Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nat Mach Intell. Author manuscript; available in PMC 2024 January 12.

Published in final edited form as:

Nat Mach Intell. 2023 March ; 5(3): 284–293. doi:10.1038/s42256-023-00627-3.

Predicting metabolomic profiles from microbial composition through neural ordinary differential equations

Tong Wang[1], **Xu-Wen Wang**[1], **Kathleen A. Lee-Sarwar**[1,2], **Augusto A. Litonjua**[3], **Scott T. Weiss**[1], **Yizhou Sun**[4], **Sergei Maslov**[5,6], **Yang-Yu Liu**[1,5,*]

[1]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA.

[2]Division of Allergy and Clinical Immunology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA.

[3]Pediatric Pulmonology, Golisano Children's Hospital, University of Rochester, Rochester, NY 14642, USA.

[4]Department of Computer Science, University of California, Los Angeles, USA.

[5]Center for Artificial Intelligence and Modeling, The Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

[6]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

Abstract

Characterizing the metabolic profile of a microbial community is crucial for understanding its biological function and its impact on the host or environment. Metabolomics experiments directly measuring these profiles are difficult and expensive, while sequencing methods quantifying the species composition of microbial communities are well-developed and relatively cost-effective. Computational methods that are capable of predicting metabolomic profiles from microbial compositions can save considerable efforts needed for metabolomic profiling experimentally. Yet, despite existing efforts, we still lack a computational method with high prediction power, general applicability, and great interpretability. Here we develop a method — mNODE (Metabolomic profile predictor using Neural Ordinary Differential Equations), based on a state-of-the-art family of deep neural network models. We show compelling evidence that mNODE outperforms existing methods in predicting the metabolomic profiles of human microbiomes and several environmental microbiomes. Moreover, in the case of human gut microbiomes, mNODE can naturally incorporate dietary information to further enhance the prediction of metabolomic profiles. Besides, susceptibility analysis of mNODE enables us to reveal microbe-metabolite interactions, which can be validated using both synthetic and real data. The presented results

*Correspondence and requests for materials should be addressed to Y.-Y. L. (yyl@channing.harvard.edu).

Author contributions. Y.-Y.L. conceived the project. T.W. and Y.-Y.L. designed the project. T.W. performed all the numerical calculations and data analysis. T.W. processed the real data with assistance from X.-W.W. and K.L.S. All authors analyzed the results. T.W. and Y.Y.L. wrote the manuscript. All authors edited and approved the manuscript.

Competing Interests. The authors declare no competing interests.

Code availability. All code for running mNODE can be found at https://github.com/wt1005203/mNODE[58].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

demonstrate that mNODE is a powerful tool to investigate the microbiome-diet-metabolome relationship, facilitating future research on precision nutrition.

Introduction

Metabolic activities of microbial communities play an important role in shaping their biological functions and their interactions with hosts or environments. For example, a key way the human gut microbiome affects host physiology involves the production of small molecules[1–3]. Microbes in the gut can digest undigestible carbohydrates[4] and eventually produce important metabolites such as essential amino acids[5] and short-chain fatty acids[2], which are crucial for human well-being[6–8] To understand and eventually control microbial communities by modulation of nutrients or probiotic administration, the first step is to connect the metabolomic profile of a microbial community with its microbial composition.

Experimental measurement of metabolomic profiles usually requires metabolomics techniques. However, metabolomics experiments are difficult because of expensive equipment[9–11], lack of automation[12,13], and limited metabolite coverage[9,14]. In contrast, amplicon or shotgun metagenomic sequencing data are more readily available for complex microbial communities because they are less expensive, more easily automated, and provide good microbial coverage[15,16]. Hence, it is desirable to develop a computational method to predict metabolic profiles based on microbial compositions. Moreover, such a method could facilitate our understanding of the interplay between microorganisms and their metabolites.

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[17], MIMOSA[18], and Mangosteen[19] that rely on genome-scale metabolic models (GEMs). As a result, they rely heavily on the completeness and accuracy of queried databases and GEMs. (2) Ecology-guided methods, which simulate metabolite consumption and byproduct generation by microbes[20–23]. Those methods heavily rely on the quality of ecological networks. (3) Machine learning (ML)-based methods, which are trained from paired microbiome and metabolome datasets, and then used to predict the metabolic profile of never-seen microbiome samples, without using any reference database or domain knowledge regarding relationships between genes and metabolites. Various ML techniques such as elastic net[24], sparsified NED (Neural Encoder-Decoder)[25], multilayer perceptron[26], and word2vec[27] have been employed to predict metabolic profiles from microbial compositions. However, none of these ML-based methods utilize state-of-the-art deep neural network models such as Neural Ordinary Differential Equation (NODE)[28], so their performance has not been fully maximized.

Residual neural networks (ResNet) are the precursor of NODE. As the most important characteristic of ResNet, the skip connection is highly analogous to the Euler step in ODE solvers[29,30]. More specifically, one residual block can be regarded as a small time-change of variables in the language of ODEs[29,30]. This analogy fostered the invention of NODE[28]. Rather than specifying a discrete sequence of hidden layers like ResNet, NODE parameterizes the derivative of the hidden state using a neural network with output computed using an ODE solver. In a sense, NODE represents continuous-depth models, has constant

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

memory cost, adapts its evaluation strategy to each input, and can explicitly trade numerical precision for speed[28,31]. As a result, higher accuracy can be achieved with fewer parameters and without explicitly introducing many neural layers[28,31].

Here we leverage the power of NODE and propose a computational method called mNODE to predict metabolomic profiles from microbial compositions. We first generated synthetic data using the microbial consumer-resource model with cross-feeding interactions to validate that (1) mNODE outperforms existing ML-based methods, and (2) supplementing the information about nutrients as an additional input variable results in more accurate metabolome predictions, especially when the nutrient similarity across samples is low. We then compared mNODE with existing ML-based methods using microbiome and metabolome data collected in real microbial communities finding that mNODE outperforms all existing ML-based methods. Moreover, we demonstrated that we could further improve the performance of mNODE by including food profiles from food frequency questionnaires as an additional input. Finally, to reveal microbe-metabolite interactions, we performed susceptibility analysis using mNODE to study how the predicted concentration of one metabolite responds to a perturbation in the relative abundance of one microbial species. We demonstrated clearly that susceptibilities for all microbe-metabolite pairs can be used to accurately predict microbe-metabolite interactions of synthetic and real data.

Results

Overview of mNODE

We aim to predict the metabolomic profile of a microbial community based on its microbial composition and the potentially available information about diets. Fig. 1 shows a hypothetical example of two training scenarios when we follow the same model training protocol for mNODE: (1) without the dietary information as part of the input and (2) with dietary information as part of the input. The performance of two scenarios is compared to understand to what extent the existence of dietary information helps our predictions. This demonstrative example system comprises 3 microbial species, 2 dietary items, 4 metabolites, and 15 samples. 10 samples are used in the training set and the remaining 5 samples are in the test set. For a given set of samples with paired microbiome and metabolome data, we randomly split samples into two non-overlapping sets: the training set and the test set. We performed a 5-fold cross-validation on the training set to determine the optimal set of hyperparameters that maximizes the prediction power. Here, for each metabolite, we can calculate the Spearman's Correlation Coefficient (SCC), denoted as ρ, between its predicted and experimentally observed concentrations across different samples. The overall prediction power of any computational method such as mNODE is evaluated as the mean SCC, $\bar{\rho}$, of all metabolites measured in experiments. After the 5-fold cross-validation, we retrained the model with the optimal set of hyperparameters on the entire training set and then used this trained model to generate predictions on the test data (Fig. 1c).

Since the numbers of microbial species (N_s) and metabolites (N_m) are generally different from each other, it is not possible to directly apply the original NODE architecture[28], which requires the input and output dimensions to be equal. Here, we leverage the flexibility

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of the Multilayer Perceptron (MLP) in the data dimension and introduce the NODE as a module in the middle, stuck between two densely connected layers (Fig. 1b). Our NODE module maps the hidden variable $\mathbf{x}\left(t=0\right) \in \mathbb{R}^{N_h}$ at time $t = 0$ to its state $\mathbf{x}\left(t=T\right) \in \mathbb{R}^{N_h}$ at time $t = T$ following the ODE $\frac{d\mathbf{x}}{dt} = f(\mathbf{x})$ where $f(\mathbf{x})$ is approximated by the learnable MLP with one hidden layer with dimension N_h (Fig. 1b, middle). Note that here the variable t is a virtual time representing the continuous transformation of the hidden units in the NODE architecture. It should not be interpreted as a real time. We choose $T = 1$ without loss of generality since the ODE can be rescaled. In this study, we did not attempt to predict changes in metabolomic profiles over time; instead, the essential idea is to use NODE to reduce memory and training time since NODE shares connection weights across continuous layers and utilizes fewer parameters. The L2 regularization is adopted to prevent overfitting. In total, there are two hyperparameters in mNODE to calibrate: the dimension of the hidden layer N_h and the weight parameter for L2 regularization λ.

Validate mNODE using synthetic data

To validate mNODE, we generated synthetic data using the Microbial Consumer-Resource Model (MiCRM) which accounts for nutrient competition and cross-feeding interactions[32]. We created each independent synthetic community ("sample") via sampling a set of species from a metapopulation pool (each species is introduced with the species sampling probability p_s) and a set of nutrients from a nutrient pool (each nutrient is introduced with its predetermined supply rate with the nutrient sampling probability p_n) to initiate the community assembly and then collecting the steady-state microbial species relative abundances as microbial compositions and steady-state nutrient concentrations as metabolomic profiles. Details can be found in see Supplementary Information section 3. Specifically, we assumed a metapopulation pool of 10 species ($N_s = 10$) and a nutrient pool of 10 nutrients ($N_n = 10$) with predetermined nutrient supply rates for all nutrients.

To explore whether mNODE generates better metabolome prediction than previously existing ML-based methods, we generated 300 samples with $p_s = 0.5$ and $p_n = 0.6$. We left out 20% of all independent samples (60 samples) as the test dataset, whereas the remaining 240 samples are used as the training dataset. Figs. 2a1–a3 show the performance comparison between mNODE and previous ML-based methods through 3 metrics: (1) mean SCC, $\overline{\rho}$ (Figs. 2a1, b1,c1), (2) the mean of the top-5 SCCs, $\overline{\rho}_5$ (Figs. 2a2, b2, c2), and (3) the number of metabolites (same as nutrients in the MiCRM) with SCCs larger than 0.5, $N_{\rho>0.5}$ (Figs. 2a3, b3, c3). We found that mNODE generates better predictions than all other methods across three performance metrics. In addition, when we incorporated nutrient supply rates as well as microbial composition as the input of mNODE (denoted as “mNODE + nutrients” in Figs. 2a1–a3), it yields the best prediction.

We also examined the influence of training sample sizes by sub-sampling the original 240 training samples and keeping the same test set. When nutrient supply rates are not used in the input, the performance reaches saturation when the training sample size reaches $8(N_n + N_s)$ (Figs. 2b1–b3). In contrast, when nutrient supply rates are added to the input, the performance saturated as the training sample size is around $4(N_n + N_s)$. Furthermore,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

we tested mNODE on synthetic data generated with larger numbers of species($N_s = 100$ or 200) and metabolites ($N_m = 100, 200,$ or 300) and found that mNODE achieves a great performance when the number of training samples is around $3N_s$, independent of the number of metabolites N_m (Extended Data Figure 5). A similar scaling happens to MelonnPan and MiMeNet but not for poor predictors such as sparse NED and ResNet.

It is natural to speculate that the performance difference may also depend on the nutrient similarity across samples because the less variable nutrient information might render it less useful in discriminating metabolomic profiles. To test this, we systematically tune the nutrient sampling probability p_n because higher p_n implies more similar nutrient supplies. Figs. 2c1–c3 show how three performance metrics change as p_n increases. We found that when nutrient supply rates are not included in the input of mNODE, all 3 metrics increase as p_n increases. When nutrient supply rates are included in the input, the performance of mNODE is consistently high over a wide range of values for p_n. When $p_n = 100\%$, there isn't any variation in nutrient supply rates to independent communities across samples and thus nutrient supply rates become redundant (closer metric values in Figs. 2c1–c3).

Superior performance of mNODE on real data

Next, we sought to test it using real datasets with paired microbiome and metabolome. We started with the dataset PRISM + NLIBD that collected fecal samples from CD (Crohn's Disease) patients, UC (Ulcerative Colitis) patients, and healthy individuals[33]. This benchmark dataset is unique due to the presence of two IBD cohorts: a 155-member cohort collected at the Massachusetts General Hospital (PRISM) and a 65-member validation cohort collected in the Netherlands (NLIBD/LLDeep). Two cohorts were collected following the same protocols and processed in the same way (see Methods and Franzosa et al[33] for more details). For this dataset, we first selected hyperparameters for mNODE based on results of the 5-fold cross-validation of the PRISM cohort. Then, we trained mNODE with the selected hyperparameters on the entire PRISM cohort and computed predictions for the NLIBD cohort's metabolome. Similar to the previous three metrics, here we assessed the model performance based on $\bar{\rho}$, $\bar{\rho}_{50}$, and $N_{\rho>0.5}$. Though all metabolites are used in the training of ML methods, we only included annotated metabolites in the metric calculation to be in line with previous studies to ensure a fair comparison[24,26]. Figs. 3a1–a3 show the performance of 5 methods on the test set NLIBD after their training process on PRISM. We found that mNODE had the highest $\bar{\rho}$ (0.287 compared to 0.191 from MiMeNet), the highest $\bar{\rho}_{50}$ (0.622 compared to 0.512 from MiMeNet), and the highest $N_{\rho>0.5}$ (69 compared to 28 from MiMeNet) among all ML-based methods. Furthermore, we directly compared ρ of annotated metabolites via a scatterplot in Extended Data Figure 1. 75.8% of metabolites are better predicted by mNODE than MiMeNet (Extended Data Figure 1c)

To examine if the superior performance of mNODE is independent of environments where microbial communities reside, we collected another four datasets: (1) lung sputum samples of 172 patients with cystic fibrosis[21], (2) 19 desert soil biocrust samples after the five continuous wetting events[34],(3) fecal samples of 340 children at the age of 3 years enrolled in the VDAART (Vitamin D Antenatal Asthma Reduction Trial) study[35–38], and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(4) blood plasma samples of the same 340 children in VDAART[35–38]. The systematic comparison of different methods across the four datasets is shown in Fig. 3b–e. Across four datasets, mNODE delivered the best performance. Finally, to check whether mNODE with different initial parameters would generate consistent performance after the training, we ran mNODE 50 times for each real microbial community dataset and found that the predictive performance is very robust (Extended Data Figure 8).

Food profiles further improve the metabolome prediction

One unique feature of VDAART is the documentation of food consumption frequency. We obtained food profiles based on the food frequency questionnaire (FFQ) that captures food and beverage consumption over time[39–42]. Later, we converted their food consumption frequencies into nutritional profiles based on the nutrient composition encoded in the FNDDS (USDA's Food and Nutrient Database for Dietary Studies)[43] database. Details about FFQs and FNDDS are specified in the Supplementary Information section 6.

Equipped with the dietary information, we explored the relative importance of the gut microbial composition, food consumption, and nutrient intake in determining the metabolomic profiles by training the mNODE in different data combinations and later comparing their differences in the prediction power. Using food profiles or nutritional profiles alone as the input produced almost equally poor performance (Fig. 4). The combination of microbial composition and food profiles seems to be the most successful approach, as indicated by consistently the best performance across three metrics and two environments. This indicates that the variation of food consumption across individuals may help us to predict the metabolomic profiles better. One reason why the addition of food profiles may be more helpful than adding nutritional profiles is that the Bray-Curtis dissimilarity of nutritional profiles across paired samples is much lower than that of food profiles (Extended Data Figure 2).

Inferring microbe-metabolite interactions

To interpret mNODE, we proposed a method that perturbs the relative abundance of species i (x_i) by a small amount Δx_i for well-trained mNODE, re-predicts the concentration of metabolite α, and measures the deviation from the original prediction (we denoted the deviation as Δy_α). We defined the susceptibility of metabolite α to species i as $s_{\alpha i} = \frac{\Delta y_\alpha}{\Delta x_i}$ (Fig. 5a). Our definition of susceptibility follows the definition of differential susceptibility in physics and is very similar to the concept of "sensitivity" in engineering sciences (which is a property of a system that indicates how the system reacts to stimuli). $s_{\alpha i} < 0$ means that a higher abundance for species i leads to a lower predicted concentration for metabolite α and might imply that species i can consume metabolite α. Similarly, $s_{\alpha i} > 0$ probably corresponds to the production of metabolite α by species i. More technical details about how to compute $s_{\alpha i}$ are in the Methods section.

We first validated the idea on the synthetic data used in Figs. 2a1–a3 since we can directly compare the susceptibility matrix (Fig. 5b) with the ground-truth consumption interactions (Fig. 5c) and production interactions (Fig. 5d) in MiCRM. It is visually evident

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

that many ground-truth interactions in Fig. 5c are correctly revealed in the susceptibility matrix (indicated by having red colors in Fig. 5b). For a given susceptibility threshold s_{thres}, we classified microbe-metabolite pairs into either (1) consumption interactions if $s < s_{\text{thres}}$ or (2) production interactions if $s > s_{\text{thres}}$. Then we changed the threshold s_{thres}, performed the classification task for each threshold, computed TP (True-Positive) rates and FP (False-Positive) rates for all thresholds, and measured the AUC (Area Under Curve) for the ROC (Receiver Operating Characteristic) curve. We found that, for synthetic data, the susceptibility can predict consumption interactions with great performance (AUC=0.95) and production interactions with decent performance (AUC=0.69). When we imposed the constraint that production interactions don't overlap with consumption interactions in MiCRM, the prediction power for production interactions is almost equivalent to that for consumption interactions (Extended Data Figure 3e).

We eventually applied the validated susceptibility method to all real-life datasets and susceptibility values for all datasets are provided as Supplementary data/tables. Here we focused on the dataset PRISM + NLIBD because gut microbiomes of human adults have been investigated more extensively. We first tested the robustness of susceptibility values by computing them from five mNODE models trained with different initial parameters for the PRISM+NLIBD dataset and found that susceptibility values from five training repeats are highly correlated (Extended Data Figure 9). Here we focused on the bile acid metabolism which is relatively well-studied[44–48] and is shown to be associated with human gastrointestinal diseases such as gastrointestinal cancers[47] and IBD[49,50]. *Bacteroides vulgatus* ATCC 8482 has large positive susceptibilities for cholate and chenodeoxycholate. This is supported by the genomic evidence that *Bacteroides vulgatus* ATCC 8482 contains the *bsh* (bile salt hydrolase) gene, which encodes the BSH enzyme responsible for the deconjugation of conjugated primary bile acids to primary bile acids cholate and chenodeoxycholate[44,48]. Similarly, *Lachnospiraceae bacterium* 5 1 57FAA has large positive susceptibilities for deoxycholate and lithocholate, as evidenced by the presence of the *bai* (bile-acid-induced) operon in its genome, which transforms primary bile acids cholate and chenodeoxycholate to secondary bile acids deoxycholate and lithocholate respectively[48]. By constrast, all susceptibilities of *Streptococcus salivarius* are much closer to zero compared with other entries in the susceptibility matrix and its genome indeed contains no genes related to bile acid metabolism[48].

Discussion

Computational methods, capable of predicting metabolite levels using microbial composition, offer a much more cost-effective way to quantify the distribution of metabolite concentrations across samples. Moreover, such a model might help us better understand the interaction of microbes and metabolites. In this study, we developed the mNODE method, which produces better metabolome predictions than almost all previously developed methods. Furthermore, we further improved the performance of mNODE by including food profiles together with the microbial composition as the input of mNODE. Finally, we showed that mNODE infers microbe-metabolite interactions accurately on synthetic data and a human dataset (PRISM+NLIBD) via susceptibility analysis. We would like to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

highlight whether inferred interactions do exist has to be tested experimentally. After all, the prediction performance is far from perfect ($\bar{\rho}$ smaller than 0.5 for most cases). In the future, if more omics data such as the metatranscriptome and metaproteome are available, they can also be easily incorporated as inputs of mNODE to investigate the relative importance of each data type in predicting the metabolome.

Compared to reference-based methods[17–19] such as MAMBO[17],MIMOSA[18], as well as Mangosteen[19], 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 $N_{\rho>0.3}$) without using metabolic models, much higher than 20 for MIMOSA[24]. The performance of reference-based methods is limited due to the lack of complete GEMs for many microbial species.

The current method of documenting diet information via the FFQ has some limitations[51]. On one hand, the systematic error in the consumption amount occurs because the FFQ only surveys the frequency of eating one food item, rather than the amount of each food item eaten per meal. Digital documentation of the amount of food eaten for every meal may alleviate this error[51]. On the other hand, the FFQ lacks detailed information about food product brands and food preparation. A similar issue arises when food profiles in FFQs are converted into nutritional profiles due to limited nutrient items and the lack of detailed nutrient composition. A detailed metabolomic analysis of a comprehensive set of foods might be important for the advancement of precision nutrition[52].

We notice that the prediction accuracy for microbial communities collected from lungs[21] and soils[34] is higher than that for human gut microbiomes (i.e. PRISM + NLIBD[33] and VDAART[35–38]), regardless of computational approaches utilized. This implies that metabolomic profiles of some microbiomes are harder to be predicted than others and we suspect that this is due to the difference in the ratio between the input dimension (N_s) and the output dimension (N_m). For instance, for lung samples[21], the number of microbial taxa (1119) is much larger than the number of metabolites (168). As a comparison, for fecal samples in PRISM + NLIBD, the number of microbial taxa (200) is much smaller than the number of metabolites (8848). This comparison just showed the complexity of inferring metabolomic profiles for human gut microbiomes.

Methods

Datasets.

PRISM + NLIBD fecal samples.—This dataset was reported by a study related to the fecal microbiome and metabolome samples of IBD patients[33]. There are two cohorts involved in the study. The first cross-sectional cohort of individuals was collected by a study PRISM (the Prospective Registry in IBD Study at MGH). This cohort includes 68 patients with Crohn’s disease (CD), 53 patients with ulcerative colitis (UC), and 34 healthy controls. The second study cohort (NLIBD/LLDeep) was independently collected in the Netherlands and consists of 20 CD patients, 23 UC patients, and 22 healthy controls. Fecal samples from both cohorts were collected following the same protocol and later both the microbiome

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and metabolome are profiled and analyzed in the same way. In total, 200 microbial taxa were generated using the metagenomic shotgun sequencing and 8848 unique metabolites were obtained by four LC-MS (Liquid Chromatography-Mass Spectrometry) metabolomics methods. Out of 8848 unique metabolites, 466 are annotated. In all machine learning tasks, all data (microbial composition and metabolomic profiles) related to individuals in PRISM are used as the training set and the external validation cohort NLIBD/LLDeep is used as the test set.

Cystic fibrosis lung samples.—lung sputum samples from 172 patients were collected in a study that investigates how the chemical gradient drives the shift in the microbiome structure and the pattern of metabolite production[21]. 1119 microbial features were determined by the 16S rRNA gene sequencing and profiles of 168 metabolites were generated by the tandem liquid chromatography-mass spectrometry (LC-MS/MS) and gas chromatography-mass spectrometry (GC-MS) metabolomics. For all machine learning tasks, an 80/20 train-test split with the random state set as 42 is utilized to guarantee a fair comparison.

Desert soil biocrust samples.—Swenson et al describes how biocrusts from four successional stages were wet up and sampled at five time points for each stage[34].466 dominant taxa were determined by the shotgun metagenomic sequencing that measures their single-copy gene markers. The liquid chromatography-mass spectrometry (LC/MS) soil metabolomics identified 85 metabolites that changed at least two-fold across both wetting and successional stages. An 80/20 train-test split with the random state set as 42 is utilized for all models.

VDAART children's fecal samples, plasma samples, and FFQ data.—VDAART (Vitamin D Antenatal Asthma Reduction Trial) is a trial of prenatal vitamin D supplementation to prevent asthma and wheezing offspring[35,36]. Fecal samples and blood plasma samples of 340 children at the age of 3 years were collected. The 16S rRNA gene sequencing identified the microbiome profiles of 209 microbial taxa in fecal samples[38]. Primer and adapter trimming was performed using Skewer. Chimera checking and filtering were performed using Qiime2[53]. Reads were denoised using DADA2 as implemented in Qiime2[54]. The metabolomic profiles of both fecal and blood samples are generated via the using ultra-high performance liquid chromatography-mass spectrometry (UPLC-LC/MS) as previously described[37,38,55]. Identification of known chemical entities was based on a comparison to metabolomic library entries of purified standards based on chromatographic properties and mass spectra. 1298 metabolites were determined in the fecal samples and 1064 metabolites were determined in blood samples. Besides, we extracted their diet information from the section in the 36 Months Quarterly Infant Follow-up Questionnaire that documented the food frequency consumed by children[37]. Child diet was evaluated at age of 3 years when parents completed a modified version of a semi-quantitative 87-item food frequency questionnaire (FFQ) that was previously validated in preschool-age children[56]. The food frequency table can be further converted to the nutritional profiles using FNDDS (USDA's Food and Nutrient Database for Dietary Studies) which is a database that is used to convert food and beverages consumed into gram amounts and to determine their

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

nutrient values. All 340 children are divided into either the train set or the test set with an 80/20 ratio with the random state set as 42.

mNODE.

The NODE (Neural Ordinary Differential Equations) is a deep learning method that combines explicit layers with implicit layers where the states of hidden layers are described by Ordinary Differential Equations. Our mNODE introduces the NODE as a module in the middle.

- Data processing: The CLR (Centered Log-Ratio) transformation is applied to microbial abundances and metabolite concentrations.
- Model detail: 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) Neural ODE module[28] where the first-order time derivative is approximated by a one-hidden-layer MLP with the hidden layer dimension the same as N_h, 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: Adam optimizer[57] is used for the gradient descent. The training stops if the mean SCC (Spearman Correlation Coefficient) of annotated metabolites $\bar{\rho}$ on the validation/test set starts to decrease within the past 20 epochs. The criterion for the decrease is judged by whether the number of decreases of the mean SCC $\bar{\rho}$ in the past 20 epochs (i.e., $\bar{\rho}$ at epoch i minus $\bar{\rho}$ at epoch i-1 is smaller than 0) is larger than 12.
- 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 the L2 penalty with weight parameter λ. N_h is chosen from [32, 64, 128], and λ is selected from $[10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}]$.

Inferring microbe-metabolite interactions via susceptibility.

The well-trained mNODE takes relative abundances of all species (relative abundance of species i is denoted as x_i) as inputs and generate predictions for metabolite concentrations (concentration of metabolite α is written as y_α). For the sample m in the training set, $x_i^{(m)}$ for all i is provided as the input vector to the trained mNODE and mNODE can predict concentrations $y_\alpha^{(m)}$ for all metabolites. To investigate the influence of species i on metabolite α for a particular sample m, we perturb $x_i^{(m)}$ by setting it as the mean of relative abundances for species i across samples $\bar{x}_i$ while keeping values of $x_{j\neq i}^{(m)}$ intact. Thus the perturbation amount for species i is $\Delta x_i^{(m)} = \bar{x}_i - x_i^{(m)}$. This perturbed vector for microbial relative abundances is provided to the trained mNODE to regenerate predictions for metabolite concentration $y_\alpha^{(m)} + \Delta y_\alpha^{(m)}$ where $\Delta y_\alpha^{(m)}$ is the deviation of newly predicted

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

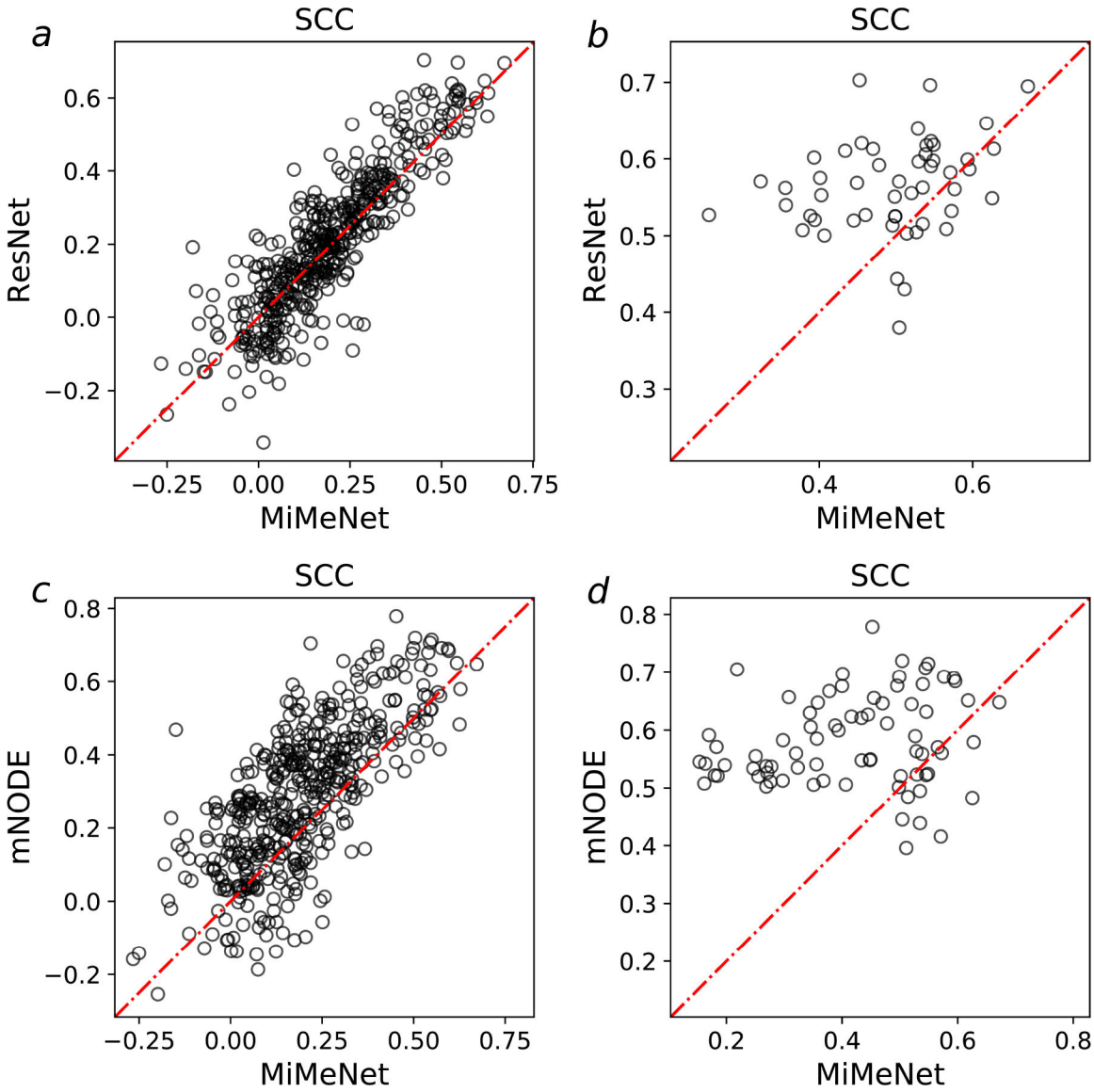
concentration for metabolite α when mNODE uses the perturbed input vector from that when mNODE uses the unperturbed input vector. For the sample m, the susceptibility of metabolite α to species i can be defined as $s_{\alpha i}^{(m)} = \frac{\Delta y_{\alpha}^{(m)}}{\Delta x_{i}^{(m)}}$. To properly take into account all training samples, we average susceptibility over samples and obtain the overall susceptibility of metabolite α to species i, $s_{\alpha i} = \frac{\sum_m s_{\alpha i}^{(m)}}{N_{\text{train}}}$, where N_{train} is the number of samples in the training set.

Statistics.

To calculate correlation coefficients throughout the study, we used Spearman's rank correlation coefficient. Wherever we used P values, we explained in the Methods how we calculated them, since for all such measurements in the study, we calculated the associated null distributions from scratch. All statistical tests were performed using standard numerical and scientific computing libraries in the Python programming language (version 3.7.3), the Julia programming language (version 1.6.2) and Jupyter Notebook (version 6.1.0).

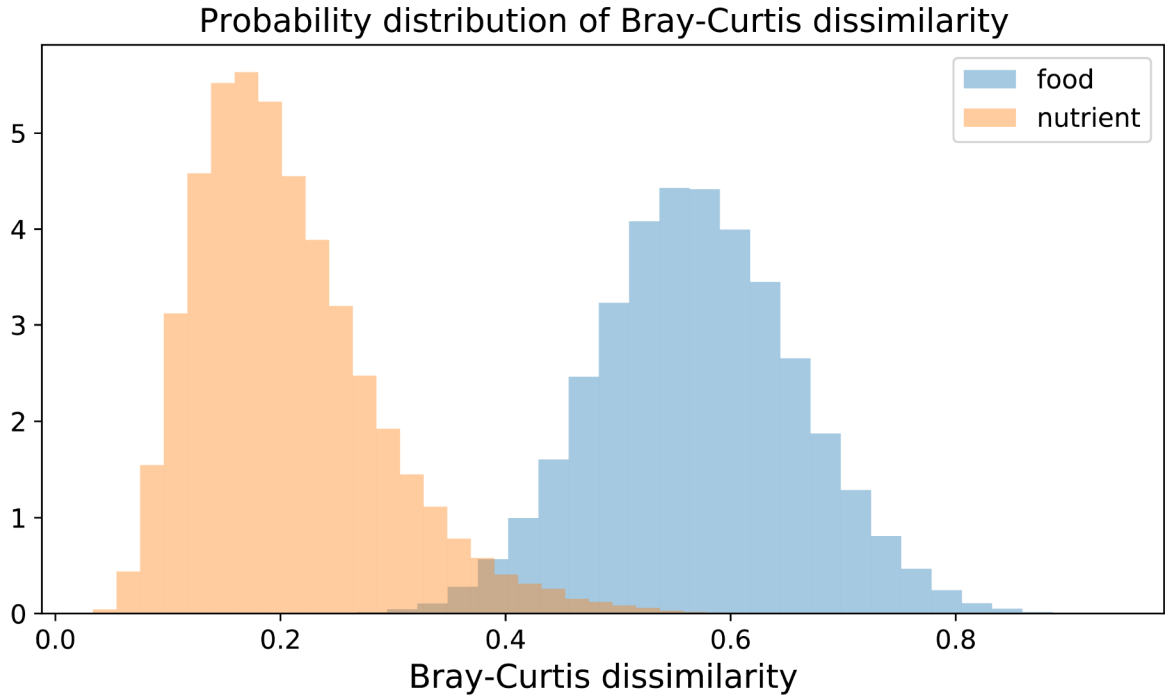
Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Extended Data



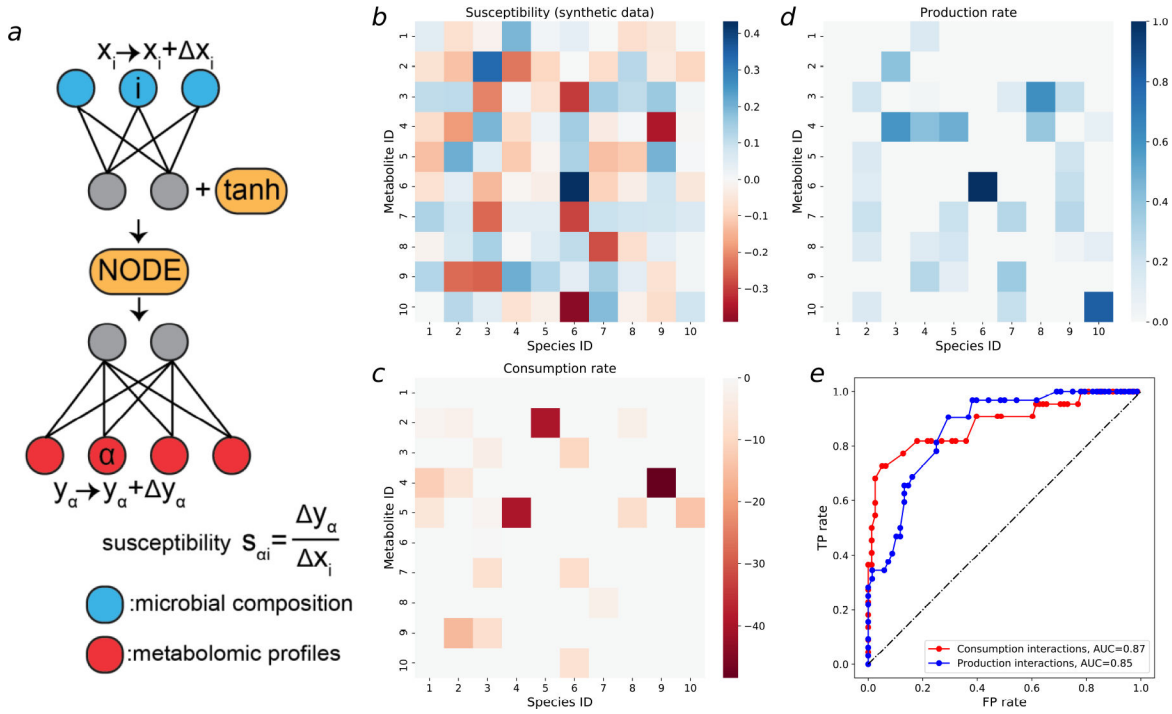
Extended Data Figure 1: Comparison of SCCs of annotated metabolites on the test set NLIBD.
For each annotated metabolite, its SCC between its predicted values and true values across samples is computed for all computational methods. **a** Comparison of SCCs of all annotated metabolites between MiMeNet and ResNet. **b** Comparison of SCCs of all well-predicted annotated metabolites between MiMeNet and ResNet. Well-predicted metabolites are metabolites that have SCCs larger than 0.5 according to either MiMeNet or ResNet. **c** Comparison of SCCs of all annotated metabolites between MiMeNet and mNODE. **d** Comparison of SCCs of all well-predicted annotated metabolites between MiMeNet and mNODE. Well-predicted metabolites are metabolites that have SCCs larger than 0.5 according to either MiMeNet or mNODE.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 2: Bray-Curtis dissimilarity of the food consumption profiles in FFQs and nutritional profiles across samples in VDAART.

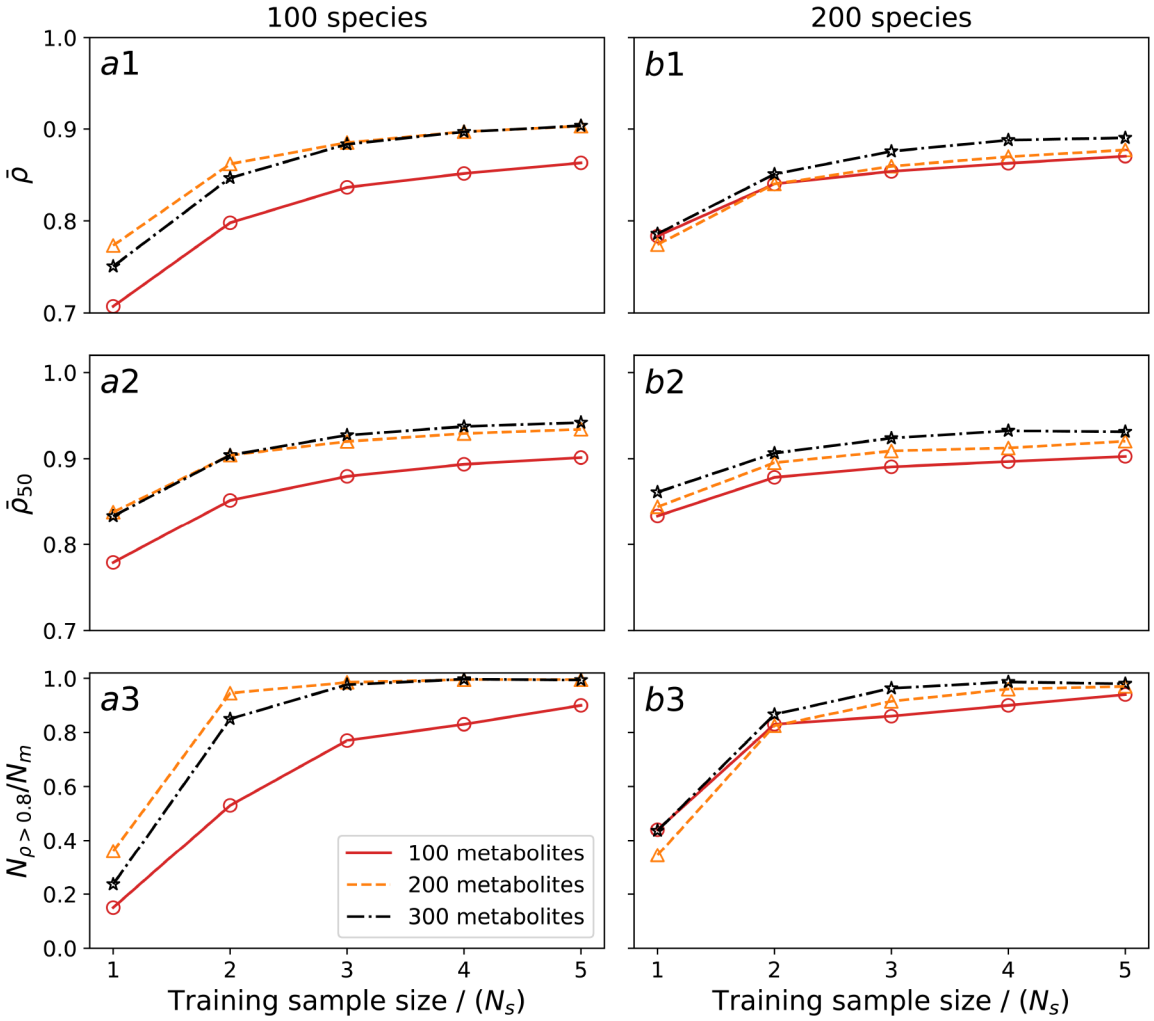
The distribution of Bray-Curtis dissimilarity for all paired food consumption profiles in FFQs and nutritional profiles.



Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Extended Data Figure 3: Using susceptibility of metabolite concentrations to microbial relative abundances of well-trained mNODE to infer microbe-metabolite interactions on new synthetic data.

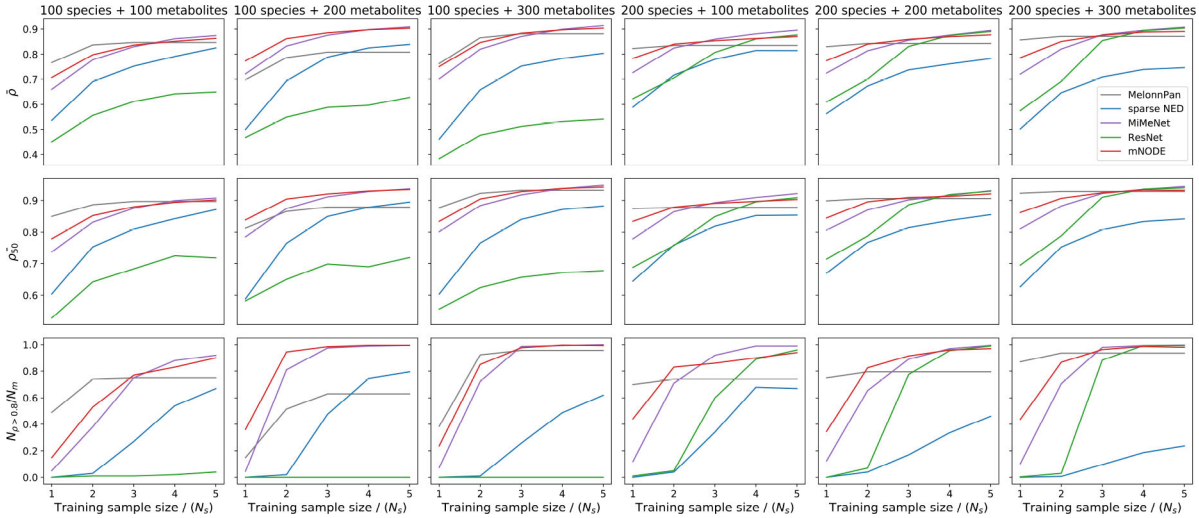
New synthetic data in this figure are generated by the microbial consumer-resource model with species-specific byproduct generations and without the overlap between consumption and production interactions between microbes and metabolites. See the Supplementary Information section 4 for more details of this model. **a** The susceptibility of the concentration of metabolite $\alpha(y_\alpha)$ to the relative abundance of species $i(x_i)$, denoted as $s_{\alpha i}$, is defined as the ratio between the deviation in the concentration of metabolite $\alpha(\Delta y_\alpha)$ and the perturbation amount in the relative abundance of species $i(\Delta x_i)$. **b** Susceptibility values for all microbe-metabolite pairs in the synthetic data. **c** The ground-truth consumption matrix and corresponding rates in synthetic data. All consumption rates are shown as negative values for the convenience of comparison with panel b. **d** The ground-truth production matrix and corresponding rates in synthetic data. **e** The ROC (Receiver Operating Characteristic) curve based on TP (True Positive) rates and FP (False Positive) rates which are obtained by setting different susceptibility thresholds for classifications of interactions.



Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Extended Data Figure 4: The training sample size needed to reach a great predictive performance scales linearly with the number of species in synthetic data.

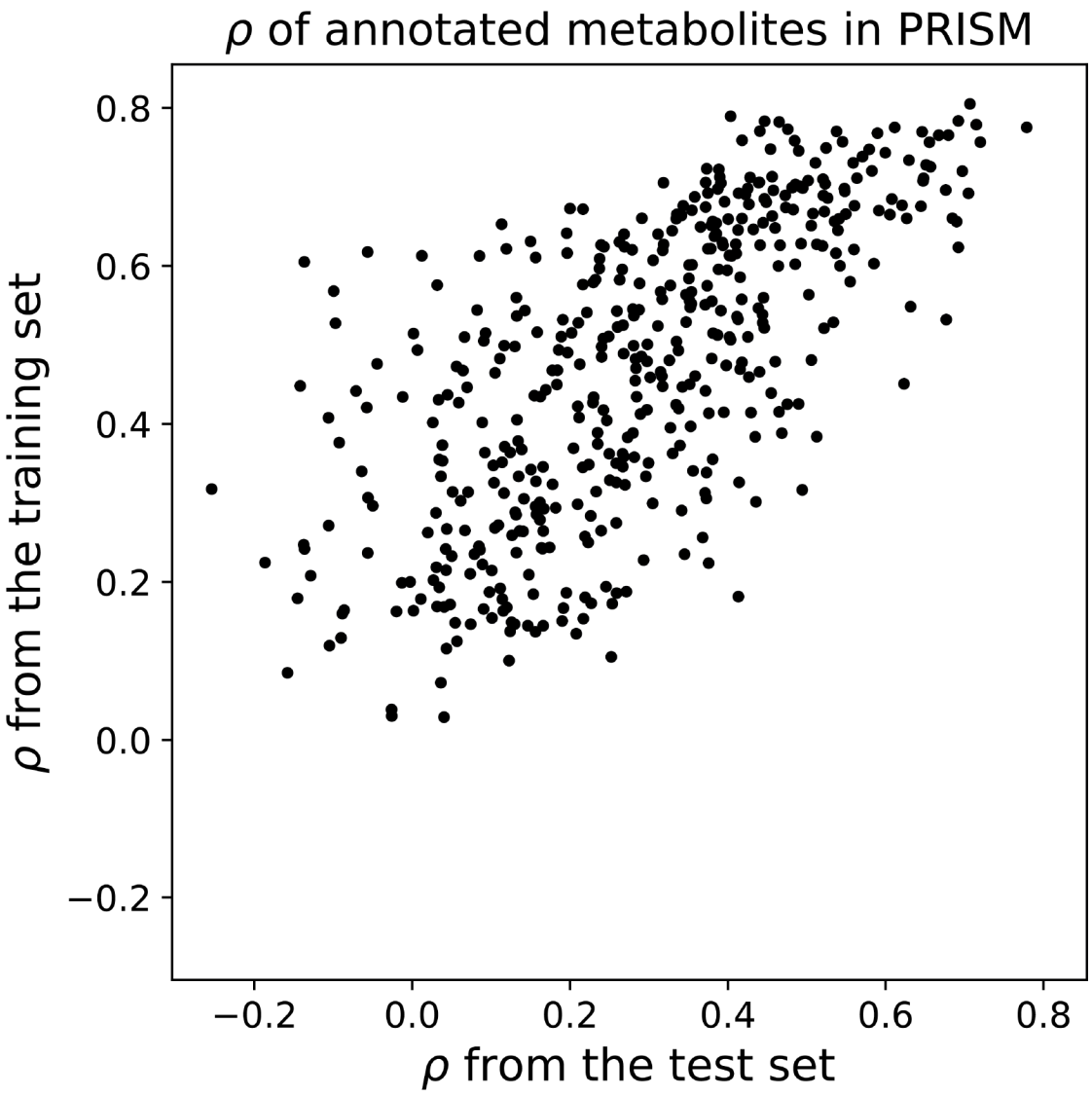
Synthetic data in this figure are generated by the microbial consumer-resource model with nutrient sampling probability $p_n = 1.0$. For the case with 100 species and varying number of metabolites (100, 200, or 300), three metrics are used for comparing model performances: a1 the mean SCC $\bar{\rho}$, **a2** the top-50 mean SCC $\bar{\rho}_{50}$, and **a3** the number of metabolites with SCCs larger than 0.8 divided by the number of metabolites $N_{\rho>0.8}/N_m$. **b1-b3** The performance metrics for the case with 200 species and varying numbers of metabolites (100, 200, or 300).



Extended Data Figure 5: The training sample size needed to reach a great predictive performance scales linearly with the number of species in synthetic data for MelonnPan, MiMeNet, and mNODE.

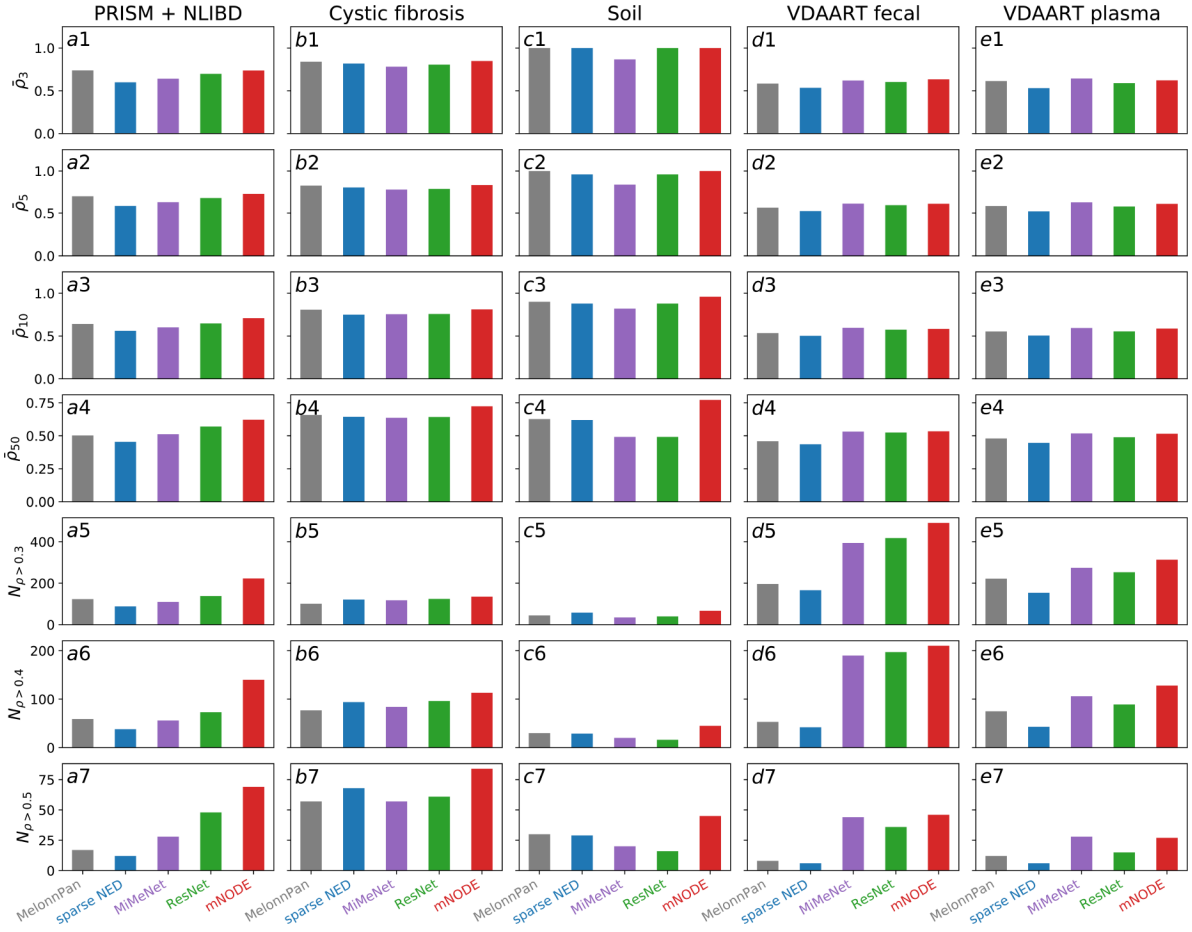
Synthetic data in this figure are generated by the microbial consumer-resource model with nutrient sampling probability $p_n = 1.0$. For the case with 100 species and varying number of metabolites (100, 200, or 300), three metrics are used for comparing model performances: the mean SCC $\bar{\rho}$, the top-50 mean SCC $\bar{\rho}_{50}$, and the number of metabolites with SCCs larger than 0.8 divided by the number of metabolites $N_{\rho>0.8}/N_m$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 6: The Spearman correlation coefficients of metabolites ρ between its predicted and true values from the test set well correlate with those from the training set for the PRISM+NLIBD dataset.

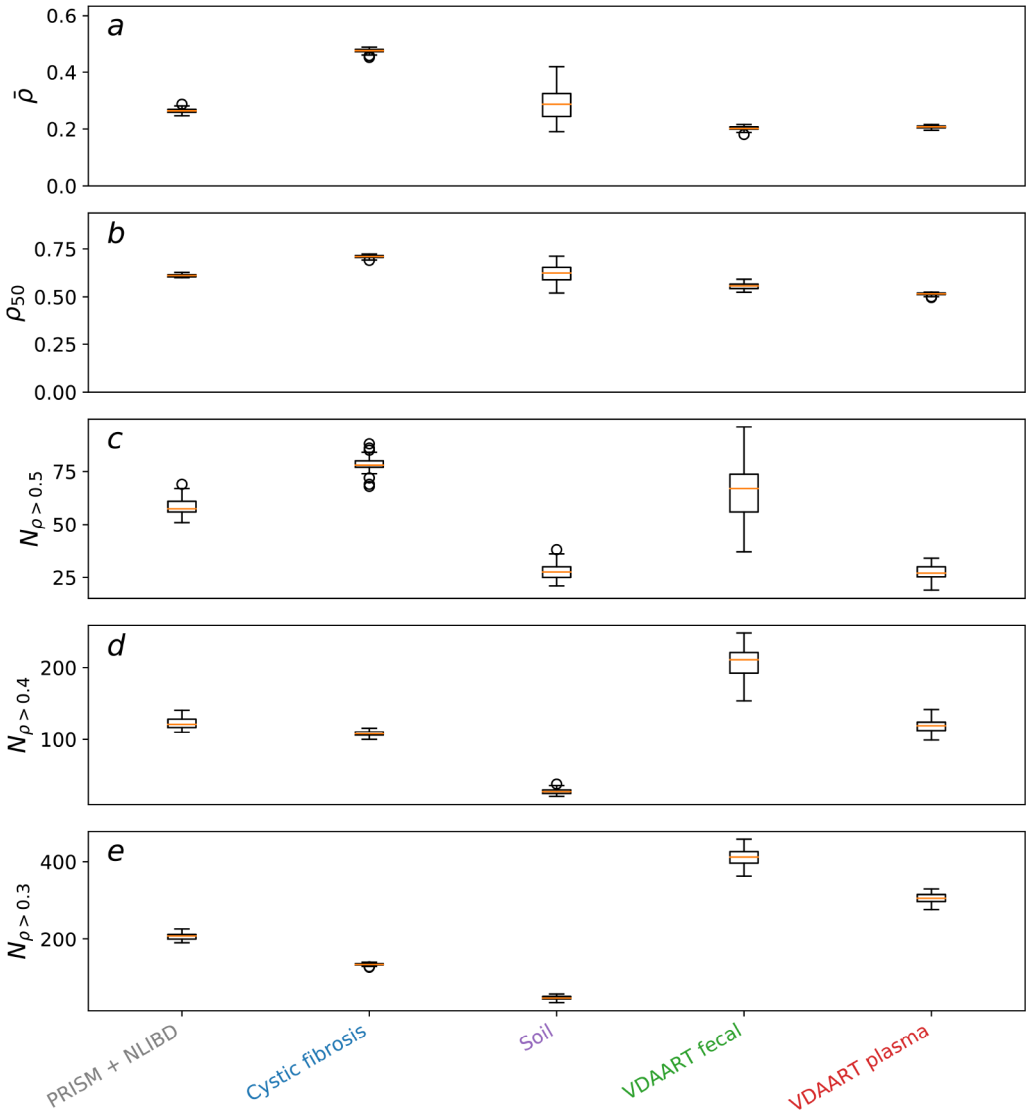
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 7: More performance metrics to compare mNODE with existing methods on five real microbial community datasets.

For the dataset PRISM+NLIBD[33], seven performance metrics are used: **a1** the top-3 mean SCC $\bar{\rho}_3$, **a2** the top-5 mean SCC $\bar{\rho}_5$, **a3** the top-10 mean SCC $\bar{\rho}_{10}$, **a4** the top-50 mean SCC $\bar{\rho}_{50}$, **a5** the number of metabolites with SCCs larger than 0.3 $N_{\rho>0.3}$, **a6** the number of metabolites with SCCs larger than 0.4 $N_{\rho>0.4}$, and a7 the number of metabolites with SCCs larger than 0.5 $N_{\rho>0.5}$. **b1-b7** Performance of methods measured by seven metrics on the data from lung samples of patients with cystic fibrosis[21]. **c1-c7** Performance of methods on the data from soil biocrust samples after 5 wetting events34. d1-d7 Performance of methods on the data from fecal samples of children at age 3[35–38]. **e1-e7** Performance of methods on the data from blood plasma samples of children at age 3[35–38].

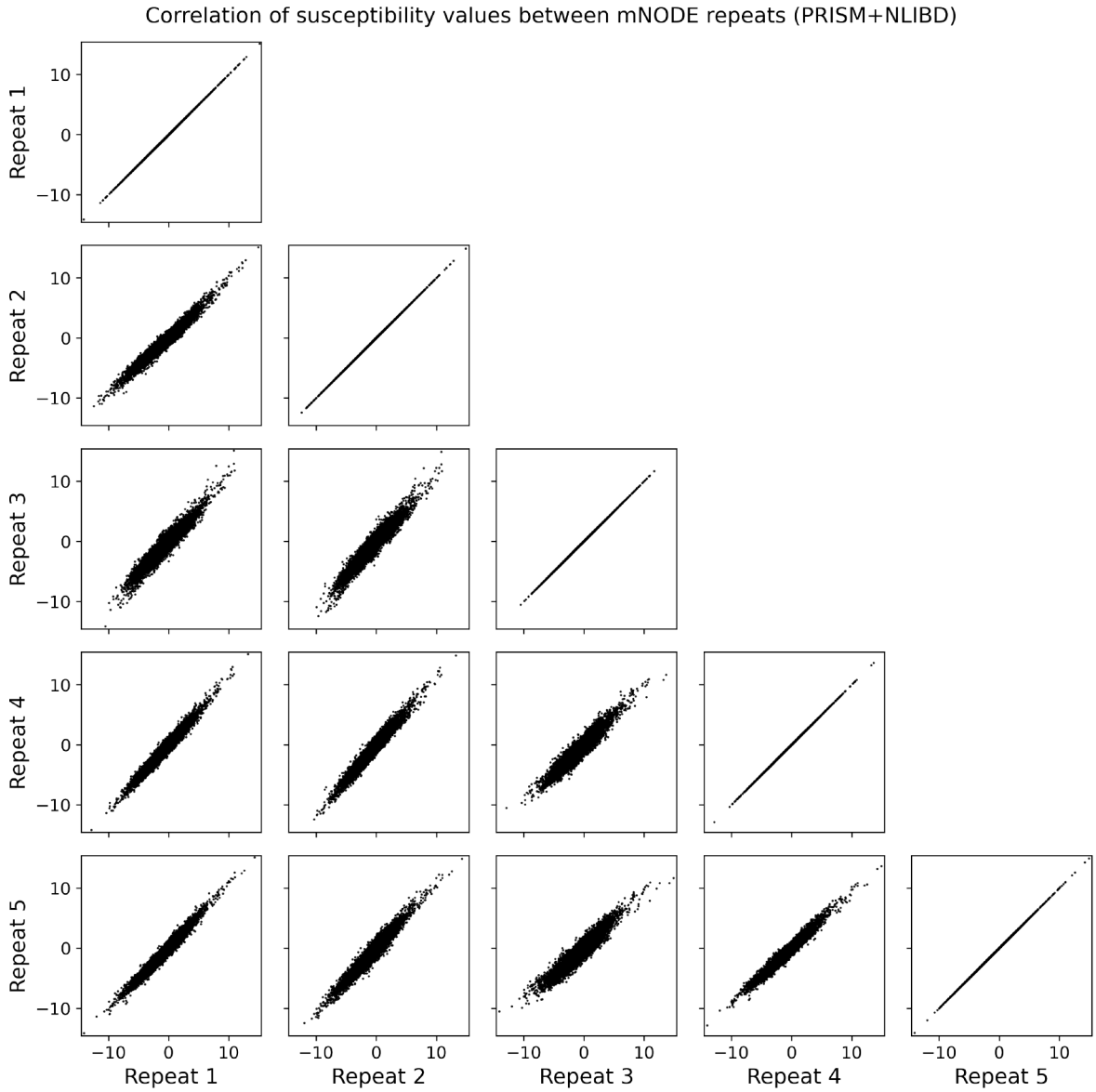
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 8: 50 mNODE training repeats with different initializations have very consistent predictive performances on real microbial community datasets.

For the dataset PRISM+NLIBD[33], three performance metrics are adopted for comparing model performances: the mean SCC $\bar{\rho}$, the top-50 mean SCC $\bar{\rho}_{50}$, and the number of metabolites with SCCs larger than 0.5, 0.4, and 0.3 (denotated as $N_{\rho>0.5}$, $N_{\rho>0.4}$, and $N_{\rho>0.3}$ respectively). All datasets[21, 33–38] are randomly divided into training and test sets with the 80/20 ratio except for the PRISM and NLIBD dataset[33]. In all boxplots, the middle orange line is the median, the box extends from the first quartile (Q1) to the third quartile (Q3) of the data, the black whiskers extend from the box by 1.5 × IQR (where IQR is the interquartile range), and outlier unfilled black circles are those beyond the range defined by two whiskers. The sample size $n = 50$ for all boxplots.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 9: Computed susceptibility values from five mNODE training repeats with different initializations are highly correlated with each other on the PRISM+NLIBD dataset[33].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements.

Y.-Y.L. acknowledges grants from the National Institutes of Health (R01AI141529, R01HD093761, RF1AG067744, UH3OD023268, U19AI095219, and U01HL089856). We thank Nancy Laranjo for VDAART data support.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Data availability.

The datasets of PRISM + NLIBD[33], lung samples[21], and soil biocrust samples[34] can be found at https://github.com/YDaiLab/MiMeNet/tree/master/data. The microbiome sequencing data, food frequency data, and metabolomics data from VDAART[35–38] are part of the ECHO (Environmental influences on Child Health Outcomes) consortium and ECHO consortium members and other interested scientists can obtain the data directly from the ECHO DCC (Data Coordinating Center). The FNDDS (USDA's Food and Nutrient Database for Dietary Studies)[43] can be found at https://data.nal.usda.gov/dataset/food-and-nutrient-database-dietary-studies-fndds.

References

1. Donia MS & Fischbach MA Small molecules from the human microbiota. Science 349 (2015).
2. Koh A, De Vadder F, Kovatcheva-Datchary P & Bäckhed F From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. Cell 165, 1332–1345 (2016). [PubMed: 27259147]
3. Koppel N, Rekdal VM & Balskus EP Chemical transformation of xenobiotics by the human gut microbiota. Science 356 (2017).
4. Myhrstad MC, Tunsjø H, Charnock C & Telle-Hansen VH Dietary fiber, gut microbiota, and metabolic regulation—current status in human randomized trials. Nutrients 12,859 (2020). [PubMed: 32210176]
5. Lin R, Liu W, Piao M & Zhu H A review of the relationship between the gut microbiota and amino acid metabolism. Amino Acids 49, 2083–2090 (2017). [PubMed: 28932911]
6. Tyson GW et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428, 37–43 (2004). [PubMed: 14961025]
7. Flint HJ, Scott KP, Louis P & Duncan SH The role of the gut microbiota in nutrition and health. Nature reviews Gastroenterology & hepatology 9, 577–589 (2012). [PubMed: 22945443]
8. Lloyd-Price J et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569, 655–662 (2019) [PubMed: 31142855]
9. Yang Q et al. Metabolomics biotechnology, applications, and future trends: a systematic review. RSC Advances 9, 37245–37257 (2019). [PubMed: 35542267]
10. Castelli FA et al. Metabolomics for personalized medicine: the input of analytical chemistry from biomarker discovery to point-of-care tests. Analytical and Bioanalytical Chemistry 1–31 (2021). [PubMed: 33432464]
11. Dias-Audibert FL et al. Combining machine learning and metabolomics to identify weight gain biomarkers. Frontiers in bioengineering and biotechnology 8, 6 (2020). [PubMed: 32039191]
12. Zheng C, Zhang S, Ragg S, Raftery D & Vitek O Identification and quantification of metabolites in 1h nmr spectra by bayesian model selection. Bioinformatics 27, 1637–1644 (2011). [PubMed: 21398670]
13. Association IRM et al. Bioinformatics: Concepts, Methodologies, Tools, and Applications (IGI Global, 2013)
14. Johnson CH & Gonzalez FJ Challenges and opportunities of metabolomics. Journal of cellular physiology 227, 2975–2981 (2012) [PubMed: 22034100]
15. Ayling M, Clark MD & Leggett RM New approaches for metagenome assembly with short reads. Briefings in bioinformatics 21, 584–594 (2020). [PubMed: 30815668]
16. Brumfield KD, Huq A, Colwell RR, Olds JL & Leddy MB Microbial resolution of whole genome shotgun and 16s amplicon metagenomic sequencing using publicly available neon data. PLoS One 15, e0228899 (2020) [PubMed: 32053657]
17. Garza DR, van Verk MC, Huynen MA & Dutilh BE Towards predicting the environmental metabolome from metagenomics with a mechanistic model. Nature microbiology 3, 456–460 (2018).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

18. 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).
19. Yin X et al. A comparative evaluation of tools to predict metabolite profiles from microbiome sequencing data. Frontiers in microbiology 11, 3132 (2020).
20. Kettle H, Louis P, Holtrop G, Duncan SH & Flint HJ Modelling the emergent dynamics and major metabolites of the human colonic microbiota. Environmental microbiology 17, 1615–1630 (2015). [PubMed: 25142831]
21. Quinn RA et al. Niche partitioning of a pathogenic microbiome driven by chemical gradients. Science advances 4, eaau1908 (2018). [PubMed: 30263961]
22. 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). [PubMed: 31856158]
23. 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).
24. Mallick H et al. Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. Nature communications 10, 1–11 (2019).
25. Le V, Quinn TP, 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).
26. Reiman D, Layden BT & Dai Y Mimenet: Exploring microbiome-metabolome relationships using neural networks. PLoS Computational Biology 17, e1009021 (2021). [PubMed: 33999922]
27. Morton JT et al. Learning representations of microbe–metabolite interactions. Nature methods 16, 1306–1314 (2019) [PubMed: 31686038]
28. Chen RT, Rubanova Y, Bettencourt J & Duvenaud D Neural ordinary differential equations. arXiv preprint arXiv:1806.07366 (2018).
29. Lu Y, Zhong A, Li Q & Dong B Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations. In International Conference on Machine Learning, 3276–3285 (PMLR, 2018).
30. Qiu C, Bendickson A, Kalyanapu J & Yan J Accuracy and architecture studies of residual neural network solving ordinary differential equations. arXiv preprint arXiv:2101.03583 (2021).
31. Dutta S, Rivera-Casillas P & Farthing MW Neural ordinary differential equations for data-driven reduced order modeling of environmental hydrodynamics. arXiv preprint arXiv:2104.13962 (2021).
32. 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). [PubMed: 30721227]
33. Franzosa EA et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nature microbiology 4, 293–305 (2019).
34. Swenson TL, Karaoz U, Swenson JM, Bowen BP & Northen TR Linking soil biology and chemistry in biological soil crust using isolate exometabolomics. Nature communications 9, 1–10 (2018).
35. Litonjua AA et al. Effect of prenatal supplementation with vitamin d on asthma or recurrent wheezing in offspring by age 3 years: the vdaart randomized clinical trial. Jama 315, 362–370 (2016). [PubMed: 26813209]
36. Litonjua AA et al. Six-year follow-up of a trial of antenatal vitamin d for asthma reduction. New England Journal of Medicine 382, 525–533 (2020). [PubMed: 32023372]
37. Lee-Sarwar KA et al. Integrative analysis of the intestinal metabolome of childhood asthma. Journal of Allergy and Clinical Immunology 144, 442–454 (2019). [PubMed: 30914378]
38. Lee-Sarwar K et al. Association of the gut microbiome and metabolome with wheeze frequency in childhood asthma. Journal of Allergy and Clinical Immunology 147, AB53 (2021).
39. Harvard T Harvard willett food frequency questionnaire. TH Chan School of Public Health, Department of Nutrition, Harvard University: Boston, MA, USA.
40. 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).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

41. Nelson KM, Reiber G & Boyko EJ 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). [PubMed: 12351468]
42. Marriott BP, 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). [PubMed: 20301013]
43. Moshfegh A Food and nutrient database for dietary studies (fndds).
44. Ridlon JM, Kang D-J & Hylemon PB Bile salt biotransformations by human intestinal bacteria. Journal of lipid research 47, 241–259 (2006). [PubMed: 16299351]
45. Bachmann V et al. Bile salts modulate the mucin-activated type vi secretion system of pandemic vibrio cholerae. PLoS neglected tropical diseases 9, e0004031 (2015). [PubMed: 26317760]
46. Ramírez-Pérez O, Cruz-Ramón V, Chinchilla-López P & Méndez-Sánchez N The role of the gut microbiota in bile acid metabolism. Annals of hepatology 16, 21–26 (2018).
47. Jia W, Xie G & Jia W Bile acid–microbiota crosstalk in gastrointestinal inflammation and carcinogenesis. Nature reviews Gastroenterology & hepatology 15, 111–128 (2018). [PubMed: 29018272]
48. Heinken A et al. Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease. Microbiome 7, 1–18 (2019). [PubMed: 30606251]
49. Duboc H et al. Connecting dysbiosis, bile-acid dysmetabolism and gut inflammation in inflammatory bowel diseases. Gut 62, 531–539 (2013). [PubMed: 22993202]
50. Thomas JP, Modos D, Rushbrook SM, Powell N & Korcsmaros T The emerging role of bile acids in the pathogenesis of inflammatory bowel disease. Frontiers in Immunology 246 (2022).
51. Kristal AR, Peters U & Potter JD Is it time to abandon the food frequency questionnaire? (2005).
52. Scalbert A et al. The food metabolome: a window over dietary exposure. The American journal of clinical nutrition 99, 1286–1308 (2014). [PubMed: 24760973]
53. Bolyen E et al. Qiime 2: Reproducible, interactive, scalable, and extensible microbiome data science. Tech. Rep., PeerJ Preprints (2018).
54. Callahan BJ et al. Dada2: High-resolution sample inference from illumina amplicon data. Nature methods 13, 581–583 (2016). [PubMed: 27214047]
55. Evans AM et al. High resolution mass spectrometry improves data quantity and quality as compared to unit mass resolution mass spectrometry in high-throughput profiling metabolomics. Metabolomics 4, 1 (2014).
56. Blum RE 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). [PubMed: 10746756]
57. Kingma DP & Ba J Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
58. Wang T wt1005203/mnode: Initial release (2023). URL 10.5281/zenodo.7602940.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

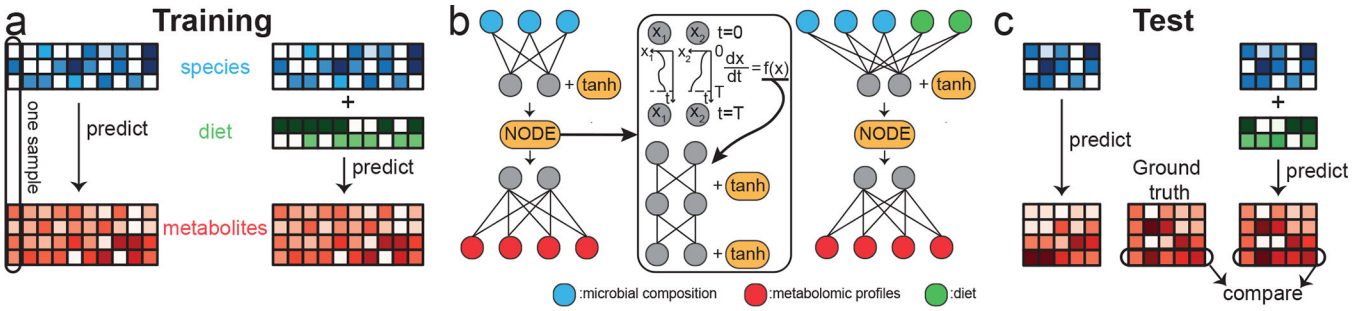


Figure 1: The mNODE workflow to predict metabolomic profiles from species microbial compositions and other dietary information.

Across all panels, blue-colored arrays represent the microbial composition. Green-colored arrays represent the dietary information. Red-colored arrays represent the metabolomic profiles. None of the arrays has missing data and white cubes in all arrays just mean small values. 15 hypothetical samples in total with 3 species, 2 dietary items, and 4 metabolites are used to illustrate the idea. The samples are divided into training and test sets with the 2/1 ratio. **a** 10 samples in the training set are used for training machine learning models. There are two ways to predict metabolomic profiles: one without diets included in the input and the other one with diets included in addition to microbial compositions. **b** The architecture of mNODE for two training approaches. The neural ODE is a module in the middle of the architecture and it computes the time evolution of ODEs whose first-order time derivatives are approximated by an MLP with one hidden layer. Grey nodes represent neurons in hidden layers of mNODE. **c** Fully trained mNODE can generate predictions for metabolomic profiles in the test set. For each metabolite, the Spearman's rank Correlation Coefficient ρ between its predicted profiles across samples and its true profiles in the test set is computed to quantify its predictability.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

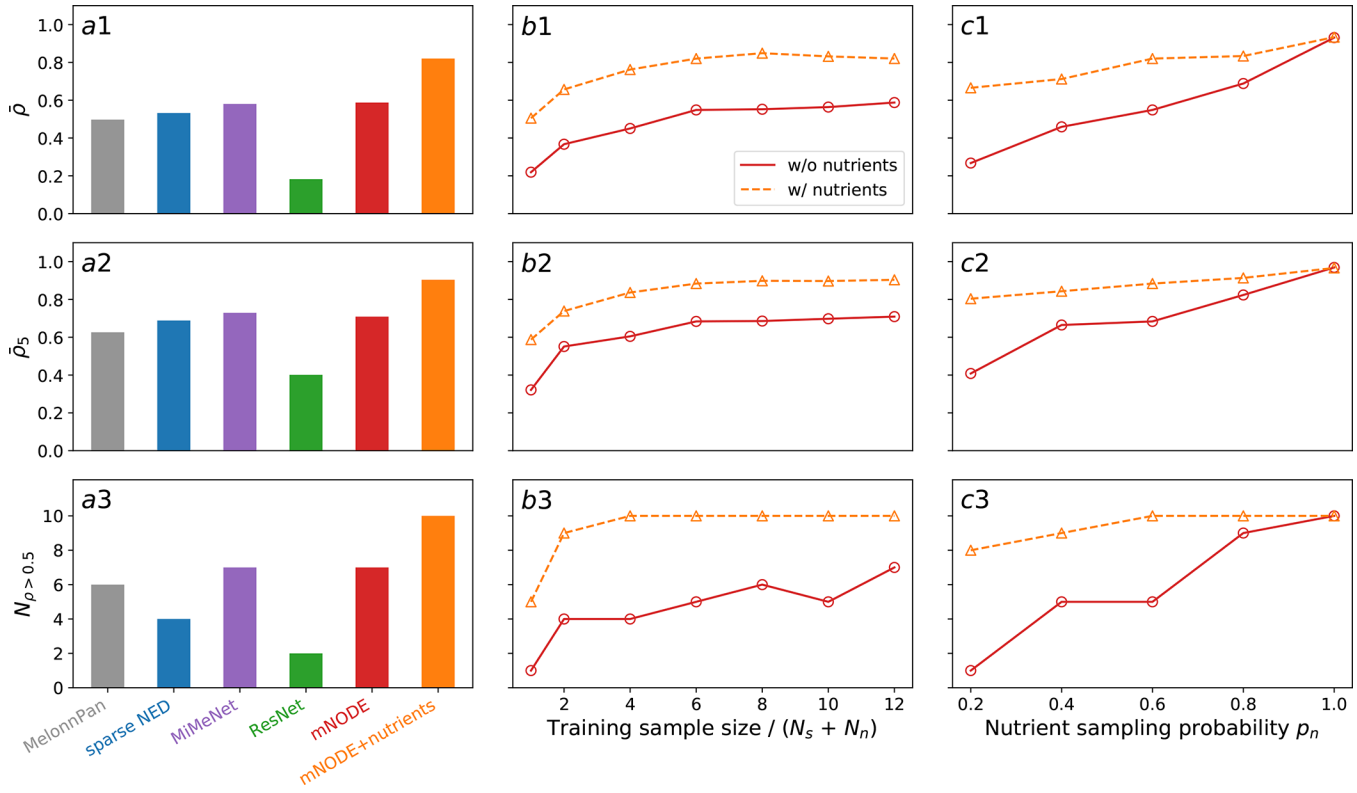


Figure 2: Model comparison and validation of mNODE on synthetic data generated by the microbial consumer-resource model.

The predictive performance of mNODE is compared to other methods through three metrics: a1 the mean SCC (Spearman Correlation Coefficient) $\bar{\rho}$, **a2** the top-5 mean SCC $\bar{\rho}_5$, and **a3** the number of metabolites with an SCC larger than 0.5 $N_{\rho>0.5}$. **b1-b3** The predictive performance of mNODE as a function of training sample size when the nutrient sampling probability (i.e. the fraction of nutrients externally supplied) $p_n = 0.6$ and the species sampling probability (i.e. the fraction of species introduced) $p_s = 0.5$. **c1-c3** The predictive performance of mNODE when the training sample size is 240 and the species sampling probability $p_s = 0.5$. Solid lines with circles are predicted results when nutrient supply rates (i.e. diets in the microbial consumer-resource model) are not included in the input of mNODE. Dashed lines with triangles are predicted results when nutrient supply rates are included in the input of mNODE.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

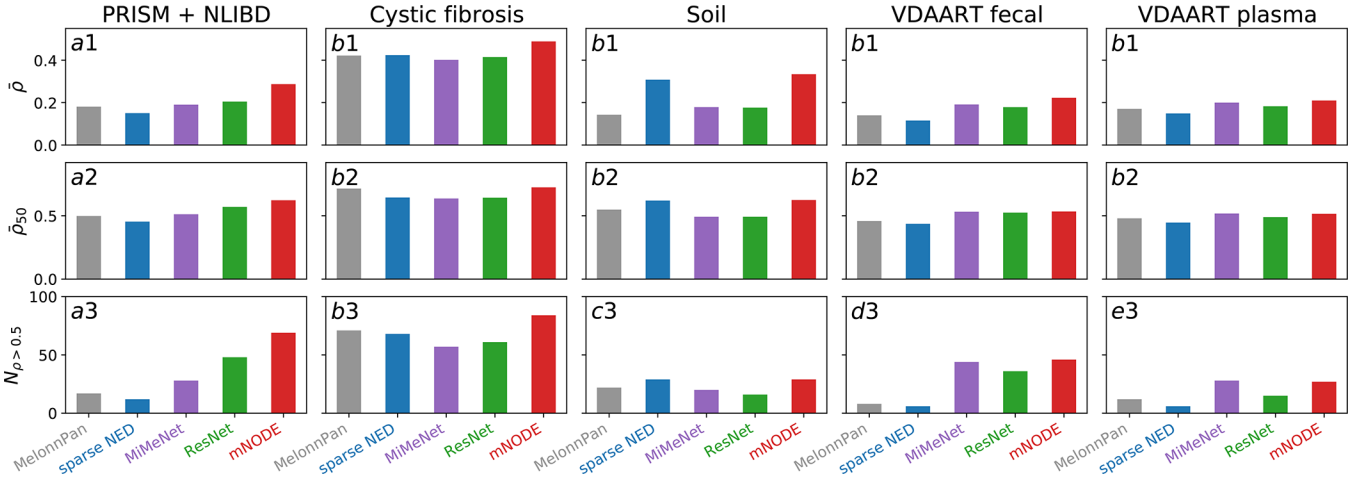


Figure 3: Performance comparison between mNODE and existing methods on real microbial community datasets.

Three metrics are adopted for comparing model performances: the mean SCC $\bar{\rho}$, the top-50 mean SCC $\bar{\rho}_{50}$, and the number of metabolites with SCCs larger than 0.5 $N_{\rho>0.5}$. All datasets[21,33–38] are randomly divided into training and test sets with the 80/20 ratio except for the PRISM and NLIBD dataset[33]. **a1-a3** Performance of methods after training on the PRISM and test on the NLIBD[33]. **b1-b3** Performance of methods on the data from lung samples of patients with cystic fibrosis[21]. **c1-c3** Performance of methods on the data from soil biocrust samples after 5 wetting events[34]. **d1-d3** Performance of methods on the data from fecal samples of children at age3[35–38]. **e1-e3** Performance of methods on the data from blood plasma samples of children at age 3[35–38].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

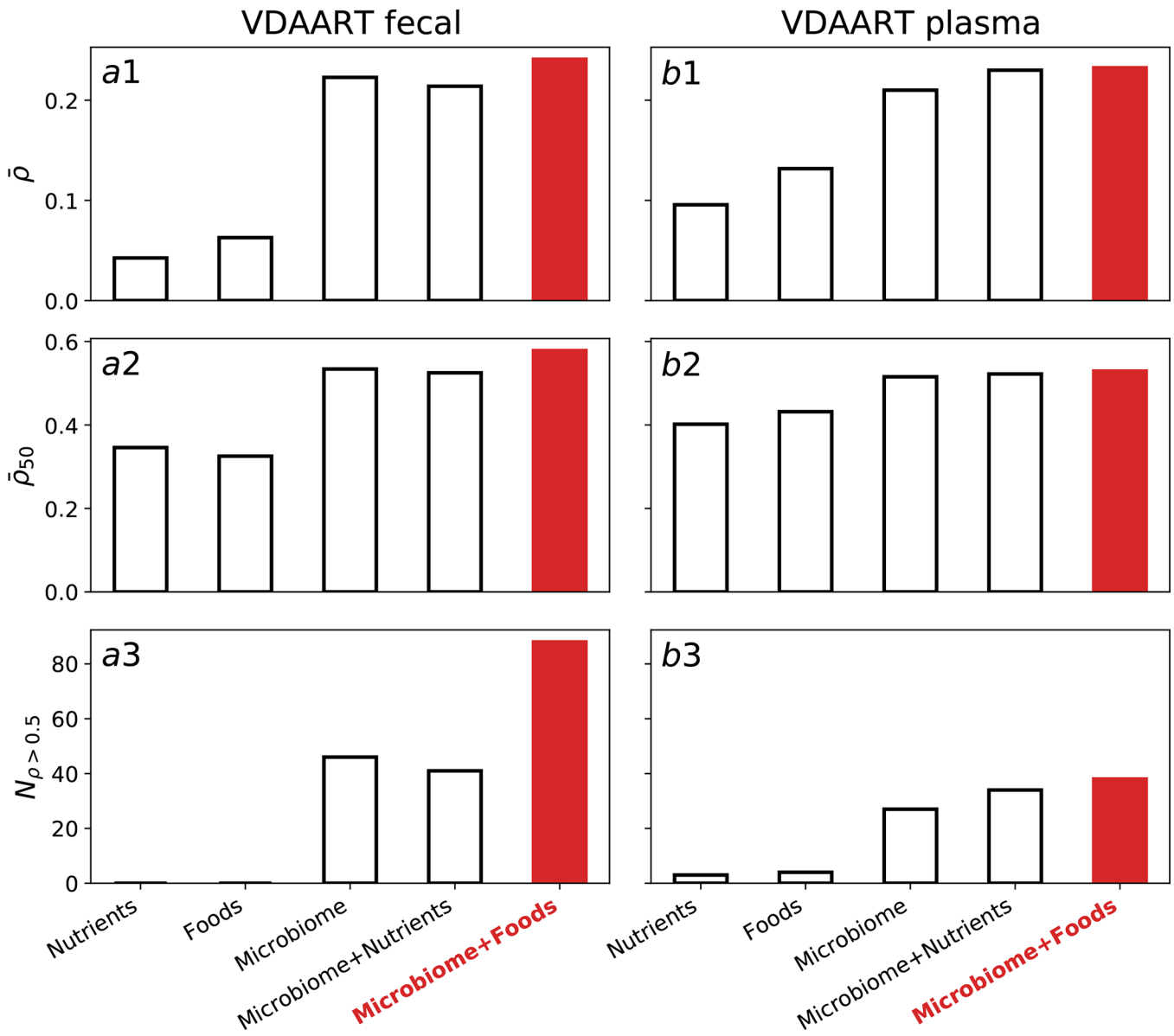


Figure 4: Performance of mNODE with different combinations of data types (microbial compositions, food profiles, and nutritional profiles) included in the input.

Five different combinations are used: foods only, nutritional profiles only, the microbial composition only, foods with microbial composition, and nutritional profiles with microbial composition. Three metrics are adopted for comparing model performances: the mean SCC $\bar{\rho}$, the top-50 mean SCC $\bar{\rho}_{50}$, and the number of metabolites with SCCs larger than 0.5 $N_{\rho>0.5}$. All datasets are randomly divided into the training set and test set with the 80/20 ratio. **a1-a3** Performance of methods on the data from fecal samples of children at age3[35–38]. **b1-b3** Performance of methods on the data from blood plasma samples of children at age 3[35–38].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

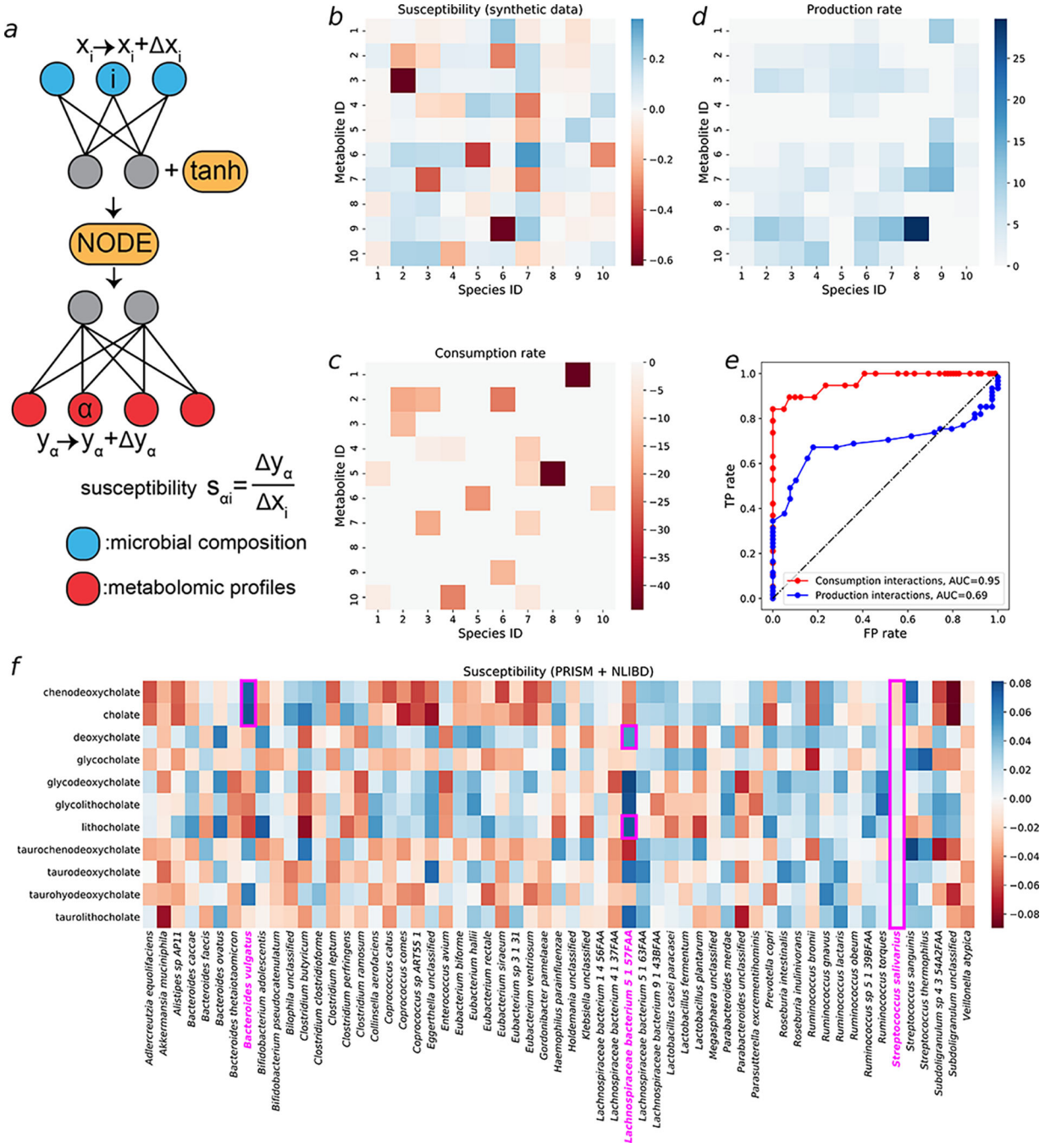


Figure 5: Using susceptibility of metabolite concentrations to microbial compositions of well-trained mNODE to infer microbe-metabolite interactions on both synthetic and real data (PRISM+NLIBD).

a The susceptibility of the concentration of metabolite α (y_α) to the relative abundance of species i (x_i), denoted as $s_{\alpha i}$, is defined as the ratio between the deviation in the concentration of metabolite α (Δy_α) and the perturbation amount in the relative abundance of species i (Δx_i). **b** Susceptibility values for all microbe-metabolite pairs in the synthetic data used in Fig. 2a1–a3. **c** The ground-truth consumption matrix and corresponding rates in synthetic data. All consumption rates are shown as negative values for the convenience of comparison with

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

panel b. **d** The ground-truth production matrix and corresponding rates in synthetic data. Details about how the production matrix is obtained can be found in the Supplementary Information section 2. **e** The ROC (Receiver Operating Characteristic) curve based on TP (True Positive) rates and FP (False Positive) rates which are obtained by setting different susceptibility thresholds for classifications of interactions. **f** The susceptibility values for all microbe-metabolite pairs in PRISM +NLIBD[33] used in Fig. 3a1–a3.