Meta-analysis of the human gut microbiome

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

[1]Center for Genomics and Systems Biology, Department of Biology, New York University,
[2]Biostatistics & Bioinformatics Branch, NICHD, NIH,
[3]Prescient Design, Genentech

Many clinical studies showed associations between disease and the human gut microbiome [1], whereas the understanding of the role that microbiota plays in disease remains limited.

Motivation

- Meta-analysis combines has increased power due to the increased number of cases.
- Standardized processing and analysis methods make compare microbial signatures of samples across studies easier.

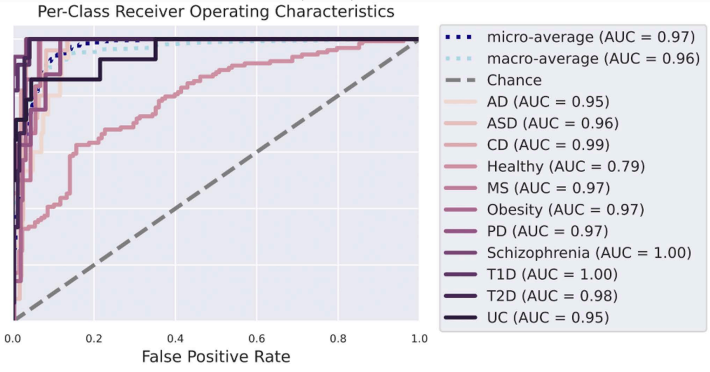
Methods

- Curate shotgun metagenomic datasets.
- Build a pipeline that can process data consistently.
- Process datasets with the pipeline.
- Build per-disease classifiers/multi-class disease classifier.
- Interpret classifiers with Shapley additive explanation (SHAP) [2].

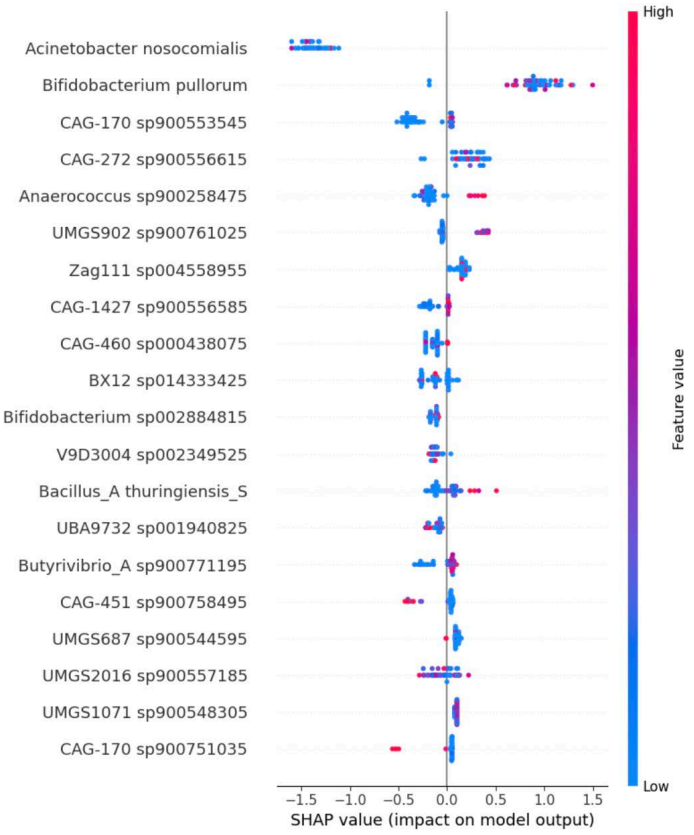
Results

- Datasets curated: 12 datasets covering 10 diseases:

Disease	Case	Control
Alzheimer's Disease(AD)	75	75
Autism Spectrum Disorder(ASD)	125	125
Crohn's Disease(CD)	54	54
Multiple Sclerosis(MS)	30	30
Obesity	36	36
Parkinson's Disease(PD)	40	40
Schizophrenia	81	81
Type 1 Diabetes(T1D)	53	53
Type 2 Diabetes(T2D)	76	76
Ulcerative Colitis (UC)	59	59



- Training a multi-class disease classifier with combined dataset boosted classification accuracy by 20% on average compared to per-disease classifiers.



- Shapley values from the gradient boosting classifier of ASD. Each dot corresponds to one individual from the datasets. The dot's position on the *x* axis shows the impact that feature has on the model's prediction for that individual. Several species from *Bifidobacterium*, *Anaerococcus*, and *Bacillus*, are of great influence for some individuals.

References

[1] Schroeder, B.O. and Bäckhed, F. (2016) 'Signals from the gut microbiota to distant organs in physiology and disease', Nature medicine, 22(10), pp. 1079–1089.

[2] Lundberg, S.M. et al. (2020) 'From local explanations to global understanding with explainable AI for trees', Nature Machine Intelligence, pp. 56–67. Available at: https://doi.org/10.1038/s42256-019-0138-9.

Contact me at dj2080@nyu.edu