**Analysis Tutorial Prospectus**

Julianne Fazekas

**1. Title**

Indole-Associated Microbial Analysis in CRC Patients Using R

**2. Research question(s)**

What is the abundance of indole-producing microbes in CRC patients compared to healthy controls? How can we use R to compare microbial compositions and visualize indole-producing species in CRC?

**3. Objective(s)**

i. Develop an R script to analyze and compare the relative abundance of indole-producing bacteria in microbiome datasets of CRC patients and healthy controls.  
ii. Use visualizations (boxplots, heatmaps, PCA) to explore microbial diversity and compare species across groups.  
iii. Provide a basic analysis of differential abundance between CRC patients and healthy controls, focusing on indole-producing species.

**4. Approach**

This project involves loading publicly available microbiome data (e.g., an OTU/ASV table or taxonomic data) into R, filtering for indole-producing species using a pre-compiled list, and conducting basic analyses to compare the microbiomes of CRC patients versus healthy controls. We will use basic R packages like phyloseq for data manipulation and ggplot2 for visualizations. The analysis will include summary statistics, differential abundance testing, and exploratory visualizations, all presented in an accessible format for non-specialist audiences. The goal is to develop a script that could serve as a foundation for further analysis in the dissertation work, focusing on indole-producing microbes.

**5. Selected References**

* R Core Team. 2024. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
* McMurdie, P. J., & Holmes, S. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8(4), e61217.
* Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
* ChatGPT, OpenAI. 2024. Help with R script for microbiome analysis.