Functional Activity of the Human Gut Microbiome to Classify Colorectal Cancer

Kelly L. Sovacool
July 2020

Specific Aims

Changes in the taxonomic composition and metabolic activity of human microbiomes have been observed in several diseases. In the case of colorectal cancer (CRC), evidence of toxigenic activity by gut microbes implies that these changes are not only a response to disease, but may also play a role in disease etiology. Taxonomic composition is commonly defined by amplicon sequencing of the 16S rRNA gene and clustering sequences into Operational Taxonomic Units (OTUs). Previous studies have built OTU-based machine learning models to classify stool samples as normal or cancerous, to serve as a less invasive diagnostic tool for CRC than colonoscopy. Efforts to find consistent changes in taxonomic composition of microbiomes between normal and dysbiotic states have found mixed success, in part because interpersonal variability in taxonomic composition sometimes exceeds the variability between disease states. Variability of microbiome composition between individuals with the same disease status may be explained by functional redundancy, where different microbial species carry out the same functions and thus can replace each other with little effect on the overall function of the community.

Sequencing whole metagenomes to identify the genes present and annotate known gene functions is commonly used to build a profile of functional potential of the microbiome. Combining taxonomic composition from OTUs with functional potential from metagenomes allows one to characterize functional redundancy across communities, where communities with similar functional potential have different taxonomic composition. Untargeted mass spectrometry can validate the functional potential characterized from metagenomics by identifying metabolites that are active in a community, thus painting a more precise picture of active microbial functions. Here, I propose to investigate the impacts of taking functional redundancy and active metabolites into account on human stool sample classification for CRC diagnosis.

Aim 1. Assess the impact of functional redundancy of the gut microbiome on CRC classification.

Hypothesis: Using functional gene profiles instead of only taxonomic profiles improves the classification modeling of samples as CRC or non-cancerous because of functional redundancy in the gut microbiome.

A. Build taxonomic profiles with OTUs from 16S rRNA gene sequences and build profiles of functional gene potential from metagenomes.
B. Compare taxonomic composition to functional gene potential of microbiomes within and between disease states to determine presence and degree of functional redundancy.
C. Build machine learning models to classify samples as CRC or non-cancerous with taxonomic composition, functional gene potential profiles, or both as model features and compare performance.

Aim 2. Assess the impact of integrating active metabolites with functional gene potential on CRC classification.

Hypothesis: Using active metabolic pathways confirmed with mass spectrometry instead of all potential metabolic pathways from metagenomes improves the classification modeling of samples as CRC or non-cancerous.

A. Annotate compounds from untargeted mass spectrometry with the GNPS database and select those known to be products of bacterial metabolic pathways with the MetaCyc database.
B. Calculate the intersection of pathways associated with active metabolites and the pathways from functional potential profiles from metagenomes.
C. Build machine learning models to classify samples as CRC or non-cancerous with all potential metabolic pathways or only confirmed active metabolic pathways as model features and compare performance.

Dataset

Stool samples were collected from patients undergoing colonoscopy as part of the GLNE 007 study (https://clinicaltrials.gov/ct2/show/study/NCT00843375). 211 individuals were diagnosed with CRC and 223 were confirmed non-cancerous. 16S rRNA gene amplicon sequencing was performed and remaining stool was kept frozen. Part of the remaining stool will be used for whole metagenome shotgun sequencing and untargeted tandem mass spectrometry to complete these aims.

Background and Motivation

Significance

Research Design and Methods

Aim 1. Functional redundancy of the gut microbiome

1A) Build profiles of taxonomic composition and functional gene potential of CRC and non-cancerous samples.

1B) Determine presence and degree of functional redundancy in CRC and non-cancerous gut microbiomes.

1C) Build and compare CRC classification models with taxonomic composition and functional gene potential.

Aim 2. Integrating active metabolites with functional gene potential

2A) Annotate compounds from untargeted mass spectrometry and select known products of bacterial metabolism.

2B) Calculate the intersection of pathways from active metabolites and functional potential profiles.

2C) Build and compare CRC classification models with all potential pathways or only confirmed active pathways.

Potential Outcomes and Conclusions

References