Cross-study analysis of ulcerative colitis microbiome identifies altered bacterial strains across multiple cohorts

Jayamary Divya Ravichandar, Sunit Jain, Kathryn Iverson, Nicole Narayan, Cheryl-Emiliane Chow, Shoko Iwai, Todd DeSantis

Abstract
The gut microbiota has emerged as an important modulator in disease onset and exacerbation in inflammatory bowel disease, including ulcerative colitis (UC) and Crohn’s disease. While several studies present insight into the landscape of microbial shifts in disease, they are influenced by cohort, sequencing-technology, and data analysis methods. Further, individual studies are often under-powered to detect microbes differentially abundant in health compared to disease for therapeutic development. We compared mucosa-associated 16S-rRNA datasets from UC and healthy subjects from five individual cohorts and three different methods (NGS, PhyloChip, Sanger), a total of seven datasets. Leveraging our multi-faceted bioinformatics platform, which enables appropriate method-specific quality filtering and statistical testing to identify differentially abundant bacteria at the strain-level, we performed a random effects model-based meta-analysis on data from multiple cohorts and multiple microbiome-profiling technologies. The 23 strains we identified as concordantly depleted in ulcerative colitis present opportunities for development of therapeutic targets.

Study-wise analysis

- **Raw sequence data**
  - Sanger
  - Illumina
  - PhyloChip

- **Sequence processing**
  - StrainSelect + UPARSE
  - DiDA2 + StrainSelect
  - Sortmerna + StrainSelect

- **Differential abundance testing**
  - DESeq2
  - Limma

- Identification of datasets and data curation
- Identification of strains concordantly enriched in health across studies

Cross-study meta-analysis
- **Meta-analysis identified 23 strains significantly enriched in controls and 1 strain significantly depleted in controls.**
- Significance: Benjamini-Hochberg-corrected REM p values < 0.05.

Methods

- **Strains depleted in ulcerative colitis across multiple studies**
  - Meta-analysis identified health-associated strains that were not inferred in independent analysis of each study.

Conclusions
- Our meta-analysis pipeline is capable of integrating data sequenced on different platforms and identifying trends not inferred in independent analysis of studies.
- The bacterial strains identified here as significantly depleted in ulcerative colitis across multiple cohorts present novel therapeutic opportunities for mining proteins modulating disease outcome.