Functional Activity of the Human Gut Microbiome to Classify Colorectal Cancer

Kelly L. Sovacool
August 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

we don't know enough about functional redundancy in the human gut microbiome [1]

many published studies claiming to have found functional redundancies in microbial systems lack quantitative analyses of redundancy [2, 3].

no one agrees on exact definition of functional redundancy [4, 1, 5, 6]. "Stability in ecosystem function with increasing microbial diversity is often considered an empirical indication of functional redundancy" [6]

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.

trait contribution evenness (TCE): "the evenness in relative contribution of that trait among taxa within the community... This definition has several appealing properties including: TCE is an extension of established diversity theory, functional redundancy measurements from communities with different richness and relative trait contribution by taxa are easily comparable, and any quantifiable trait data (genes copies, protein abundance, transcript copies, respiration rates, etc.) is suitable for analysis." [6] it's just like Shannon index if you used alpha=1.

how to quantify fcnl redundancy overall within and between disease states? pairwise comparison of TCE for all traits? for each trait, compare median TCE among cancer vs non-cancer samples? what statistical test to use?

should consider each trait a function, i.e. metabolic pathway, *not* a gene.

instead of TCE, consider using the fraction of taxa that are capable of producing the trait?

"Functional redundancy is a measure of the number of different populations within a community that are able to perform the same functions. Functional redundancy can increase functional resilience, in case perturbations affect the taxonomic community structure; this allows for a return to community function, and therefore can increase stability." [1]

"This functional redundancy is further reflected in the fraction of the observed microbial community capable of participating in each metabolic step, with no statistically significant difference between the boreholes, except for ammonia oxidation (Figure 6)." (used Student's t-test and Wilcoxon rank sum) [5] if no sig diff, then the communities are redundant for that function. count number of redundant functions?

Use TCE; compare TCE between communities just like Shannon diversity are compared in Hannigan paper. box plot of TCE for traits (potential functions / pathways) for cancer vs healthy per trait. "the statistical significance of results of comparisons between the disease state clusters was assessed using analysis of similarity (ANOSIM) with a post hoc multivariate Tukey test." [7] however, with Hannigan paper, there is 1 Shannon index per community. did he do ANOSIM on bray-curtis only, or also on Shannon div? did he do ANOSIM on differences in mean Shannon diversities? for this, there will be 1 TCE per trait per community. Do statistical test for every trait? idea: for each trait, calc mean TCE in groups (mean cancer TCE and mean non-cancer TCE), do Wilcoxon rank-sum test for significance? - similar to ANCOM? or do Wilcoxon rank-sum on proportion of community capabable of performing each function, like [5]?

Why not look at Jaccard index of traits (potential functions / pathways) to get at beta div? probably jaccard and not bray-curtis because bray takes abundance into account, but pathway presene weighted by taxa abundance might not correspond to actual abundance of gene products. that may stretch limits of functional potential. ANOSIM on pairwise jaccard dissimilarity.

How does humann2 recommend comparing fcnl potential b/e communities? Calculate diversity metrics on function abundances to find "contributional diversity". Alpha (within-sample): Gini-Simpson. Beta (between-sample): Bray-Curtis. Fig 2A in [8] plotted beta vs alpha for samples. I'd like to do ANOSIM + NMDS on Bray-Curtis distances.

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

[1] A. Heintz-Buschart and P. Wilmes. Human Gut Microbiome: Function Matters. *Trends in Microbiology*, 26(7):563–574, July 2018. PMID: 29173869.

[2] R. C. Souza, M. Hungria, M. E. Cantão, A. T. R. Vasconcelos, M. A. Nogueira, and V. A. Vicente. Metagenomic analysis reveals microbial functional redundancies and specificities in a soil under different tillage and crop-management regimes. *Applied Soil Ecology*, 86:106–112, February 2015.

[3] M. Ferrer, A. Ruiz, F. Lanza, S.-B. Haange, A. Oberbach, H. Till, R. Bargiela, C. Campoy, M. T. Segura, M. Richter, M. v. Bergen, J. Seifert, and A. Suarez. Microbiota from the distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear differences in community structure. *Environmental Microbiology*, 15(1):211–226, 2013.

[4] S. Louca, M. F. Polz, F. Mazel, M. B. N. Albright, J. A. Huber, M. I. O'Connor, M. Ackermann, A. S. Hahn, D. S. Srivastava, S. A. Crowe, M. Doebeli, and L. W. Parfrey. Function and functional redundancy in microbial systems. *Nature Ecology & Evolution*, 2(6):936–943, June 2018.

[5] B. J. Tully, C. G. Wheat, B. T. Glazer, and J. A. Huber. A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer. *The ISME Journal*, 12(1):1–16, January 2018.

[6] T. M. Royalty and A. D. Steen. A quantitative measure of functional redundancy in microbial ecosystems. *bioRxiv*, page 2020.04.22.054593, April 2020.

[7] G. D. Hannigan, M. B. Duhaime, M. T. Ruffin, C. C. Koumpouras, and P. D. Schloss. Diagnostic Potential and Interactive Dynamics of the Colorectal Cancer Virome. *mBio*, 9(6), December 2018. PMID: 30459201.

[8] E. A. Franzosa, L. J. McIver, G. Rahnavard, L. R. Thompson, M. Schirmer, G. Weingart, K. S. Lipson, R. Knight, J. G. Caporaso, N. Segata, and C. Huttenhower. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods*, 15(11):962–968, November 2018.