Supplementary Information: Hierarchical Sparse Bayesian Multitask Learning for Disease Prediction in Pooled Microbiome Studies

Haonan Zhu[1,2], Andre R. Goncalves[1], Car Reen Kok[1], Camilo Valdes[1], Hiranmayi Ranganathan[1], Boya Zhang[1], Jose Manuel Martí[1], Monica K. Borucki[1], Nisha J. Mulakken[1], James B. Thissen[1], Crystal Jaing[1], Alfred Hero[2], Nicholas A. Be[1*]

[1]Lawrence Livermore National Laboratory, Livermore, CA, USA.
[2]Electrical and Computer Engineering, University of Michigan, Ann Arbor, MI, USA.

*Corresponding author(s). E-mail(s): be1@llnl.gov;
Contributing authors: haonan@umich.edu; goncalves1@llnl.gov;

1 Dataset References

Table 1 lists all studies used in this work along with their metadata, including sample sizes, disease categories, and sequencing details. Some studies were obtained from the curatedMetagenomicData R package [1], which provides uniformly processed and annotated metagenomic datasets. All datasets used in this analysis are publicly available at https://zenodo.org/records/17315984.

Table 1: Summary of microbiome studies by disease category. Two studies cover multiple conditions: [2] includes age-related macular degeneration and cardiovascular disease, while [3] covers six diseases — diabetes, neurological disease, irritable bowel syndrome, inflammatory bowel disease, gastrointestinal infection, and diarrhea. Both are split into separate entries by disease category for our analysis.

Disease Category	**Ref.**	**Host Body Site**	**Geographic Location**	**Control and Disease Splits (C/D)**
Diabetes	[4]	Fecal	Europe	43/102
	[5]	Fecal	United States	17/19
	[6]	Fecal	Luxembourg	26/27
	[3]	Fecal	United States	359/19
Cirrhosis	[7]	Fecal	China	113/110
	[8]	Fecal	United States	77/9
Cancer	[9]	Fecal	Austria	61/93
	[10]	Fecal	China	53/75
	[11]	Fecal	North America	21/36
	[12]	Fecal	Japan	250/365
	[13]	Fecal, Oral	Not Available	470/90
Neurological Disease	[14]	Oral	United States	16/16
	[15]	Fecal	United States	50/50
	[16]	Fecal	China	40/40
	[17]	Fecal	China	81/90
	[18]	Fecal, Oral	South Korea	74/81
	[19]	Fecal	Germany	100/75
	[20]	Fecal	United States	233/491
	[3]	Fecal	United States	359/253
Diarrhea	[21]	Fecal, Rectal Swab	Bangladesh	7/23
	[22]	Fecal	Bangladesh	9/18
	[3]	Fecal	United States	359/60
Dermatologic Disease	[23]	Skin	Asia	40/38
	[24]	Skin	United States	53/238
	[25]	Skin	Singapore	50/69
	[26]	Fecal	United States	15/33
Gastrointestinal Infection	[27]	Fecal	Canada	182/27
	[28]	Fecal	Indonesia, Liberia	5/19
	[29]	Fecal	Cameroon	86/89
	[3]	Fecal	United States	359/5
Inflammatory Bowel	[30]	Fecal	United States	71/175
	[31]	Fecal	Not Available	56/162
	[32]	Fecal	Not Available	354/888

Disease Continued on next page

Table 1 – continued from previous page

Disease Category	Ref.	Host Body Site	Geographic Location	Control and Disease Splits (C/D)
	[33]	Fecal	United States	27/8
	[3]	Fecal	United States	359/9
Cardiovascular Disease	[34]	Fecal	China	171/214
	[35]	Fecal	China	41/155
	[36]	Fecal	China	56/60
	[37]	Fecal	United States	27/127
	[38]	Fecal	China	10/10
	[2]	Fecal	Switzerland	30/29
Oral Disease	[39]	Oral	Australia	37/48
	[40]	Oral	Italy	51/23
	[41]	Oral	Hungary	8/19
Autoimmune Disease	[42]	Fecal	China	45/20
	[43]	Fecal	China	62/100

Continued on next page

Table 1 – continued from previous page

Disease Category	Ref.	Host Body Site	Geographic Location	Control and Disease Splits (C/D)
Immune Disease	[44]	Fecal	Spain	27/129
Pulmonary Disease	[45]	Fecal	China	31/46
	[46]	Cough Swabs	Germany	48/49
	[47]	Fecal	China	15/55
	[48]	Nasal	Sweden	20/37
	[49]	Fecal	China	63/47
	[50]	Lung (Bronchioalveolar Lavage Fluid)	China	16/45
	[51]	Fecal	China	69/138
	[52]	Fecal	China	21/20
Metabolic Disease	[53]	Fecal	Mexico	10/10
Hormonal Disorder	[54]	Fecal	China	43/50
Genitourinary Disease	[55]	Endocervical, Vaginal, Rectal	Fiji	11/14
Seafaring Syndrome	[56]	Fecal	Not Available	99/55
Irritable Bowel Syndrome	[3]	Fecal	United States	359/52
Age-Related Macular Degeneration	[2]	Fecal	Switzerland	33/57

2 CAVI update derivation

This section includes the derivations of CAVI updates for Algorithm I.

Recall Equation 18 of [57] states that if we are to approximate a general posterior distribution $p(\boldsymbol{\xi} \mid \text{data})$ with a mean-field approximation $q(\boldsymbol{\xi}) := \prod_j q_j(\xi_j)$, the CAVI update for j-th latent variable ξ_j (i.e., the optimal solution $q_j^\star(\xi_j)$) is proportional to the exponentiated conditional expected log of the joint:

$$q_j^\star \propto \exp\left(\mathbb{E}_{\boldsymbol{\xi}_{-j} \sim q_{-j}}\left[\log\left(p\left(\xi_j, \boldsymbol{\xi}_{-j} \mid \text{data}\right)\right)\right]\right). \tag{1}$$

where $\boldsymbol{\xi}_{-j}$ corresponds to all but the j-th latent variable.

Update for α, β.

Based on Eqn. 1, the exponentiated conditional expectation of all the parameters except θ up to a constant scaling factor:

$$\begin{aligned} q^{\star}(\theta) &\propto \exp\left((\alpha_0 - 1)\log\theta + (\beta_0 - 1)\log(1-\theta)\right) \\ &\quad \exp\left(\left(\sum_j \phi_j\right)\log\theta + \left(d - \sum_j \phi_j\right)\log(1-\theta)\right) \\ &= \exp\left(\left(\alpha_0 - 1 + \sum_j \phi_j\right)\log\theta\right) \\ &\quad \exp\left(\left(\beta_0 + d - \sum_j \phi_j - 1\right)\log(1-\theta)\right). \end{aligned}$$

This implies $q^{\star}(\theta)$ follows a beta distribution with parameters:

$$\begin{aligned} \alpha &= \alpha_0 + \sum_j \phi_j, \\ \beta &= \beta_0 + d - \sum_j \phi_j. \end{aligned}$$

Update for v and $\boldsymbol{V}$. Based on Eqn. 1, the exponentiated conditional expectation of all the parameters except $\boldsymbol{\Sigma}_0^{-1}$ up to a constant scaling factor:

$$\begin{aligned} q^{\star}\left(\boldsymbol{\Sigma}_0^{-1}\right) &\propto \exp\left(-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}_0^{-1}\boldsymbol{\Sigma}_0^{-1}\right)\right) \\ &\quad \exp\left(-\frac{v_0 + d - T - 1}{2}\log\det\left(\boldsymbol{\Sigma}_0^{-1}\right)\right) \\ &\quad \exp\left(-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_0^{-1}\left(\sum_j \boldsymbol{m}_{(j)}\boldsymbol{m}_{(j)}^{\top} + \boldsymbol{\Sigma}_j\right)\right)\right). \end{aligned}$$

This implies $q^{\star}\left(\boldsymbol{\Sigma}_0^{-1}\right)$ follow a Wishart distribution with parameters:

$$\begin{aligned} v &= v_0 + d, \\ \boldsymbol{V} &= \left(\boldsymbol{V}_0^{-1} + \sum_j \boldsymbol{m}_{(j)}\boldsymbol{m}_{(j)}^{\top} + \boldsymbol{\Sigma}_j\right)^{-1}. \end{aligned}$$

Update for $\boldsymbol{\Sigma}_j$. all the terms involve $\boldsymbol{\Sigma}_j$ in ELBO approximation (Eqn. 3, section III.A):

$$-\frac{1}{2}\operatorname{tr}\left(v\boldsymbol{V}\boldsymbol{\Sigma}_j\right)-\frac{1}{8}\sum_t\left(\boldsymbol{\Sigma}_j\right)_{t,t}\phi_j\sum_i\left(x_t^{ij}\right)^2+\frac{1}{2}\log\det\left(\boldsymbol{\Sigma}_j\right).$$

Rewrite the second term:

$$-\frac{1}{8}\operatorname{tr}\left(\boldsymbol{\Sigma}_j\operatorname{diag}\left(\left[\textstyle\sum_i\phi_j\left(x_1^{ij}\right)^2,\cdots,\sum_i\phi_j\left(x_T^{ij}\right)^2\right]^\top\right)\right).$$

Denote the diagonal matrix as $\tilde{\boldsymbol{X}}_j$. For every j, we have a constrained optimization problem:

$$\max_{\boldsymbol{\Sigma}\in\mathbb{S}_{++}^T}\log\det\left(\boldsymbol{\Sigma}\right)-\operatorname{tr}\left(\boldsymbol{\Sigma}\left(v\boldsymbol{V}+\frac{1}{4}\tilde{\boldsymbol{X}}_j\right)\right).$$

which admits a closed form solution:

$$\boldsymbol{\Sigma}_j^\star=\left(v\boldsymbol{V}+\frac{1}{4}\tilde{\boldsymbol{X}}_j\right)^{-1}. \tag{2}$$

for $j=1,\ldots,d$.

Update for $\boldsymbol{m}_{(j)}$. all the terms involved $\boldsymbol{m}_{(j)}$ in ELBO approximation (Eqn. 3, section III.A):

$$\begin{aligned}
&-\frac{v}{2}\left\langle\boldsymbol{m}_{(j)},\boldsymbol{V}\boldsymbol{m}_{(j)}\right\rangle-\frac{1}{8}\left\langle\boldsymbol{m}_{(j)},\tilde{\boldsymbol{X}}_j\boldsymbol{m}_{(j)}\right\rangle\\
&+\left\langle\boldsymbol{m}_{(j)},\left[\cdots,\ \phi_j\textstyle\sum_i\left(y_t^i-\tilde{y_t^i}\right)x_t^{ij},\ \cdots\right]^\top\right\rangle\\
&+\frac{1}{4}\left\langle\boldsymbol{m}_{(j)},\left[\cdots,\ \phi_j\textstyle\sum_i\left\langle\boldsymbol{w}_t'\circ\boldsymbol{z}',\boldsymbol{x}_t^i\right\rangle x_t^{ij},\ \cdots\right]^\top\right\rangle\\
&-\frac{1}{4}\left\langle\boldsymbol{m}_{(j)},\left[\cdots,\ \phi_j\textstyle\sum_i x_t^{ij}\sum_{l\neq j}\phi_l x_t^{il}m_{tl},\ \cdots\right]^\top\right\rangle.
\end{aligned}$$

This problem is quadratic with a negative definite Hessian matrix, hence by stationary condition (i.e., zero gradient) we have closed form updates:

$$\begin{aligned}
\boldsymbol{\Sigma}_j^\star(&\left[\cdots\ \phi_j\textstyle\sum_i\left(y_t^i-\tilde{y_t^i}\right)x_t^{ij},\ \cdots\right]^\top\\
&+\frac{1}{4}\left[\cdots,\ \phi_j\textstyle\sum_i\left\langle\boldsymbol{w}_t'\circ\boldsymbol{z}',\boldsymbol{x}_t^i\right\rangle x_t^{ij},\ \cdots\right]^\top\\
&-\frac{1}{4}\left[\cdots,\ \phi_j\textstyle\sum_i x_t^{ij}\sum_{l\neq j}\phi_l x_t^{il}m_{tl},\ \cdots\right]^\top).
\end{aligned} \tag{3}$$

For all $j = 1 \ldots, d$. When the reference point of quadratic lower bound $\boldsymbol{w}' \circ \boldsymbol{z}'$ is set to be the mean parameters from the previous iteration, we can simplify Eqn. 4:

$$\begin{aligned} \boldsymbol{\Sigma}_j^{\star}(&\left[\phi_j \sum_i \left(y_1^i - \tilde{y_1^i}\right) x_1^{ij}, \cdots, \phi_j \sum_i \left(y_T^i - \tilde{y_T^i}\right) x_T^{ij}\right]^{\top} \\ &+ \frac{1}{4}\left[\phi_j^2 \sum_i \left(x_1^{ij}\right)^2 m_{1j}^{(k)}, \cdots, \phi_j^2 \sum_i \left(x_T^{ij}\right)^2 m_{Tj}^{(k)}\right]^{\top}. \end{aligned} \tag{4}$$

Update for ϕ_j.

All the terms involve ϕ_j in ELBO approximation (Eqn. 3, section III.A):

$$\begin{aligned} f(\phi_j) := \phi_j(\psi(\alpha) - \psi(\beta)) &+ \phi_j \sum_t \sum_i \left(y_t^i - \tilde{y_t^i}\right) m_{tj} x_t^{ij} \\ &+ \frac{\phi_j}{4} \sum_t \sum_i \left(m_{tj} x_t^{ij}\right) \langle \boldsymbol{w}_t' \circ \boldsymbol{z}', \boldsymbol{x}_t^i \rangle \\ &- \frac{\phi_j}{4} \sum_t \sum_i m_{tj} x_t^{ij} \sum_{l \neq j} m_{tl} \phi_l x_t^{il} \\ &- \frac{\phi_j}{8} \sum_t \left((\Sigma_j)_{tt} + m_{tj}^2\right) \sum_i \left(x_t^{ij}\right)^2 \\ &- \phi_j \log(\phi_j) - (1 - \phi_j) \log(1 - \phi_j). \end{aligned}$$

Observe $f(\phi_j)$ is a smooth strictly concave function, so we can solve for $\phi_j^{\star}$ by stationary condition (i.e., 0 derivatives), which admit a closed form update:

$$\phi_j^{\star} = \sigma(a). \tag{5}$$

where $\sigma(t) := \frac{1}{\exp(-t)+1}$ denote the sigmoid function, and:

$$\begin{aligned} a = \psi(\alpha) - \psi(\beta) &+ \sum_t \sum_i \left(y_t^i - \tilde{y_t^i}\right) m_{tj} x_t^{ij} \\ &+ \frac{1}{4} \sum_t \sum_i \left(m_{tj} x_t^{ij}\right) \langle \boldsymbol{w}_t' \circ \boldsymbol{z}', \boldsymbol{x}_t^i \rangle \\ &- \frac{1}{4} \sum_t \sum_i m_{tj} x_t^{ij} \sum_{l \neq j} m_{tl} \phi_l x_t^{il} \\ &- \frac{1}{8} \sum_t \left((\Sigma_j)_{tt} + m_{tj}^2\right) \sum_i \left(x_t^{ij}\right)^2. \end{aligned}$$

When reference point of quadratic lower bound $\boldsymbol{w}' \circ \boldsymbol{z}'$ is set to be the mean parameters from the previous iterations, Eqn. 5 is simplified:

$$\phi_j^{k+1} = \sigma(a). \tag{6}$$

where

$$a = \psi(\alpha) - \psi(\beta) + \sum_t \sum_i \left(y_t^i - \tilde{y_t^i}\right) m_{tj} x_t^{ij} + \frac{1}{8} \sum_t \left(m_{tj}^2 \left(2\phi_j^{(k)} - 1\right) - (\Sigma_j)_{tt}\right) \sum_i \left(x_t^{ij}\right)^2.$$

3 Additional Experimental Results

This section includes additional experimental results. We exam the methods using more evaluation metrics including: accuracy, balanced accuracy, average precision, F1 score, F2 score, and Matthews correlation coefficient (MCC). Table 2 summarizes the definitions of these metrics.

Metric	**Definition**
TP	$\sum_{i=1}^{N} \mathbf{1}(y_i = 1 \wedge \hat{y}_i = 1)$
TN	$\sum_{i=1}^{N} \mathbf{1}(y_i = 0 \wedge \hat{y}_i = 0)$
FP	$\sum_{i=1}^{N} \mathbf{1}(y_i = 0 \wedge \hat{y}_i = 1)$
FN	$\sum_{i=1}^{N} \mathbf{1}(y_i = 1 \wedge \hat{y}_i = 0)$
Precision	$\frac{\mathrm{TP}}{\mathrm{TP+FP}}$
Recall	$\frac{\mathrm{TP}}{\mathrm{TP+FN}}$
Accuracy	$\frac{\mathrm{TP+TN}}{N}$
Balanced Accuracy	$\frac{1}{2}\left(\frac{\mathrm{TP}}{\mathrm{TP+FN}} + \frac{\mathrm{TN}}{\mathrm{TN+FP}}\right)$
F1 Score	$\frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision+Recall}}$
F2 Score	$\frac{5 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{4 \cdot \mathrm{Precision+Recall}}$
MCC	$\frac{(\mathrm{TP} \cdot \mathrm{TN}) - (\mathrm{FP} \cdot \mathrm{FN})}{\sqrt{(\mathrm{TP+FP})(\mathrm{TP+FN})(\mathrm{TN+FP})(\mathrm{TN+FN})}}$

Table 2: Definitions of classification metrics and the intermediate variables given ground truth labels $\boldsymbol{y}$ and predicted labels $\hat{\boldsymbol{y}}$. $\wedge$ denotes the "and" operation, and $\mathbf{1}(\cdot)$ is the indicator function, which is 1 if the condition inside is true and 0 otherwise.

3.1 Synthetic Datasets

This subsection include the detailed experimental results on the synthetic datasets summarized in Table 3

Dataset	Metrics	BayesMTL	MTFL	MSSL	STL-LC	Pooled-LC
dataset1 (dense, balanced)	Accuracy	0.322 (0.06)	**0.793** (0.05)	**0.793** (0.05)	0.688 (0.05)	0.71 (0.04)
	Balanced Accuracy	**0.565** (0.03)	0.5 (0)	0.5 (0)	0.525 (0.05)	0.52 (0.03)
	Average Precision	**0.819** (0.05)	0.793 (0.05)	0.793 (0.05)	0.803 (0.04)	0.800 (0.05)
	F1 Score	0.261 (0.06)	**0.884** (0.03)	**0.884** (0.03)	0.801 (0.04)	0.821 (0.03)
	F2 Score	0.183 (0.05)	**0.950** (0.01)	**0.950** (0.01)	0.800 (0.05)	0.834 (0.03)
	MCC	**0.156** (0.06)	0 (0)	0 (0)	0.0565 (0.09)	0.0427 (0.07)
dataset2 (sparse, balanced)	Accuracy	**0.882** (0.04)	0.225 (0.03)	0.225 (0.03)	0.438 (0.08)	0.364 (0.02)
	Balanced Accuracy	**0.768** (0.05)	0.5 (0)	0.5 (0)	0.616 (0.05)	0.548 (0.04)
	Average Precision	**0.597** (0.10)	0.225 (0.03)	0.225 (0.03)	0.279 (0.06)	0.244 (0.02)
	F1 Score	**0.681** (0.09)	0.367 (0.03)	0.367 (0.03)	0.431 (0.07)	0.383 (0.03)
	F2 Score	0.602 (0.09)	0.590 (0.03)	0.590 (0.03)	**0.634** (0.06)	0.579 (0.04)
	MCC	**0.638** (0.10)	0 (0)	0 (0)	0.227 (0.09)	0.104 (0.07)
dataset3 (ultra sparse, balanced)	Accuracy	**0.988** (0.02)	0.0350 (0.01)	0.0350 (0.01)	0.437 (0.08)	0.199 (0.04)
	Balanced Accuracy	**0.947** (0.06)	0.5 (0)	0.5 (0)	0.708 (0.04)	0.526 (0.09)
	Average Precision	**0.801** (0.20)	0.035 (0.01)	0.035 (0.01)	0.0584 (0.01)	0.0372 (0.01)
	F1 Score	**0.876** (0.13)	0.0674 (0.02)	0.0674 (0.02)	0.110 (0.02)	0.070 (0.02)
	F2 Score	**0.886** (0.12)	0.152 (0.04)	0.152 (0.04)	0.235 (0.05)	0.154 (0.04)
	MCC	**0.879** (0.13)	0 (0)	0 (0)	0.154 (0.02)	0.0199 (0.08)
dataset4 (dense, unbalanced)	Accuracy	0.441 (0.12)	**0.795** (0.04)	**0.795** (0.04)	0.689 (0.09)	0.693 (0.04)
	Balanced Accuracy	**0.584** (0.06)	0.5 (0)	0.5 (0)	0.554 (0.07)	0.517 (0.04)
	Average Precision	**0.827** (0.04)	0.795 (0.04)	0.795 (0.04)	0.814 (0.04)	0.801 (0.04)
	F1 Score	0.483 (0.16)	**0.885** (0.02)	**0.885** (0.02)	0.794 (0.08)	0.808 (0.03)
	F2 Score	0.393 (0.15)	**0.951** (0.01)	**0.951** (0.01)	0.788 (0.12)	0.813 (0.04)
	MCC	**0.147** (0.09)	0 (0)	0 (0)	0.118 (0.12)	0.039 (0.08)
dataset5 (sparse, unbalanced)	Accuracy	**0.796** (0.07)	0.219 (0.04)	0.219 (0.04)	0.451 (0.07)	0.334 (0.03)
	Balanced Accuracy	**0.783** (0.06)	0.5 (0)	0.5 (0)	0.626 (0.036)	0.541 (0.04)
	Average Precision	**0.472** (0.13)	0.219 (0.04)	0.219 (0.04)	0.275 (0.05)	0.234 (0.03)
	F1 Score	**0.625** (0.1)	0.358 (0.05)	0.358 (0.05)	0.427 (0.06)	0.370 (0.04)
	F2 Score	**0.694** (0.05)	0.579 (0.06)	0.579 (0.06)	0.630 (0.05)	0.569 (0.04)
	MCC	**0.516** (0.13)	0 (0)	0 (0)	0.240 (0.05)	0.086 (0.08)
dataset6 (ultra sparse, unbalanced)	Accuracy	**0.917** (0.10)	0.059 (0.0239)	0.059 (0.0239)	0.478 (0.07)	0.21 (0.04)
	Balanced Accuracy	**0.905** (0.08)	0.5 (0)	0.5 (0)	0.715 (0.05)	0.553 (0.03)
	Average Precision	**0.598** (0.02)	0.059 (0.02)	0.059 (0.02)	0.097 (0.03)	0.0646 (0.02)
	F1 Score	**0.699** (0.27)	0.11 (0.04)	0.11 (0.04)	0.176 (0.05)	0.120 (0.04)
	F2 Score	**0.763** (0.20)	0.233 (0.08)	0.233 (0.08)	0.343 (0.08)	0.246 (0.07)
	MCC	**0.704** (0.26)	0 (0)	0 (0)	0.198 (0.03)	0.063 (0.03)

Table 3: Summary of the support recovery results for the simulated data. The bold number means the corresponding method is the best performing algorithm for the given metrics and dataset, and the values in parentheses represent standard deviations computed over 10 different runs. The proposed Bayesian approach outperforms the benchmark methods in all evaluation metrics when there is a shared sparsity structure across regression coefficients of different tasks. Both MSSL and MTFL prioritize the prediction performance in the cross-validation step which results in complete dense solutions (i.e all regression coefficients are non-zero), hence they have identical results.

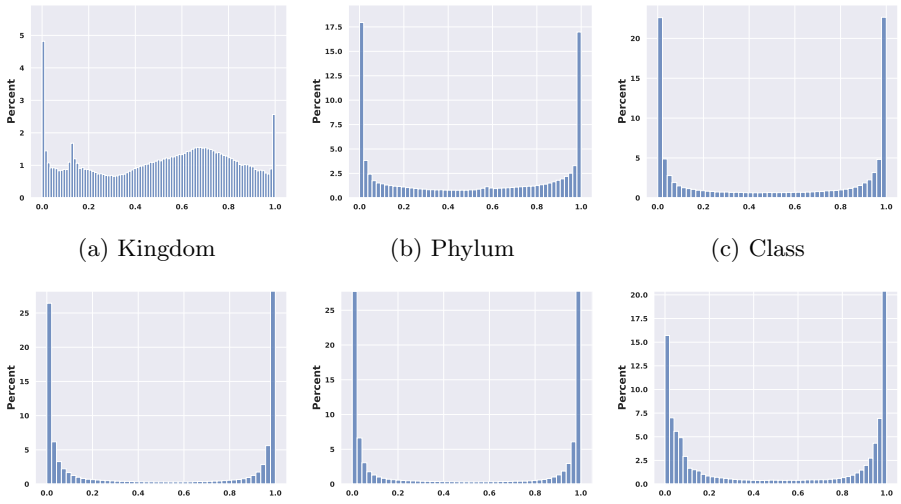
3.2 Microbiome Data

This subsection include the additional experimental results on the microbiome data: table 4 includes the evaluations of the prediction performance, Fig. 1 and Fig. 2 include the predicted probabilities on training and test data for the other taxonomic ranks

respectively, Fig. 3 includes the additional calibration curves, Fig. 4 shows the full set of disease-specific feature-sparsity visualizations at the Class rank, and Fig. 5 provides a compact cross-rank overview of the taxID (feature) selection-stability curves.

Taxon Ranks	Metrics	BayesMTL	MTFL	MSSL	Pooled-LC	STL-LC
Kingdom	Accuracy	0.681 (0.00956)	**0.685 (0.0116)**	**0.685 (0.0121)**	0.532 (0.00974)	0.618 (0.00725)
	Average Precision	0.572 (0.00897)	0.578 (0.00981)	**0.582 (0.00999)**	0.542 (0.00297)	0.57 (0.00872)
	Balanced Accuracy	0.578 (0.0131)	0.581 (0.0124)	**0.589 (0.0137)**	0.531 (0.00856)	0.584 (0.0105)
	F1 Score	0.588 (0.0163)	0.587 (0.0147)	**0.59 (0.0121)**	0.474 (0.0106)	0.523 (0.0284)
	F2 Score	**0.602 (0.0182)**	0.598 (0.0126)	0.598 (0.0104)	0.489 (0.0116)	0.547 (0.0365)
	mcc	0.171 (0.0281)	0.178 (0.0309)	**0.19 (0.0345)**	0.0622 (0.0169)	0.157 (0.0217)
	Sparsity Ratio	0.288 (0.0102)	0.541 (0.0298)	0.467 (0.0267)	0.883 (0.0408)	**0.263 (0.0203)**
Phylum	Accuracy	0.689 (0.0178)	**0.715 (0.0159)**	0.701 (0.0174)	0.552 (0.00715)	0.655 (0.01)
	Average Precision	0.598 (0.00862)	0.615 (0.00804)	**0.621 (0.00927)**	0.538 (0.00385)	0.593 (0.00802)
	Balanced Accuracy	0.618 (0.0146)	0.64 (0.0147)	**0.651 (0.0151)**	0.527 (0.0085)	0.612 (0.00993)
	F1 Score	0.614 (0.0153)	**0.63 (0.018)**	0.627 (0.0158)	0.521 (0.00613)	0.579 (0.0387)
	F2 Score	0.619 (0.0203)	**0.63 (0.0222)**	0.619 (0.0193)	0.536 (0.00786)	0.601 (0.0431)
	mcc	0.243 (0.0309)	0.294 (0.033)	**0.302 (0.033)**	0.0458 (0.0139)	0.222 (0.0223)
	Sparsity Ratio	0.182 (0.0169)	0.301 (0.00222)	0.311 (0.00286)	0.781 (0.0247)	**0.0856 (0.012)**
Class	Accuracy	0.688 (0.00922)	**0.726 (0.0099)**	0.695 (0.0156)	0.58 (0.0163)	0.668 (0.013)
	Average Precision	0.599 (0.0107)	**0.625 (0.00884)**	0.623 (0.0116)	0.548 (0.0061)	0.605 (0.00575)
	Balanced Accuracy	0.619 (0.013)	**0.659 (0.0114)**	0.656 (0.0167)	0.544 (0.0121)	0.629 (0.00728)
	F1 Score	0.619 (0.0143)	**0.652 (0.0134)**	0.632 (0.016)	0.54 (0.0155)	0.584 (0.0269)
	F2 Score	0.624 (0.0167)	**0.657 (0.0155)**	0.625 (0.0188)	0.552 (0.0159)	0.603 (0.037)
	mcc	0.245 (0.0261)	**0.326 (0.0237)**	0.307 (0.0324)	0.0803 (0.031)	0.26 (0.0151)
	Sparsity Ratio	0.116 (0.0121)	0.24 (0.00565)	0.295 (0.0531)	0.692 (0.019)	**0.0697 (0.00536)**
Order	Accuracy	0.698 (0.01)	**0.703 (0.0116)**	0.68 (0.012)	0.6 (0.00926)	0.68 (0.0142)
	Average Precision	0.609 (0.00955)	0.618 (0.00827)	**0.622 (0.00906)**	0.558 (0.0044)	0.617 (0.00803)
	Balanced Accuracy	0.634 (0.0126)	0.655 (0.0123)	**0.659 (0.0131)**	0.557 (0.00793)	0.645 (0.0119)
	F1 Score	0.629 (0.00964)	**0.649 (0.018)**	0.623 (0.0202)	0.553 (0.0147)	0.574 (0.0121)
	F2 Score	0.633 (0.0112)	**0.659 (0.0238)**	0.618 (0.0269)	0.566 (0.0169)	0.589 (0.0171)
	mcc	0.271 (0.0234)	**0.309 (0.0279)**	0.306 (0.0237)	0.112 (0.0148)	0.29 (0.023)
	Sparsity Ratio	0.109 (0.0115)	0.199 (0.00967)	0.222 (0.00103)	0.536 (0.00935)	**0.046 (0.00772)**
Family	Accuracy	0.704 (0.0218)	0.687 (0.0148)	0.665 (0.014)	0.608 (0.0162)	**0.706 (0.00802)**
	Average Precision	0.613 (0.0168)	0.625 (0.0144)	0.621 (0.0145)	0.567 (0.0104)	**0.632 (0.00636)**
	Balanced Accuracy	0.643 (0.0256)	0.658 (0.0149)	0.656 (0.0139)	0.57 (0.0179)	**0.663 (0.00637)**
	F1 Score	0.627 (0.0239)	0.62 (0.0171)	0.616 (0.0123)	0.565 (0.0172)	**0.632 (0.0184)**
	F2 Score	0.631 (0.0235)	0.609 (0.0223)	0.612 (0.0126)	0.574 (0.0164)	**0.648 (0.0183)**
	mcc	0.285 (0.0516)	0.307 (0.0286)	0.298 (0.0279)	0.134 (0.0366)	**0.326 (0.00791)**
	Sparsity Ratio	0.0429 (0.0151)	0.409 (x0.293)	0.172 (0.00179)	0.414 (0.00336)	**0.0407 (0.00736)**
x Genus	Accuracy	0.707 (0.00852)	0.65 (0.0291)	0.66 (0.0156)	0.619 (0.00865)	**0.728 (0.0158)**
	Average Precision	0.616 (0.00589) x	0.609 (0.0168)	0.62 (0.0123)	0.575 (0.00737)	**0.647 (0.00831)**
	Balanced Accuracy	0.641 (0.0109)	0.642 (0.0238)	0.657 (0.0135)	0.593 (0.0123)	**0.683 (0.00959)**
	F1 Score	0.634 (0.00774)	0.562 (0.109)	0.612 (0.0143)	0.579 (0.0167)	**0.638 (0.0267)**
	F2 Score	0.638 (0.00785)	0.562 (0.11)	0.61 (0.015)	0.589 (0.0219)	**0.649 (0.0305)**
	mcc	0.288 (0.0204)	0.272 (0.0468)	0.3 (0.0265)	0.175 (0.0246)	**0.366 (0.02)**
	Sparsity Ratio	**0.0181 (0.00677)**	0.403 (0.281)	0.152 (0.00116)	0.326 (0.00331)	0.0351 (0.00621)
Species	Accuracy	0.711 (0.0143)	0.57 (0.0205)	0.672 (0.0106)	0.648 (0.0113)	**0.744 (0.0145)**
	Average Precision	0.621 (0.015)	0.57 (0.00941)	0.628 (0.00825)	0.593 (0.00687)	**0.66 (0.0119)**
	Balanced Accuracy	0.65 (0.0168)	0.584 (0.0158)	0.671 (0.0117)	0.626 (0.0103)	**0.7 (0.0132)**
	F1 Score	0.635 (0.0168)	0.301 (0.0674)	0.622 (0.0108)	0.611 (0.00954)	**0.651 (0.0222)**
	F2 Score	0.634 (0.0146)	0.297 (0.0656)	0.618 (0.0136)	0.625 (0.0128)	**0.661 (0.0244)**
	mcc	0.305 (0.0333)	0.16 (0.0288)	0.325 (0.023)	0.234 (0.0202)	**0.401 (0.0279)**
	Sparsity Ratio	**0.0289 (0.0053)**	0.149 (0.0275)	0.149 (0.000836)	0.279 (0.00367)	0.0319 (0.00481)

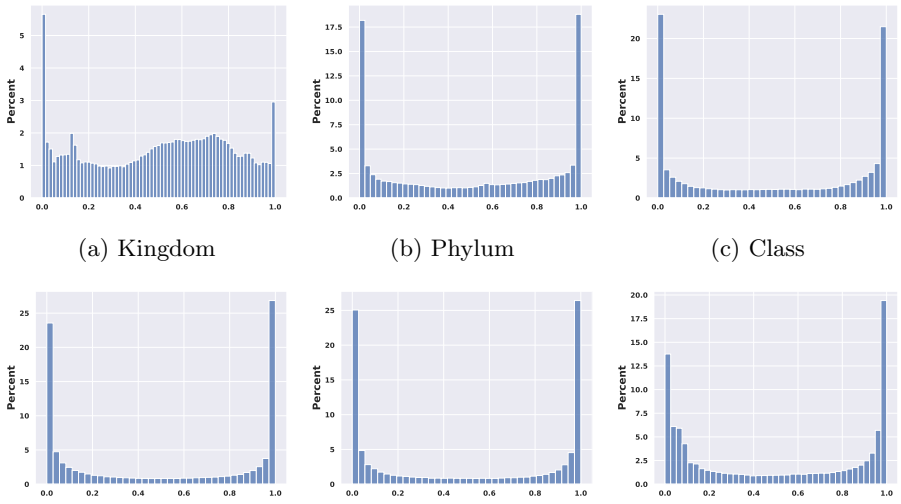
Table 4: Summary of the prediction performance. The bold number means the corresponding method is the best performing algorithm for the given metrics and taxonomic rank, and the values in parentheses represent standard deviations computed over 5 different runs. Due to the heterogeneous nature of the data, we do not see an improvement of the proposed approach over single-task model. However, the proposed approach is the only multitask method that provides a sparse solution i.e identify common bacteria across studies of the same disease categories that are informative for the predictions.



(a) Kingdom (b) Phylum (c) Class

(d) Family (e) Genus (f) Species

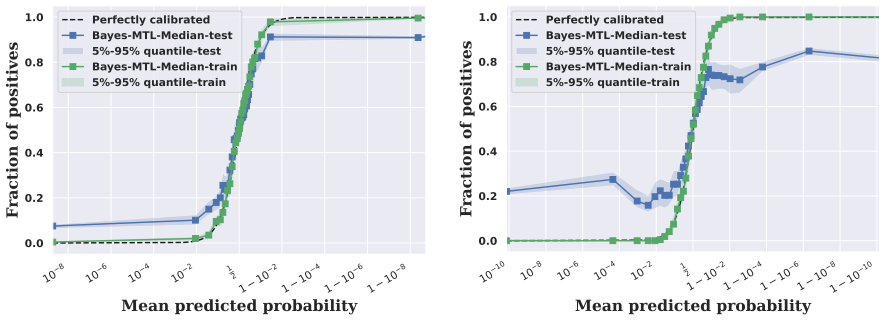
Fig. 1: Histogram of predicted probabilities on *training* data for different taxonomic ranks.



(a) Kingdom (b) Phylum (c) Class

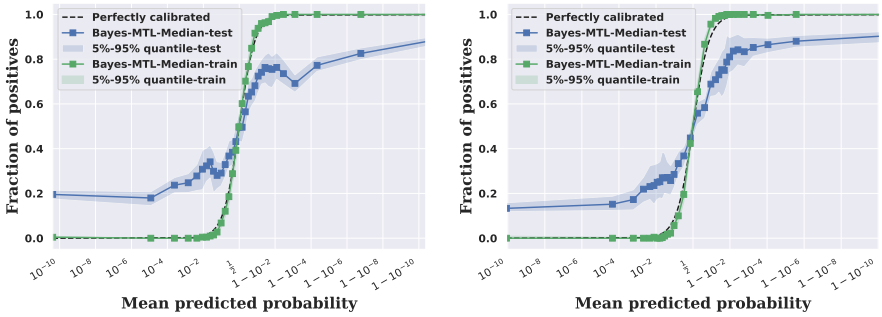
(d) Family (e) Genus (f) Species

Fig. 2: Histogram of predicted probabilities on *test* data for different taxonomic ranks.



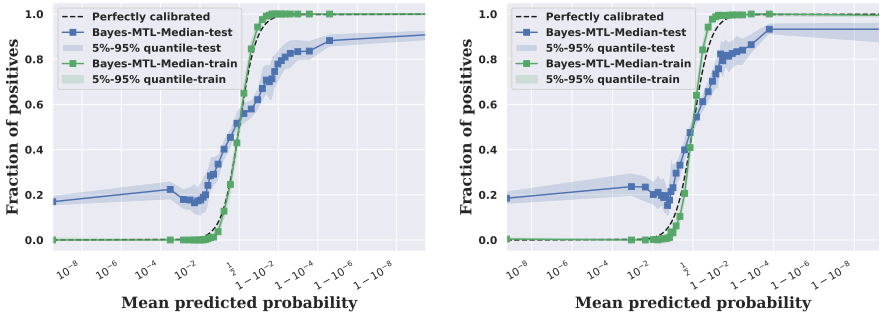
(a) Kingdom

(b) Phylum



(c) Class

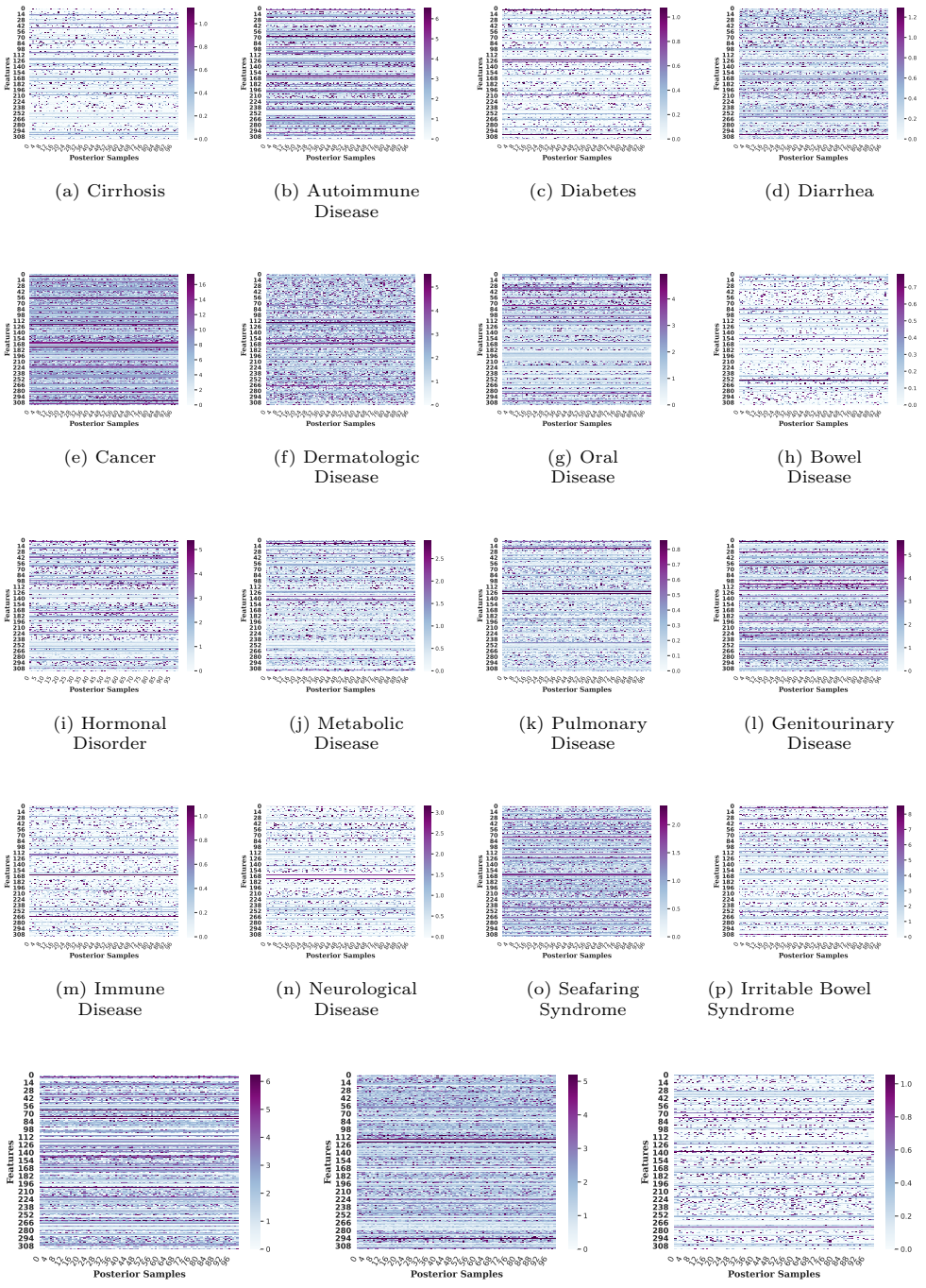
(d) Family



(e) Genus

(f) Species

Fig. 3: Calibration curves for the proposed BayesMTL model for different taxonomic ranks.



(a) Cirrhosis (b) Autoimmune Disease (c) Diabetes (d) Diarrhea

(e) Cancer (f) Dermatologic Disease (g) Oral Disease (h) Bowel Disease

(i) Hormonal Disorder (j) Metabolic Disease (k) Pulmonary Disease (l) Genitourinary Disease

(m) Immune Disease (n) Neurological Disease (o) Seafaring Syndrome (p) Irritable Bowel Syndrome

(q) Age-related Macular Degeneration (r) Gastrointestinal Infection (s) Cardiovascular Disease

Fig. 4: Feature sparsity visualization across 19 different disease categories of Class taxonomic rank. The x-axis corresponds to different samples drawn from the posterior distribution and the y-axis correspond to different taxIDs. The gradation from white to black for variable color corresponds to its increasing importance weight, and the darker shaded horizontal lines represent the sparse features selected by the algorithm. For diabetes and diarrhea, few taxIDs are considered informative for the health prediction task by the model, while for cardiovascular disease the majority of taxIDs are considered informative.

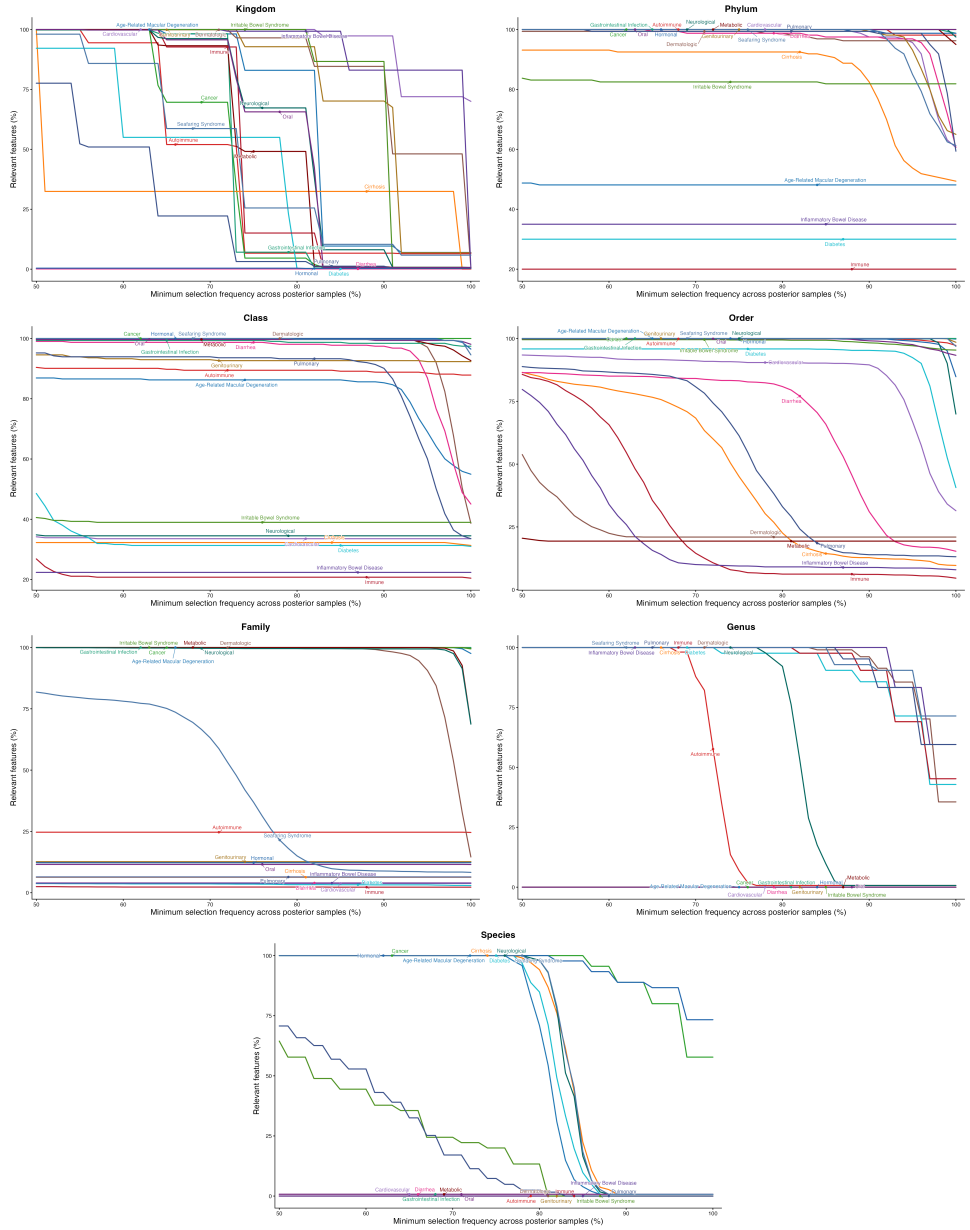


Fig. 5: TaxID selection-stability curves across taxonomic ranks. Within each panel, the y-axis gives the percentage of relevant features (taxID) exceeding the recovered value threshold and the x-axis gives the minimum proportion of posterior samples in which a feature must remain selected. Rank titles are embedded in the individual panels.

References

[1] Pasolli E, Schiffer L, Manghi P, Renson A, Obenchain V, Truong DT, et al. Accessible, curated metagenomic data through ExperimentHub. Nature methods. 2017;14(11):1023–1024.

[2] Zysset-Burri DC, Keller I, Berger LE, Largiadèr CR, Wittwer M, Wolf S, et al. Associations of the intestinal microbiome with the complement system in neovascular age-related macular degeneration. NPJ genomic medicine. 2020;5(1):34.

[3] McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, et al. American gut: an open platform for citizen science microbiome research. Msystems. 2018;3(3):10–1128.

[4] Karlsson FH, Tremaroli V, Nookaew I, Bergström G, Behre CJ, Fagerberg B, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature. 2013;498(7452):99–103.

[5] Sankaranarayanan K, Ozga AT, Warinner C, Tito RY, Obregon-Tito AJ, Xu J, et al. Gut microbiome diversity among Cheyenne and Arapaho individuals from western Oklahoma. Current Biology. 2015;25(24):3161–3169.

[6] Heintz-Buschart A, May P, Laczny CC, Lebrun LA, Bellora C, Krishna A, et al. Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. Nature microbiology. 2016;2(1):1–13.

[7] Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, et al. Alterations of the human gut microbiome in liver cirrhosis. Nature. 2014;513(7516):59–64.

[8] Loomba R, Seguritan V, Li W, Long T, Klitgord N, Bhatt A, et al. Gut microbiome-based metagenomic signature for non-invasive detection of advanced fibrosis in human nonalcoholic fatty liver disease. Cell metabolism. 2019;30(3):607.

[9] Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, et al. Gut microbiome development along the colorectal adenoma–carcinoma sequence. Nature communications. 2015;6(1):6528.

[10] Yu J, Feng Q, Wong SH, Zhang D, yi Liang Q, Qin Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut. 2017;66(1):70–78.

[11] Hannigan GD, Duhaime MB, Ruffin IV MT, Koumpouras CC, Schloss PD. Diagnostic potential and interactive dynamics of the colorectal cancer virome. MBio. 2018;9(6):10–1128.

[12] Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. Nature medicine. 2019;25(6):968–976.

[13] Nagata N, Nishijima S, Kojima Y, Hisada Y, Imbe K, Miyoshi-Akiyama T, et al. Metagenomic identification of microbial signatures predicting pancreatic cancer from a multinational study. Gastroenterology. 2022;163(1):222–238.

[14] Castro-Nallar E, Bendall ML, Pérez-Losada M, Sabuncyan S, Severance EG, Dickerson FB, et al. Composition, taxonomy and functional diversity of the oropharynx microbiome in individuals with schizophrenia and controls. PeerJ. 2015;3:e1140.

[15] Nagy-Szakal D, Williams BL, Mishra N, Che X, Lee B, Bateman L, et al. Fecal metagenomic profiles in subgroups of patients with myalgic encephalomyelitis/chronic fatigue syndrome. Microbiome. 2017;5(1):1–17.

[16] Qian Y, Yang X, Xu S, Huang P, Li B, Du J, et al. Gut metagenomics-derived genes as potential biomarkers of Parkinson's disease. Brain. 2020;143(8):2474–2489.

[17] Zhu F, Ju Y, Wang W, Wang Q, Guo R, Ma Q, et al. Metagenome-wide association of gut microbiome features for schizophrenia. Nature communications. 2020;11(1):1612.

[18] Jo S, Kang W, Hwang YS, Lee SH, Park KW, Kim MS, et al. Oral and gut dysbiosis leads to functional alterations in Parkinson's disease. npj Parkinson's Disease. 2022;8(1):87.

[19] Laske C, Müller S, Preische O, Ruschil V, Munk MH, Honold I, et al. Signature of Alzheimer's disease in intestinal microbiome: results from the AlzBiom study. Frontiers in Neuroscience. 2022;16:792996.

[20] Wallen ZD, Demirkan A, Twa G, Cohen G, Dean MN, Standaert DG, et al. Metagenomics of Parkinson's disease implicates the gut microbiome in multiple disease mechanisms. Nature communications. 2022;13(1):6958.

[21] David LA, Weil A, Ryan ET, Calderwood SB, Harris JB, Chowdhury F, et al. Gut microbial succession follows acute secretory diarrhea in humans. MBio. 2015;6(3):10–1128.

[22] Kieser S, Sarker SA, Sakwinska O, Foata F, Sultana S, Khan Z, et al. Bangladeshi children with acute diarrhoea show faecal microbiomes with increased Streptococcus abundance, irrespective of diarrhoea aetiology. Environmental microbiology. 2018;20(6):2256–2269.

[23] Chng KR, Tay ASL, Li C, Ng AHQ, Wang J, Suri BK, et al. Whole metagenome profiling reveals skin microbiome-dependent susceptibility to atopic dermatitis flare. Nature microbiology. 2016;1(9):1–10.

[24] Byrd AL, Deming C, Cassidy SK, Harrison OJ, Ng WI, Conlan S, et al. Staphylococcus aureus and Staphylococcus epidermidis strain diversity underlying pediatric atopic dermatitis. Science translational medicine. 2017;9(397):eaal4651.

[25] Tay AS, Li C, Nandi T, Chng KR, Andiappan AK, Mettu VS, et al. Atopic dermatitis microbiomes stratify into ecologic dermotypes enabling microbial virulence and disease severity. Journal of Allergy and Clinical Immunology. 2021;147(4):1329–1340.

[26] Chang HW, Yan D, Singh R, Bui A, Lee K, Truong A, et al. Multiomic analysis of the gut microbiome in psoriasis reveals distinct host–microbe associations. JID Innovations. 2022;2(3):100115.

[27] Vincent C, Miller MA, Edens TJ, Mehrotra S, Dewar K, Manges AR. Bloom and bust: intestinal microbiota dynamics in response to hospital exposures and Clostridium difficile colonization or infection. Microbiome. 2016;4:1–11.

[28] Rosa BA, Supali T, Gankpala L, Djuardi Y, Sartono E, Zhou Y, et al. Differential human gut microbiome assemblages during soil-transmitted helminth infections in Indonesia and Liberia. Microbiome. 2018;6(1):1–19.

[29] Rubel MA, Abbas A, Taylor LJ, Connell A, Tanes C, Bittinger K, et al. Lifestyle and the presence of helminths is associated with gut microbiome composition in Cameroonians. Genome biology. 2020;21(1):1–32.

[30] Hall AB, Yassour M, Sauk J, Garner A, Jiang X, Arthur T, et al. A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. Genome medicine. 2017;9:1–12.

[31] 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. Nature microbiology. 2019;4(2):293–305.

[32] Integrative H. The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. Cell host & microbe. 2014;16(3):276–289.

[33] Zuo W, Wang B, Bai X, Luan Y, Fan Y, Michail S, et al. 16S rRNA and metagenomic shotgun sequencing data revealed consistent patterns of gut microbiome signature in pediatric ulcerative colitis. Scientific Reports. 2022;12(1):6421.

[34] Jie Z, Xia H, Zhong SL, Feng Q, Li S, Liang S, et al. The gut microbiome in atherosclerotic cardiovascular disease. Nature communications. 2017;8(1):845.

[35] Li J, Zhao F, Wang Y, Chen J, Tao J, Tian G, et al. Gut microbiota dysbiosis contributes to the development of hypertension. Microbiome. 2017;5:1–19.

[36] Yan Q, Gu Y, Li X, Yang W, Jia L, Chen C, et al. Alterations of the gut microbiome in hypertension. Frontiers in cellular and infection microbiology. 2017;7:381.

[37] Polster SP, Sharma A, Tanes C, Tang AT, Mericko P, Cao Y, et al. Permissive microbiome characterizes human subjects with a neurovascular disease cavernous angioma. Nature communications. 2020;11(1):2659.

[38] Xiong Z, Peng K, Song S, Zhu Y, Gu J, Huang C, et al. Cerebral intraparenchymal hemorrhage changes patients' gut bacteria composition and function. Frontiers in Cellular and Infection Microbiology. 2022;12:829491.

[39] Espinoza JL, Harkins DM, Torralba M, Gomez A, Highlander SK, Jones MB, et al. Supragingival plaque microbiome ecology and functional potential in the context of health and disease. MBio. 2018;9(6):10–1128.

[40] Ghensi P, Manghi P, Zolfo M, Armanini F, Pasolli E, Bolzan M, et al. Strong oral plaque microbiome signatures for dental implant diseases identified by strain-resolution metagenomics. npj Biofilms and Microbiomes. 2020;6(1):47.

[41] Wirth R, Maróti G, Lipták L, Mester M, Al Ayoubi A, Pap B, et al. Microbiomes in supragingival biofilms and saliva of adolescents with gingivitis and gingival health. Oral Diseases. 2022;28(7):2000–2014.

[42] Ye Z, Zhang N, Wu C, Zhang X, Wang Q, Huang X, et al. A metagenomic study of the gut microbiome in Behcet's disease. Microbiome. 2018;6(1):1–13.

[43] Zhu Q, Hou Q, Huang S, Ou Q, Huo D, Vázquez-Baeza Y, et al. Compositional and genetic alterations in Graves' disease gut microbiome reveal specific diagnostic biomarkers. The ISME journal. 2021;15(11):3399–3411.

[44] Guillén Y, Noguera-Julian M, Rivera J, Casadellà M, Zevin AS, Rocafort M, et al. Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection. Mucosal immunology. 2019;12(1):232–246.

[45] Hu Y, Feng Y, Wu J, Liu F, Zhang Z, Hao Y, et al. The gut microbiome signatures discriminate healthy from pulmonary tuberculosis patients. Frontiers in cellular and infection microbiology. 2019;9:90.

[46] Pust MM, Wiehlmann L, Davenport C, Rudolf I, Dittrich AM, Tümmler B. The human respiratory tract microbial community structures in healthy and cystic fibrosis infants. npj Biofilms and Microbiomes. 2020;6(1):61.

[47] Zuo T, Zhang F, Lui GC, Yeoh YK, Li AY, Zhan H, et al. Alterations in gut microbiota of patients with COVID-19 during time of hospitalization. Gastroenterology. 2020;159(3):944–955.

[48] Bai X, Narayanan A, Skagerberg M, Ceña-Diez R, Giske CG, Strålin K, et al. Characterization of the upper respiratory bacterial microbiome in critically Ill COVID-19 patients. Biomedicines. 2022;10(5):982.

[49] Liu Q, Mak JWY, Su Q, Yeoh YK, Lui GCY, Ng SSS, et al. Gut microbiota dynamics in a prospective cohort of patients with post-acute COVID-19 syndrome. Gut. 2022;71(3):544–552.

[50] Xiao G, Cai Z, Guo Q, Ye T, Tang Y, Guan P, et al. Insights into the unique lung microbiota profile of pulmonary tuberculosis patients using metagenomic next-generation sequencing. Microbiology Spectrum. 2022;10(1):e01901–21.

[51] Zhang F, Wan Y, Zuo T, Yeoh YK, Liu Q, Zhang L, et al. Prolonged impairment of short-chain fatty acid and L-isoleucine biosynthesis in gut microbiome in patients with COVID-19. Gastroenterology. 2022;162(2):548–561.

[52] Zhou T, Wu J, Zeng Y, Li J, Yan J, Meng W, et al. SARS-CoV-2 triggered oxidative stress and abnormal energy metabolism in gut microbiota. MedComm. 2022;3(1):e112.

[53] Maya-Lucas O, Murugesan S, Nirmalkar K, Alcaraz LD, Hoyo-Vadillo C, Pizano-Zárate ML, et al. The gut microbiome of Mexican children affected by obesity. Anaerobe. 2019;55:11–23.

[54] Qi X, Yun C, Sun L, Xia J, Wu Q, Wang Y, et al. Gut microbiota–bile acid–interleukin-22 axis orchestrates polycystic ovary syndrome. Nature medicine. 2019;25(8):1225–1233.

[55] Bommana S, Richards G, Kama M, Kodimerla R, Jijakli K, Read TD, et al. Metagenomic shotgun sequencing of Endocervical, vaginal, and rectal samples among Fijian women with and without Chlamydia trachomatis reveals disparate microbial populations and function across anatomic sites: a pilot study. Microbiology spectrum. 2022;10(3):e00105–22.

[56] Sun Z, Zhang M, Li M, Bhaskar Y, Zhao J, Ji Y, et al. Interactions between human gut microbiome dynamics and sub-optimal health symptoms during seafaring expeditions. Microbiology Spectrum. 2022;10(1):e00925–21.

[57] Blei DM, Kucukelbir A, McAuliffe JD. Variational inference: A review for statisticians. Journal of the American statistical Association. 2017;112(518):859–877.