RESEARCH ARTICLE

MiMeNet: Exploring microbiome-metabolome relationships using neural networks

Derek Reiman[1], **Brian T. Layden**[2,3], **Yang Dai**[1]*

1 Department of Bioengineering, University of Illinois at Chicago, Chicago, Illinois, United States of America,
2 Department of Medicine, Division of Endocrinology, Diabetes, and Metabolism, University of Illinois at Chicago, Chicago, Illinois, United States of America, **3** Jesse Brown Veterans Affairs Medical Center, Chicago, Illinois, United States of America

* yangdai@uic.edu

Abstract

The advance in microbiome and metabolome studies has generated rich omics data revealing the involvement of the microbial community in host disease pathogenesis through interactions with their host at a metabolic level. However, the computational tools to uncover these relationships are just emerging. Here, we present MiMeNet, a neural network framework for modeling microbe-metabolite relationships. Using ten iterations of 10-fold cross-validation on three paired microbiome-metabolome datasets, we show that MiMeNet more accurately predicts metabolite abundances (mean Spearman correlation coefficients increase from 0.108 to 0.309, 0.276 to 0.457, and -0.272 to 0.264) and identifies more well-predicted metabolites (increase in the number of well-predicted metabolites from 198 to 366, 104 to 143, and 4 to 29) compared to state-of-art linear models for individual metabolite predictions. Additionally, we demonstrate that MiMeNet can group microbes and metabolites with similar interaction patterns and functions to illuminate the underlying structure of the microbe-metabolite interaction network, which could potentially shed light on uncharacterized metabolites through "Guilt by Association". Our results demonstrated that MiMeNet is a powerful tool to provide insights into the causes of metabolic dysregulation in disease, facilitating future hypothesis generation at the interface of the microbiome and metabolomics.

Author summary

The microbiome has shown to functionally interact with its host or environment at a metabolic level, however the exact nature of these interactions is not well understood. In addition, metabolic dysregulation caused by the microbiome is believed to contribute to the development of diseases such as inflammatory bowel disease, diabetes mellitus, and obesity. In this manuscript, we introduce a computational framework to integrate microbiome and metabolome data to uncover microbe-metabolite interactions in a data-driven manner. Our model uses neural networks to predict metabolite abundances from microbe abundances. The trained models are then used to derive microbe-metabolite feature scores, which are used for clustering microbes and metabolites into functional modules.

These module-based interactions are useful in generating biological insights and facilitating hypothesis generation for the investigation of their roles in various metabolic diseases. The software of our model is made freely available to interested researchers.

This is a *PLOS Computational Biology* Methods paper.

Introduction

The microbiome has been shown to impact both host development, normal metabolic processes, as well as the pathogenesis of various diseases [1–3]. Of particular interest is the microbiome of the human gut, which has been linked to diseases such as inflammatory bowel disease (IBD), obesity, and diabetes mellitus [4–7]. While previous studies have uncovered various microbe-disease associations, recent work also revealed the central role of bacterial metabolites and their impact on host health [8–13]. The microbiome-metabolome crosstalk is pervasive. For example, strong associations between microbes and metabolites were found in the gut and blood metabolomic profiles [14] and the gut of patients with IBD [15]. The abundance of metabolic pathways is relatively consistent despite considerable variability in taxonomic composition among individuals [16–19]. Thus, the identification of mechanisms of microbiome-metabolome interactions by modeling community metabolic activity is essential for understanding how the microbiome affects the host's health and for the development of precise therapies for the prevention or management of chronic diseases [20, 21].

Early methods for modeling have focused on mapping metagenomic features to metabolomic features due to the unavailability of metabolomic profiles. One set of methods referred to as Predictive Reactive Metabolic Turnover (PRMT) uses known enzymatic reactions from genome annotations and existing metabolic pathways to calculate the relative production and consumption of metabolites using metagenomic gene abundance profiles [22, 23]. Other methods have focused on constraint-based stoichiometric modeling using flux balance analysis to learn the flux rate of metabolites in the community [24, 25]. A major limitation of these approaches is their reliance on annotated references. Missing or incorrect annotations can lead to poor performing models. Additionally, the reliance on annotations makes it difficult to identify novel mechanisms.

More recently, several machine learning models have been developed to map metagenomic features to metabolites as both data have become increasingly available. One method, MelonnPan, uses the Elastic Net linear regression to model the relative abundance of each metabolite using metagenomic taxonomic or functional features [26]. The primary emphasis of MelonnPan is predictive modeling of metabolites so that the learned models can be used for the prediction of metabolomes in similar studies where only microbiome is available. MelonnPan displayed promising performance, however, it models each metabolite individually, missing the opportunity to use shared information across metabolomic features to boost prediction performance. Another method, mmvec, uses neural networks to estimates the conditional probability that a metabolite is present given the presence of a specific microbial sequence [27]. mmvec focuses on learning embeddings of microbial sequences and metabolites to capture microbe-metabolite interactions, thus, it cannot predict the entire metabolomic profile. Another recent model based on neural network encoder-decoder (NED) has proposed

constraints of sparsity and non-negative weights for mapping microbiomes to metabolomes [28, 29]. The use of non-negative weights in NED imposes a stringent constraint on the model, which may limit the learning capacity. Consequently, these existing models have not fully explored the microbiome-metabolome interactions that can be uncovered through integrative analysis.

In this work, we present MiMeNet (Microbiome-Metabolome Network), a multi-layer perceptron neural network (MLPNN) that models the community metabolome profile using metagenomic taxonomic or functional features obtained from a microbiome sample. The use of MLPNN allows MiMeNet to be scalable regarding the amount of metagenomic and metabolomic features in the two types of omics data. Furthermore, by learning multiple metabolites simultaneously, the underlying information can be transferred to augment the learning of metabolites with similar patterns [30], thus leading to more robust predictive models and hence more reliable microbiome-metabolome interactions. Here, we use three paired metagenomic-metabolomic datasets obtained from studies on IBD, cystic fibrosis, soil wetting environment as well as an additional external dataset of IBD to evaluate MiMeNet's ability to predict metabolomic profiles from metagenomic features by comparing with the existing methods. Using the IBD datasets, we further show how MiMeNet uses learned network models to generate hierarchies of metagenomic and metabolomic modules, which further reveal microbe-metabolite interactions and their associations to host IBD status. The MiMeNet package and the datasets used for evaluation are freely available at https://github.com/YDaiLab/MiMeNet.

Results

MiMeNet framework

MiMeNet is an integrative analysis framework for microbiome and metabolome data utilizing MLPNN, which trains models to accurately predict the metabolome based on a microbiome sample represented by the microbial taxonomic composition or microbial gene features. With learned network models, MiMeNet performs additional analysis to extract feature attribution scores between all microbe-metabolite pairs and organizes microbes and metabolites into modules according to the patterns of their attribution scores. These modules can be used to examine the microbe-metabolite interactions between modules and assess the statistical associations of modules to host phenotypes. MiMeNet takes advantage of multivariate learning by using the entire microbiome to model the abundance of all metabolites in a unified neural network architecture. This approach enables MiMeNet to train more robust models. An overview of the MiMeNet framework is shown in Fig 1.

The network architecture in MiMeNet is first determined for each paired dataset by tuning the hyperparameters for the number and size of the hidden layers, the L2 regularization penalty parameter, and the dropout rate. MiMeNet then trains multiple network models using 10-fold cross-validation and the predictive performance of a model is measured by the averaged Spearman correlation coefficients (SCCs) between the predicted and the observed abundance of the metabolites across the samples in testing sets. The protocols for MLPNN model training are outlined in the Methods section. MiMeNet then generates a background distribution of SCCs through multiple iterations of shuffling the dataset and performing a cross-validated evaluation on the shuffled set. This background distribution of SCCs is then used to determine a cutoff for significantly well-predicted metabolites. Then, using the learned network weights obtained from cross-validation training, MiMeNet constructs a score matrix of microbe-metabolite feature attributions between the microbes and well-predicted metabolites. Then MiMeNet biclusters the score matrix into microbial and metabolomic modules,
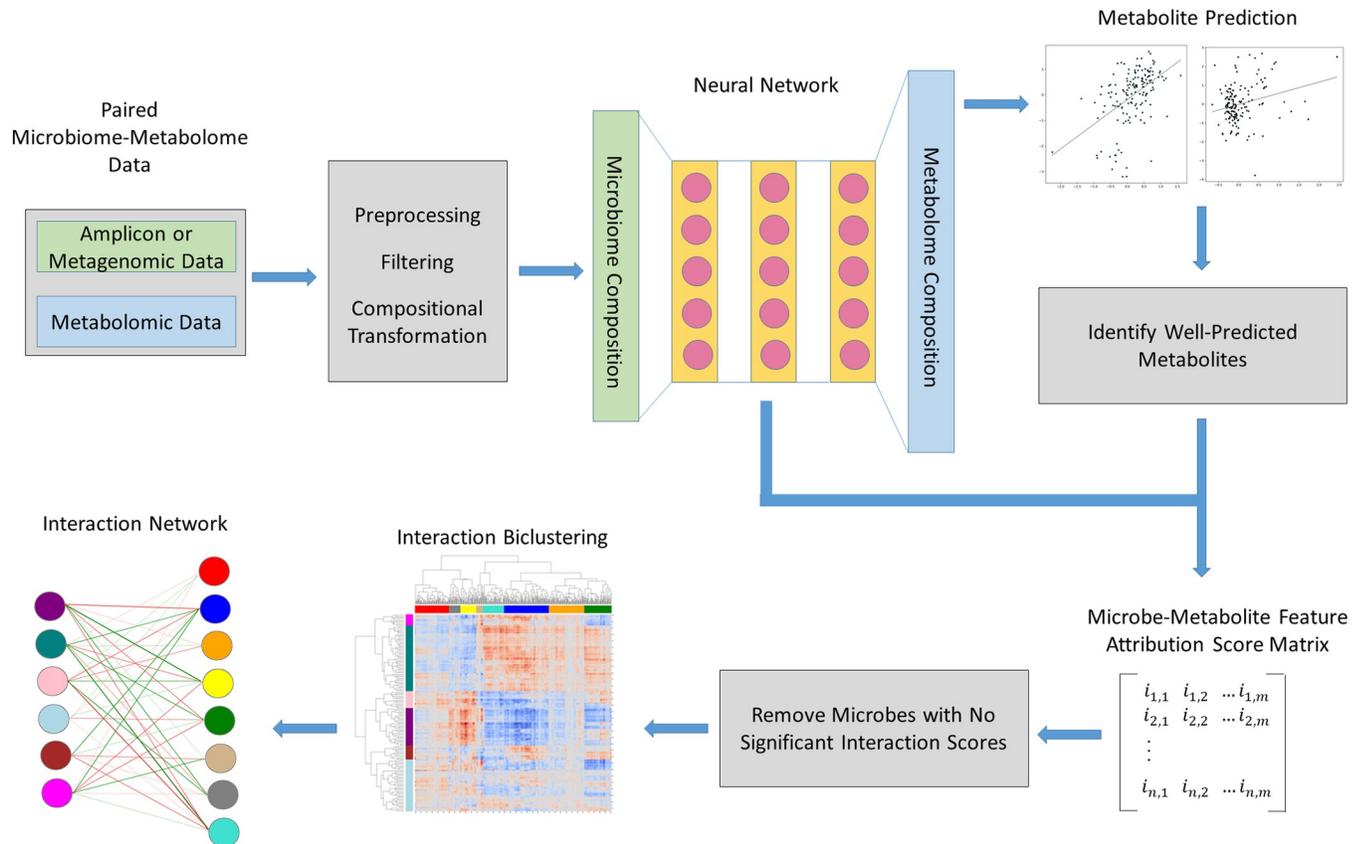
Fig 1. Framework of MiMeNet learning model. MiMeNet uses paired microbiome and metabolome data for model training. Microbiome abundance features (green) are used to train a neural network to predict metabolite abundance features (blue). Well-predicted metabolites are identified, and the trained models are used to learn a microbe-metabolite attribution score matrix. The attribution score matrix is biclustered into microbe and metabolite modules which are then used to construct a module-based interaction network.

https://doi.org/10.1371/journal.pcbi.1009021.g001

grouping microbes or metabolites with shared feature attribution patterns and constructs a module-based interaction network. Additionally, MiMeNet trains a final predictive model to be used to predict the metabolomic profile of an external set with similar microbial features (Methods).

MiMeNet identifies significantly correlated metabolites

We used three datasets for the development and evaluation of MiMeNet (Table 1). The first dataset was taken from a published study of patients enrolled in PRISM (the Prospective Registry in IBD Study at Massachusetts General Hospital) containing 121 IBD patients and 34 controls [15]. Additionally, it includes another set of 43 IBD and 20 control subjects from two

Table 1. Summary of the datasets used in evaluation. The patients with IBD have two subtypes: Crohn's Disease (CD) and Ulcerative Colitis (UC).

	#Case	#Control	#Microbes	#Metabolites
IBD (PRISM)	68 (CD), 53 (UC)	34	201	8848
IBD (External)	20 (CD), 23 (UC)	20	201	8848
Cystic Fibrosis	172	-	657	168
Soil	-	19	446	85

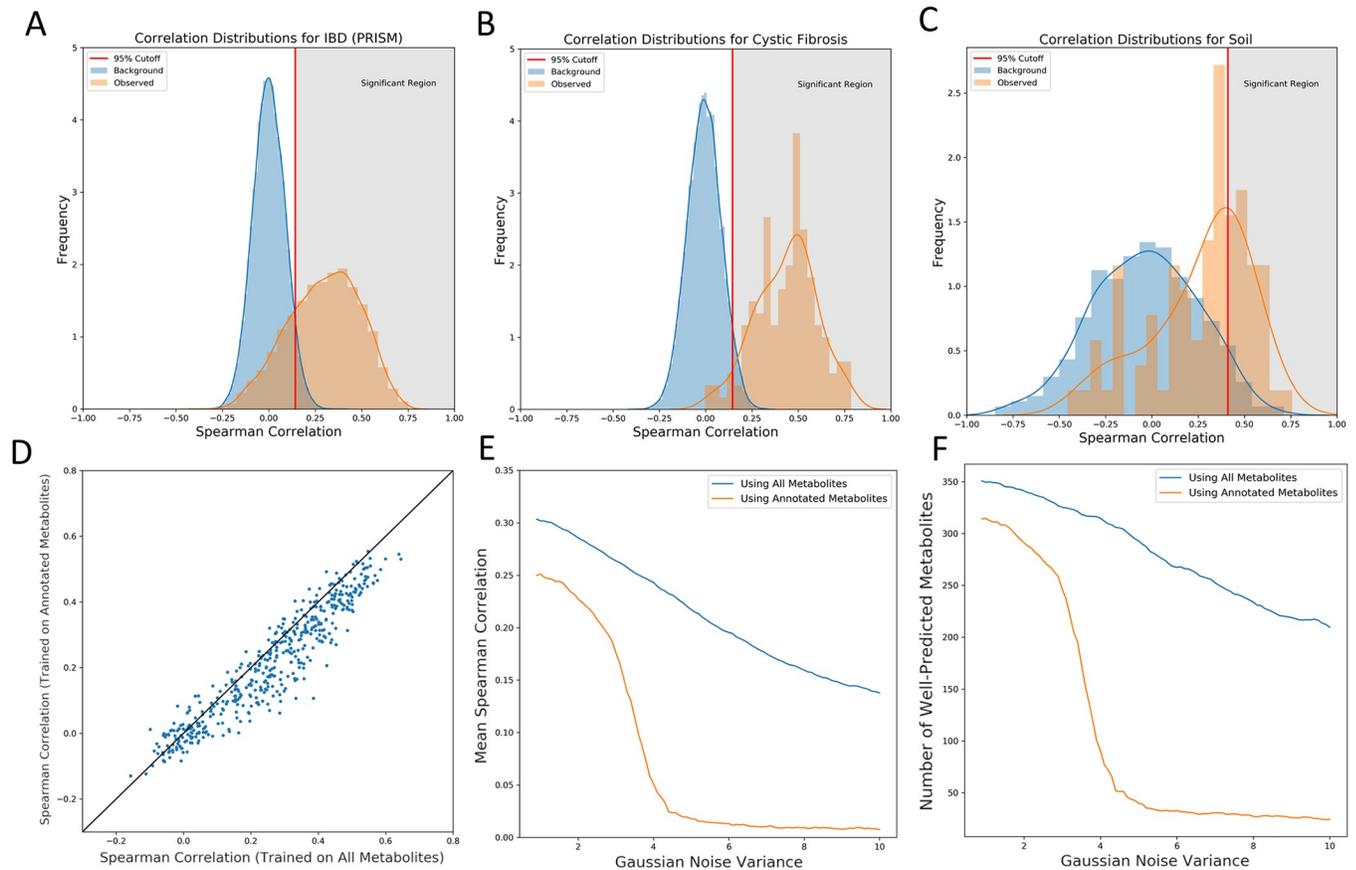https://doi.org/10.1371/journal.pcbi.1009021.t001

Fig 2. Distributions of correlations (Background and observed) and evaluation of multivariate learning. Background (blue) and observed (orange) distributions are shown for the (A) IBD (PRISM), (B) cystic fibrosis, and (C) soil datasets. The red vertical line denotes the 95th percentile of the background correlations and the gray area represents the well-predicted region using this threshold. (D) Scatter plot comparing the annotated metabolite correlations between models trained on just the annotated set and models trained on the full set of metabolites. (E) Mean correlation and (F) number of well-predicted metabolites found in models trained on the annotated set of metabolites and full set of metabolties as Gaussian noise is added to the annotated metabolite set input. All results in (D)-(F) are for prediction of the annotated metabolites.

other cohorts. This set will be used for external evaluation. The second dataset was taken from a study that collected lung sputum samples from 172 patients with cystic fibrosis [31]. The third dataset was taken from biocrust soil water and captures microbial and metabolic activity caused by soil wetting at five-time points across four biocrust successional stages [32]. For IBD, cystic fibrosis, and soil datasets, the numbers of microbes are 201, 657, and 446, respectively; and the numbers of metabolites are 8848, 168, and 85 respectively (see Methods for details).

Using the background distributions generated by MiMeNet (Method), the cutoffs for SCCs between the predicted and observed abundances of metabolites were found to be 0.136, 0.129, and 0.410 for the IBD (PRISM), cystic fibrosis, and soil datasets, respectively. Based on these cutoff values, MiMeNet identified metabolites to be well-predicted for 6857 (77.50%) of the 8848 metabolites in the IBD (PRISM) dataset, 143 (94.08%) of the 152 metabolites in the cystic fibrosis dataset, and 29 (34.12%) of the 85 metabolites in the soil dataset. The distributions of the SCCs in the background and observed data are shown in Fig 2A–2C. The soil dataset had the lowest percent of well-predicted metabolites, which could be due to the larger cutoff. We suspect that this is from the bootstrapping procedure being performed on the dataset of small size as well as the fact that the dataset is longitudinal and samples may be correlated with each

other. Our evaluation shows the strong predictability of the MiMeNet models trained on data with reasonable sample sizes.

Multivariate learning boosted MiMeNet's prediction performance

To evaluate if multivariate learning improves the prediction of the metabolomic profiles, we trained two separate models using 10 iterations of 10-fold cross-validation using the IBD (PRISM) dataset. The first model was trained to predict the entire set of metabolites, and the second model was trained to predict the 466 annotated set of metabolites without including the rest of the metabolites. We then compared the SCCs of the 466 metabolites from both models and observed that by training on the entire set of metabolites, the number of well-predicted metabolites for the annotated set increased from 333 to 366. Additionally, the SCCs of the annotated metabolites significantly increased from 0.259 to 0.309 when using all the metabolites to train MiMeNet ($P < 10^{-47}$, the Wilcoxon signed-rank test). The scatter plot comparing the prediction correlation performances is shown in Fig 2D.

Next, we evaluated the robustness of MiMeNet by gradually increasing noise to the annotated set of metabolites. Specifically, with 10-fold cross-validation, we trained models using all the metabolites and using only the annotated set to predict the 466 annotated set. For each partition of the cross-validated training, we added Gaussian noise to the annotated metabolites within the training data. We observed that the two models performed similarly under small amounts of noise. However, once the noise increased to higher levels and had a variance greater than 2, the models trained only on the annotated set collapsed and could no longer predict the annotated metabolites. On the other hand, the models trained using all the metabolites were much more robust to the noise at higher levels and could predict the annotated metabolites to a much greater degree compared to those trained using only the annotated set (Fig 2E and 2F). These results show MiMeNet benefited from multivariate learning.

MiMeNet is robust to training dataset size and hyper-parameter selection

To evaluate MiMeNet performance on different sizes of data for training and testing, we compared the k-fold cross-validated prediction correlations (k = 10, 5, 3, and 2) using the IBD (PRISM) and cystic fibrosis datasets (the soil dataset was excluded from this analysis due to the small data size). In the IBD (PRISM), we only observed a slight decrease in performance (mean correlation coefficient decrease from 0.297 to 0.218) as the number of partitions decreased (Fig 3A). Similarly, in the cystic fibrosis dataset, the correlation dropped slightly from 0.457 to 0.410 (Fig 3B). Additionally, we evaluated performance on random subsetting of 100%, 80%, 60%, and 40% of the entire datasets. For each level of subsetting, 3 random sets of subset data were generated. Then, for each set of data, network hyper-parameters were tuned and 10 iterations of 10-fold cross-validation was performed to evaluate how reducing the number of overall samples affected the prediction correlation (Fig 3C). As the size of the dataset decreased, we observed a decrease in the IBD (PRISM) dataset from a mean correlation of 0.287 to 0.179, and a decrease in the cystic fibrosis dataset from a mean correlation of 0.443 to 0.364. Moreover, we evaluated the IBD (External) dataset for each MiMeNet model trained on the IBD (PRISM) dataset and observed a decrease in mean correlation from 0.222 to 0.162. Even though there was a decrease in overall correlations as expected, we show that MiMeNet can still predict the metabolomic profiles when using smaller sets of training data.

We also compared the performance of the prediction using two types of microbial abundance representations: relative abundance (RA) and the centered log-transformation of abundance (CLR). The prediction correlations in the IBD (PRISM) dataset were comparable between both transformations, however, we saw an increase in correlations in the cystic
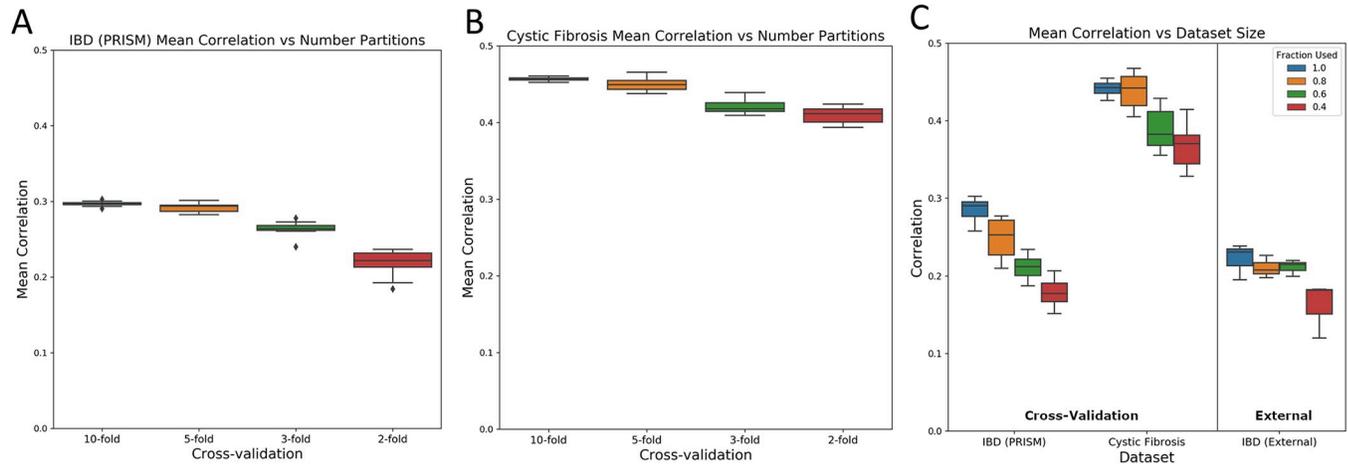
Fig 3. Mean correlation analysis using different amounts of training data. Correlations for 10-, 5-, 3-, and 2- fold cross-validation evaluations are shown for the (A) IBD (PRISM) and (B) cystic fibrosis datasets. (C) Subsets of the IBD (PRISM) and cystic fibrosis corresponding to 100%, 80%, 60%, and 40% of the total samples are used as an input for MiMeNet. Three random datasets for each level of subsetting were created and then mean correlation using 10 iterations of 10-fold cross-validation is calculated across the three. In addition, models trained on the complete subsets of the IBD (PRISM) data are used to evaluate the IBD (External) test set.

https://doi.org/10.1371/journal.pcbi.1009021.g003

fibrosis and soil datasets when using CLR. In addition, we observed an improvement in prediction performance on the IBD (External) test set when using the CLR transformation (S1 Fig).

Lastly, we evaluated if sharing the learned hyper-parameters across all cross-validated partitions in MiMeNet lead to overfitting. Although performing a single run of hyper-parameter tuning that is shared allows for much more computational efficiency, it could potentially be a source of bias. We evaluated the IBD (PRISM) and cystic fibrosis datasets using a single shared hyper-parameter set learned on the first partition against cross-validation where hyper-parameters are tuned every partition. Using the IBD (PRISM) dataset, we observed an increase in mean SCC when tuning every iteration, while in the cystic fibrosis dataset, we observed a decrease in mean SCC. Despite the decrease of performance in the cystic fibrosis dataset, 141 of the 143 significantly correlated metabolites were still identified. Comparisons of the two evaluations for each dataset are shown in S2 Fig. Together, MiMeNet shows a robust performance using the proposed parameter-tuning procedure.

MiMeNet outperforms models in MelonnPan

For benchmarking, we first compared MiMeNet to MelonnPan, a recent model that uses Elastic Net linear regression to predict metabolite abundance from microbial abundance features [26]. Elastic net regression applies a linear combination of both L1 and L2 regularizations to avoid overfitting. In the case of MelonnPan, a linear model is trained for one metabolite at a time and cannot benefit from multivariate learning. In our study, MelonnPan was evaluated using the same data partitions of the 10 iterations of 10-times cross-validation for each dataset. However, in the case of the IBD (PRISM) dataset, only the annotated metabolites were predicted due to the large running time for the entire metabolite set. On the other hand, MiMeNet was trained to predict all metabolites in the IBD (PRISM) dataset. We observed that in each of the datasets trained using cross-validation (Fig 4A–4C), MiMeNet has a higher correlation for prediction across all metabolites when compared to MelonnPan. In the IBD (PRISM) dataset, the mean correlation increased from 0.108 to 0.309 when evaluating the annotated metabolites. When training MiMeNet only on the annotated metabolites, we observed a similar result with
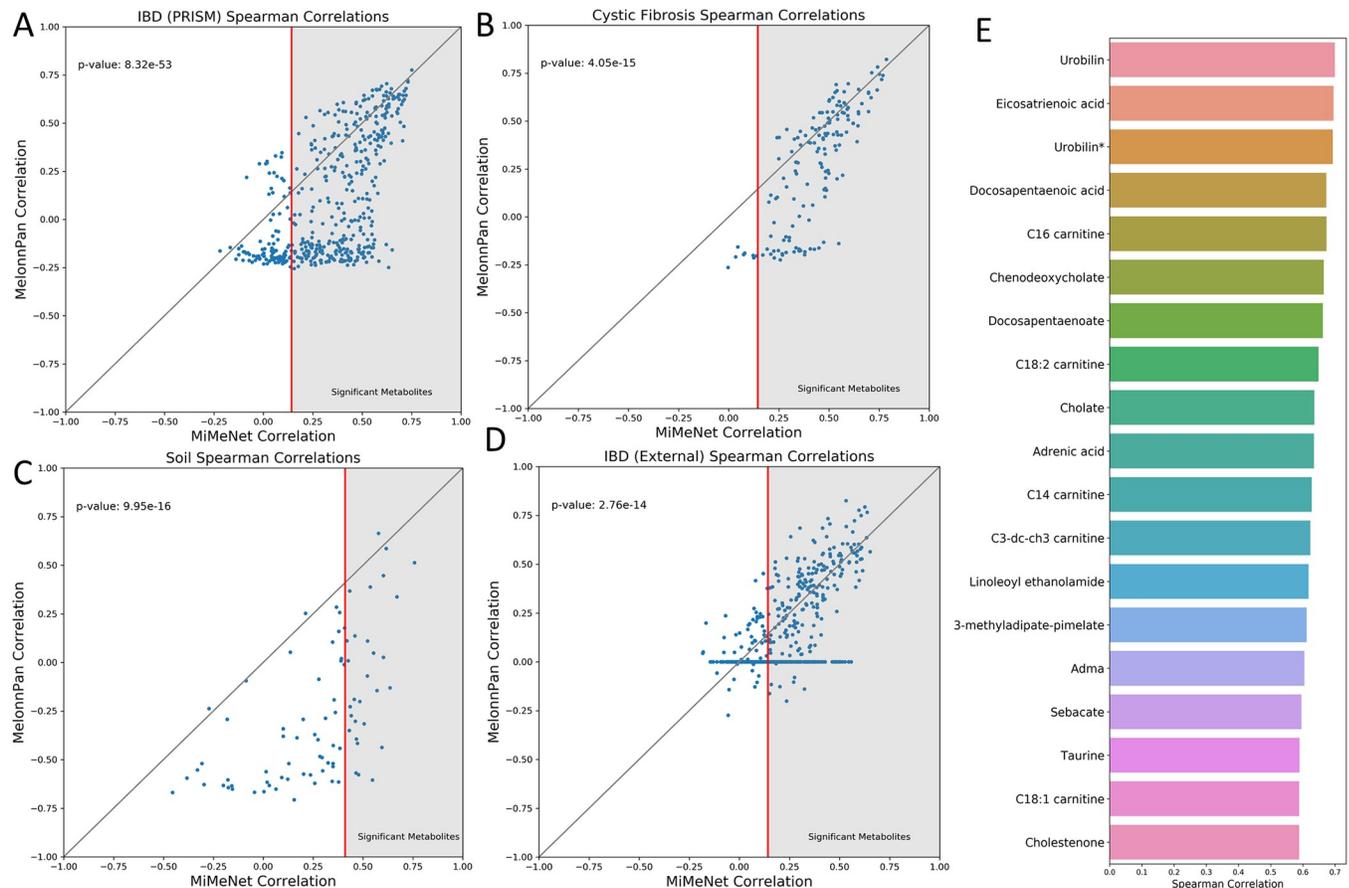
Fig 4. Scatter plots comparing MiMeNet to MelonnPan prediction correlations. Scatter plots showing the metabolite prediction correlations of MiMeNet against MelonnPan for (A) PRISM IBD dataset, (B) cystic fibrosis dataset, and (C) soil dataset, each trained using 10 iterations of 10-fold cross-validation. (D) In addition, 10 iterations of both models were trained on the PRISM dataset to evaluate the external IBD dataset. Each point represents the average correlation of a metabolite across 10 iterations of training. (E) The top 20 best correlated metabolites identified from the PRISM IBD dataset are shown.

https://doi.org/10.1371/journal.pcbi.1009021.g004

an increased correlation to 0.259 (S3 Fig). In the cystic fibrosis dataset, the mean correlation increased from 0.276 to 0.457. In the soil dataset, the mean correlation from MelonnPan was -0.272 and was increased to 0.264 using MiMeNet. Moreover, we evaluated the performance of the models obtained from MelonnPan and MiMeNet using the IBD (External) dataset on the annotated metabolites. The mean correlation of the annotated metabolites was increased from 0.168 to 0.275 (Fig 4D). Among the top 20 well-predicted metabolites from MiMeNet from the IBD (PRISM) dataset (Fig 4E), they are fatty acids (eicosatrienoic acid, docosapentaenoic acid, adrenic acid, and docosapentaenoate), and bile acids (chenodeoxycholate and cholate). Both of these classes of metabolites have been associated with IBD in previous studies [33–35].

Additionally, within the IBD (PRISM) dataset, MiMeNet identified 351 well-predicted metabolites from the 466 annotated metabolites. Even though MelonnPan uses a default correlation cutoff of 0.3, when using the same correlation cutoff derived by MiMeNet, MelonnPan identified 198 well-predicted metabolites of which 181 were identified by MiMeNet. In the cystic fibrosis dataset, MiMeNet identified 143 well-predicted metabolites while MelonnPan identified 104. In the soil dataset, MiMeNet identified 29 well-predicted metabolites while MelonnPan identified 4. When training using the entire IBD (PRISM) dataset to predict the IBD (External) test data, MiMeNet identified 308 well-predicted metabolites while MelonnPan

identified 186, of which 160 were also identified by MiMeNet (S4 Fig). When analyzing the overall prediction correlation and number of well-predicted metabolites, we observed that MiMeNet's improvement was not a global improvement across all metabolites, but rather it came from MiMeNet being able to model a large set of microbes that MelonnPan could not. For example, in the IBD (PRISM) dataset, there were 237 metabolites with a negative prediction correlation. Of these metabolites, MiMeNet was able to predict 160 with a correlation above the determined cutoff. These metabolites also made up the set of metabolites with a prediction correlation of 0 in the IBD (External) set when using MelonnPan. Upon investigation, this set of metabolites was predominantly composed of triacylglycerols, long-chain fatty acids, and bile acids. These three classes of metabolites have been shown to interact with various microbes as well as relate to IBD patients.

We observed that the running time of MiMeNet was robust and did not scale largely with the number of metabolites as all three datasets complete in similar timespans (Table 2). These results show that MiMeNet benefited from multivariate learning, the scalability of MLPNN, and the ability of MLPNN in capturing complex relationships between microbiome and metabolomes.

In addition, we benchmarked MiMeNet against other general regression models, i.e., Random Forest (RF), multivariate Elastic Net, and canonical correlation analysis (CCA) models using 10 iterations of 10-fold cross validation (S1–S3 Tables). Based on three evaluation metrics, i.e., SCC, Pearson correlation coefficient (PCC), and mean absolute error (MAE), we observed that for the IBD (PRISM) and cystic fibrosis datasets, MiMeNet and RF models performed best. For the soil dataset, we observed that CCA models performed the best, which may be due to the extremely small sample size of the soil dataset. When models were trained on the entire IBD (PRISM) dataset to predict the IBD (External) dataset, we observed that MiMeNet outperformed all other models.

MiMeNet identifies biologically important modules of microbes and metabolites

To discern what group of microbes contributed collectively to a group of metabolites, we computed the feature attribution scores for all pairs of microbes and the 6857 well-predicted metabolites using the network weights of the trained models obtained from the IBD (PRISM) data set. We identified 163 microbes that had at least one significant attribution score with a well-predicted metabolite (Methods). A positive score means that the microbe contributes positively to the prediction of the abundance of the metabolite. Likewise, a negative score contributes negatively to the prediction of the abundance of the metabolite. To reveal the pattern of attribution scores, we grouped the microbes and metabolites into modules using biclustering (Methods). We identified 8 modules of microbes and metabolites respectively (Fig 5A) and computed the module feature value as the average abundance of features within the module

Table 2. Running time for MiMeNet and MelonnPan. Running time for MiMeNet was calculated using the recommended settings from the Github page: 20 random searches for hyper-parameter tuning, 10-iterations of 10-fold cross-validation for prediction evaluation, 10-iterations of 10-fold cross-validation for background generation, and the training of 10 final candidate models. MelonnPan was run using 10-iterations of 10-fold cross-validation and for the IBD (PRISM) dataset MelonnPan only used the 466 annotated metabolomic features.

	MiMeNet Running Time (H:M:S)	MelonnPan Running Time (H:M:S)
IBD (PRISM)	1:11:39	16:33:04
Cystic Fibrosis	1:18:05	3:47:27
Soil	1:54:24	1:31:29

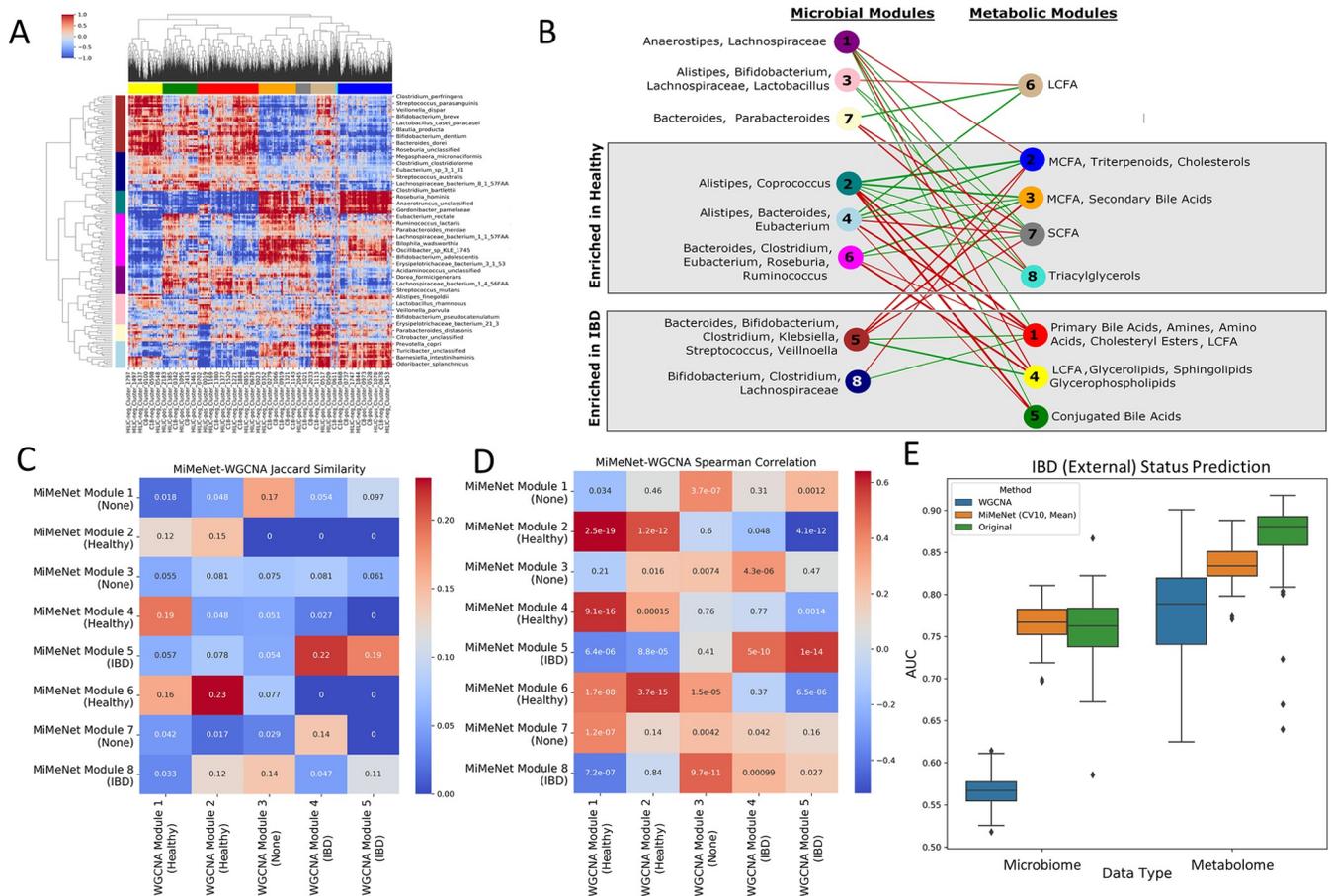https://doi.org/10.1371/journal.pcbi.1009021.t002

Fig 5. Clustering of microbes and metabolites in the IBD datasets. (A) Clustering of microbes (row) and metabolites (column) based on the feature attribution scores obtained from the IBD (PRISM) dataset. Row and column colors represent assigned modules. (B) Network connecting microbial modules with metabolomic modules. The most abundant genera are annotated for microbe modules. The most abundant metabolite classes are annotated for the metabolite modules. Red connections indicate negative attributions and green edges indicate positive attributions. Node color represents module color from (A). (C) Jaccard Index and (D) Spearman correlation between module features of WGCNA and MiMeNet. Jaccard Index values and Spearman correlation p-values shown in the boxes. (E) IBD status prediction of IBD (External) using models trained on WGCNA module feature values, MiMeNet module feature values, and original abundance from IBD (PRISM) microbial and metabolomic data.

https://doi.org/10.1371/journal.pcbi.1009021.g005

for each sample. We then constructed a bipartite graph between microbe and metabolite modules such that the attribution score was the mean attribution score found within the block identified by both modules (Fig 5B).

We further went to examine if a module is enriched for one patient group (IBD or healthy) by comparing the average normalized feature values of the members within the module between the two groups using the IBD (PRISM) samples (P<0.05, Wilcoxon rank-sum test). We observed that 7 of the 8 microbial modules were significantly different between groups using the IBD (PRISM) data. Using the IBD (External) data, two of the modules were still significantly different; and even though other modules were no longer significant, they shared similar trends in the differences between groups (Fig 6A and 6C). We also found that 7 of the 8 metabolic modules were significantly different between groups using the IBD (PRISM) data and when using the IBD (External) data, the same 7 modules were also significantly different between groups (Fig 6B and 6D). To further examine the MiMeNet's predictive performance in each metabolite module, we calculated the mean SCC and PCC values of members within each module from both the cross-validated evaluation and the evaluation on the IBD
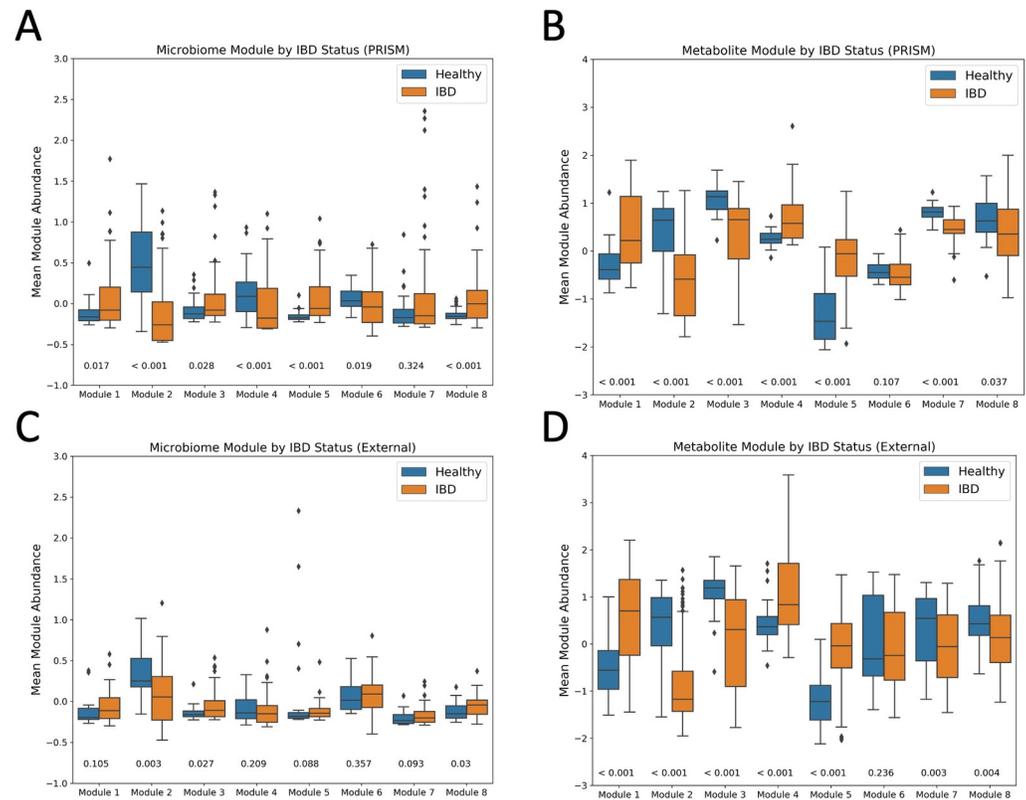
Fig 6. Microbial and metabolic module abundance by patient status in the IBD (PRISM) dataset. The mean normalized abundance of members within modules are shown here for healthy patients and IBD patients from the IBD (PRISM) dataset using (A) microbial and (B) metabolic modules and from the IBD (External) dataset using (C) microbial and (D) metabolic modules identified by MiMeNet. P-values from a two-sided Wilcoxon rank-sum test are shown on the bottom.

(External) data (S5 Fig). For the cross-validated evaluation, the mean SCC for each module ranged from 0.25 to 0.41 and the mean PCC ranged from 0.21 to 0.35, showing that each module contributed to the overall prediction performance of the MiMeNet model. On the IBD (External) evaluation, the mean SCC ranged from 0.14 to 0.28 and the mean PCC ranged from 0.06 to 0.25. Although the values decreased in the IBD (External) data, the modules with higher mean SCC and PCC values in the cross-validated evaluation were also the modules with the higher SCC and PCC values in the IBD (External) data. Taken together, these results show that the predictive ability as well as the information carried by the collective members of each module were transferrable to an external cohort of patients.

Four metabolite modules were enriched in healthy subjects, i.e., higher average module feature values. The first module (module 2) contained medium-chain fatty acids (MCFA), triterpenoids, and cholesterols. Triterpenoids, such as oleanolic acid and maslinic acid, have been shown to have an anti-inflammatory effect as well as enhance the integrity of intestinal tight junctions [36], and have been explored as therapeutic options for IBD [36–38]. In addition, both cholesterols and MCFAs have been noted to be depleted in subjects with IBD [39]. The second module (module 3) contained MCFA molecules as well as secondary bile acids such as deoxycholic acid and lithocholic acid, which was found to be reduced in IBD patients [40, 41]. The third module (module 7) was mainly composed of short-chain fatty acids (SFCA) such as propionate, butyrate, and valeric acid, all of which were shown to be protective against IBD

[34]. The final module (module 8) contained triradylcglycerols, which in previous studies have been found depleted in subjects with IBD [39].

Similarly, three metabolite modules were found enriched in IBD patients. The first module (module 1) was composed of a large portion of primary bile acids, amines, amino acids, cholesteryl esters, and long-chain fatty acids (LCFA). Primary bile acids aid in the digestion of lipids and are further deconjugated to secondary bile acids by microbes in the gut. The deconjugation of primary bile acids in subjects with IBD was previously shown to have an impaired ability causing a greater abundance of primary bile acids [40]. Additionally, these primary bile acids were shown in previous studies to bind to the farnesoid X receptor, which is linked to the elevated immune response in IBD [42]. Cholesteryl esters have also been shown to be elevated in subjects with IBD, potentially explained by lipid mobilization or by increased intestinal permeability [43]. The amine group within this module is composed of N-acylethanolamines as well as sphingosine. N-acylethanolamines have been shown to alter the gut microbiome and potentially increase levels of lipopolysaccharides in the intestines, causing inflammation [44]. Sphingosine along with fatty acids that make up ceramide were also found within this module. Ceramide is a precursor to sphingosine-1-phosphate, a signaling sphingolipid that has been implicated in increased inflammation of the gut [45]. The LCFAs found here have been implicated with IBD contained eicosapentaenoic acid, arachidonic acid, and docosapentaenoic acid, which were all identified in the original study as well [15]. The largest group in this module, however, comprised of 31 (24.4%) different amino acids and amino acid derivatives. Amino acids have been found in a previous study to be elevated in IBD patients, possibly due to the increase in the bacterial enzyme urease promoting a flux of nitrogen for the synthesis of amino acids by the host [46]. The second module (module 4) contained LCFAs such as docosahexaenoic acid, arachidonic acid, and eicosapentaenoic acid, as well as glycerolipids, glycerophospholipids, and sphingolipids. Bacterial-derived sphinogolipids have been shown to play a crucial role in the development of IBD through multiple signaling pathways [47]. The elevated glycerolipids and glycerophospholipids were also identified in the original study [15]. The third module (module 5) contained mostly conjugated bile acids, which is similar to primary bile acids have been shown to be elevated in IBD subjects due to a decreased ability to deconjugate the bile acids into secondary bile acids. In addition to disease related metabolite modules, MiMeNet also identified one module (module 6) not associated with disease status. This module was composed of a small number of LCFAs.

Lastly, we compared the microbial modules in the IBD (PRISM) dataset identified by MiMeNet to those identified by the Weighted Correlation Network Analysis (WGCNA) (Methods). The module features of a sample were calculated as the average normalized abundance of the members within the module. Using the Jaccard similarity between the members of the modules as well as the Spearman correlation coefficient between module features across samples, we observed only a small consensus between the two groupings (Fig 5C and 5D).

To evaluate if the modules embody the useful information for IBD prediction, we trained neural network models on the IBD (PRISM) dataset using three different sets of inputs: WGCNA module features, MiMeNet module feature values, and the original abundance. We trained 100 networks for each input type and each network was then evaluated on the IBD (External) dataset using the same input feature structure. For both microbial and metabolomic data, we observed that MiMeNet's modules had similar predictive power to the original features, and that the models trained using the feature values of the WGCNA modules had a much lower mean AUC value (Fig 5C). Moreover, the metabolite abundance values and their module feature values are more predictive of IBD status compared to the microbial abundance and their module feature values, respectively. Additionally, we compared the prediction of IBD status when MiMeNet uses k-fold cross-validation (k = 10, 5, 3, and 2) and observed that

modules generated using 10-fold cross-validation were most predictive of IBD status (S6 Fig). Lastly, we compared the modules of well-predicted metabolites for the IBD (PRISM) dataset when using CLR and RA transformations. We observed a large consensus of well-predicted metabolites using both transformations, but MiMeNet was able to identify slightly more well-predicted metabolites when using CLR (S7 Fig). We also observed that the modules generated using CLR were more predictive of IBD status than those generated using RA values (S6 Fig). Taken together, the analyses demonstrate that MiMeNet has organized microbes and metabolites into biologically meaningful modules.

MiMeNet identifies biologically important microbe-metabolite module interactions

We analyzed how the MiMeNet's microbe modules impacted the various metabolite modules identified by MiMeNet. SCFAs, and in particular butyrate, have been shown to be protective of inflammation in the gut and important for gut homeostasis [48]. A recent study identified ten microbial species commonly found in the gut microbiome that produce butyrate [49]. Five of these microbes, *Eubacterium biforme*, *Eubacterium hallii*, *Eubacterium rectale*, *Roseburia intestinalis*, and *Roseburia inulinivorans* were found in microbial module 6, which had a strong positive association with the metabolite module comprising of SCFAs. For comparison, we calculated the Spearman correlation between butyrate and each microbe in the IBD (PRISM) and selected significant pairs after multiple test correction (adj. P<0.05, Benjamini-Hochberg). Although the pairwise analysis identified 12 significant correlations, only one pair contained a microbe, *Anaerostipes hadrus*, from the set of butyrate-producing microbes, indicating the pairwise univariate analysis can only capture the strongest correlation pairs. This suggests that not only could MiMeNet capture more biologically relevant microbe-metabolite interactions but that it can also group them into meaningful modules based on common interactions.

In addition, we observed that metabolite module 5 had very weak positive interactions and very strong negative interactions with microbial modules 2 and 4. This module contained primarily conjugated bile acids, which are formed when primary bile acids are conjugated with either taurine or glycine in the liver. Since this process is performed independently of the gut microbiome, it would be expected to not observe a strong positive association. However, once these bile acids enter the gut, various microbial enzymes convert these conjugated bile acids into secondary bile acids through deconjugation, oxidation, and 7-dehydroxylation. Previous studies have identified multiple genera of microbes that produce enzymes responsible for this conversion, including *Bacteroides*, *Clostridium*, *Eubacterium*, and *Ruminococcus* [50], all of which constitute a large portion of microbial modules 2, 4, and 6. These modules all show a negative interaction with conjugated bile acids and a positive interaction with secondary bile acids, suggesting that microbes found within these modules regulate the conversion of conjugated bile acids to secondary bile acids.

Discussion

We have developed MiMeNet for integrative analysis of the paired microbiome-metabolome datasets using neural networks. The objectives of MiMeNet are to model microbiome-metabolome interactions and uncover functional relationships between microbes and metabolites. Using datasets obtained from the human gut of subjects with IBD, lung sputum of subjects with cystic fibrosis, and soil wetting environmental studies, we showed that MiMeNet models can produce a robust and accurate prediction of the community metabolome based on the microbial taxa abundance in both cross-validation and external validation. MiMeNet is empowered by neural networks' capability of modeling nonlinear relationships and

multivariate learning, which predicts the abundance of all metabolites at once, which can help facilitate prediction through shared information across different metabolites. Indeed, our results of the IBD data demonstrated that the MiMeNet prediction on the set of the annotated metabolites benefited from including tasks of predicting the rest of the unannotated metabolites in the metabolome profiles (Fig 2D–2F). We note that since not all metabolites may be associated with microbes, some metabolites will have lower prediction correlations, which resulted in an overall lower mean correlation across all metabolites. We also observed a higher threshold value for the soil data (Fig 3A–3C), which may be due to the longitudinal observations.

MiMeNet also facilitates additional analysis from the learned network models. Using the IBD data, we showed that the feature attribution scores derived from the network weights could be used to construct modules of microbes with similar positive or negative effects on a set of metabolites. This is a unique feature that other regression-based models (e.g. random forest) can not provide. Grouping metabolites into modules with similar feature attribution score patterns can facilitate the annotation of uncharacterized metabolites through "Guilt of Association" [15, 51]. This is extremely important in the field of metabolomics due to a large amount of current "dark matter" [52]. Untargeted metabolomics study has been hindered by many unknown metabolites and much of the newly collected information remains uninterpreted. The annotated metabolites clustered in a MiMeNet module that contains unannotated metabolites may provide a clue that those metabolites participate in similar biochemical reactions due to similar interaction patterns with some microbial modules. Those unannotated metabolites may relate to the annotated metabolite structurally or functionally. Additionally, if microbial gene features are used as input to train MiMeNet, it may further uncover gene-metabolite association from the feature attribution scores identified from the MiMeNet model. Although the MiMeNet analysis is data-driven without incorporating mechanistic knowledge, these types of evidence obtained from the integrative analysis of metagenomes and metabolomes could be used in predictive computational approaches such as MAGI and MINEs to increase the confidence of metabolite identification [53]. Furthermore, we have shown that the predictive models learned from the derived modules were equally competitive for the prediction of IBD status compared to those built on the entire microbes or metabolites, suggesting the metabolic functional relevance of microbial modules. This direction could be further explored in integrating omics data for host phenotype prediction [23].

The predictive model in MiMeNet distinguishes it from MelonnPan [26], which uses a regularized linear regression to model each metabolite separately. MiMeNet models all metabolites nonlinearly and benefits from learning the shared information. The MiMeNet model is also different from another highly relevant predictive model of NED in several aspects. NED models the metabolome using the representation of microbiome in the latent space generated by an encoder [28, 29]. It imposes an additional sparsity constraint to prevent overfitting and non-negative weights for ease of interpretation. Our comparison shows that the MiMeNet model is competitive with the NED model in terms of predictive accuracy and the ability to make well-prediction for a larger set of metabolites (S8 Fig). This suggests that the constraint of non-negative weights may diminish the capability of the NED to capture interactions where some microbes negatively affect some metabolites. Indeed, it was reported that more than 50% of associations identified between microbial metabolic pathways and species in human fecal and blood samples are negative [19]. Similar to MiMeNet, the NED generated latent microbiome space, which was shown to contain biologically relevant information useful in discrimination of Crohn's disease, ulcerative colitis, and healthy subject, and prediction of other clinical measurements such as usage of immunosuppressive therapy. However, since NED models metabolites using the microbiome latent space representation, the interpretation of

microbe-metabolite interaction is less intuitive, and it does not reveal modules in microbes and metabolites with shared interaction patterns. Nevertheless, we note that NED belongs to a broad class of integrative approaches to omics data analysis using latent components derived from various statistical and machine learning models [54].

The scope for integrative analysis of microbiome and metabolomics data developed so far are diverse, ranging from the identification of statistical associations using univariate and multivariate correlation analyses to predictive modeling based on machine learning and metabolic network-based modeling [8, 14, 15, 18, 55–57]. For example, a recently released webserver, M2IA [58], is an excellent tool to streamline the statistical analysis, such as overall similarity assessment of the two omics datasets, pairwise correlation analysis between microbes and metabolites, a heatmap visualization tool for revealing positive and negative correlation patterns. It also includes other components including WGCNA for identifying clusters of metabolites based on the pairwise Spearman's correlations of metabolites, supervised multivariate analyses of detection of disease associations on the microbes, or metabolites or functional annotations. M2IA, however, does not provide predictive models to quantify the impact of the abundance of multiple microbes on a metabolite abundance. This is the major distinction between MiMeNet and other machine learning-based methods (MelonnPan [26], mmvec [27], and NED [28, 29]) specifically designed for microbiome and metabolome integrative analysis. Despite the progress, all these methods including MiMeNet, are still limited in providing biological plausibility and mechanistic insights. Future direction for MiMeNet extension could be the development of procedures to discern the statistical association of microbe-metabolites interactions to host diseases and detect modules of metabolites that require specific microbial taxa in healthy and disease samples. In addition, datasets including additional metagenomics data could be integrated to allow for predicting metabolites from genes. This could potentially reveal richer functional connections that are not currently annotated as well as provide a richer functional exploration of the impact of the microbiome on the host metabolome. Moreover, neural networks could be utilized to model how the host metabolome influences gut microbial composition.

Methods

Data for evaluation

Concurrent profiles of microbiome and metabolome from varying environments were used for evaluation (Table 1). The first dataset was taken from a published study of patients with inflammatory bowel disease (IBD) [15]. It includes one cohort from the Prospective Registry in IBD Study at MGH (PRISM), which enrolled patients with a diagnosis of IBD based on endoscopic, radiographic, and histological evidence of either Crohn's Disease (CD) or Ulcerative Colitis (UC). This dataset has 121 IBD patients and 34 controls and is named as IBD (PRISM). Additionally, it includes an external validation dataset using two other cohorts. One consists of 20 healthy subjects who participated in LifeLines-DEEP, a general population-based study in the northern Netherlands (NLIBD) [59]. The second cohort consists of 43 subjects with IBD taken from the Department of Gastroenterology and Hepatology at the University Medical Center in Groningen, Netherlands. This dataset is named as IBD (External). The processing of the stool samples collected is described in the original study [15]. A total of 201 microbial species and 8848 metabolites were identified for the IBD (PRISM) and IBD (External) datasets.

The second dataset was taken from a study that collected 172 lung sputum samples from patients with cystic fibrosis [31]. Microbial features were generated using 16S rRNA gene sequencing and abundance was collected at the genus level, resulting in 657 unique microbial

features. Metabolomic data were generated using LC-MS/MS technology, resulting in 168 unique metabolites.

The third dataset represents microbial and metabolic activity caused by soil wetting at five-time points across four biocrust successional stages [32]. Biocrust soil water for each sample was analyzed by LC/MS for metabolite detection. Metagenomic shotgun sequencing was used to profile the microbial community and the authors used the 50S ribosomal protein L15 to map microbial taxa. A total of 466 microbes and 85 metabolites were detected.

Any input or output feature that is present in less than 10% of samples was removed. Microbiome and metabolomic data were then transformed using the centered log-ratio (CLR) transformation:

$$CLR(x) = \left[\log \frac{x_1}{g(x)}, \ldots, \log \frac{x_m}{g(x)} \right]$$

where x is the abundance vector of a sample, $g(x)$ is the sum of entry values in x, and m is the number of features. A pseudocount of 1 was added to each entry of x before the CLR transformation to prevent taking the log of 0 values. The only exception was for the IBD (PRISM) and IBD (External) microbe values, which were obtained in relative abundance (RA).

MiMeNet architecture and training protocol

An MLPNN model is composed of multiple fully connected hidden layers composed of perceptrons. The values h_{l+1} of layer l+1 are defined as

$$h_{l+1} = \varphi(h_l W_l + b_l)$$

Here W_l are the weights connecting the perceptrons of the lth layer with values h_l with those of layer l+1, b_l is the bias values between layer l and l+1, and φ is a non-linear function. In MiMeNet, φ is set as the rectified linear unit (ReLU). We selected this activation function since previous studies have shown that it is resilient to the problems of exploding and vanishing gradient in MLPNN training [60]. We used the L2 regularization in order to prevent the model from having large weights. Additional regularization was applied through dropout at each hidden layer, where a portion of the nodes and their weights are masked for a given epoch. MiMeNet was trained using the ADAM optimizer and the mean squared error (MSE) loss function. The entire loss function for MiMeNet is defined as,

$$Loss = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_l)^2 + \lambda \sum_{l \in L} \| W_l \|_2$$

Here, N is the number of training samples and L is the total number of hidden layers. The first term represents the mean squared error (MSE) of the observed metabolites y and the predicted metabolites $\hat{y}$. The second term represents the L2-regularization penalty which is controlled by the parameter λ.

The overall evaluation of MiMeNet prediction was conducted using 10 iterations of the 10-fold cross-validation, and the average of the correlations between the predicted and observed values for metabolites was reported. More explicitly, during the 10-fold cross-validation, each dataset was partitioned into two subsets: 90% for training and 10% for testing. For each training partition, the 90% of the data was further split into 80% for model training and 20% for validation. After finishing one iteration of 10-fold cross-validation, the SCC between the predicted and the observed was calculated for each metabolite. To prevent overfitting, MiMeNet models were trained using early stopping. After each iteration of updating network

Table 3. The optimal hyper-parameters for the neural network model in each dataset.

	Layer Size	Number of Layers	L2 Penalty (λ)	Dropout
IBD (PRISM)	512	1	0.001	0.5
Soil	128	2	0.005	0.3
Cystic Fibrosis	128	1	0.0001	0.5

weights using the 80% of the training set, the loss of the validation set was calculated. The training process was terminated when the loss of the validation set has not improved within 40 iterations, and the network weight parameters were set to the values of the best performing model on the validation set. Finally, the average of the SCC values was calculated after repeating the 10-fold cross-validation procedure for 10 times. For the IBD datasets, a final model trained on the full IBD (PRISM) dataset was then evaluated on the IBD (External) data set.

Hyper-parameter tuning was performed on the first training partition during cross-validation. To determine the optimal set of hyper-parameters (number of layers, layer size, λ, and dropout rate), we performed a cross-validated random search using a nested 5-fold cross-validation. We allowed for 1, 2, and 3 hidden layers of sizes 32, 128, and 512. The L2 regularization parameter (λ) was selected from 10 different values between 0.0001 and 0.1, evenly spaced on a log scale. Dropout was selected from 0.1, 0.3, and 0.5. The average SCC was calculated after a model was trained. We evaluated 20 sets of hyper-parameters and selected the best performing set for the rest of the 10-fold cross-validation. The optimal hyper-parameters are shown in Table 3.

Identifying well-predicted metabolites

To identify which metabolites are well predicted by MiMeNet, we generated a background distribution of SCC by training 100 models of 10-fold cross-validation where the samples in microbiomes and metabolomes were each randomly shuffled. From each of the 100 models generated, we collected the SCC for each metabolite and combined the values across all metabolites to construct the background distribution. We then defined a metabolite as well-predicted if its SCC is above the 95th percentile of the background correlations.

Calculation of microbe-metabolite feature attribution scores

Microbe-metabolite feature attribution scores are calculated using Olden's method for understanding variable contributions in neural network models [61]. Olden's method generates these scores by multiplying the weights of each hidden layer together, resulting in a single matrix where each row represents an input feature, and each column represents an output feature. More specifically, for each trained network model, a microbe-metabolite feature attribution score matrix S is calculated as

$$S = \prod_{l \in L} W_l$$

where l is the current layer in the set of L layers, and W_l is the weight matrix connecting layer l −1 and layer l. Each element in S represents a microbe-metabolite feature attribution score. A positive value indicates that an increase of the microbe will lead to an increase of the metabolite and a negative value indicates that an increase of the microbe will lead to a decrease of the metabolite. For the subsequent procedure, we only retained the columns of the feature attribution matrices representing the well-predicted metabolites.

Identifying microbes with significant microbe-metabolite interactions

Denoting S_i as the feature attribution score matrix for the ith trained model ($n = 100$ models resulted from the 10 iterations of 10-fold cross-validation), we calculated the mean feature attribution matrix $\bar{S}$ as

$$\bar{S} = \frac{1}{n}\sum_{i=1}^{n} S_i$$

To identify microbes with significant associations, we further calculated feature attribution score matrices from the network models used to generate the background correlation distributions and calculated the mean feature attribution score matrix, which was then flattened into a vector and a threshold was set at the 97.5 percentile. Any feature attribution score in the observed dataset with an absolute value above the threshold was considered significant. Finally, any microbe with at least one significant feature attribution score with any metabolite was considered to be significant and the rows representing non-significant microbes were filtered out from $\bar{S}$ as well as from all feature attribution score matrices S_i used in subsequent analyses.

Clustering and visualization of microbe-metabolite interactions

We normalized the values in each feature attribution score matrix S_i by dividing the significant threshold score identified from the background and clipped values to be between -1 and 1. In doing so, every significant attribution score was treated with equal magnitude. We recalculated $\bar{S}$ using the normalized S_i so that each element in $\bar{S}$ is also between -1 and 1. The normalized matrix S_i was then used to cluster microbes (rows) and metabolites (columns) separately based on the Euclidean distance and complete linkage using Seaborn's *clustermap* function in Python. Modules were constructed by cutting each dendrogram at a given height. To determine the number of clusters for microbes, for each fixed k, ranged from 2 to 20, a k-clustering of the rows using each normalized S_i was generated. Then a consensus matrix $M^{(k)}$ was calculated as the mean connectivity matrix across all k-clustering results ($n = 100$),

$$M^{(k)} = \frac{1}{n}\sum_{i=1}^{n} C_i^{(k)}$$

$$C_i^{(k)}(a,b) = 1 \; \textit{if features a and b are in the same cluster}$$

$$C_i^{(k)}(a,b) = 0 \; \textit{if features a and b are not in the same cluster}$$

where $C_i^{(k)}$ is the connectivity matrix of the clustering using k clusters on S_i. We further calculated the area under the cumulative distribution function (CDF) for the consensus matrix of each clustering,

$$A^{(k)} = \sum_{j=1}^{m} (x_j - x_{j-1})\varphi(x_j)$$

where x_j is the jth value from the set {0.01, 0.02, 0.03,. . .,0.99, 1.0} and $\varphi(x_j)$ is the proportion of entries in consensus matrix $M^{(k)}$ that are less than x_j. Lastly, we calculated the proportional change in area as the number of clusters changed,

$$\Delta k = \frac{A^{(k)} - A^{(k-1)}}{A^{(k-1)}}$$

This value represents how much cleaner the consensus matrix gets if we increase the number of clusters by 1. We set a threshold of $\Delta k = 0.025$, indicating that increasing the cluster by 1 more would give less than a 2.5% increase in the area under the CDF. The best cluster number k^* was selected as the largest value of k that resulted in Δk larger than the threshold. Further details of this analysis can be found in *Monte et al.* [62] and an example demonstrating the changes of $A^{(k)}$ is shown in S9 Fig. The number of metabolite clusters was determined using the same procedure. The best cluster numbers for microbes and metabolites are then denoted as k_1^* and k_2^*, respectively. The final set of microbial and metabolite modules are then determined by biclustering $\bar{S}$ using k_1^* and k_2^* to cluster the rows and columns respectively.

For visualization of the microbe-metabolite interaction network, the score between a pair of microbe and metabolite modules was calculated as the average normalized feature attribution score in $\bar{S}$ between each microbe and metabolite within the two modules. For visualization purposes only we removed any score whose absolute value was less than 0.25. Networks showing microbe and metabolite modules and the interactions between them were constructed using Cytoscape [63].

Determination of the final model for external metabolite prediction

For evaluating the external test data, multiple neural network models were trained on the entire training dataset. For each model, biclustering is performed on the extracted feature attribution score matrix using k_1^* and k_2^* as the number of microbial and metabolite clusters respectively, and microbe and metabolite match scores are assigned as the number of co-clustered microbes or metabolites that also occur in the previously identified modules. The model with the highest combined match scores was selected as the final model for external evaluation.

Machine learning models for benchmarking

MelonnPan and NED models were obtained from their respective GitHub repositories and executed using default parameters as according to their tutorials. Random Forest, multivariate Elastic Net, and Canonical correlation analysis models were implemented using Python's *scikit-learn* package. Similar to MiMeNet, these models can predict the entire set of metabolites at once, and all models were evaluated using 10 iterations of 10-fold cross-validation. Random Forest models were implemented using *RandomForestRegressor* with the default parameter settings of 100 tree estimators. Multivariate Elastic Net models were implemented using *ElasticNet* and *GridSearchCV* using 5-fold internal cross-validation for hyper-parameter tuning where the hyper-parameter grid contained $\alpha \in \{0.0001, 0.001, 0.01, 0.1\}$ and *l1 ratio* $\in \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$. To keep computational time in line with MiMeNet, the maximum iterations was set to 10. Canonical correlation analyses were implemented using *CCA* with 10, 20, and 40 components.

Weighted correlation network analysis (WGCNA)

WGCNA analysis of microbial features was performed using the *WGCNA* library in R [64] on each dataset. The adjacency matrix was created with a power of 2 for the microbiome data and 5 for the metabolomic data. To compare the WGCNA microbial modules to the MiMeNet microbial modules derived from the IBD (PRISM) dataset, we calculated the Jaccard similarity between the modules as well as the Spearman correlation.

Neural networks for host phenotype prediction

For each module derived from MiMeNet or WGCNA using the IBD (PRISM) dataset, the module feature value was calculated as the mean normalized abundance value of the features comprising the module for each subject in both IBD datasets. Network models for the IBD status prediction were trained on the IBD (PRISM) dataset using the original microbial OTUs (using either relative or center log-transformed abundance), the module feature values of the MiMeNet and WGCNA microbial modules, respectively. Similarly, the network models were trained on the IBD (PRISIM) data using the normalized metabolomic abundance and metabolomic module feature values. We trained a 3-layer neural network model with 32 nodes in each layer for each input format described above. The evaluation was done by training 100 neural network models evaluating on the IBD (External) test set using the area under the receiver operating characteristic curve (AUC).

Other existing tools

The MelonnPan was downloaded from https://github.com/biobakery/melonnpan and executed using the given instructions. The NED model was trained using code downloaded from https://github.com/vuongle2/BiomeNED.

Supporting information

S1 Fig. Comparison of prediction correlation when using relative abundance and centered log-ratio. Scatterplots comparing metabolite correlation prediction between data transformed to relative abundance (RA) and centered log-ratio (CLR) for (A) IBD (PRISM), (B) cystic fibrosis, (C) soil datasets using 10 iterations of 10-fold cross-validation, and (D) IBD (External) test predictions using models trained on the full IBD (PRISM) dataset.
(TIF)

S2 Fig. Performance comparison of models trained using shared hyper-parameters against models trained with tuning hyper-parameters every cross-validated partition. Using 10-iterations of 10-fold cross-validation, evaluations using shared hyper-parameters tuned from the first partition (Tune Once) were compared against evaluations with tuning for each partition (Tune Every Partition) for the IBD (PRISM) and cystic fibrosis dataset. Each point represents the mean SCC of a metabolite and the red lines represent the determined SCC threshold for significantly well-predicted metabolties.
(TIF)

S3 Fig. Prediction correlation comparison between models of MiMeNet and MelonnPan trained only on the annotated metabolites. Scatterplot of mean predicted Spearman's correlations over 10 iterations of the 10-fold cross-validation for each metabolite between MiMeNet and MelonnPan when MiMeNet is trained only on the annotated metabolites in the IBD (PRISM) dataset.
(TIF)

S4 Fig. Overlap of well-predicted metabolites identified by MiMeNet and MelonnPan. Using the correlation cutoff identified by MiMeNet, the overlap between the well-predicted metabolites is shown between MiMeNet and MelonnPan for (A) IBD (PRISM) dataset, (B) cystic fibrosis dataset, (C) and soil dataset. (D) In addition, the overlap of well-predicted metabolites is shown when training on the entire IBD (IBD) dataset and predicting on the IBD (External) dataset.
(TIF)

S5 Fig. Mean Spearman and Pearson correlation per metabolite module. For each metabolite module, the (A) mean SCC and (B) mean PCC values of the members within the module are shown using the cross-validated evaluation on IBD (PRISM) as well as when evaluating the IBD (External) data.
(TIF)

S6 Fig. IBD status of prediction for different module construction methods evaluated on the IBD (External) dataset. Module values were constructed using WGCNA and MiMeNet. For metabolomic modules constructed by MiMeNet, the values within the parentheses represent the compositional transformation, number of folds for cross-validation, and the aggregation method respectively. Mean aggregation calculates the mean normalized abundance value. PCA aggregation uses the first principal component of the members' values from that module. Microbiome modules were constructed similarly with the exception that RA was always used.
(TIF)

S7 Fig. Overlap of significant metabolites and module membership when using relative abundance and centered log-ratio. (A) Overlap of well-predicted metabolites when using relative abundance and centered log-ratio for the IBD (PRISM) dataset. Heatmaps show Jaccard similarity between membership of (B) metabolite and (C) microbial modules in the IBD (PRISM) dataset. (D) Overlap of well-predicted metabolites when using relative abundance and centered log-ratio for the cystic fibrosis dataset. Heatmaps show Jaccard similarity between membership of (E) metabolite and (F) microbial modules in the cystic fibrosis dataset.
(TIF)

S8 Fig. Comparison of MiMeNet to NED. Scatterplot comparing metabolite prediction correlation between MiMeNet and NED on the IBD (External) dataset validation. The red line indicates the correlation threshold identified by MiMeNet and the gray area represents well-predicted metabolites. The one-tailed p-value (Wilcoxon sign-rank) comparing MiMeNet's values to NED's is shown in the top right.
(TIF)

S9 Fig. Using consensus clustering analysis for cluster number determination for the IBD (PRISM) dataset. (A) The cumulative distribution functions (CDF) for varying cluster numbers and (B) change in area under the CDF is shown for clustering on the microbial features. (C) The cumulative distribution functions (CDF) for varying cluster numbers, and (D) change in area under the CDF is shown for clustering on the metabolic features.
(TIF)

S1 Table. Cross-validated evaluation using centered log-ratio data. Benchmarking results using MiMeNet, Random Forest (RF), Multivariate Elastic Net, and Canonical correlation analysis (CCA) of the three datasets using 10 iterations of 10-fold cross-validation. Mean and standard deviation values for Spearman correlation coefficient (SCC), Pearson correlation coefficient (PCC), and mean absolute error (MAE) are shown. All data were transformed using centered log-ratio exception for IBD microbial input, which was obtained in relative abundance.
(XLSX)

S2 Table. Cross-validated evaluation relative abundance data. Benchmarking results using MiMeNet, RF, Multivariate Elastic Net, and CCA of the three datasets using 10 iterations of 10-fold cross-validation. Mean and standard deviation values for SCC, PCC, and MAE are

shown. All data are transformed using relative abundance.
(XLSX)

S3 Table. Evaluation of the IBD (External) data. Benchmarking results using MiMeNet, Random Forest (RF), Multivariate Elastic Net, and CCA of the three datasets using the IBD (PRISM) data to predict the IBD (External) data. Values for SCC, PCC, and MAE are shown.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Derek Reiman, Yang Dai.

Data curation: Derek Reiman.

Formal analysis: Derek Reiman, Brian T. Layden, Yang Dai.

Investigation: Derek Reiman, Yang Dai.

Methodology: Derek Reiman.

Project administration: Yang Dai.

Resources: Yang Dai.

Software: Derek Reiman.

Supervision: Yang Dai.

Writing – original draft: Derek Reiman, Yang Dai.

Writing – review & editing: Derek Reiman, Brian T. Layden, Yang Dai.

References

1. Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI. Human nutrition, the gut microbiome and the immune system. Nature. 2011; 474(7351):327–36. https://doi.org/10.1038/nature10213 PMID: 21677749

2. Wang J, Jia H. Metagenome-wide association studies: fine-mining the microbiome. Nature Reviews Microbiology. 2016; 14(8):508–22. https://doi.org/10.1038/nrmicro.2016.83 PMID: 27396567

3. Ghaisas S, Maher J, Kanthasamy A. Gut microbiome in health and disease: Linking the microbiome–gut–brain axis and environmental factors in the pathogenesis of systemic and neurodegenerative diseases. Pharmacology & therapeutics. 2016; 158:52–62. https://doi.org/10.1016/j.pharmthera.2015.11.012 PMID: 26627987

4. Brown CT, Davis-Richardson AG, Giongo A, Gano KA, Crabb DB, Mukherjee N, et al. Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. PloS one. 2011; 6(10).r https://doi.org/10.1371/journal.pone.0025792 PMID: 22043294

5. Tilg H, Kaser A. Gut microbiome, obesity, and metabolic dysfunction. The Journal of clinical investigation. 2011; 121(6):2126–32. https://doi.org/10.1172/JCI58109 PMID: 21633181

6. Kho ZY, Lal SK. The Human Gut Microbiome–A Potential Controller of Wellness and Disease. Frontiers in Microbiology. 2018; 9(1835). https://doi.org/10.3389/fmicb.2018.01835 PMID: 30154767

7. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome biology. 2012; 13(9):R79. Epub 2012/09/28. https://doi.org/10.1186/gb-2012-13-9-r79 PMID: 23013615; PubMed Central PMCID: PMC3506950.

8. McHardy IH, Goudarzi M, Tong M, Ruegger PM, Schwager E, Weger JR, et al. Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. Microbiome. 2013; 1(1):17. https://doi.org/10.1186/2049-2618-1-17 PMID: 24450808

9. Feng Q, Liu Z, Zhong S, Li R, Xia H, Jie Z, et al. Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. Scientific reports. 2016; 6(1):1–14. https://doi.org/10.1038/s41598-016-0001-8 PMID: 28442746

10. Parker A, Lawson MA, Vaux L, Pin C. Host-microbe interaction in the gastrointestinal tract. Environmental microbiology. 2018; 20(7):2337–53. https://doi.org/10.1111/1462-2920.13926 PMID: 28892253

11. Barton W, Penney NC, Cronin O, Garcia-Perez I, Molloy MG, Holmes E, et al. The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level. Gut. 2018; 67(4):625–33. https://doi.org/10.1136/gutjnl-2016-313627 PMID: 28360096

12. Cui X, Ye L, Li J, Jin L, Wang W, Li S, et al. Metagenomic and metabolomic analyses unveil dysbiosis of gut microbiota in chronic heart failure patients. Scientific reports. 2018; 8(1):1–15. https://doi.org/10.1038/s41598-017-17765-5 PMID: 29311619

13. Martin AM, Sun EW, Rogers GB, Keating DJ. The Influence of the Gut Microbiome on Host Metabolism Through the Regulation of Gut Hormone Release. Frontiers in Physiology. 2019; 10(428). https://doi.org/10.3389/fphys.2019.00428 PMID: 31057420

14. Zierer J, Jackson MA, Kastenmüller G, Mangino M, Long T, Telenti A, et al. The fecal metabolome as a functional readout of the gut microbiome. Nature genetics. 2018; 50(6):790–5. https://doi.org/10.1038/s41588-018-0135-7 PMID: 29808030

15. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat Microbiol. 2019; 4(2):293–305. Epub 2018/12/10. https://doi.org/10.1038/s41564-018-0306-4 PMID: 30531976.

16. Benson AK. The gut microbiome-an emerging complex trait. Nat Genet. 2016; 48(11):1301–2. Epub 2016/10/28. https://doi.org/10.1038/ng.3707 PMID: 27787511.

17. Lee-Sarwar KA, Lasky-Su J, Kelly RS, Litonjua AA, Weiss ST. Metabolome–Microbiome Crosstalk and Human Disease. Metabolites. 2020; 10(5). https://doi.org/10.3390/metabo10050181 PMID: 32369899

18. Bassett SA, Young W, Fraser K, Dalziel JE, Webster J, Ryan L, et al. Metabolome and microbiome profiling of a stress-sensitive rat model of gut-brain axis dysfunction. Scientific Reports. 2019; 9(1):14026. https://doi.org/10.1038/s41598-019-50593-3 PMID: 31575902

19. Visconti A, Le Roy CI, Rosa F, Rossi N, Martin TC, Mohney RP, et al. Interplay between the human gut microbiome and host metabolism. Nature Communications. 2019; 10(1):4505. https://doi.org/10.1038/s41467-019-12476-z PMID: 31582752

20. Shreiner AB, Kao JY, Young VB. The gut microbiome in health and in disease. Current opinion in gastroenterology. 2015; 31(1):69. https://doi.org/10.1097/MOG.0000000000000139 PMID: 25394236

21. Skelly AN, Sato Y, Kearney S, Honda K. Mining the microbiota for microbial and metabolite-based immunotherapies. Nature Reviews Immunology. 2019; 19(5):305–23. https://doi.org/10.1038/s41577-019-0144-5 PMID: 30858494

22. Larsen PE, Collart FR, Field D, Meyer F, Keegan KP, Henry CS, et al. Predicted Relative Metabolomic Turnover (PRMT): determining metabolic turnover from a coastal marine metagenomic dataset. Microbial informatics and experimentation. 2011; 1(1):4. https://doi.org/10.1186/2042-5783-1-4 PMID: 22587810

23. Larsen PE, Dai Y. Metabolome of human gut microbiome is predictive of host dysbiosis. Gigascience. 2015; 4(1):s13742-015-0084-3. https://doi.org/10.1186/s13742-014-0042-5 PMID: 25653850

24. Biggs MB, Medlock GL, Kolling GL, Papin JA. Metabolic network modeling of microbial communities. Wiley Interdisciplinary Reviews: Systems Biology and Medicine. 2015; 7(5):317–34. https://doi.org/10.1002/wsbm.1308 PMID: 26109480

25. Gottstein W, Olivier BG, Bruggeman FJ, Teusink B. Constraint-based stoichiometric modelling from single organisms to microbial communities. J R Soc Interface. 2016; 13(124):20160627. https://doi.org/10.1098/rsif.2016.0627 PMID: 28334697.

26. Mallick H, Franzosa EA, McLver LJ, Banerjee S, Sirota-Madi A, Kostic AD, et al. Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. Nature Communications. 2019; 10(1):3136. https://doi.org/10.1038/s41467-019-10927-1 PMID: 31316056

27. Morton JT, Aksenov AA, Nothias LF, Foulds JR, Quinn RA, Badri MH, et al. Learning representations of microbe–metabolite interactions. Nature Methods. 2019; 16(12):1306–14. https://doi.org/10.1038/s41592-019-0616-3 PMID: 31686038

47. Abdel Hadi L, Di Vito C, Riboni L. Fostering Inflammatory Bowel Disease: Sphingolipid Strategies to Join Forces. Mediators of Inflammation. 2016; 2016:1–13. https://doi.org/10.1155/2016/3827684 PMID: 26880864

48. Parada Venegas D, De la Fuente MK, Landskron G, González MJ, Quera R, Dijkstra G, et al. Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. Front Immunol. 2019; 10(277). https://doi.org/10.3389/fimmu.2019.00277 PMID: 30915065

49. Louis P, Flint HJ. Formation of propionate and butyrate by the human colonic microbiota. Environmental microbiology. 2017; 19(1):29–41. https://doi.org/10.1111/1462-2920.13589 PMID: 27928878

50. Gérard P. Metabolism of cholesterol and bile acids by the gut microbiota. Pathogens. 2013; 3(1):14–24. https://doi.org/10.3390/pathogens3010014 PMID: 25437605.

51. Erbilgin O, Rübel O, Louie KB, Trinh M, Raad Md, Wildish T, et al. MAGI: A Method for Metabolite Annotation and Gene Integration. ACS chemical biology. 2019; 14(4):704–14. https://doi.org/10.1021/acschembio.8b01107 PMID: 30896917

52. Peisl BYL, Schymanski EL, Wilmes P. Dark matter in host-microbiome metabolomics: Tackling the unknowns–A review. Analytica Chimica Acta. 2018; 1037:13–27. https://doi.org/10.1016/j.aca.2017.12.034 PMID: 30292286

53. Jeffryes JG, Colastani RL, Elbadawi-Sidhu M, Kind T, Niehaus TD, Broadbelt LJ, et al. MINEs: open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics. J Cheminform. 2015; 7:44. Epub 2015/09/01. https://doi.org/10.1186/s13321-015-0087-1 PMID: 26322134; PubMed Central PMCID: PMC4550642.

54. Pedersen HK, Forslund SK, Gudmundsdottir V, Petersen AØ, Hildebrand F, Hyötyläinen T, et al. A computational framework to integrate high-throughput '-omics' datasets for the identification of potential mechanistic links. Nature Protocols. 2018; 13(12):2781–800. https://doi.org/10.1038/s41596-018-0064-z PMID: 30382244

55. Franzosa EA, Hsu T, Sirota-Madi A, Shafquat A, Abu-Ali G, Morgan XC, et al. Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. Nature reviews Microbiology. 2015; 13 (6):360–72. Epub 2015/04/29. https://doi.org/10.1038/nrmicro3451 PMID: 25915636; PubMed Central PMCID: PMC4800835.

56. Chong J, Xia J. Computational Approaches for Integrative Analysis of the Metabolome and Microbiome. Metabolites. 2017; 7(4):62. https://doi.org/10.3390/metabo7040062 PMID: 29156542.

57. Chatterjee I, Lu R, Zhang Y, Zhang J, Dai Y, Xia Y, et al. Vitamin D receptor promotes healthy microbial metabolites and microbiome. Scientific Reports. 2020; 10(1):7340. https://doi.org/10.1038/s41598-020-64226-7 PMID: 32355205

58. Ni Y, Yu G, Chen H, Deng Y, Wells PM, Steves CJ, et al. M2IA: a web server for microbiome and metabolome integrative analysis. Bioinformatics. 2020; 36(11):3493–8. https://doi.org/10.1093/bioinformatics/btaa188 PMID: 32176258

59. Tigchelaar EF, Zhernakova A, Dekens JAM, Hermes G, Baranska A, Mujagic Z, et al. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. BMJ Open. 2015; 5(8):e006772–e. https://doi.org/10.1136/bmjopen-2014-006772 PMID: 26319774.

60. Hara K, Saito D, Shouno H, editors. Analysis of function of rectified linear unit used in deep learning. 2015 International Joint Conference on Neural Networks (IJCNN); 2015: IEEE.

61. Olden J, Joy M, Death R. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecological Modelling. 2004; 178:389–97. https://doi.org/10.1016/j.ecolmodel.2004.03.013

62. Monti S, Tamayo P, Mesirov J, Golub T. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Machine Learning. 2003; 52:91–118. https://doi.org/10.1023/A:1023949509487

63. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13(11):2498–504. https://doi.org/10.1101/gr.1239303 PMID: 14597658.

64. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9(1):559. https://doi.org/10.1186/1471-2105-9-559 PMID: 19114008