

UC San Diego

2023 UC San Diego Undergraduate Library Research Prize

Title

Diagnosis of Inflammatory Bowel Disease Using a Robust 14 Species
Signature

Permalink

<https://escholarship.org/uc/item/8zh6b7k9>

Authors

Li, Sky

Boland, Brigid S.

Uzelac, Matthew

et al.

Publication Date

2023

Diagnosis of Inflammatory Bowel Disease Using A Robust 14 Species Signature

Yuxiang Li 1,2, †, Brigid S. Boland, †, Matthew Uzelac 1,2, †, Wei Tse Li 1,2,4, Jaideep Chakladar 1,2, Michael Andrew Yu, Jessica Wang-Rodriguez 2,6,7, Weg M. Ongko 1,2,*

1. Division of Otolaryngology-Head and Neck Surgery, Department of Surgery, University of California, San Diego, La Jolla, CA, USA
2. Research Service, VA San Diego Healthcare System, San Diego, California, CA, USA
3. Division of Gastroenterology, Department of Medicine, University of California, San Diego, CA, USA
4. School of Medicine, University of California, San Francisco, CA, USA
5. Division of Gastroenterology, Department of Internal Medicine, Emory University School of Medicine, Atlanta, GA, 30322, USA
6. Department of Pathology, University of California San Diego, La Jolla, CA 92093, USA.
7. Pathology Service, VA San Diego Healthcare System, San Diego, CA 92161, USA.

VA San Diego
HEALTHCARE SYSTEM

UC San Diego
SCHOOL OF MEDICINE

Abstract

Inflammatory bowel diseases (IBD) are disorders of the gastrointestinal system affecting a considerable portion of Western populations. With nearly 6.8 million individuals affected globally, the ability to accurately and efficiently diagnose the disease in patients is crucial. The gastrointestinal microbiome composition has been associated with several diseases, disorders, and cancers. Several microbial signature-based machine learning IBD predictive models have been developed but have yet to be implemented clinically due to poor external accuracy. This study analyzes the microbial compositions of 1,612 stool samples from 665 patients across seven distinct shotgun metagenomic cohorts in constructing an IBD diagnostic model capable of high predictive accuracy and robustness. We created a 14-species diagnostic ML panel and validated the performance of the model on 3 independent patient cohorts. The diagnostic model exhibited a ROCAUC score of 0.821 and a PRAUC score of 0.902. Testing the ML model against other microbiome-associated diseases, such as colorectal cancer and type 2 diabetes, suggested that the diagnostic model is specific to IBD. The model's reliance on abundance levels of only 14 biomarker species, in addition to its non-invasive nature, makes diagnosis relatively inexpensive and convenient. Furthermore, the signature is stable across different patient cohorts, sequencing pipelines, and sample preparation methods. To examine potential mechanisms by which the microbiome influences IBD, we identified 23 metabolic pathways that were consistently differentially enriched across cohorts, 2 microbes of the greatest contribution to differences in pathway abundance, and metabolites correlated to the presence of the 14 biomarker species.

Figure 1: Study Schematic and Workflow



Table 1: Table 1: Overview of Metagenomic Sequencing Databases after sample exclusion.

Cohort #	Cohort Name	Region	Year published	Age (AVG, SD)	BMI (AVG, SD)	Sex (M %, F %)	# Healthy Control	# Crohn's Disease	# Ulcerative Colitis	ENA ID
1	HMP2-IBD	North America	2019	40.21, 14.55	NA, 24.5	50.5, 49.5	72	14	35	PR000639
2	EBH	China	2017	24.5, 8.9	18.2, 13.1	84.5, 15.5	183	54	49	PRJEB15571
3	PRISM	Spain	2014	40.6, 12.4	23.9, 4.1	39.7, 60.3	141	39	13	PRJEB1220
4	PRISM	North America	2013	42.0, 17.0	NA, 24.5	NA, 50.5	155	34	68	PRJNA400072
5	UCSD IBD	North America	2021	41.3, 15.7	25.2, 5.4	50.3, 50.7	129	0	50	79
6	UMCQIBD	North America	2017	48.0, 17.8	NA, 24.5	NA, 50.5	43	0	20	23
7	LLDeep	North America	2015	41.0, 18.9	NA, 24.5	NA, 50.5	22	22	0	PRJNA400072

Figure 2: Cross-prediction matrix reporting ROCAUC and PRAUC scores obtained by training a LGBM on species-level relative abundances and functional pathways profiled from metagenomic shotgun sequencing.

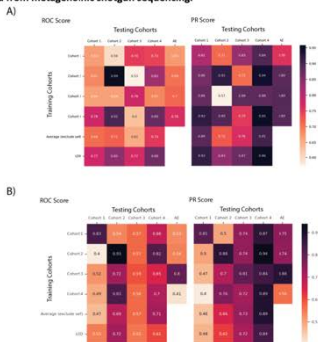
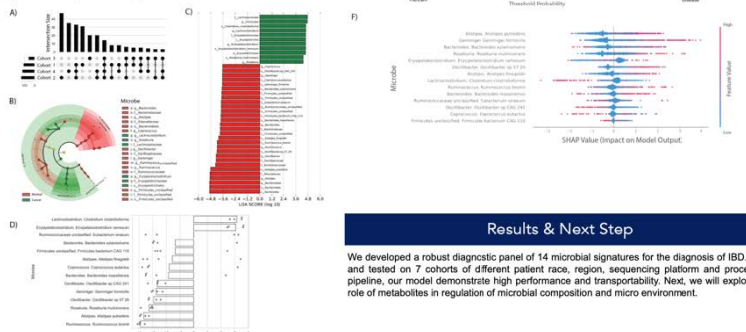


Figure 3: Differentially abundant species identified - LESe



Results & Next Step

We developed a robust diagnostic panel of 14 microbial signatures for the diagnosis of IBD. Train and tested on 7 cohorts of different patient race, region, sequencing platform and processing pipeline, our model demonstrate high performance and transportability. Next, we will explore the role of metabolites in regulation of microbial composition and micro environment.