**Analysis Overview**

Colorectal Cancer (CRC) is the third most common cancer in the United States, with about 150,000 new cases annually (American Cancer Society). While the rates of late-onset (age>50) CRC (LOCRC) incidence has decreased substantially over the past decade, the opposite is true for early-onset (age<50) CRC (EOCRC) (Akimoto *et al*., 2021). With these alarming statistics, researchers and physicians have investigated causes for this increased incidence of EOCRC. The gut microbiome has been associated with CRC and other gastrointestinal diseases and has thus been the center of this research. Several studies have highlighted the importance of the gut microbiome in CRC initiation and progression (Pandey, *et al*., 2023). Specifically, disruptions in normal gut microbiota populations have been correlated with CRC risk (Zhao *et al*., 2021). Dysbiosis in the gut has been observed in cases where there is a decrease in the diversity of commensal bacteria (Rebersek, 2021). At the same time, the enrichment of pathogenic bacteria has also been associated with gut dysbiosis and CRC progression (Nezhadi *et al*., 2025). There is an urgent need to further understand diversity patterns of the gut microbiome (Fusco *et al*., 2024).

The code I developed seeks to make a way of analyzing the alpha diversity of CRC patients at various stages of disease progression. The code outputs two graphs; a Shannon Diversity (SD) plot and a Principal Component Analysis plot. The SD plot aims to visualize the gut microbial diversity of CRC patients at each stage (Stage I – Stage IVB). This is aimed at being able to visualize a general trend of microbial diversity along CRC progression, providing a visualization of diversity for each stage of CRC. This visualization is followed by a PCA plot, which has the goal of visualizing microbial diversity at the individual level (i.e. observing the differences of each patient’s gut microbiome). The dataset used for this project comes from The Cancer Microbiome Atlas, found on the dataset repository for Duke University, and I chose to analyze this dataset because it’s specific to my field and freely accessible.

**References**

Akimoto, N., Ugai, T., Zhong, R., Hamada, T., Fujiyoshi, K., Giannakis, M., Wu, K., Cao, Y., Ng, K., & Ogino, S. (2020). Rising incidence of early-onset colorectal cancer — a call to action. *Nature Reviews Clinical Oncology*, *18*(4). https://doi.org/10.1038/s41571-020-00445-1

American Cancer Society. (2023, January 13). *Colorectal Cancer Statistics | How Common Is Colorectal Cancer?* Www.cancer.org; American Cancer Society. <https://www.cancer.org/cancer/types/colon-rectal-cancer/about/key-statistics.html>

Dohlman, A., Arguijo Mendoza, D., Ding, S., Gao, M., Dressman, H., Iliev, I., Lipkin, S., & Shen, X. (2020). Data from: The cancer microbiome atlas (TCMA): A pan-cancer comparative analysis to distinguish organ-associated microbiota from contaminants. Duke Research Data Repository. <https://doi.org/10.7924/r4rn36833>

Fusco, W., Bricca, L., Kaitsas, F., Tartaglia, M. F., Venturini, I., Rugge, M., Gasbarrini, A., Cammarota, G., & Ianiro, G. (2024). Gut microbiota in colorectal cancer: From pathogenesis to clinic. *Best Practice & Research Clinical Gastroenterology*, *72*(101941), 101941. https://doi.org/10.1016/j.bpg.2024.101941

Nezhadi, J., Kafil, H. S., Sadrkabir, M., Mahdavi, F., Moaddab, S. Y., Nouri, R., Mohammadzadeh-Asl, Y., Sattarpour, S., & Rezaee, M. A. (2025). The relationship between pathogenic bacteria and different stages of colorectal cancer. *Letters in Applied Microbiology*, *78*(2). https://doi.org/10.1093/lambio/ovaf017

Pandey, H., Tang, D. W. T., Wong, S. H., & Lal, D. (2023). Gut Microbiota in Colorectal Cancer: Biological Role and Therapeutic Opportunities. *Cancers*, *15*(3), 866. https://doi.org/10.3390/cancers15030866

Rebersek, M. (2021). Gut Microbiome and Its Role in Colorectal Cancer. *BMC Cancer*, *21*(1). https://doi.org/10.1186/s12885-021-09054-2

Zhao, L., Cho, W. C., & Nicolls, M. R. (2021). Colorectal Cancer-Associated Microbiome Patterns and Signatures. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.787176