Meta-analysis of the human gut microbiome uncovers shared microbial signatures between diseases

Dong-Min Jin[1], James T. Morton[2], Richard Bonneau[3]

[1]Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY [2]Flatiron Institute, Simons Foundation, New York, NY [3]Prescient Design, Genentech, New York, NY

CONTACTS

dj2080@nyu.edu

ABSTRACT

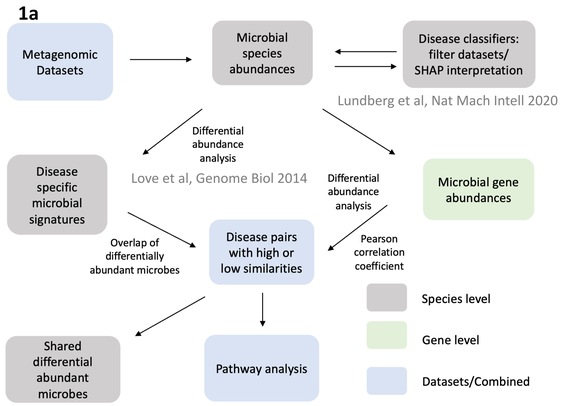
Microbiome studies have shed light on the potential roles of gut microbiota on human health and diseases. However, many studies focus on a single disease within a cohort. Previous cancer meta-analysis showed increased power due to the increased number of samples, it was able to find robust microbials associations across cohorts. Understanding the relationships between diseases is critical for understanding disease similarities, which is important for treatment development. We developed a meta-analysis workflow to analyze gut microbiome profiles, and used data from 12 studies covering 10 diseases, including samples (n = 1258) from healthy controls and patients with disorders ranging from neurological, autoimmune, to metabolic and gastrointestinal. Differential abundance analysis found diseases show similarities at the microbial species or gene level. Our results demonstrate that understanding complex diseases in the context of population heterogeneity is key to improving the specificity of reported differentially abundant microbes and metabolic pathways.

OBJECTIVES

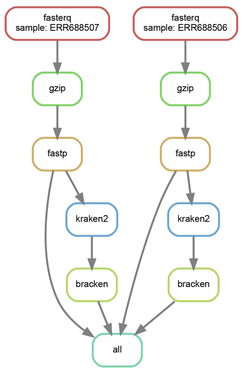
- Associations between complex human diseases at the microbial level
- What diseases are similar the microbial species level?
- What diseases are similar the microbial gene level?

MATERIALS & METHODS

1a



1b



1c

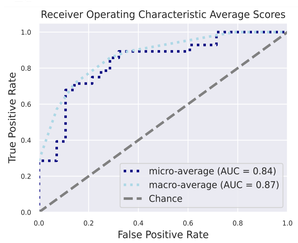


Fig. 1 The overall design and data analysis pipeline.
a Overall workflow. **b** Snakemake pipeline for processing metagenomic data.
c Binary classifier AUC.

RESULTS

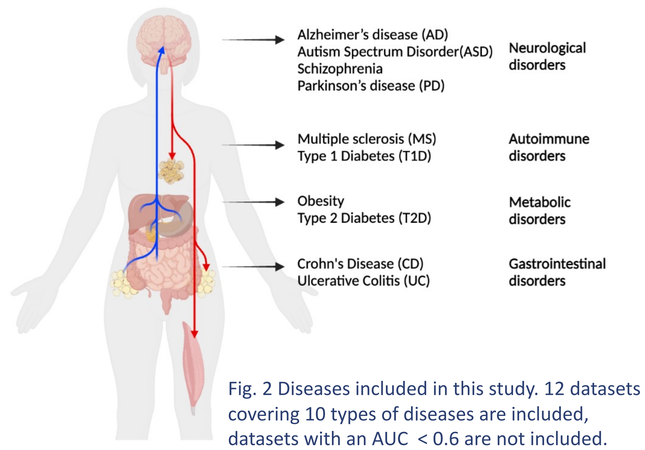


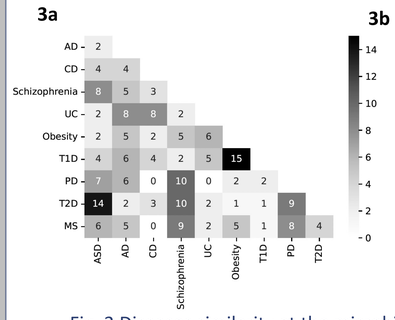
Fig. 2 Diseases included in this study. 12 datasets covering 10 types of diseases are included, datasets with an AUC < 0.6 are not included.

Table 1 Datasets and AUC for classifiers.

Datasets	AUC of binary classifier	AUC of multi-class classifier
AD	0.66	0.97
ASD_Dan	0.65	0.97
ASD_Wang	1	0.97
ASD_Wan	0.62	0.97
CD	0.9	0.98
MS	0.83	0.99
Obesity	0.8	0.95
PD	0.98	0.99
Schizophrenia	0.71	0.99
T1D	0.88	0.99
T2D	0.77	0.98
UC	0.87	0.99

- Multi-class classifier showed improved classification accuracy.

3a



3b

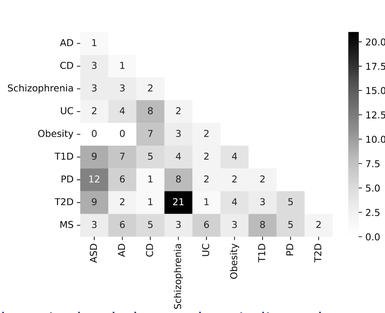


Fig. 3 Diseases similarity at the microbial species level, the numbers indicate the overlap of differential abundant microbes among the top 100 microbes. **a** overlap of disease-associated microbes **b** overlap of healthy-associated microbes.

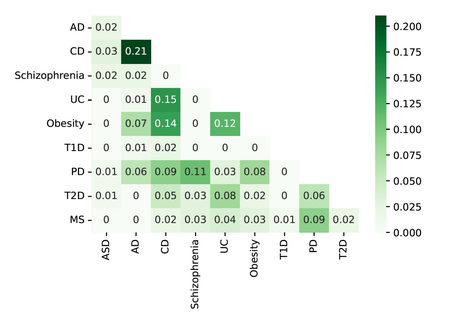
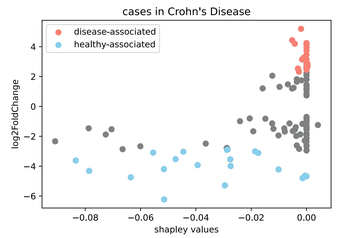
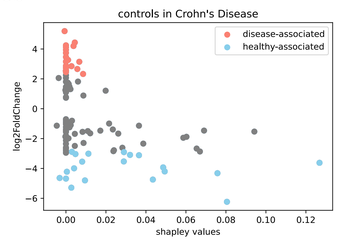


Fig. 4 Disease similarity at the microbial gene level, the numbers represent the R^2 values.

5a



5b



5c

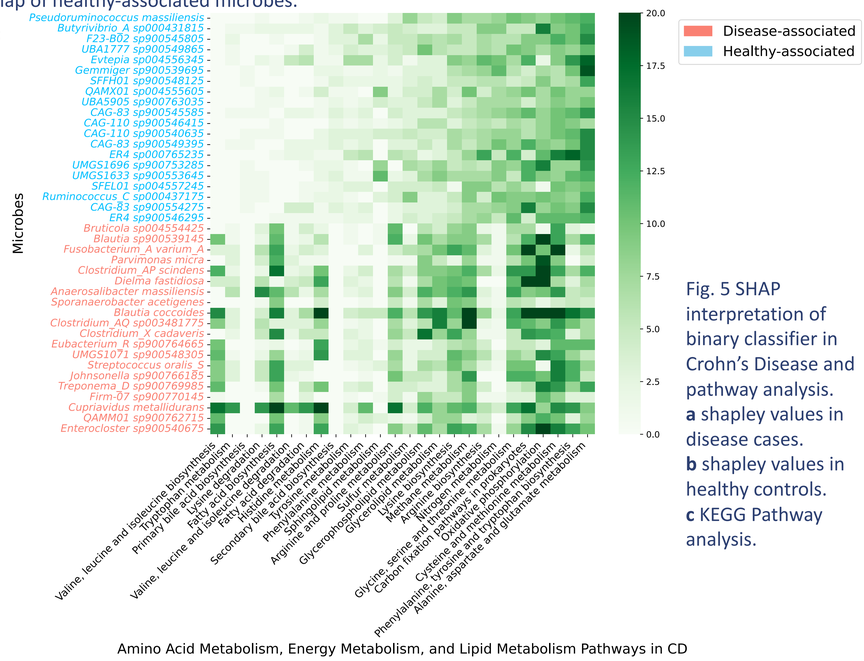
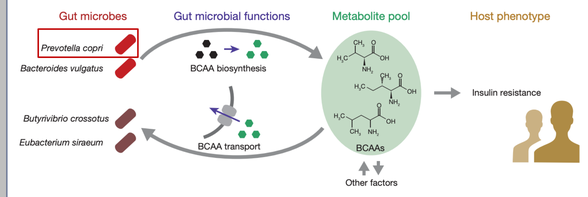


Fig. 5 SHAP interpretation of binary classifier in Crohn's Disease and pathway analysis.
a shapley values in disease cases.
b shapley values in healthy controls.
c KEGG Pathway analysis.

Amino Acid Metabolism, Energy Metabolism, and Lipid Metabolism Pathways in CD

RELEVANT STUDIES



Pedersen et al, Nature 2016

- *Prevotella copri* can induce insulin resistance.
- The expansion of *Prevotella copri* and other *Prevotella* spp.correlates with enhanced susceptibility to Rheumatoid arthritis. Alpizar-Rodriguez et al, Ann. Rheum. Dis 2019
- The prevalence of **T2D** in patients with **Schizophrenia** is about 2 to 3 times higher than in the general population. Mizuki et al, Int. J. Neuropsychopharmacol 2021
- Population-based cohort studies found that **Alzheimer's disease** occurs more frequently in patients with **Crohn's Disease.** Zhang et al, Gut 2021
- **Crohn's Disease** and **Ulcerative Colitis** are two subtypes of inflammatory bowel diseases. Pascal et al. Gut 2017

CONCLUSIONS

- Multi-class disease classifier has the potential for non-invasive diagnosis for many complex human disease.
- The increased abundances of *Prevotella copri* and other *Prevotella* spp. in disease cases contributes to the similarity between Obesity and T1D.
- The decreased abundance of microbes from the genus *Lachnospira* and *Haemophilus* in disease cases contribute to the similarity between Schizophrenia and T2D.
- The top two disease pairs showed high similarity at microbial gene level are Crohn's Disease (CD) vs Alzheimer's disease (AD), CD vs Ulcerative Colitis (UC).
- Pathway analysis found that disease-specific differential abundant microbes can be involved to affect the host in the same metabolic pathways.
- Pathways showed higher incidence in disease-associated microbes unique to CD and AD include Tryptophan metabolism and Lysine degradation, while Fatty acid biosynthesis is unique to CD and UC.