Supplemetary Material

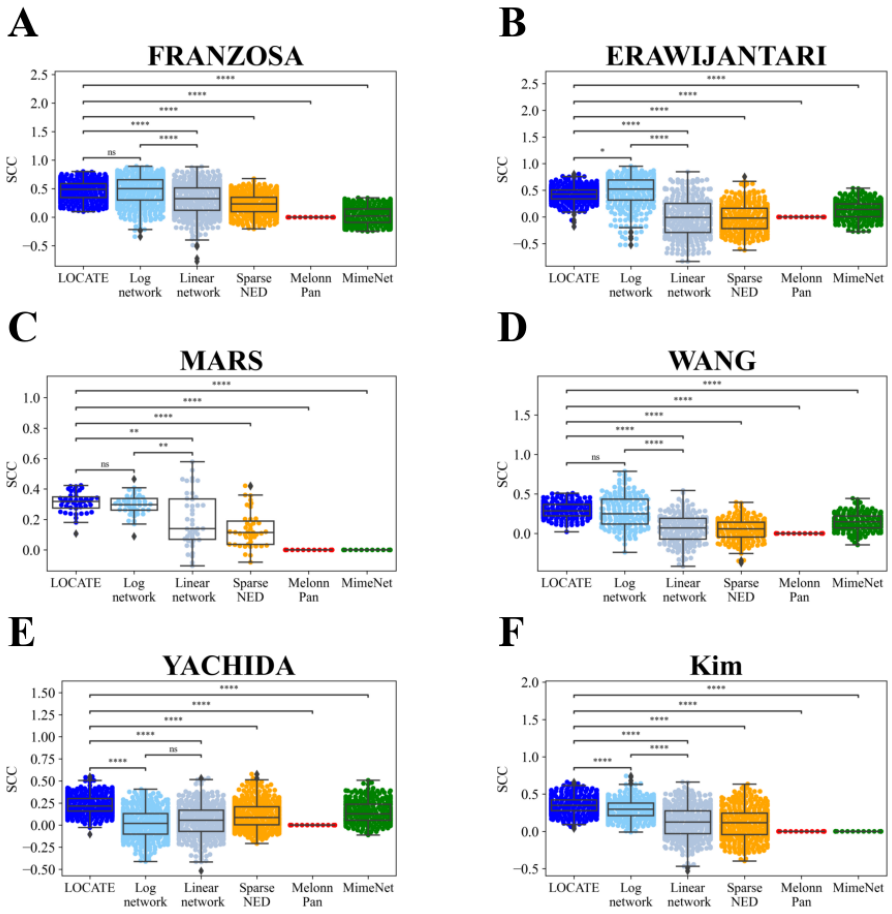

Figure 1: Latent representation of microbiome can be used to predict metabolites in each dataset separately better than all existing methods. **A - E.** Comparison between LOCATE and all state-of-the-art metabolites prediction models as well as a Linear network and a Log network over the different datasets FRANZOSA **(A)**, ERAWIJANTARI**(B)**, MARS **(C)**, WANG **(D)** and YACHIDA **(E)**.**F.** Comparison between LOCATE and all state-of-the-art metabolites prediction models as well as a Linear network and a Log network over the Kim dataset.

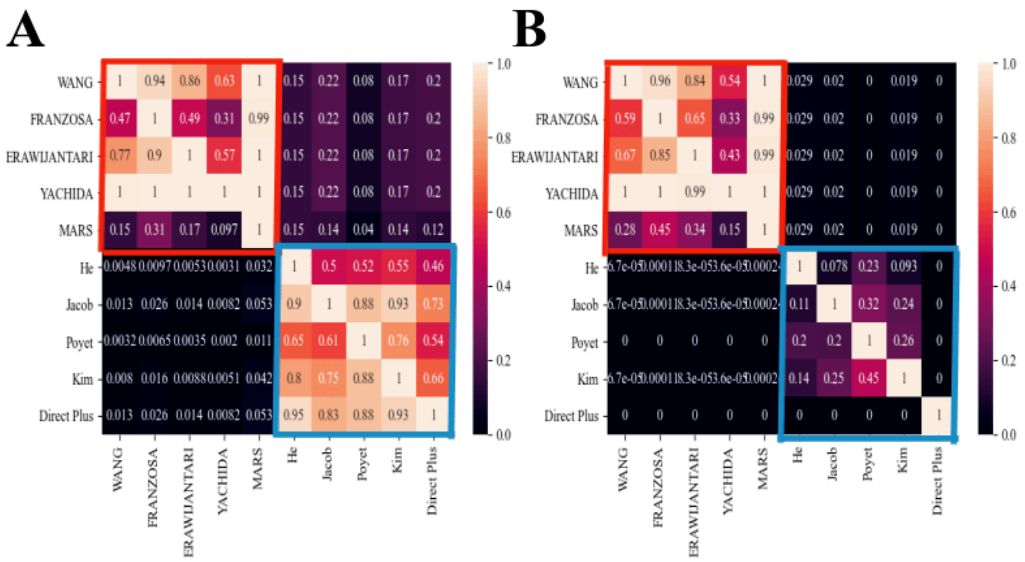

Figure 2: Intersections between pairs of cohorts of 16S and WGS at the order taxonomic level **(A)**, and at the species level **(B)**. The overlap between the pairs of the WGS datasets (red) is much higher than the overlap in the 16S datasets (blue), especially at the species level. The overlap between 16S and 16S is higher than the overlap between 16S and WGS, although the number of taxa in WGS is much higher than 16S, and one could expect the 16S taxa to be included in the WGS.

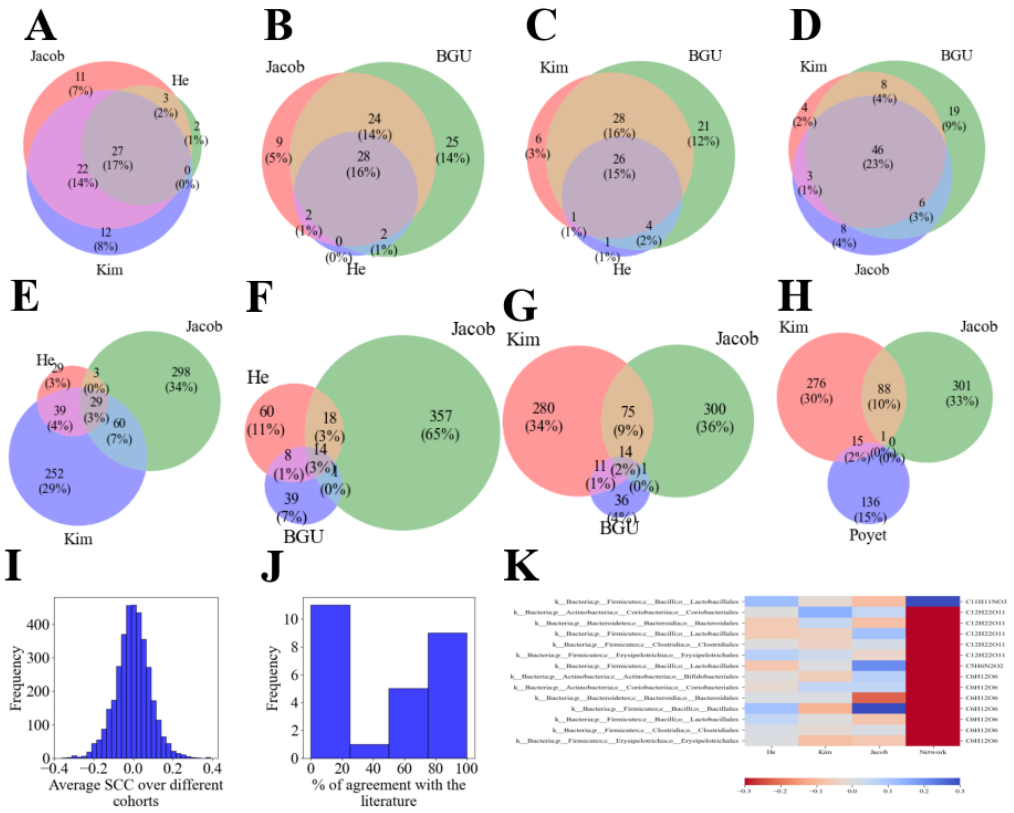

Figure 3: Low intersection between the orders microbiome and metabolites of different cohorts. **A - D.** Venn diagrams of the microbiome of triads 16S datasets. **E - H.** Venn diagrams of the metabolites of triads 16S datasets. Each color represents a dataset, and the intermediate colors represent the intersection. **I.** Histogram of average SCCs between each microbe and each metabolite that appears at least at 2 cohorts (of the 16S cohorts). The histogram's peak is at 0.0, what emphasizes the inconsistent SCCs cross datasets. **J.** Histogram of percent of agreement with the correlations reported in the literature and the correlations found in the cohorts. Most of the correlations do not agree with the literature. **K.** Heatmap of NMF coefficients between microbes and metabolites over different datasets (He, Kim and Jacob) vs the relations that are reported in the literature. Blue/Red colors represent positive/negative correlations. The relations vary between different datasets and do not preserve the known relations from the literature [1].

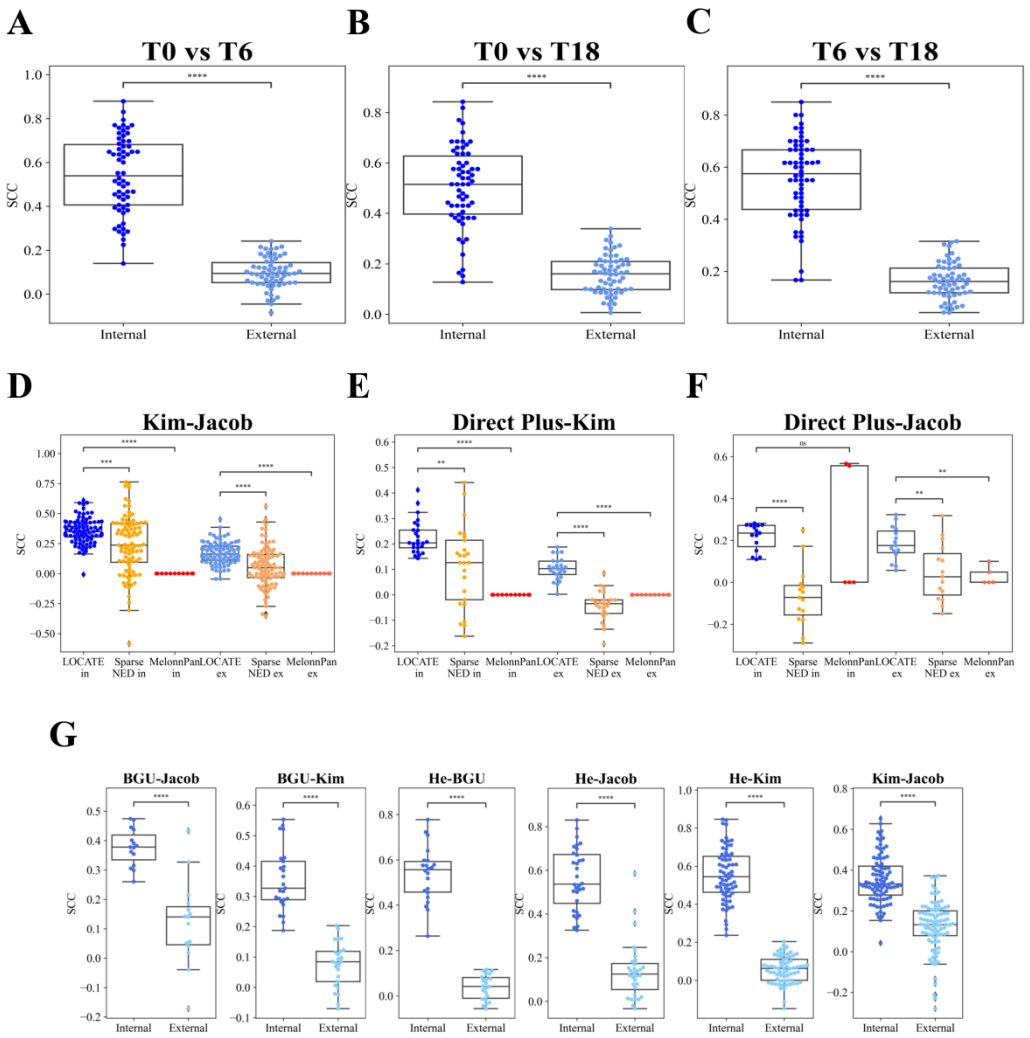

Figure 4: Microbiome-metabolite relations are dataset specific. **A - C.** Swarm plots of LOCATE's predicted metabolites SCCs in the cross-times test over the Direct Plus cohort. The dark blue points represent the SCCs of the "in-learning", referred as "Internal", where only one time point was used for the training and the testing, by 10 CV approach. The lightblue points represent the SCCs of the "ex-learning", referred as "External", where LOCATE is trained on one time point and is tested on another one. There is a decrease in the accuracy of the ex-learning vs the in-learning. The stars follow all other figures. **D - F.** Swarm plots of all of the cross-datasets learning between couples of datasets, Kim-Jacob **(D)**, Direct Plus-Kim**(E)**, Direct Plus-Jacob **(F)**. **G.** Swarm plots of all of the cross-datasets learning between couples of datasets of the Log network model. The decline in performance between the "in-learning" and "ex-learning can be seen here too.

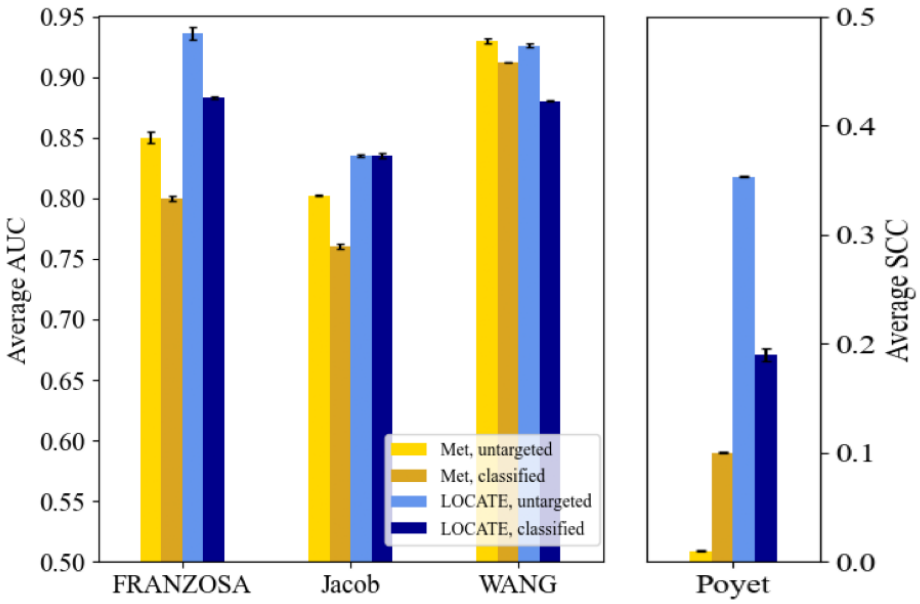

Figure 5: Phenotype predictions based on targeted metabolites vs untargeted metabolites. In each cohort with untargeted metabolites, the phenotype is predicted once based on models which are trained only on classified metabolites (the dark bars), and once on all the metabolites including unclassified ones. LOCATE based on untargeted metabolites (lightblue) outperforms all the other methods. The performance is measured as the average AUC (for binary phenotypes) and SCC for continuous phenotypes on a test set over 10 CVs. The black error bars represent the standard errors within the 10 CVs.

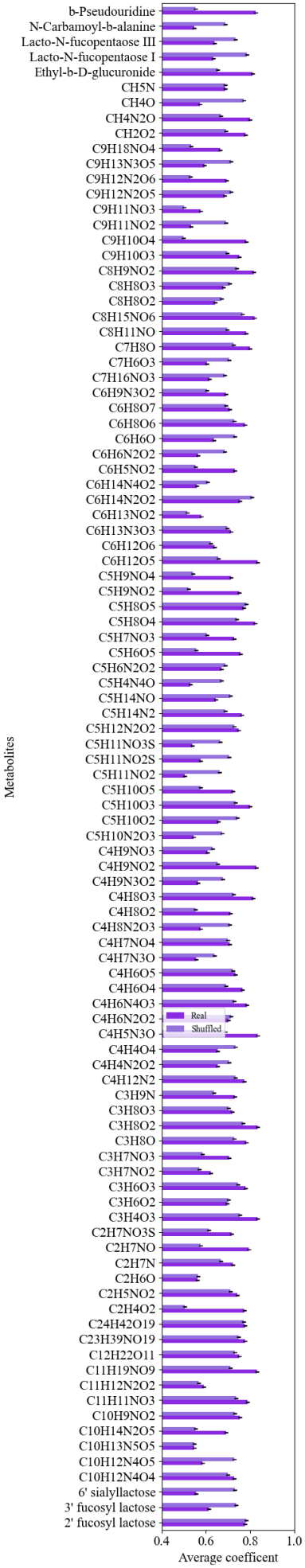
b-Pseudouridine
N-Carbamoyl-b-alanine
Lacto-N-fucopentaose III
Lacto-N-fucopentaose I
Ethyl-b-D-glucuronide
CH5N
CH4O
CH4N2O
CH2O2
C9H18NO4
C9H13N3O5
C9H12N2O6
C9H12N2O5
C9H11NO3
C9H11NO2
C9H10O4
C9H10O3
C8H9NO2
C8H8O3
C8H8O2
C8H15NO6
C8H11NO
C7H8O
C7H6O3
C7H16NO3
C6H9N3O2
C6H8O7
C6H8O6
C6H6O
C6H6N2O2
C6H5NO2
C6H14N4O2
C6H14N2O2
C6H13NO2
C6H13N3O3
C6H12O6
C6H12O5
C5H9NO4
C5H9NO2
C5H8O5
C5H8O4
C5H7NO3
C5H6O5
C5H6N2O2
C5H4N4O
C5H14NO
C5H14N2
C5H12N2O2
C5H11NO3S
C5H11NO2S
C5H11NO2
C5H10O5
C5H10O3
C5H10O2
C5H10N2O3
C4H9NO3
C4H9NO2
C4H9N3O2
C4H8O3
C4H8O2
C4H8N2O3
C4H7NO4
C4H7N3O
C4H6O5
C4H6O4
C4H6N4O3
C4H6N2O2
C4H5N3O
C4H4O4
C4H4N2O2
C4H12N2
C3H9N
C3H8O3
C3H8O2
C3H8O
C3H7NO3
C3H7NO2
C3H6O3
C3H6O2
C3H4O3
C2H7NO3S
C2H7NO
C2H7N
C2H6O
C2H5NO2
C2H4O2
C24H42O19
C23H39NO19
C12H22O11
C11H19NO9
C11H12N2O2
C11H11NO3
C10H9NO2
C10H14N2O5
C10H13N5O5
C10H12N4O5
C10H12N4O4
6' sialyllactose
3' fucosyl lactose
2' fucosyl lactose
Real
Shuffled
Metabolites
0.4
0.6
0.8
1.0
Average coefficent

Figure 6: Average coefficients of each metabolite in the real dataset (dark bar) and in the shuffled one (light bar). The black error bars are for the standard errors.

1 Supp. Mat. Tables

Table 1: Summary of current state-of-the-art methods

Model	Advantages	Disadvantages	Ref
PRMT	1. First framework. 2. Significant correlations between PRMT scores and relative abundances of selected environmental measurements.	1. Based on biological known networks. 2. Limited only to the KEGG database. 3. Performance is limited. 4. Not transferable (mixed datasets, between datasets).	[2]
MIMOSA	1. Gives information about the contribution of each taxon to the metabolites. 2. Succeed to predict relations in real and simulated datasets. 3. Freely available web server.	1. Based on biological known networks. 2. Limited only to the KEGG database. 3. Requires the original sequences data. 4. Performance is limited (worse than MelonnPan). 5. Not transferable (mixed datasets, between datasets).	[3, 4]
Mangosteen	1. Success in specific metabolites.	1. Based on biological known networks. 2. Limited only to the KEGG database. 3. Performance is limited (worse than MelonnPan and MIMOSA). 4. Not transferable (mixed datasets, between datasets).	[5]
MelonnPan	1. Best performance among existing state-of-the-arts. 2. Quite a good definition for well-predicted metabolites. 3. Independent from previous biological knowledge.	1. Long running times (run for each metabolite separately). 2. Cannot cope with all metabolites. 3. Sometimes returns the same prediction to all samples for specific metabolites. 4. Not transferable (mixed datasets, between datasets).	[6]
MiMeNet	1. Learning multiple metabolites simultaneously enables us to find relations between the metabolites. 2. They claim it is better than existing state-of-the-art methods (PRMT , MelonnPan, and SparseNED). 3. Independent from previous biological knowledge.	1. Long-running time. 2. Performance is limited in the external test within a dataset. 3. Not transferable (mixed datasets, between datasets).	[7]
SparseNED	1. Learning multiple metabolites simultaneously enables us to find relations between the metabolites. 2. Quite short running times. 3. Independent from previous biological knowledge.	1. Performance is limited within a dataset (worse than MelonnPan). 2. Not transferable (mixed datasets, between datasets).	[8]

Table 2: Datasets details

Dataset	Cohort description	16S or WGS	N (subjects)		N (samples)		Targeted / untargeted	Ref
			Case	Control	Case	Control		
Direct Plus	18-month randomized clinical trial, we assigned 294 participants with abdominal obesity/dyslipidemia into healthy dietary guidelines (HDG), MED and green-MED weight-loss diet groups, all accompanied by physical activity.	16S	NA	294	NA	784	Targeted	[9]
Kim	Patients with advanced colorectal adenomas, colorectal cancer, and controls.	16S	138	102	138	102	Untargeted	[10]
He	Infants over several time points during the 1st year of life, either breast-fed, formula-fed, or experimental formula fed.	16S	NA	80	NA	277	Targeted	[11]
Jacob	Inflammatory bowel disease patients and their first degree (healthy) relatives.	16S	36	54	36	54	Untargeted	[12]
Poyet	Longitudinal samples from healthy donors to the Broad Institute-OpenBiome Microbiome Library (BIO-ML).	16S	NA	83	NA	164	Untargeted	[13]
ERAWIJANTARI	Patients who underwent colonoscopy, half of which with a history of gastrectomyfor gastric cancer and no signs of gastric cancer recurrence.	WGS	42	54	42	54	Targeted	[14]
FRANZOSA	Inflammatory bowel disease patients and controls (PRISM cohort + A validation cohort).	WGS	164	56	164	56	Untargeted	[15]
MARS	Longitudinal samples (over 6 months) from patients with Irritable Bowel Syndrome and controls.	WGS	51	24	305	139	Targeted	[16]
WANG	Adults with end-stage renal disease (ESRD) and controls.	WGS	220	67	220	67	Untargeted	[17]
YACHIDA	Patients who underwent colonoscopy, with findings from normal to stage 4 colorectal cancer.	WGS	220	127	220	127	Targeted	[18]

Table 3: LOCATE's hyperparameters used

	BGU	**He**	**Jacob**	**Kim**	**Poyet**
Activation function	Tanh	elU	Tanh	Tanh	elU
Dropout	0.002	0.070	0.209	0.002	0.079
Weight decay	0.127	0.030	0.138	0.120	0.020
Learning rate	0.001	0.001	0.05	0.001	0.001
Number of neurons layer1	90	20	20	90	20
Number of neurons layer 2	80	10	30	80	10
Representation size	10	10	10	10	10
Optimizer	Adam	Adam	Adam	Adam	Adam
Max epochs	1000	1000	1000	1000	1000

Table 4: Metadata of each cohort

Dataset	**Metadata used**
Direct Plus	Diet, sex, height
He	Diet, age, sex
Jacob	Sex, age,pedigree
Poyet	Travel abroad last year, seasonal Pollen allergy, weight, height, BMI country birth, sex, relationship status, pet allergy, diet, age
Kim	Age, sex, race, smoking history
ERAWIJANTARI	Smoking status, lung cancer, drinking status, breast cancer, glucose, liver cancer, total cholesterol, diabetes med, analgesic, anticoagulant, gastric acid medication, high blood pressure, uterine cancer, sex, alcohol consumption, age
FRANZOSA	Age, antibiotic, immunosuppressant, mesalamine, steroids,
WANG	Age, BMI, Creatinine , Urea, eGFR, sex
MARS	Age, BMI, sex, antibiotics
YACHIDA	Age, sex, BMI, alcohol

Table 5: WGS 4 different clusters cross-datasets WGS

Pair	Cluster num	Color
s__Pauljensenia turicensis-C5H11NO2	1	lightgrey
s__Collinsella sp900551195-C6H13NO2	1	lightgrey
s__Collinsella sp900551605-C6H13NO2	1	lightgrey
s__Collinsella sp900759435-C4H4N2O2	1	lightgrey
s__Eggerthella lenta-C6H13NO2	1	lightgrey
s__Eggerthella sp014287365-C6H13NO2	1	lightgrey
s__Prevotella sp000431975-C4H4N2O2	1	lightgrey
s__Alistipes sp002428825-C4H4N2O2	1	lightgrey
s__Alistipes sp900021155-C4H4N2O2	1	lightgrey
s__Tidjanibacter inops_A-C4H4N2O2	1	lightgrey
s__Confluentibacter sp003258355-C5H4N4O	1	lightgrey
s__Clostridium saudiense-C6H13NO2	1	lightgrey
s__Clostridium sp900543325-C6H13NO2	1	lightgrey
s__Acetatifactor sp002431915-C4H4N2O2	1	lightgrey
s__Acetatifactor sp900771995-C4H4N2O2	1	lightgrey
s__Acetatifactor sp900772845-C4H4N2O2	1	lightgrey
s__Coprococcus sp900548315-C4H4N2O2	1	lightgrey
s__Lachnospira sp900547255-C4H4N2O2	1	lightgrey
s__UBA11774 sp003507655-C3H5O3-	1	lightgrey
s__UBA7182 sp002491115-C4H4N2O2	1	lightgrey
s__Acutalibacter sp009936035-C4H4N2O2	1	lightgrey
s__Acutalibacter sp900543305-C4H4N2O2	1	lightgrey
s__Ruminococcus_E sp900315195-C6H14N2O2	1	lightgrey
s__Ruminococcus_E sp902797655-C6H14N2O2	1	lightgrey
s__UBA737 sp900554525-C4H4N2O2	1	lightgrey
s__CAG-170 sp000432135-C4H4N2O2	1	lightgrey
s__Dysosmobacter sp900752075-C4H4N2O2	1	lightgrey
s__UBA5446 sp900543085-C4H4N2O2	1	lightgrey
s__Ruminococcus sp900540005-C4H4N2O2	1	lightgrey
s__CAG-145 sp900545135-C4H4N2O2	1	lightgrey
s__Emergencia sp900551775-C4H4N2O2	1	lightgrey
s__NSJ-50 sp014385105-C4H4N2O2	1	lightgrey
s__UBA2862 sp902790525-C3H7NO2	1	lightgrey
s__Christensenella massiliensis-C26H43NO6	1	lightgrey
s__UBA2897 sp002350105-C6H14N2O2	1	lightgrey
s__Fusobacterium_A sp900015295-C3H7NO2	1	lightgrey
s__D16-34 sp009911635-C3H5O2-	2	darkgrey
s__Alistipes sp002428825-C5H4N4O	2	darkgrey
s__Alistipes sp900549305-C5H4N4O	2	darkgrey
s__Parabacteroides sp011038785-C4H4N2O2	2	darkgrey
s__RC9 sp900546445-C5H4N4O	2	darkgrey
s__Streptococcus hyointestinalis-C6H13NO2	2	darkgrey
s__Streptococcus parasanguinis_A-C6H13NO2	2	darkgrey
s__Streptococcus parasanguinis_B-C6H13NO2	2	darkgrey
s__Streptococcus parasanguinis_C-C6H13NO2	2	darkgrey

Table 5: WGS 4 different clusters cross-datasets WGS

Pair	Cluster num	Color
s__Streptococcus parasanguinis_D-C6H13NO2	2	darkgrey
s__Streptococcus sp000314795-C6H13NO2	2	darkgrey
s__Streptococcus sp000448565-C6H13NO2	2	darkgrey
s__Streptococcus sp900543065-C6H13NO2	2	darkgrey
s__Streptococcus sp900766505-C6H13NO2	2	darkgrey
s__UBA9502 sp004554205-C5H4N4O	2	darkgrey
s__NSJ-32 sp014384895-C5H4N4O	2	darkgrey
s__CAG-110 sp003525905-C5H4N4O	2	darkgrey
s__CAG-83 sp900545585-C5H4N4O	2	darkgrey
s__Dysosmobacter sp001916835-C5H4N4O	2	darkgrey
s__ER4 sp900552015-C5H4N4O	2	darkgrey
s__Flavonifractor sp900549795-C5H4N4O	2	darkgrey
s__HGM12998 sp900756495-C5H4N4O	2	darkgrey
s__Intestinimonas butyriciproducens-C5H4N4O	2	darkgrey
s__Intestinimonas massiliensis-C5H4N4O	2	darkgrey
s__UBA3855 sp902783005-C3H5O2-	2	darkgrey
s__UMGS1889 sp900556055-C5H4N4O	2	darkgrey
s__Emergencia sp900551775-C26H43NO6	2	darkgrey
s__Phil1 sp001940855-C5H4N4O	2	darkgrey
s__UMGS692 sp900544545-C3H5O2-	2	darkgrey
s__Firm-10 sp001603025-C3H5O2-	2	darkgrey
s__HGM11575 sp002068815-C3H5O2-	2	darkgrey
s__UBA2862 sp900315585-C3H5O2-	2	darkgrey
s__UBA2862 sp900318045-C3H5O2-	2	darkgrey
s__UBA2862 sp902798105-C3H5O2-	2	darkgrey
s__QALW01 sp003150515-C26H43NO6	2	darkgrey
s__Akkermansia sp004167605-C3H5O2-	2	darkgrey
s__Porphyromonas sp900539155-C5H11NO2	3	dimgrey
s__Alistipes putredinis-C5H11NO2	3	dimgrey
s__Alistipes senegalensis-C5H11NO2	3	dimgrey
s__Alistipes shahii-C5H11NO2	3	dimgrey
s__Alistipes sp900021155-C5H11NO2	3	dimgrey
s__Alistipes sp900541585-C5H11NO2	3	dimgrey
s__Alistipes_A indistinctus-C5H11NO2	3	dimgrey
s__UBA940 sp900768115-C5H11NO2	3	dimgrey
s__W3P20-009 sp004552385-C5H11NO2	3	dimgrey
s__NSJ-32 sp014384895-C5H11NO2	3	dimgrey
s__Acutalibacter timonensis-C5H11NO2	3	dimgrey
s__NSJ-40 sp014384705-C5H11NO2	3	dimgrey
s__UMGS856 sp900760305-C5H11NO2	3	dimgrey
s__CAG-390 sp000437015-C5H11NO2	3	dimgrey
s__CAG-390 sp900753295-C5H11NO2	3	dimgrey
s__CAG-841 sp000437375-C5H11NO2	3	dimgrey
s__HGM12650 sp900761725-C5H11NO2	3	dimgrey
s__UMGS1002 sp900547565-C5H11NO2	3	dimgrey

Table 5: WGS 4 different clusters cross-datasets WGS

Pair	Cluster num	Color
s__UMGS1696 sp900763885-C5H11NO2	3	dimgrey
s__SFLA01 sp004553575-C5H11NO2	3	dimgrey
s__UCG-010 sp900754535-C5H11NO2	3	dimgrey
s__CAG-103 sp900317855-C5H11NO2	3	dimgrey
s__CAG-110 sp900546415-C5H11NO2	3	dimgrey
s__CAG-110 sp900548795-C5H11NO2	3	dimgrey
s__CAG-110 sp900551495-C5H11NO2	3	dimgrey
s__CAG-110 sp900554625-C5H11NO2	3	dimgrey
s__CAG-110 sp900769995-C5H11NO2	3	dimgrey
s__CAG-170 sp000436735-C5H11NO2	3	dimgrey
s__CAG-170 sp002437575-C5H11NO2	3	dimgrey
s__CAG-170 sp900548625-C5H11NO2	3	dimgrey
s__CAG-83 sp900548615-C5H11NO2	3	dimgrey
s__Flavonifractor massiliensis_A-C5H11NO2	3	dimgrey
s__Marseille-P3106 sp900169975-C5H11NO2	3	dimgrey
s__Pseudoflavonifractor sp900079765-C5H11NO2	3	dimgrey
s__Anaerotruncus rubiinfantis-C5H11NO2	3	dimgrey
s__Anaerotruncus sp014385085-C5H11NO2	3	dimgrey
s__Massilimaliae massiliensis-C5H11NO2	3	dimgrey
s__UBA1394 sp002305725-C5H11NO2	3	dimgrey
s__BX12 sp009911365-C5H11NO2	3	dimgrey
s__CAG-145 sp900545135-C5H11NO2	3	dimgrey
s__CAG-238 sp002439735-C5H11NO2	3	dimgrey
s__Phil1 sp001940855-C5H11NO2	3	dimgrey
s__Firm-11 sp900548145-C5H11NO2	3	dimgrey
s__SFFS01 sp004557805-C5H11NO2	3	dimgrey
s__NSJ-63 sp014384805-C5H11NO2	3	dimgrey
s__Alistipes onderdonkii-C6H13NO2	4	k
s__Alistipes sp002358415-C6H13NO2	4	k
s__Alistipes sp002362235-C6H13NO2	4	k
s__Alistipes sp900290115-C6H13NO2	4	k
s__Alistipes sp900541585-C6H13NO2	4	k
s__Alistipes sp902388705-C6H13NO2	4	k
s__UMGS2068 sp900769635-C6H13NO2	4	k
s__Anaerofustis stercorihominis-C6H13NO2	4	k
s__Blautia_A luti-C3H5O3-	4	k
s__Blautia_A sp900540785-C4H4N2O2	4	k
s__CAG-317 sp011960265-C4H4N2O2	4	k
s__HGM11523 sp900756545-C6H13NO2	4	k
s__Agathobaculum sp900291975-C6H13NO2	4	k
s__Lawsonibacter sp900066825-C6H13NO2	4	k
s__Lawsonibacter sp900764755-C6H13NO2	4	k
s__Anaerotruncus colihominis-C6H13NO2	4	k
s__UBA1409 sp002305045-C24H40O5	4	k
s__BX12 sp009911365-C6H13NO2	4	k

Table 5: WGS 4 different clusters cross-datasets WGS

Pair	Cluster num	Color
s__CAG-145 sp900545135-C6H13NO2	4	k
s__Emergencia sp009935805-C6H13NO2	4	k
s__Mogibacterium timidum-C4H4N2O2	4	k
s__RUG100 sp900315555-C6H13NO2	4	k
s__Fusobacterium_A varium-C4H4N2O2	4	k

Table 6: Acronym table

Acronym	**Meaning**
LOCATE	Latent Of miCrobiome And meTabolites rElations
ML	Machine Learning
SCFA	Short Chain Fatty Acids
T1D	Type 1 Diabetes
IBD	Inflammatory bowel disease
T2D	Type 2 Diabetes
DNN	Deep Neural Networks
CNN	Convolutional neural networks
PRMT	Predicted Relative Metabolomic Turnover
MIMOSA	Model-based Integration of Metabolite Observations and Species Abundance
MLPNN	multiple-layer perceptron neural network
WGS	whole genome shotgun sequencing
HDG	healthy dietary guidelines
MRS	magnetic resonance spectroscopy
DSC	deep subcutaneous
SSC	superficial subcutaneous
VAT	Visceral adipose tissue
CD	Crohn's disease
UC	Ulcerative colitis
ESRD	end-stage renal disease
NMF	Non Negative Matrix Factorization
NNI	Neural Network Intelligence
MSE	Mean Square Error
SCC	Spearman Correlation Coefficient
AUC	Area Under the ROC Curve
CCA	Canonical-Correlation Analysis
SVD	Singular value decomposition
CRC	Colorectal cancer

References

[1] Roktaek Lim, Josephine Jill T. Cabatbat, Thomas L.P. Martin, Haneul Kim, Seunghyeon Kim, Jaeyun Sung, Cheol-Min Ghim, and Pan-Jun Kim. Large-scale metabolic interaction network of the mouse and human gut microbiota. *Scientific Data*, 7(1):1–8, 2020.

[2] Peter E. Larsen, Frank R. Collart, Dawn Field, Folker Meyer, Kevin P. Keegan, Christopher S. Henry, John McGrath, John Quinn, and Jack A. Gilbert. Predicted relative metabolomic turnover (prmt): determining metabolic turnover from a coastal marine metagenomic dataset. *Microbial Informatics and Experimentation*, 1(1):1–11, 2011.

[3] Cecilia Noecker, Alexander Eng, Sujatha Srinivasan, Casey M. Theriot, Vincent B. Young, Janet K. Jansson, David N. Fredricks, and Elhanan Borenstein. Metabolic model-based integration of microbiome taxonomic and metabolomic profiles elucidates mechanistic links between ecological and metabolic variation. *MSystems*, 1(1):e00013–15, 2016.

[4] Cecilia Noecker, Alexander Eng, Efrat Muller, and Elhanan Borenstein. Mimosa2: a metabolic network-based tool for inferring mechanism-supported relationships in microbiome-metabolome data. *Bioinformatics*, 38(6):1615–1623, 2022.

[5] Xiaochen Yin, Tomer Altman, Erica Rutherford, Kiana A. West, Yonggan Wu, Jinlyung Choi, Paul L. Beck, Gilaad G. Kaplan, Karim Dabbagh, Todd Z. DeSantis, et al. A comparative evaluation of tools to predict metabolite profiles from microbiome sequencing data. *Frontiers in Microbiology*, 11:3132, 2020.

[6] Himel Mallick, Eric A. Franzosa, Lauren J. Mclver, Soumya Banerjee, Alexandra Sirota-Madi, Aleksandar D. Kostic, Clary B. Clish, Hera Vlamakis, Ramnik J. Xavier, and Curtis Huttenhower. Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. *Nature Communications*, 10(1):1–11, 2019.

[7] Derek Reiman, Brian T. Layden, and Yang Dai. Mimenet: Exploring microbiome-metabolome relationships using neural networks. *PLOS Computational Biology*, 17(5):e1009021, 2021.

[8] Vuong Le, Thomas P. Quinn, Truyen Tran, and Svetha Venkatesh. Deep in the bowel: highly interpretable neural encoder-decoder networks predict gut metabolites from gut microbiome. *BMC Genomics*, 21(4):1–15, 2020.

[9] Anat Yaskolka Meir, Ehud Rinott, Gal Tsaban, Hila Zelicha, Alon Kaplan, Philip Rosen, Ilan Shelef, Ilan Youngster, Aryeh Shalev, Matthias Blüher, et al. Effect of green-mediterranean diet on intrahepatic fat: the direct plus randomised controlled trial. *Gut*, 70(11):2085–2095, 2021.

[10] Minsuk Kim, Emily Vogtmann, David A. Ahlquist, Mary E. Devens, John B Kisiel, William R. Taylor, Bryan A. White, Vanessa L. Hale, Jaeyun Sung, Nicholas Chia, et al. Fecal metabolomic signatures in colorectal adenoma patients are associated with gut microbiota and early events of colorectal cancer pathogenesis. *MBio*, 11(1):e03186–19, 2020.

[11] Xuan He, Mariana Parenti, Tove Grip, Bo Lönnerdal, Niklas Timby, Magnus Domellöf, Olle Hernell, and Carolyn M Slupsky. Fecal microbiome and metabolome of infants fed bovine mfgm supplemented formula or standard formula with breast-fed infants as reference: a randomized controlled trial. *Scientific Reports*, 9(1):1–14, 2019.

[12] Jonathan P Jacobs, Maryam Goudarzi, Namita Singh, Maomeng Tong, Ian H. McHardy, Paul Ruegger, Miro Asadourian, Bo-Hyun Moon, Allyson Ayson, James Borneman, et al. A disease-associated microbial and metabolomics state in relatives of pediatric inflammatory bowel disease patients. *Cellular and Molecular Gastroenterology and Hepatology*, 2(6):750–766, 2016.

[13] Mathilde Poyet, Mathieu Groussin, Sean M. Gibbons, J. Avila-Pacheco, Xiaofang Jiang, Sean M. Kearney, A.R. Perrotta, B. Berdy, S. Zhao, T.D. Lieberman, et al. A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nature Medicine*, 25(9):1442–1452, 2019.

[14] Pande Putu Erawijantari, Sayaka Mizutani, Hirotsugu Shiroma, Satoshi Shiba, Takeshi Nakajima, Taku Sakamoto, Yutaka Saito, Shinji Fukuda, Shinichi Yachida, and Takuji Yamada. Influence of gastrectomy for gastric cancer treatment on faecal microbiome and metabolome profiles. *Gut*, 69(8):1404–1415, 2020.

[15] Eric A. Franzosa, Alexandra Sirota-Madi, Julian Avila-Pacheco, Nadine Fornelos, Henry J. Haiser, Stefan Reinker, Tommi Vatanen, A Brantley Hall, Himel Mallick, Lauren J. McIver, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature Microbiology*, 4(2):293–305, 2019.

[16] Ruben AT Mars, Yi Yang, Tonya Ward, Mo Houtti, Sambhawa Priya, Heather R Lekatz, Xiaojia Tang, Zhifu Sun, Krishna R. Kalari, Tal Korem, et al. Longitudinal multi-omics reveals subset-specific mechanisms underlying irritable bowel syndrome. *Cell*, 182(6):1460–1473, 2020.

[17] Xifan Wang, Songtao Yang, Shenghui Li, Liang Zhao, Yanling Hao, Junjie Qin, Lian Zhang, Chengying Zhang, Weijing Bian, LI Zuo, et al. Aberrant gut microbiota alters host metabolome and impacts renal failure in humans and rodents. *Gut*, 69(12):2131–2142, 2020.

[18] Shinichi Yachida, Sayaka Mizutani, Hirotsugu Shiroma, Satoshi Shiba, Takeshi Nakajima, Taku Sakamoto, Hikaru Watanabe, Keigo Masuda, Yuichiro Nishimoto, Masaru Kubo, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nature Medicine*, 25(6):968–976, 2019.