# Taxonomic Resolution Matters for Microbiome-Based Classification of Colorectal Cancer

Courtney R. Armour, Begüm D.Topçuoğlu, Andrea Garretto, Patrick D. Schloss

Department of Microbiology and Immunology, University of Michigan, Ann Arbor

Correspondence: Patrick D. Schloss

**observation format - max 1200 words, 2 figures, 25 ref**

## Abstract (max 250 words)

## Importance (max 150 words)

Colorectal cancer is one of the most common cancers in men and women and a leading cause of cancer related deaths in the United States [@siegel2020]. Early detection and treatment are essential to increase survival rates, but for a variety of reasons including the invasiveness and high cost of screening (i.e. colonoscopy), many people do not comply with recommended screening guidelines [@garcía2011a] prompting a need for low cost, non-invasive detection methods. A growing body of research points to the gut microbiome as a promising target for non-invasive screening to detect screen relevant neoplasia (SRNs) consisting of pre-cancerous polyps (i.e. advanced adenomas) and carcinomas. Efforts to realize the diagnostic potential of the gut microbiome in detecting SRNs have focused on machine learning (ML) methods using abundances of operational taxonomic unit (OTU) classifications based on amplicon sequencing of the 16S rRNA gene {}. However, whether this is the optimal taxonomic resolution for classifying SRNs from microbiome data is unknown. Additionally, recent work has pushed for the use of amplicon sequence variants (ASVs) to replace OTUs for marker-gene analysis because of the improved resolution with ASVs [@callahan2017]. However, whether the additional resolution provided by ASVs is useful for ML classification is unclear {}. Since ML classification relies on consistent differences between groups, its possible that the resolution at the ASV level is too individualized to accurately differentiate groups. Topçuoğlu *et al* [@topçuolu2020] recently demonstrated how to effectively apply machine learning (ML) methods to microbiome based classification problems and developed a framework for applying ML practices in a more reproducible way (mikropml). This analysis utilizes the reproducible framework developed by Topçuoğlu *et al* to determine which ML method and taxonomic level produce the best performing classifier for detecting SRNs from microbiome data.