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

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

[1] Center for Genomics and Systems Biology, New York University, New York, NY, USA
[2] Center for Computational Biology, Flatiron Institute, Simons Foundation, New York, NY, USA
[3] Genentech, New York, NY, USA

CONTACT: dj2080@nyu.edu

ABSTRACT

Microbiome studies have revealed gut microbiota's potential impact on complex diseases. However, many studies often focus on one disease per cohort. We developed a meta-analysis workflow for gut microbiome profiles and analyzed shotgun metagenomic data covering 11 diseases. Using interpretable machine learning and differential abundance analysis, our findings reinforce the generalization of binary classifiers for Crohn's disease (CD) and colorectal cancer (CRC) to hold-out cohorts and highlight the key microbes driving these classifications. We identified high microbial similarity in disease pairs like CD vs ulcerative colitis (UC), CD vs CRC, Parkinson's disease vs type 2 diabetes (T2D), and schizophrenia vs T2D. We also found strong inverse correlations in Alzheimer's disease vs CD and UC. These findings detected by our pipeline provide valuable insights into these diseases.

OBJECTIVES

Our goal is to provide a computational pipeline that can measure disease similarity based on microbiome composition.

METHODS

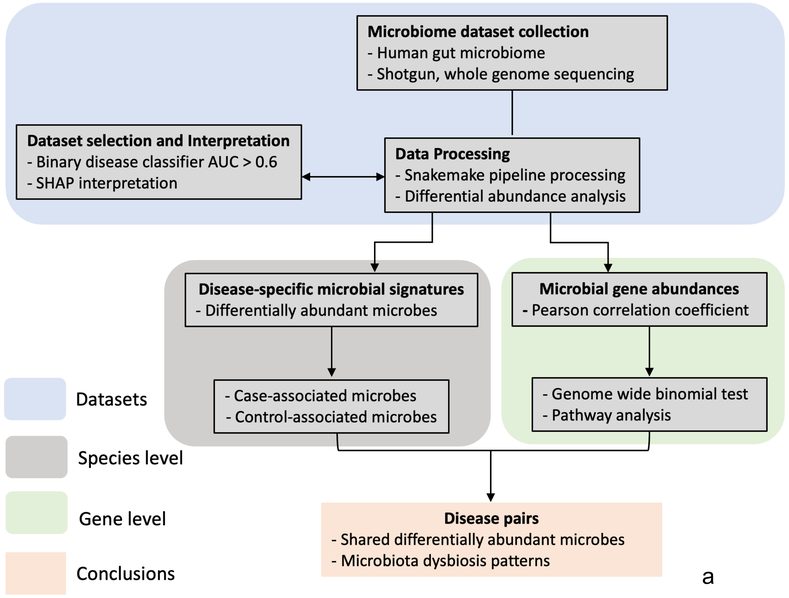


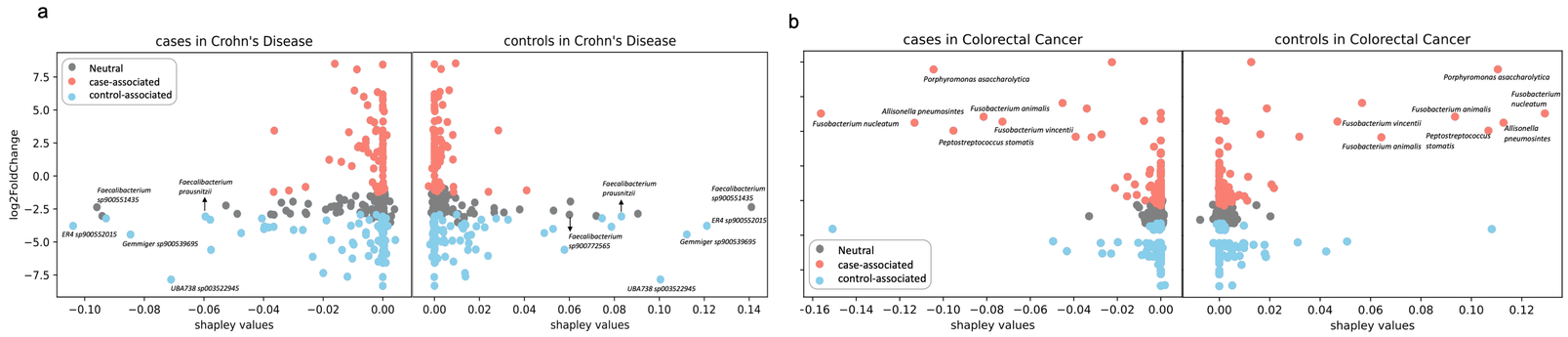
Fig 1. The overall design and data analysis pipeline

Diseases included:

- Alzheimer's disease (AD)
- Autism spectrum disorder (ASD)
- Schizophrenia
- Parkinson's disease (PD)
- Crohn's disease (CD)
- Ulcerative colitis (UC)
- Colorectal cancer (CRC)
- Multiple sclerosis (MS)
- Type 1 diabetes (T1D)
- Obesity
- Type 2 diabetes (T2D)

RESULTS & CONCLUSIONS

We compared the shared microbial signature between CD and CRC first as a sanity check: Population-based cohort studies have found that CD is a risk factor for CRC. We asked here: what are the shared microbial signatures between CD and CRC?



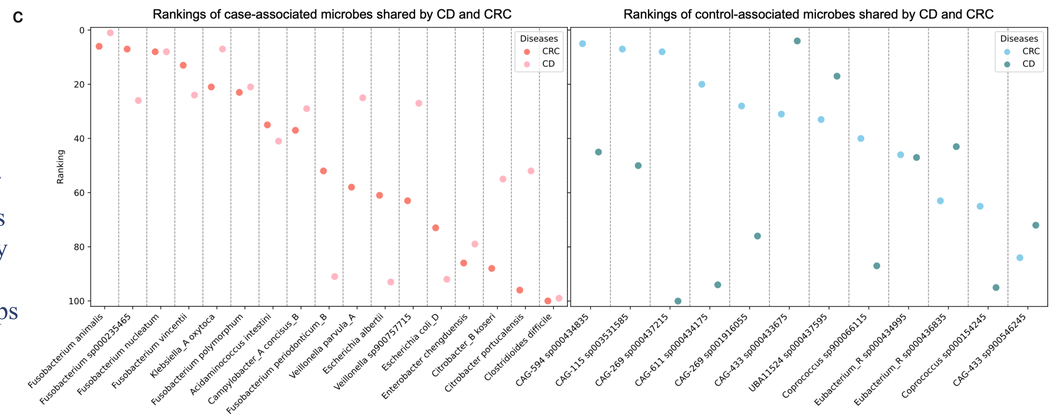


Fig 2. Interpretation of binary classifiers and differentially abundant microbes overlaps in CD and CRC.

- Case-associated microbes contribute more to classify CD cases from controls (Fig. 2a).
- Control-associated microbes contributed more to classify CRC cases from controls (Fig. 2b).
- Overlapped case-associated microbes between CD and CRC contributed more to the similarities of these two diseases, such as *Fusobacterium spp.* and *Veillonella spp* (Fig. 2c).

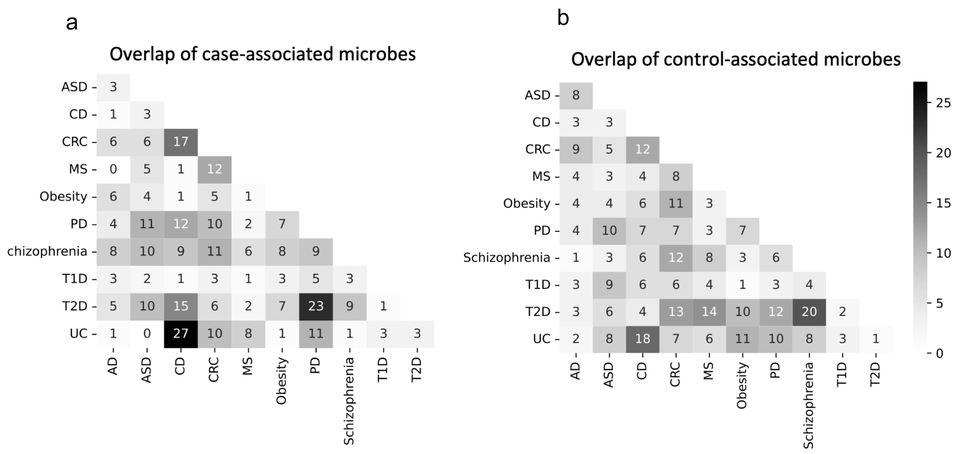


Fig 3. Disease similarity at the microbial species level.

- Top disease-pairs exhibiting the most significant overlap of case-associated microbes: CD vs UC, PD vs T2D (Fig. 3a).
- Top disease-pairs exhibiting the most significant overlap of control-associated microbes: Schizophrenia vs T2D, CD vs UC (Fig. 3b).

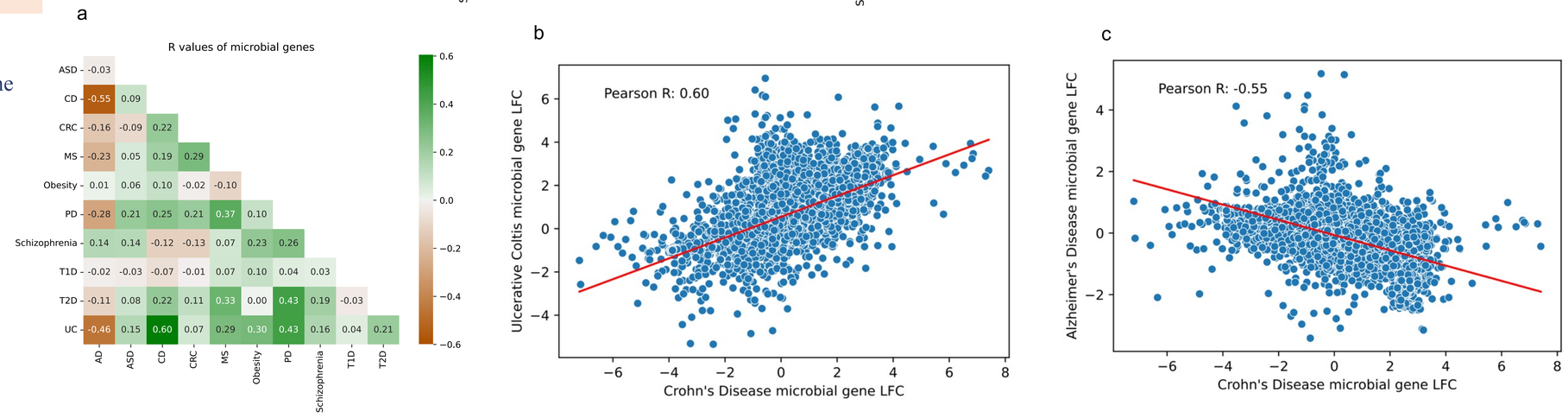


Fig. 4 Disease similarity at the microbial gene level.

- There is consistency between the similarity observed at the microbial species level and that at the microbial gene level (Fig. 4a).
- The R value for CD vs UC stands at 0.6, representing the highest positive correlation observed across all disease pairs (Fig. 4b).
- AD has a strong negative correlation between differential gene abundances with CD (R=-0.55) (Fig. 4c)