Developing tools for characterizing the taxonomic and functional composition of host-associated microbiomes

Kelly L. Sovacool
March 2021

intro paragraph introducing microbiome research and techniques: 16S sequencing and metabolomics.

16S tells us the taxonomic composition, i.e. which microbes are present in a sample. Can then compare taxonomic composition between individuals with different disease states (e.g. colorectal cancer). There is high interpersonal variability in taxonomic composition of the human gut microbiome. Ordination plots (e.g. NMDS) do not show any separation between disease states. ML models built on taxonomic composition do an okay job of discriminating stool samples from healthy and cancerous colons (AUROC 0.7 for RF model on 490 samples).

LC-MS/MS tells us the metabolites present – the inputs and outputs of chemical reactions. Can use metabolomics to build profiles of functional composition, i.e. what the microbes (and the host) are doing. omics integration: can we identify which microbes are performing which functions?

with metabolomics, only 2 percent of untargeted lc-ms/ms features can be annotated with known metabolites. why should we throw out that data? microbial ecology gets around this type of problem by using database-independing approaches for clustering sequences into OTUs. let's take this approach and apply it to metabolomics, cluster ms features into Operational Metabolomic Units. with metagenomics, only a small fraction of microbial genes are annotated with known functions.

aim 1: build a tool mums2 to cluster untargeted LC-MS/MS features into OMUs.

aim 2: build ML models with OTUs, OMUs, and both. hypothesis: incorporating both OMUs and OTUs improves ML model performance for CRC classification.

aim 3: functional redundancy!