Supplementary Information

1 Existing machine learning (ML) methods

MelonnPan. MelonnPan is a computational method based on the regularized linear regression for predicting metabolite composition from microbiome sequencing data[1].

- Data processing: It applies the rank-transformation to the normalized microbial abundances. Specifically, it utilizes the quantile transformation that transforms features to follow a normal distribution. Normalized metabolite profiles are arcsine square root transformed as the outputs to be predicted by the model.

- Model detail: The model is a linear regression model with elastic net regularization. The elastic net regularization is a hybrid method that linearly combines the L1 (i.e., LASSO) and L2 (i.e., ridge) penalties. It has the elastic net mixing parameter α (the fraction of L2 between L1 and L2) to adjust the ratio between two types of regularizations and sparsity parameter λ (penalty weight) to modify the overall parameter sparsity. For each metabolite, one elastic net model is designed to achieve the best performance.

- Hyperparameter selection: Two hyperparameters are selected based on the 5-fold cross-validation results (the mean Spearman's rank correlation coefficients) on the training set: the elastic net mixing parameter α and sparsity parameter λ. α is selected from [0.1, 0.5, 0.9] and λ is selected from [10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}].

Sparse NED. Le et al proposed an MLP (Multiple-Layer Perceptron) model with one hidden layer with fewer model parameters to predict metabolite clusters[2].

- Data processing: It applies the CLR (Centered Log-Ratio) transformation[3] to both microbial abundances and metabolite concentrations since both omics data are compositional. After that, the Z-score transformation, widely used in machine learning, is applied to both data types.

- Model detail: The model is a sparsified MLP model with one hidden layer. the dimension of the hidden layer is N_{h}. The learning process is made of two steps: the screening stage

and the training stage. During the screening stage, the MLP model with fully connected weights is trained and connections that are most useful in extracting the information needed to predict metabolite concentrations from microbe abundances are identified. The connection importance is measured by the normalized magnitude of the derivatives. The connection with a larger magnitude for its derivative is considered to be more important. Only connections ranked as the top β percentile are kept and β is a sparsity parameter that can be adjusted to change the percent of weights kept. Then in the training stage, the MLP model with less important connection deactivated (or masked) from the forward-feed and backpropagation operations.

- Training method: Adam (Adaptive Moment Estimation) optimizer[4] is used for the gradient descent. the training stops after 100 epochs.

- Activation function: the hyperbolic tangent function tanh.

- Hyperparameter selection: Two hyperparameters are selected based on the 5-fold cross-validation results (the mean Spearman's rank correlation coefficients) on the training set: the dimension of the hidden layer N_{h} and sparsity parameter β. N_{h} is selected from [32, 64, 128] and β is selected from [0.05, 0.1, 0.2, 0.5].

MiMeNet. MiMeNet (Microbiome-Metabolome Network) is a method based on the multilayer perceptron neural network (MLPNN) with two ways to prevent the overfitting: L2 regularization and dropout[5]. It attempted to predict the entire metabolite profile and learn the microbe-metabolite interactions using the feature attribution scores[5].

- Data processing: The Z-score transformation is applied to the CLR (Centered Log-Ratio) transformed microbial abundances and metabolite concentrations.

- Model detail: The model is an MLP model with one hidden layer or several hidden layers. The L2 regularization with weight parameter λ is adopted to sparsify the number of model parameters. In addition, the dropout with a rate r at each hidden layer (i.e., a random fraction of nodes and corresponding weights are masked temporarily) is applied to further regularize the MLP model.

- Training method: Adam optimizer[4] is used for the gradient descent. The training stops if the loss on the validation/test set has not improved within the past 40 epochs.

- Activation function: Rectified Linear Unit (ReLU).

- Hyperparameter selection: Four hyperparameters are selected based on the 5-fold cross-validation results (the mean Spearman's rank correlation coefficients) on the training set: the dimension of the hidden layer N_{h}, the number of hidden layers N_{ℓ}, the L2 penalty with weight parameter λ, and the dropout rate r. N_{h} is selected from [32, 128, 512], N_{ℓ} is selected from [1, 2, 3], λ is selected from [10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}], and r is selected from [0.1, 0.3, 0.5].

ResNet. The ResNet (Residual neural Network) is a deep learning method based on the idea of residual blocks and skip connections.

- Data processing: The Z-score transformation is applied to the CLR (Centered Log-Ratio) transformed microbial abundances and metabolite concentrations.

- Model detail: We adapted the ResNet (Residual neural Networks) architecture from Goyal et al[6]. Goyal et al proposed the Linear-QuadraticResidual Network (LQResNet), which linearly combines the linear and quadratic mappings with the residual neural network to model the first-order time-derivative of variables[6]. The architecture consists of 3 connected modules: (1) one fully connected layer that maps the input (such as the microbial composition) to the hidden layer with dimension N_{h} followed by an activation function, (2) N residual blocks with each block made of a one-hidden-layer MLP with the layer dimension the same as N_{h} plus the skip connection, and (3) one fully connected layer that maps from the hidden layers to the output (i.e., the metabolomic profiles). The L2 regularization with the weight parameter λ is assumed to prevent overfitting.

- Training method: RAdam (Rectified Adam) optimizer[7], which utilizes a warm-up strategy to rectify the variance of the adaptive learning rate, is used for the gradient descent. The training stops after 100 epochs.

- Activation function: Exponential Linear Unit (ELU).

- Hyperparameter selection: Three hyperparameters are selected based on the 5-fold cross-validation results (the mean Spearman's rank correlation coefficients) on the training set: the dimension of the hidden layer N_{h}, the number of residual blocks N, and the L2 penalty with weight parameter λ. N_{h} is chosen to be the same as, 2 times, or 3 times the input dimension, N is selected from [2, 3, 4], and λ is selected from [1, 5, 20].

2 Microbial Consumer-Resource Model (MiCRM) with cross-feeding interactions

Similar to the formalism of ecological models developed for microbial communities[8], we considered how the nutrients supplied to a microbial ecosystem are consumed and other byproduct nutrients are further produced by microbes. The supply rate of nutrient α is h_α. Also, the system is assumed to be constantly diluted with a dilution rate δ. For each microbial species i, the consumption rate of nutrient α per microbial density per nutrient concentration is assumed to be $a_{i\alpha}$. As a result, the overall consumption rate of nutrient α by species i is $J_{i\alpha}^{con} = a_{i\alpha} N_i R_\alpha$ where N_i is the density of species i and R_α is the concentration of nutrient α. Upon consumption, a fraction of consumed nutrients $1 - l$ are assumed to contribute to the growth of microbes and the remaining fraction l is converted to other byproducts. The byproduct conversion flux from nutrient α to nutrient β is encoded by $D_{\beta\alpha}$ and thus the matrix D encodes the partitioning of one nutrient to other byproduct nutrients. To conserve fluxes of nutrients, $\sum_\beta D_{\beta\alpha} = 1$ is assumed. Therefore, the total consumption rate of nutrient α by all species is $J_\alpha^{con} = \sum_i J_{i\alpha}^{con} = \sum_i a_{i\alpha} N_i R_\alpha$, while the total production rate of nutrient α is $J_\alpha^{pro} = l \sum_\beta D_{\alpha\beta} J_\beta^{con} = l \sum_\beta \sum_i D_{\alpha\beta} a_{i\beta} N_i R_\beta$. For each species, the growth rate $g_i N_i = \frac{(1-l)\sum_\alpha a_{i\alpha} N_i R_\alpha}{Y}$, which is the sum of consumption rates on all nutrients that are not converted to byproducts divided by the yield Y. g_i is also termed as the specific growth rate of species i. Overall, the dynamics of the dynamics for concentrations of nutrients R_α and microbial species abundance N_i:

$$\frac{dR_\alpha}{dt} = h_\alpha - \delta R_\alpha - J_\alpha^{con} + J_\alpha^{pro}, \tag{S1}$$

$$\frac{dN_i}{dt} = -\delta N_i + g_i N_i. \tag{S2}$$

To obtain the production matrix used in Fig. 5d, we multiply the consumption matrix with the byproduct conversion matrix D. More specifically, the production flux of byproduct α by species i is written as $P_{i\alpha} = \sum_\beta D_{\alpha\beta} a_{i\beta}$ which is a sum of all possible byproducts produced by species i when it consumes all available nutrients.

3 Synthetic data from MiCRM

We used the above MiCRM with 10 microbial species and 10 nutrients to generate the synthetic cross-sectional data. All MiCRM parameters such as consumption rates, dilution rates, and byproduct generation rates are assumed to be the same. Different samples are considered as a random sampling of microbial species and nutrients to assemble. Overall, our procedure for generating synthetic data can be divided into three steps: (1) the metapopulation establishment stage: for the metapopulation with all potential species and nutrients, model parameters are drawn from their corresponded probability distributions, (2) the subsampling stage: a fraction of microbial species and nutrients are selected to start the community assembly, and (3) simulation stage: dynamics of sampled species and nutrients are simulated as specified in MiCRM until the synthetic community reaches a steady state. For this community, we collected steady-state relative abundances for all microbial species as the microbial composition, steady-state nutrient concentrations as the corresponding metabolomic profile, and nutrient supply rates as the diet. This 3-stage procedure is repeated many times to create microbial compositions, metabolomic profiles, and diets to form independent samples in the synthetic data.

More specifically, during the first stage, model parameters are determined as follows:

- the chance of one random species consuming one random nutrient is assumed to be 20%. The rate $a_{i\alpha}$ is drawn from the uniform distribution between 0 and 100. After the random drawing, the rate $a_{i\alpha}$ is divided by the number of nutrients the species i can consume to prevent the existence of superbugs.

- the connectance of the byproduct conversion matrix D is assumed to be 50%. In practice,

each entry in the matrix D has a probability of 50% to be non-zero. The non-zero entries are drawn from the uniform distribution between 0 and 1. After the drawing, the normalization is imposed to guarantee $\sum_\beta D_{\beta\alpha} = 1$.

- the byproduct fraction $l = 0.5$ for all cases.
- the dilution rate $\delta = 0.2$ hour^{-1}.
- the yield $Y = 1$.
- the nutrient supply rate h_α is drawn from the uniform distribution between 0 and 1.

During the second stage, for each sample, 50% of species are randomly chosen to be introduced in the initial pool (i.e., $p_{\rm s} = 0.5$) and nutrients are randomly chosen with the sampling probability $p_{\rm n}$ to have non-zero supplies (as defined in the nutrient supply rate h_i in the first stage).

4 Modified MiCRM with species-specific byproduct generation and no overlap between consumption and production interactions

The MiCRM above assumes a universal byproduct generation that is encoded by the byproduct conversion matrix D and preserved for all species. This would lead to a case where one metabolite can be consumed and produced by one species at the same time (i.e. overlap between consumption and production interactions for the same metabolite-species pairs). To avoid the overlap, here we considered a more generic case where the byproduct generation is specific to each species. After microbial species i consumes all available nutrients in the community, all consumed nutrients are divided into two parts: (1) a fraction l of total consumed nutrients by species i contributes to the biomass growth, and (2) the other fraction $1 - l$ of total consumed nutrients by species i are converted to byproducts, with their production fluxes proportional to the pre-specified production rate by species i (written as $P_{i\alpha}$). $\sum_\alpha P_{i\alpha} = 1$ is imposed to conserve the total flux of nutrients. Other dynamics such as growth dynamics, nutrient consumption, and dilutions are the same as the previous MiCRM. Overall, the dynamics of nutrient concentrations

R_α and microbial species abundances N_i are as follows:

$$\frac{dR_\alpha}{dt} = h_\alpha - \delta R_\alpha - \sum_i a_{i\alpha} N_i R_\alpha + \sum_i P_{i\alpha} l \sum_\beta a_{i\beta} N_i R_\beta, \tag{S3}$$

$$\frac{dN_i}{dt} = -\delta N_i + \frac{(1-l)\sum_\alpha a_{i\alpha} N_i R_\alpha}{Y}. \tag{S4}$$

All parameters follow the same definitions specified in the previous MiCRM except for species-specific byproduct production rates $P_{i\alpha}$. The protocol for generating synthetic data by this new version of MiCRM follows the protocol for the previous MiCRM except:

- the chance of one random species producing one random byproduct is assumed to be 50%. $P_{i\alpha}$ is drawn from the uniform distribution between 0 and 1. To avoid the overlap between consumption and production interactions between the same microbe-metabolite pairs, any overlapped entries in the production matrix will be set as zero. After the random drawing and setting values of overlapped entries as zero, the normalization over each species is imposed to guarantee $\sum_\alpha P_{i\alpha} = 1$.

5 A summary of ML-based and non-ML computational methods to predict metabolomic profiles from microbial compositions

Numerous computational methods have been proposed to achieve this goal, and they can be divided into the following three categories. (1) Reference-based methods such as MAMBO[9], MIMOSA[10], and Mangosteen[11]. In MAMBO (Metabolomic Analysis of Metagenomes using flux Balance analysis and Optimization), reference microbial genomes are used to reconstruct genome-scale metabolic models (GEMs). Then the microbial composition is correlated with the biomass production of the GEMs (obtained through the flux balance analysis). Finally, the correlation is optimized by multiple iterations of a Monte Carlo-based sampling algorithm. Both MIMOSA (Model-based Integration of Metabolite Observations and Species Abundances) and Mangosteen (Metagenome-based Metabolome Prediction) are reference-based, gene-to-metabolite prediction methods. Note that all those reference-based methods rely heavily

on the completeness and accuracy of queried databases and GEMs. (2) Ecology-guided methods, where abundances of both microbes and metabolites are considered as end-products of the metabolic cascade, propagating through ecological networks of microbial communities (i.e., metabolite consumption and byproduct generation reactions by microbes)[12–15]. Those methods heavily rely on the completeness and accuracy of ecological networks of microbial communities. (3) Machine learning (ML)-based methods, which are trained from paired microbiome and metabolome datasets, and then used to predict the metabolic profile of a never-seen microbiome sample based on its microbial composition, without using any reference database or domain knowledge regarding relationships between genes and metabolites. Various ML techniques such as elastic net[1], sparsified NED (Neural Encoder-Decoder)[2], multilayer perceptron[5], and word2vec[16] have been employed to predict metabolic profiles from microbial compositions. Despite the fact that these ML-based methods circumvent limitations of reference-based or ecology-guided methods discussed above and have been shown to generate promising results in various contexts, none of these ML-based methods utilize state-of-the-art deep neural network models such as the Neural Ordinary Differential Equation (NODE)[17], so their performance has not been fully maximized.

Compared to reference-based methods[9–11], ML-based methods are better at predicting metabolomic profiles and do not require complete GEMs as inputs. For example, it has been previously shown that MelonnPan produces 130 well-predicted KEGG metabolites for the PRISM+NLIBD dataset (measured by having a Spearman correlation coefficient ρ larger than 0.3) without using metabolic models, much higher than 20 for MIMOSA[1]. The performance of reference-based methods is limited due to the lack of complete GEMs for many microbial species. Therefore, it is hard to find a close match of metabolic models for each microbial species in the experimental data. By contrast, ML-based models do not require genomic information as the prerequisite input and yet yield much better predictions for metabolomic profiles because of their flexibility to learn metabolomic activities based on various microbial features. In addition, well-trained ML-based models can be interpreted eventually to suggest some unknown microbe-metabolite interactions. The susceptibility measure proposed here presents a way to reveal potentially existing microbe-metabolite interactions. However, whether suggested interactions do exist has to be tested experimentally. Some suggested microbe-metabolites

interactions might be false positives. After all, the prediction performance is far from perfect (the mean Spearman correlation coefficient $\bar{\rho}$ smaller than 0.5 for most cases).

6 FFQ (Food frequency questionnaire) and FNDDS (USDA's Food and Nutrient Database for Dietary Studies)

The food frequency questionnaire (FFQ) is commonly used to capture food and beverage consumption over time[18–21]. The questionnaire consists of a finite list of foods and beverages with different choice answers that reveal the typical frequency of consumption over the specific time interval queried, such as the frequency of broccoli eaten in a week for the past year. FFQs come in many varieties, such as the Harvard Willett FFQ[18] and the NHANES (National Health and Nutrition Examination Survey) FFQ[19–21]. In VDAART, the FFQ followed and slightly modified the 87-item validated FFQ in preschool-age children[22]. Specifically, the FFQ in VDAART appears as a section in the 36 Months Quarterly Infant Follow-up Questionnaire[23]. Later, we converted their food consumption frequencies into the nutritional profiles based on the nutrient composition of each documented food. The conversion relies on the FNDDS (USDA's Food and Nutrient Database for Dietary Studies)[24] database which encodes the detailed amounts of nutrient components in food and beverage items.

7 The flexibility of NODE and mNODE

NODE (Neural Ordinary Differential Equations) integrates the well-developed ODE solving techniques of the past 120 years with deep learning by "unrolling" the implicit layer that approximates first-order time derivatives of the dynamical systems. Because of this, researchers are increasingly interested in the possibility of solving dynamical systems problems using the architecture of NODE. For instance, recently Dutta et al showed that NODE can generate the solutions for the various evolution problems of different fluid dynamics and further provide a promising potential to extrapolate[25]. Another attempt was to use NODE to learn ecological and evolutionary processes based on the time-series data generated by traditional ecological models[26]. Microbial communities are dynamical systems, in which microbes interact primarily

through the metabolite consumption and production of byproduct metabolites[8,27]. Nutrients provided periodically (such as dietary fibers in the diet for gut microbiomes) or continuously (such as nutrient fluxes in chemostats) to a microbial community can be consumed by microbes and converted to other byproducts by the microbes' metabolism. The experimentally measured metagenome and metabolome of samples from microbial communities at different times can be considered as the profiling of microbial abundances and metabolites concentrations at corresponding times. Consequently, we expect that leveraging the NODE framework with the correct input data types to determine the community dynamics of microbes and their metabolism would generate better performance. Out of many input data types, we believe that diet/nutrient information is an important one.

It is worth noting that our mNODE method and the original NODE[17] is very generic, which makes it easy to apply it to other biological problems where the dynamics of the system play an essential role. For example, it is possible to apply this model framework to predict one omics data type such as the metatranscriptome and metaproteome from the other omics data such as the metagenome. For many bacteria like *Staphylococcus aureus* and *Bacillus subtilis*, the expression of genes in the DNA to mRNAs and its further translation to proteins are dynamic processes that are tightly regulated[28]. Since the protein levels reflect metabolic activities of microbes better than their genomic information, if we can design a model to predict metaproteome from metagenome, we might be one step closer to unraveling the function of microbes in their communities. Also, such a connection between microbial genome and their proteome may enable us to predict phenotypes of microbes such as their specific growth rates in different environments if environmental data is available.

References

[1] Mallick, H. *et al.* Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. *Nature communications* **10**, 1–11 (2019).

[2] Le, V., Quinn, T. P., Tran, T. & Venkatesh, S. Deep in the bowel: highly interpretable neural encoder-decoder networks predict gut metabolites from gut microbiome. *BMC genomics* **21**, 1–15 (2020).

[3] Quinn, T. P. *et al.* A field guide for the compositional analysis of any-omics data. *GigaScience* **8**, giz107 (2019).

[4] Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[5] Reiman, D., Layden, B. T. & Dai, Y. Mimenet: Exploring microbiome-metabolome relationships using neural networks. *PLoS Computational Biology* **17**, e1009021 (2021).

[6] Goyal, P. & Benner, P. Lqresnet: A deep neural network architecture for learning dynamic processes. *arXiv preprint arXiv:2103.02249* (2021).

[7] Liu, L. *et al.* On the variance of the adaptive learning rate and beyond. *arXiv preprint arXiv:1908.03265* (2019).

[8] Marsland III, R. *et al.* Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities. *PLoS computational biology* **15**, e1006793 (2019).

[9] Garza, D. R., van Verk, M. C., Huynen, M. A. & Dutilh, B. E. Towards predicting the environmental metabolome from metagenomics with a mechanistic model. *Nature microbiology* **3**, 456–460 (2018).

[10] Noecker, C. *et al.* Metabolic model-based integration of microbiome taxonomic and metabolomic profiles elucidates mechanistic links between ecological and metabolic variation. *MSystems* **1**, e00013–15 (2016).

[11] Yin, X. *et al.* A comparative evaluation of tools to predict metabolite profiles from microbiome sequencing data. *Frontiers in microbiology* **11**, 3132 (2020).

[12] 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).

[13] Quinn, R. A. *et al.* Niche partitioning of a pathogenic microbiome driven by chemical gradients. *Science advances* **4**, eaau1908 (2018).

[14] Wang, T., Goyal, A., Dubinkina, V. & Maslov, S. Evidence for a multi-level trophic organization of the human gut microbiome. *PLoS computational biology* **15**, e1007524 (2019).

[15] Goyal, A., Wang, T., Dubinkina, V. & Maslov, S. Ecology-guided prediction of cross-feeding interactions in the human gut microbiome. *Nature communications* **12**, 1–10 (2021).

[16] Morton, J. T. *et al.* Learning representations of microbe–metabolite interactions. *Nature methods* **16**, 1306–1314 (2019).

[17] Chen, R. T., Rubanova, Y., Bettencourt, J. & Duvenaud, D. Neural ordinary differential equations. *arXiv preprint arXiv:1806.07366* (2018).

[18] Harvard, T. Harvard willett food frequency questionnaire. *TH Chan School of Public Health, Department of Nutrition, Harvard University: Boston, MA, USA* .

[19] for Health Statistics (US), N. C. *Plan and operation of the third National Health and Nutrition Examination Survey, 1988-94*. 32 (National Ctr for Health Statistics, 1994).

[20] Nelson, K. M., Reiber, G. & Boyko, E. J. Diet and exercise among adults with type 2 diabetes: findings from the third national health and nutrition examination survey (nhanes iii). *Diabetes care* **25**, 1722–1728 (2002).

[21] Marriott, B. P., Olsho, L., Hadden, L. & Connor, P. Intake of added sugars and selected nutrients in the united states, national health and nutrition examination survey (nhanes) 2003—2006. *Critical reviews in food science and nutrition* **50**, 228–258 (2010).

[22] Blum, R. E. *et al.* Validation of a food frequency questionnaire in native american and caucasian children 1 to 5 years of age. *Maternal and child health journal* **3**, 167–172 (1999).

[23] Lee-Sarwar, K. A. *et al.* Integrative analysis of the intestinal metabolome of childhood asthma. *Journal of Allergy and Clinical Immunology* **144**, 442–454 (2019).

[24] Moshfegh, A. Food and nutrient database for dietary studies (fndds) .

[25] Dutta, S., Rivera-Casillas, P. & Farthing, M. W. Neural ordinary differential equations for data-driven reduced order modeling of environmental hydrodynamics. *arXiv preprint arXiv:2104.13962* (2021).

[26] Bonnaffé, W., Sheldon, B. C. & Coulson, T. Neural ordinary differential equations for ecological and evolutionary time-series analysis. *Methods in Ecology and Evolution* (2021).

[27] Gonze, D., Coyte, K. Z., Lahti, L. & Faust, K. Microbial communities as dynamical systems. *Current opinion in microbiology* **44**, 41–49 (2018).

[28] Duval, M., Simonetti, A., Caldelari, I. & Marzi, S. Multiple ways to regulate translation initiation in bacteria: mechanisms, regulatory circuits, dynamics. *Biochimie* **114**, 18–29 (2015).

Supplementary Table Legends

Supplementary Table: Susceptibility values for PRISM + NLIBD, lung samples from patients with cystic fibrosis, soil biocrust samples, fecal samples of children in VDAART, and blood plasma samples of children in VDAART.