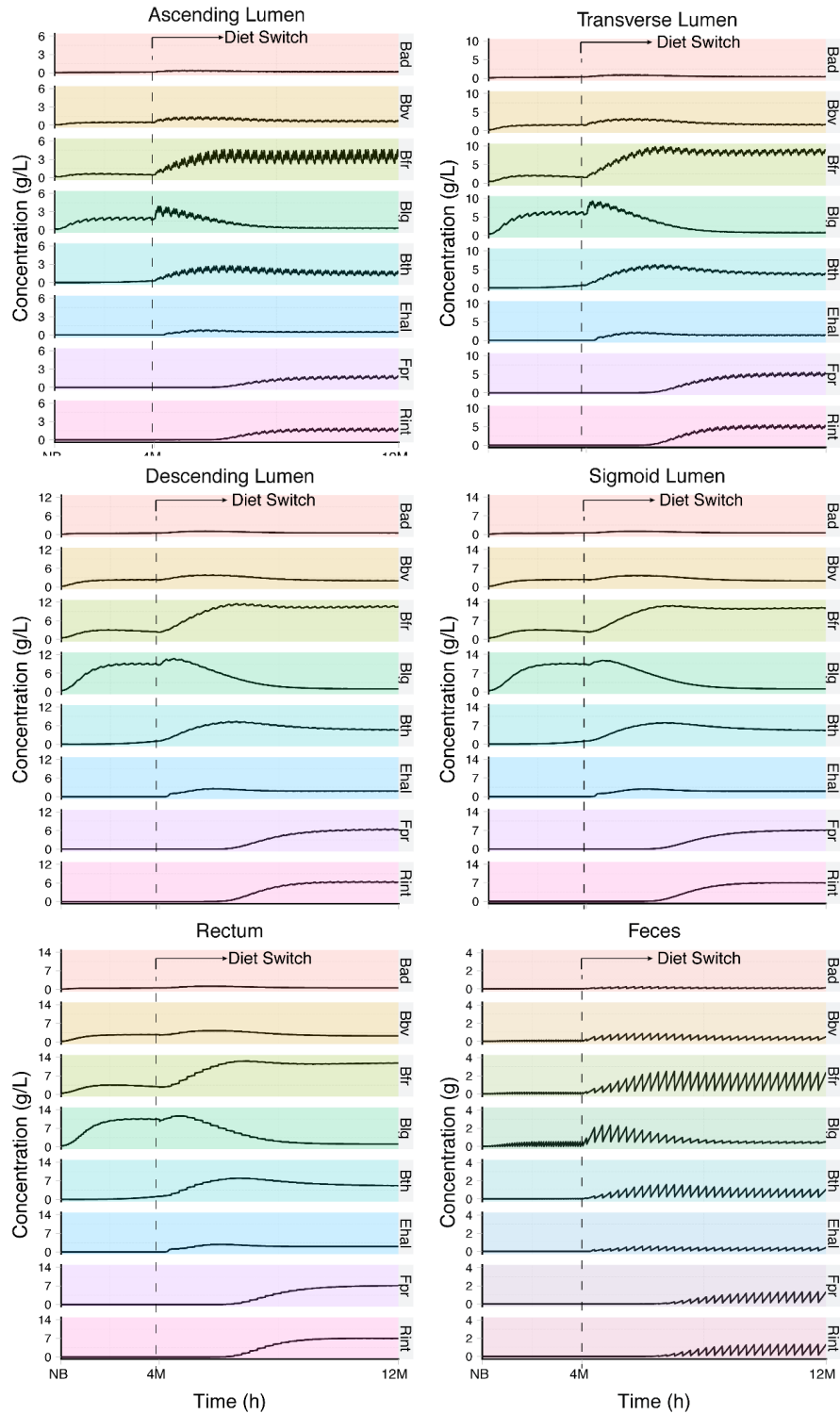
**Fig. S14 | Modeled Dynamic evolution profiles of microbiota downstream the luminal colon phases, rectum and feces storage.** Data is shown to present the modeled dynamic transient microbiota biology from newborn to stable colonization under breastmilk-feeding, and then switching to solid-food feeding. The diet-switching point is annotated by the arrows in each panel.
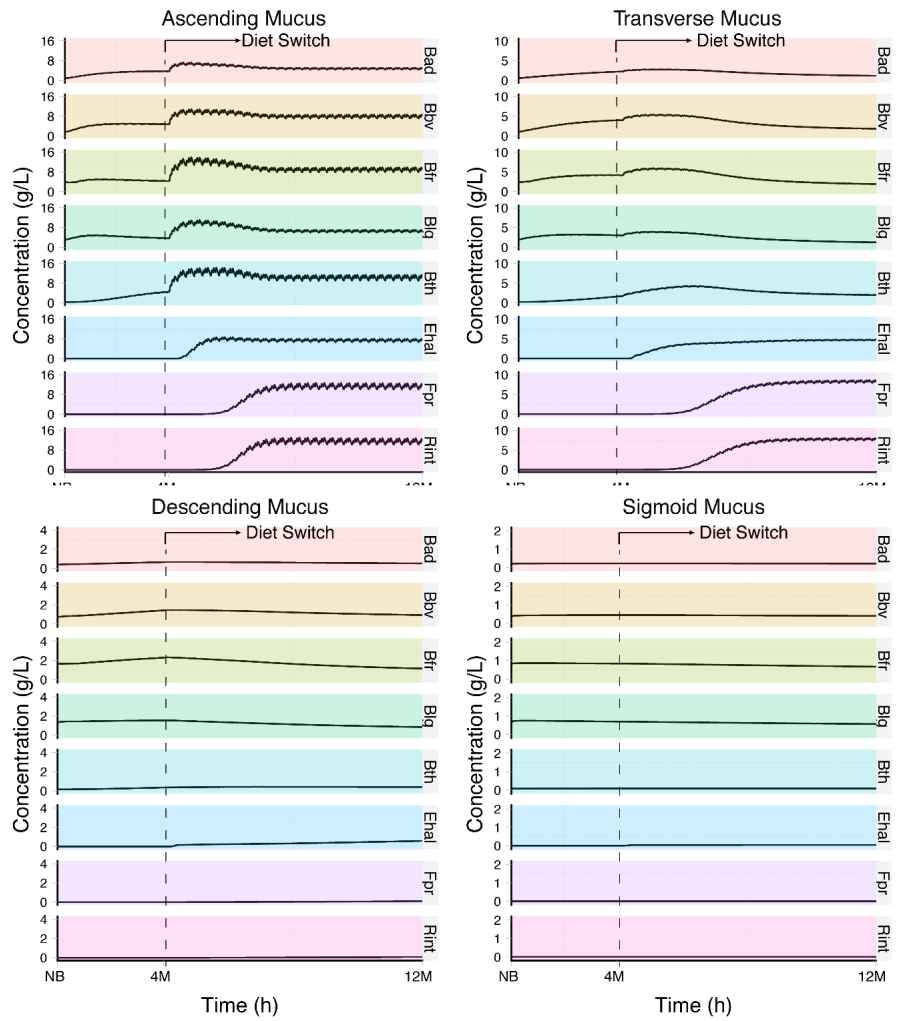
**Fig. S15 | Modeled dynamic evolution profiles of microbiota in the mucus compartments, longitudinally from cecum to sigmoid regions.** Data is shown from newborn to stable colonization under breastmilk, and then switching to solid food. The diet-switching point is annotated by the arrows in each panel.
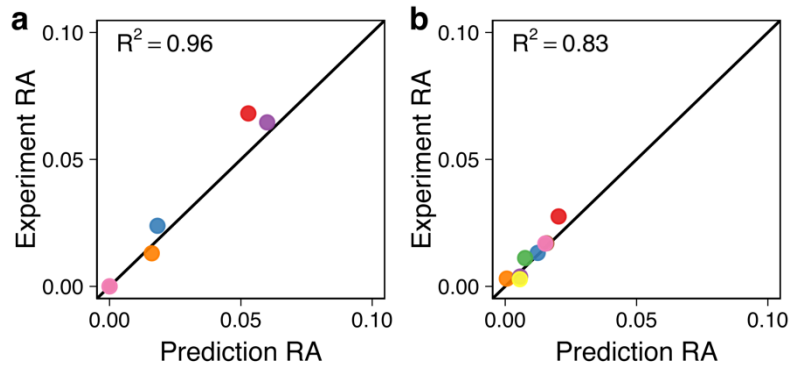
**Fig. S16 | Evaluation of CODY predictions on feces microbiota configurations under breastmilk feeding (4th month) and solid-food feeding (12th month).** (**a**) $R^2$ between microbiota relative abundance (RA) predicted by CODY in the feces compartment as compared to *in vitro* metagenomics measurement of fecal samples, at 4th month. (**b**) $R^2$ between microbiota relative abundance (RA) predicted by CODY in the feces compartment as compared to the reference data taken from *in vitro* metagenomics measurement of fecal samples, at 12th month.
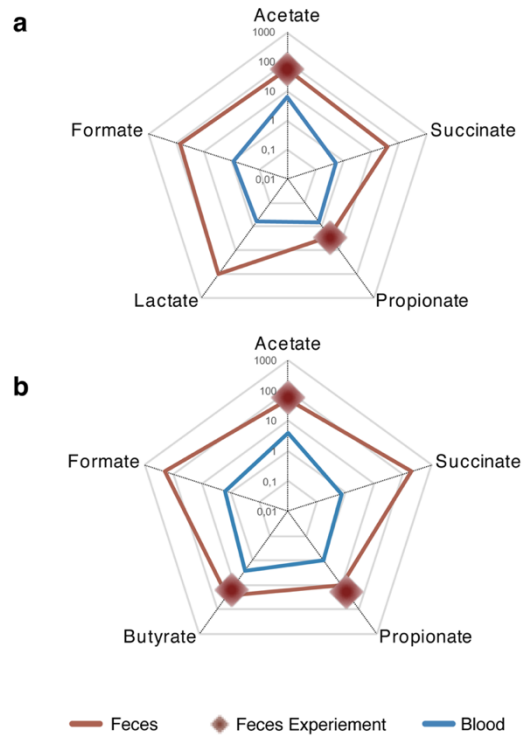
**Fig. S17 | Evaluation of CODY predictions on the SCFA excretion in the feces and plasma.**
For feces compartment, the unit is mmol/kg feces. For the blood compartment, the unit is mmol/L.
(**a**) Evaluation of CODY predictions of the excretion of acetate, propionate, succinate, and formate in the plasma (blue lines) and feces (dark red lines) of 4th month infant. Lines denote model predictions, filled diamond symbols denote reference data. (**b**) Evaluation of CODY predictions of the excretion of acetate, succinate, propionate, butyrate and formate, in the plasma (blue lines) and feces (dark red lines) of the 12th month infant. Lines denote model predictions, filled diamond symbols denote reference data.

**Fig. S18 | Modeled dynamic evolution profiles of microbiota-associated metabolites in the ascending and transverse regions of lumen compartments.** Data is shown from newborn to stable colonization under breastmilk, and then switching to solid food. The diet-switching point is annotated by the arrows in each panel.
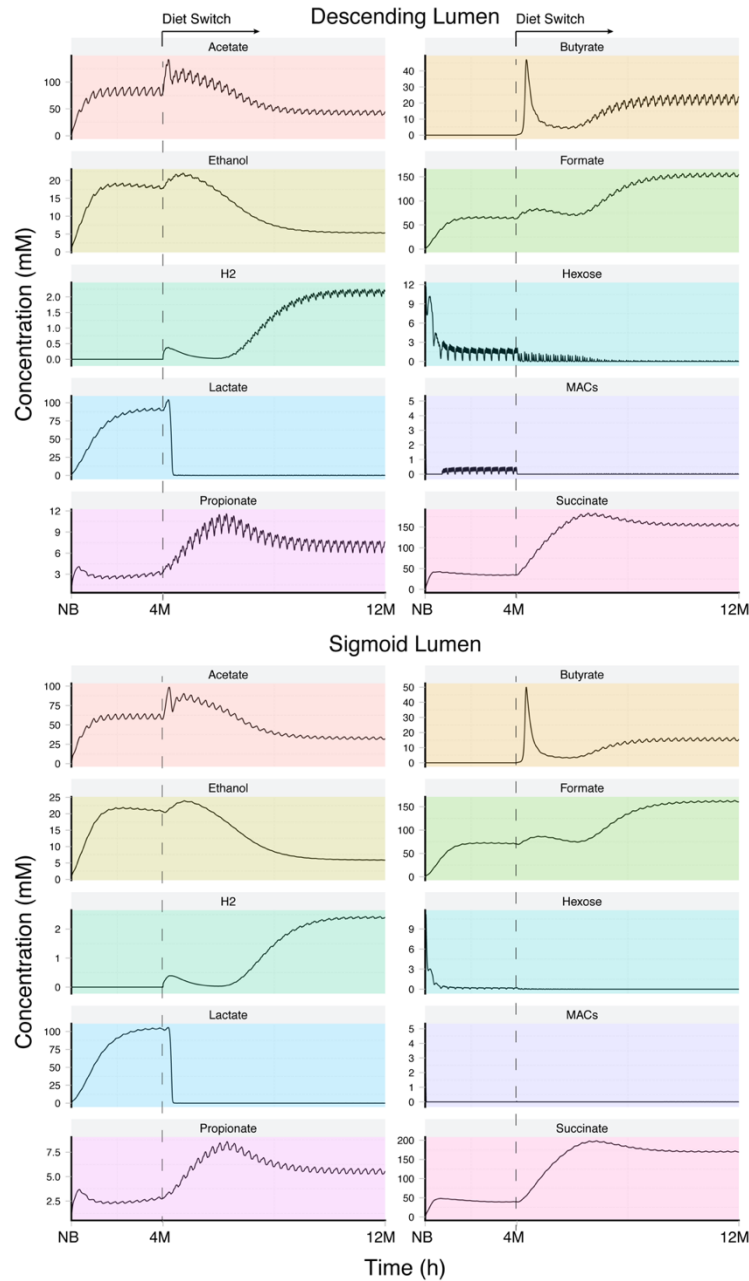
**Fig. S19 | Modeled dynamic evolution profiles of microbiota-associated metabolites in the descending and sigmoid regions of lumen compartments.** Data is shown from newborn to stable colonization under breastmilk, and then switching to solid food. The diet-switching point is annotated by the arrows in each panel.
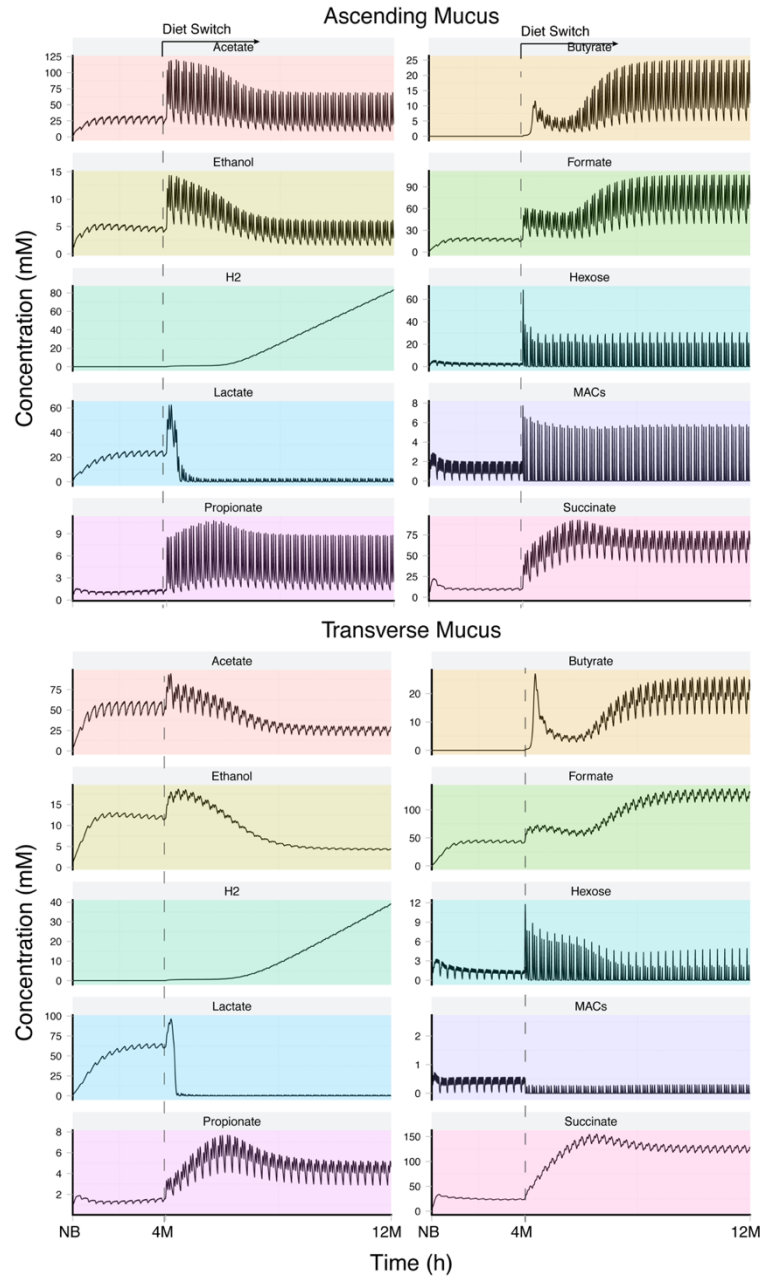
**Fig. S20 | Modeled dynamic evolution profiles of microbiota-associated metabolites in the ascending and transverse regions mucus compartments.** Data is shown from newborn to stable colonization under breastmilk, and then switching to solid food. The diet-switching point is annotated by the arrows in each panel.
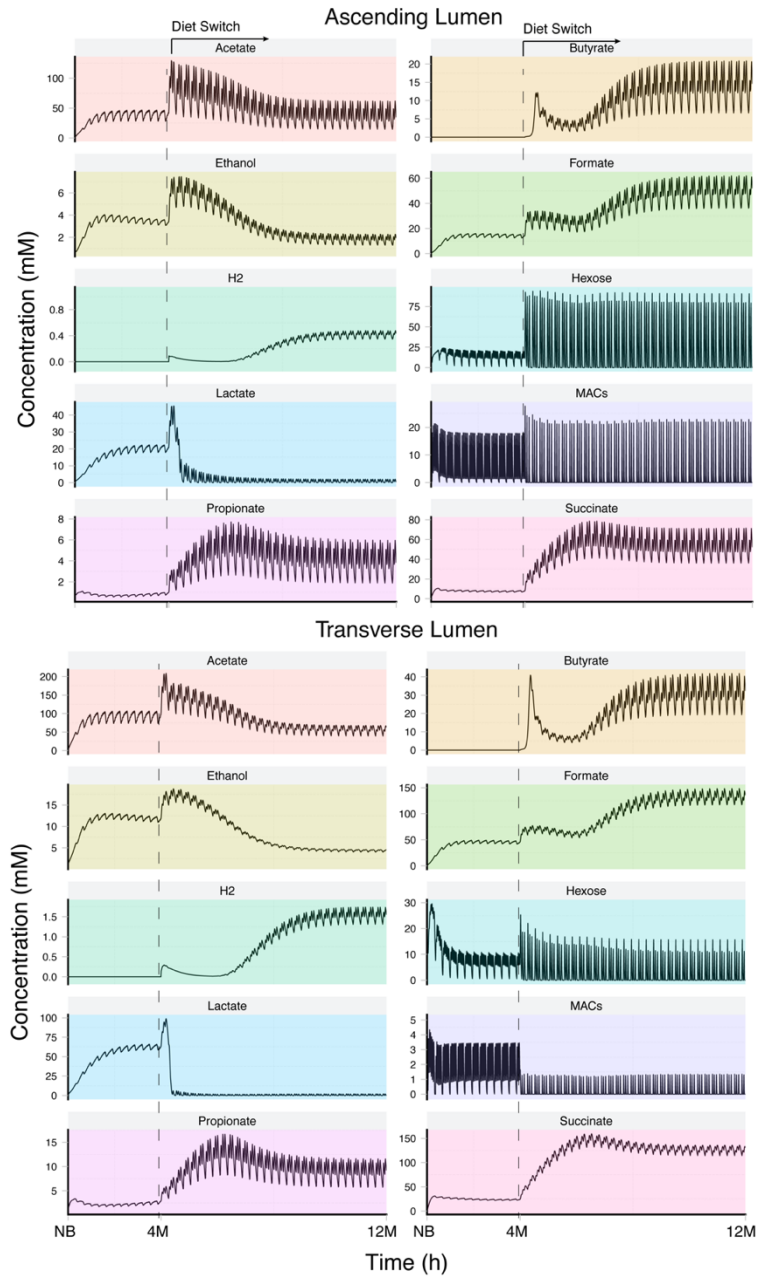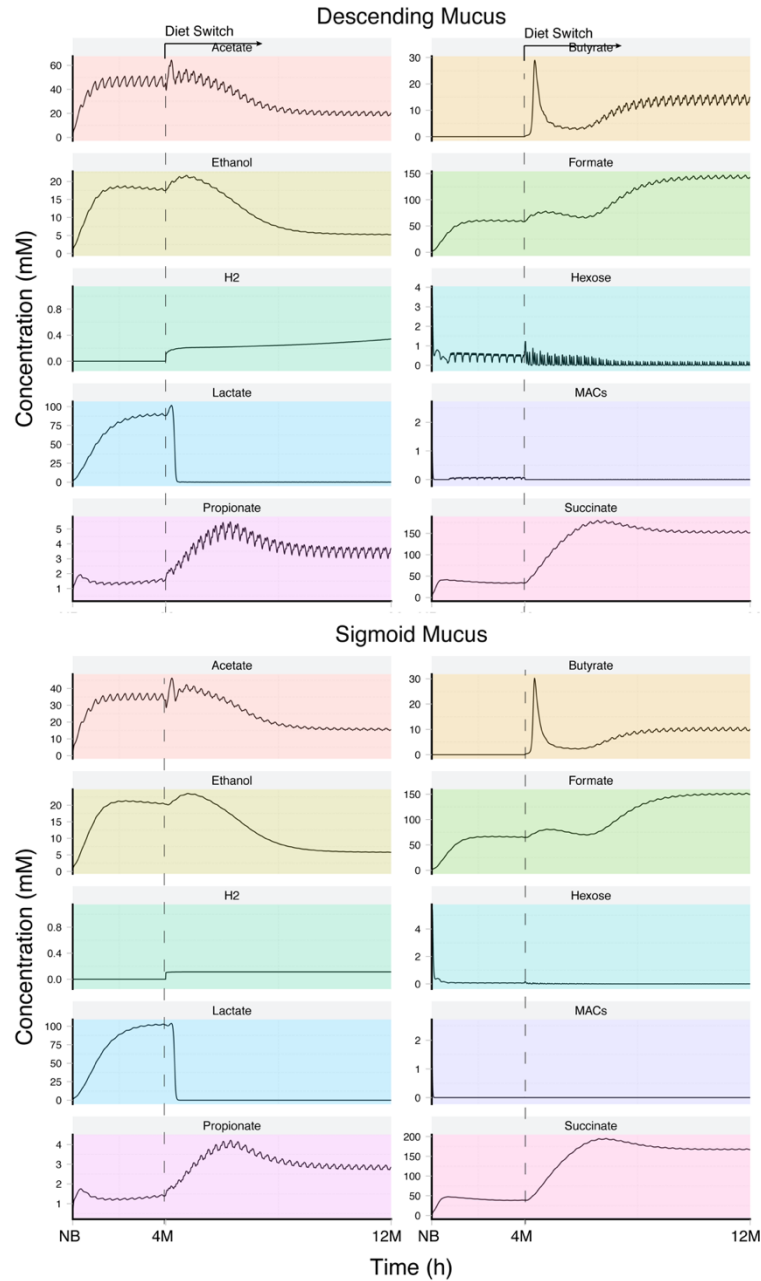
**Fig. S21 | Modeled dynamic evolution profiles of microbiota-associated metabolites in the ascending and transverse regions mucus compartments.** Data is shown from newborn to stable colonization under breastmilk, and then switching to solid food. The diet-switching point is annotated by the arrows in each panel.
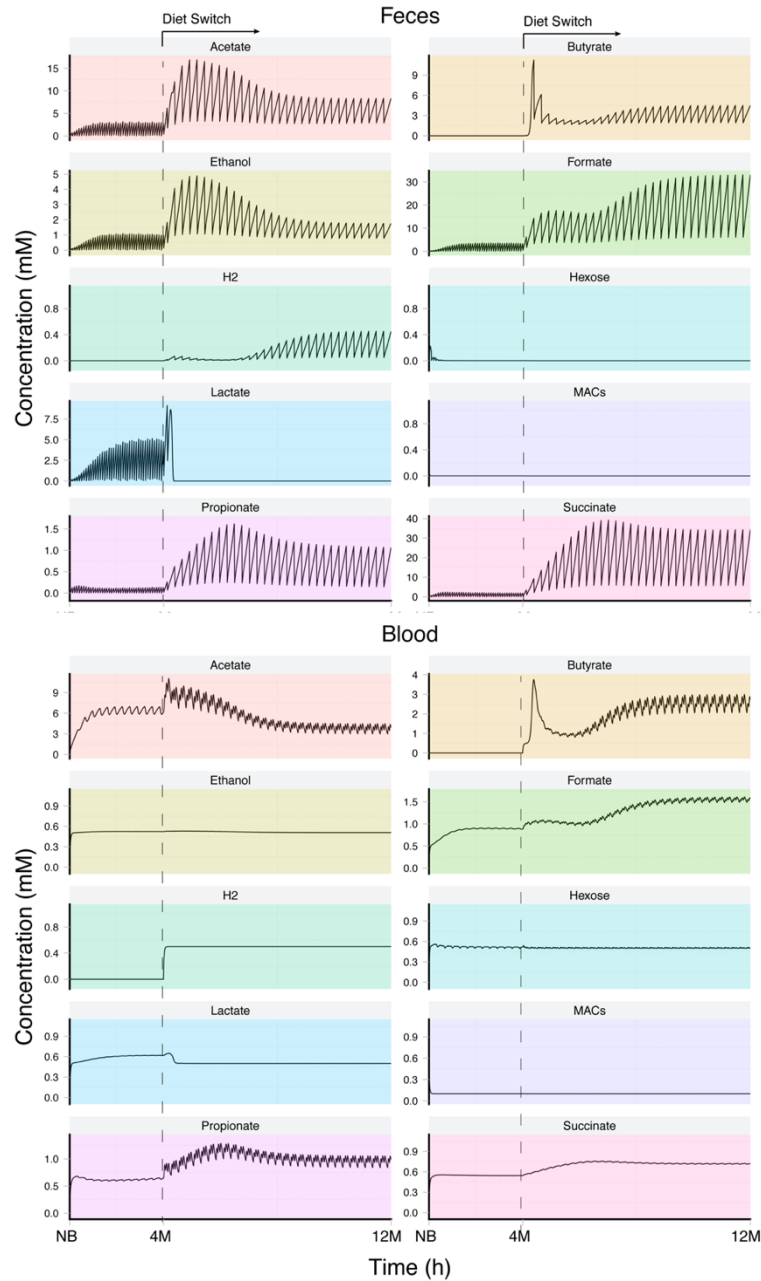
**Fig. S22 | Modeled dynamic evolution profiles of microbiota-associated metabolites in the feces and BCS phases.** Data is shown from newborn to stable colonization under breastmilk, and then switching to solid food. The diet-switching point is annotated by the arrows in each panel.
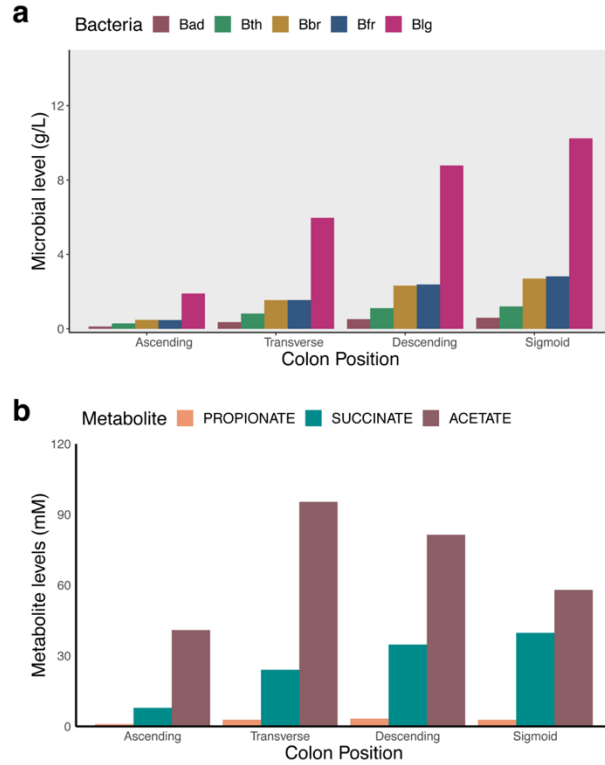
**Fig. S23 | Comparison of the predicted post-stabilized levels of bacteria and SCFA in each region of the lumen compartments during breastmilk feeding period. (a**) The post-stabilized levels of bacteria in each region of lumen colon increase mainly in the proximal colon and maintain relatively stable when moving towards the distal colon. (**b**) SCFA level increase in the proximal colon and decrease when reaching the distal colon due to absorption.
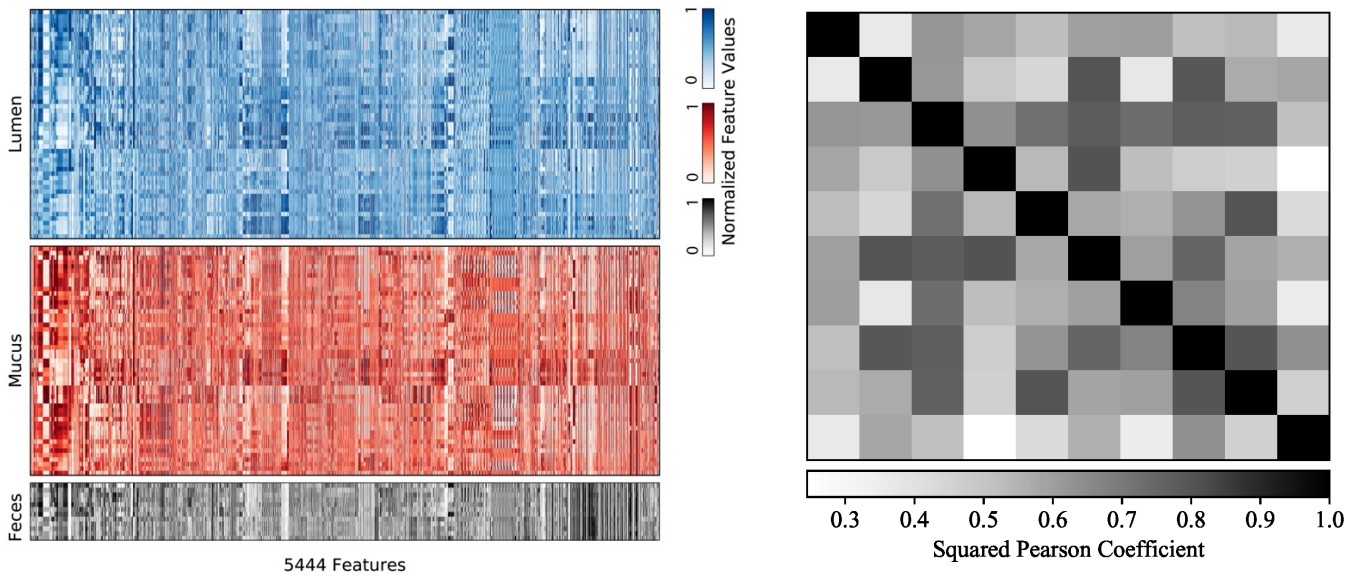
**Fig. S24 | Feature extraction and the pairwise association among the top 10 features, performed by using a computational framework for the time-series phenotyping using massive feature extraction (hctsa) toolbox to distinguish the dynamic development profiles of microbiota with respect to lumen, mucus and feces. Left:** Using hctsa to quantify and interpret phenotypic difference, the package facilitates massive feature extraction to compare over 5000 features of each time series, derived from an interdisciplinary time-series analysis literature. The feature matrix contains the result of this feature extraction, where each row represents a time series and each column represents a feature that encapsulates some property of that time series (e.g., measures of its autocorrelation structure, entropy, etc.). Colored (blue and red) labels denote the three types of data: dynamic evolution process of bacteria development profiles in lumen, mucus and feces; dark/light labels high/low values of each feature, revealing rich structure in the dynamical properties of the dataset. **Right:** Pairwise correlation map among the top 10 features identified by hctsa for distinguishing the two phenotypes of time-series data. The top 10 features span a wide range of time-series analysis techniques, and form sets of high diversity of correlated groups.

**Fig. S25 | The statistical values corresponding to the top 10 features of three phenotypes of time-series data, each bar shows the property value across each phenotype of time-series data (lumen, mucus or feces), error bar means standard deviation across each phenotype of time-series data.** The property names and their potential meanings corresponding to top ten features are (*29, 30*):

(**a**) MF_AR_arcov_5_a3: Using time-series dataset to fit a fifth-order Autoregressive model applying Matlab Signal processing toolbox, returning third-order coefficient of an autoregressive AR(5) model fitted to the time-series dataset;
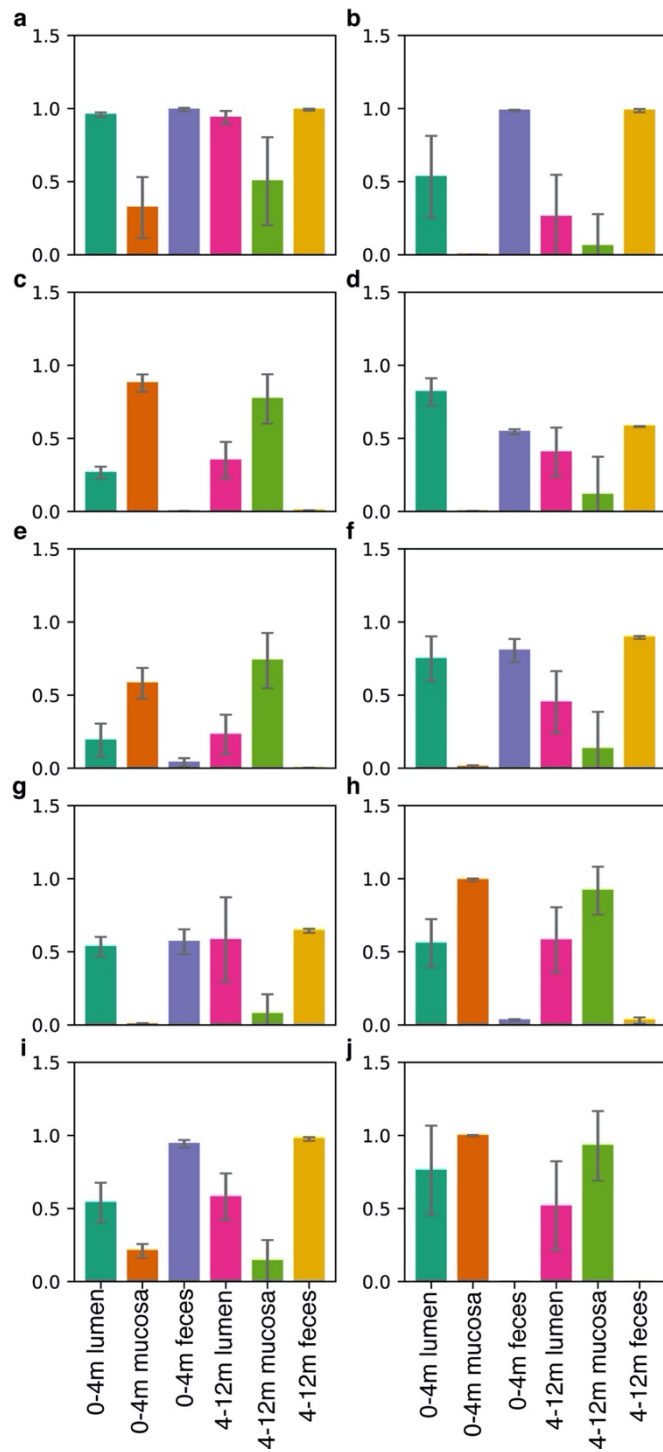
(**b**) MF_AR_arcov_5_a4: Using time-series dataset to fit a fifth-order Autoregressive model applying Matlab Signal processing toolbox, return the fourth-order coefficient of an autoregressive AR(5) model fitted to the time-series dataset

(**c**) MF_AR_arcov_5_a5: Using time-series dataset to fit a fifth-order Autoregressive model applying Matlab Signal processing toolbox, return the fifth-order coefficient of an autoregressive AR(5) model fitted to the time-series dataset

(**d**) MF_CompareTestSets_y_ar_4_rand_25_01_1_ac1s_mean: using the time-series dataset to fit a fourth-order AR model which is then used to repeatedly predict 25 times randomly chosen subsegments with 10% length and one-step ahead in the test sets, and then return the mean autocorrelation coefficient of prediction residuals at lag one phase computed by 'Fourier' method.

(**e**) MF_CompareTestSets_y_ar_4_rand_25_01_1_ac1s_median: using the time-series dataset to fit a fourth-order AR model which is then used to repeatedly predict 25 times  randomly chosen subsegments with 10% length and one-step ahead in the test sets, and then return the median autocorrelation coefficient of prediction residuals at lag one phase computed by 'Fourier' method.

(**f**) MF_StateSpace_n4sid_2_05_1_minfpe: Using time-series dataset to fit a second-order state-space model and then train the first 50% of dataset to predict the rest data with 1 step ahead. MF_ResidualAnaysis would return the minfpe value which is the minimal fpe (final prediction error) derived by fitting a zero-mean AR model to residuals between modeling fitting and rest dataset of 1 step ahead.

(**g**) MF_armax_3_1_05_1_minfpe: using time-series dataset to fit an 3-1 order of Ar-max model, train the model with 50% portion data and testing the rest portion with 1 step ahead. Returning the goodness of model fit in the training data

(**h**) MF_armax_2_2_05_1_p3_5: statistical tests on the residuals from using the fitted model to predict the testing data. P3_5 means identify any low-frequency trends in residuals from proportion of power spectrum in third and fourth

(**i**) MF_armax_2_2_05_1_p4_5: statistical tests on the residuals from using the fitted model to predict the testing data. p4_5 means identify any low-frequency trends in residuals from proportion of power spectrum in fourth and fifth

(**j**) MF_armax_2_2_05_1_minfpe: Using time-series dataset to fit a 2-2 order of AR Max model, train the mode with 50% portion data and testing the rest portion with 1 step ahead. Returning the goodness of model fit in the training data by the minPFE, by fitting a linear model and see if it picks up any structure.
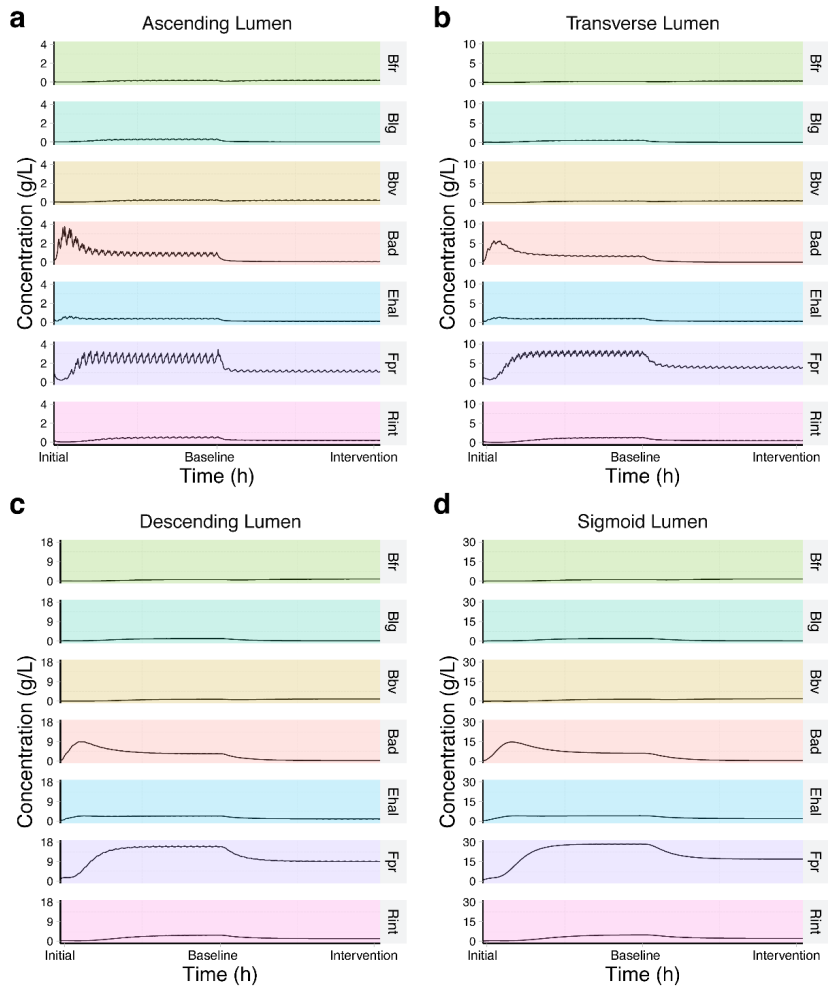
**Fig. S26 | Modeled dynamic evolution profiles of microbiota downstream the luminal colon phases, for the adult cohort experiencing diet-intervention.** Data is shown to present the dynamic transient microbiota development under normal diet and achieve stability, further experiencing the diet-intervention treatment, specifically, in the (**a**) Ascending lumen region; (**b**) Transverse lumen region; (**c**) Descending lumen region and (**d**) Sigmoid lumen region.
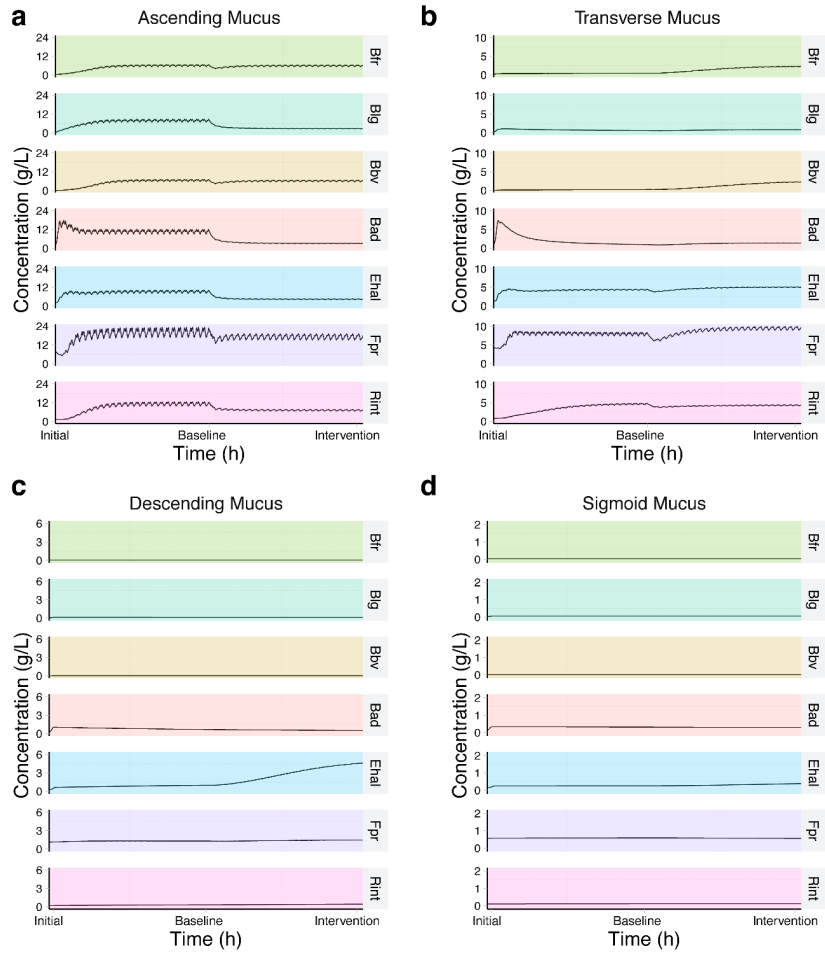
**Fig. S27 | Modeled dynamic evolution profiles of microbiota in the mucus compartments, for the adult cohort experiencing diet-intervention.** Data is shown to present the dynamic transient microbiota development under normal diet and achieve stability, further experiencing the diet-intervention treatment, specifically, in the (**a**) Ascending mucus region; (**b**) Transverse mucus region; (**c**) Descending mucus region and (**d**) Sigmoid mucus region.
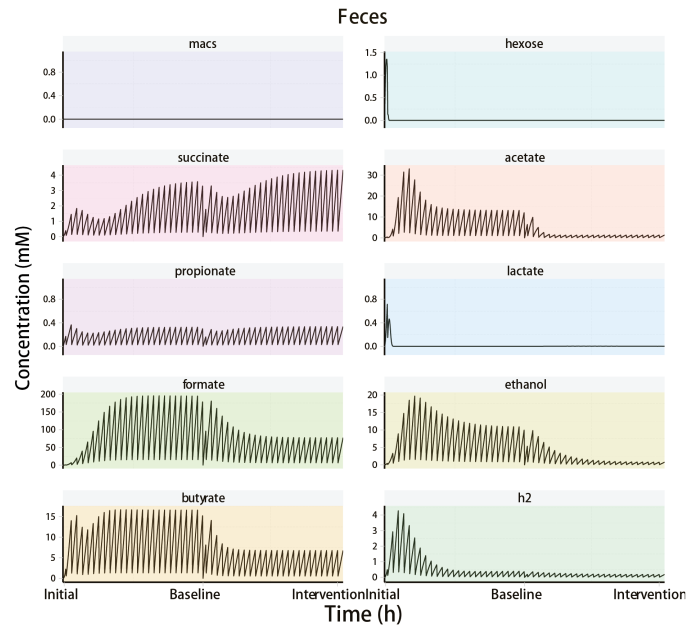
**Fig. S28 | Modeled dynamic evolution profiles of microbiota-associated metabolites in the feces region for the adult cohort.** Data is shown development under normal diet and achieve stability, further experiencing the diet-intervention treatment, in the feces region.
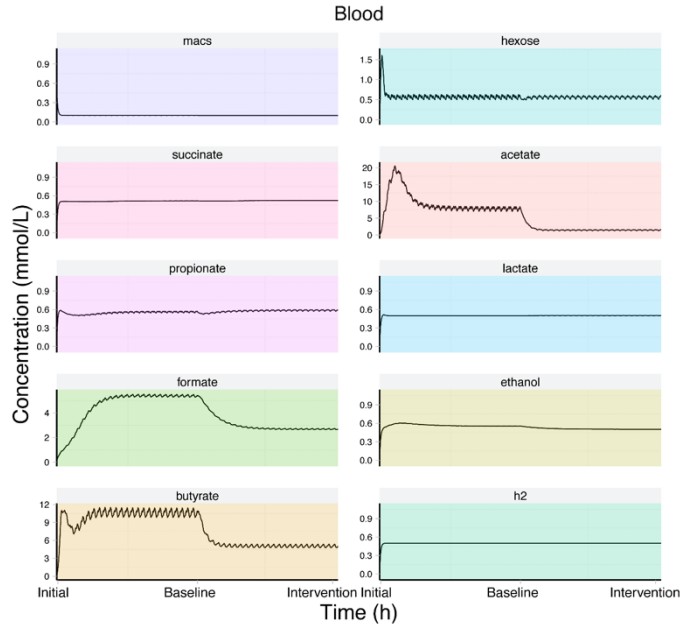
**Fig. S29 | Modeled dynamic evolution profiles of microbiota-associated metabolites in the blood compartment for the adult cohort.** Data is shown the microbiota-associated metabolite profiles under normal diet and achieve stability, further experiencing the diet-intervention treatment, in the blood compartment.

# Supplementary Tables

**Table S1 |** Summary of abbreviations used in the main text and the corresponding full names, as well as the associated representations.

| Abbreviation | Full name | Representation |
|---|---|---|
| ECMF | Enzyme-centric metabolic framework | Dynamic framework that describe the bacterial biology for monoculture |
| HRAF | Hierarchical resource allocation framework | Dynamic framework that describe the bacterial biology for consortia |
| SPCF | Spatiotemporal compartmentalized framework | Dynamic framework that capture the colon structure and physical forces |
| CODY | Multiscale framework for computing the dynamics of microbiota | Comprehensive dynamic model that predict the dynamic gut microbiota reprogramming |
| SCFA | Short chain fatty acid | The major products of the microbiota |
| GEMs | Genome-scale metabolic models | The representation of microbial metabolism |
| MPMs | Metabolic pathway modules | The active set of metabolic pathways of microbe that represent the metabolic capacity |
| FBA | Flux balance analysis | A common approach in system biology that identify one metabolic pathway with maximal yield, i.e., biomass |
| BCS | Blood circulation system | The blood compartment considered in SPCF |
| $R^2$ | Determination of coefficient | Evaluation of the accuracy of model predictions |
| MACs | Microbial accessible carbohydrates | Microbial accessible carbohydrates |
| TIM | TNO intestinal model | A classical in vitro experimental setup used to simulate human colon environment |
| Bth | *Bacteroides thetaiotaomicron* | The microbial ecosystem member |
| Bfr | *Bacteroides fragilis* | The microbial ecosystem member |
| Blg | *Bifidobacterium longum* | The microbial ecosystem member |
| Bbv | *Bifidobacterium breve* | The microbial ecosystem member |
| Bad | *Bifidobacterium adolescentis* | The microbial ecosystem member |
| Ehal | *Eubacterium hallii* | The microbial ecosystem member |
| Fpr | *Faecalibacterium prausnitzii* | The microbial ecosystem member |
| Rint | *Roseburia intestinalis* | The microbial ecosystem member |

| RA | Relative abundance | Metric of the microbiota composition |

**Table S2** | Summary of information of original genome scale metabolic models (GEMs), the referred literature, and information of the reduced GEMs.

| Species | Information of original GEMs | | | | | Reduced GEMs | | |
|---|---|---|---|---|---|---|---|---|
| Name | Genome-scale metabolic model | No. of rxns[1] | No. of mets[2] | No. of genes | References | No. of rxns[1] | No. of mets[2] | No. of genes |
| *B. thetaiotao micron* | iBth801 v1.00 | 1045 | 877 | 801 | Cell Metabolism, 4;22(2):320-331; Cell host & Microbe, 10(4):336-347 | 109 | 98 | 79 |
| *B. fragilis* | Bacteroides_fragilis_NCTC_9343 | 1045 | 877 | 801 | Nature Biotechnology, 35(1):81-89; Cell host & Microbe, 10(4):336-348 | 109 | 98 | 71 |
| *B. longum* | Bifidobacterium_longum_infantis_ATCC_15697 | 715 | 610 | 452 | Nature Biotechnology, 35(1):81-89; Gene & Nutrition, 6(3): 285-304 | 89 | 81 | 65 |
| *B. breve* | Bifidobacterium_breve_UCC2003_NCIMB8807 | 715 | 610 | 452 | Nature Biotechnology, 35(1):81-89; Gene & Nutrition, 6(3): 285-305 | 89 | 81 | 68 |
| *B. adolescentis* | iBif452.V01.00 | 715 | 610 | 452 | Cell Metabolism, 4;22(2):320-331; Gene & Nutrition, 6(3): 285-306 | 89 | 81 | 68 |
| *E. hallii* | iEre400 v2 00 | 450 | 381 | 400 | Cell Metabolism, 4;22(2):320-33; FEMs Microbiology Letters, 294(1):1:8 | 100 | 91 | 97 |
| *F. praustznii* | iFap484.V01.00 | 729 | 620 | 484 | Cell Metabolism, 4;22(2):320-331; J Bac. 196(18):3289-3302 | 87 | 106 | 87 |
| *R. intestinalis* | Roseburia_intestinalis_L1_82 | 729 | 620 | 484 | Nature Biotechnology, 35(1):81-89; FEMs Microbiology Letters, 294(1):1:8 | 87 | 106 | 66 |

Notes: [1] Number of reactions in associated models.
Notes: [2] Number of metabolites in associated models.

**Table S3 |** Summary of datasets used for evaluation of the three modeling frameworks comprised in CODY, as well for evaluation of CODY, which are employed to compare with corresponding model prediction results.

| Datasets employed for model evaluation | | | | |
|---|---|---|---|---|
| ECMF | HRAF | | SPCF | CODY |
| | Monoculture [40] | Coculture [40] | | |
| *Bacteroides thetaiotaomicron*[33] | *Bacteroides thetaiotaomicron* | *Bacteroides thetaiotaomicron* | Colon physiology[63] | Longitudinal metagenomics study of dynamic infant gut microbiota development during first year of life [1] |
| *Bacteroides fragilis*[33] | *Bifidobacteria adolesentis* | *Bifidobacteria adolesentis* | SCFA production in in vitro TIM setup[44] | Longitudinal metagenomics study of dynamic gut microbiota changes of hepatic steatosis adults under diet-intervention [45] |
| *Bifidobacteria longum*[33] | | | | |
| *Bifidobacteria breve*[34] | | | | |
| *Bifidobacteria adolesentis*[36] | | | | |
| *Eubacterium hallii*[37] | | | | |
| *Faecalibacterium prausztnii*[38] | | | | |
| *Roseburia intestinalis*[30] | | | | |

**Table S4 |** Metabolic capacity of microbial ecosystem members, including growth rate, production of key metabolites and degradation of microbial accessible carbohydrates, where negative values mean corresponding metabolites are consumed. The four categories of polysaccharides used include: HDP representing host-derived polysaccharides, MDP representing breastmilk derived polysaccharide and DDP representing diet-derived polysaccharide, respectively.

| MACs (Microbial-accessible carbohydrates) | Host-derived mucin | Breastmilk oligosaccharide | Dietary fiber | Resistant starch |
|---|---|---|---|---|
| Category | HDP | MDP | DDP | DDP |
| Composition | | | | |
| 4th month | 1 g/d | 14 g/d | - | - |
| 12th month | 1.5 g/d | - | 12.5 g/d | 10g/d |
| Adult control | 2 g/d | - | 33.2g/d | 23.8g/L |
| Adult intervention | 2 g/d | - | 19.7g/d | 4g/d |
| Degradation by microbiota (kd/Kcap) | | | | |
| *B. thetaiotaomicron* | 3/90 | 2/50 | 1/90 | 1/90 |
| *B. fragilis* | 10/30 | 16/40 | 8/20 | 8/50 |
| *B. longum* | 5/30 | 16/20 | 2/50 | 6/30 |
| *B. breve* | 1/90 | 1/90 | 1/90 | 1/90 |
| *B. adolescentis* | 2/50 | 6/30 | 1/80 | 5/50 |
| *E. hallii* | - | - | 9/60 | 3/50 |
| *F. praustznii* | - | - | 15/30 | 2/50 |
| *R. intestinalis* | - | - | 5/60 | 8/15 |

Metabolic Capacity of microbial species

| Representatives | $\mu_{max}(h^{-1})$ | $Y_{acetate}$ | $Y_{butyrate}$ | $Y_{propionate}$ |
|---|---|---|---|---|
| *B. thetaiotaomicron* | 0.5087 | 0.65 | 0 | 0.08 |
| *B. fragilis* | 0.4321 | 0.49 | 0 | 0.11 |
| *B. longum* | 0.5082 | 1.47 | 0 | 0 |
| *B. breve* | 0.4746 | 1.02 | 0 | 0 |
| *B. adolescentis* | 0.4347 | 1.71 | 0 | 0 |
| *E. hallii* | 0.5028 | -0.449 | 0.697 | 0 |
| *F. praustznii* | 0.6837 | -0.63 | 1.09 | 0 |

| | | | | |
|---|---|---|---|---|
| *R. intestinalis* | 0.6632 | -0.899 | 1.175 | 0 |

**Table S5 |** Summary of parameters in mono culture evaluation of two species, i.e., *B.thetaiotaomicron* and *B.adolescentis* under multiple substrates, i.e., maltose, cellubiose and glucose. The parameters are used the same in both the mono culture and coculture of the two species in the evaluation of HRAF. The unit of kmax is mM/gDW/h. The unit of $K_M$ is mM.

| Substrate-associated MPMs | *B. thetaiotaomicon* | | *B. adolescentis* | |
|---|---|---|---|---|
| **Glucose-associated MPMs** | $k_{max}$ | $K_M$ | $k_{max}$ | $K_M$ |
| MPM$_1$ | 80.0807 | 10 | 0.1831 | 10 |
| MPM $_2$ | 51.8401 | 10 | 0.018 | 10 |
| MPM $_3$ | 21.2521 | 10 | 1.8277 | 10 |
| MPM $_4$ | 0.0042 | 10 | 0.0002 | 10 |
| MPM $_5$ | 20.9395 | 10 | 0.0002 | 10 |
| MPM $_6$ | 25.8386 | 10 | 0.0078 | 10 |
| MPM $_7$ | 24.2382 | 10 | 0.5642 | 10 |
| MPM $_8$ | 13.0745 | 10 | 0.4039 | 10 |
| **Maltose-associated MPMs** | $k_{max}$ | $K_M$ | $k_{max}$ | $K_M$ |
| MPM $_9$ | 182.3542 | 50 | 411.9073 | 50 |
| MPM $_{10}$ | 1.8564 | 50 | 4.172 | 50 |
| MPM $_{11}$ | 23.5593 | 50 | 1.8656 | 50 |
| MPM $_{12}$ | 0.4113 | 50 | 8.3734 | 50 |
| MPM $_{13}$ | 53.4456 | 50 | 0.2354 | 50 |
| MPM $_{14}$ | 0.2271 | 50 | 1.0372 | 50 |
| MPM $_{15}$ | 4.8425 | 50 | 0.6833 | 50 |
| MPM $_{16}$ | 182.3542 | 50 | 2.5583 | 50 |
| **Cellubiose - associated MPMs** | $k_{max}$ | $K_M$ | $k_{max}$ | $K_M$ |
| MPM $_{17}$ | 38.3886 | 50 | 1.529 | 50 |
| MPM $_{18}$ | 48.3502 | 50 | 0.8782 | 50 |
| MPM $_{19}$ | 83.2366 | 50 | 1.41 | 50 |
| MPM $_{20}$ | 3.4449 | 50 | 66.7082 | 50 |
| MPM $_{21}$ | 6.8404 | 50 | 3.5839 | 50 |
| MPM $_{22}$ | 58.8268 | 50 | 4.7126 | 50 |
| MPM $_{23}$ | 56.5179 | 50 | 0.5204 | 50 |
| MPM $_{24}$ | 52.1582 | 50 | 1.7026 | 50 |

| | | | | |
|---|---|---|---|---|
| **MPM** $_{25}$ | 0.0047 | 50 | 79.2421 | 50 |
| **MPM** $_{26}$ | - | | 2.1551 | 50 |

**Table S6** | Summary of evaluation results of SPCF by comparing its predictions under different scenarios to TIM experimental observations.

| Variables | TIM model [44, 63] | SPCF colon model Simulations |
|---|---|---|
| Volume | 0.12 L | 0.12L |
| Water absorption | Dialysis | Diffusion |
| SCFA absorption | Dialysis | Diffusion |
| Inflow rate | 4ml/h | 4ml/h |
| Outlet flow rate | 2ml/h | 2ml/h |
| Retention time | 24~72h[63] | 26~50h |
| Peristaltic mixing | Computer-controlled | Back-mixing |
| Sampling position | Only one | Spatiotemporal |
| Carbohydrate fermentation | 2.5g/10ml | 2.5g/10ml |
| Total SCFA produced by FOS in 2hrs | 1.69 mmol | 1.67 mmol |
| Total SCFA produced by Pectin in 2hrs | 1.54 mmol | 1.59 mmol |
| Total SCFA produced by Lactulose in 2hrs | 1.64 mmol | 1.71 mmol |
| Total SCFA produced by Lactitol in 2hrs | 0.52 mmol | 0.43 mmol |

**Table S7** | Summary of colon physiology associated parameters used in the SPCF which are determined based on literature-based knowledges.

| Model Parameters of colon physiology | Symbol | Value |
|---|---|---|
| Volume of total lumen phase | $V_k$ | 0.5 L |
| Volume of total mucus phase | $v_k$ | 0.1 L |
| Volume of rectum | $V_{rectum}$ | [0.05L, 0.15L] |
| Volume of Blood phase | $V_B$ | 1 L |
| Inflow nutrients flow rate | $F_0$ | 0.1 L/h |
| Nutrients concentration during breastfeeding | $S^0$ | 50 g/L |
| Nutrients concentration during solid food feeding | $S^0$ | 80 g/L |
| Flow rate of blood compartment | $F_B$ | 0.5 L/h |
| Thickness of colon wall-membrane that radial dispersion passes across | $a_v$ | 0.5/cm |
| Maximal detachment rate | $k_d$ | [0.01/h, 0.05/h] |
| Pseudo-mass transport coefficient accounting for radial dispersion | $k_m$ | 0.1/h |
| thickness of mucus | $l_m$ | 400 µm |
| Permeability constant of acetate | $\pi_{n,ac}$ | $0.2*10^{-8}$m/s |
| Permeability constant of butyrate | $\pi_{n,but}$ | $0.3 *10^{-8}$m/s |
| Permeability constant of propionate | $\pi_{n,prop}$ | $0.2 *10^{-8}$m/s |
| feces density | $\rho$ | 0.7 g/L |
| Back mixing flow rate due to peristalsis movement | $q_k$ | 0.01 L/h |
| Integration time step | $dt$ | 0.01 h |
| Simulation time during colonization | - | 302 h |
| Simulation time during maturation | - | 600 h |
| Simulation time for adult normal condition | - | 302 h |
| Simulation time for adult intervention | - | 600 h |
| Bacterial biomass typical molecular weight | $M_{biom}$ | 113 g/mol |
| Time interval of two meals | $t$ | 1 h |
| Defecation time period | $T$ | 5 min |
| Molar mass of acetate | $M_{ac}$ | 59.04 g/mol |
| Molar mass of butyrate | $M_{but}$ | 87. 098 g/mol |

**Table S8 |** Summary of top features used in PCA analysis for three phenotypic data of dynamic trajectories of gut microbiota development, i.e., lumen, mucus and feces.

| Feature name | Keywords [3] |
|---|---|
| MF_AR_arcov_5_a3 | model, fit |
| MF_AR_arcov_5_a4 | model, fit |
| MF_AR_arcov_5_a5 | model, fit |
| MF_CompareTestSets_y_ar_4_rand_25_01_1_ac1s_mean | Model prediction, stationarity |
| MF_CompareTestSets_y_ar_4_rand_25_01_1_ac1s_median | Model prediction, stationarity |
| MF_StateSpace_n4sid_2_05_1_minfpe | model |
| MF_armax_3_1_05_1_minfpe | model |
| MF_armax_2_2_05_1_p3_5 | model |
| MF_armax_2_2_05_1_p4_5 | model |
| MF_armax_2_2_05_1_minfpe | model |

Note: [3] The meaning for all features could be referred to the figure legend of the Fig. S25.

## Supplementary References

1.  Bäckhed, F. et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host & Microbe* **17**, 690-703 (2015).
2.  Wiener, N. Cybernetics: or Control and Communication in the Animal and the Machine. (1948).
3.  Ramkrishna, D. & Song, H.-S. Dynamic models of metabolism: Review of the cybernetic approach. *AIChE Journal* **58**, 986-997 (2012).
4.  Shoaie, S. et al. Quantifying Diet-Induced Metabolic Changes of the Human Gut Microbiome. *Cell Metabolism* **22**, 320-331.
5.  Magnusdottir, S. et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat Biotech* **35**, 81-89 (2017).
6.  Fischbach, M.A. & Sonnenburg, J.L. Eating For Two: How Metabolism Establishes Interspecies Interactions in the Gut. *Cell host & microbe* **10**, 336-347 (2011).
7.  Pokusaeva, K., Fitzgerald, G.F. & van Sinderen, D. Carbohydrate metabolism in Bifidobacteria. *Genes & Nutrition* **6**, 285-306 (2011).
8.  Louis, P. & Flint, H.J. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiology Letters* **294**, 1-8 (2009).
9.  Macfarlane, G.T. & Gibson, G.R. in Gastrointestinal Microbiology: Volume 1 Gastrointestinal Ecosystems and Fermentations. (eds. R.I. Mackie & B.A. White) 269-318 (Springer US, Boston, MA; 1997).
10. Rivière, A., Gagnon, M., Weckx, S., Roy, D. & De Vuyst, L. Mutual Cross-Feeding Interactions between Bifidobacterium longum subsp. longum NCC2705 and Eubacterium rectale ATCC 33656 Explain the Bifidogenic and Butyrogenic Effects of Arabinoxylan Oligosaccharides. *Applied and Environmental Microbiology* **81**, 7767-7781 (2015).
11. Kettle, H., Louis, P., Holtrop, G., Duncan, S.H. & Flint, H.J. Modelling the emergent dynamics and major metabolites of the human colonic microbiota. *Environmental Microbiology* **17**, 1615-1630 (2015).
12. Heinken, A. et al. A functional metabolic map of Faecalibacterium prausnitzii, a beneficial human gut microbe. *Journal of Bacteriology* (2014).
13. Agren, R. et al. The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum. *PLOS Computational Biology* **9**, e1002980 (2013).
14. Song, H.-S. & Ramkrishna, D. Reduction of a set of elementary modes using yield analysis. *Biotechnology and Bioengineering* **102**, 554-568 (2009).
15. Young, J.D. & Ramkrishna, D. On the Matching and Proportional Laws of Cybernetic Models. *Biotechnology Progress* **23**, 83-99 (2007).
16. Vavilin, V.A., Fernandez, B., Palatsi, J. & Flotats, X. Hydrolysis kinetics in anaerobic degradation of particulate organic material: An overview. *Waste Management* **28**, 939-951 (2008).
17. Muñoz-Tamayo, R., Laroche, B., Walter, É., Doré, J. & Leclerc, M. Mathematical modelling of carbohydrate degradation by human colonic microbiota. *Journal of Theoretical Biology* **266**, 189-201 (2010).
18. Sonnenburg, J.L., Angenent, L.T. & Gordon, J.I. Getting a grip on things: how do communities of bacterial symbionts become established in our intestine? *Nature Immunology* **5**, 569 (2004).
19. Gayer, C.P. & Basson, M.D. The effects of mechanical forces on intestinal physiology and pathology. *Cellular signalling* **21**, 1237-1244 (2009).
20. O.Levenspiel Chemical Reaction Engineering. *J. Wiley & Sons, USA* (1999).

21. Pritchard, S.E. et al. Fasting and postprandial volumes of the undisturbed colon: normal values and changes in diarrhea-predominant irritable bowel syndrome measured using serial MRI. *Neurogastroenterology & Motility* **26**, 124-130 (2014).

22. Hounnou, G., Destrieux, C., Desmé, J., Bertrand, P. & Velut, S. Anatomical study of the length of the human intestine. *Surgical and Radiologic Anatomy* **24**, 290-294 (2002).

23. Helander, H.F. & Fändriks, L. Surface area of the digestive tract – revisited. *Scandinavian Journal of Gastroenterology* **49**, 681-689 (2014).

24. Ababneh, Z.Q. et al. In vivo lipid diffusion coefficient measurements in rat bone marrow. *Magnetic Resonance Imaging* **27**, 859-864 (2009).

25. Phillips, S.F. & Giller, J. The contribution of the colon to electrolyte and water conservation in man. *The Journal of Laboratory and Clinical Medicine* **81**, 733-746.

26. SANDLE, G.I. Salt and water absorption in the human colon: a modern appraisal. *Gut* **43**, 294-299 (1998).

27. Christl, S.U. & Scheppach, W. Metabolic Consequences of Total Colectomy. *Scandinavian Journal of Gastroenterology* **32**, 20-24 (1997).

28. Chang, E.B. & Leung, P.S. in The Gastrointestinal System: Gastrointestinal, Nutritional and Hepatobiliary Physiology. (ed. P.S. Leung) 107-134 (Springer Netherlands, Dordrecht; 2014).

29. Taghipoor, M., Barles, G., Georgelin, C., Licois, J.R. & Lescoat, P. Digestion modeling in the small intestine: Impact of dietary fiber. *Mathematical Biosciences* **258**, 101-112 (2014).

30. Duncan, S.H., Barcenilla, A., Stewart, C.S., Pryde, S.E. & Flint, H.J. Acetate Utilization and Butyryl Coenzyme A (CoA):Acetate-CoA Transferase in Butyrate-Producing Bacteria from the Human Large Intestine. *Applied and Environmental Microbiology* **68**, 5186-5190 (2002).

31. Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. & Gilles, E.D. Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420**, 190-193 (2002).

32. Kamp, A.V. & Schuster, S. Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics* **22**, 1930-1931 (2006).

33. Van der Meulen, R., Makras, L., Verbrugghe, K., Adriany, T. & De Vuyst, L. In Vitro Kinetic Analysis of Oligofructose Consumption by Bacteroides and Bifidobacterium spp. Indicates Different Degradation Mechanisms. *Applied and Environmental Microbiology* **72**, 1006-1012 (2006).

34. Rios-Covian, D. et al. Interactions between Bifidobacterium and Bacteroides Species in Cofermentations Are Affected by Carbon Sources, Including Exopolysaccharides Produced by Bifidobacteria. *Applied and Environmental Microbiology* **79**, 7518-7524 (2013).

35. Van der Meulen, R., Adriany, T., Verbrugghe, K. & De Vuyst, L. Kinetic Analysis of Bifidobacterial Metabolism Reveals a Minor Role for Succinic Acid in the Regeneration of NAD(+) through Its Growth-Associated Production. *Applied and Environmental Microbiology* **72**, 5204-5210 (2006).

36. Falony, G. et al. In Vitro Kinetic Analysis of Fermentation of Prebiotic Inulin-Type Fructans by Bifidobacterium Species Reveals Four Different Phenotypes. *Applied and Environmental Microbiology* **75**, 454-461 (2009).

37. Muñoz-Tamayo, R. et al. Kinetic modelling of lactate utilization and butyrate production by key human colonic bacterial species, Vol. 76. (2011).

38. Rios-Covian, D., Gueimonde, M., Duncan, S.H., Flint, H.J. & de los Reyes-Gavilan, C.G. Enhanced butyrate formation by cross-feeding between Faecalibacterium prausnitzii and Bifidobacterium adolescentis. *FEMS Microbiology Letters* **362** (2015).

39.    Belenguer, A. et al. Two Routes of Metabolic Cross-Feeding between Bifidobacterium adolescentis and Butyrate-Producing Anaerobes from the Human Gut. *Applied and Environmental Microbiology* **72**, 3593-3599 (2006).

40.    Das, P., Ji, B., Kovatcheva-Datchary, P., Bäckhed, F. & Nielsen, J. In vitro co-cultures of human gut bacterial species as predicted from co-occurrence network analysis. *PLOS ONE* **13**, e0195161 (2018).

41.    Sundararaj, S. et al. The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of Escherichia coli. *Nucleic Acids Research* **32**, D293-D295 (2004).

42.    Freilich, S. et al. Competitive and cooperative metabolic interactions in bacterial communities. *Nature Communications* **2**, 589 (2011).

43.    Faust, K. & Raes, J. Microbial interactions: from networks to models. *Nat Rev Micro* **10**, 538-550 (2012).

44.    Minekus, M. et al. A computer-controlled system to simulate conditions of the large intestine with peristaltic mixing, water absorption and absorption of fermentation products, Vol. 53. (2000).

45.    Mardinoglu, A. et al. An Integrated Understanding of the Rapid Metabolic Benefits of a Carbohydrate-Restricted Diet on Hepatic Steatosis in Humans. *Cell Metabolism* **27**, 559-571.e555 (2018).

46.    Underwood, M.A., German, J.B., Lebrilla, C.B. & Mills, D.A. Bifidobacterium longum subspecies infantis: champion colonizer of the infant gut. *Pediatric Research* **77**, 229 (2014).

47.    He, Y., Liu, S., Leone, S. & Newburg, D.S. Human colostrum oligosaccharides modulate major immunologic pathways of immature human intestine. *Mucosal Immunology* **7**, 1326 (2014).

48.    Macfarlane, S., Woodmansey, E.J. & Macfarlane, G.T. Colonization of Mucin by Human Intestinal Bacteria and Establishment of Biofilm Communities in a Two-Stage Continuous Culture System. *Applied and Environmental Microbiology* **71**, 7483-7492 (2005).

49.    Bingham, S.A. et al. Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): an observational study. *The Lancet* **361**, 1496-1501 (2003).

50.    Salyers, A.A., Vercellotti, J.R., West, S.E. & Wilkins, T.D. Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon. *Applied and Environmental Microbiology* **33**, 319-322 (1977).

51.    Derrien, M. et al. Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut Microbes* **1**, 254-268 (2010).

52.    Barboza, M. et al. Glycoprofiling Bifidobacterial Consumption of Galacto-Oligosaccharides by Mass Spectrometry Reveals Strain-Specific, Preferential Consumption of Glycans. *Applied and Environmental Microbiology* **75**, 7319-7325 (2009).

53.    Marcobal, A. et al. Consumption of Human Milk Oligosaccharides by Gut-related Microbes. *Journal of agricultural and food chemistry* **58**, 5334-5340 (2010).

54.    Asakuma, S. et al. Physiology of Consumption of Human Milk Oligosaccharides by Infant Gut-associated Bifidobacteria. *Journal of Biological Chemistry* **286**, 34583-34592 (2011).

55.    Mirande, C. et al. Dietary fibre degradation and fermentation by two xylanolytic bacteria Bacteroides xylanisolvens XB1AT and Roseburia intestinalis XB6B4 from the human intestine. *Journal of Applied Microbiology* **109**, 451-460 (2010).

56.    Flint, H.J., Scott, K.P., Duncan, S.H., Louis, P. & Forano, E. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **3**, 289-306 (2012).

57.    Martens, E.C. et al. Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts. *PLOS Biology* **9**, e1001221 (2011).

58.    Birt, D.F. et al. Resistant Starch: Promise for Improving Human Health. *Advances in Nutrition* **4**, 587-601 (2013).

59.    Ze, X., Duncan, S.H., Louis, P. & Flint, H.J. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *The Isme Journal* **6**, 1535 (2012).

60.    Duncan, S.H. et al. Proposal of Roseburia faecis sp. nov., Roseburia hominis sp. nov. and Roseburia inulinivorans sp. nov., based on isolates from human faeces. *International Journal of Systematic and Evolutionary Microbiology* **56**, 2437-2441 (2006).

61.    Fulcher, B.D. & Jones, N.S. hctsa: A Computational Framework for Automated Time-Series Phenotyping Using Massive Feature Extraction. *Cell Systems* **5**, 527-531.e523.

62.    Fulcher, B.D., Little, M.A. & Jones, N.S. Highly comparative time-series analysis: the empirical structure of time series and their methods. *Journal of The Royal Society Interface* **10** (2013).

63.    Venema, K. in The Impact of Food Bioactives on Health: in vitro and ex vivo models. (eds. K. Verhoeckx et al.) 293-304 (Springer International Publishing, Cham; 2015).